An Introduction to Spatial Analysis and Decision Assistance (SADA)

Environmental Applications for Version 5

User Guide

Robert Stewart
The Institute for Environmental Modeling
University of Tennessee, Knoxville, TN

Chris Welsh
The Institute for Environmental Modeling
University of Tennessee, Knoxville, TN

Tom Purucker
Gainesville College, Athens, GA
Athens, GA

Debra Stewart
The Institute for Environmental Modeling
University of Tennessee, Knoxville, TN

Table of Contents

This page left intentionally blank

Preface

Spatial Analysis and Decision Assistance is nearing its 15th year of development (2009) and distribution. During that time the code has been applied by a wide array of users to a wide array of applications. In order to assist users with their work, we have offered help in several ways. There has been classroom instruction, online help, a user's group, a constantly evolving help file, and personal contact with users. This user guide represents an explanation and ongoing discussion of SADA that has been tested and refined over time in all of these venues. Of course, we have always offered a user's guide for each version we've released. Previous user guides were essentially the contents of the topic-oriented help file arranged into chapters. As a result, they had little or no additional content.

This time, I decided to write the user guide from an application perspective rather than an enumeration of every feature SADA has to offer. From my own experience with help documentation, I have found it difficult to search through lists of topics to create an analysis that suits my needs. It requires that you know exactly what you're doing and exactly what the feature names mean. This is usually not the case. Such topic-oriented listings are important and the help file will always be a source for that type of learning.

In this text, applications drive the discussion, and features are introduced when they become relevant. If you are a veteran user and interested in a specific feature or model, then the help file will likely be the best choice for you. If you are interested in learning SADA through application and hands on guidance, then this user guide is probably best.

While we have many users across many different fields, environmental pollution assessment and remediation continues to be our focus. The text is therefore written within the language of environmental characterization and remediation. Many of the methods have more far-reaching application, and users from other walks of life can still benefit from this text.

There is a lot of emphasis on getting started with SADA. Such things as getting comfortable with the interface, importing your data, and just succeeding at some very basic things. Many users learn SADA on the job where there is never enough time. We wrote the chapters to begin as simple as possible and increase in complexity as the chapters progress. If you need to jump to a later chapter quickly and have never learned SADA, you should still focus as much as possible on the early chapters.

The writing style is very "folksy" in nature, using a lot of "you" and "we" throughout the discussion. The idea was to create a comfortable learning environment similar to our classroom setting (the technical editors really hate us). Hopefully this will help. Comments are always welcome at sada@tiem.utk.edu.

Robert Stewart

Project Lead

Part I: Overview

Chapter 1: Introduction

Background

Spatial Analysis and Decision Assistance (SADA) began as a grass roots effort in 1996 between the University of Tennessee and the Oak Ridge National Laboratory's Environmental Restoration program. The restoration program at ORNL was associated with the cleanup of legacy waste across the Oak Ridge Reservation from activities related to WWII and the Cold War arms race. The purpose of the effort was to develop tools that would integrate human health and ecological risk assessment with geospatial processes in a manner that could directly impact environmental restoration decisions. This would require the software to be generalized, easy to use, and incorporate the risk models necessary for ongoing cleanup operations.

In the late 1990s and early 2000s, the US EPA continued to support SADA as it grew in popularity. Shortly thereafter, SADA received support from the US Nuclear Regulatory Commission. Through each of these major sponsors, SADA has grown significantly with deployment of newer versions, national conferences, training at numerous facilities, classroom education, collaboration with experts and other institutions, and a continually growing body of documents, applications, and most importantly users. As of the development of this document, SADA has over 15,000 users in the US and abroad.

Through its many development stages, the authors have attempted to maintain the original principles of the project: everyday applicability and ease of use. This has resulted in the type of interface and body of documentation available to users today. Because of its generalized interface, the types of applications have extended beyond environmental remediation efforts into areas of civil engineering, geology, ecology, and many other unexpected fields. For this reason, SADA is also an excellent platform for developing new methods or utilizing existing methods in new ways, particularly within a spatial context.

The current version of SADA represents a collaborative effort between authors, sponsors, and users. To participate in this effort, please visit the SADA website (http://www.tiem.utk.edu/~sada/index.shtml) at the University of Tennessee, where you can learn how to join the SADA User's Group or directly communicate with developers via sada@tiem.utk.edu.

Here, we provide a broad laundry list of SADA's capabilities. Each of these capabilities can be used independently or in an integrated manner.

- Data Exploration and Visualization
- Geographic Information System
- Statistical Analysis
- Human Health Risk Assessment
- Ecological Risk Assessment
- Data Screening and Decision Criteria
- Geospatial Interpolation
- Uncertainty Analysis
- Decision Analysis
- Sample Designs

- MARSSIM module

While SADA was written within the context of environmental analysis, many of the processes were broadly constructed to deal with a wide array of problems concerning spatially distributed information. For simplicity, this manual will be written in the language of environmental assessment and will focus on example applications.

We will assume you are familiar with the Windows environment and comfortable with the topics covered in this manual. References are made to other resources outside of the manual for topics that may not be familiar to everyone. While some concepts may be discussed briefly, the manual is written primarily to show how to implement these concepts within the software.

A number of files have been included on the distribution disks to help you get started. They include comma-delimited data files (.csv), example SADA files (.sda), and some GIS layer files (.dxf). The data sets contained in these files are intended for demonstration purposes only; they were not originally sampled at the locations seen in the GIS overlays and have been altered. ToxicologicalDatabases.mdb and ScenarioDatabases.mdb are also included for setting up the human health risk module in SADA, while a benchmark database is included for ecological risk assessment setup and a custom database is included for the custom analysis setup. These file types will be discussed and demonstrated throughout the manual.

The general philosophy for how to use SADA can be described like this. You will create a SADA file. Most Windows applications operate on files. For example, Excel has Excel Files (.xls), Word has Word files (.doc) and SADA has SADA Files (.sda). Everything you need will be stored in your SADA file. When a SADA file is first created, it will not contain anything. Just like when you create a new Excel workbook or Word document, there will be nothing but a blank slate.

With your blank SADA file, you can begin to define the location of your site, maybe bring in some GIS layers, and maybe import some data. SADA has easy ways to bring all this information into the file. You will be able to see all these things in a simple graphics viewer. After you have your information in SADA, you can begin to perform some statistics or visualization. You can even run geospatial models like kriging, use the human health and ecological risk models, and begin to see how all these things can be used to support different kinds of decisions. When you've run your analysis, SADA keeps track of what you did so you can create instant documentation and maintain complete transparency in your work. There are also a number of ways that SADA can export your results to Excel, Word, or GIS systems such as ArcMap.

Before we get too far ahead, you need to install SADA.

Installing SADA

Before installing SADA, check for the following minimal requirements:

- CPU: Pentium IV
- Disk Space: 400 MB
- RAM: 500 MB
- OS: 2000, XP, VISTA
- Clock: 1GHz

The following requirements, however, are recommended to enhance performance of geospatial analyses:

- CPU: Pentium IV or higher
- Disk Space: 1 GB
- RAM: 1GB or higher
- OS: 2000, XP, VISTA
- Clock: 2 GHz

Very important: you must be logged in as an administrator on your computer when installing SADA. If you are not the administrator, then SADA will not properly install or run. If you have no login prompt on your computer, the chances are good that you are in fact the administrator. However, if you are not certain, ask your IT folks.

Open up your web browser and visit the SADA website (http://www.tiem.utk.edu/~sada/download.shtml) to download the latest copy of SADA. You will be asked to fill out a simple questionnaire. This questionnaire does not check information, so if you feel uncomfortable providing any information, you can just enter nonsensical information into the window and it will accept it. This information is used only to help us understand the location and interests of our users. This helps us continue to receive sponsorship and to provide SADA as a free product.

Save the downloaded file to your desktop.

Run the installation setup.

You might be asked to install the .NET framework. This is required by SADA and is a normal part of the installation process.

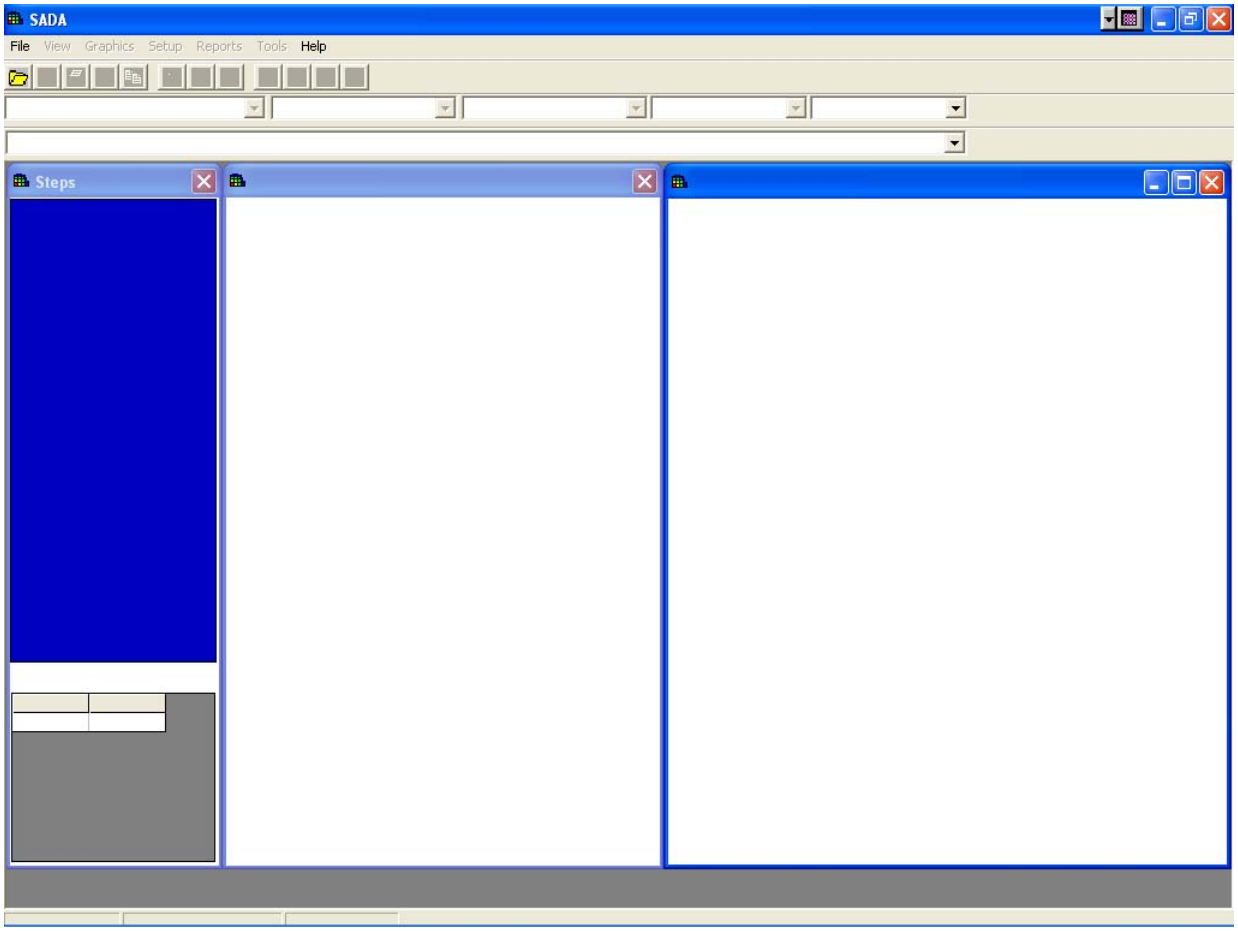
Follow the remaining installation instructions.

You can install SADA on as many machines as you like. Be aware though that SADA will not install correctly when installed over a network or when run over a network. For example, you cannot install SADA on a single network machine and allow users to run it remotely. It was never designed with this in mind and serious errors are likely to occur.

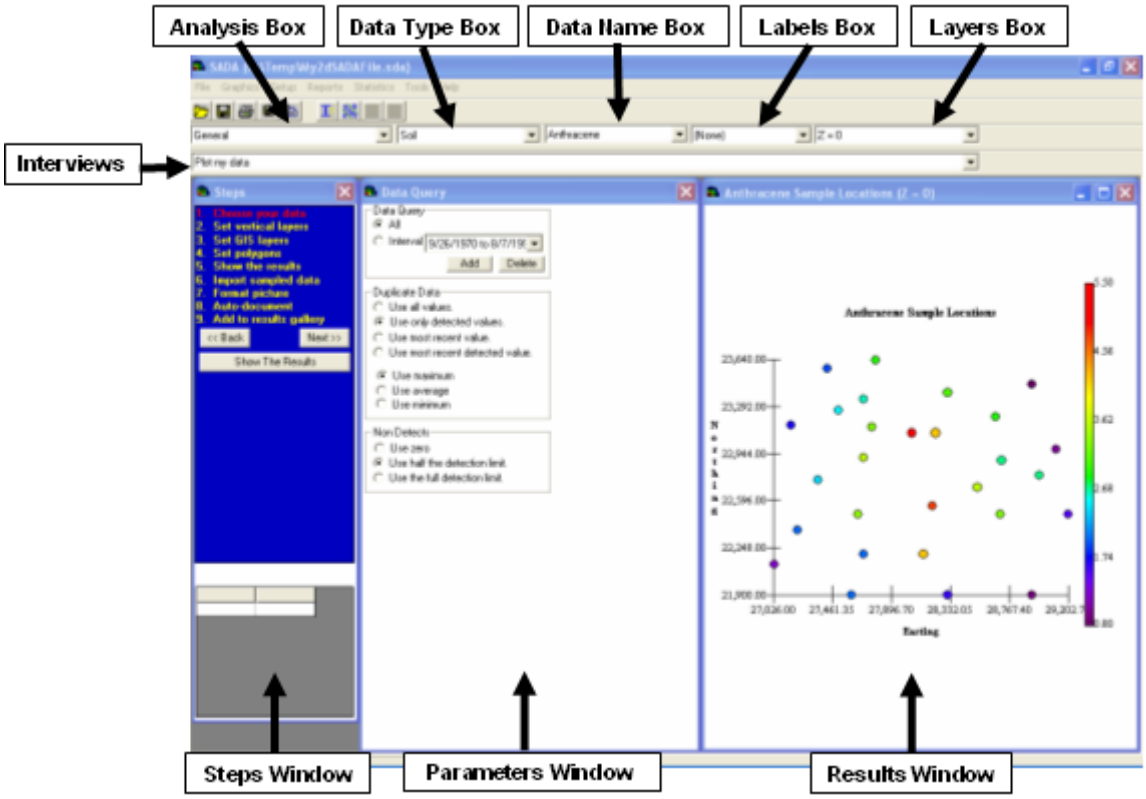
SADA Layout

Before doing anything in SADA, let's take a moment to understand the main components that make up the software: list boxes, interviews, windows, buttons, and menus. When SADA is first opened, a welcome screen will appear. To do this, we need to first open SADA and then open a SADA file. Start SADA by clicking on your Start button, then program files, then SADA 5.0 and then finally SADA. When SADA opens, close the little welcome screen, and from the main menu choose *File→Open*.

Behind the welcome screen are three empty windows. Only the *Open* button and the *File* and *Help* menus will be enabled.



After a SADA file is created, the menus and buttons will be enabled and the windows and list boxes will display the corresponding data. In a moment, we will create a SADA file, but for now, just keep reading. In the following image, the file ThreeDimensional.sda, which is distributed with SADA, has been opened (any file name with a .sda extension is a SADA file).

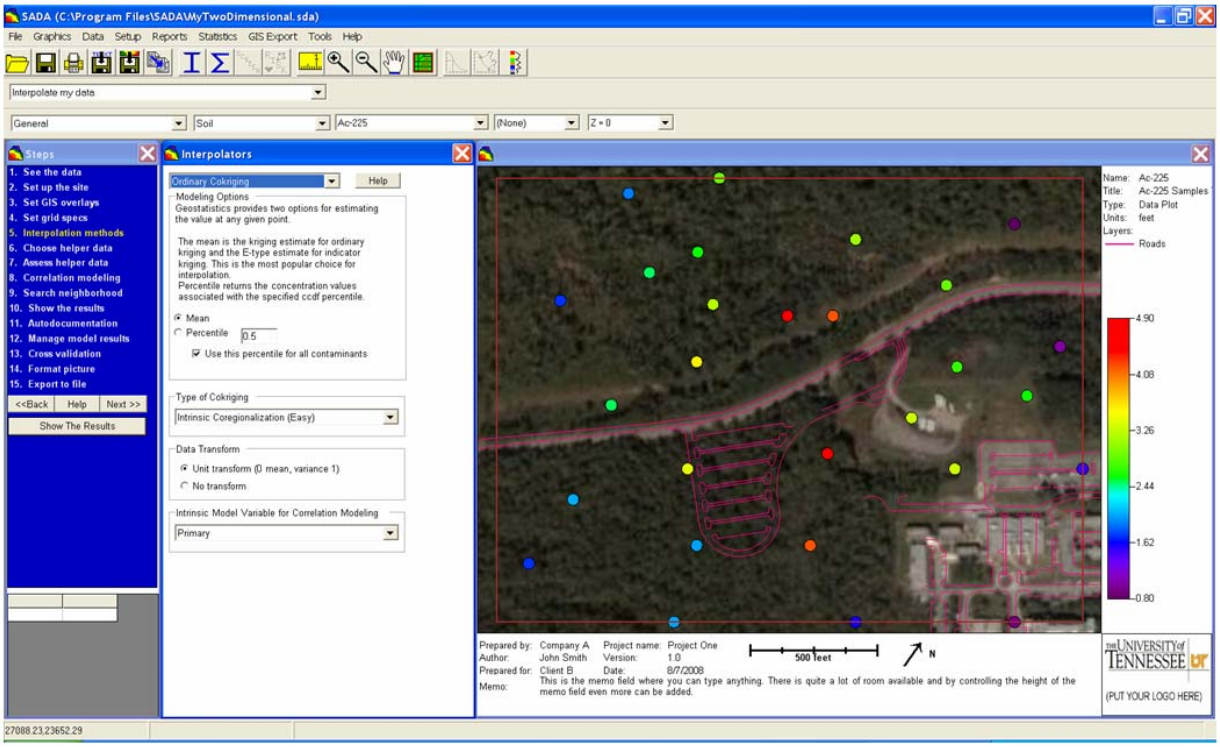


The design for SADA's interface attempts to accomplish the following:

- Be as intuitive as possible.
- Maintain the same look and feel regardless of the modeling or assessment under consideration.
- In a seeming contrast to the second objective, make the interface conform itself to the interests and needs of the user at any given time. In particular, hide all but the most relevant functions and options for any given modeling effort.
- Provide the user with some interface guidance on how to accomplish things but not force them into a strict interview type format.

Let's take a look at the interface and see how these were accomplished.

Navigate to the SADA installation directory and open the file TwoDimensional.sda. Any file ending with the suffix (.sda) is a SADA file and contains virtually everything you'll need to perform your analysis. SADA will respond by opening the SADA file and showing you the following (your graphics on the right side might look different, but that's okay).



At first glance, there seem to be a lot of things to consider in the interface; however, the interface will remain in this same format regardless of the types of analyses and modeling that you may do. Let's begin by dividing the interface into its major parts.

SECTION A

This section is the starting point for how the interface is controlled. It is comprised of:

Interview Box

This box contains all the major things you can "do" with SADA written in "plain English". These are referred to as interviews largely because each of them will produce a set of steps or questions that you can go through (see next section). The number and type of interviews that appear in this box depends on what particular analysis you might be interested in or what type of data you are currently examining. For example, the interview "Contour Risk" only appears as an option when you are doing a risk analysis. Another example is the interview that includes map smoothing, which only appears when you are examining a spatial model. In this way, SADA "hides" unnecessary functions or invalid functions in the interface, and the interface dynamically adapts itself to your interests.

Analysis Box

This box displays the type of analysis (e.g., *General, Human Health, Ecological*, or *Custom*). For a new file, only *General* is available. *General* is the default choice and simply means that the data will be used in a straightforward way without applying any other framework or models, such as risk. Later, you'll learn how to set up human health risk, ecological risk, MARSSIM, and even your own custom analysis. After selecting an analysis, the interviews will change to display the appropriate actions for that analysis, and the windows will change to reflect the data.

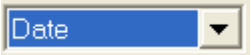
Data Type Box

This box has been known by previous SADA users as the media box. It had this name originally because SADA could only deal with one type of data, sampled point values, where media was often specified. Media refers to soil, surface water, sediment, or groundwater. SADA, however, can now deal with a number of data types and so the box name had to be generalized. For example, elevation data can now be imported. Therefore, this box is now called Data Type instead of media. SADA organizes and presents your data by its data type.

Data Set Box

This box contains each data set found under the selected Data Type (see above). Historically, this was called the analyte box because SADA was explained only in the language of environmental contaminant measurements (even though it was applied in many different kinds of applications). Now, even within an environmental context, we'll have to let go of the term analyte box since SADA includes imported models, user-defined models, and elevation maps. In this example, we can see we have a data set of Ac-225 data, and we know it's a soil type data set because that is our selection in the data type box shown above.

Label Box



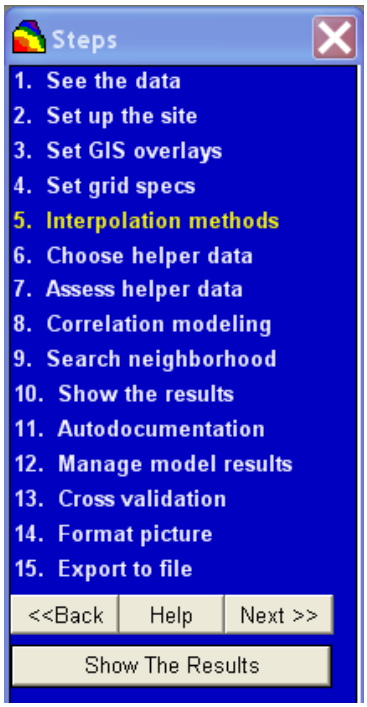
For the measured sample or point type data sets, users may import additional metadata with their core required values. This label box allows you to see these metadata plotted alongside your point values in the map. If any other data type is selected, it will be disabled.

Layer Box

SADA is actually a 3d tool in which you can do 2d analyses. So, all spatial information has a vertical component (even if it is 2d) that must be addressed. SADA deals with the vertical direction by allowing you to define a vertical layering scheme. This layering scheme is then used in modeling and visualization alike. In this layer box, you can control which layer you are looking at in the graphics window. You will learn a great deal about vertical layering later.

SECTION B – INTERVIEW STEPS

This is the interview steps section. When you select an interview in the interview box, SADA responds by providing necessary steps to accomplish the objective as well as a number of potentially interesting or helpful steps that might also be used.

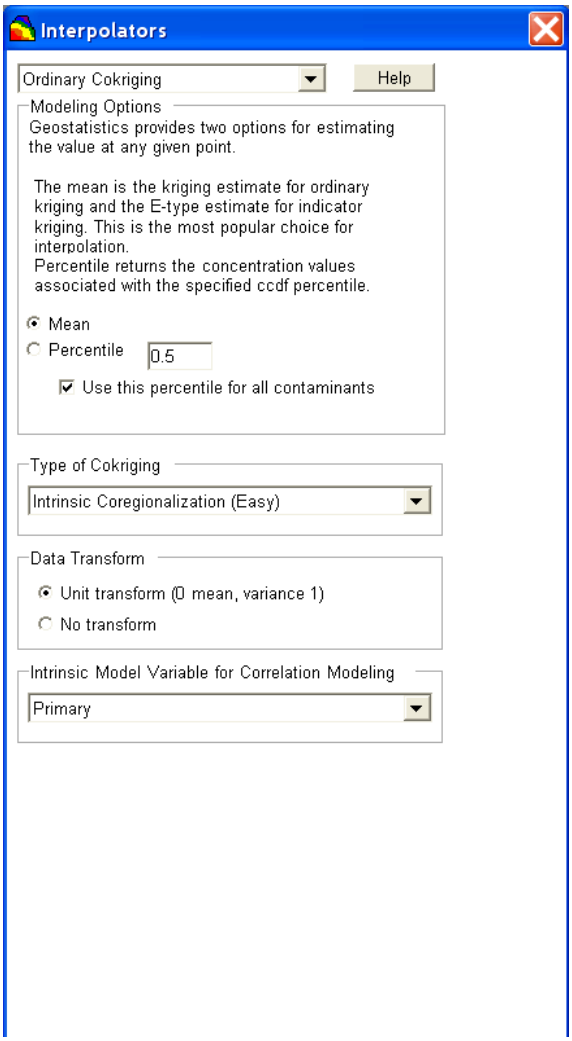


When you click on any given step, the parameter window (Section C) changes to display questions or options related to that particular step. The relationship does run both ways. Certain answers or options in a parameter step may affect which additional steps may be needed. For example, if you select ordinary kriging (a geostatistical method), the interview steps expand to include correlation modeling (an interim step important to ordinary kriging).

SADA will attempt to carry out the objective of the interview when you visit the *Show The Results* step or press the *Show The Results* button.

SECTION C – PARAMETER WINDOW

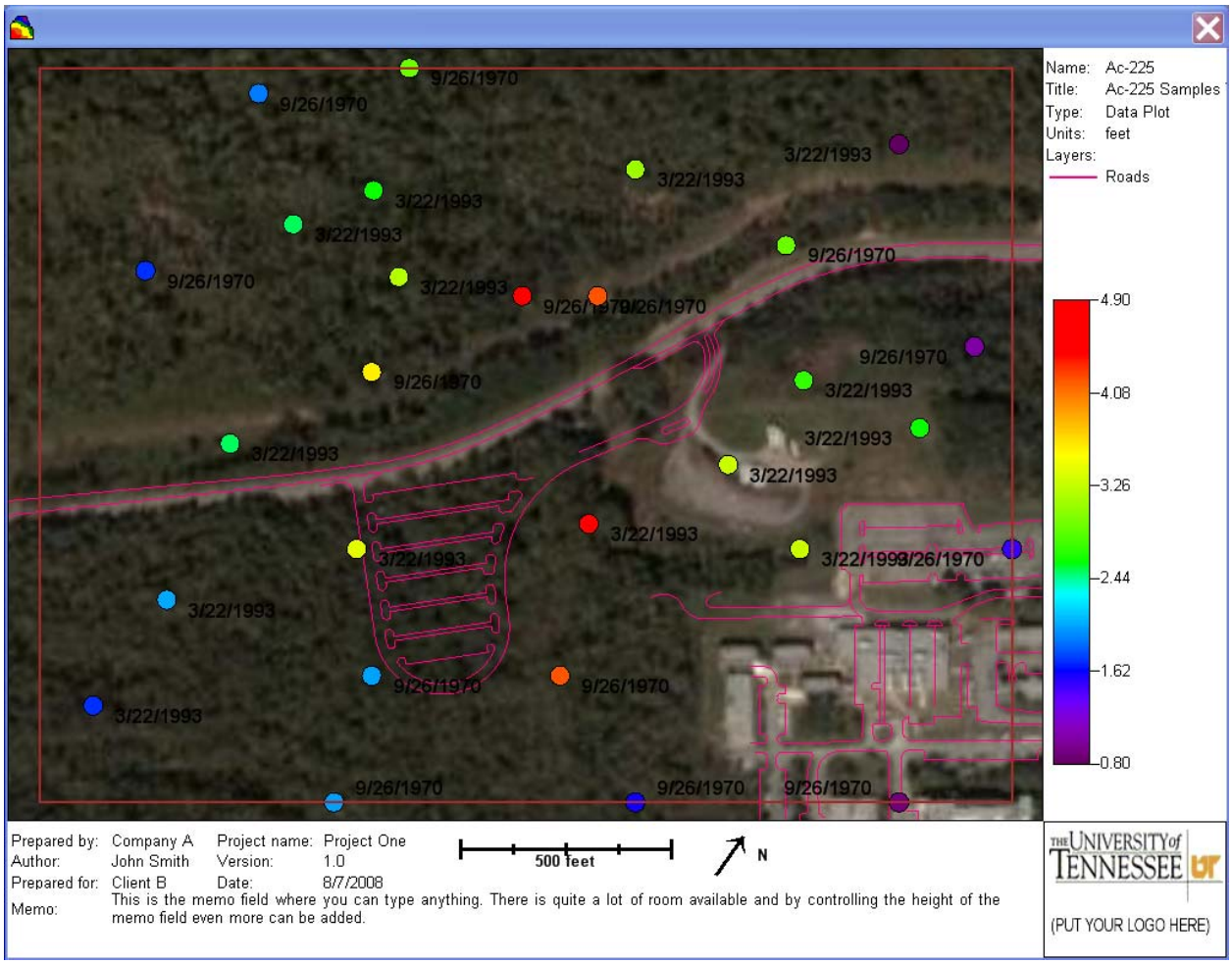
The parameter window's contents are dependent on the currently selected step. The following image displays the parameter window for the *Interpolation methods* step (when *Ordinary Cokriging* is selected in top drop-down list).



The parameters for each step are discussed under the help topic associated with specific steps.

SECTION D – GRAPHICS VIEWER

Virtually all graphical results will be shown in this window. The graphics viewer is a very powerful tool and discussion of its features is handled throughout the help file.



SECTION E – MENUS AND TOOLBAR BUTTONS

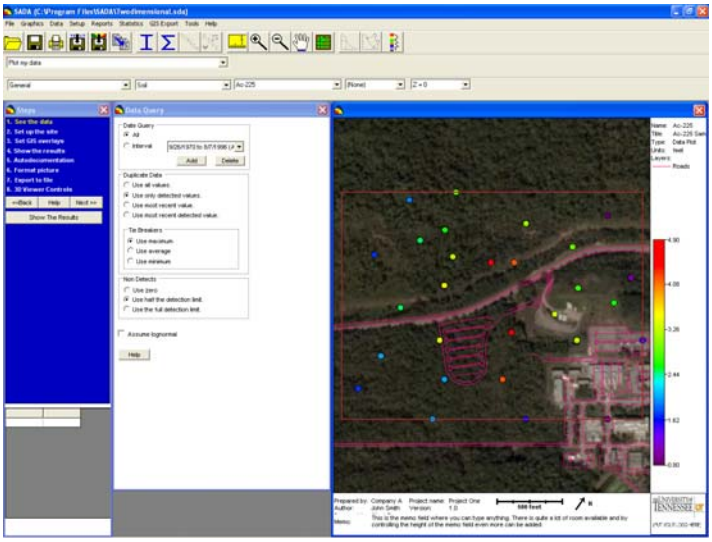
This section is comprised of the menus and toolbar buttons. The menu section in particular changes with the type of analysis as well as data type.

Each of these interviews, toolbar buttons, menus and so forth will be introduced as we go along. For a quick reference, we recommend you use the SADA help file. The *Help* file is organized around topics (e.g., *Tools* menu) whereas this user's guide is organized more around applications (e.g., determining where to place new samples).

See Chapter 2, "A Quick Tour," for a quick flyover of SADA.

Chapter 2: A Quick Tour

We will now take a quick tour of SADA, showing you very quickly some of the major features and lines of thought people have often taken when using the code. This will familiarize you with what SADA can do and also give you some practice moving around in the software. Each of these features will be taken up in great detail later. If you haven't already done so, open SADA and open the file Twodimensional.sda (*File→Open*). It should be located where SADA was installed (probably c:\Program Files\SADA 5).



Virtually all spatial data exploration, spatial modeling, and any results that are based on spatial outcomes are organized and made available to you through interviews. SADA "interviews" you to acquire the information it needs from you and then carries out the objective you had in mind. You are "interviewed" through a series of logical steps one would follow in order to accomplish some task. Each step in the process presents you with questions or options or both that you will need to answer as you go along. Some steps may actually run small interim models that are needed prior to the final result. Let's run through a couple of interviews to get a feel for what each entails.

Suppose you have some environmental data, measurements taken in the soil on a particular contaminant of interest. You would like to interpolate these data to create a continuous map of contaminant concentrations. There are a series of questions you need to answer to accomplish this:

- What data are you going to use?
- If your data are 3d, what type of vertical layers do you want in the interpolation?
- Do you want to focus on specific areas of the site?
- What type of grid would you like to use in interpolating?
- Which method of interpolation do you prefer?

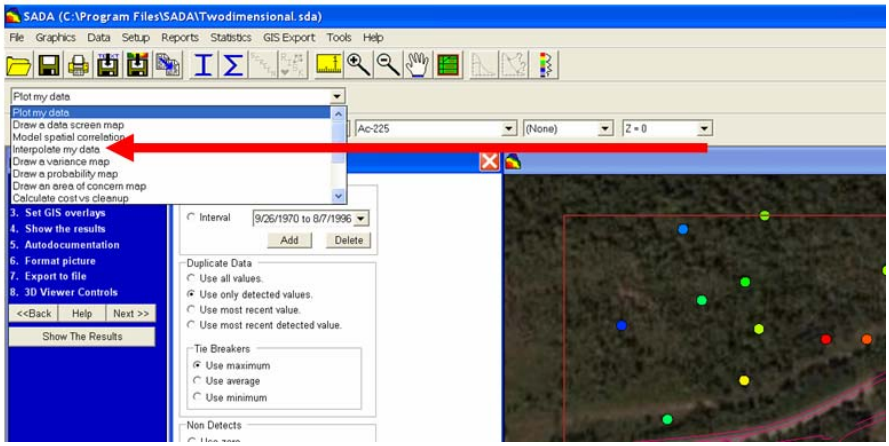
After you have generated an interpolation map, there may be some follow up questions that are helpful, such as:

- Would you like to format your picture?
- Do you want to store these results?

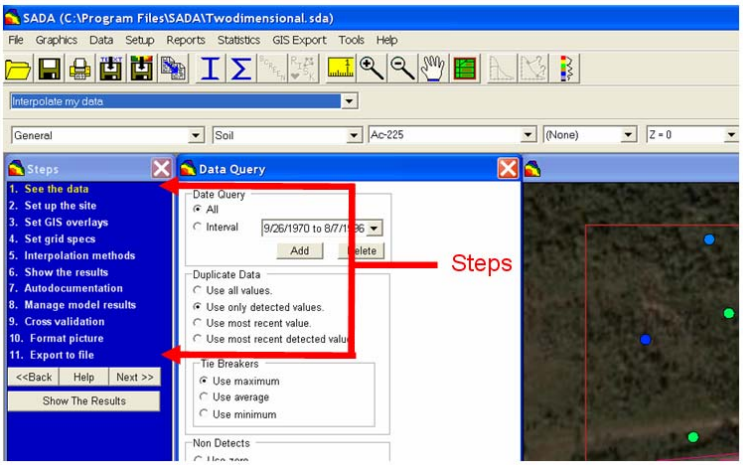
- Do you want to make a quick report explaining how you got this result?

Let's work through this using Twodimensional.sda as an example.

Click on the interview box down arrow and select *Interpolate my data*.



In the steps window (blue box) is a series of steps to go through to interpolate your data. Look them over and see what types of questions correspond to what steps.

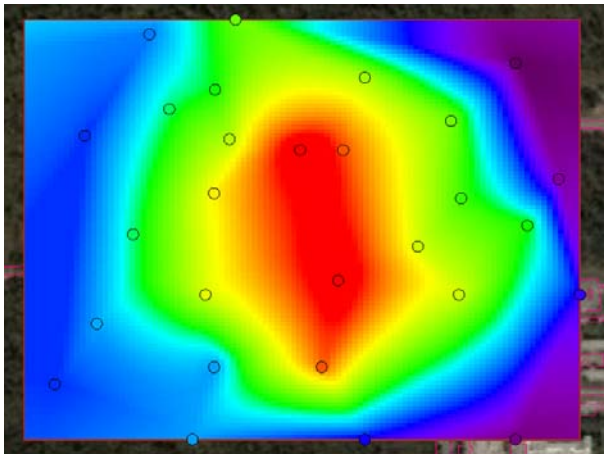


Click on each step and you will see the parameters window change to reflect the types of information the step is asking for:

1. *See the data* (What data are you going to use?)
2. *Set up the site* (What kind of vertical layers might you have? Do you want to focus on any particular area in the interpolation?)
3. *Set GIS overlays* (Do you have GIS layers like roads or photos to add?)
4. *Set Grid specs* (What type of grid do you want to use? – grids are the backbone of any interpolation.)
5. *Interpolation methods* (What type of interpolation method do you prefer?)
6. *Show the results* (Run the interpolation. Really a command more than a question.)

This first set of steps really leads up to the function. The *Show the results* step represents a break between what you have to do and what you might like to do after the fact. The steps that follow afterward are “after the fact” steps you might want to consider.

For our example, before you press the *Show The Results* button, make sure you have selected *Natural Neighbor* in the *Interpolation methods* step. This is the easiest interpolation (estimation) method in SADA. To do this, click on the step 5, Interpolation methods. In the drop-down window, switch to *Natural Neighbor*. Press the *Show The Results* button or click on the *Show the results* step (they do the same thing). SADA completes the interpolation, turning the discrete data map into a continuous data map that should look something like this:



You’ve just created your first geospatial map!

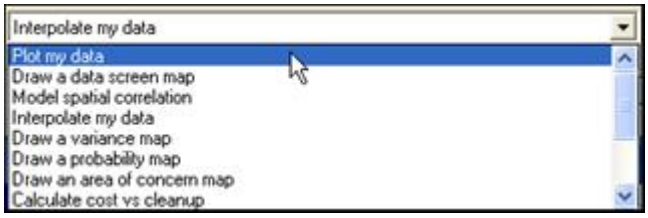
Each interview has its own series of steps. The first two or three steps are the same for many interviews (i.e., *See the data*, *Set up the site*, *Set GIS overlays*). SADA remembers the answers to these, so you don’t have to revisit them every time you switch to a new interview (unless you need to make a change). Some questions SADA asks ARE NOT dependent on any particular data set or interview. For example, if you import a GIS layer called Roads.dxf, you probably want to use it no matter what kind of spatial modeling you’re doing. SADA recognizes this and shows the roads layer for all interviews until you say otherwise.

Some questions SADA will ask you ARE dependent on the particular data set you are using. For example, suppose you want to compare your data set with specified limits established by the EPA. These limits are different for each contaminant, so you’ll need to answer this question at least once for each contaminant individually. (Chapter 5, “Importing Sampled Data,” discusses how SADA identifies separate contaminants.) SADA recognizes these types of questions and answers, and does not assume the same answer for every data set.

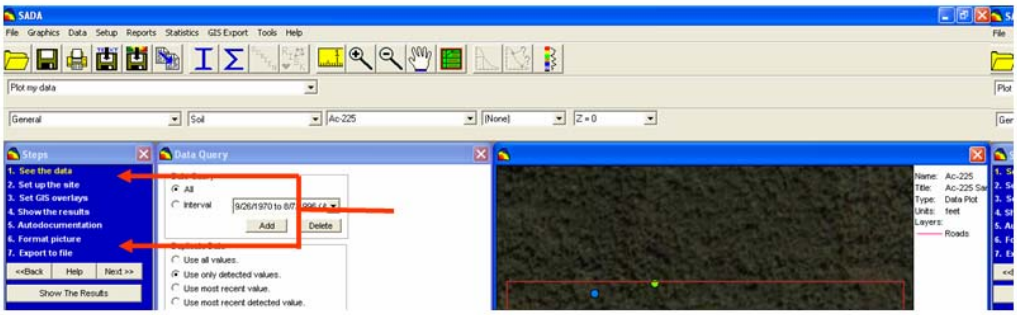
Now let’s look at the “after the fact” steps:

7. *Autodocumentation* (Would you like to make a quick report that helps explain this picture?)
8. *Manage model results* (Want to store this picture?)
9. *Cross validation* (Want to see how well the model did?)
10. *Format picture* (Want to pretty the picture up?)
11. *Export to file* (Want to dump this out into a file another system might use?)

We've just skimmed through one interview on how to interpolate data to get a continuous data map. Now let's go through another interview called *Plot my data*. This is one of the more basic interviews, and it is the default interview after you import sampled data. As we go through, we will look at features of three steps that are common to many interviews and important when setting up your site. Let's begin by selecting the *Plot my data* interview from the Interview Box

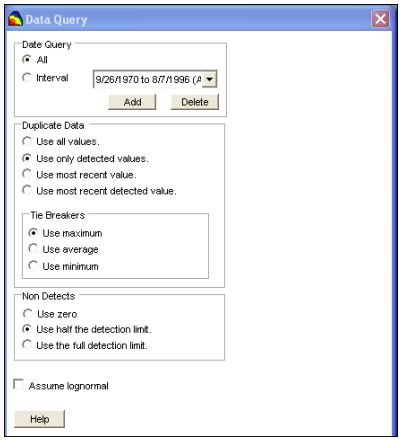


Notice that many of the steps for this interview are the same as those for the *Interpolate my data* interview.



Let's start with the first step and go through them in a little more detail.

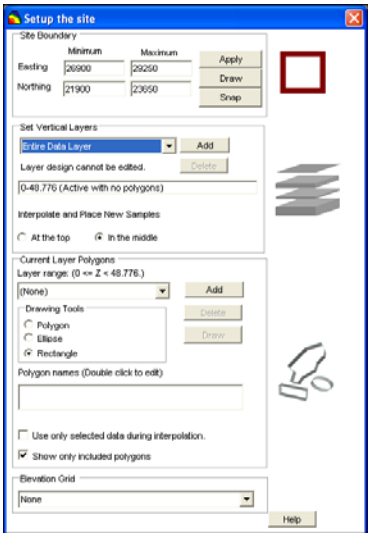
See the data gives you control over the data with which you are working. The parameter window is divided into sections. *Date Query* lets you choose all data or a subset with a given date range. The *Duplicate Data* section lets you decide how to deal with data from duplicate samples. Finally, the section on *Non-Detects* lets you specify the value to be used when the measured value is below detection limits. Essentially, this step defines how SADA will query and resolve your sample results for the data set with which you are working. In this case, we are working with *Ac-225* measurements in the *Soil*.



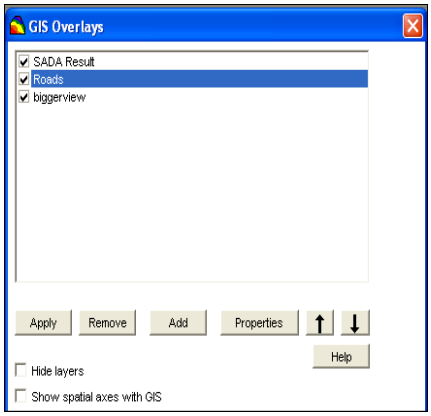
Set up the site is where you will define the horizontal and vertical extents of your site. You can set vertical layers, set site boundaries, and use different shape features to include or exclude parts of your study area. In particular, the site boundary section allows you to specify the extent

of the site based on coordinates, by manually drawing the boundaries on the map, or simply by snapping the boundaries to the limits of the data. The section *Set Vertical Layers* lets you specify various layer designs to be used in your analyses. In other words, you can define how you will divide up the subsurface. SADA is both a 2d and 3d tool. If all you have is 2d data, then simply specify surface only, for example.

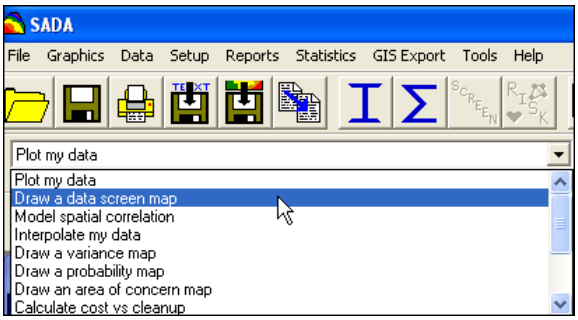
The polygon section lets you spatially aggregate subsets of your data for analysis by drawing features directly on the data map. These features can be rectangles, ellipses, or polygons, and they can be used to define irregular site boundaries as well as focus in on certain areas of interest while excluding others.



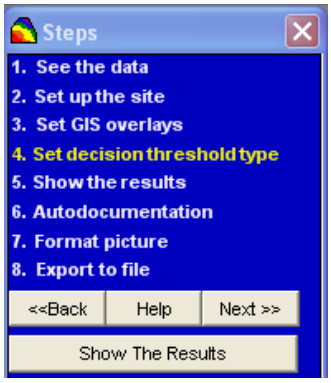
Set GIS overlays allows you to pull in GIS layers to display with your data. In this example, we've imported layers with roads at the site and an aerial photograph. You control how and whether the layers are displayed. You can hide all layers or select or deselect individual layers by clicking the associated boxes. The *Properties* button lets you change layer colors and draw widths.

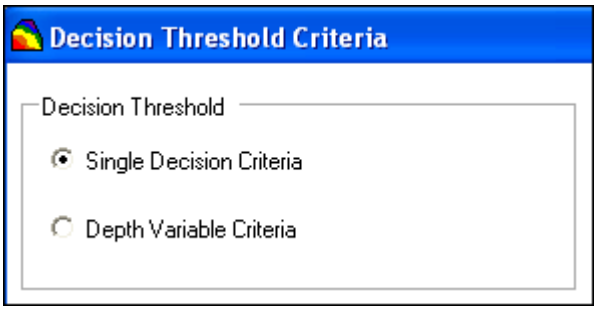


We've taken a quick look at the *Interpolate my data* and *Plot my data* interviews. Now, suppose you want to see which samples from your site exceed some threshold level. The *Draw a data screen map* interview is designed to give you a quick route to visualize both the number and distribution of exceedances at your site. To see how it works, select *Draw a data screen map* from the interview list.



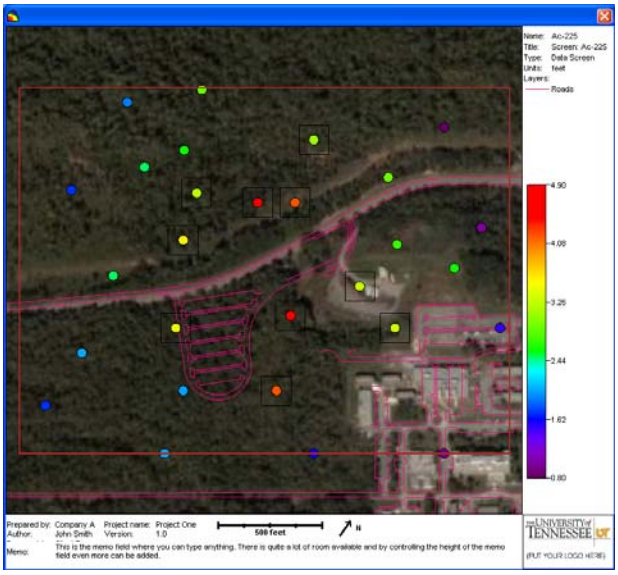
Again, notice many steps here are the same as for *Plot my data*. We don't need to revisit the set up steps, but step 4, *Set decision threshold type* is important (more on this later). Select *Set decision threshold type* and the *Decision Threshold* choices will pop up.





Click the *Show the Results* button and SADA will ask you to specify a decision goal. Here our goal is 3pCi/g. We got this number from some outside source. It could have been calculated outside SADA, or it could be part of some regulatory guidance. At any rate, this is the screening value. SADA has more sophisticated ways to import screening criteria or calculate it for you. You'll learn more about these later in Chapter 14.

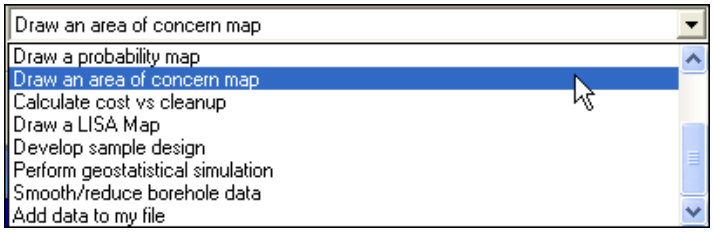
Click *OK* and a box appears telling you 10 stations exceeded the decision goal. Clicking OK again results in a data map with boxes that mark stations with values >3.



Just a note about units: SADA does not force or expect any particular type of unit for measured values. You are responsible for making sure all your data is in the units you expect. The only exception for SADA is when you are doing human health risk or ecological risk. In those cases, SADA assumes your measured values are in pCi/g, mg/kg, mg/L etc. See the risk chapters (Chapters 17-27) for greater detail.

You might notice that the black boxes around those points exceeding 3pCi/g are difficult to see with the dark photograph in the background. This is easy to fix. From the main menu, select Graphics→Set Various Colors→Screening Box. Choose orange as the color. You should see the boxes more clearly now.

Suppose you want to see not just which stations exceed a criterion but what area on the map would be predicted to exceed the criterion. SADA combines the concepts behind the *Interpolate my data* interview and the *Draw a data screen map* interview into the *Draw an area of concern map* interview. An area of concern can be defined in different ways, but the underlying meaning is consistent. An area of concern is a spatially defined portion of the site that is causing a concern because it is in exceedance of a decision criterion or is significantly contributing to a decision. For environmental applications, this is usually viewed negatively, as it represents an area that might require some sort of remediation or intervention. For mining applications, this might represent a lucrative body of ore. Either way, it represents an area of interest to the investigator.

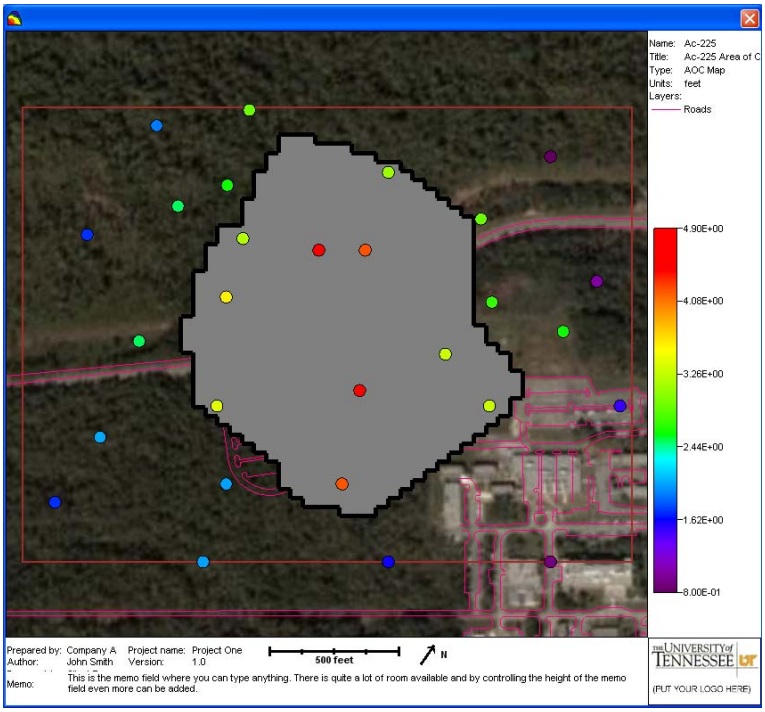


Draw an area of concern map contours the data and then shows the region on the map that exceeds the criterion. From an environmental viewpoint, the result is an area of concern map that can be used in setting remediation boundaries.

Let's take a look at the steps available for this interview. They should look familiar since they contain many of the steps already seen for *Interpolate my data* and *Draw a data screen map*.

Some important new steps appear between steps 4 and 7. At the fourth step, we have a question about the grid specification. You'll learn later that contouring algorithms rely on specifying a grid that is stretched across the study area. Then, the center of each cell is estimated, forming a continuous map. You can make the cells as small or as large as you like, and there are different motivations for choosing one cell size over another.

Click on *Set grid specs*. You can see that we've set grid specs at default of 50 cells x 50 cells. Click on the *Interpolation methods* step. Make sure this is set to *Natural Neighbor* again. Click on *Specify decision criteria*. This parameter window contains quite a few options for defining decision maps. We'll accept the default choices there and continue to the step *Set decision threshold type*. In SADA 5, you can choose a single decision criterion or specify the decision criterion as a function of depth. In our case, we will choose to go with a single decision criterion. Press the *Show The Results* button and enter a value of 3 when prompted.



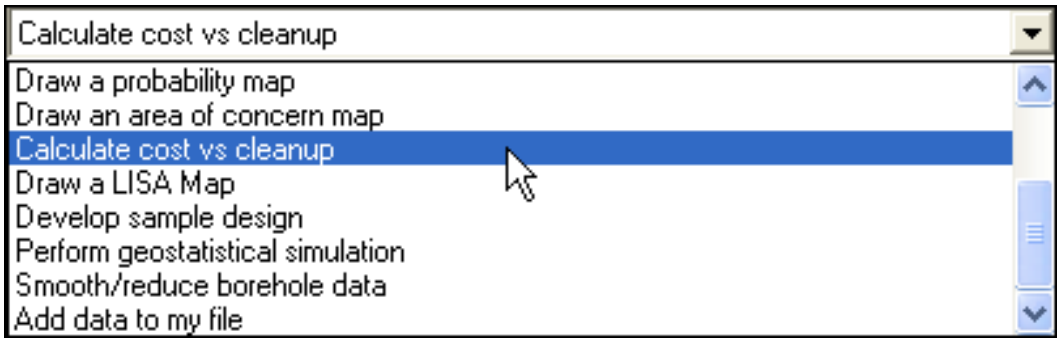
The area in gray represents the area of concern. The thick black line represents the current boundary on the area of concern. Later, you will learn how to visualize uncertainties about the boundary on the area of concern.

SADA also computes a volume estimate for the area of concern. This appears at the bottom of the screen when the area of concern map appears.

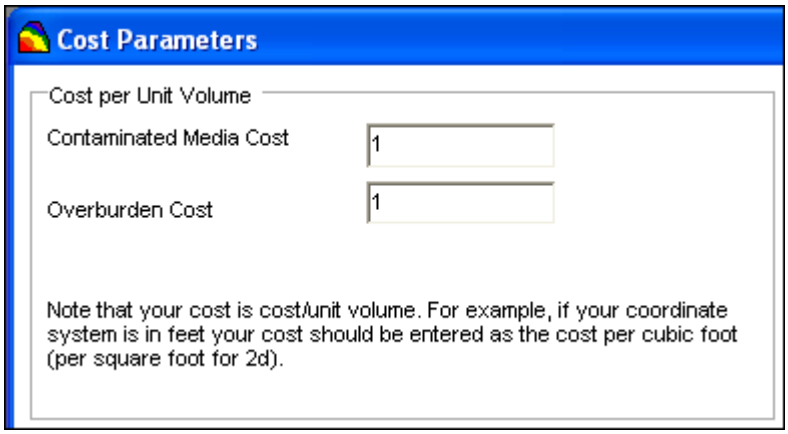
Clean 782 blocks. Volume = 1286390 cubic (square for 2d) units. Mass = 1286390

The number of blocks to clean for a block scale analysis is the number of grid cells estimated to exceed the decision goal. Volume is computed based on the units used to plot the data (i.e., cubic feet, cubic meters, etc., or squared units if the data layer does not include depth). Mass is computed using assumptions on the density of the material to be removed. In this example, mass = 1.

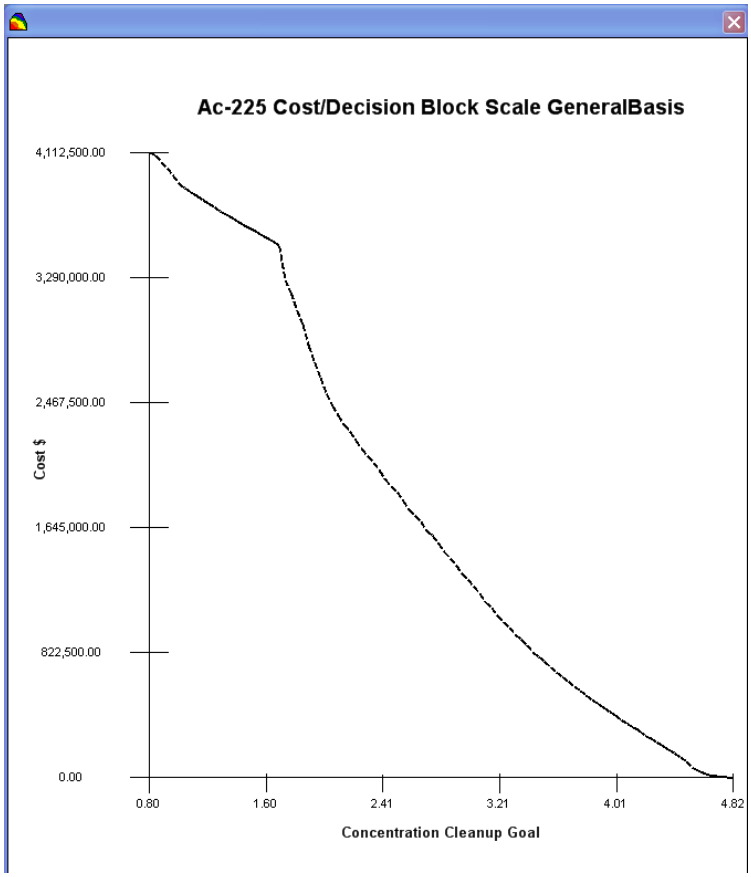
So now we've estimated an area of concern based on a decision goal of 3 pCi/g. What are the costs of cleaning up to this level? How would cost change if the goal were raised or lowered? The *Calculate cost vs cleanup* interview provides you with the means to answer these questions. Select this interview now.



The only step we haven't seen yet is step 7, *Set cost information*. This step allows you to set the cost/unit volume (i.e., $/cubic foot) of removing the contaminated media and of removing uncontaminated overburden.

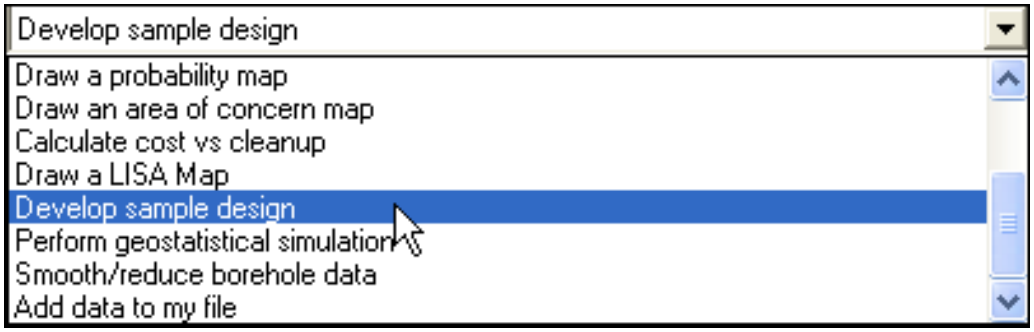


Press *Show The Results* button and you will see a cost-benefit graph like the following.

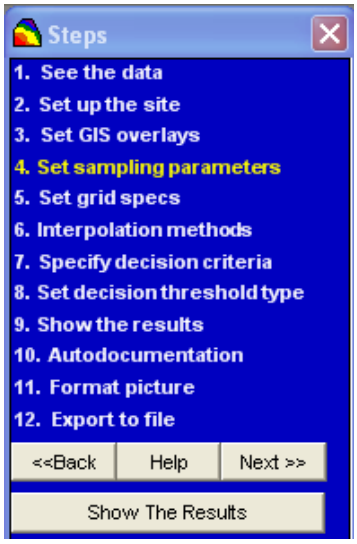


This graph is very useful in showing how small changes in the decision criteria can result in large changes in cost. For example, in this plot, the cost changes very dramatically between 1.8 pCi/g and about 2 pCi/g. Later, in the chapter on assessing cost (Chapter 36), you will learn about some tools for interpreting and reading this graph.

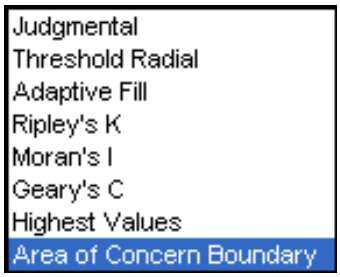
Let's continue to follow this problem out. We've interpolated the data, determined an area of concern, estimated volume and cost of clean up. One thing we might want to do is take some samples to improve the area of concern calculation. The *Develop sample design* interview assists you in deciding where to place additional samples to get the most "bang for your buck." From the interview box, select *Develop sample design*.



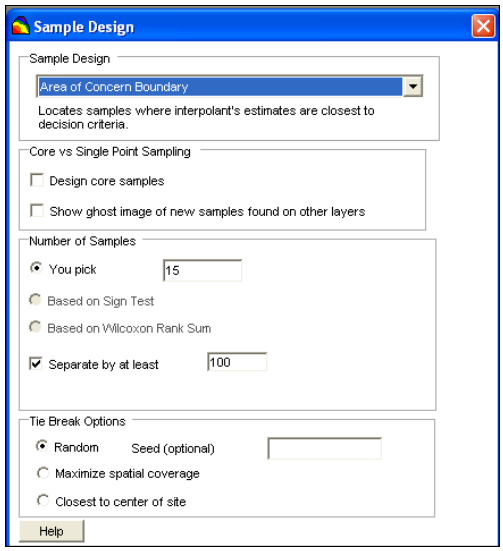
Looking at the list of steps for this interview, we see *Set sampling parameters* is the only new one. This allows you to select the sampling design you would like to use.



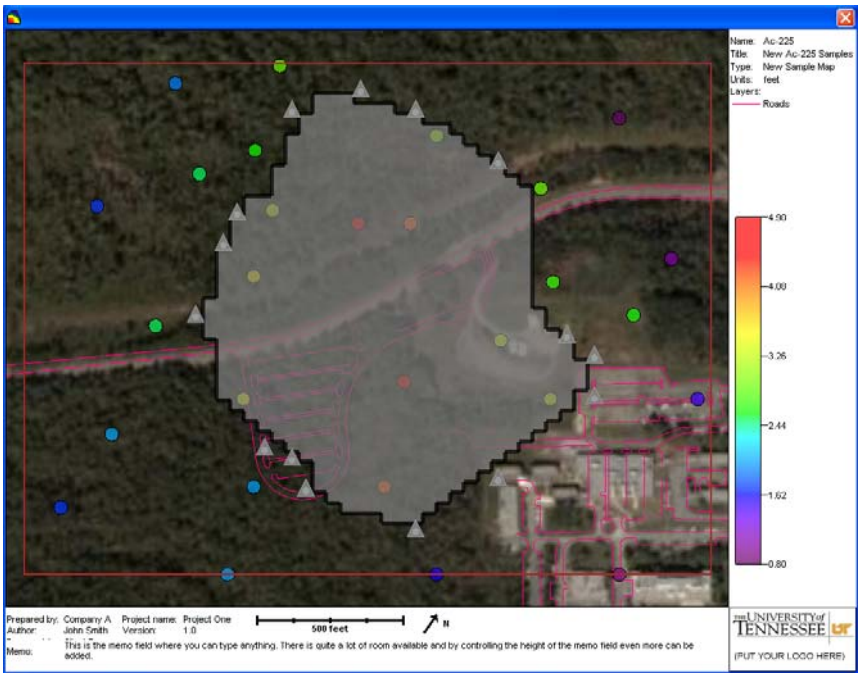
In this example, we want to improve our estimates around the area of concern boundary, so we'll select that as our sample design and fill in the appropriate parameters.



The parameters window allows you to specify features that will control where the samples are placed, whether to design core samples or single point samples, the number and spacing of samples, and how to decide between two locations when they are tied for the next best place to sample.



Enter 15 for the number of samples. Make sure you've selected *Separate by at least* and enter a value of 100 for this option (more on this in Chapters 37-39). Press the *Show The Result* button, enter a value of 3 for your decision criteria and you should see a result similar to the following image.



The graphic now shows triangular shapes where new samples would optimally be located to better bound or investigate this area of concern. Notice they are not necessarily uniform. This is because the model needs more information in certain areas than in others.

There are many, many other features that SADA provides not included here in this flyover. Hopefully by now you are starting to feel more comfortable switching between different interviews and perusing through some steps. In the next chapter, we'll start from the beginning and closely investigate what all this really means.

If you wish to take a break, you can close SADA now by selecting *File→Close*. Select *No* if asked to save. If you are continuing, then you can keep SADA open.

Chapter 3: Understanding SADA Files

Like many other Windows software packages, SADA creates and uses SADA files denoted by the suffix .sda. With few exceptions, everything is stored in a SADA file. You can then easily pass this file to other folks who can open it and see the work you've done. Any data you wish to use is imported into the file (model parameters, risk parameters, decision parameters, and so forth). Virtually everything that you create, import, or calculate can be found in your SADA file.

When you install SADA, a handful of SADA files are included in the installation to help you get started using the code. Let's start by opening a SADA file and see what it looks like. If you do not have SADA open or the twodimensional.sda file open, please close SADA and follow the directions.

Start SADA (e.g., *Start→Program Files→SADA 5→SADA5* or you may have the shortcut on your desktop). The SADA icon is as follows

The first time you open SADA you will be presented with the welcome screen.

If you do not see this screen, continue reading to find out what you should do next. If you do see this screen, then it is easy to create a new SADA file or open an existing file. Finally, if you do not like this welcome screen to show up each time, select the "Don't show me this again option" and press *Close* (don't do this for now).

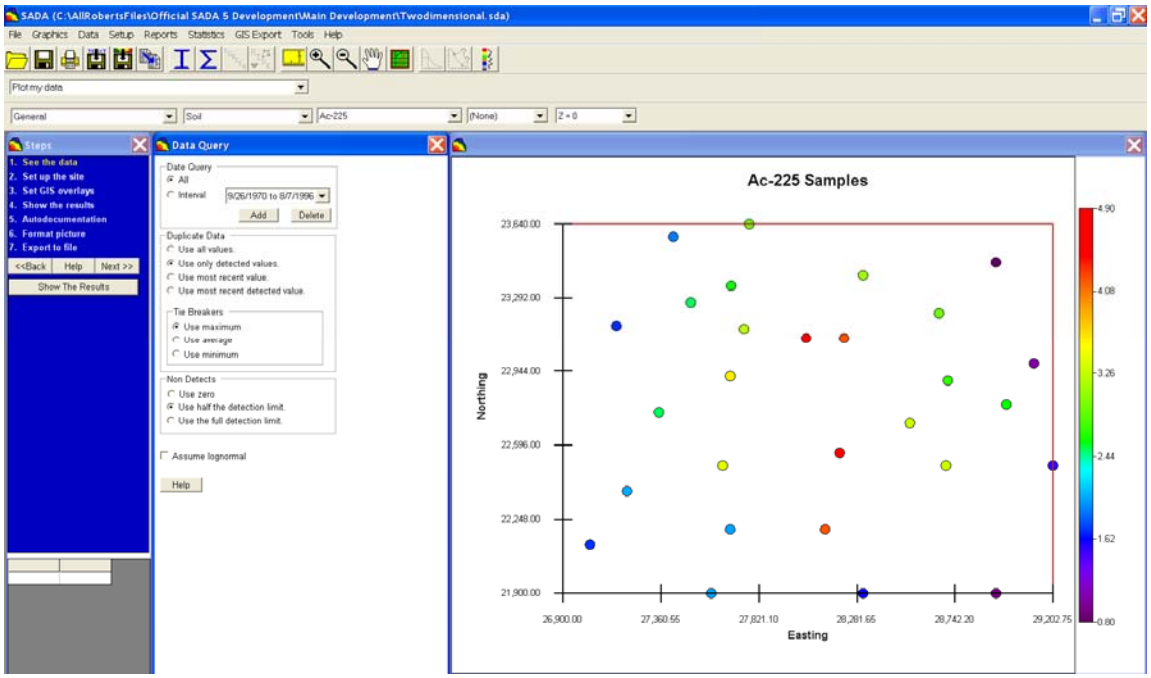
For now, let's open the file Twodimensional.sda. Select the *Open existing file* option. Use the Browse button to navigate to the Twodimensional.sda file location. The file should be located in the same directory where SADA was installed. If you are unable to find your installation files, you can access them again by visiting the SADA website (http://www.tiem.utk.edu/~sada/index.shtml).

If you did not see this welcome screen, it means that someone has previously selected to turn this option off. To make sure you see it in the future, simply go to the menu path:

Tools→Show Open Dialogue at Startup

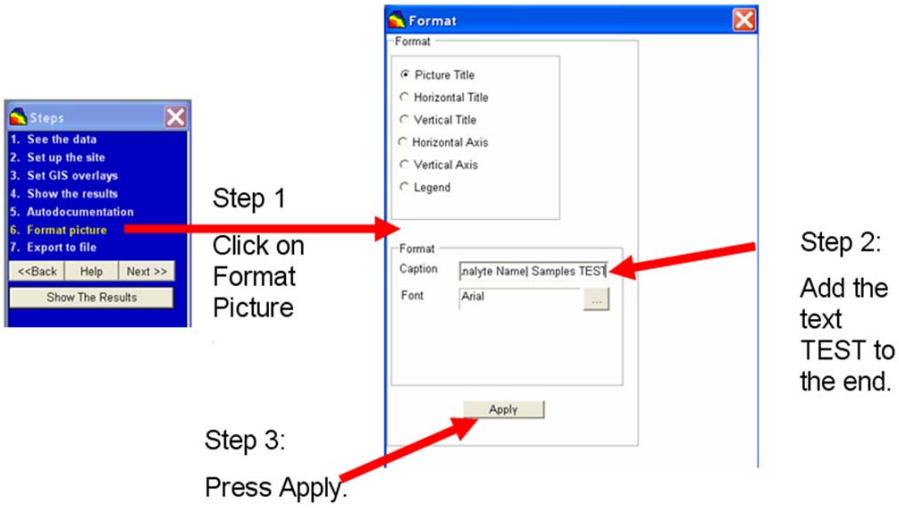
The next time you open SADA, the welcome screen will appear. In the meantime, you will select *Files→Open* from the main menu. Navigate using the open window dialogue to Twodimensional.sda and press *Open*.

The SADA file is now opened and you should see something like the following (depending on your version of Twodimensional.sda, you may see a more graphically rich display).



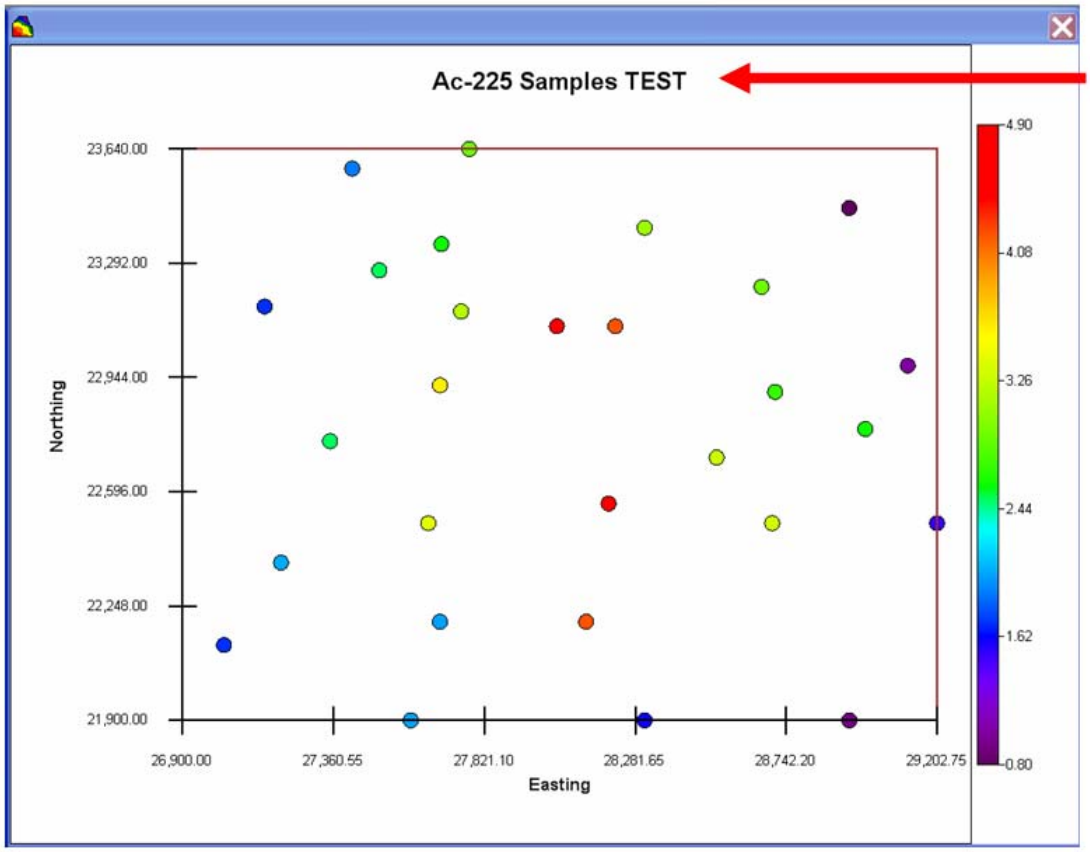
At the top of the SADA interface is the name of your SADA file (Twodimensional.sda). Much of the rest of SADA is now enabled, and you can see the first contaminant in the data set (Ac-225). If you need to refresh your memory on the SADA layout, revisit Chapter 1.

What does it mean to edit your SADA file? Well, we'll now make a trivial change so you can see what it means. On the left hand side of the screen is a blue window called *Steps*. You should see a step called *Format picture*. Click on this step.



The window on the right will change. This is the parameters window, and it adjusts itself to reflect whatever step you are on. You can learn more about this later. In the parameters window, there is a text box called *Caption*. Add the word "TEST" to the end of the text you see there. Press the *Apply* button. Notice in the graphics window that the title of the picture has

changed and now includes the word TEST. This is part of the format feature, and you can learn more about this later.



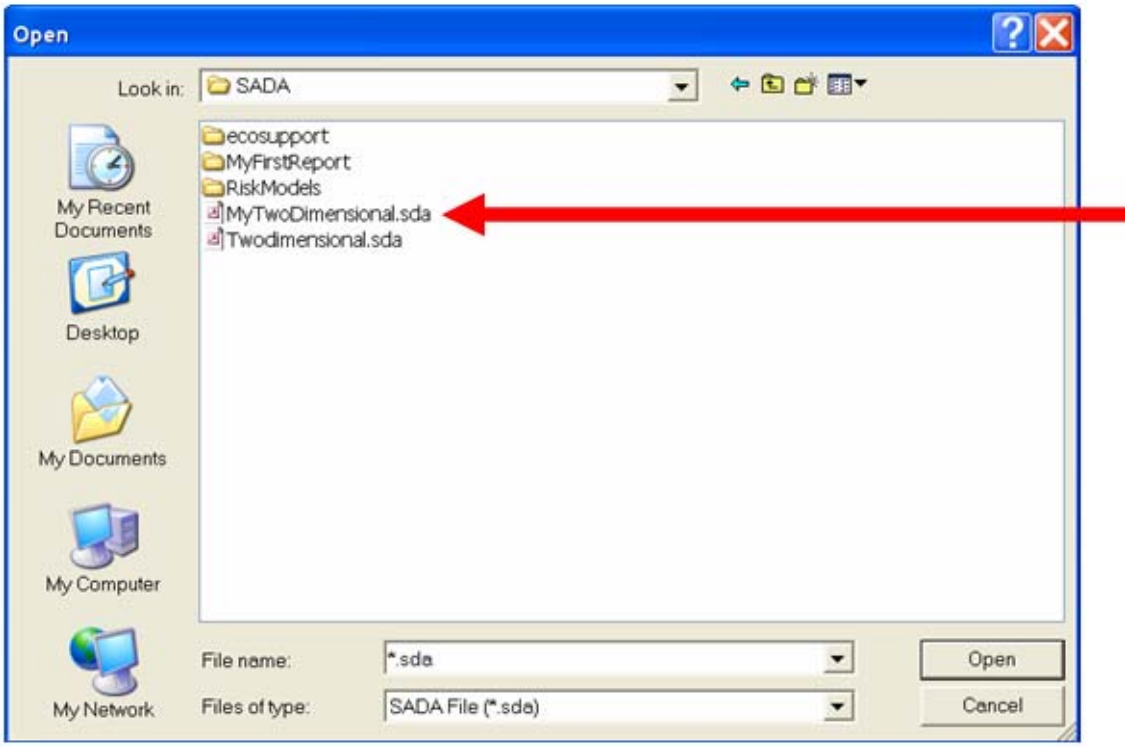
Now select from the main menu *File→Save As…*

Save the Twodimensional.sda file as MyTwoDimensional.sda.

Now select *File→ Close*.

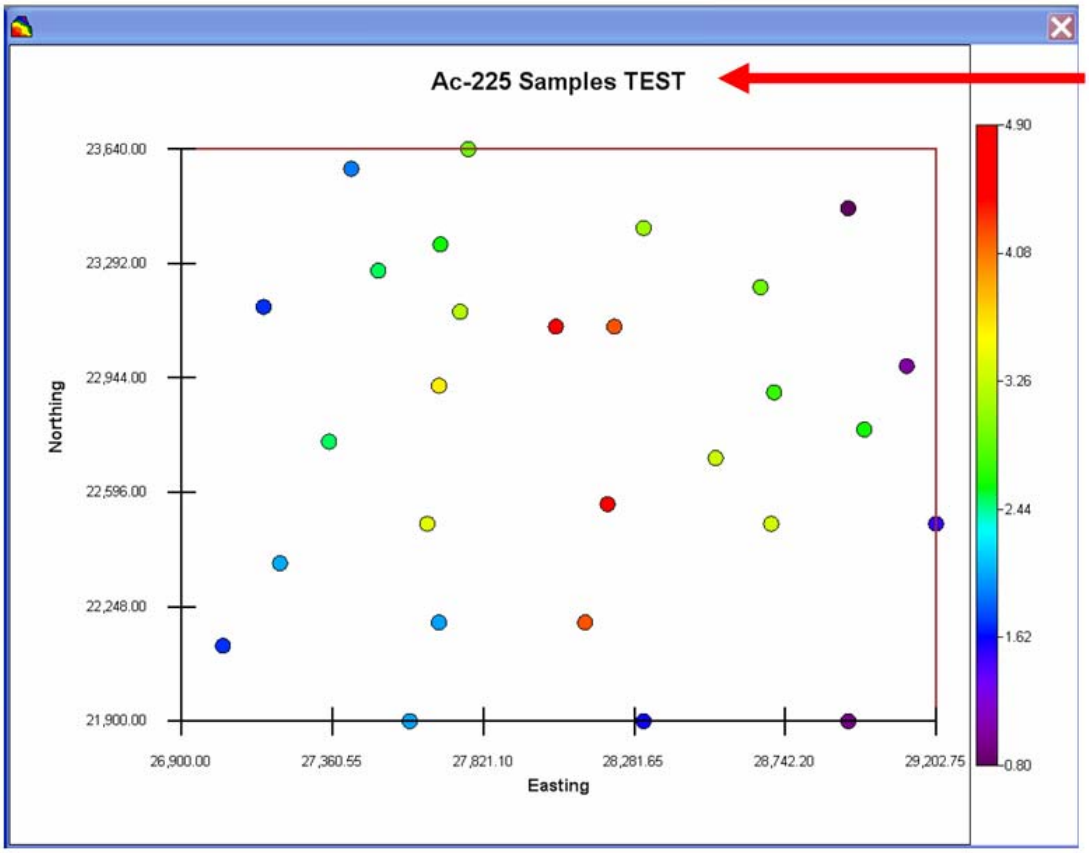
SADA asks one more time if you would like to save the file. We just saved it, so say *NO*.

Now select *File→Open*, and navigate to where you saved MyTwoDimensional.sda.



Select MyTwoDimensional.sda and press *Open*.

Notice that the modification to the content of your project ("TEST") was saved.



This is a very trivial example of how a .sda file stores everything you do in SADA. The important thing to know about SADA files is the following: Nothing is stored in the SADA file until you save the file. This is consistent with common Windows applications such as Excel or Word.

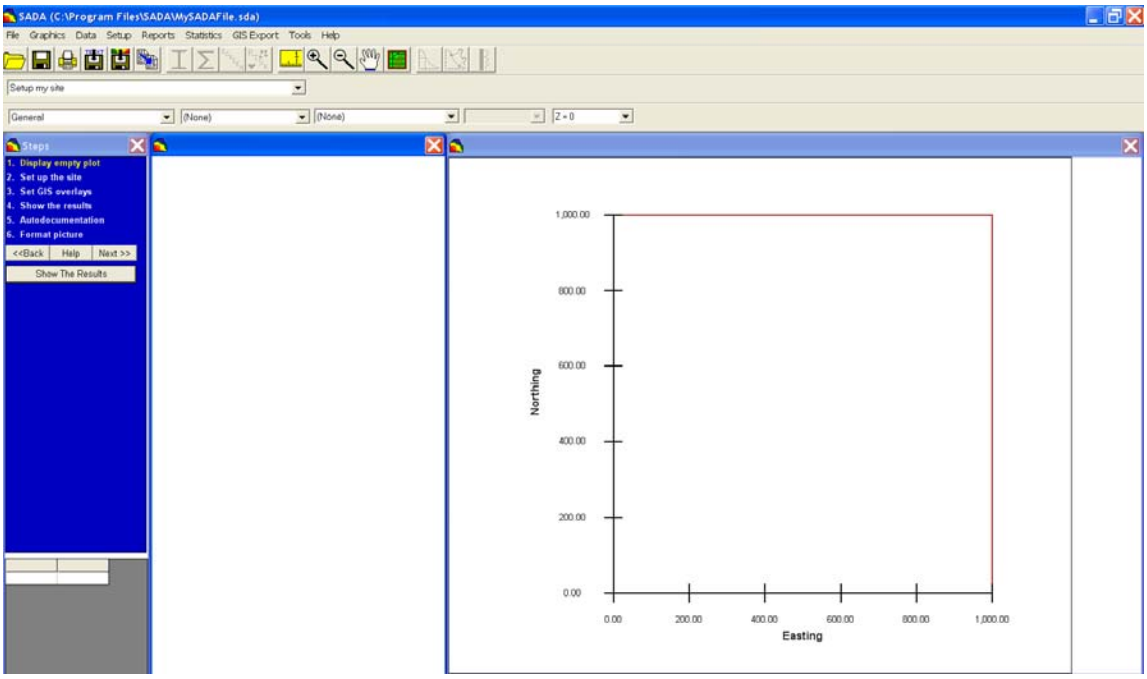
Any data that you import into the SADA file is completely contained within the SADA file. What this means is that any changes to the data you may make in the SADA file DOES NOT affect the original data file or database from which it came. The change is only internal to the SADA file. Conversely, any change in the original source file or data base will not affect the data values in the SADA file.

Throughout SADA, you will have opportunities to import a variety of things, such as toxicity values, custom screening values, ecological benchmarks and so forth. These values are copied and brought into the SADA file just like data. Therefore, any changes to these parameters done inside SADA are not reflected in the original file or data base. The converse is also true.

New Files

You can also create a new SADA file very easily. From your welcome screen, press the *New* button. Or alternatively, from the main menu, select *File→New*.

You will be presented with a Windows Save dialogue box. Enter a name for your SADA file (e.g., MySADAFile.sda) and press *Save*. SADA will respond by creating a "blank canvas" SADA file. The file is essentially empty because you have not brought any data into it yet. Therefore, a lot of functions are disabled.



You can do a few preliminary things, such as set up your site, bring in GIS layers, or create a new sample design. You'll learn more about these in the next chapter. If you like, you can close your SADA file now (*File→Close*) and choose to not save.

Chapter 4: Setting up a SADA File

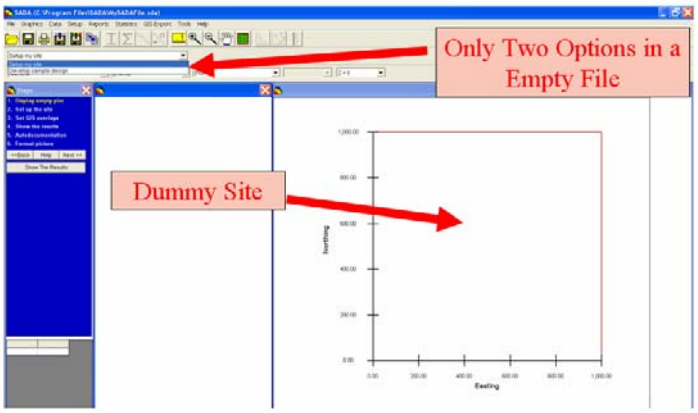
Let's start out by creating a brand new SADA file. This is very easy to do. Setting up the site after the file has been created, however, can take a bit more effort. When you first start, there may be some data already collected that can be used to help setup site boundaries, establish vertical layers, and so forth. Other times, you may not have any data but want to set the site up to perhaps create an initial sample design. In either case, the way you create a SADA file is the same and is relatively easy. When you have data available, there are some extra handy tricks (see Chapter 5, "Importing Sampled Data").

Let's start with the very beginning - you have no data and you want to create essentially an empty SADA file. There are a lot of things that can be done in a file that has no data. You can set horizontal and vertical extents of your site, bring in GIS overlays, and set boundaries of your site with the polygon tools. You can also develop an initial sample design that can be exported and submitted to the sampling team or import existing data.

Create a SADA File

If you do not have SADA running, open it up now. At the Welcome Window, select *Close*. If you already have SADA running, make sure you close out any file you may now have open by choosing from the main menu *File→Close*. If asked whether you would like to save or not say *No*.

From the main menu, select *File→New*. Navigate in the explorer window to where you have installed SADA. This is likely C:\Program Files\SADA 5 or something similar. In the file name box, enter the name MySADAFile.sda and click *Save*. SADA will go through a creation process and present you with a blank file. SADA will look something like the following picture.



The interview box will only have two interviews: *Setup the site* or *Develop a sample design*. The graphics window will show a dummy site. SADA needs to show something in the graphics window, so it starts you off with a dummy site on an arbitrarily selected coordinate system of 0-1000 in both the easting and northing direction. Next, we'll show you how to setup a meaningful site.

Notice that you are automatically in the interview *Setup the site*. If you are not on this interview, switch to it now. You are automatically set to the first step in the interview, which right now is *Display empty plot*. The first step in any interview is usually an action step, meaning if you click on it some action is immediately taken. In this case, every time you click on the first step, it immediately redraws or refreshes the empty plot.

Let's move to the *Set up the site* step. Here is where we can change our dummy site into something useful. This is done by specifying the bounding box, called the "site boundary." It is not likely that your site is a perfect rectangle. That's OK. This is really just a bounding box that needs to be big enough to encompass your entire site. You can use the drawing shapes found further down to really define the boundary of your site. So to get started, all we need to know is the coordinate boundaries for our site. Unfortunately, we often may not know this right off hand. We could also set up our vertical layering scheme, but this is usually secondary to getting the horizontal extent correct. If we don't have a horizontal bounding box, we can't draw any meaningful polygons. If, for some reason, you do know the coordinates for your set and can enter them in now, you are in luck and you should definitely do so. But for the rest of us, what can we do?

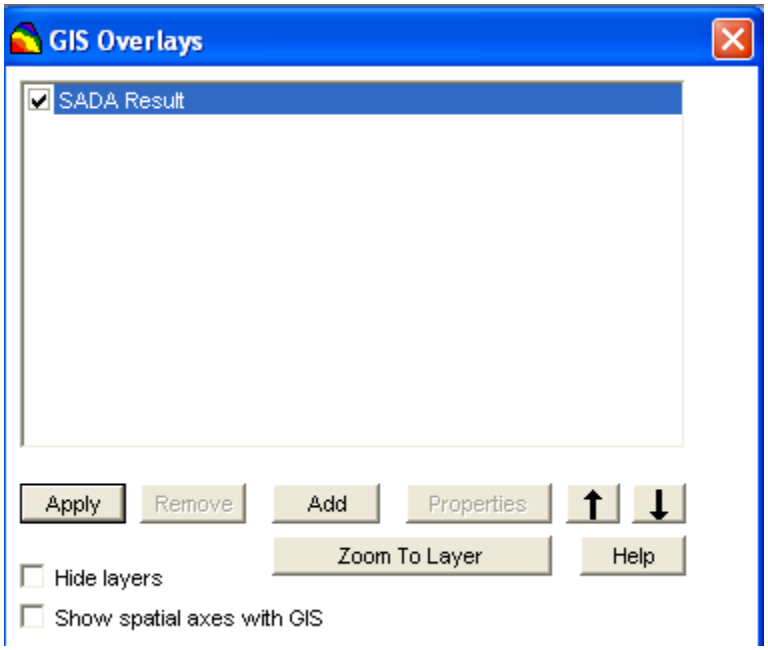
We need some context; that is we need a map or a photograph to use as a reference. Then we have a good chance of recognizing our site and being able to realistically define it for SADA. If you don't have a map or any kind of photograph, maybe you have some sample points to import. They might provide you some context for setting the boundaries of your site. We'll show you how to do that next. For now, we'll assume you have no data, and we will move forward without it for now.

To get some context, we need to look to the *Set GIS overlays* step. This is the third step, but we have seriously considered moving it to the second step as this is often where users go first.

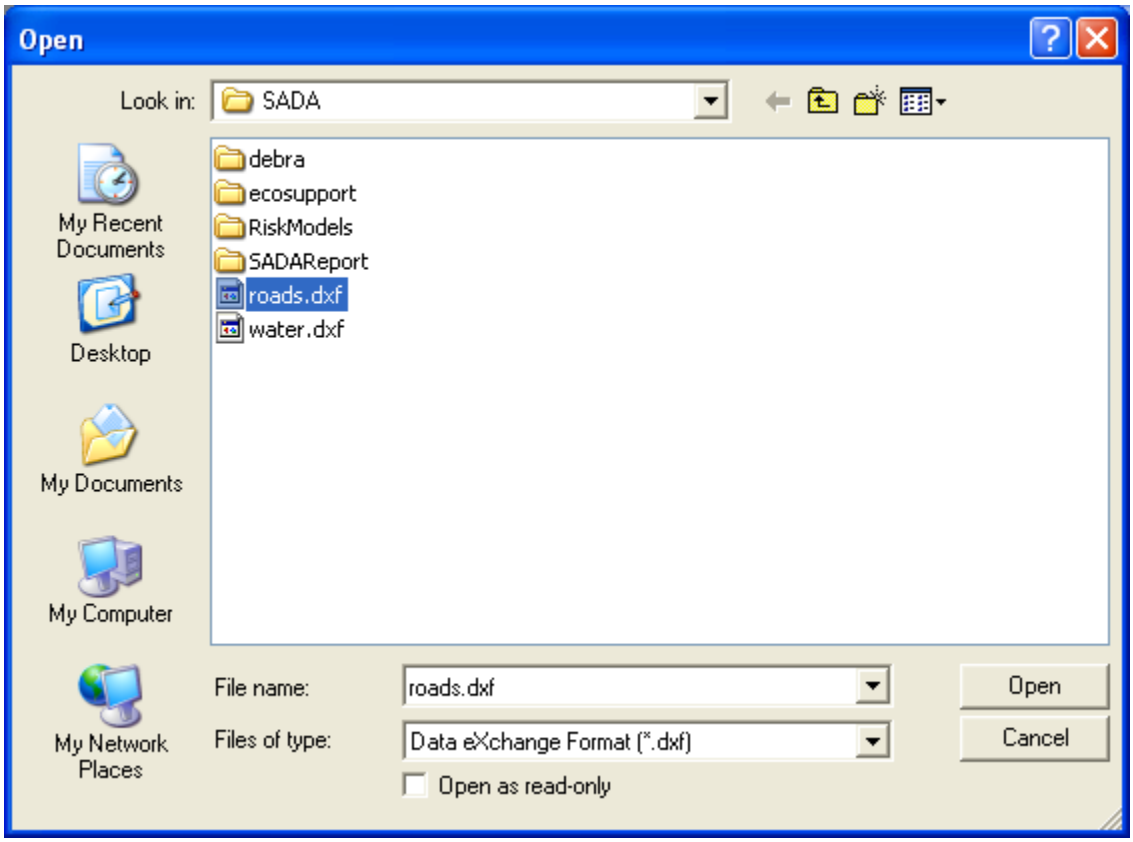
Working with GIS Layers

SADA can read and use map layers that have been produced by a Geographic Information System (GIS) and saved in a Data eXchange Format (.dxf) or Shape File (.shp) format. For example, one layer may show all roads and another may show all buildings. In addition, SADA can read and overlay photographic/raster type images such as .JPG, .GIF, .TIF, and .BMP files. These photographs can be dropped onto a site and shifted/stretched/compressed to match existing GIS or data layers. SADA does not allow users to edit or query GIS layers. They serve strictly as maps to provide reference. *As a note to GIS folks, SADA can also import raster images such as geospatial models and elevation in a ASCII raster format or a ESRI grid format. Elevation can also be brought in as a Digital Elevation Model (DEM). This type of layer can be used directly in SADA decision and modeling processes and is considered separately from the types of layers presented here.*

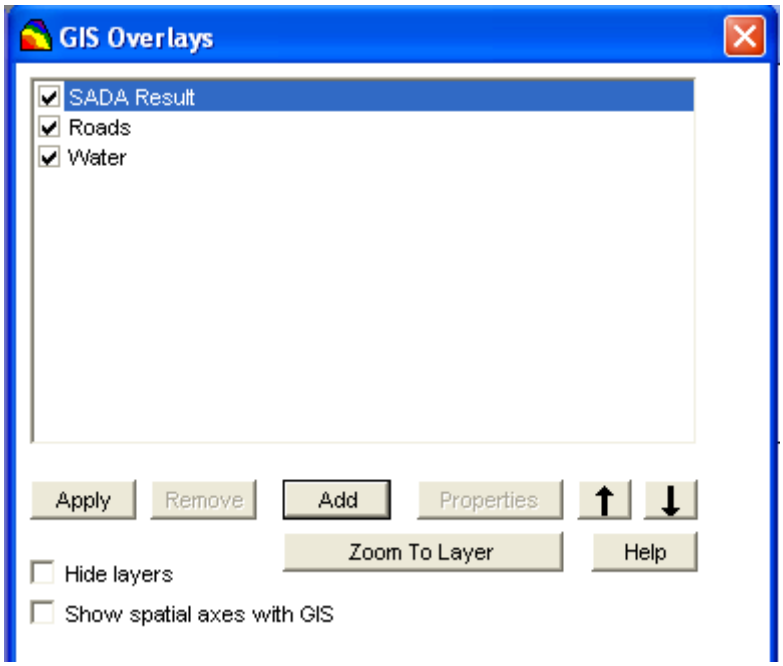
Set GIS overlays is one of the basic steps and is usually step 3. Click on this step now and SADA will display the parameters window.



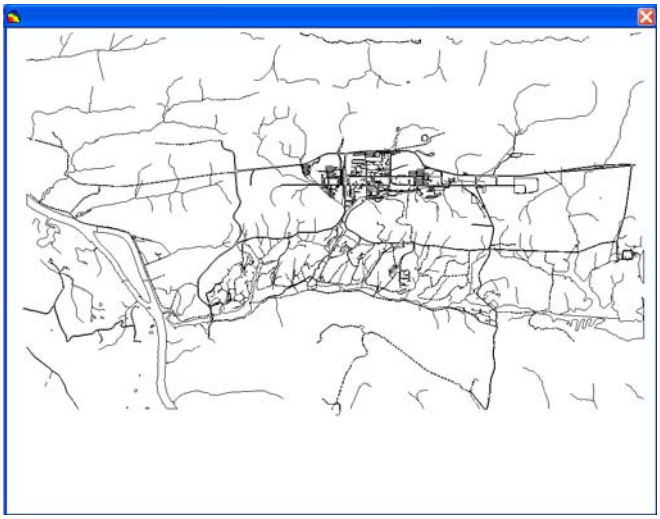
To add an overlay, press the *Add* button. SADA can import Data eXchange Format (.dxf), shape format (.shp), .jpg, .gif, .tif, and .bmp files. In the explorer window, make sure the *Files of type* option is set to Data eXchange Format. Select the roads.dxf file and press *Open*. Repeat this and choose water.dxf as well.



After adding overlays, the parameters window will look like the following.

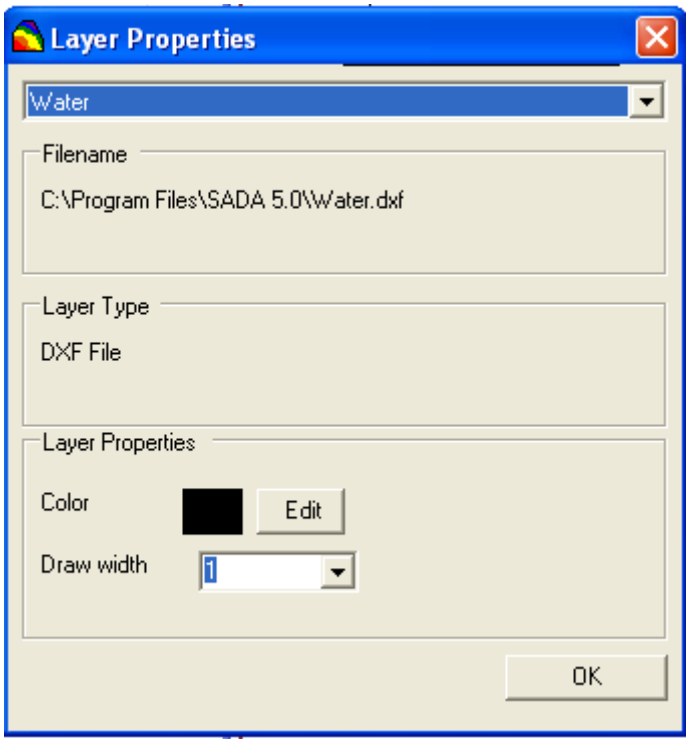


Chances are pretty good that you don't see anything but a blank graphics screen. This is because the coordinates for these two map layers are nowhere near the 0-1000 easting and 0-1000 northing range. This is a common problem in any GIS system, that is, matching coordinates properly. We don't really need to match our 0-1000 dummy site to anything because it is, after all, a dummy site. What we want to do is find the layers we just imported. To do this, select *Roads* in the parameters window and press the *Zoom To Layer* button. You should see something like this:

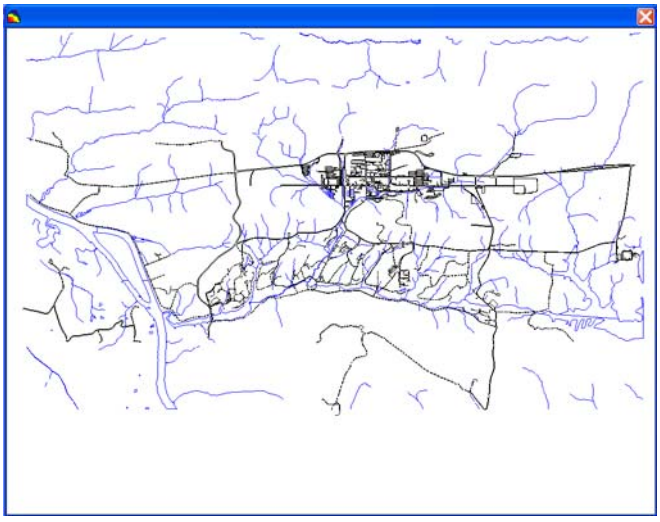


Move your mouse over the picture and notice in the lower-left-hand corner that SADA is reporting the exact coordinate of your mouse pointer. Notice that these are nowhere near 0-1000. This means that our dummy site is way off in the lower-left-hand side way out of view. We haven't yet changed our site. We've only shifted our view to the GIS layers we just imported. We can use these layers to get some kind of reference for where our site resides.

Right now though, both the roads and the water layers are colored in black. Let's take a look at the properties feature of each layer. Click on the *Properties* button. In the window that comes up, choose *Water*.



SADA provides you with some useful information in this window. First of all, we see the original location of the .dxf file for water. Note that SADA has copied this file and added it to your SADA file. So, SADA isn't actually using this file anymore. Any changes to this file will not be seen in SADA until you reapply it. We see that it is a DXF file. Below that are the layer properties we can change. To change the color, press the *Edit* button and change the water color from black to blue. Press *OK*. To change the draw width, type a number in the box or select one of the predefined sizes. Press *OK*. To see the changes apply in your graphics window, press the *Apply* button.

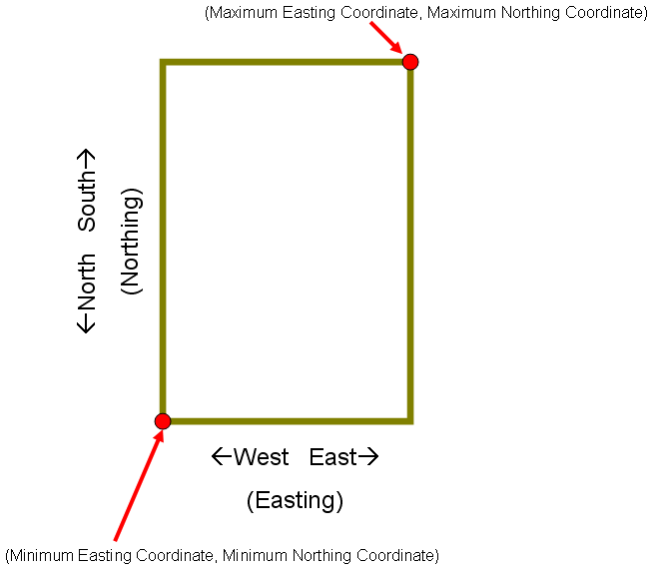


To remove an overlay, select the overlay and press the *Remove* button. To make an overlay visible, check the box next to the name. To turn off the GIS overlay altogether, click in the box next to *Hide Layers*. Repeat this process to turn the GIS back on (*Hide Layers* will be unchecked).

Setting Site Boundaries

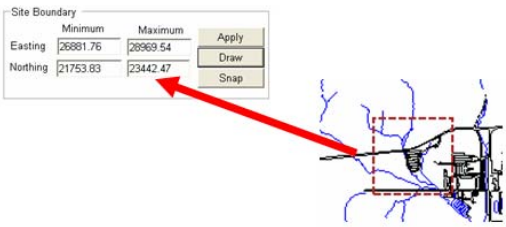
Let's now define the real site of interest. First, make sure the graphics viewer still shows the roads and water layer. If it doesn't show anything for some reason, in the parameter window, click on *Roads* and press *Zoom to Layer*. Click on the step *Setup the site*.

So, let's spend just a moment for vocabulary. The left to right direction in the graphics viewer is referred to as the easting, meaning this is the east and west direction. The up and down direction is the northing, and it refers to the north-south direction. Now, the site boundary box is defined as follows.

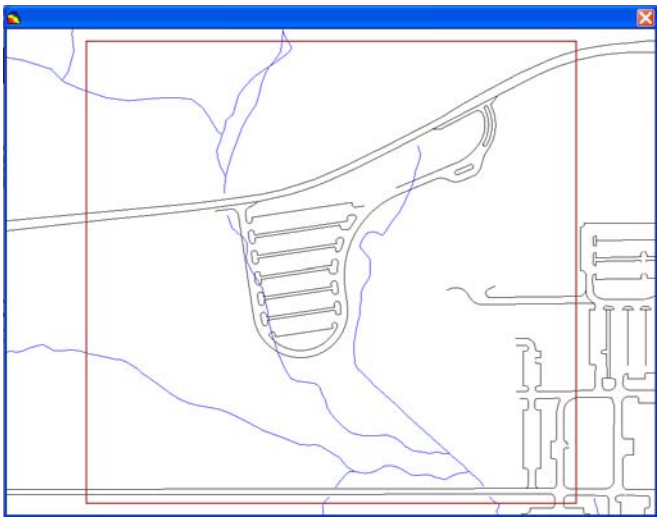


So, we need to specify the minimum easting and northing value as well as their maximum counterparts. We can do this in three ways. First, if we know them, you can just type them in and press the *Apply* button (don't try this now). Second, we could draw them. This is the one we want to demonstrate now.

To draw your site boundaries, press the *Draw* button and move your mouse over your graphics viewer. To draw, choose a place on your map, then left mouse click down and hold it. Then, move your mouse around and you will see a brown box open up and follow you around. This is your site boundary box. When you have it sized how you want, release the mouse and notice that the proper coordinates have been entered for you.



Now you may have drawn your site differently, and that's okay. When you release the mouse, SADA snaps the site into view. In this demonstration, the following site was created.

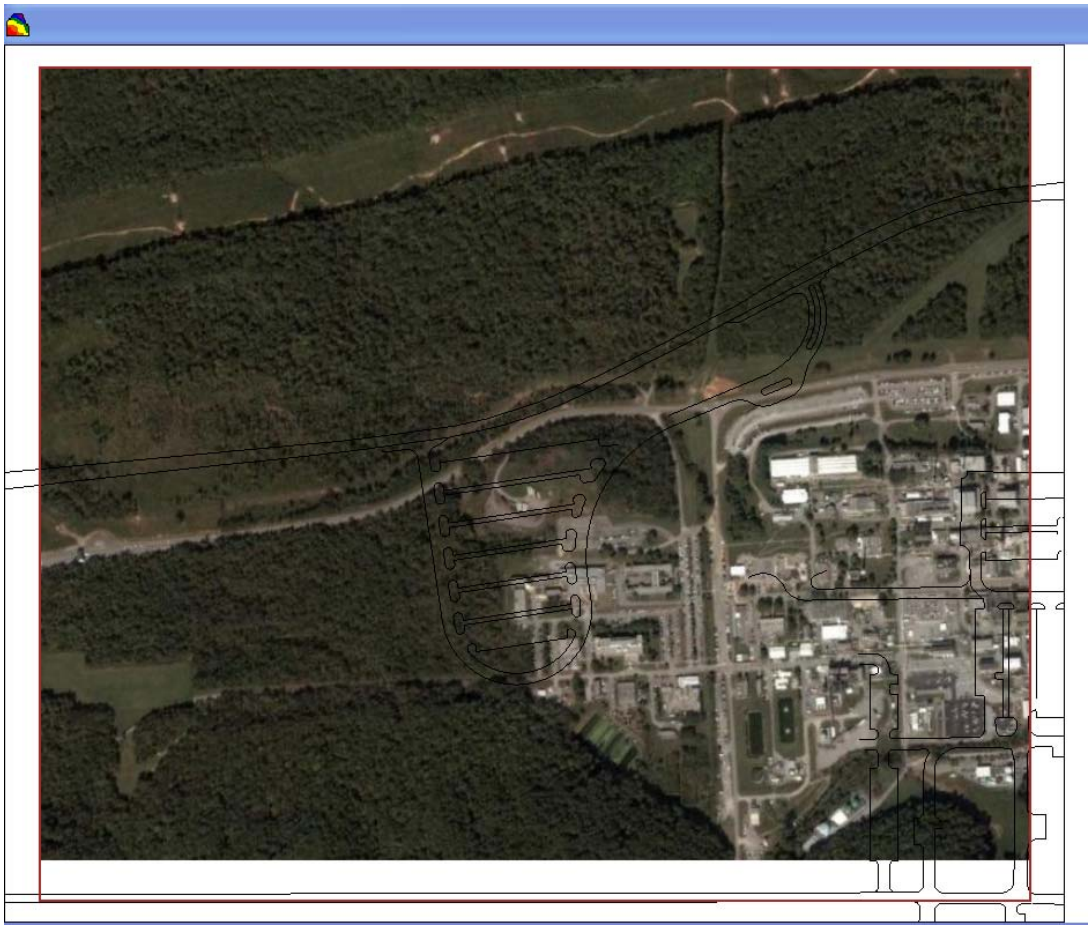


Before you go any further, draw your site approximately like the one you see above. Then save the file by choosing from the main menu *File→Save*. Next let's bring in a photograph.

Working with Photographs

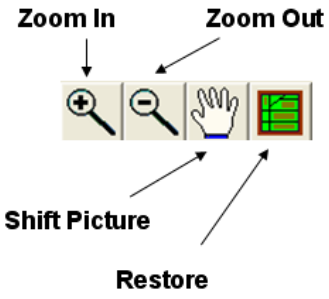
While photographic/raster layers can be brought in the same way, they do not have coordinates to orient them relative to the data or other layers (for example, TIFF files may have a coordinates file, but SADA won't use the coordinate information at this time). It is definitely important to have your site boundary box setup before you proceed with bringing in photographs.

You can bring in any type of photograph. In the examples we present here, we simple did screen captures of Google Earth and saved them as .jpg files. To show how this works, return to the *Set GIS overlays* step and click on *Add* again. Navigate to Program Files\SADA 5.0 and change your file type options at the bottom of your open dialogue to .jpg. Choose biggerview.jpg and press open. When a photographic or other raster type image is added as a layer in SADA, the image is dropped directly onto the site without any reference. This is just the starting point.



First, let's change the color of the roads layer to red so we can see it better against the photo background. Click on the roads layer, push the properties button, and change the color to red.

Now let's talk about zooming, shifting, and restoring a bit. You'll find these features on the main toolbar.

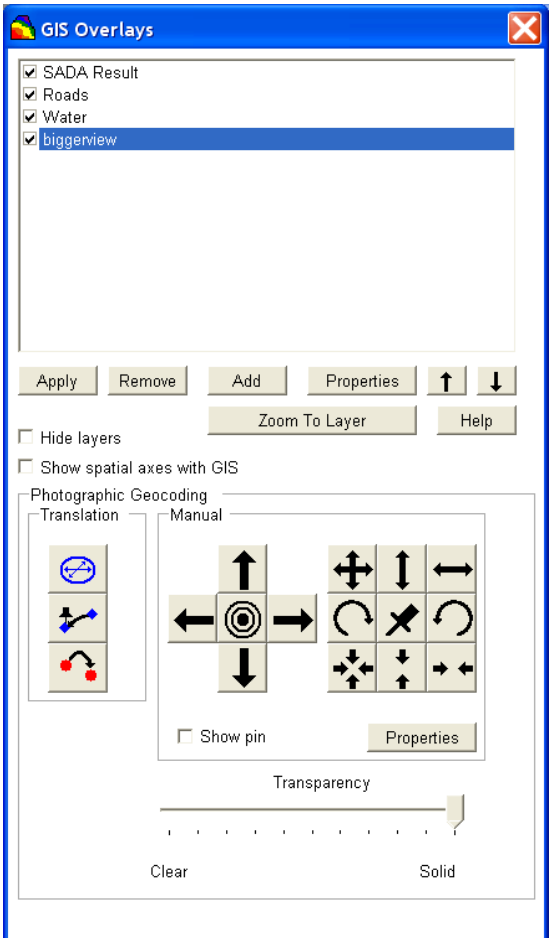


The *Zoom In* and *Zoom Out* buttons work much the same way as drawing site boundaries. To try it, press the *Zoom In* button. Most of the SADA interface will become disabled (SADA doesn't' like you multi-tasking). Move your mouse over the graphics window. Then, left mouse click down and drag open a zoom window. *Zoom Out* works the same but makes the view wider. To use the shift feature, press the button with the hand on it. When you move your mouse over the graphics window it will turn into a hand. Left-mouse-click down and move the image around. When you release the mouse, it will refresh the image. Now if you really lose

yourself in all this and want to get back to your site, press the *Restore* button. Press the *Restore* button now.

Now back to photographs. First, zoom out a little so you see all the road features you see in the following image.

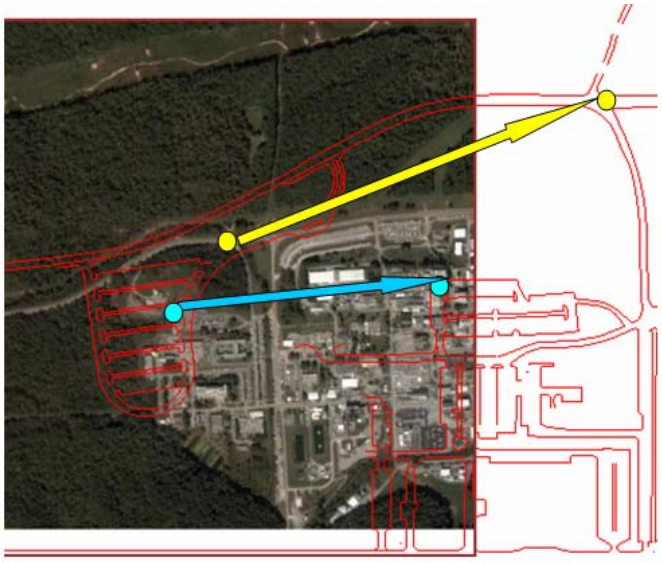
Select the *biggerview* layer from the parameters window, and the parameters window will display the photographic geocoding options available in SADA.



There are a number of features here that allow you to shift, stretch, and otherwise position your photograph. It is a bit overwhelming at first, but really the only feature you need to know is the rubber sheeting function accessed by the button  .

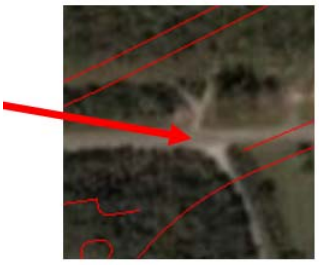
Rubber sheeting translation allows orientation of the photograph by graphically specifying the location of two points. This is the recommended approach and can likely better orient your photograph than any other positioning tools found in SADA. Essentially, you match up a pair of points on the site, and the picture is reoriented to match the rest of your data.

To begin geo-referencing our photograph with the other layers, we first need to locate points that appear on both the photograph and one of the other layers (building corner, road intersections, etc.). The following image shows the two points of interest in this demonstration. The yellow arrow connects the T intersection in the photograph with the T intersection on the *Roads* layer. Similarly, the blue arrow connects the corner of a parking lot with that same corner on the *Roads* layer. Before you begin, make sure you are zoomed out enough to see both of these locations in both the photograph and the *Roads* layer inside your graphics viewer.

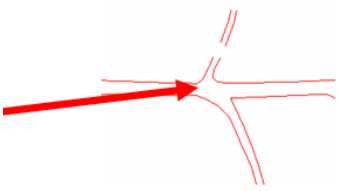


Next, click on the rubber sheeting button  in the parameters window. SADA will prompt you to start rubber sheeting. Click on a point in the photo and draw a line by moving the cursor to where the point should be. It is important to remember during the rubber sheeting process to click on a point in the photo and draw from it to the other layer rather than moving a point on the other layer to a location on the photo.

At the prompt, click *OK*. Then, click on the T intersection of the photograph.



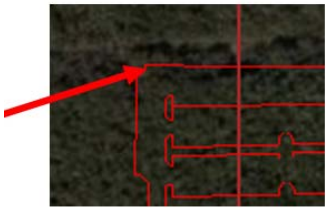
Drag the line over to the same T intersection on the *Roads* layer and click again.



SADA pops up the coordinates of the point. Click *OK*. You'll then see a prompt to specify a second location for rubber sheeting. Click *OK*, and you'll see the photograph has been reoriented to match the first point you specified on the photo to the location on the *Roads* layer.

Now click on the edge of the parking lot in the photograph.

Drag the line and click on the same corner of the same parking lot in the *Roads* layer. Be careful.



After you click on the location where the second photograph point should be, you'll see a window with the coordinates for that location. Click OK, and SADA will redraw the image, stretching the photograph such that the first and second points you specified on the photograph now fall at the locations you specified on the *Roads* layer.

If you made an error, you are in good company. You can do one of two things. First just repeat the process again. But if things are really out of line, press the *photograph restore* button. This is found on the parameter window and looks like a bull's eye  .

This will reset the photograph into your site boundary box. An easy mistake to make is to do the reverse. That is, select the location on the *GIS* layer and draw it to the correct location on the photo. This mistake can lead to some surprising (and even disorienting) results.

Another feature is the transparency control. This feature allows you to control how well you can “see through” the photograph. This is useful if you should need to overlap photos. For example, one photo may be actually taken from an on-line road map service, like Map Quest, and a second photo might be from an old engineering drawing.

The remaining geo-referencing tools on the parameter window are intuitive and are discussed quickly. If you are not able to use rubber sheeting to align your photograph, you are not likely to succeed using these tools either. These tools were intended to be refinement actions to make very fine-tuned adjustments.

The buttons are described briefly as follows.

 - shifts picture up.

 - shifts picture right.

 - shifts picture down.

 - shifts picture left.

 - restores picture to default orientation.

 - sets a pin point on the map where all expansions originate from.

 - stretches equally in both directions from the pinpoint location.

 - stretches equally in up/down direction from the pinpoint location.

 - stretches equally in left/right direction from the pinpoint location.

 - rotates clockwise around pinpoint.

 - rotates counter-clockwise around pinpoint.

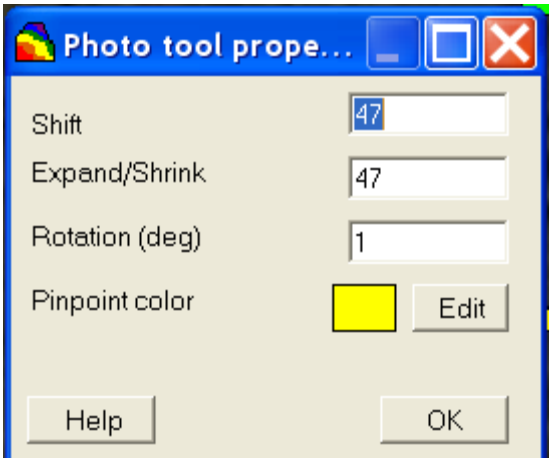
 - contracts equally in all directions relative to pinpoint.

 - contracts vertically relative to pinpoint.

 - contracts horizontally relative to pinpoint.

When the *show pin* box is checked, the location of the pin point around which you are rotating and stretch/shrinking will be displayed.

The properties button under this set of tools will allow you to set the distances that you shrink, expand, or move with each button push. Similarly, you can set the angle you would like to rotate with each push of the button. The smaller these values are, the more refined your adjustments will be. You can also control the color of your pin-point when it is displayed.



GIS/Photograph Tips

It can be complicated to geo-reference a photograph to other data layers if the scales are disparate. Something to consider when you first pull a photograph into SADA is whether you can see specific points on the photo and another layer to use in geo-referencing. In this tutorial, the scale of the site is much larger than that of the photograph, making it more difficult to find matching points. A simple solution is to remove the photo layer, broaden the site boundary, and then re-import the photo layer. When the photo drops onto the larger site, you should be able to see more locations--road intersections--for instance, to use in geo-referencing. Having the scales more similar from the start makes geo-referencing easier.

It is helpful to set up the site boundaries prior to importing the photographic layers, as this defines where the layer will appear. If no site boundaries are set, the layer will appear in the default 0-1000 easting/northing, which may not coincide with the location of data or other layers for the site.

The *GIS overlays* parameter window shows the current layers in order of priority. Thelayer that appears first is plotted over those that appear later. If you pull in a photograph before pulling in other GIS layers, the other layers will be plotted behind the photograph and you won't be able to see them. Use the up or down arrows in the *GIS overlays* parameter window to move layers up or down on the list. Putting your photograph at the bottom makes it the backdrop for all the other overlays.

There is an unmentioned player here called projections. Because coordinate systems may be in different projections, it is unlikely that your photograph will exactly match every feature in your GIS layer. Even in true GIS systems with projections accounted for, there will be small discrepancies. This is unfortunately the reality of GIS tools and the quality of the data they are required to use. What is important is to match the photo as well as possible to the area of interest. Because of these discrepancies, we DO NOT recommend that you use geo-referencing of this sort when it is extremely important that the photo or layers are exactly positioned. If issues of safety arise (e.g., things like gas lines, highly dangerous buried contaminants, etc.), you should not rely on this method but rather turn to a true GIS system with extreme local accuracy.

Finally, sometimes GIS layer order matters. Sometimes the SADA result will be a contour result and may hide the road or water layers. In this case, you'll want to use the up and down arrows to specify the order in which the layers are displayed with the layers higher on the list being plotted on top of the layers below them.

SADA doesn't automatically show the spatial axes associated with the plot when displaying GIS layers. If you want to see the axes with the GIS layers, check the box.

Setting Vertical Layers

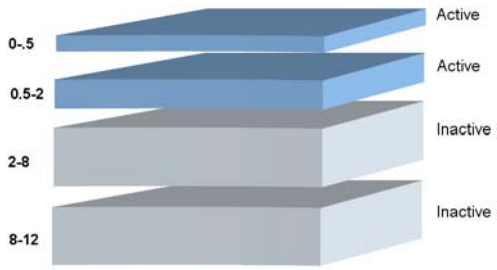
SADA is ultimately a 3d tool, so it is a good time to talk about how we deal with the subsurface. We've spent a great deal of time on how to establish the horizontal extents. There are times when you will need to deal with the 3d extents as well. Unlike the horizontal direction, the vertical direction is divided into layers.

In the following picture, you see a vertical layering design that divides the subsurface into the first half foot, then the next 1.5 feet, then 2-8 feet and finally 8-12 feet. These measurements are always relative to the surface (i.e., 0-0.5 means 0.5 feet from the surface; 0.5-2.0 means from a depth of 0.5 feet below the surface to a depth of 2 feet below the surface). As always, the units are not tracked by SADA, but they should match the same units as that in which the horizontal is measured. For example, if the horizontal is reported in meters, then the depth should be reported in meters as well.

In the image below, the layers have been visually separated so you can see them better in the image. But in reality, they are stacked right on top of each other.

Perhaps the most important thing to note about this design is that the vertical layers have irregular thicknesses. The first layer is only ½ foot, meter, etc. The second layer is 1.5 units thick and so forth. This is very different than many modeling packages where the subsurface is divided evenly. From a risk perspective, this is particularly important as different exposure scenarios are associated with different depths. For example, residential is typically associated with the first ½ foot while industrial or worker scenario may be associated with the first 2 feet.

Dividing the subsurface this way affects the visualization of data as well as the modeling features. It is also possible to turn layers off and on. In the following picture, we see a representation of the subsurface layers where active layers in blue are turned on. Gray or inactive layers are turned off. If a layer is turned off, then data points found in that layer are not considered in the analysis (an exception is for geospatial analysis, where you have the option to continue including them even if you are not modeling the layer). So for risk assessment, the data would not be included. For modeling, these layers would not be contoured. For decision analysis, they would not be considered in computing the remedial volume.



A layer design is actually a collection of layers. You can create as many layering designs as you like and assign each one a unique name. SADA comes with two special case layer designs.

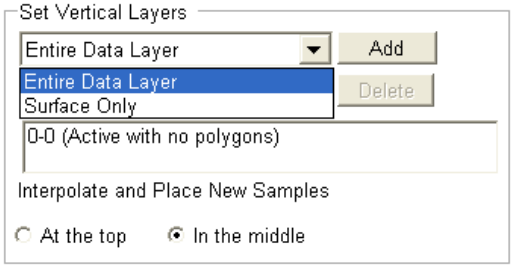
The first one is called *Entire Data Layer*. This layer is dynamically created and maintained by SADA. You cannot edit it yourself. This layer design is a single thick layer that is always large enough to view all your data at once. This is especially useful when you are importing new data and want to make sure your layers are deep enough to see all the data.

The second layer design is called *Surface Only*. This layer is always at depth = 0 and cannot be changed. Certain features, such as the MARSSIM analysis, depend on its presence.

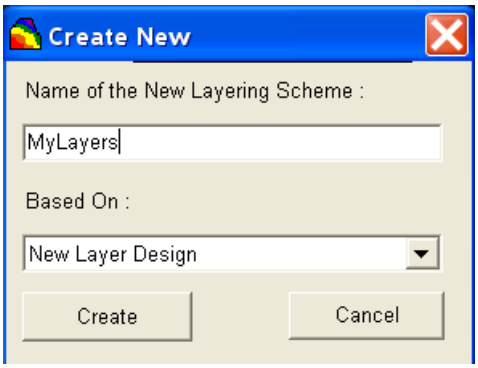
To make sure you are on the same page as this text, close out any file you might have open in SADA and say *No* when prompted to save. If you do not have SADA open right now, go ahead and start it and close out the Welcome screen.

Now let's open the file MySADAFileWithGIS.sda. Recall that to open a file you choose from the main menu *File→Open*. Navigate to the location where SADA is installed and choose MySADAFileWithGIS.sda. This file picks up where our previous file creation activities left off.

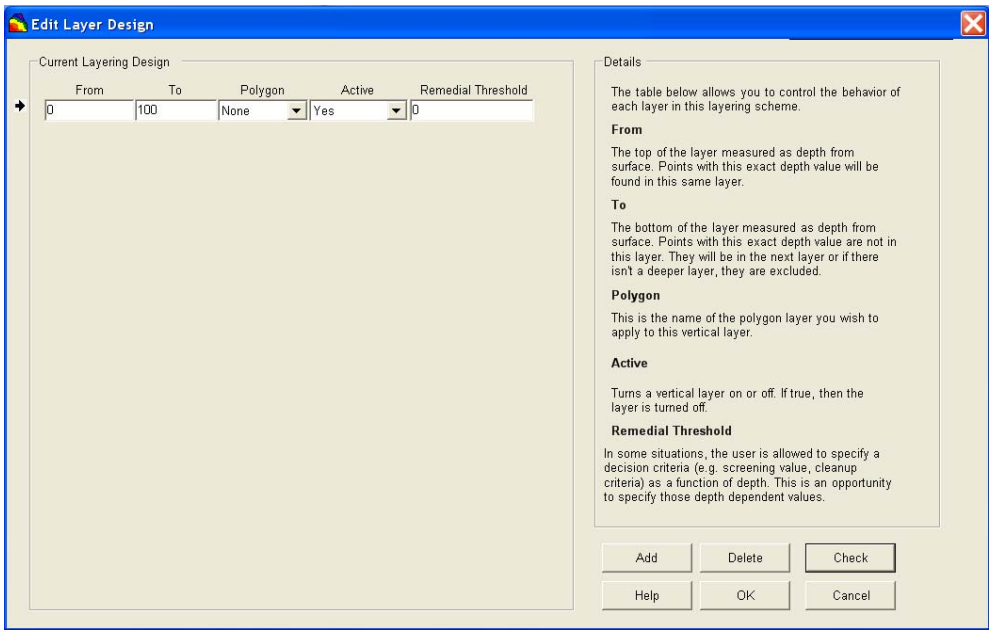
Click on the *Setup the site* step and notice the *Set Vertical Layers* section.



If you click on the layering design box you will see the two default layers. Let's add a new layer. Press the *Add* button and you will be presented with the new name window.

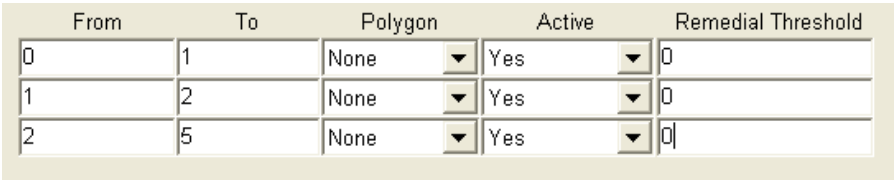


At the top of this window, type in “MyLayers.” This is the name of your new layering design. Underneath this you can select a previously created design to use as a template for your new design. For now, we’ll just keep the selection “New Layer Design.” Press *Create* and you will be presented with the vertical layering window.

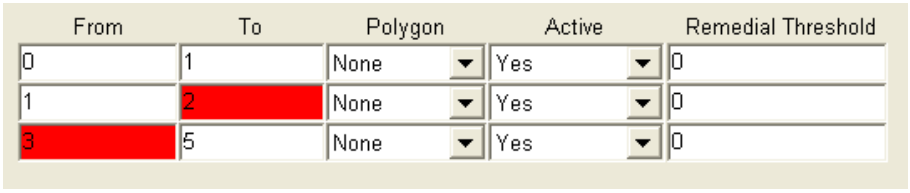


On the left hand side is a single row of entry boxes. On the right is some explanation of what to do. The window is fairly intuitive. You will enter values in the *From* and *To* columns to define each layer depth. Press *Add Layer* to add a blank layer to the layer scheme. If polygons have already been created, select the applicable polygon layer to display in each vertical layer (you don’t have any right now). Specify whether you want the layer to be turned off by choosing *Yes* or *No* in the active column. A *Yes* means the layer is turned on. In the final column you can enter a depth-specific decision criterion.

Click the *Add* button twice, creating two more rows (or two more layers). We’ll create three layers by filling out the entry boxes as follows.



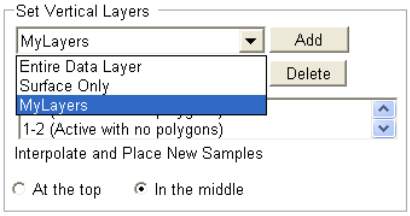
If you have any errors in your design, SADA will highlight these with red boxes when you press the check button. A common mistake is to have gaps in your layers, as in the following picture.



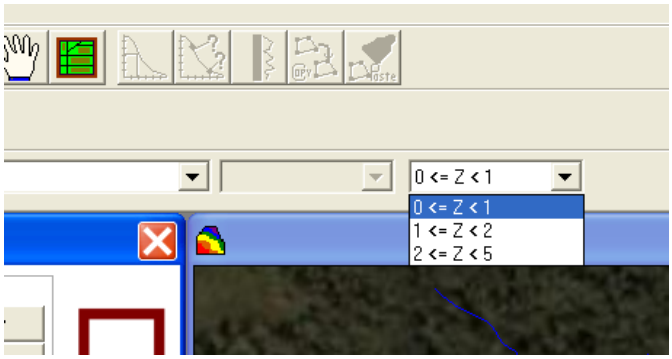
Here, we've typed in 0 to 1, 1 to 2, and 3 to 5, leaving a vertical gap between 2 and 3. This is not allowed. If you want a gap to appear, you'll need to define a layer for 2 to 3 and then turn the layer off.

You can delete a row by clicking anywhere in the row and pressing the delete button. Make sure you have the design specified as in the first picture and press *OK*.

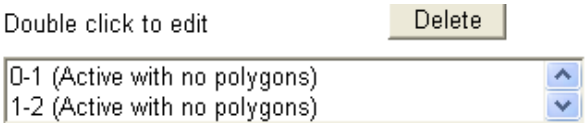
You will see your new design in the list of layer designs. Select *MyLayers* from the drop-down list.



Since we have no data in our SADA file, not much has changed when you selected your new layer design. One thing that is important to point out is that your depth layer box now has your layer intervals in them. Note that a layer from 0 to 1 actually means 0<=z<1 (meaning that 1 is not included in the 0 to 1 layer; it would be included in the next layer 1<=z<2).

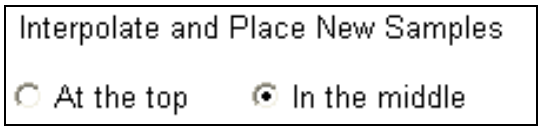


You can always edit your layers by double clicking in the layer list box.



This will return you to the layer creation screen we just left.

Finally, when SADA places new samples or interpolates a 3d grid, it will need to know where in the vertical layer to place the samples. The options are in the center of the layer or the top. This option is selected in the last part of the *Set Vertical Layers* section.

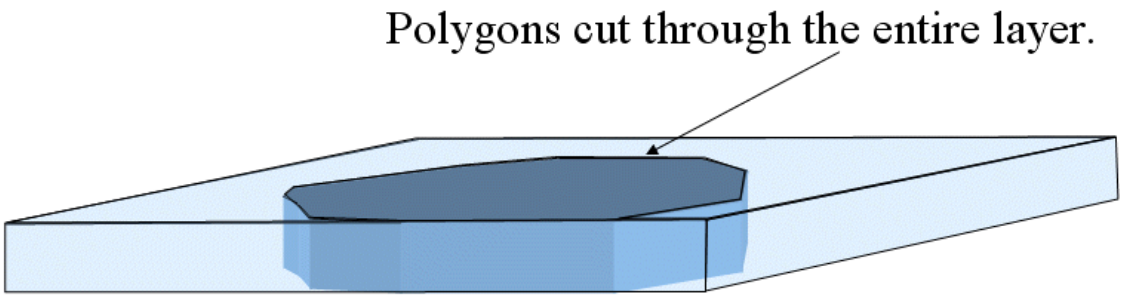


The “in the middle” option is the most popular, but sometimes users want an estimated value exactly at the surface. If the top layer has a nonzero thickness, say 0-1ft, then this would require you to choose “At the top” in order to have values placed at the top of the layer. Otherwise, they would be placed at z = 0.5.

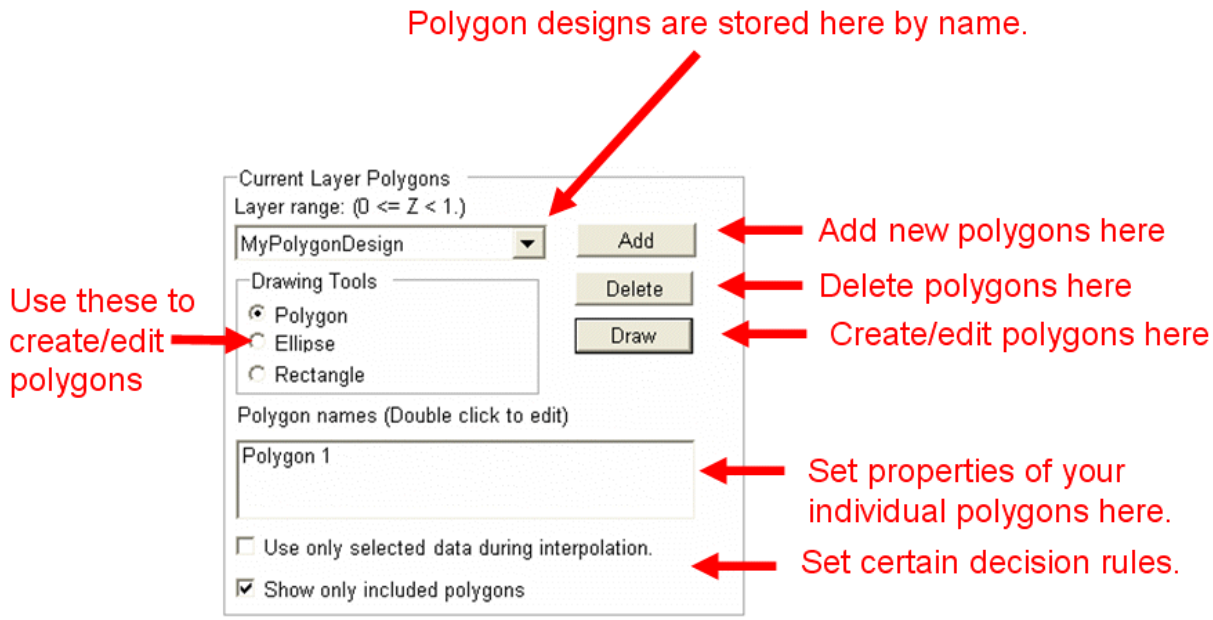
Working with Polygons

Within a spatial framework, it is often necessary to select certain items or identify a particular area in space. SADA allows you to draw various shapes for defining areas and selecting items in space. These include ellipses, rectangles, and polygons. While it would be more accurate to refer to these as shape tools, they have been historically referred to as polygon tools and will therefore remain with this nomenclature. Polygons are very flexible and can be used for defining irregularly shaped spaces. When a polygon is drawn, items or space located inside the polygon is either included or excluded, depending on the type of polygon.

Polygons are created and maintained as "GIS-like" layers themselves. In fact, we refer to them as polygon designs. Each one can be assigned to one or more vertical layers. A polygon layer is "embedded" into a vertical layer and cuts through the entire layer



The part of the parameter window that deals with this is shown here with more detail:



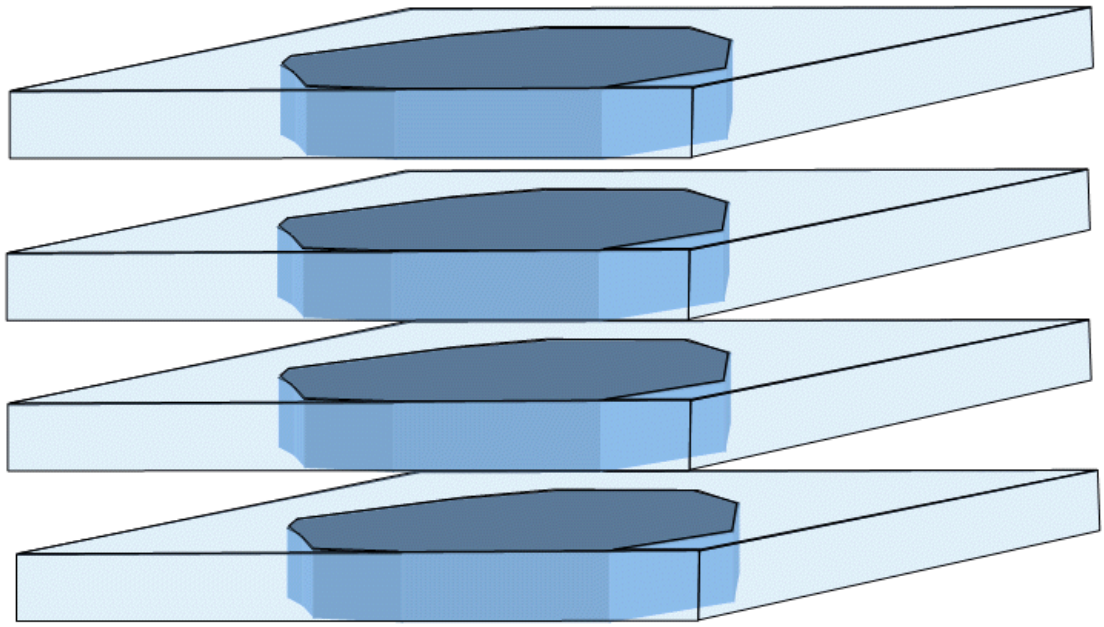
Creating a Polygon Design

Let's begin by creating a polygon of our own. Press the *Add* button to add a new polygon layer.

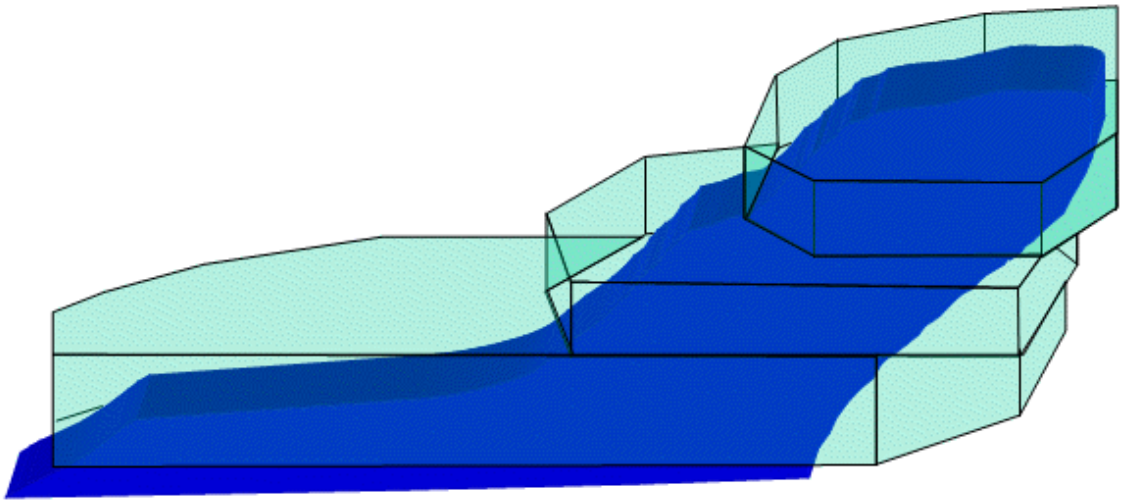
Enter the name "test polygon" for the design name. Just like in vertical layers, this polygon design could be based on an existing design or template. In this example, let's choose *None* for the *Based On* option. This means we'll start from scratch.

Then press ***OK***. If you have more than one vertical layer in your layer design, SADA will ask if the new polygon should be applied to all layers in the design.

This is an important question and the answer depends on what you are doing. If you are just defining the boundary of your site at every depth, you want to say *Yes*. This turns the polygon into a "cookie cutter" in both directions, above and below the current layer. The following shows what a cookie cutter polygon design does:



If you are interested in bounding something like a contaminant plume, however, you will want to say *NO*. You may want a different polygon design for each vertical layer.



Your polygon will now be active in the list of polygons, if selected; however, you won't see any polygons on your site until you draw them.

Drawing A Polygon

Notice that there are 3 drawing tools. To draw a polygon, select the appropriate tool.

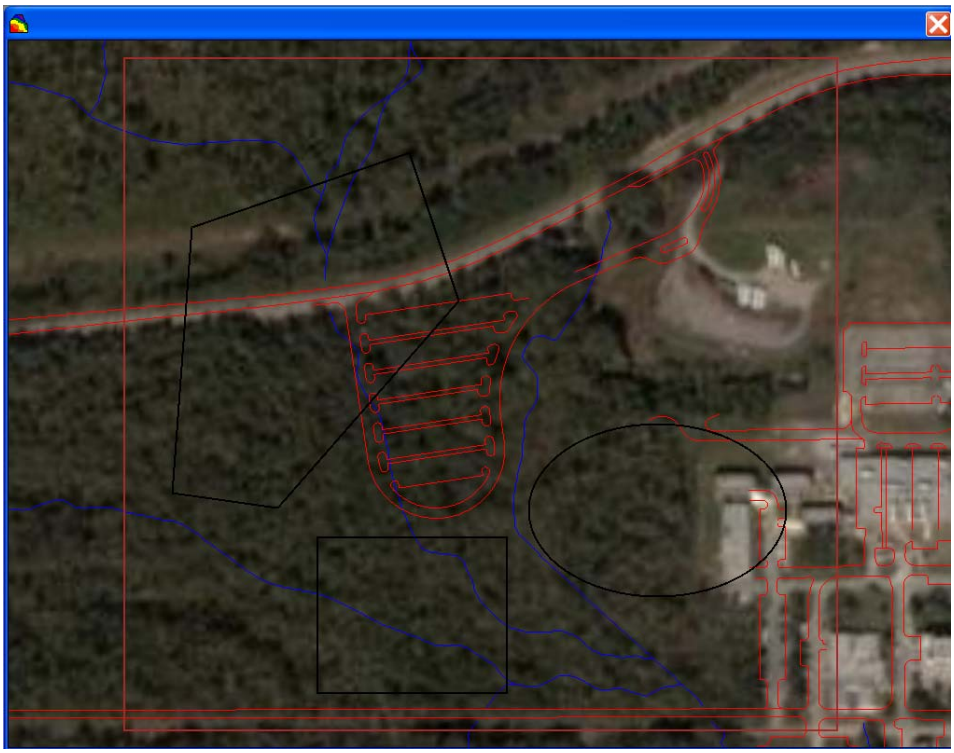
- Polygon – Allows the user to create a many-sided polygon to partition the data.
- Ellipse – Allows the user to create an ellipse to partition the data.
- Rectangle – Allows the user to create a rectangle to partition the data.

To start drawing, press the *Draw* button. It will change to *Done*. Notice that everything becomes disabled in SADA during polygon creation.

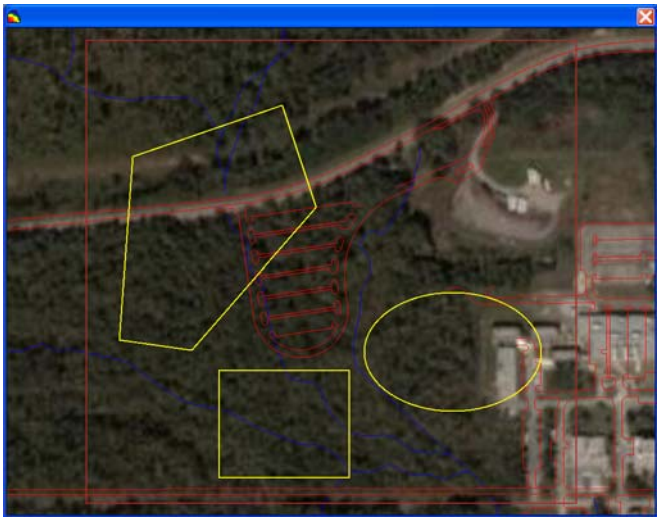
For the polygon tool, move the mouse over the map and note that the coordinates appear on the status bar at the bottom of the screen. Each time the left mouse button is clicked, a new vertex is added. To complete the polygon, double click the left mouse button. For the other two tools, hold the mouse button down and drag the mouse to create the ellipse or rectangle.

Add as many polygons per layer as desired. To change the layer, select the new layer from the layer box and add polygons to that layer in the same manner.

When you are finished with drawing, press the *Done* button. The polygon(s) will now display over the map and the rest of SADA will become enabled again.



At times, it might be difficult to see your polygon, particularly if you have a photograph behind it. To see your polygons better, you may want to change their color. From the main menu select *Graphics→Set Various Colors→Polygons*. From the color palette choose yellow and press *OK*. The shapes should be easier to see now.

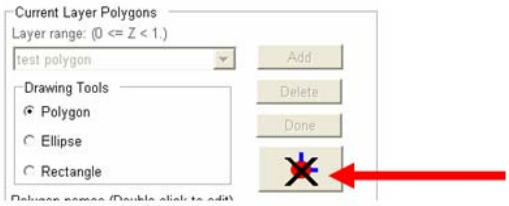


Polygons may be edited at any time during or after their creation.

Edit Polygons

After a polygon has been drawn, vertices can be removed, added, or moved to another location. Similarly, an entire polygon can be moved to a new location or deleted entirely. To get started, press the *Draw* button again. You can now continue drawing new polygons or edit existing polygons.

To edit an existing polygon, left-click the mouse to highlight the shape in a blue color. To select a particular vertex, left click inside the vertex and it will turn red. Notice when a vertex is red, a new button, the *Delete* button appears in the Polygon area.



To delete the vertex, press this *Delete* button.

To move the vertex, hold the left mouse button down inside the vertex, move to the desired location, and release. To add a new vertex, click anywhere on a line between two vertices.

To select an entire polygon, right click inside any vertex and the entire polygon will turn red. To delete the polygon, press the Delete button. To move the entire polygon, hold the right mouse button down inside any vertex, move to the desired location, and release.

When you have a polygon selected (blue vertices are showing), notice that the *copy polygon* button is now available on the main toolbar.

Simply press this button and SADA makes a copy of the polygon. You can now paste this shape anywhere you like: on the current design or in another polygon design altogether. When a polygon is available for pasting, the *paste polygon* button will also become enabled.

Simply press this paste button whenever you are in draw mode and SADA will drop it exactly on the horizontal position in which it was copied. Note that neither button will be enabled until you are in draw mode.

The same type of actions work for both rectangle and ellipse as well. The exception is that obviously you can't delete vertices, or you would no longer have a rectangle or ellipse.

In the following image, you can see each of the three types of polygons or shapes applied over a data set. Later we will work with data in depth, but for this discussion, everywhere you see a circle is where a sampled value was taken for Ac-225. The brighter the color, the higher the concentration.

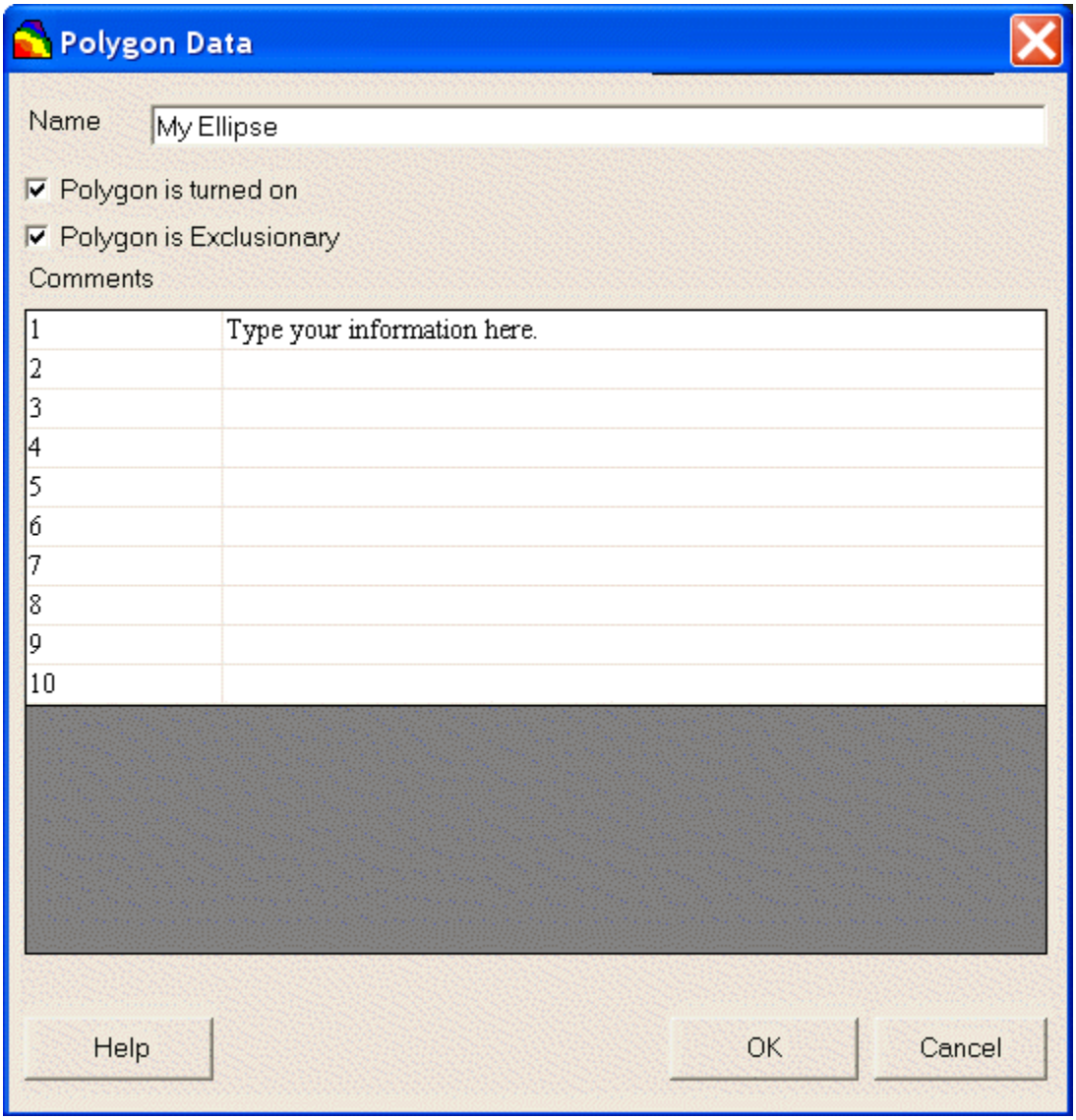
Polygons can be either inclusive or exclusive, and a data plot is the best way to demonstrate this principle. Notice that data found inside each shape is colored and data outside the polygons is white. The empty circles indicate that they are currently not included. When you draw a polygon, it is by default an inclusive polygon (meaning that anything inside of it is included). You can also have exclusive polygons, but to change a polygon from inclusive to exclusive or back again you'll need to read on about polygon properties.

Polygon Properties

You have a good deal of control over the individual polygons or shapes you create. To see the property window for any given shape, just double click the name for that polygon from the list of polygon names for the given polygon design.

Select the polygon collection from the drop-down box on the parameters window and double-click on the polygon name that appears in the list box below. You may have different shapes and may wish to use different names. This is OK.

The *Polygon Data* window will appear.



At the top of the window, you can change the name of your shape from its default (e.g., Ellipse1) to something more meaningful. In this example, it has been changed to MyEllipse.

Just below the name is the opportunity to turn this individual polygon off. Keep in mind that if you turn it off, it will be turned off in any layer for which you are using it. When a polygon is turned off, it isn't deleted; it's just literally turned off for the moment and doesn't affect any analysis.

Below this is the place to set a polygon to be exclusive. If a polygon is exclusive, then anything inside it is NOT included in the analysis. This is useful in carving out buildings away from the analysis of outdoor contamination. IMPORTANT: exclusive polygons take precedence over inclusive polygons. So, if you have a polygon that is inclusive overlapping on one that is exclusive, any point that is common to both will be UNINCLUDED.

You can also add comments about the polygon in the 10 rows of open text boxes. Click the *OK* button to close the window.

Returning now to the parameter window, if the *Show only included polygons* option is checked, then SADA will remove all "turned off" polygons from the graphics window. If this option is unchecked, all turned off polygon names will still appear as a gray outline in the graphics window. These polygons, however, will not be included in any analyses and only remind users that they are there.

The option *Use only selected data during interpolation* is very important. Sometimes, even though a section of the site, such as polygonal area or layer, has been turned off/exclusionary, you may not want to interpolate in those areas, but the data from those areas maybe crucial to the understanding of the spatial process. In most applications, this is the case. For example, if some of your data were say inside a building and some were outside a building, it may make sense to not include the inside data to estimate the outside values. In this case, make sure that this option is selected.

To turn off the polygon collection, simply select *None* from the main *Polygon* layer list.

Elevation Grid

At the bottom of the parameters window is an opportunity to select an elevation grid that you may have imported. If you have one you would like to use, select it now.

There will be no apparent change in the interface. But the next time SADA draws a result, it will include elevation. If you want to see your elevation data, you will need to select it by choosing *Elevation* from the data type box and your elevation dataset from the dataset box. Elevation is only applied to a result following its analysis. In other words, all geospatial models are modeled as a function of depth below surface. This is advantageous for a number of reasons. Later, the model is corrected for elevation in the 3d viewer.

Summary

This fairly lengthy chapter showed how to create a SADA file, use GIS layers to set up the site, create vertical layers, and use polygons to further aggregate our site. The next few chapters will talk about how to import different kinds of data into your file. You can also use imported data to help set up your site. We'll show how to do that as well, briefly revisiting some the principles discussed here.

Part II: Importing Data

Chapter 5: Importing Sampled Data

There are five types of data that can be imported or created in SADA: sampled data (discrete sampled values), gridded data (models created outside SADA), elevation data (formulated outside SADA, digital elevation models, etc.), and two user created types called “standard” and “probability.” In this chapter, we will concentrate on importing sampled data.

Getting Your Data Ready

Sampled data are the most common form of data users will likely use. These are the usual sample measurements taken at specific points around your site. Before SADA can import the data values, it needs the data in a particular format. Fortunately, the format is very easy to work with; most of the time environmental data comes in this type of configuration. Simply put, data needs to be in a tabular format with each row representing a single measurement value of a single contaminant (or attribute). Each column is information about the measurement, such as the easting coordinate, the northing coordinate, depth, name of the contaminant, measured value at that location, and so forth. Many times, users look at this data with a spreadsheet program like Excel. Here is an example data set in Excel.

	A	B	C	D	E	F	G	H	I	J	K	L
1	Easting	Northing	Depth	Casnumber	Name	Value	Detected	Media	Date	SampleID	Region	
2	27596.25	21900	0	14265851	Ac-225	2	1	SO	9/26/1970	1	1	
3	28310.25	21900	0	14265851	Ac-225	1.6	1	SO	9/26/1970	2	1	
4	28935	21900	0	14265851	Ac-225	0.9	1	SO	9/26/1970	3	1	
5	27685.5	22200	0	14265851	Ac-225	2	1	SO	9/26/1970	4	1	
6	28131.75	22200	0	14265851	Ac-225	4.2	1	SO	9/26/1970	5	1	
7	29202.75	22500	0	14265851	Ac-225	1.5	1	SO	9/26/1970	6	1	
8	27150	23160	0	14265851	Ac-225	1.7	1	SO	9/26/1970	7	1	
9	27685.5	22920	0	14265851	Ac-225	3.6	1	SO	9/26/1970	8	1	
10	28042.5	23100	0	14265851	Ac-225	4.9	1	SO	9/26/1970	9	1	
11	28221	23100	0	14265851	Ac-225	4.2	1	SO	9/26/1970	10	1	
12	28667.25	23220	0	14265851	Ac-225	2.9	1	SO	9/26/1970	11	1	
13	29113.5	22980	0	14265851	Ac-225	1	1	SO	9/26/1970	12	1	
14	27417.75	23580	0	14265851	Ac-225	1.9	1	SO	9/26/1970	13	1	
15	27774.75	23640	0	14265851	Ac-225	2.9	1	SO	9/26/1970	14	1	
16	28310.25	23400	0	14265851	Ac-225	3.1	1	SO	3/22/1993	15	1	
17	28935	23460	0	14265851	Ac-225	0.8	1	SO	3/22/1993	16	1	
18	28200	22560	0	14265851	Ac-225	4.8	1	SO	3/22/1993	17	1	
19	28700	22500	0	14265851	Ac-225	3.3	1	SO	3/22/1993	18	1	
20	27200	22380	0	14265851	Ac-225	2.03	1	SO	3/22/1993	19	1	
21	28984	22787	0	14265851	Ac-225	2.6	1	SO	3/22/1993	4	1	
22	27350	22750	0	14265851	Ac-225	2.5	1	SO	3/22/1993	3	1	
23	27026	22129	0	14265851	Ac-225	1.7	1	SO	3/22/1993	2	1	
24	27690	23350	0	14265851	Ac-225	2.6	1	SO	3/22/1993	4	2	
25	27500	23270	0	14265851	Ac-225	2.5	1	SO	3/22/1993	44f	2	
26	28709	22900	0	14265851	Ac-225	2.7	1	SO	3/22/1993	43	2	
27	27650	22500	0	14265851	Ac-225	3.4	1	SO	3/22/1993	33	2	
28	28530	22700	0	14265851	Ac-225	3.3	1	SO	3/22/1993	11	2	
29	27750	23145	0	14265851	Ac-225	3.2	1	SO	3/22/1993	12	2	
30	27596.25	21900	0	7440393	Barium	42.7	1	SO	8/7/1996	13	2	
31	28310.25	21900	0	7440393	Barium	35.1	1	SO	8/7/1996	14	2	
32	28935	21900	0	7440393	Barium	18.4	1	SO	8/7/1996	15	2	
33	27685.5	22200	0	7440393	Barium	43.5	1	SO	8/7/1996	16	2	
34	28131.75	22200	0	7440393	Barium	87.9	1	SO	8/7/1996	17	2	

SADA can import measured values using two different kinds of files. A Comma Separated Value (.csv) file is the most commonly used file type. Anyone with a copy of Excel or a similar

spreadsheet program can easily save their file as a .csv file and SADA can read it. Even if you don't have a spreadsheet program, you can use simple text editors (e.g., Notepad that comes with Windows) to create or format your data set. If you are using a .csv file, the first row must be the title row. Punctuation is not allowed in the column (field) names. The second way to import measured values is to use a Microsoft Access database. Let's examine each of the types of information now.

Easting and Northing Columns (required)

SADA must have these two columns. They tell where on the site your sample was collected. SADA does not require or enforce the use of any predefined coordinate system. There are a number of them available, such as UTMs, state plane, lat/long, and so forth. Many sites have their own localized system. In some cases, you'll see sample measurements relative to the corner of the site or a building. These are all OK. (Note to lat/long users: you'll need to make sure you use the negative of your latitude values or your data will plot backwards.)

What is important to note here is that these values must be in the same coordinate system as any GIS layer you plan on using. For example, the roads.dxf layer seen throughout these examples is in State Plane Coordinates for Tennessee. You would need to make sure that the easting and northing coordinates are using the same system or there will be big problems. Namely, your data will not plot correctly on your GIS layer. Now, if you are not using GIS, then you have no worries with your coordinate system.

Blank values and non-numeric values are not permitted in either of these columns.

Depth Column (may or may not be required)

If the data are taken over depth, an additional column with the depth value is required. No empty values are allowed for any of these columns, and non-numerical values are not permitted for any coordinate or sample value. There are two important things to check about your depth column. First, make sure your depth is *depth below surface* and NOT elevation measurements. SADA operates on the concept of depth below surface. Elevation is used in SADA, but only as a 3d view feature. Second, make sure that your depth values are in the same units as your horizontal extents. For example, it is nonsensical to use lat/long for horizontal and meters for depth. Later on, when you do 3d geospatial analysis, this really has bizarre implications. If the horizontal measurements are in feet, make sure the depth is in feet and so forth. SADA is not going to do any conversion for you.

Sampled Values (required)

This field is where your measurement value is found. There are no units enforced by SADA because of the many and varied types of applications for SADA. You will need to make sure that if you are doing human health or ecological risk assessment in SADA you MUST have your measured units as follows.

- Soil, Sediment, and Biota: mg/kg for nonradionuclides, pCi/g for radionuclides
- Surface/Groundwater: mg/L for nonradionuclides, pCi/L for radionuclides

The sample values column cannot contain any blanks or non-numeric values. If a measurement value was a nondetect, you must enter the detection limit into the value column. For example, you cannot enter for example <5. You would instead enter a value of 5 and then not enter this in the detection field (see below). Values like NA are obviously not allowed either.

Name (required)

This column provides the name of the attribute (contaminant) you are measuring in each record. The field cannot contain any blank values. When you are working with environmental contaminants, it is common to work with contaminant names that contain commas. These are a serious problem for comma delimited files, as you might imagine. For example, if you have the contaminant name 1,2,3-Trichlorobenzene, it would be recognized as 3 separate columns: 1, 2, and finally 3-Trichlorobenzene. This will cause an error in import. Typically, this error states that SADA has reached the end of the file unexpectedly. To correct this, make sure your name field contains the name within quotation marks: "1,2,3-Trichlorobenzene." Spreadsheet programs like Excel normally do a good job of taking care of this for you by placing quotations around the contaminant name. To be sure, however, you may want to open up your .csv file in a simple text editor, such as Notepad (shipped with Windows), to make sure that you are not running into this problem. Typically, SADA will complain about reaching the end of the file too soon in this situation.

It is also important to make sure that each contaminant is consistently named. For example, make sure you don't have 1,2,3-Trichlorobenzene also spelled as Trichlorobenzene-1,2,3. SADA may see these as two different contaminants.

CAS Numbers (not required)

A column containing the CAS numbers can be useful when you are setting up different kinds of analyses, such as human health. It isn't required, though. If you do have CAS numbers and want to use them, make sure that they do not contain any dashes or spaces. This will render them virtually useless later as SADA tries to match them up to its various data bases. A simple search and replace in the CAS column using Excel can take care of this pretty quick.

Detection Qualifiers (not required)

Another field of some importance is the detection field. Instead of the normal Us and Js and so forth, SADA works with either a zero value (sample was below detection limit) or a 1 (sample was detected). As the user, you must decide whether a value should be considered detected or not before bringing it into SADA. This involves an area of the environmental process we are not involved in; however, we do provide some guidance for risk assessment in the tables below if you have no other instructions.

Flag	Meaning	Use for risk?
R	Rejected	No
B	Blanks contaminated	Treat as non-detect
J	Estimated	Yes, treat as detect
UJ	Estimated non-detect	Yes, treat as non-detect
K	Biased high	Yes, treat as detect
L	Biased low	Yes, treat as detect
U	Non-detect	Yes, treat as non-detect

Now for measured values that are considered non-detects, you should enter the detection limit of the instrument into the column containing measured values. Blank or missing values in the value column are not permitted.

Date Column (not required)

The date field is not required; but if you do use it, later you can query your data by date to see how things are evolving over time. This can be useful in monitoring situations. If you do use the date field you must use some combination of the m, mm, d, dd, yy, yyyy format. Examples include m/d/yy and mm/dd/yyyy. Dates that include any text (other than /) are not permitted. For example, September 26th, 2004 is not permitted. You would need to use 9/26/2004. In Europe and other areas that use day/month/year format, this should be ok as long as your data set and your operating system are consistent.

Media (may be required)

In the event that you plan on doing either a human health or ecological risk assessment, you MUST have a media column. Proper media identification qualifiers are as follows: Soil – SO, Sediment – SD, Groundwater – GW, Surface water – SW, Air – AIR, Biota – BIO, and Background – Background.

Beyond these columns, you can include any other information you like, but the total number of columns may not exceed 250.

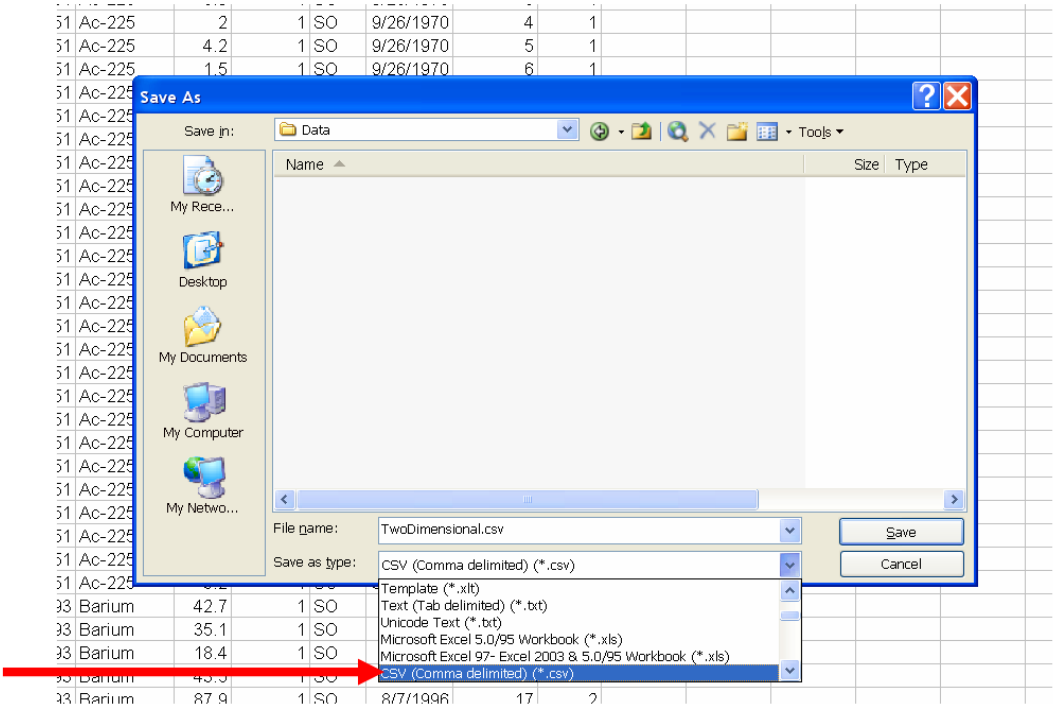
Final Tips

Now a few general requirements regardless of the type of information follow. First, you must use quotation marks around items that contain a comma. As mentioned earlier in the description of the Name column, you might have a contaminant name like 1,2,3-Trichlorobenzene that would be recognized as 3 separate columns (1, 2, and finally 3-

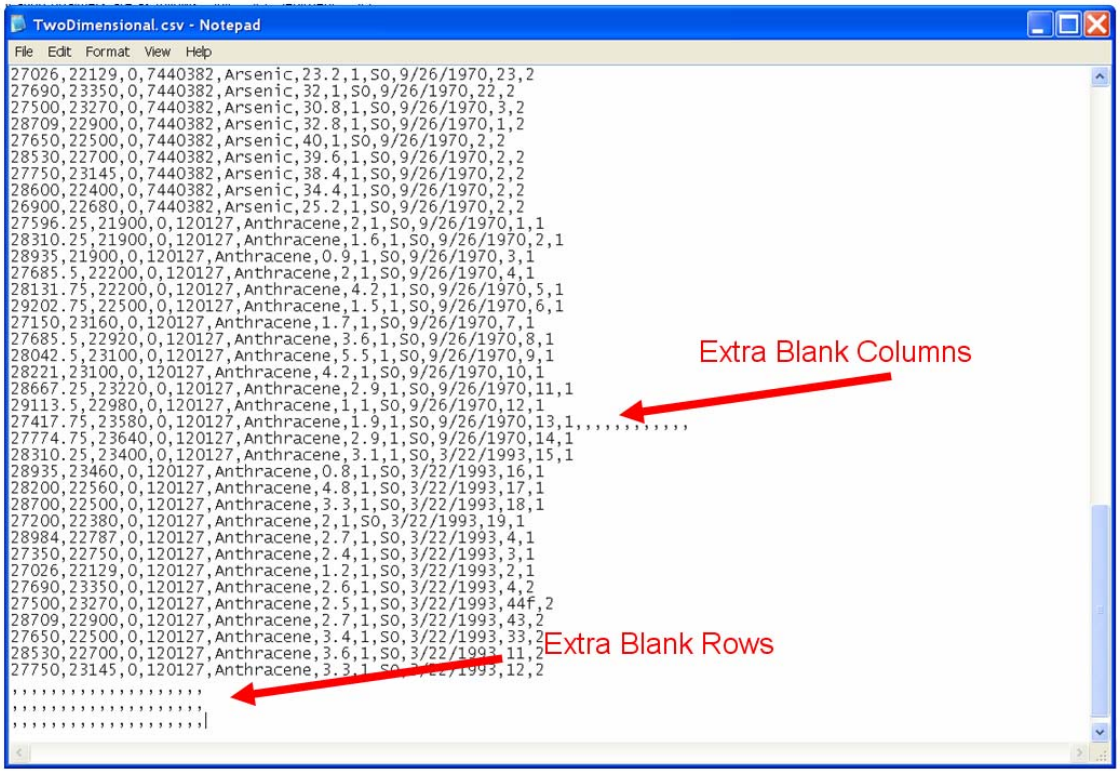
Trichlorobenzene). This will cause an error in import. Typically, this error states that SADA has reached the end of the file unexpectedly. To correct this, make sure your name field is contained within quotation marks: “1,2,3-Trichlorobenzene.” Spreadsheet programs like Excel normally do a good job of taking care of this for you by placing quotations around the contaminant name; however, you may want to open up your .csv file in a simple text editor, such as Notepad (shipped with Windows), to make sure that you are not running into this problem.

Quotations themselves can be problematic. For example, suppose you had a column that described where the sample was taken. Suppose for one of the samples this column’s value says *Sample located on "C" Street.* Because of the quotation marks around C, SADA sees this as three separate column values. So it accepts it as: *Sample located*, *C*, and finally *Street.* In this situation, make sure your field value reads *Sample located on C Street*. This is not something a spreadsheet like Excel can anticipate. You will need to make sure that this kind of thing is cleaned up on your own.

Saving your file as a .csv out of Excel is relatively easy. Open Excel, open your file, select the sheet you want, and select *File→Save As…* The names vary for .csv files, but it is usually easy to determine the right type to select.



Unfortunately, spreadsheet programs can do odd things when saving to a .csv file. For example, extra commas may be added and entire extra rows added with nothing in them. It is well worth your time to open this .csv file in a simple text editor, like Notepad (comes with Windows), and look at the data yourself for a final check. The following image shows the type of common problems you may encounter. These are simple fixes that when left uncorrected can cause great grief during the import process.



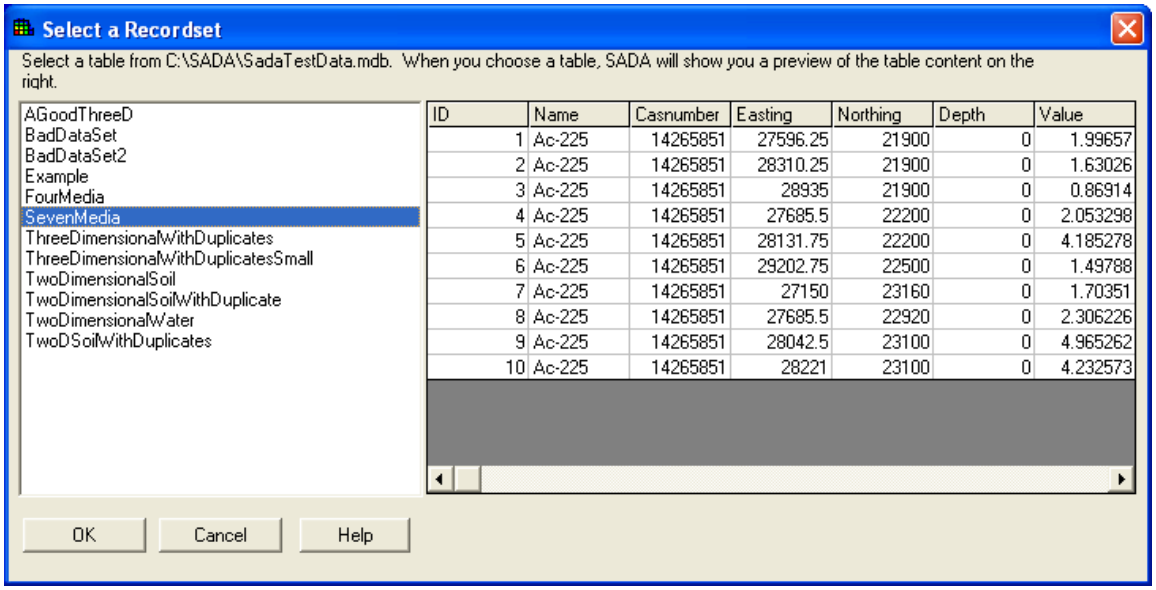
Use your text editor to delete or correct any such anomalies and then proceed with import.

Import the Data

We will show this process with a simple example. Begin by opening SADA and creating a SADA file called MyPointData.sda. This will give us a clean blank slate to work from.

There are two ways to add new sampled values. Choose from the main menu *Data→Import* Sampled data. SADA responds with the Open file window. The next step is to use the window to locate the Microsoft Access or comma-delimited ASCII text file that contains the sample data and select it by double-clicking the name (or type the name of the file in the file name box). Navigate to where you've installed SADA (probably C:\Program Files\SADA 5.0). Note that the external file itself is not affected by the conversion process.

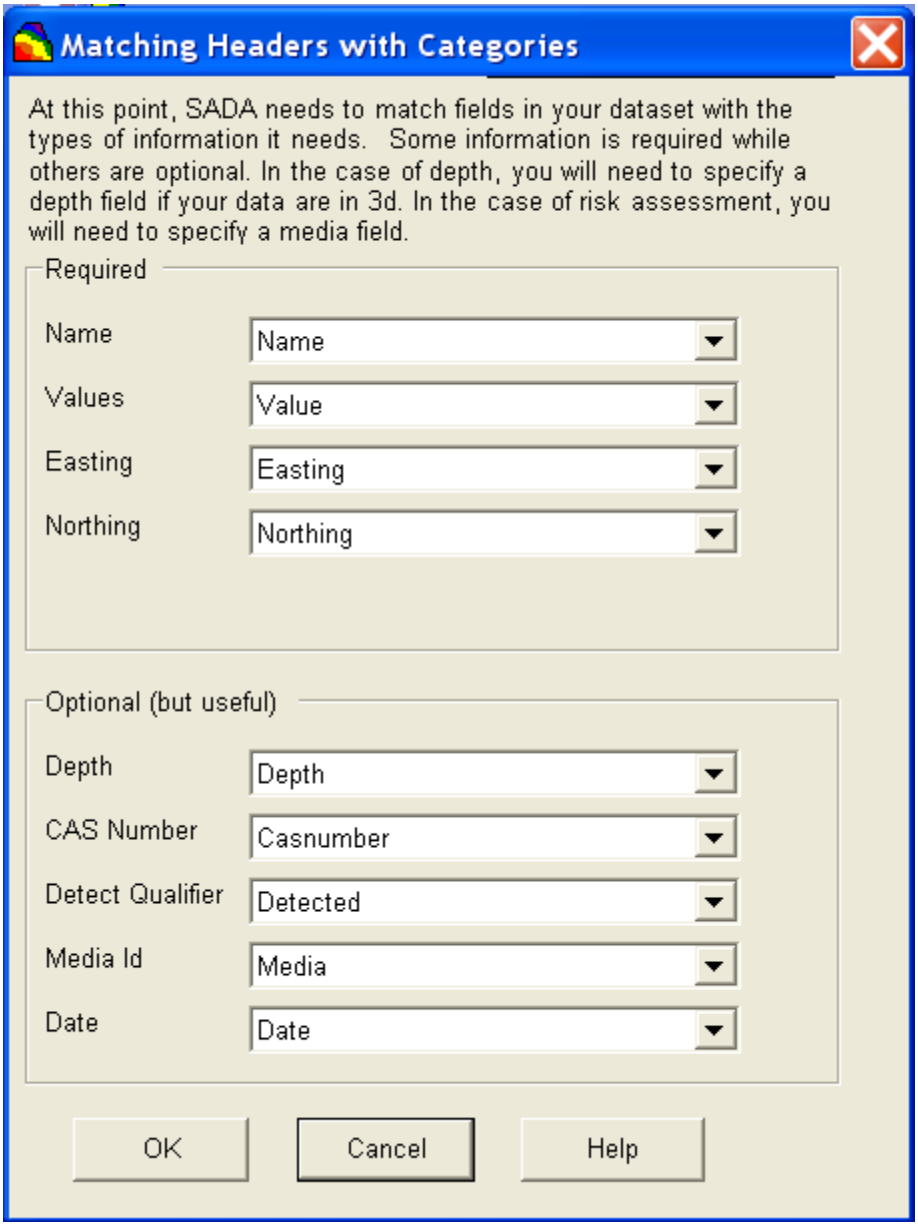
After you press *Open* (we will do this in a moment, just keep reading), if your input file is an Access database, the following window will appear:



The pane on the left shows all the available tables in your database. You should select the table with the needed information and press *OK*.

The next step in the process, or the first step if your dataset is a .csv file, is to identify the columns of information in the ASCII data file and match these columns of information to information categories that are required or may be useful in SADA.

In our current example, select the .csv file TwoDimensional.csv and press *Open*. You should be presented with the matching headers window.



At the top of the window is some useful text to help remind you what's going on. This step is actually fairly simple. There are certain fields that SADA requires and others that are sometimes required or perhaps just helpful in some way. These are divided into the two boxes you see on this window. The first box holds the required fields. You must have a *Name* field, a *Value* field, an *Easting* field, and a *Northing* field. On the right side you see in each drop down list the names of the fields in your data set. Match the correct field in your data set to what SADA needs for each piece of required information.

The lower box holds some useful and sometimes required fields (depending on what you want to do). The same principles apply. On the left are types of information SADA recognizes. On the right are the corresponding fields. Notice that each of the drop down lists in this group includes the option "none." This is because you may not have a field that corresponds to the information SADA is asking about.

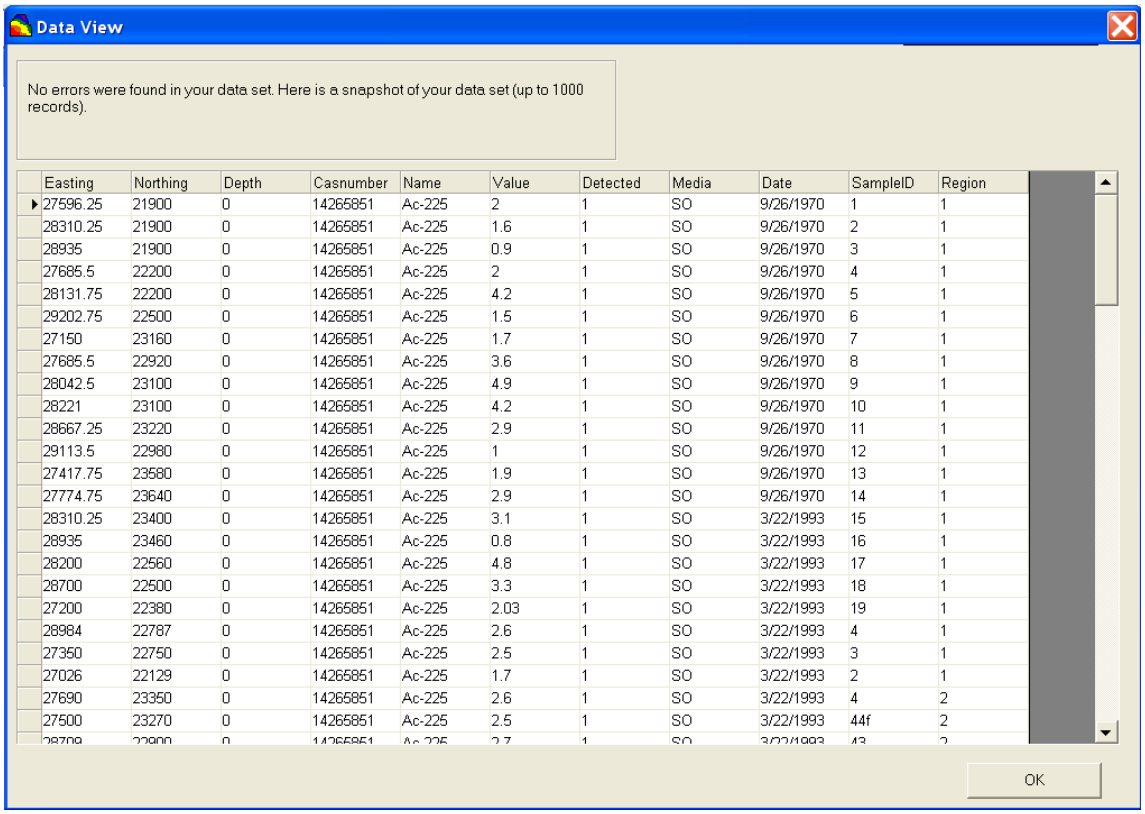
In particular, the depth category is required only when data exist at varying depths. If the *Detect Qualifier* is not assigned, the data are assumed to be all detects. If the media column is not assigned, SADA adds an artificial media column titled *Basic*.

WARNING:

If Media ID, which denotes the type of media the contaminants are sampled in (e.g., soil or groundwater) is not defined, then the human health risk and/or ecological risk modules cannot be setup later. The media is a critical information category to the risk modules. Also, remember that SADA expects certain units for measured values in the risk modules.

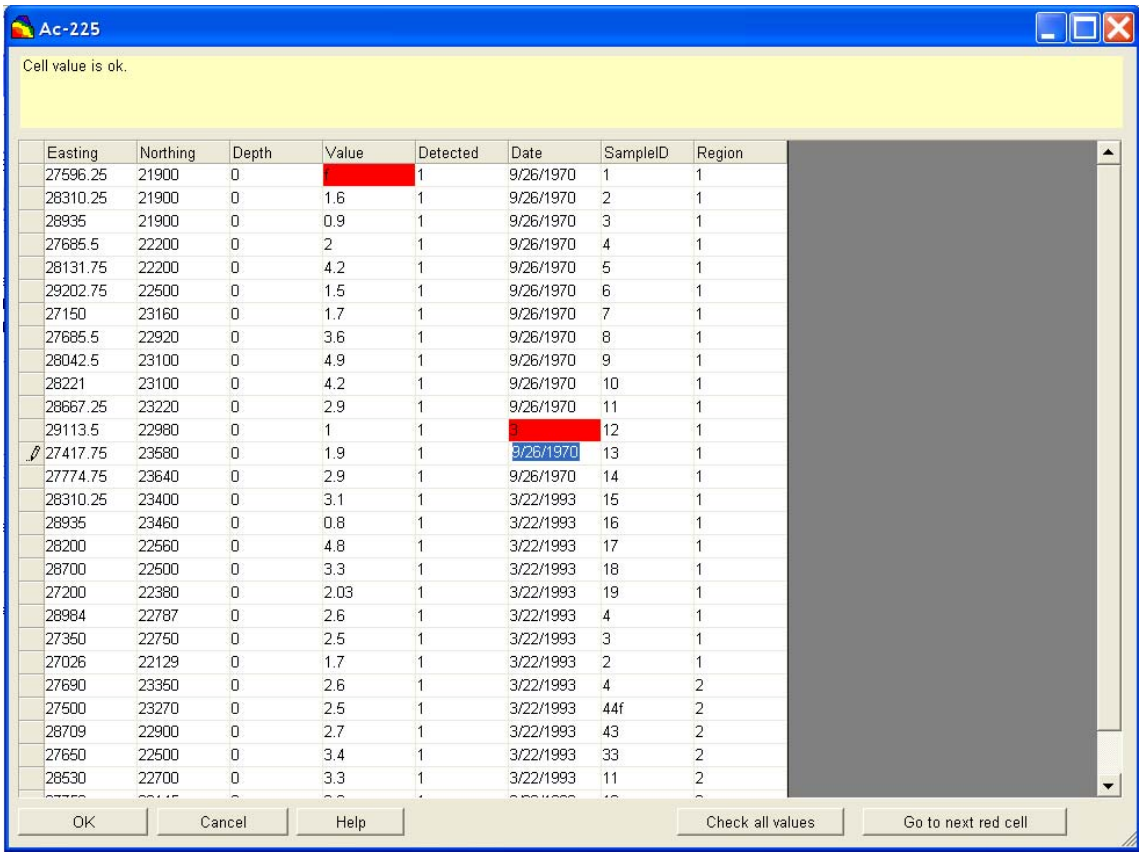
In addition to these pieces of required and optional information, you can bring in up to 250 fields of metadata from your record set as part of the import process. Later you can see them in a variety of ways.

After the columns have been set, press *OK***.** SADA begins the import process. If your data set contains no errors, you will be presented with a snap shot of the data set (up to 1000 records). For our current import example, you should get the following screen.



Easting	Northing	Depth	Casnumber	Name	Value	Detected	Media	Date	SampleID	Region
27596.25	21900	0	14265851	Ac-225	2	1	SO	9/26/1970	1	1
28310.25	21900	0	14265851	Ac-225	1.6	1	SO	9/26/1970	2	1
28935	21900	0	14265851	Ac-225	0.9	1	SO	9/26/1970	3	1
27685.5	22200	0	14265851	Ac-225	2	1	SO	9/26/1970	4	1
28131.75	22200	0	14265851	Ac-225	4.2	1	SO	9/26/1970	5	1
29202.75	22500	0	14265851	Ac-225	1.5	1	SO	9/26/1970	6	1
27150	23160	0	14265851	Ac-225	1.7	1	SO	9/26/1970	7	1
27685.5	22920	0	14265851	Ac-225	3.6	1	SO	9/26/1970	8	1
28042.5	23100	0	14265851	Ac-225	4.9	1	SO	9/26/1970	9	1
28221	23100	0	14265851	Ac-225	4.2	1	SO	9/26/1970	10	1
28667.25	23220	0	14265851	Ac-225	2.9	1	SO	9/26/1970	11	1
29113.5	22980	0	14265851	Ac-225	1	1	SO	9/26/1970	12	1
27417.75	23580	0	14265851	Ac-225	1.9	1	SO	9/26/1970	13	1
27774.75	23640	0	14265851	Ac-225	2.9	1	SO	9/26/1970	14	1
28310.25	23400	0	14265851	Ac-225	3.1	1	SO	3/22/1993	15	1
28935	23460	0	14265851	Ac-225	0.8	1	SO	3/22/1993	16	1
28200	22560	0	14265851	Ac-225	4.8	1	SO	3/22/1993	17	1
28700	22500	0	14265851	Ac-225	3.3	1	SO	3/22/1993	18	1
27200	22380	0	14265851	Ac-225	2.03	1	SO	3/22/1993	19	1
28984	22787	0	14265851	Ac-225	2.6	1	SO	3/22/1993	4	1
27350	22750	0	14265851	Ac-225	2.5	1	SO	3/22/1993	3	1
27026	22129	0	14265851	Ac-225	1.7	1	SO	3/22/1993	2	1
27690	23350	0	14265851	Ac-225	2.6	1	SO	3/22/1993	4	2
27500	23270	0	14265851	Ac-225	2.5	1	SO	3/22/1993	44f	2

On the other hand, if there were errors in the data set, you will be presented with the data editor and the complete data set will be presented to you along with any errors highlighted in red. You will need to correct these errors before proceeding.



A detailed explanation of how to correct these kinds of errors in the data editor is provided in Chapter 9, “Managing Data.” Briefly speaking, the process is simple. The data editor operates much like a spreadsheet, allowing you to change values in order to meet different data requirements. Whenever you do not meet a particular data requirement, the particular cell will appear red. If you hold your cursor over it, the yellow box at the top will tell you what’s wrong. Now, if you have a lot of red cells, you probably want to Cancel the process and return to your normal spreadsheet or data management system to correct the errors. This simple spreadsheet is designed only with the idea that a handful of errors need to be corrected.

Tips: First of all, any errors corrected in this sheet do not change the original data values in your .csv or .mdb file. Remember that SADA grabs the data all at once and pulls it into your SADA file, where you can further manage it. It might be better in the long run to cancel the process and fix the errors in the source data file.

Second, if you are presented with the data editor and an entire column is red, it is quite likely that you have mismatched columns. For example, you may have set the *Easting* column to be the *Name* column in your data set. You’ll want to cancel in that case and start over. The data editor will parenthetically tell you in each column what it considers the column to be. So you wouldn’t want to see something like *Names* (= Easting). This would mean you assigned the name to your *Easting* column.

After the errors have been corrected or the snap shot has been closed, SADA will proceed with the import. For very large data sets, you may be given the opportunity to "thin" the data some.

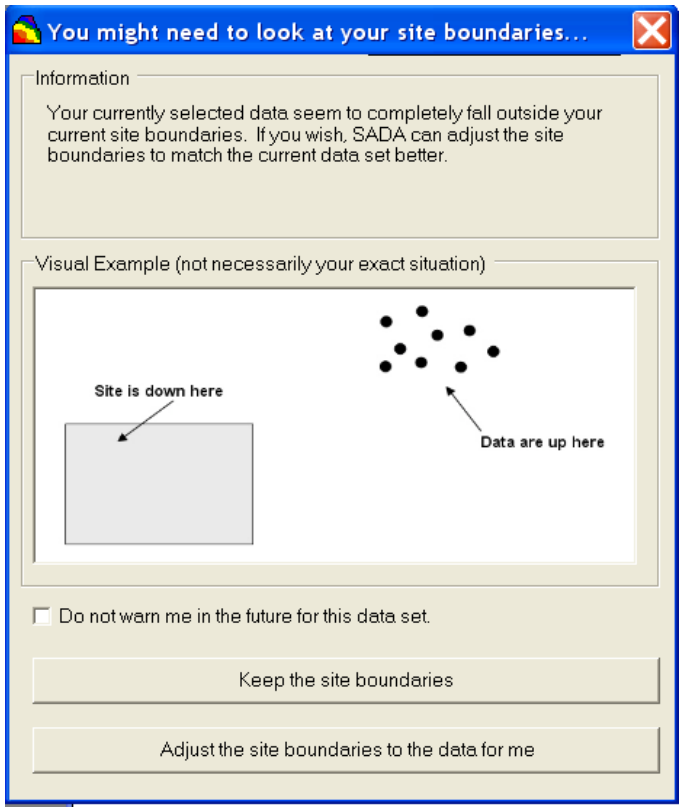
Thinning The Data

As SADA finishes the import process, you may be provided a warning about your site boundaries. This is common and you can find a discussion of this in Chapter 8, “Importing Elevation Data.” The same principles apply here and anywhere large data sets are imported into SADA. For this discussion, we’ll move forward with our smaller data set.

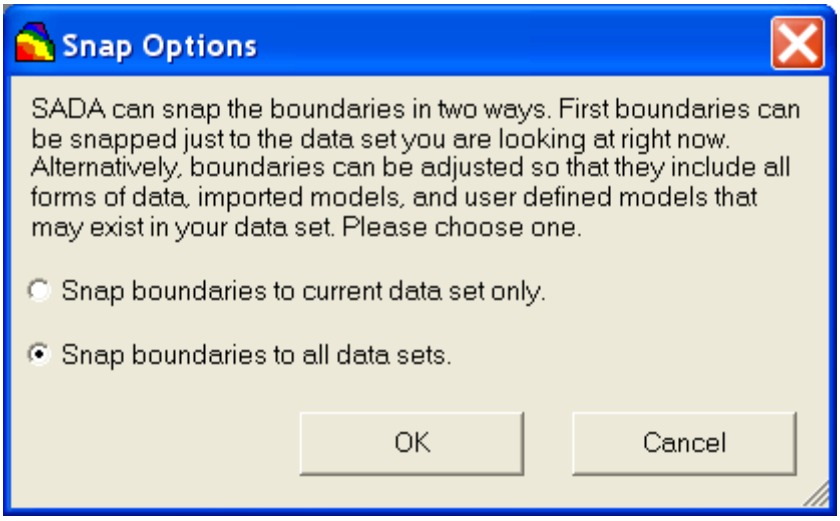
Adjusting Site Boundaries

If you import information into your SADA file that has horizontal spatial extents that are beyond the current site boundaries, you will be presented with an opportunity to change the site boundaries to include the current data. This often happens when you have just created a new SADA file and you are using the dummy site (see Chapter 4, “Setting up a SADA File”).

This is an action that you may or may not wish to take. For example, certain elevation sets exceed site boundaries, but you do not want to include the entire elevation set. When SADA sees this situation, you may be presented with the following opportunity:

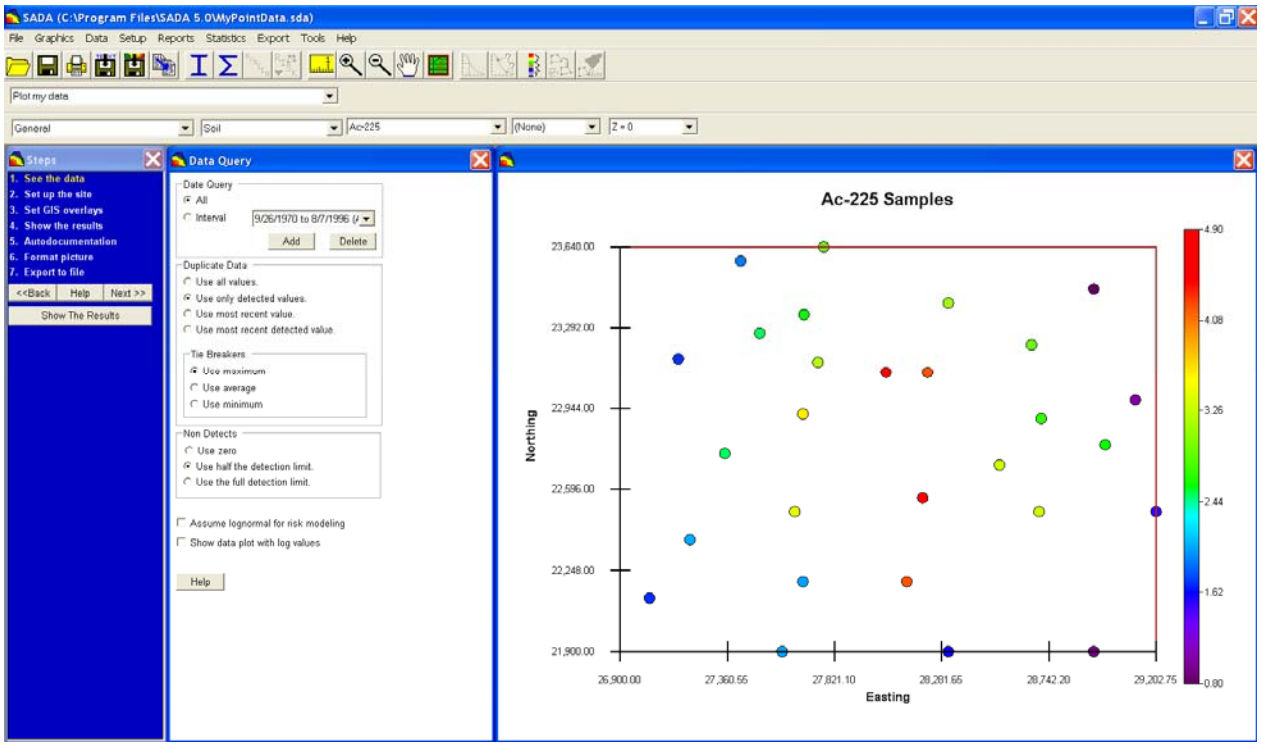


You may select either *Keep the site boundaries* or *Adjust the site boundaries to the data for me*. For data sets that you regularly visit but do not wish to be warned about, select the box for *Do not warn me in the future for this data set*. If you regret selecting this, you can reverse it by visiting *Tools→Warning Options→Site Boundary Problems→Turn Warnings On*. If you choose to *Adjust site boundaries*, you are presented with another option.



Here, you can choose *Snap the boundaries to the current data set only* or snap them once-and-for-all to all data sets found in your SADA file: *Snap boundaries to all data sets*. This will automatically update your *Site Boundaries* information found in your *Setup the site* step. Further information on adjusting site boundaries is available in Chapter 4, "Setting up a SADA File."

When you are presented with this option in our current example, agree to *Adjust the site boundaries* and select *Snap boundaries to all data sets*. SADA should now look similar to the following image:



The first thing you should do is save your file. If you are interested in learning how to import external models or elevation, please continue to the next chapter. If you would like to jump ahead and learn how to work with your data, please skip to Chapter 9, "Managing Data."

Chapter 6: Import New Data Over Existing Data

The import process can be done again and again on the same SADA file, regardless of the data type (e.g., sampled, gridded, elevation, etc.). Also, there is no difference whether the file contains previous data or not. For sampled data, with its richer array of information types, however, some additional things need to be considered. If your file already contains some sampled point values and you wish to import more sample data values, you will see a slightly different import process.

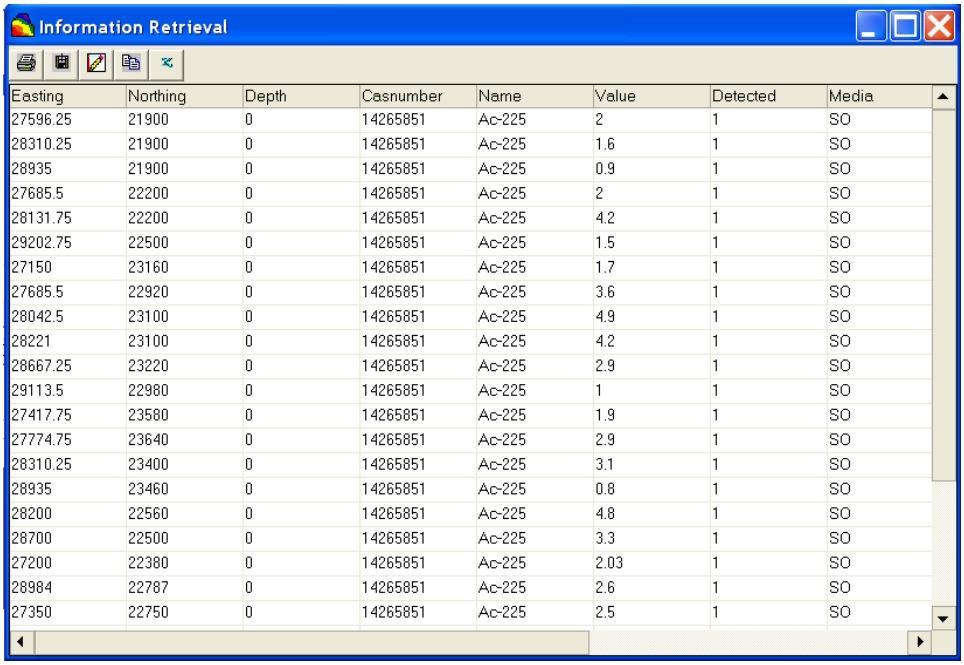
Let's begin by opening SADA (if it is not already open). Then open the file Twodimensional.sda from the SADA installation directory. To begin, Save the File as TwoDimensionalWithMore.sda so we don't affect the original file. This file already has the two dimensional soil data set imported from twodimensional.csv. We will now import a new set of three-dimensional groundwater data into the file. In this new data set, there is only one data set of Chlordane measurements.

Select from the main menu *Data→Import Sampled Data* and choose from the file ThreeDimensional.csv from the SADA installation directory (likely C:\Program Files\SADA 5.0).

SADA will first scan your file and identify the major column headers. The file looks like this in Excel.

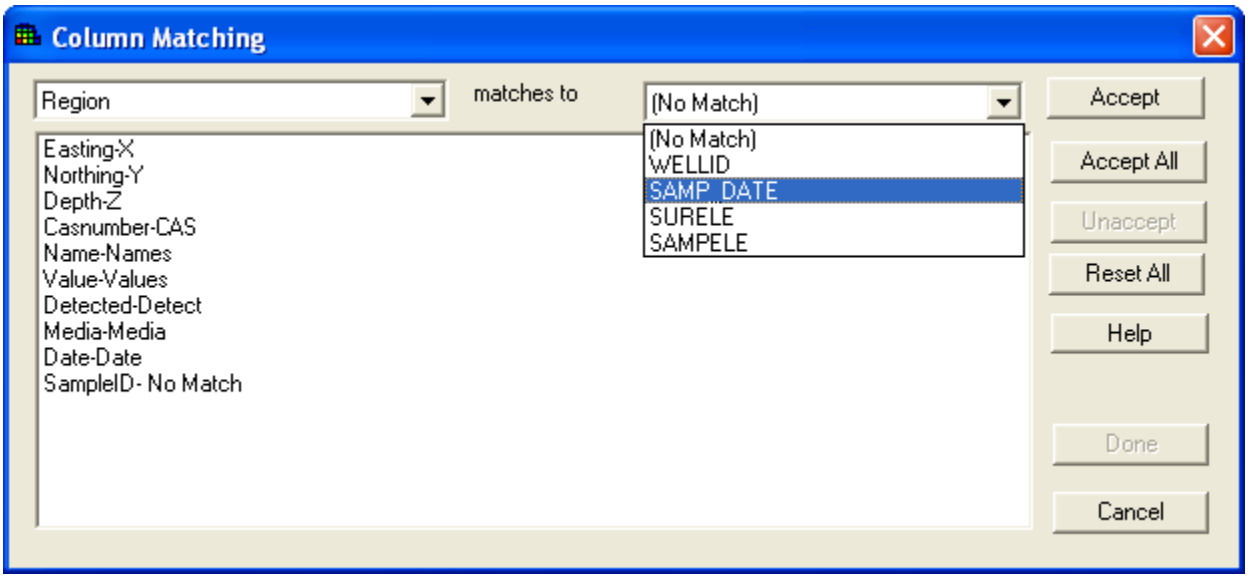
	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	X	Y	Z	CAS	Names	Values	Detect	Media	Date	WELLID_[	WellID	SAMP_DA	SURELE	SAMPELE
2	27273.89	23261.16	2.6	1234123	Chlordane	0.00015	1	GW	3/22/1994	DP-102-1	DP-102	4/8/1997	89.8	32
3	27273.89	23261.16	5.2	1234123	Chlordane	0.00027	1	GW	3/22/1994	DP-102-1	DP-102	4/8/1997	89.8	27
4	27273.89	23261.16	7.8	1234123	Chlordane	0.00048	1	GW	3/22/1994	DP-102-1	DP-102	4/8/1997	89.8	22
5	27273.89	23261.16	10.4	1234123	Chlordane	0.00045	1	GW	3/22/1994	DP-102-1	DP-102	4/8/1997	89.8	17
6	27273.89	23261.16	13	1234123	Chlordane	0.00028	1	GW	3/22/1994	DP-102-1	DP-102	4/8/1997	89.8	12
7	27273.89	23261.16	15.6	1234123	Chlordane	0.00012	1	GW	3/22/1994	DP-102-1	DP-102	4/8/1997	89.8	7
8	27273.89	23261.16	18.2	1234123	Chlordane	0.00004	1	GW	3/22/1994	DP-102-1	DP-102	4/8/1997	89.8	2
9	27273.89	23261.16	20.8	1234123	Chlordane	0	1	GW	3/22/1994	DP-102-1	DP-102	4/8/1997	89.8	-3
10	27273.89	23261.16	23.4	1234123	Chlordane	0	1	GW	3/22/1994	DP-102-1	DP-102	4/8/1997	89.8	-8
11	27273.89	23261.16	26	1234123	Chlordane	0	1	GW	3/22/1994	DP-102-2	DP-102	4/8/1997	89.8	-13
12	27512.91	23260.99	2.6	1234123	Chlordane	0.46147	1	GW	3/22/1994	DP-103-1	DP-103	4/8/1997	89.8	32
13	27512.91	23260.99	5.2	1234123	Chlordane	0.69144	1	GW	3/22/1994	DP-103-1	DP-103	4/8/1997	89.8	27
14	27512.91	23260.99	7.8	1234123	Chlordane	0.73024	1	GW	3/22/1994	DP-103-1	DP-103	4/8/1997	89.8	22
15	27512.91	23260.99	10.4	1234123	Chlordane	0.38323	1	GW	3/22/1994	DP-103-1	DP-103	4/8/1997	89.8	17
16	27512.91	23260.99	13	1234123	Chlordane	0.12108	1	GW	3/22/1994	DP-103-1	DP-103	4/8/1997	89.8	12
17	27512.91	23260.99	15.6	1234123	Chlordane	0.02729	1	GW	3/22/1994	DP-103-1	DP-103	4/8/1997	89.8	7
18	27512.91	23260.99	18.2	1234123	Chlordane	0.00492	1	GW	3/22/1994	DP-103-1	DP-103	4/8/1997	89.8	2
19	27512.91	23260.99	20.8	1234123	Chlordane	0.00076	1	GW	3/22/1994	DP-103-1	DP-103	4/8/1997	89.8	-3
20	27512.91	23260.99	23.4	1234123	Chlordane	0.00011	1	GW	3/22/1994	DP-103-1	DP-103	4/8/1997	89.8	-8
21	27512.91	23260.99	26	1234123	Chlordane	0.00001	1	GW	3/22/1994	DP-103-2	DP-103	4/8/1997	89.8	-13
22	27512.91	23260.99	28.6	1234123	Chlordane	0	1	GW	3/22/1994	DP-103-2	DP-103	4/8/1997	89.8	-18
23	27512.91	23260.99	31.2	1234123	Chlordane	0	1	GW	3/22/1994	DP-103-2	DP-103	4/8/1997	89.8	-23
24	27782.53	23262.52	2.808	1234123	Chlordane	0.13306	1	GW	3/22/1994	DP-104-1	DP-104	4/8/1997	90.2	32
25	27782.53	23262.52	5.408	1234123	Chlordane	0.20684	1	GW	3/22/1994	DP-104-1	DP-104	4/8/1997	90.2	27
26	27782.53	23262.52	8.008	1234123	Chlordane	0.23958	1	GW	3/22/1994	DP-104-1	DP-104	4/8/1997	90.2	22
27	27782.53	23262.52	10.608	1234123	Chlordane	0.14084	1	GW	3/22/1994	DP-104-1	DP-104	4/8/1997	90.2	17
28	27782.53	23262.52	13.208	1234123	Chlordane	0.04967	1	GW	3/22/1994	DP-104-1	DP-104	4/8/1997	90.2	12
29	27782.53	23262.52	15.808	1234123	Chlordane	0.01232	1	GW	3/22/1994	DP-104-1	DP-104	4/8/1997	90.2	7

The column names in the current SADA file, however, look like the following image (you can get this result by pressing the information button on the main toolbar):



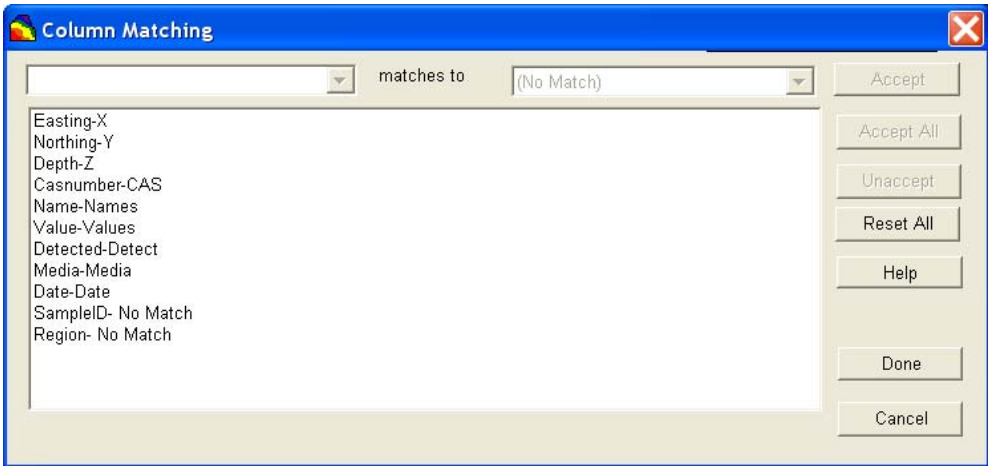
Easting	Northing	Depth	Casnumber	Name	Value	Detected	Media
27596.25	21900	0	14265851	Ac-225	2	1	SO
28310.25	21900	0	14265851	Ac-225	1.6	1	SO
28935	21900	0	14265851	Ac-225	0.9	1	SO
27685.5	22200	0	14265851	Ac-225	2	1	SO
28131.75	22200	0	14265851	Ac-225	4.2	1	SO
29202.75	22500	0	14265851	Ac-225	1.5	1	SO
27150	23160	0	14265851	Ac-225	1.7	1	SO
27685.5	22920	0	14265851	Ac-225	3.6	1	SO
28042.5	23100	0	14265851	Ac-225	4.9	1	SO
28221	23100	0	14265851	Ac-225	4.2	1	SO
28667.25	23220	0	14265851	Ac-225	2.9	1	SO
29113.5	22980	0	14265851	Ac-225	1	1	SO
27417.75	23580	0	14265851	Ac-225	1.9	1	SO
27774.75	23640	0	14265851	Ac-225	2.9	1	SO
28310.25	23400	0	14265851	Ac-225	3.1	1	SO
28935	23460	0	14265851	Ac-225	0.8	1	SO
28200	22560	0	14265851	Ac-225	4.8	1	SO
28700	22500	0	14265851	Ac-225	3.3	1	SO
27200	22380	0	14265851	Ac-225	2.03	1	SO
28984	22787	0	14265851	Ac-225	2.6	1	SO
27350	22750	0	14265851	Ac-225	2.5	1	SO

So we'll need to do some column matching. You will have to match the columns in the new file to their counterpart in the current data set. This is done in the Column Matching window.

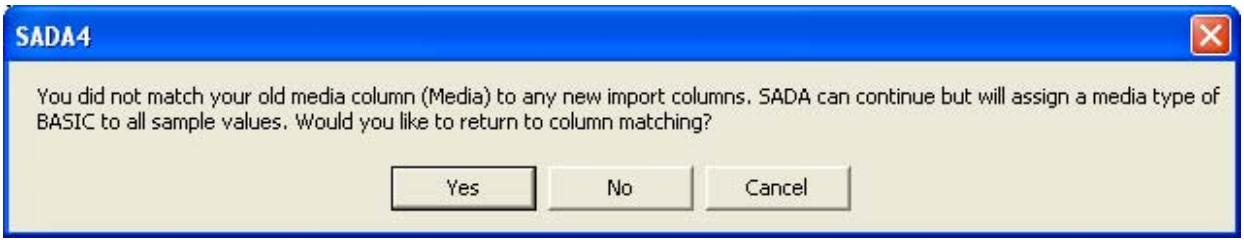


Match the columns in the current dataset to the columns of the new file using the drop down arrows. To accept a match, press *Accept*, and the match will appear in the window below. To accept all the matches (if you know that the columns are lined up), press *Accept All*. Select *Unaccept* to remove one match or *Reset All* to return all matches back to their respective lists.

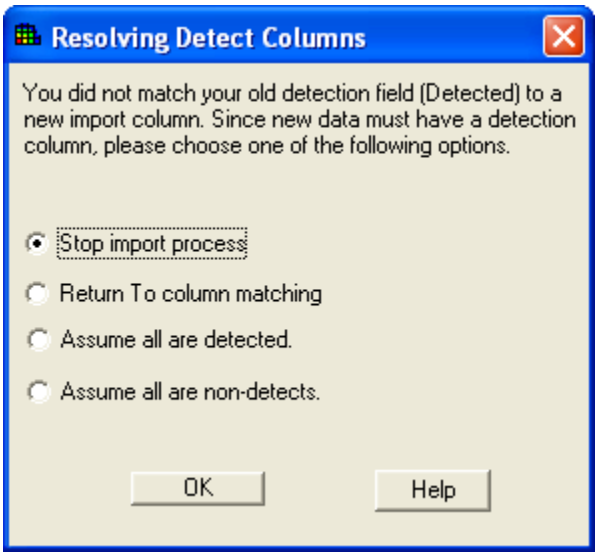
When you are done, make sure your matches look exactly like the following image:



If you did not have a media column in your new data set, SADA would present the following message:



Additionally, detect columns have to be matched or SADA will present the following message:



If you specified a date column in your first data set, you MUST specify one in this new dataset. You cannot proceed unless you do.

After the columns have been set, press *Done*. SADA begins the conversion process and presents the data as it will be imported into the Data Editor.

Just like the first time you imported your sampled data, if there are no errors, you will be given a snapshot of the first 1000 records. If there are errors, you will be presented with the data editor. In the data editor, you can change a handful of errors (more information is available in Chapter 9, “Managing Data”). If you have too many errors, you should consider correcting them in the

source file first. Remember that any changes to the data inside SADA will not be made to the source file on the outside.

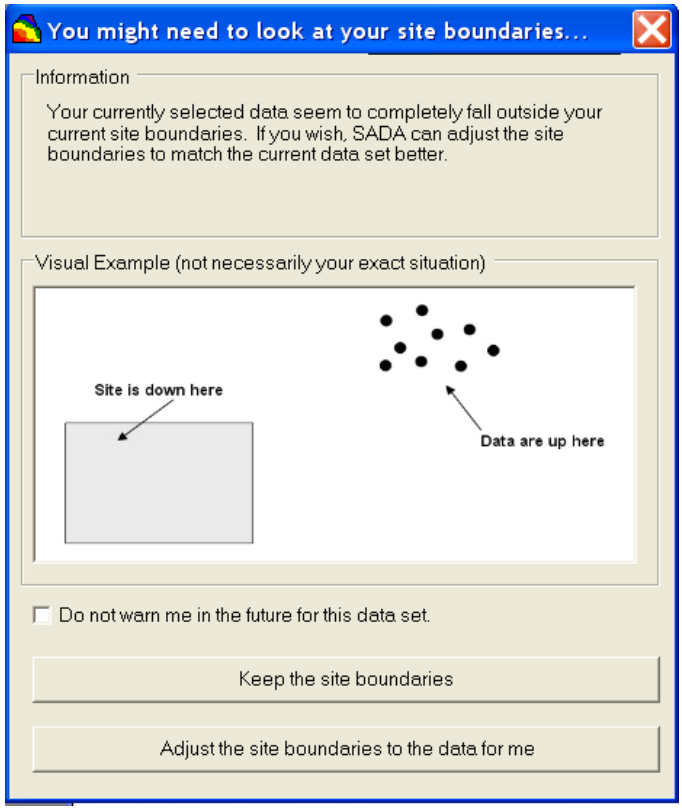
Thinning the Data

As SADA finishes the import process, you may be provided a warning about your site boundaries. This is common and you can find a discussion of this in Chapter 8, "Importing Elevation Data." The same principles apply here and anywhere large data sets are imported into SADA. For this discussion, we'll move forward with our smaller data set.

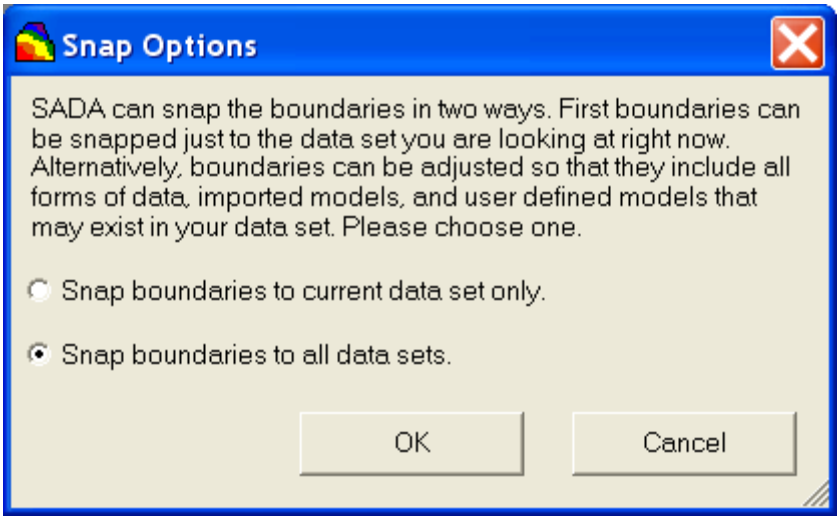
Adjusting Site Boundaries

If you import information into your SADA file that has horizontal spatial extents that are beyond the current site boundaries, you will be presented with an opportunity to change the site boundaries to include the current data. This often happens when you've just created a new SADA file and you're using the dummy site (see chapter 4, "Setting up a SADA File").

This is an action that you may or may not wish to take. For example, certain elevation sets exceed site boundaries, but you do not want to include the entire elevation set. When SADA sees this situation, you may be presented with the following opportunity:



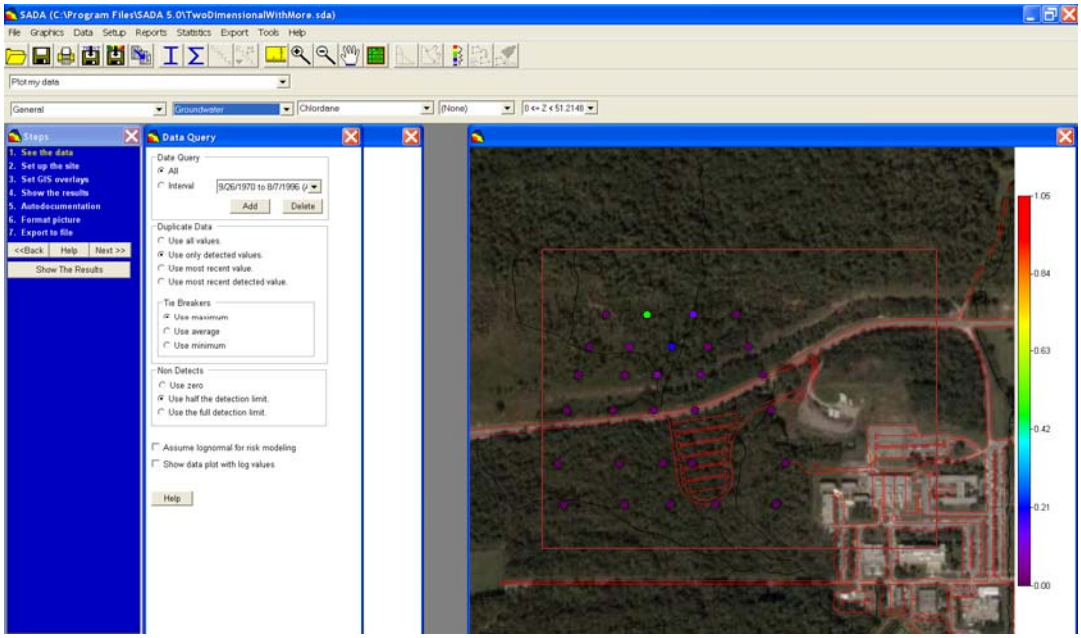
You may select either *Keep the site boundaries* or *Adjust the site boundaries to the data for me*. For data sets that you regularly visit but do not wish to be warned about, select the box for *Do not warn me in the future for this data set*. If you regret selecting this, you can reverse it by visiting *Tools→Warning Options→Site Boundary Problems→Turn Warnings On*. If you choose to *Adjust site boundaries*, you are presented with another option.



Here, you can choose *Snap the boundaries to the current data set only* or snap them once-and-for-all to all data sets found in your SADA file: *Snap boundaries to all data sets*. This will automatically update your *Site Boundaries* information found in your *Setup the site* step. Further information on adjusting site boundaries is available in Chapter 4, "Setting up a SADA File."

When you are presented with this option in our current example, agree to *Adjust the site boundaries* and select *Snap boundaries to all data sets*.

After submitting data through the Data Editor, SADA adds the new data set to your file. When SADA returns, you will still be looking at the Ac-225 soil data. To see your new data set, change the data type from *Soil* to *Groundwater*. You should see the following.



You can save your file now. This new data set is a three dimensional data set. If you have not already done so, please read Chapter 11, "Visualizing and Exploring Your Data."

Chapter 7: Importing Modeled or Gridded Data

SADA provides a suite of geospatial modeling and contouring tools that are flexible enough to handle a wide variety of applications. If you are more comfortable using another contouring package, however, you can still bring the modeled results into SADA and use them within the decision and risk frameworks. Also, you may have geophysical or remote sensing data that have been created/gridded outside of SADA and wish to use them in your SADA file. Import and use of these results is very easy.

SADA recognizes three types of gridded data formats: SADA grid (*.csv), float grid (*.hdr), and ASCII grid (*.txt). The float grid and ASCII grid formats are industry standard formats that most GIS or contouring packages typically export. These are 2d format files, though. SADA also allows you to import three dimensional models, and different packages have different export formats. Because SADA permits variable depth layers, we need to use a slightly expanded format of the typical “grid then results” format. This is the SADA grid format.

SADA grid format

This format permits the type of grids SADA supports, where there can be multiple layers and each layer can vary in depth. The file must be provided in a comma separated value format (.csv).

The contents of the file must adhere to certain specifications. The file begins with the definition of the grid upon which the model was built. So the first two lines define this grid.

X Start, Number Of X, and X Size

Y Start, Number Of Y, and Y Size

The next line must specify the number of vertical layers in the model output.

Number of Z

For the next NumberOfZ lines, the file must contain the beginning and ending z values,

Z1, Z2

Z2, Z3

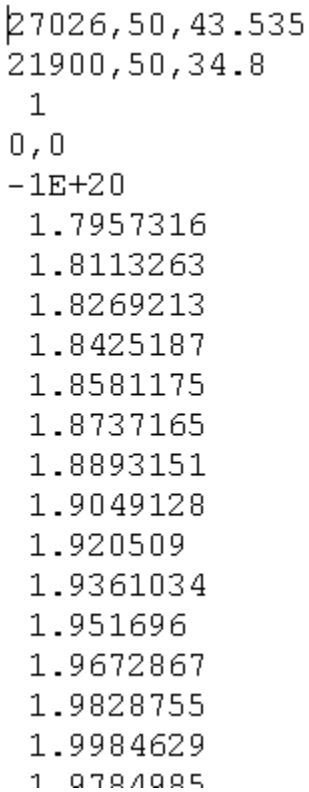
Z3, Z4

where SADA recognizes the layer as inclusive on the lower bound and exclusive on the upper bound. So, if Z1 = 0 and Z2 = 1, then the layer contains all points Z1<=0<Z2.

The next line contains the missing value parameter. If there are any points in the model that were not calculated for some reason, enter the value you are using to represent these missing values here.

From this point forward, the model values are provided one per line, cycling fastest on x, then y, and finally by layer. So, the first value would be the cell result on the first column, first row, on the first layer. The second value would be the cell result on the second column, first row, first layer, and so forth.

The following image shows an example two dimensional model in a SADA grid format:



Float Grid (*.hdr/*.flt)

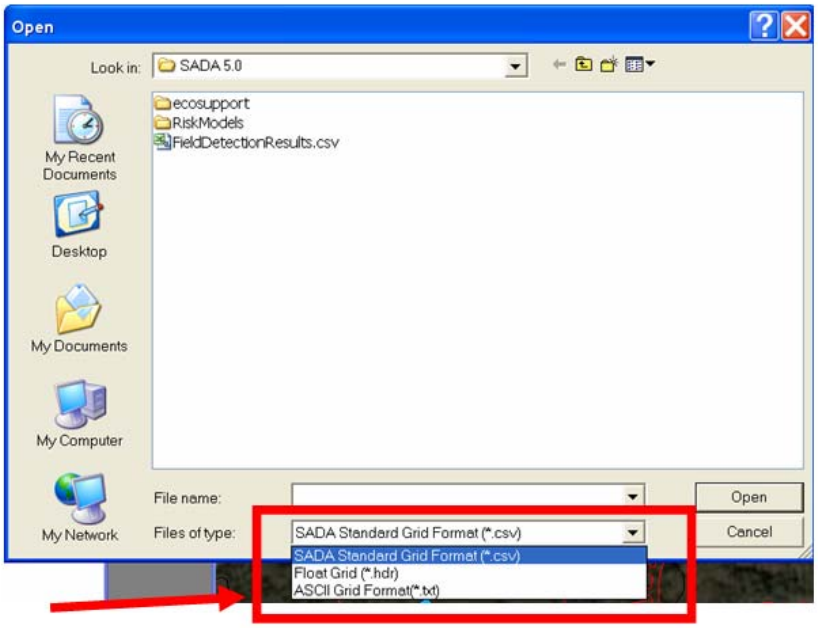
This is an industry standard grid supported by most major GIS systems, such as ESRI Arcmap. See the ESRI documentation for converting any raster set in ArcMap into a Float Grid. You will need to provide both the *.hdr file and the *.flt file in the same directory when importing into SADA.

ASCII Grid (*.txt)

This is a standard grid format for storing raster data and is supported by most major GIS systems, including ESRI ArcMap.

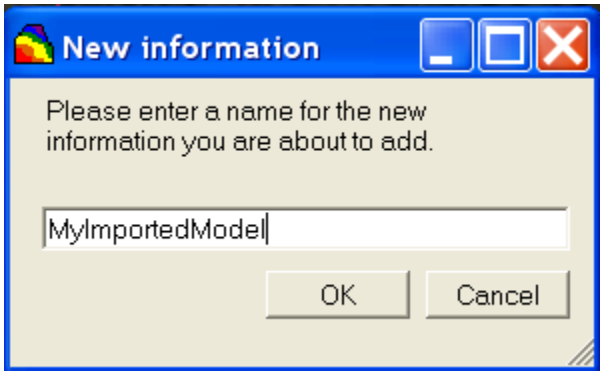
Importing the Gridded Data

If you have SADA open, please open the file Twodimensional.sda (*File→Open* and navigate to where SADA is installed). Regardless of the type of gridded data you have, select from the main menu *Data→Import Gridded Data*, and SADA responds with a standard open window. For this example, we will import the file FieldDetectionResults.csv. This file is a two dimensional model of a gamma walk over survey.



To pick the type of gridded map you wish to import, click on *Files of type* and select accordingly. In this case, we will choose *SADA Standard Grid Format*. Select the file named FieldDetectionResults.csv and press *Open*.

SADA responds by asking for a name to give to the model result.



Press *OK* and SADA will import the model and add it to your list of data sets.

Thinning The Data

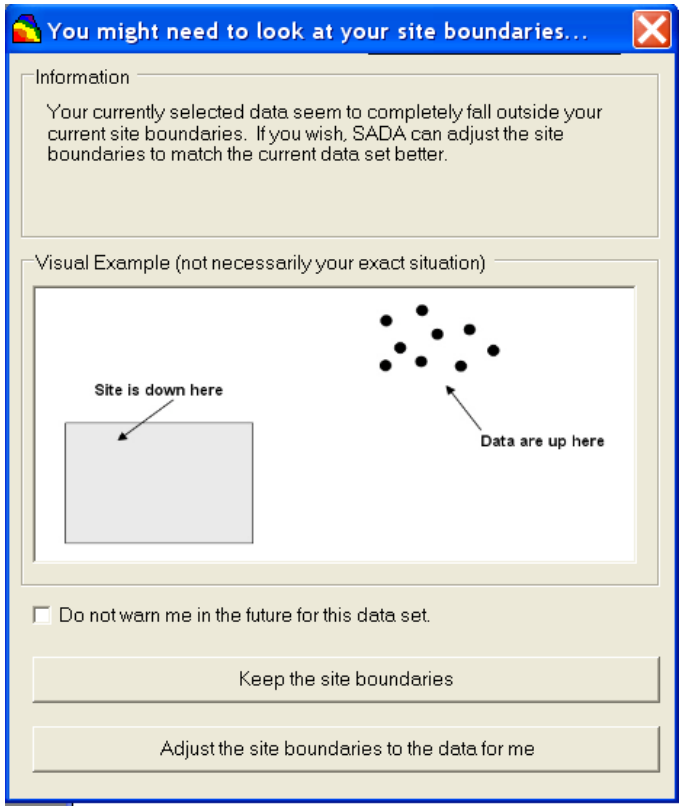
For very dense models, you may be given the opportunity to "thin" the model some. This is normal, and a discussion is found in Chapter 8, “Importing Elevation Data.” In addition, as SADA finishes the import process, you may be provided a warning about your site boundaries.

Adjusting Site Boundaries

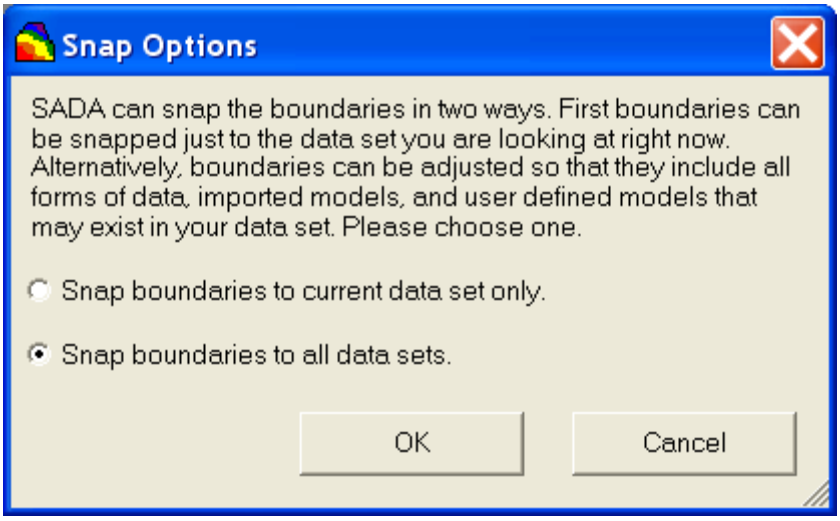
If you import information into your SADA file that has horizontal spatial extents that are beyond the current site boundaries, you will be presented with an opportunity to change the site boundaries to include the current data. This often happens when you’ve just created a new

SADA file and you're using the dummy site (see Chapter 4, "Setting Up a SADA File"). In this case, you will not see this option, as the site is horizontally configured to include our field detection results model.

In general though, this is an action that you may or may not wish to take. For example, certain elevation sets exceed site boundaries, but you do not want to include the entire elevation set. When SADA sees this situation, you may be presented with the following opportunity:

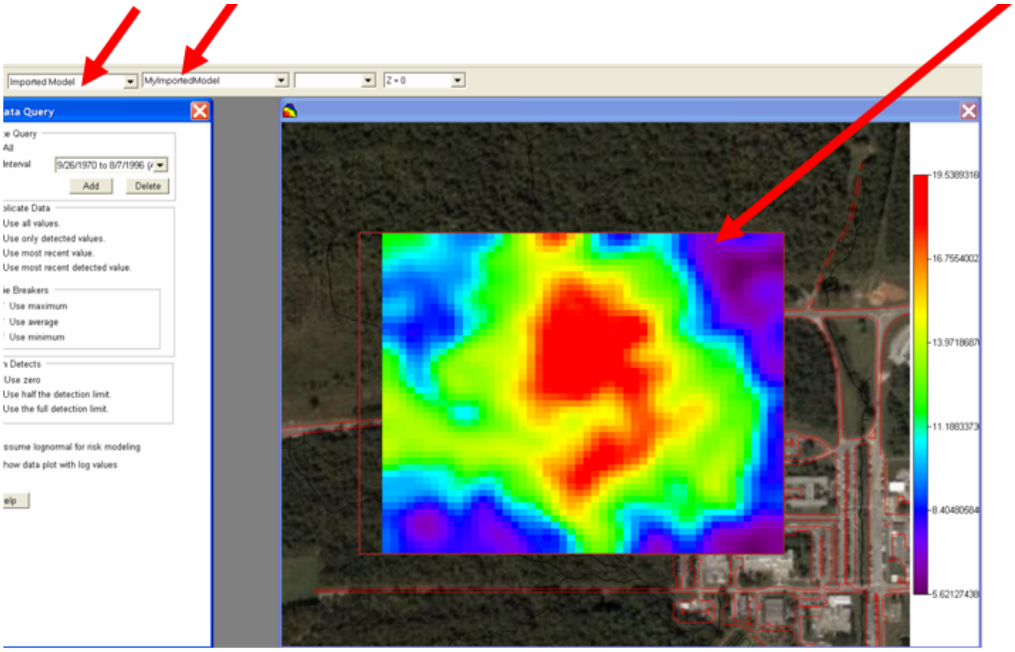


You may either keep the site boundaries or adjust them. For data sets that you regularly visit but do not wish to be warned about, select the "Do not warn me in the future for this data set" option. If you regret selecting this you can reverse it by visiting *Tools→Warning Options→Site Boundary Problems→Turn Warnings On*. If you choose to adjust the site boundaries, you are presented with snap option.



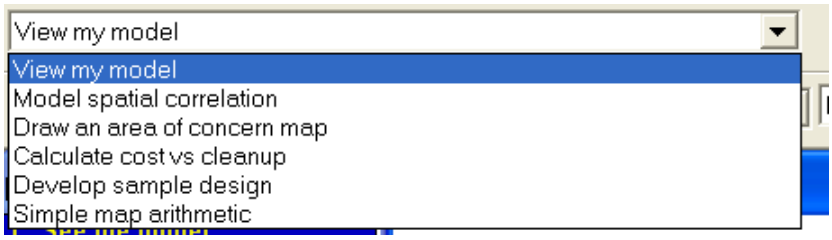
Here, you can choose *Snap the boundaries to the current data set only* or snap them once-and-for-all to all data sets found in your SADA file: *Snap boundaries to all data sets*. This will automatically update your *Site Boundaries* information found in your *Setup the site* step. Further information on adjusting site boundaries is available in Chapter 4, “Setting up a SADA File.”

The result of your import can be seen by selecting *Imported Model* from the list of data types and selecting your recently added model from the list of data sets.



If you now look at your list of interviews, notice how many geospatial tools are available (determine areas of concern, sampling designs, and cost benefit analyses). You will notice some new features (simple map arithmetic) as well.

These tools will behave just as if you had created this model using one of SADA’s contouring packages. To learn more about these, visit the applicable chapters.



You may now close your file. When asked to save, say *No*.

Chapter 8: Importing Elevation Data

SADA permits users to bring elevation data into their analysis. The same types of grid file formats for importing gridded data discussed in the previous chapter apply to elevation grid data. SADA allows users to import Digital Elevation Models or DEMS. *Note: Regardless of the type, be advised that SADA DOES NOT perform any type of projection or coordinate conversions. The user is responsible for making sure the elevation information and any sampled or modeled values applied are reconciled before bringing them into SADA.*

It may be that you don't have access to local elevation information and may want to construct your elevation contours from elevation point measurements taken across your site. If this is the case, simply import the elevation data through the standard sampled or point data format (see Chapter 5, "Importing Sampled Data"). You can then contour the elevation measurements and store the model. If you do have elevation information from outside SADA, the following formats are permissible.

Standard SADA Grid, Float Grid, and ASCII Grid

Just like in the import model case, SADA recognizes three types of gridded data formats for elevation: SADA grid (*.csv), float grid (*.hdr), and ASCII grid (*.txt). The float grid and ASCII grid formats are industry standard formats that most GIS or contouring packages typically export. You can also bring in elevation in a SADA file format, but this isn't really necessary since that format is usually associated with 3d data.

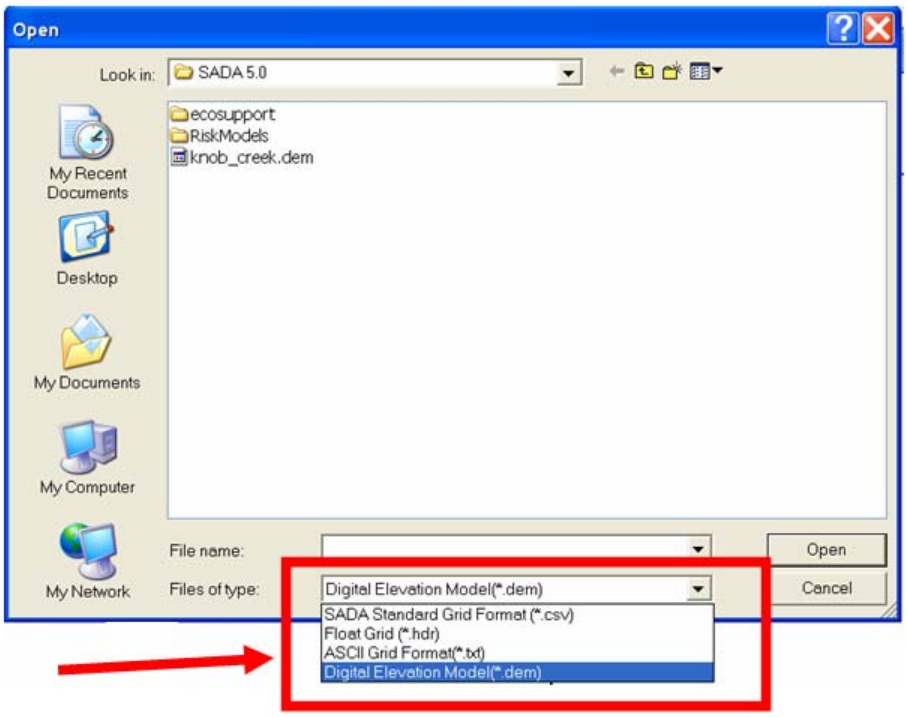
Digital Elevation Model (DEMS)

A digital elevation model (DEM) is a standard GIS format for storing often very dense elevation data. The USGS provides a great deal of elevation data as DEMS, typically very dense and often in UTMs. *Note: SADA does not convert UTMS and unless your data set is in UTMs or you have made the conversion outside of SADA, you should not use DEMS.*

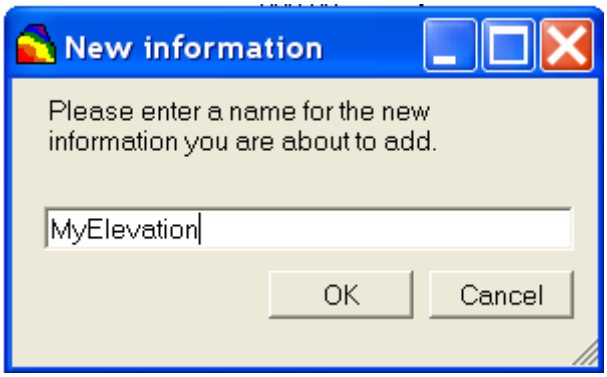
Importing Elevation Data

We'll use a quick example to show how to import some elevation data into SADA. First, open SADA and open the SADA file MyEmptyFile.sda. This is an empty file, so SADA will ask us questions about adjusting site boundaries and so forth when we bring in the elevation data.

From the main menu, select *Data→Import Elevation Data*. SADA presents the standard open windows form. As in the case of importing gridded data, click on the *File of type* to choose the type of file you wish to import.

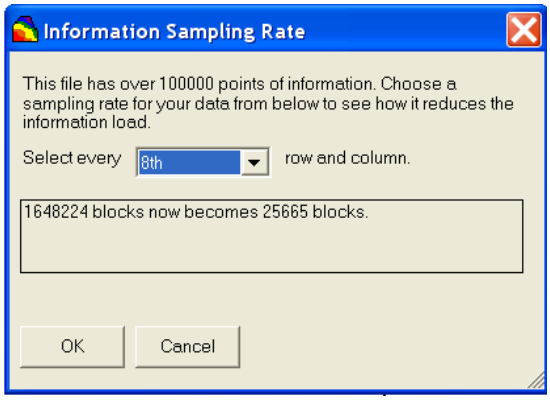


Next, select the file knob_creek.dem and press *Open*. This is elevation data in the East Tennessee area. In the next window, provide the name you wish to call your elevation set. In this case call it "MyElevation." Press *OK*.



Thinning The Data

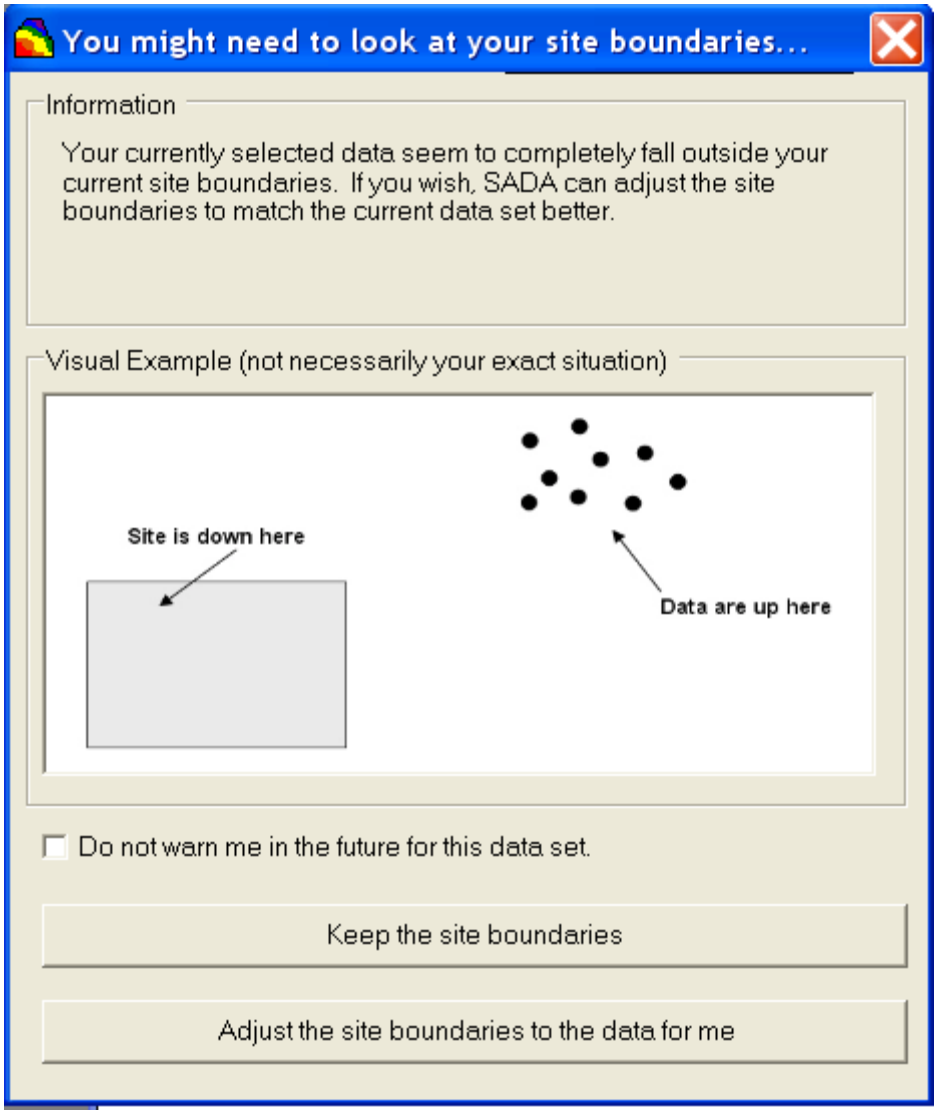
If your information set has a very large number of data, it may be practical and even desirable to thin the data set some so that it does not consume an inordinate amount of resources. This is likely to be the case when importing geophysical data, elevation data, and so forth. If this is the case, SADA will present you with a window for selecting a subset of the data.



Thinning is done geometrically by selecting every other value, every fourth value, and so forth. By changing the sampling rate, you can see the impact on the number of samples or grid cells. Select *8th* from the drop-down list and press *OK*.

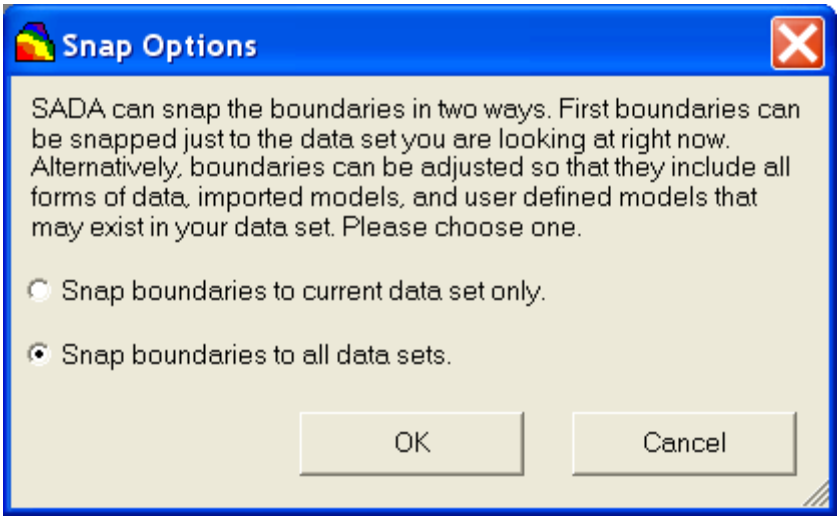
Adjusting Site Boundaries

As you know from reading previous chapters, if you import information into your SADA file that has horizontal spatial extents that are beyond the current site boundaries, you will be presented with an opportunity to change the site boundaries to include the current data. This is an action that you may or may not wish to take. For example, certain elevation sets exceed site boundaries, but you do not want to include the entire elevation set. When SADA sees this situation, you may be presented with the following opportunity:



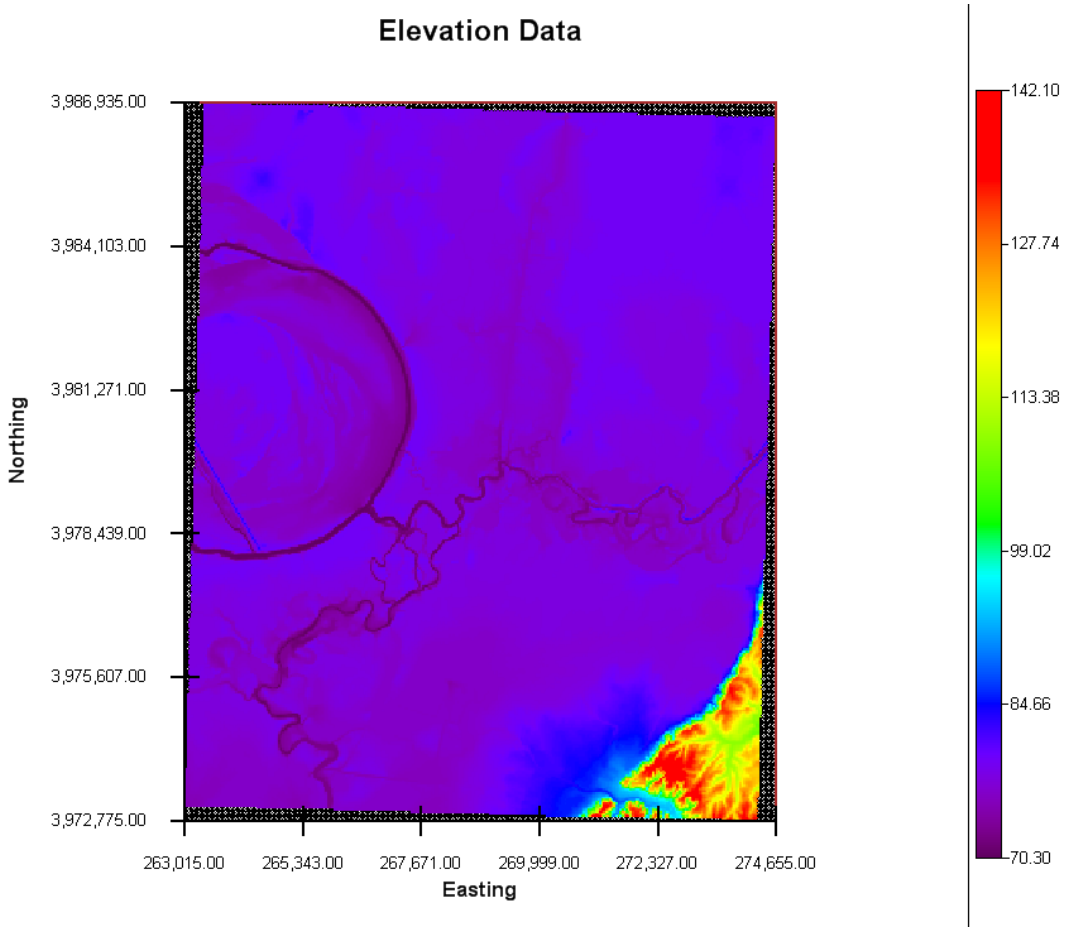
You may either keep the site boundaries or adjust them. *Note: For data sets that you regularly visit but do not wish to be warned about, select the "Do not warn me in the future for this data set" option. If you regret selecting this, you can reverse it by visiting Tools→Warning Options→ Site Boundary Problems→Turn Warnings On.*

If you choose to adjust the site boundaries, you are presented with another option. For this example, choose *Adjust the site boundaries to the data for me.*



Here, you can choose to *Snap boundaries to current data set only* or *Snap boundaries to all data sets* (once-and-for-all to all data sets found in your SADA file). This will automatically update your *Site Boundaries* information found in your *Setup the site* step. Choose to *Snap boundary to all data sets.* (We only have one – MyElevation, so selecting either option here is really the same thing.)

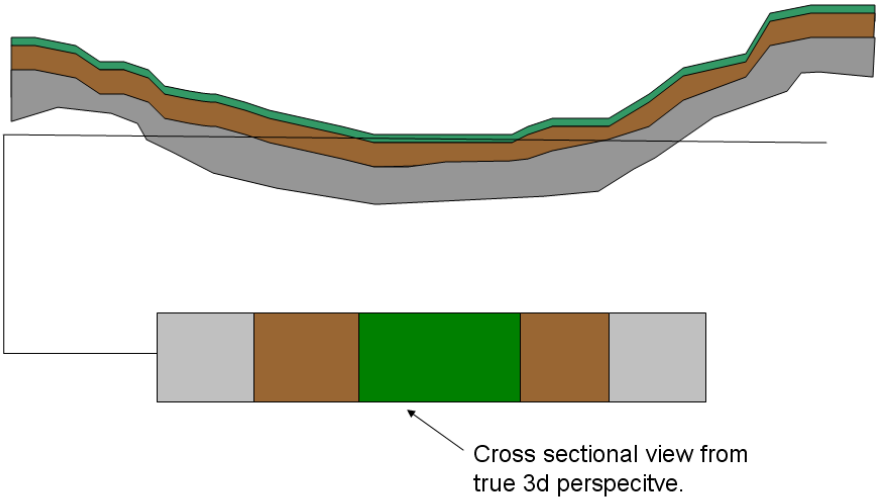
You are finally presented with the results of your elevation map.



Using Elevation Data in Your Work

Elevation in SADA is used primarily in the 3d viewer. There is a chapter on using the 3d viewer (Chapter 44) that you may want to look at later in your readings. Essentially, this 3d viewer will use elevation to properly position data points and model cell values in correct relation to each other vertically, so that land contours are represented in the final outcome.

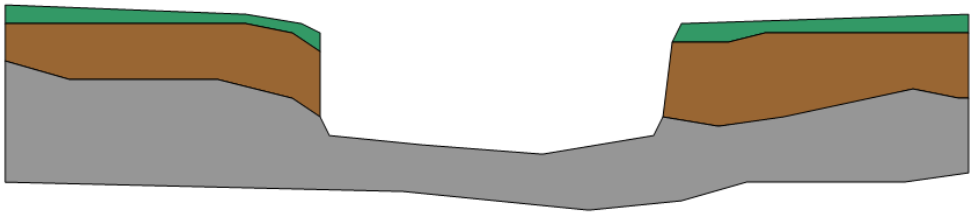
Note: SADA's basis is always "depth below surface." This is how vertical layers are defined and how geospatial modeling is done. Using depth below surface to produce, for example, geospatial models can be advantageous over using elevation. Consider the following scenario. We have a shallow valley with a subsurface geology that is in a typical pancake formation.



From a geospatial perspective, this would be a difficult way to assess correlation structures, particularly anisotropic conditions. In order to assess the spatial structure, it would need to be broken into subsections in which the lower part of the valley sees fairly horizontal correlation structures, the left side of the valley sees correlation structure on a downward angle, and the right side of the valley sees structures on an upward angle.

Instead, if you view the modeling from depth below surface, you can see that really you have fairly consistent layers when measured by the depth, so that a single correlation structure (and geospatial model) may suffice to model the entire domain. So, you first model the domain and then correct in the visualization for elevation.

A situation where we will need a little more attention is seen in the following schematic:



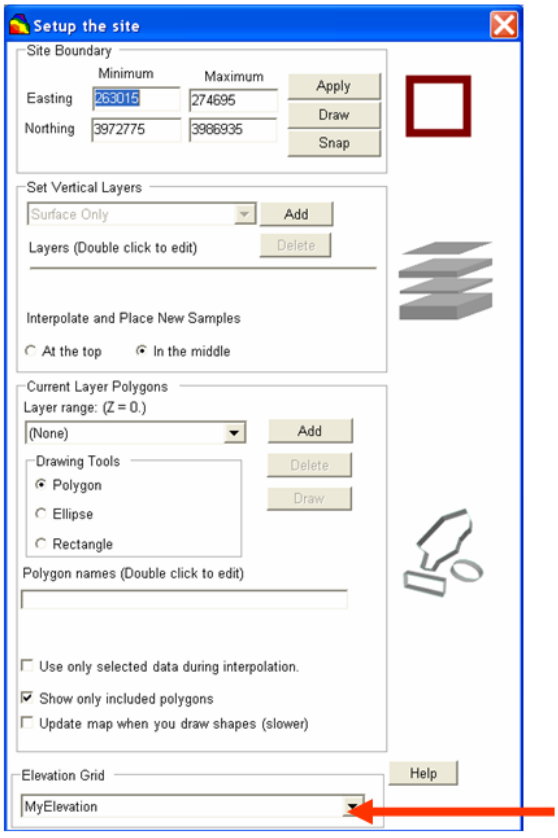
Here, we have a ditch, canyon, or perhaps an excavation situation. In this scenario, you can still attack it with the depth below surface approach. You'll just need to use your polygon tools to exclude the center portion.

To turn the elevation on in your 3d viewer, you need to visit the step *Setup the site*. In the lower portion of the parameter window, you'll see a drop-down list of your elevation models. You would select *MyElevationModel* in this case. No change will be apparent until you actually use the 3d viewer (see Chapter 44).

Generating Elevation Coverage From Scratch

In many cases, users don't have access to elevation data or don't know how to properly project it before bringing it into SADA. This is actually an advanced GIS operation; however, users may have quite a bit of elevation data stored in their measured samples data set. Sometimes the elevation is recorded with each sample. Users could extract this information into a separate data set using Excel. In this separate data set, the information would be created with the following fields: Easting, Northing, Depth, Elevation Value, and Name.

Depth is meaningless but should be added to the data set with a value of zero in each record. Name is meaningless but should be added with a value of "elevation" for each record. Both of these fields are included only because SADA requires them in the next step. Now one can import this into SADA as a point file. Once the elevation data is in SADA as a set of point values, use a fine grid and natural neighbor interpolation to contour the elevation values. If the sampling is representative enough of elevation changes, the result may be adequate. Once this is created, export the model as elevation.csv. SADA will automatically export the results as a standard SADA grid. You can now bring this file back in using the SADA Grid format method described above to create an elevation data set for yourself.



You can now close your file and save it if you like.

Chapter 9: Managing your Data Sets

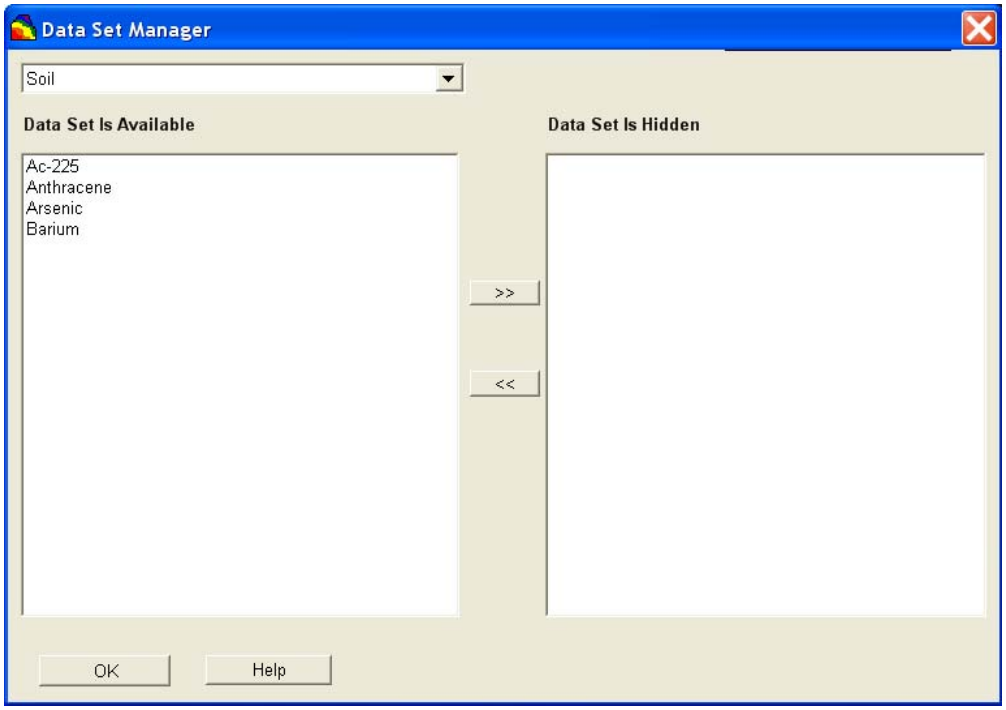
Often, users will have a SADA file with many, many data sets (e.g., contaminants), some of which may no longer be of much use. For example, in a risk assessment, they be identified as not contributing to risk and may be excluded from further analysis. There may be many other reasons why you may want to exclude or delete a data set. In addition, you can edit your data values after they've been brought into SADA. You'll use the data editor to do this. Just remember that any changes made to your data inside of SADA will not affect the source data file used in the import process.

Hiding Data Sets

There are two ways to exclude a data set from your file. The first method is to just hide a particular data set using the contaminant manager. This is a non-permanent way to remove data in order maintain an uncluttered environment.

Open Twodimensional.sda. Save the file as TwoDimensional_Managed.sda to make sure you don't alter the original file, as it is used in other examples. From the main menu, select *Data→ Data Set Management→Choose Included Data Sets.*

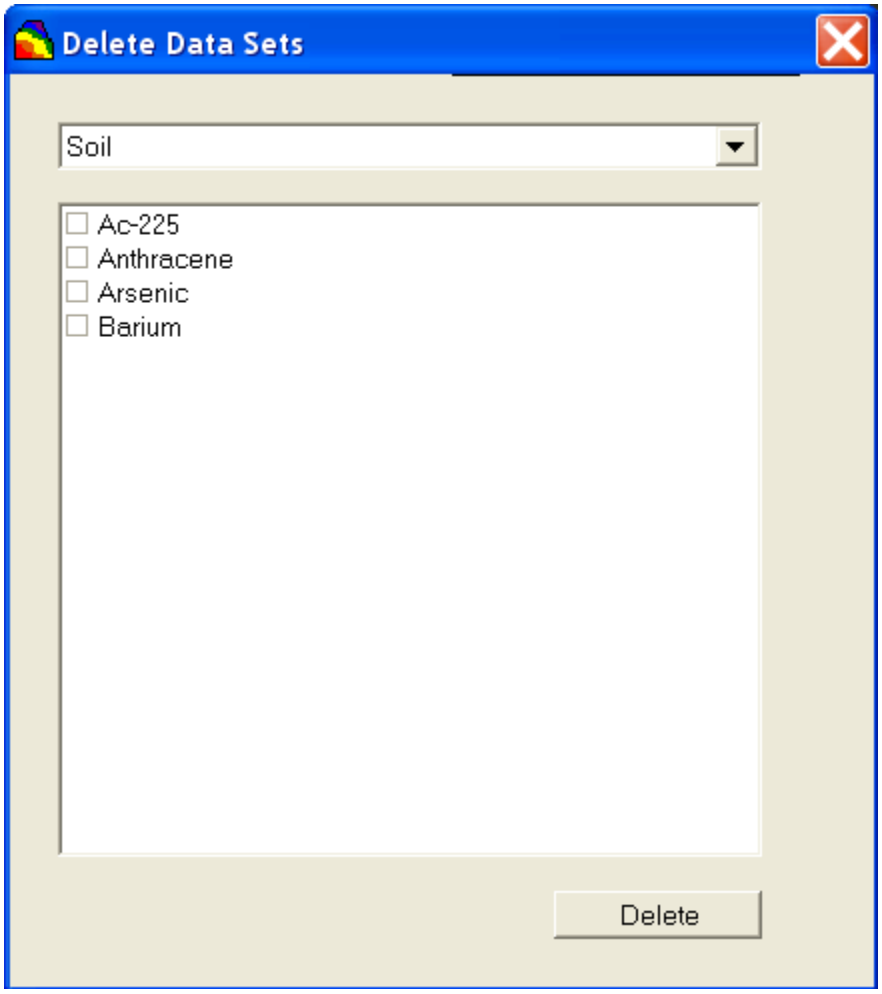
SADA presents the data set manager.



In the upper-left-hand corner, select the data type of interest. The rest of the window displays the data sets that are available (visible) on the left side and data sets that are hidden on the right. Move data sets using the arrows to the appropriate side. When you are done, press the *OK* button. You will no longer see your hidden contaminants anywhere in the interface. They are not deleted though, and you can restore them at any time.

Deleting Data Sets

The second way to exclude data is to permanently delete the contaminant. Again, select from the main menu *Data→Data Set Management→Delete Data Sets*.



This brings up the *Delete Data Setst* window. At the top of the window are the available data types. Select the one you are interested in (e.g., in the figure above, we have *Soil* data selected).

SADA fills the lower box with the data sets found in that data type group. So, in the figure above, we have all the soil measurements. Place a check mark in the box next to the contaminant you want to delete. Press the Delete button. SADA will ask for confirmation. If you confirm the delete, the data set is permanently removed from the file.

Note: Keep in mind that these changes are not permanent until you save the SADA file.

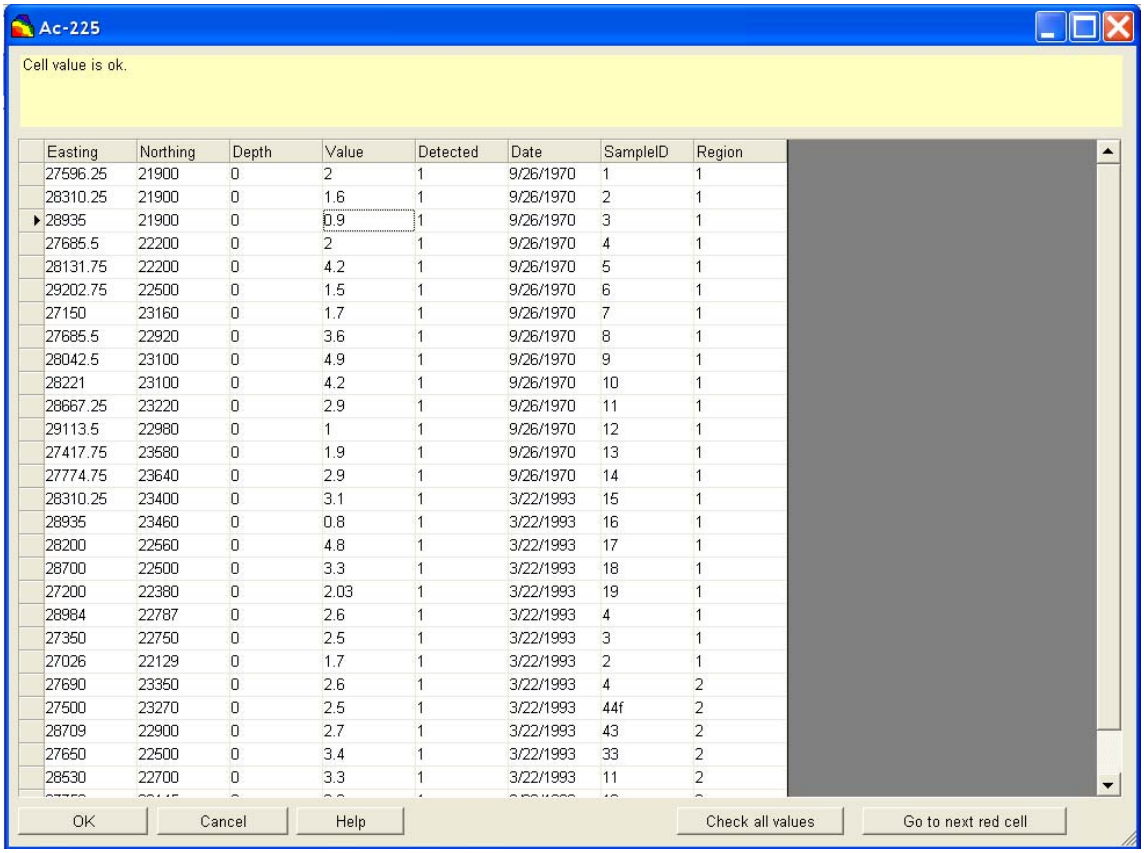
Using The Data Editor

A very simple data editor is provided to allow minor changes to a point or measured value data set. If you have a large data set with a large number of errors (for example, during import), you should seriously consider fixing it outside of SADA as the editor is limited in features.

The data editor is presented to you in two locations. First, if there are errors when you import point data values, you will be presented with the data editor and an opportunity to correct the errors at that time.

To manually open the data editor, select from the main menu *Data→Data Management→Edit Sampled Data Table...*

Open Twodimensional.sda. Select from the menu *Data→Data Management→Edit Sampled Data Table….* SADA presents you with the Data Editor.

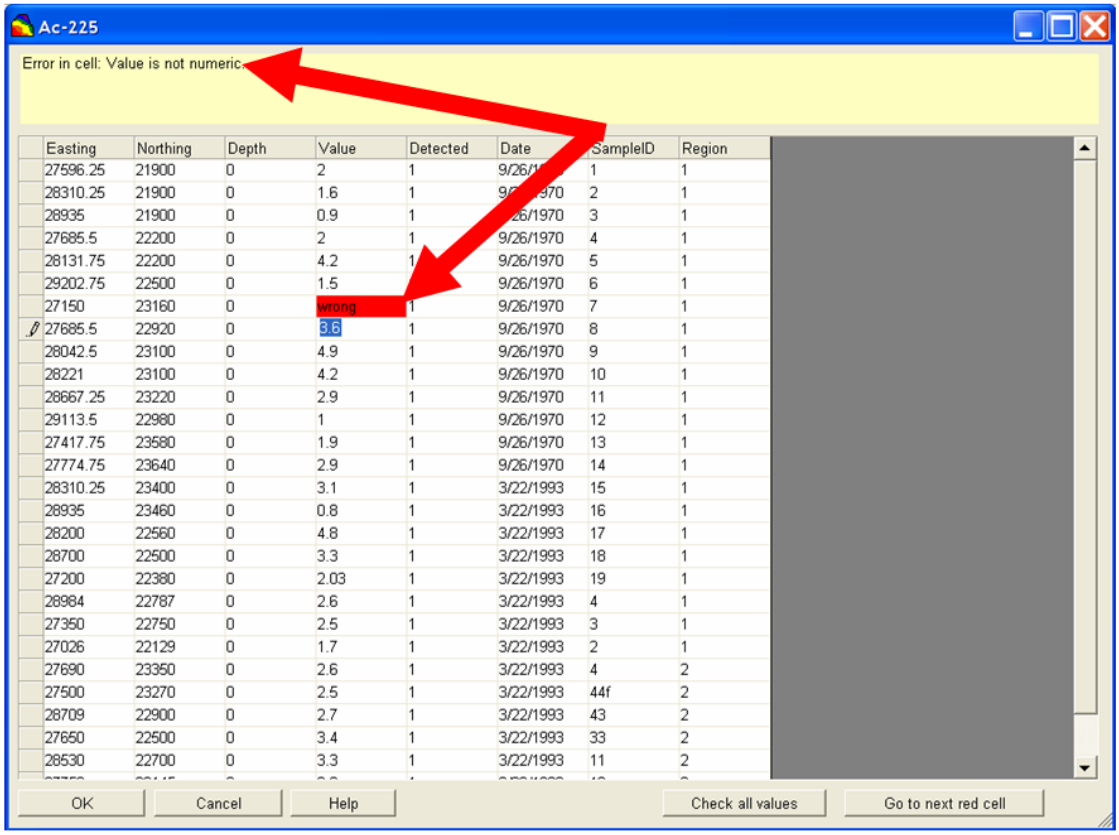


Easting	Northing	Depth	Value	Detected	Date	SampleID	Region
27596.25	21900	0	2	1	9/26/1970	1	1
28310.25	21900	0	1.6	1	9/26/1970	2	1
28935	21900	0	0.9	1	9/26/1970	3	1
27685.5	22200	0	2	1	9/26/1970	4	1
28131.75	22200	0	4.2	1	9/26/1970	5	1
29202.75	22500	0	1.5	1	9/26/1970	6	1
27150	23160	0	1.7	1	9/26/1970	7	1
27685.5	22920	0	3.6	1	9/26/1970	8	1
28042.5	23100	0	4.9	1	9/26/1970	9	1
28221	23100	0	4.2	1	9/26/1970	10	1
28667.25	23220	0	2.9	1	9/26/1970	11	1
29113.5	22980	0	1	1	9/26/1970	12	1
27417.75	23580	0	1.9	1	9/26/1970	13	1
27774.75	23640	0	2.9	1	9/26/1970	14	1
28310.25	23400	0	3.1	1	3/22/1993	15	1
28935	23460	0	0.8	1	3/22/1993	16	1
28200	22560	0	4.8	1	3/22/1993	17	1
28700	22500	0	3.3	1	3/22/1993	18	1
27200	22380	0	2.03	1	3/22/1993	19	1
28984	22787	0	2.6	1	3/22/1993	4	1
27350	22750	0	2.5	1	3/22/1993	3	1
27026	22129	0	1.7	1	3/22/1993	2	1
27690	23350	0	2.6	1	3/22/1993	4	2
27500	23270	0	2.5	1	3/22/1993	44f	2
28709	22900	0	2.7	1	3/22/1993	43	2
27650	22500	0	3.4	1	3/22/1993	33	2
28530	22700	0	3.3	1	3/22/1993	11	2

The top of the window displays the name of your data set. Here, it's the Ac-225 soil data. Underneath this, there is a yellow box. This yellow box will report errors you may have in your data set.

Below the yellow box is the grid itself. To edit any value, simply place your mouse in the cell and click. Enter the value you like and then click in another cell to continue.

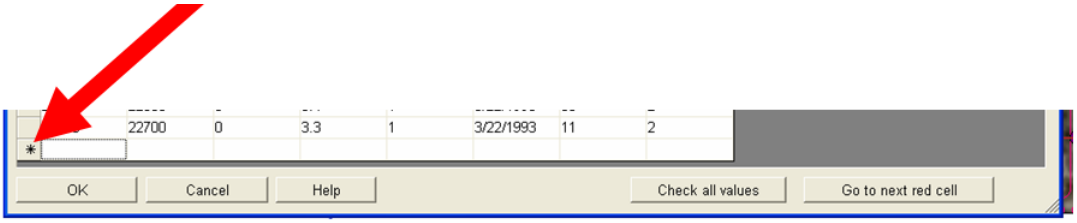
If you type a value that is invalid, SADA will turn the cell red. When you place your mouse over the red cell, the yellow box reports the problem with the value.



In the image above, we've entered the text "wrong" into the value column. This column requires only numeric values. Placing the mouse over this cell updates the yellow box to report that the value is not numeric. To correct any errors, visit each red cell and enter a valid entry. If you have a number of errors scattered throughout your data set, you can use the *Go to next red cell* button at the bottom right of screen to navigate to the next row that contains a red cell.

You can delete a row of data by selecting the row selection field on the far left. A black arrow will appear in the row you selected. Press the *Delete* key on the keyboard.

You can add new data by visiting the bottom of the grid. You will find the last row has an asterisk in the row selection field.



Simply start typing in this row to add a record.

Show Only Registered Contaminants

In SADA, you have an opportunity to set up different analyses, such as human health, ecological risk, and custom analysis, through the *Setup Menu*. For measured point values, you are required to register each data set (e.g., contaminant) as part of the setup. The chapters on risk and custom analysis will take you through this process, so don't worry if you are not yet familiar with this process. For users that are familiar, sometimes not all of your data sets (e.g., contaminants) will be registered because you don't want to register them or because they have no counterpart in the setup you are running.

For example, when a human health setup is complete, you can switch to human health using the analysis box. SADA responds by placing as a default only those data sets in the data set box that were registered during the setup. This is designed to simplify your experience and de-clutter the interface with data sets that are not currently relevant. If you would like to see all the data sets (e.g., contaminants), whether they were registered or not, simply select, from the main menu, *Data→Data Set Management→Show Only Registered Contaminants*.

This menu item is actually a toggle switch. If it is checked, then you will see only registered data sets for the current analysis. If it is not checked, you will see all available data sets. You may now close your SADA file.

Chapter 10: Automatically Importing Information into SADA from Your Database Management System

Your institution may have a centralized or official database management infrastructure in place. From this system, you query and export data sets so you can manipulate them in a manner that they can be imported into SADA. This chapter talks about different options for sending information to SADA behind the scenes in an automated way, so that the transition from your database system to SADA can appear seamless. Similarly, this automation can allow SADA to be called and controlled by any external modeling or management system so that it can be laid in-line with the institution's modeling toolbox. This chapter is really targeted at the IT folks in your institution more than the end SADA user. If you are not an IT person, then what is really important is that you know this capability exists.

Writing Automation Script

The lion's share of this task is really on your in-house programmers or information technology folks. What they will need to do is write a small script or code that exports the information out of the main system into a .csv file automatically. Typically, this is not difficult to accomplish, as the requirements bar for data format is fairly low. Upon successfully exporting the information to a .csv file, this same script or code will execute the SADA application using a "shell" type command or equivalently with a command line prompt. They will then need to provide some additional command line arguments that tell SADA exactly what file was just created and what to do with it. If you are not familiar with this type of language, it's OK; your IT folks will likely be familiar with it. This sounds hard, but it's really quite trivial.

Command Line Prompts

When SADA opens, it looks for a command line prompt. If there is one, the first thing it does is look at the last four characters of the prompt. These four characters tell it what to do next. They are:

Command Line Argument	Meaning	Example
".sda"	Means open the SADA file I have sent you.	Sada5.exe MySADAFile.sda
" efo"	Create an empty file with the name I've given you and then open SADA up showing this file.	SADA5.exe MyEmptyFile.sda efo
" efc"	Create an empty file with the name I've given you but don't open SADA up (completely silent).	SADA5.exe MyEmptyFile.sda efc
" dfo"	Create a file with the data I'm sending you as defined by the .sdx file. Then, open it up in SADA.	SADA5.exe MyDataToImport.sdx dfo
" dfe"	Create a file with the data I'm sending you as defined by the .sdx file. Then, don't open SADA (completely silent).	SADA5.exe MyDataToImport.sdx dfe

Now, the sdx file has the following formats.

SDX File Format

Line: Parameter Description

1: The new sada file name with full path

2: The source data file (*.csv) with full path

3: Data Type (Elevation model = 11, Standard Model = 10, Sample values=6)

4: Data Subtype (Elevation model = 11, Standard Model = 10, Sample values=6)

If Data Type = 11 or 10 then

5: Name to give the model you are importing

6: Sampling Rate Power for reading in the model (for example 2^0 1 = every value, 2^1 = 2 = every other sample, 2^2 = 4 every fourth sample and so on where N is the sampling rate power)

If Data type = 6 (sample values)

5: Sampling Rate Power for reading in the model (for example 2^0 1 = every value, 2^1 = 2 = every other sample, 2^2 = 4 every fourth sample and so on where N is the sampling rate power)

Beginning with line 6 and 7, you must enter the name of the primary field SADA is looking for, then a comma, and then the column number where it is located in your data set. The field numbers are 1 based. That is, the first column is column =1, the second column is column = 2,

and so forth. The recognized field names are: Easting, Northing, Depth, CAS, Name, Value, Detect, Media, Date.

Your fields can be named anything you like; these are just string based identifiers to tell SADA where the needed information is. For example, Easting = 1 means that the easting coordinates are in the first column

Automation Examples

Here is an example of a SDX file to directly import the threedimensional.csv file packed with SADA. You can use this file to create the file MyThreeDimensional.sda automatically. Note: your directory structure may vary. In this example, every required field or field of interest that SADA might use is included and is coincidentally in order.

```
C:\Program Files\SADA 5.0\MyThreeDimensional.sda
C:\Program Files\SADA 5.0\ThreeDimensional.csv
0
6
0
Easting, 1
Northing, 2
Depth, 3
CAS, 4
Name, 5
Value, 6
Detect, 7
Media, 8
Date, 9
```

Here is an example of an SDX file where only a few of the base required columns are present.

```
C:\Program Files\SADA 5.0\MyTwoDimensional.sda
C:\Program Files\SADA 5.0\TwoDimensional.csv
0
6
0
Easting, 1
Northing, 2
Name, 5
Value, 6
```

This example will automatically import the elevation model “knob_creek.dem” packed with SADA, and it will sample every 16th (2^4) value (it’s a huge file).

C:\Program Files\SADA 5.0\TwoDimensionalElevation.sda

C:\Program Files\SADA 5.0\knob_creek.dem

1

11

MyElevation

4

Part III: Exploring Data

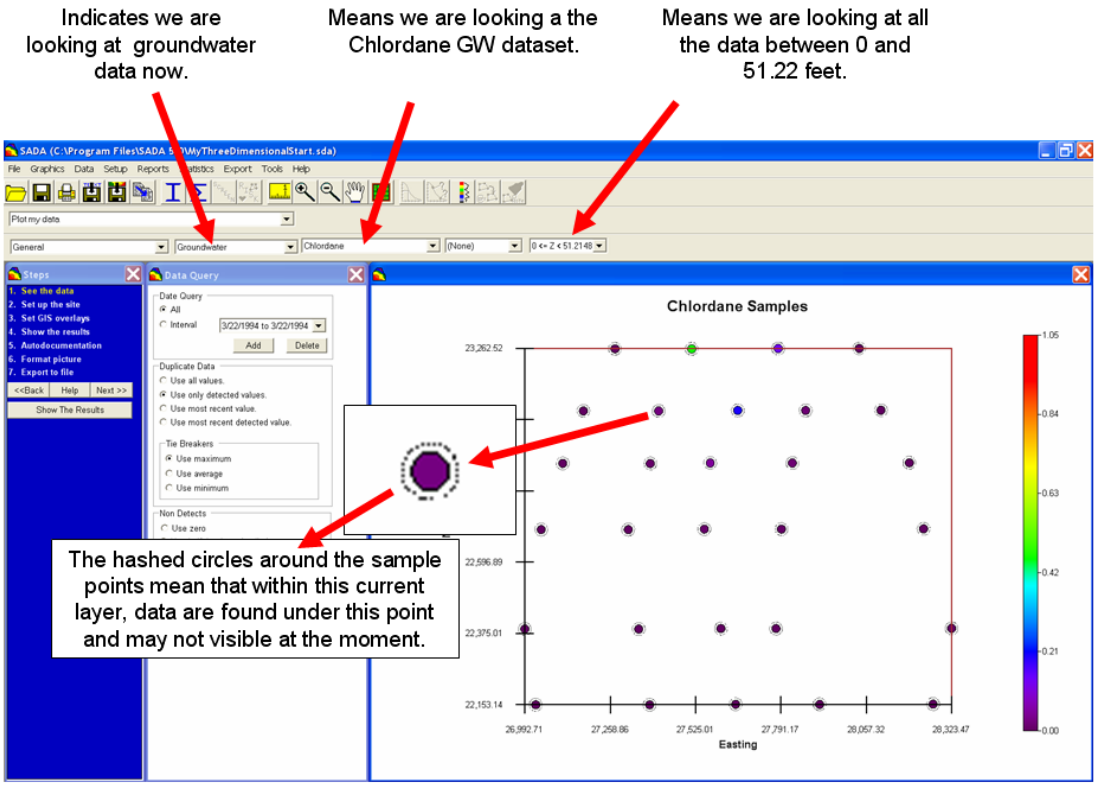
Chapter 11: Visualizing and Exploring Your Data

Once you have imported your data set into SADA, you will likely want to start by exploring your data. SADA provides ways visualize your data, query some information, and perhaps look at some basic statistical results. This chapter highlights some of the most important tools for spatially examining your data. Let's begin by opening a SADA file that has already been populated with some data. Open SADA and select from the main menu *File→Open*. Navigate to the directory where SADA has been installed (probably C:\Program Files\SADA 5.0) and select AThreeDimensionalStart.sda. After the file opens, save it as MyThreeDimensionalStart.sda using the main menu *File→Save As…*

This file is in a very early stage of development. After creating the file, the Chlordane groundwater data from ThreeDimensional.csv was imported. Beyond that, nothing has been done. We'll begin first by looking at the 2d Graphics window.

Understanding the 2d Viewer Graphics Window

With the file MyThreeDimensionalStart.sda open (see above), we'll begin with discussing what we are looking at in the graphics window. First of all, SADA defaults to the "2d" graphics window. This is the normal graphical view that most users use. SADA also has a 3d viewer window and we'll see how that works very soon. At the moment, the SADA interface looks like this.



First of all, you can tell from the data type box that we are looking at groundwater data. If you select the arrow next to *Groundwater*, you'll see that this is the only type of data we have at the moment. You can also tell that we are looking at Chlordane data. In the data set box, you'll see the name *Chlordane*. Selecting the drop-down arrow on this box shows only *Pooled* data and

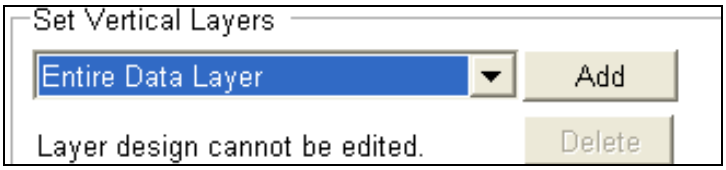
None. Pooled data is not a data set per se and is something we'll talk about in an upcoming example. In the far right drop-down box, SADA indicates that we are looking at a very thick layer from 0 to 51.2+ feet. So, if you look in the graphics window, you'll see a spatial plot of the sample values that occupy a vertical depth of 0 to 51.2+ feet. Notice that virtually all the samples are encircled by a hashed circle. Anytime you see a point value drawn like this, it means that within the current layer there are samples located below this point (but still within the selected layer) that you cannot see. The term "within the current layer" is important to note here. These hashed circles are only drawn when data within the same layer overlap each other. If you don't see the circles, it doesn't mean there is no data at all anywhere below this point (within the selected layer). In fact there may be points at lower or higher layers, but they would be seen only when those layers are selected. *Note: For review please see the section in Chapter 4 on setting up vertical layers.*

The circles are filled according to the legend on the far right. Notice that the higher the measured value, the brighter the color. Shortly, you'll see how to make your own legends.

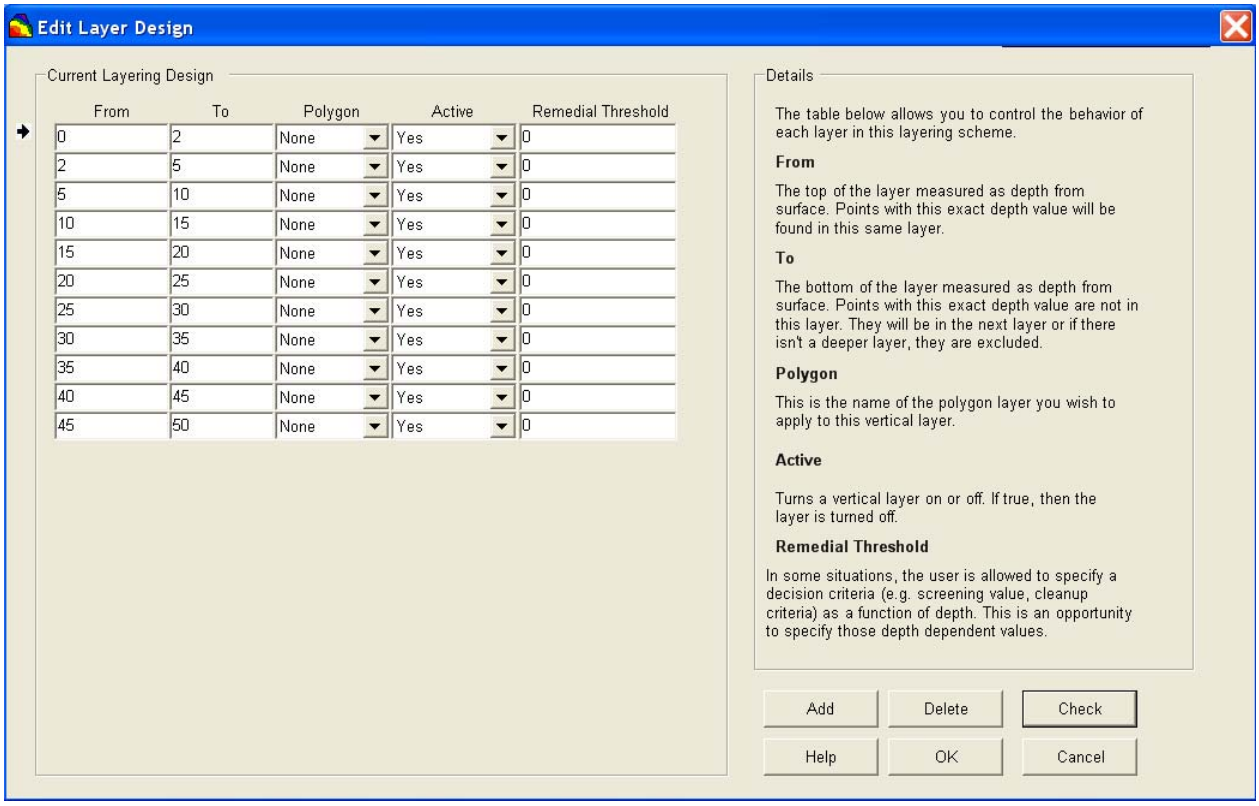
Try moving your mouse over the graphics window and notice that in the lower left hand corner of your screen SADA is reporting the current coordinate of your mouse pointer. This can be very helpful in your data exploration.

Coming up with reasonable vertical layers

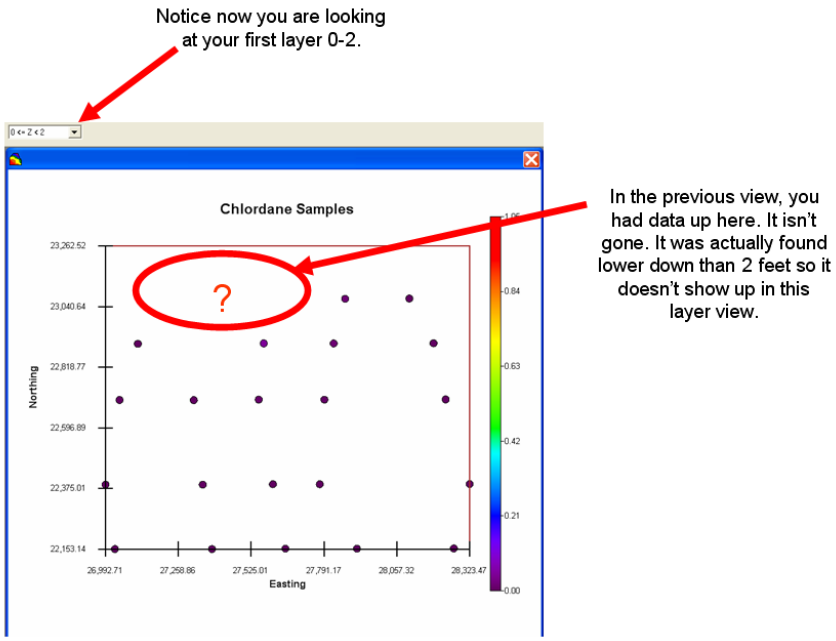
Let's return to the thick layer we're currently using. It turns out that the very thick layer (0<=z<51.2148) is the layer design *Entire Data Layer*. This is a special layering design presented in Chapter 4, "Setting up a SADA File." This layer design is dynamically controlled by SADA to always be large enough to show all the data at once. This is helpful when you first bring a data set in and want to make sure all your sampling stations were properly imported. To find out what layering design you are currently using, click on the step *Setup the site*. Recall that here is where vertical layering designs are managed. In the center of the parameter window you'll notice that *Entire Data Layer* has been selected.



Suppose now we want to come up with our own vertical layering scheme that allows us to see the data in thinner layer divisions. There are different motivations for choosing a layering scheme. You may want to use layer thicknesses that correspond to a particular land use in a human health scenario. You may just want to set your vertical layers so that you no longer have so many data plotted on top of each other in the graphics window. Let's use this later reasoning and setup a new layering design called MyThinLayers. In Chapter 4, we covered how to setup a layering design. If you can't recall how to do this, please review that chapter. Very briefly though, press the *Add* button next to the box showing Entire Data Layer. Enter the name MyThinLayers and base the design on a new layer design. In the layer editor, create a 0-2 layer, a 2-5 layer, and then thereafter all 5 foot intervals. Before finishing, make sure your entries match the following:



When you press *OK*, you'll see your new layering design has been applied. Now you can peruse through the data set and see a little more clearly what is going on. Not all the hash marks are gone, but it's definitely an improvement. Notice that in the upper-left-hand corner it looks like some of the data we saw in the previous view using the *Entire Data Layer* has now been lost. This is not the case. Those data points are just located in a deeper layer than our current 0<=z<2 layer. To see for yourself, just choose a different layer from the *Set Vertical Layers* drop-down box.

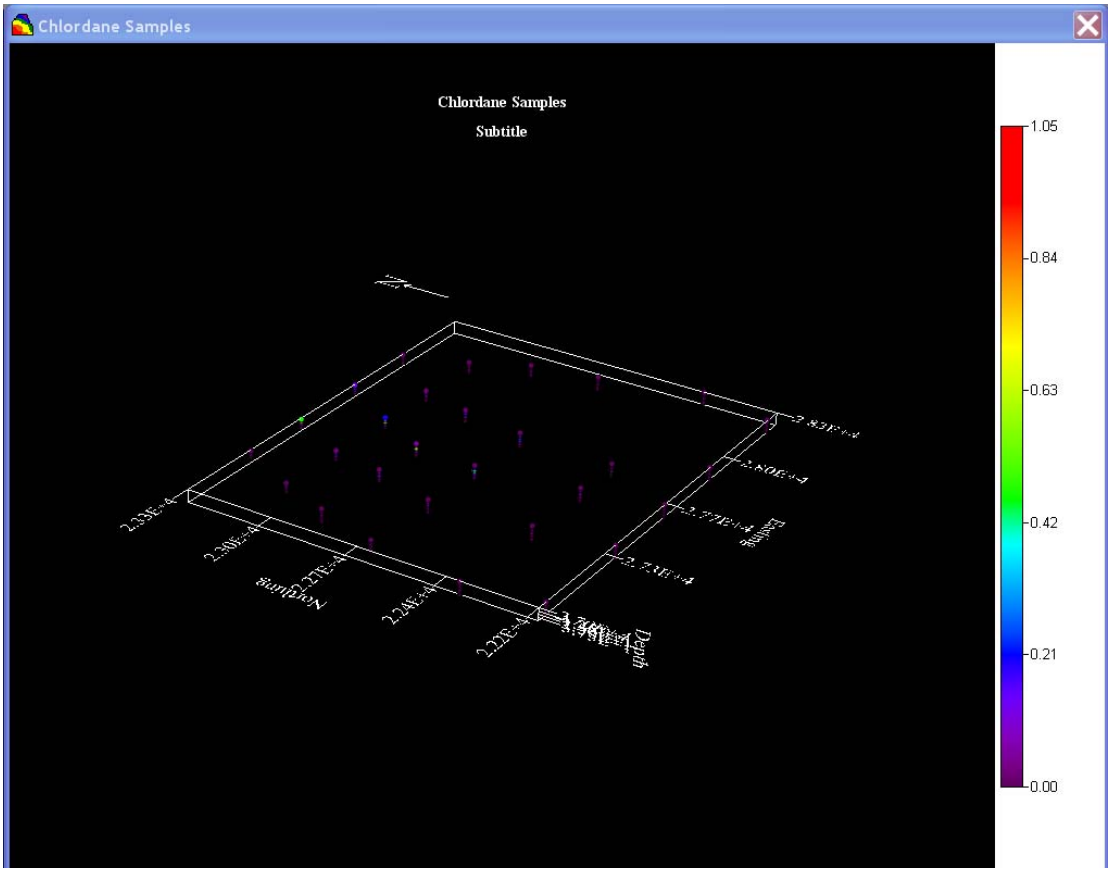


If you select deeper and deeper layers, you can see that we have an elevated set of data values in the center of the site. The elevated values seem to move south as you move deeper. This is very important to notice, as it likely indicates the presence of a plume. Go ahead now and save your file.

A brief look at the 3d Viewer

You can look at the data by slice or look at it all at once using the 3d view. Be aware that 3d viewers are worlds of their own, worthy of their chapter (see Chapter 44, "The 3D Viewer"). Let's turn on the 3d viewer to show you it's there and show you a few tricks that are worth knowing. To see the 3d viewer, select from the *Graphics Menu→Show 3d View*. You should see something like the following diagram.

Note: 3d viewers take longer to render than 2d viewers, and SADA is no different. We recommend that when you are doing most of your analyses you leave the 3d viewer turned off. In addition, many tools are not available when the 3d viewer is on. The 2d viewer should be your primary base of operation.

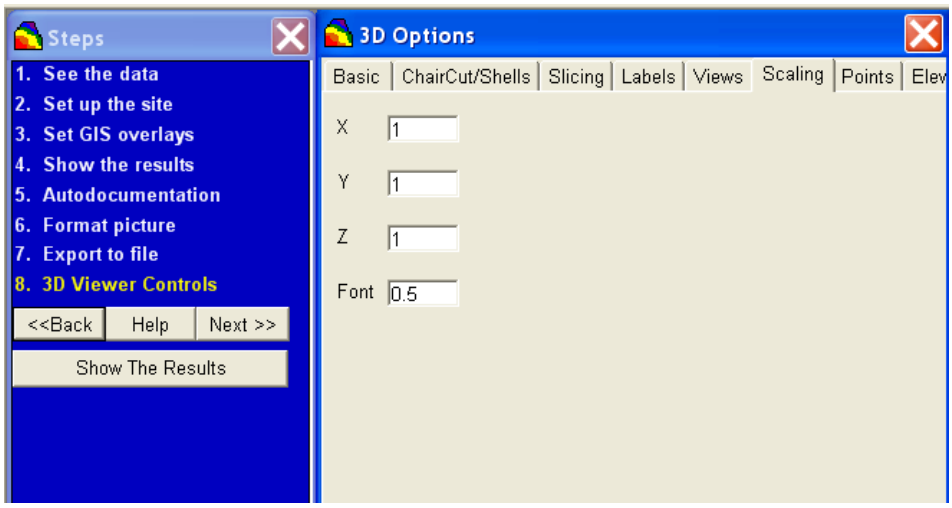


IMPORTANT TIP!!!

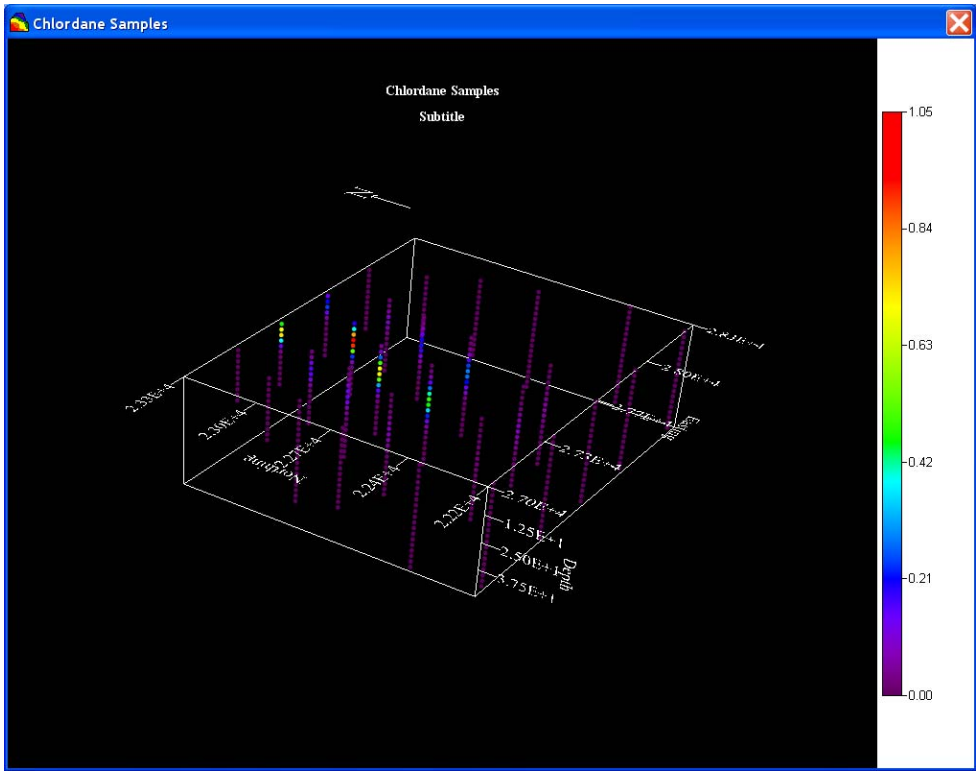
Depending on the video card you have, the 3d viewer may act temperamental. If you don't see this view or see "remnants" of the last window you had open in its place, just click anywhere in the 3d viewer to refresh it.

Most of the time, for environmental data, the vertical extent is considerably smaller than the horizontal extent. For this reason, when you view the results in 3d your site looks "pancaked" or mashed flat. So, the first thing you'll want to do is exaggerate the z direction.

Click on the step *3d Viewer Controls* (this is a new step that popped up when you turned the 3d viewer on). This parameter window has quite a few tabs and can be intimidating at first. These features are taken up in greater detail in Chapter 44. Really though, most of the time you'll repeatedly use only a few of the features. Scroll to the right in the window and click on the *Scaling* tab. You should have the following view.



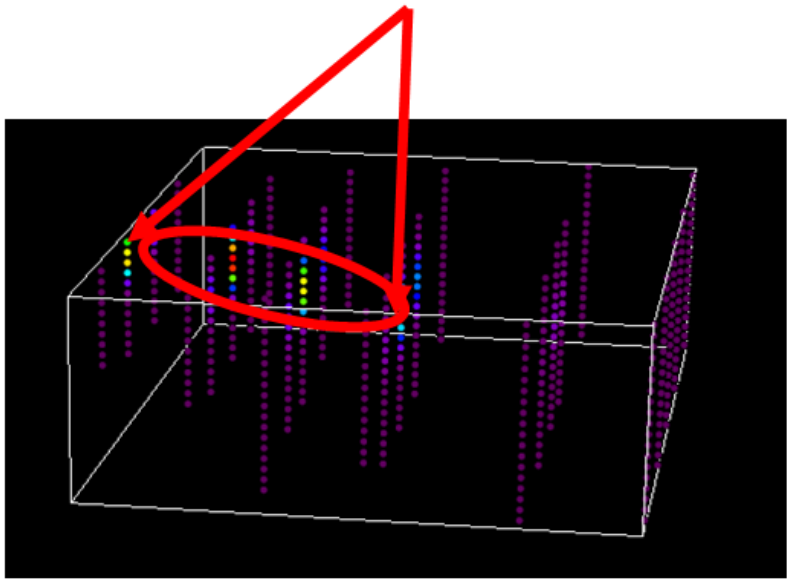
In the Z parameter box, enter a value of 10 and press the Enter key on your keyboard. You should see the following change in your 3d view. Again, if you don't see a change, just mouse click anywhere in the 3d view.



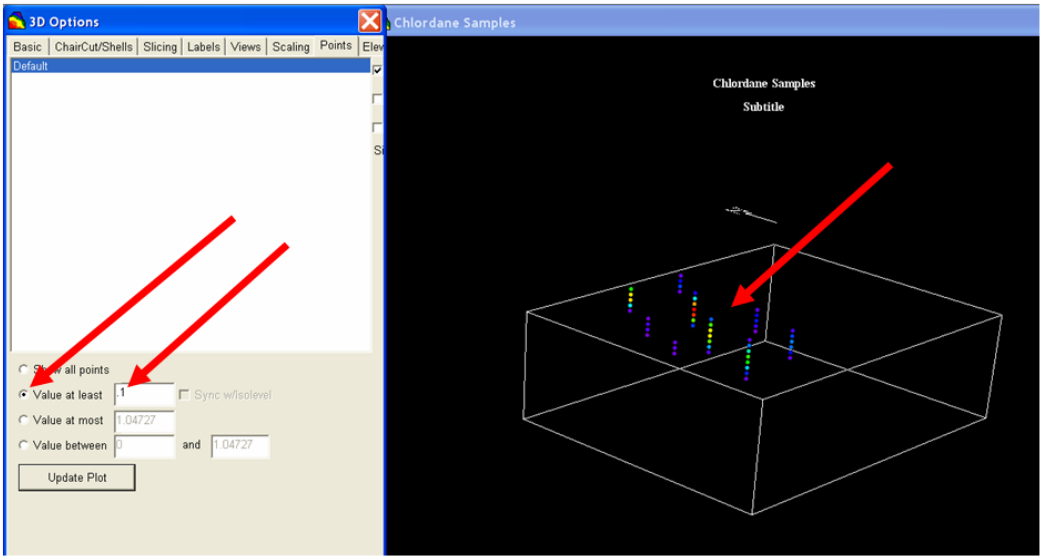
Rotating the view is easy. Just left mouse click and hold anywhere in the 3d viewer area. Now drag your mouse and you'll see the image begin rotating. The best way to learn how to rotate is to just play around with moving your mouse in different directions.

If the axis labels are annoying, you can get rid of them by right-mouse-clicking on the axis they belong too. They turn off and on accordingly.

You can now see the location of the elevated zone in the subsurface very easily by rotating the 3d results into different positions.



You can now turn off all the lower value points so you see your results better. To the right of the *Scaling* tab you'll find the *Points* tab. Click on this tab. Select the *Value at least* option and enter a value of 0.1. This will eliminate any measured values below a value of 0.1 mg/L (recall that SADA doesn't track units; these are selected by the user. The exception is when doing risk assessment.)



Now you can clearly see the elevated values in the 3d viewer. Unfortunately, we don't have a GIS layer at this point for reference. If you did, the 3d viewer would draw it on the top of the 3d cube. *Note: Sometimes you need to click on the 3d viewer window to refresh or recalculate your view.*

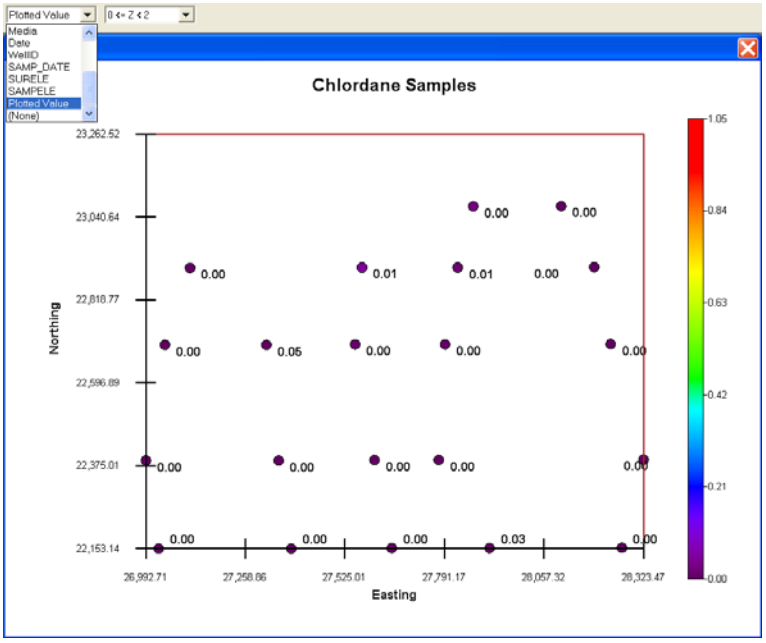
Let's turn the 3d viewer off by choosing *Graphics→Show 3d Viewer*. The menu item acts like a toggle and will turn the 3d viewer off and on.

Getting Information About Your Data

There are a number of ways to query the SADA file for information about the data you are seeking. The first way is to use the meta labels drop-down box. Remember that SADA allows you to import all required or useful fields as well as any other field you may have.

Meta labels drop down

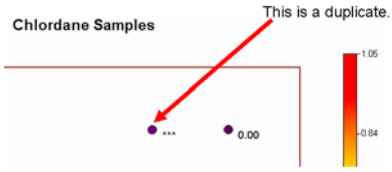
Let's begin with an option in the drop down box that actually *isn't* a field in your data base. If you scroll to the bottom of the fields you'll see an entry called *Plotted Value*. This is a dynamically labeled result, whereby the exact value of the currently plotted value at each location is reported. Let's select this item now.



You'll see the concentration values that are currently plotted appear as small labels to the lower right of each point. Now select the field named *Values*. This is actually the field in the data set that was matched to SADA's required Values information type. Nothing much has changed. So the question arises, what is the difference between *Plotted Value* and *Values*? *Values* comes from your data set and is the actual measured value for each point. *Plotted Value* is the value that is plotted at each sample location in the current view. In this example, they are the same, but they are not always equal. Consider a point risk plot. Each plotted value might be a carcinogenic risk value and not a concentration value.

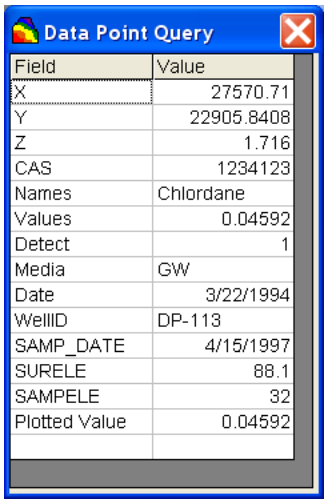
In actuality, even in this example, there is a difference between the two. With *Values* selected as your meta label, notice the label result "***" in the upper-right-hand portion of the picture. Anytime you see the three asterisks like this, it means this sample location contained a duplicate value. Therefore, SADA doesn't know what to provide as the meta label value. The only exception would be the selection *Plotted Value*. In that case, SADA has followed the rules regarding duplicates and resolved the duplicate values to a single value that can be plotted. For

all other cases, however, there is more than one possible answer. In this case, the asterisks are used. This is an excellent way to quickly visualize where your duplicates are found.



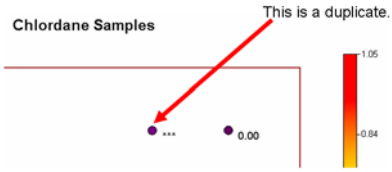
You can switch to other labels and view a single database result for all the visible samples at once. The numerical values are formatted according to whatever format is used for the legend.

Alternatively, you can look at all the meta data values for a single point. To do this, simply click on a data point of interest. SADA responds by showing you a popup window containing all the information about the point you clicked on. If the point you clicked still has hash marks around it, (indicating deeper samples within the same layer are also there) you'll get all the information for each value found at the easting/northing location within the current layer.

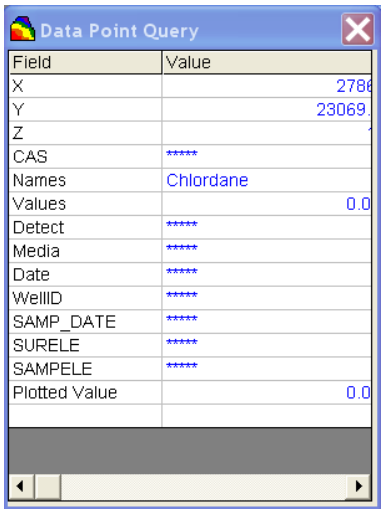


Field	Value
X	27570.71
Y	22905.8408
Z	1.716
CAS	1234123
Names	Chlordane
Values	0.04592
Detect	1
Media	GW
Date	3/22/1994
WellID	DP-113
SAMP_DATE	4/15/1997
SURELE	88.1
SAMPELE	32
Plotted Value	0.04592

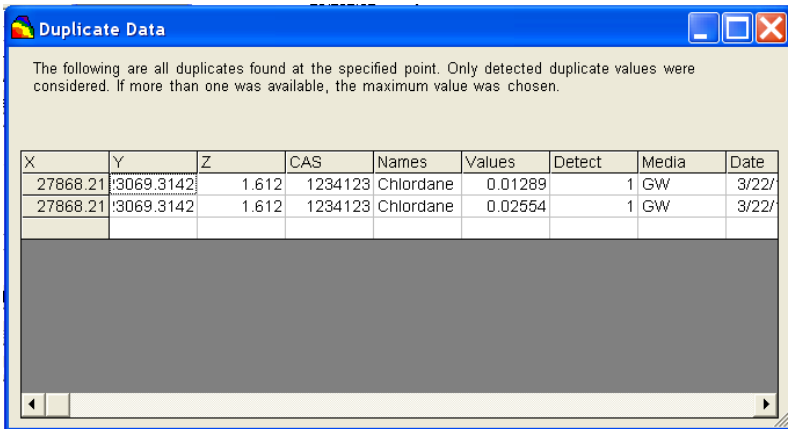
Try now also clicking on the duplicate location.



You will see instead the following in your data point query window:

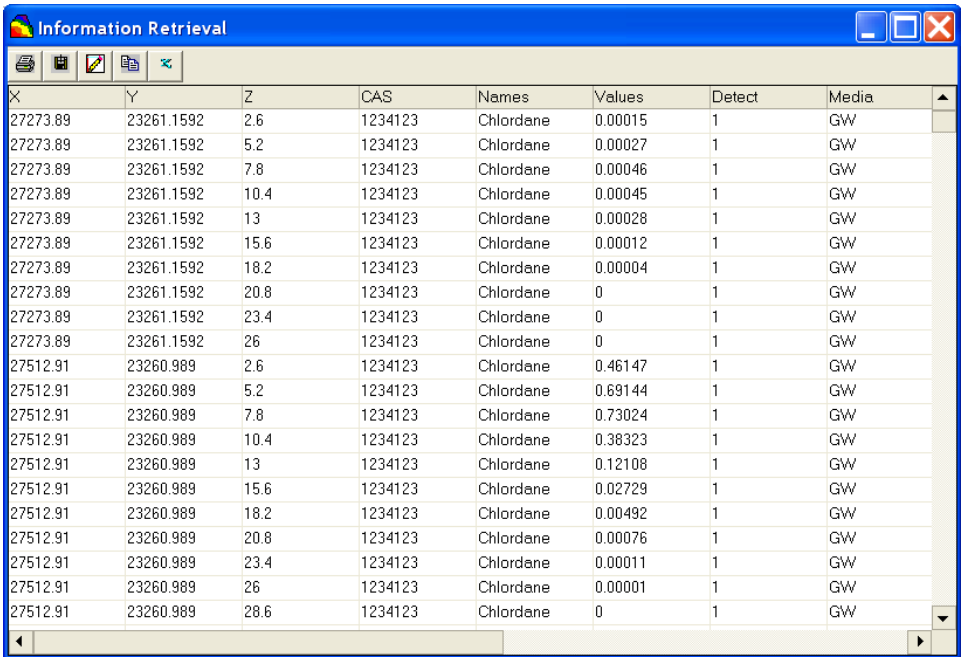


Blue values indicate that this result is a constructed result and is based on multiple values. The blue also indicates that, much like a hyperlink, you can click on the result in this window and "drill down" to see what values were used to construct this result. Click now on any blue text in this window and you'll see a second window pop up with information on the two duplicates that were resolved at this location.



Close both of these windows now.

You can also get all the information for all the points at once by choosing from the main menu *Tools→Information*. Alternatively, you can simply press the *Information* button on the main toolbar (see the help file for more information).You will be presented with the data grid showing you results for all the data.

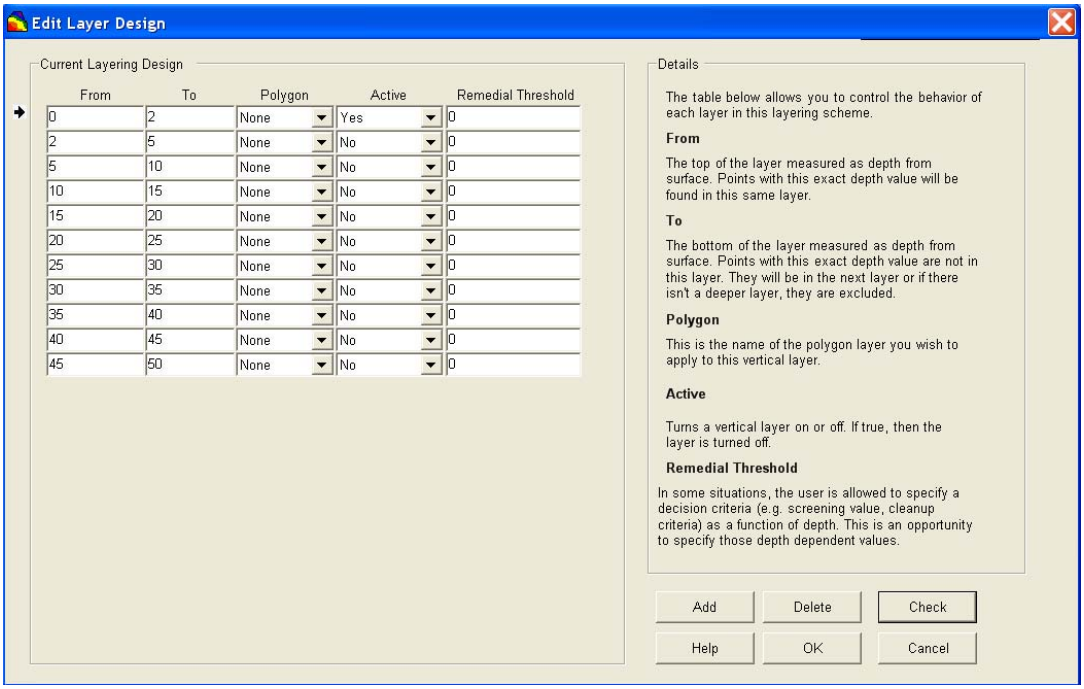


X	Y	Z	CAS	Names	Values	Detect	Media
27273.89	23261.1592	2.6	1234123	Chlordane	0.00015	1	GW
27273.89	23261.1592	5.2	1234123	Chlordane	0.00027	1	GW
27273.89	23261.1592	7.8	1234123	Chlordane	0.00046	1	GW
27273.89	23261.1592	10.4	1234123	Chlordane	0.00045	1	GW
27273.89	23261.1592	13	1234123	Chlordane	0.00028	1	GW
27273.89	23261.1592	15.6	1234123	Chlordane	0.00012	1	GW
27273.89	23261.1592	18.2	1234123	Chlordane	0.00004	1	GW
27273.89	23261.1592	20.8	1234123	Chlordane	0	1	GW
27273.89	23261.1592	23.4	1234123	Chlordane	0	1	GW
27273.89	23261.1592	26	1234123	Chlordane	0	1	GW
27512.91	23260.989	2.6	1234123	Chlordane	0.46147	1	GW
27512.91	23260.989	5.2	1234123	Chlordane	0.69144	1	GW
27512.91	23260.989	7.8	1234123	Chlordane	0.73024	1	GW
27512.91	23260.989	10.4	1234123	Chlordane	0.38323	1	GW
27512.91	23260.989	13	1234123	Chlordane	0.12108	1	GW
27512.91	23260.989	15.6	1234123	Chlordane	0.02729	1	GW
27512.91	23260.989	18.2	1234123	Chlordane	0.00492	1	GW
27512.91	23260.989	20.8	1234123	Chlordane	0.00076	1	GW
27512.91	23260.989	23.4	1234123	Chlordane	0.00011	1	GW
27512.91	23260.989	26	1234123	Chlordane	0.00001	1	GW
27512.91	23260.989	28.6	1234123	Chlordane	0	1	GW

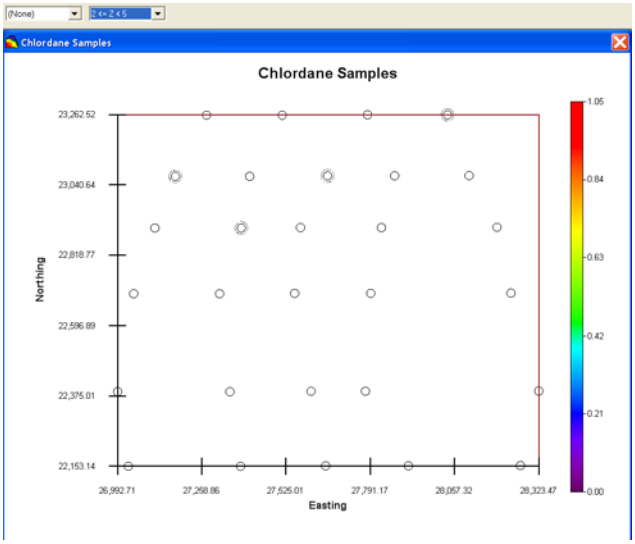
Scroll to the bottom of the grid, and you'll see the duplicate value again represented by blue text. Click on this line in the grid to drill down further and see the duplicates involved in the processes, as before.

Recall that you can use polygon tools to subset areas of the data or even turn layers on and off to exclude or include data. Let's see how that works. Close this window and return *to Set up the site*.

Underneath your selection of MyThinLayers, double click in the layer list box to bring up the layer editor window again. Turn all the layers off (choose *No* in the *Active* column) except the top layer. Your window should look like this before you close it.



Notice that in all layers except 0 to 2, the sample points are not colored. This indicates that they are not currently included. Only those points 0<=z<2 will be included.

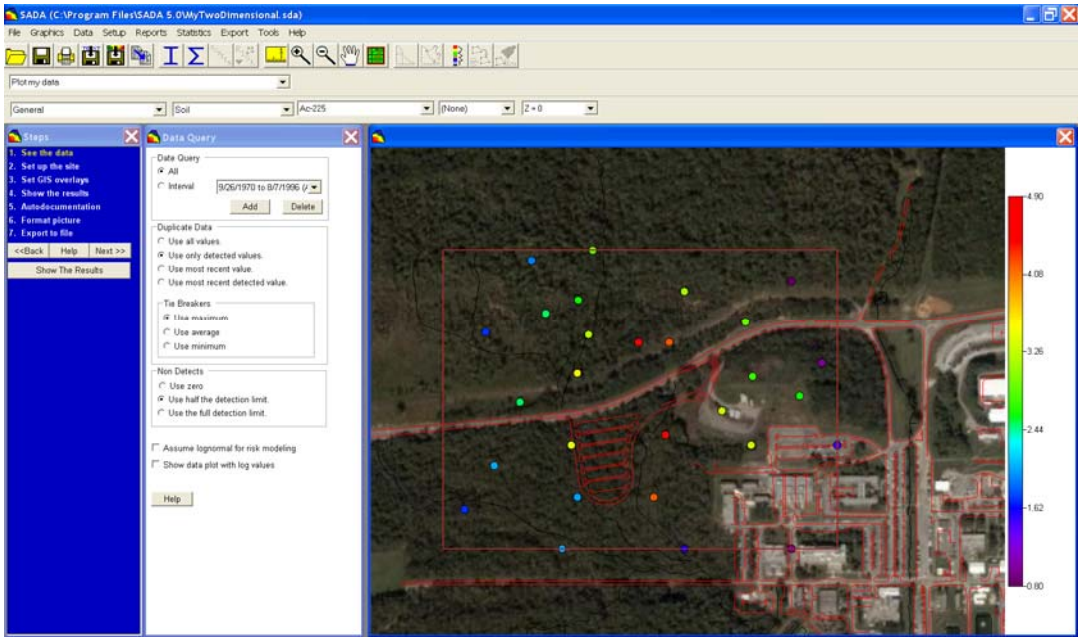


Now press your information button or choose *Tools→Information*. Notice that the data grid now contains only the top layer of points. Close the window.

We will now continue the discussion on data exploration using another file that highlights some different exploration tools. Close MyThreeDimensionalStart.sda and when asked to save, say *Yes*.

Pooling your data sets

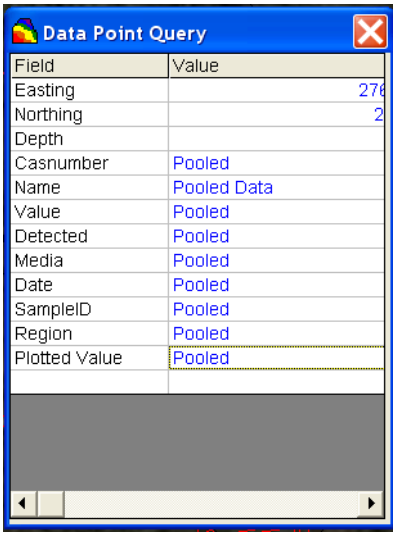
Open the file Twodimensional.sda and save it as MyTwoDimensional.sda. If the file already exists, you can overwrite it. When SADA first brings a file up, you will always be presented with the results of one of your data sets. In this case, we see Ac-225 soil.



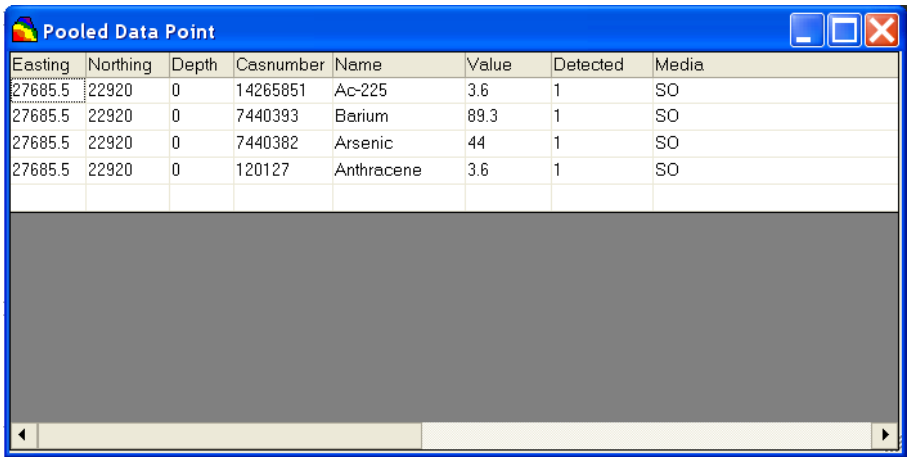
We saw in the previous section how to get information on a single dataset, like Chlordane or Ac-225. You can also get information on all contaminants at once in exactly the same way. Using your data set drop down list, switch from *Ac-225* to *Pooled Data*. The sample locations will turn black.

With this photo in the background, it is difficult to see where they are. To change the color of your pooled data points, select from the main menu *Graphics→Set Various Colors→Pooled Data Point Color* and choose instead yellow.

Save your file. Using your mouse, click on any one of the sample locations and you will be presented with the *Data Point Quer*y window filled mostly with the words *Pooled*.

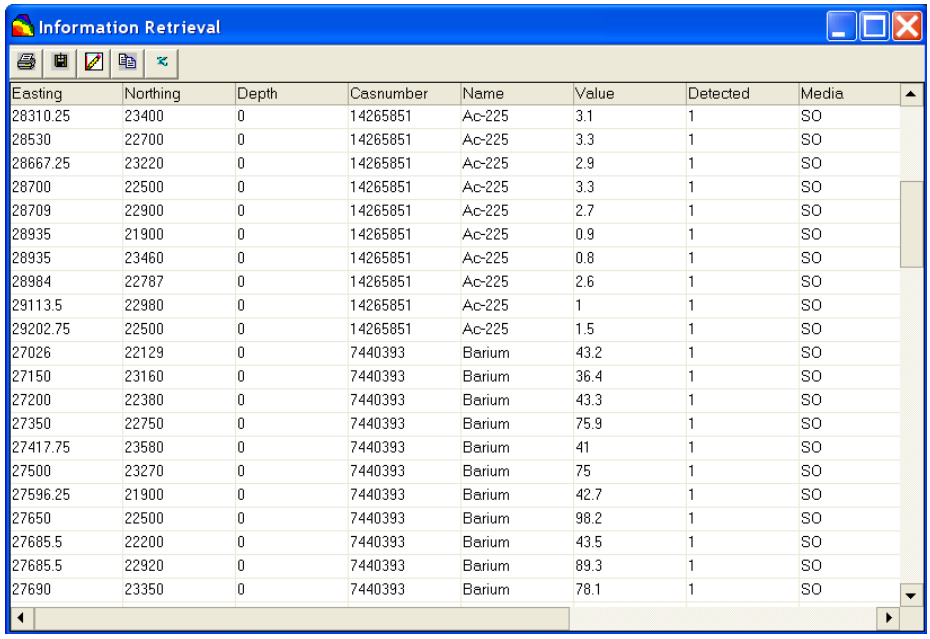


Like in the case of duplicates, since there are multiple points at this location (in this case multiple contaminants), it is impossible to report field values. Again, click on one of the blue text entries to drill down and see what data sets are located at that point.



Close both of these windows now.

The *Information* button (or *Tools→Information*) will now show all the contaminants at once. Select this option now and you should see the following result:



Easting	Northing	Depth	Casnumber	Name	Value	Detected	Media
28310.25	23400	0	14265851	Ac-225	3.1	1	SO
28530	22700	0	14265851	Ac-225	3.3	1	SO
28667.25	23220	0	14265851	Ac-225	2.9	1	SO
28700	22500	0	14265851	Ac-225	3.3	1	SO
28709	22900	0	14265851	Ac-225	2.7	1	SO
28935	21900	0	14265851	Ac-225	0.9	1	SO
28935	23460	0	14265851	Ac-225	0.8	1	SO
28984	22787	0	14265851	Ac-225	2.6	1	SO
29113.5	22980	0	14265851	Ac-225	1	1	SO
29202.75	22500	0	14265851	Ac-225	1.5	1	SO
27026	22129	0	7440393	Barium	43.2	1	SO
27150	23160	0	7440393	Barium	36.4	1	SO
27200	22380	0	7440393	Barium	43.3	1	SO
27350	22750	0	7440393	Barium	75.9	1	SO
27417.75	23580	0	7440393	Barium	41	1	SO
27500	23270	0	7440393	Barium	75	1	SO
27596.25	21900	0	7440393	Barium	42.7	1	SO
27650	22500	0	7440393	Barium	98.2	1	SO
27685.5	22200	0	7440393	Barium	43.5	1	SO
27685.5	22920	0	7440393	Barium	89.3	1	SO
27690	23350	0	7440393	Barium	78.1	1	SO

Scroll down through the rows to see all the soil contaminants. Close the window when you're done.

The method of using meta-data labels is somewhat less helpful. Only in those cases where the field value is exactly the same for all entries will the result show anything other than *Pooled*.

Select *Media* from the drop-down meta labels list.

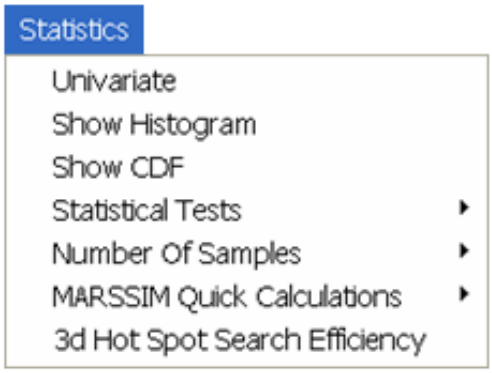
Again the black text is difficult to read. Select *Graphics→Set Various Label Colors→Label Colors* and change to yellow. You can now see that every point has the symbol SO next to it. This means that for every contaminant measured at that location, every result had a media field value exactly equal to SO. This in fact should be the case, since we are looking at soil data. Now switch to *Value*. Notice that every sample point except for two has the familiar three asterisks.

The reason that two of the sample points do report values of 25.20 and 34.40 doesn't mean that all soil contaminants measured there were exactly these values. In this case, it turns that there was only one contaminant measured at both those locations (Arsenic).

You may close your file now. If you wish, you may choose to save the file.

Chapter 12: Statistics

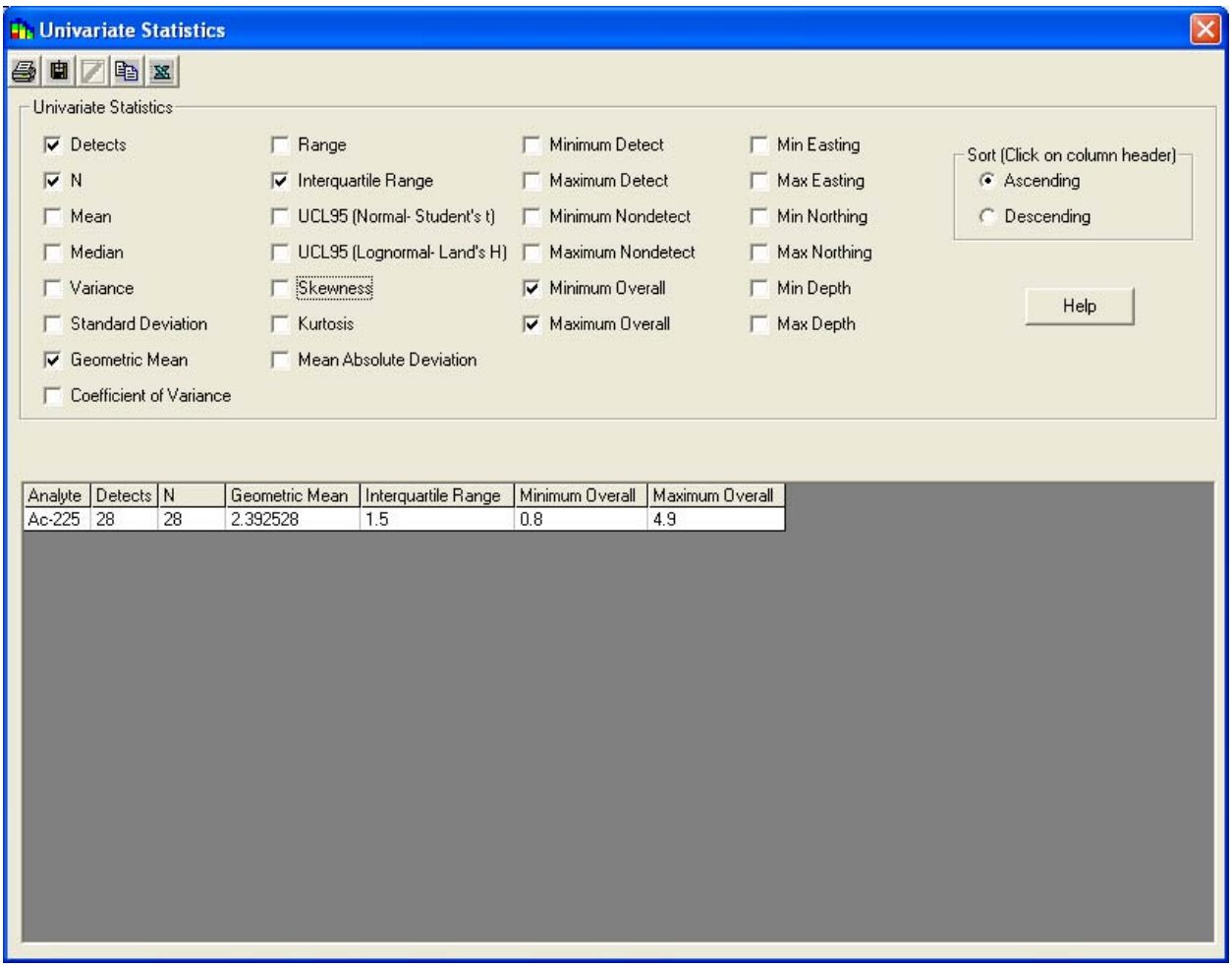
Once you have imported your data or created a geospatial model, you may wish to calculate some simple statistics, run some simple tests, or see some traditional plots. On the main menu, you'll find the *Statistics* menu item, which contains most of the standard statistical features provided.



Let's begin by opening the file Statistics.sda (*File→Open* and navigate to where SADA is installed). SADA opens up your file with the soil measurements for Ac-225.

Univariate

This feature will show you a collection of standard statistical endpoints, such as mean, variance and so forth, for the currently selected contaminant. In this case, we have Ac-225 selected. Choose *Statistics→Univariate* and the following window is presented. Or you can press the statistics button  on the toolbar.



You can choose which statistical values you want to see by selecting or deselecting them in the top half of the window. Also, you can sort columns by different values. Simply click on the column you want to sort and SADA will sort them according to the sort selection (found on the right hand side). Continue reading for details on each statistical endpoint.

Detects

The number of total samples that were detected, presented in the form (detected/total N).

N

The total number of values.

Mean

The sum of the values of a variable divided by the number of values.

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Median

The value below which 50% of the data values fall.

odd n $\tilde{X} = X_{([n+1]/2)}$

even n $\tilde{X} = \frac{X_{(n/2)} + X_{([n/2]+1)}}{2}$

Variance

A parameter that measures how dispersed a random variable's probability distribution is, the mean of the squares of the differences between the respective samples and their mean.

$$S^2 = \frac{\sum \left(X - \overline{X}\right)^2}{n-1}$$

Standard Deviation

The positive square root of the variance.

$$S = \sqrt{S^2}$$

Geometric Mean

The mean of n numbers expressed as the nth root of their product.

$$\exp\left[\frac{1}{n}\sum_{i=1}^{n}\ln x_i\right]$$

Coefficient of Variance

The ratio of the variance over the mean .

$$CV = s / \overline{X} = \frac{\left[\ \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2\ \right]^{1/2}}{\frac{1}{n}\sum_{i=1}^{n}X_i}$$

Range

The difference between the lowest and highest values.

R = maximum - minimum

Interquartile Range

The central portion of a distribution, calculated as the difference between the third quartile and the first quartile; this range includes about one-half of the observations in the set.

IQR = y(75) - y(25)

UCL95 (Normal – Student's t)

Upper 95% confidence limit on the mean concentration of a normal distribution.

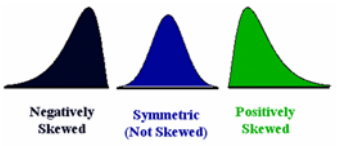
UCL95 (Lognormal – Land's H)

Upper 95% confidence limit on the mean concentration of a lognormal distribution.

Skewness

A measure of symmetry, skewness for a normal distribution is zero, and any symmetric data should have skewness near zero. Negative values for the skewness indicate data that are skewed left and positive values for the skewness indicate data that are skewed right.

$$b_1 = \frac{1}{N}\sum_{j=1}^{N}\left(\frac{x_j - \overline{x}}{\sigma}\right)^3$$



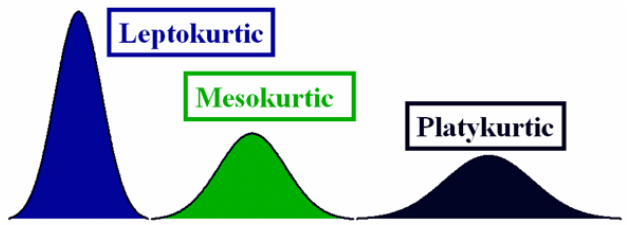
Kurtosis

Kurtosis is a measure of whether the data are peaked or flat relative to a normal distribution. Data sets with high kurtosis tend to have a distinct peak near the mean, decline rather rapidly, and have heavy tails. Data sets with low kurtosis tend to have a flat top near the mean rather than a sharp peak.

$$b_2 = \frac{1}{N}\sum_{j=1}^{N}\left(\frac{x_j - \overline{x}}{\sigma}\right)^4$$

Kurtosis can be grouped into three broad categories

- Leptokurtic: high and thin (high kurtosis values)
- Mesokurtic: normal in shape
- Platykurtic: flat and spread out (low kurtosis values)



Mean Absolute Deviation

The mean of the absolute values of the differences between the respective samples and their mean.

$$M.A.D. = \frac{\sum |X - \mu|}{N}$$

Minimum Detect

The lowest detect value found in the dataset.

Maximum Detect

The highest detected value found in the dataset.

Minimum Non-detect

The lowest detect value found in the dataset.

Maximum Non-detect

The highest detected value found in the dataset.

Minimum Overall

The lowest detect value found in the dataset.

Maximum Overall

The highest detected value found in the dataset.

Min Easting

Smallest easting coordinate value.

Max Easting

Largest easting coordinate value.

Min Northing

Smallest northing coordinate value.

Max Northing

Largest northing coordinate value.

Min Depth

Smallest depth below surface value.

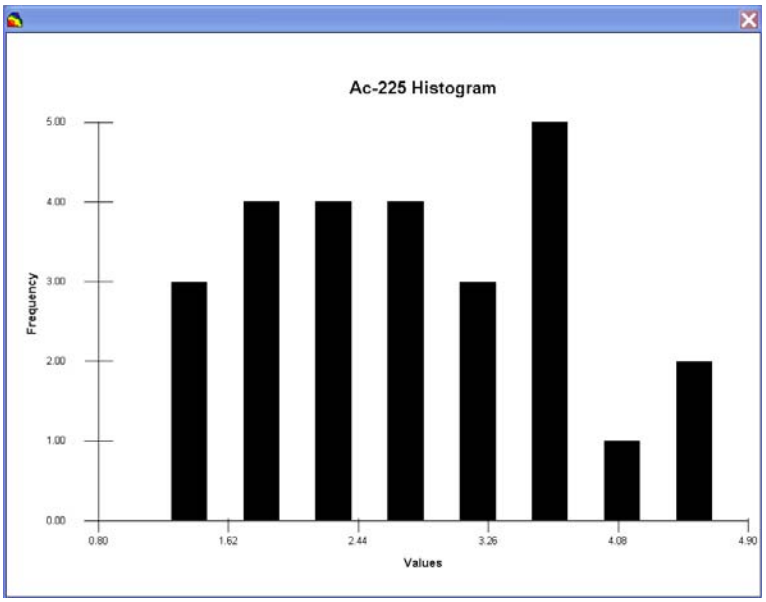
Max Depth

Largest depth below surface value.

Obviously, certain univariate statistics would have little meaning when applied to a set of model values. For example, the upper confidence limit on the mean is meaningless, since it is a function of the number of values, and the number of cells in a grid can be easily varied.

Show Histogram

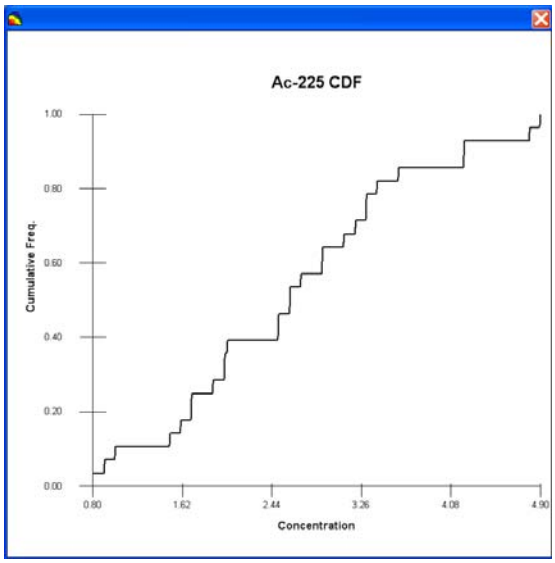
Looking at the histogram of values is very valuable in getting a visual sense for the distribution of your results. The histogram is always shown for the current result in your map. Of course, histograms are not available for everything that may appear in your viewer (e.g., cost benefit analysis). The menu item *Statistics → Show Histogram* will show the histogram view. Select this menu now.



The *Show Histogram* menu actually acts as a toggle. Select it again to switch back to your spatial view.

Show CDF

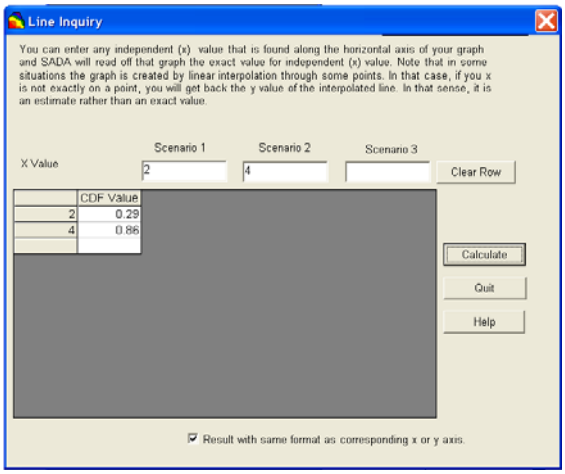
The cumulative distribution function, or CDF, is a natural companion to the histogram and is the focus of evaluation for certain geospatial models (e.g., indicator kriging). To see the CDF, select *Statistics→Show CDF* and the following result appears in your viewer for Ac-225.



On the x-axis is the range of Ac-225 concentration values. On the y-axis is the percentage or, more accurately, the proportion of data values that fall below a given concentration value. So for an x-axis reading of 0.80, we have no data that fall below this value, since this is the smallest Ac-225 reading in the data set. The top reading of 4.90 pCi/g shows a value near 1.0, as this is the highest value in the data set.

You can use the *Line Query* tool to directly read up to 3 values from the line. Select this button now.

You will be presented with this window.



Here, Scenario just means a particular x value. Enter 2 and 4 into Scenario 1 and Scenario 2, respectively. Press *Calculate* and SADA calculates that the proportion of sample values below 2

is 0.29 and 4 is 0.86. Note that you can match the format of reported CDF values with the y axis in the CDF plot by checking the option at the bottom of the window. The Line Query tool can be used with cost benefit plots as well.

Press the *Quit* button.

Statistical Tests

A statistical test is a procedure for deciding whether a hypothesis about a population of values is true or false. For example, one might hypothesize that the mean of the population is less than 5 pCi/g. Of course, the only way to know for certain is to have access to the entire population of values. Here, the population would be the exhaustive set of Ac-225 measurements (e.g., a measured value of Ac-225 for every single location on the site). This of course is impossible. Instead, only a sample of the population is available. Therefore, statistical tests are used to draw a reasonable conclusion about the hypothesis when the entire data set is not available.

Furthermore, statistical tests are instrumental in separating significant effects from mere luck or random chance. For example, suppose that you flipped a coin twice, landing on tails both times. We know that a fair coin should see heads 50% of the time. With so few flips (only 2) we cannot conclude whether the coin is fair or not. We might have tails two times in a row by mere luck, or it might be true that the coin is biased so badly we would get tails every single time. A statistical test would inform us that we cannot separate out luck from true bias in this situation. After 100 flips, however, a test could discern the difference. This is the power of a statistical test.

Of course, without the full population at hand, any conclusion drawn is at risk for error. Fortunately this risk is quantifiable. Two main types of error can occur:

1. A type I error occurs when a true hypothesis is rejected (a false negative in terms of the null hypothesis).
2. A type II error occurs when a false hypothesis is accepted (a false positive in terms of the null hypothesis).

Decision Based on Sample Data	**True Condition**	
	Baseline is True	Alternative is True
Decide baseline is true	Correct Decision	*Decision Error (False Acceptance)*
Decide alternative is true	*Decision Error (False Rejection)*	Correct Decision

*Taken from Guidance for the Data Quality Objectives Process USEPA 2000 QA/G-5, EPA/600/R-96/055

Officially speaking, a null hypothesis is a statistical hypothesis that is tested for possible rejection. The alternative hypothesis is the hypothesis contrary to the null hypothesis.

The process for a statistical test is broadly conducted as follows. First, form a hypothesis about some statistical endpoint (e.g., the mean < 5) of your population. This is your null hypothesis and the test will attempt to reject it by evaluation of the data. Next, draw a random sample from the population (this is the Ac-225 data you have currently selected). Typically, an assumption about the underlying distribution is made (e.g., normal distribution). Using this assumption, choose an appropriate statistical test and compare the result to a critical level.

There are two broad types of statistical tests: parametric and non-parametric. A parametric test makes an assumption about the underlying distribution of observed data. A non-parametric test makes no such assumption.

In particular, non-parametric tests do not assume normal distributions, can handle non-detects (which often cause truncated distribution shapes), are insensitive to outliers and work nearly as well as their parametric counterparts when applied to normally distributed data.

SADA currently implements two non-parametric tests used by the MARSSIM processes: the Sign Test and the Wilcoxon Rank Sum Test.

Whether parametric or nonparametric, there are two different types of test you can conduct: two sided and one sided. A two sided hypothesis states that there is a difference between the two groups (or values) being tested, but does not specify in advance what direction this difference will be. In an environmental context, one group might be Ac-225 measurements in the contaminated area. The second group could be Ac-225 measurements in background. The hypothesis could be that there is no difference between the two groups. A one sided hypothesis states a specific direction (e.g., the site concentrations are greater than the reference site concentrations).

We now present the two nonparametric tests that SADA provides: Sign and Wilcoxon Rank Sum (WRS) Test.

Sign Test

In a sign test, we do the following. First, take the difference between each measured value and the decision criterion. Some measured values will be less than the criterion and will produce a negative difference. Some measured values will be greater than the criterion and will produce a positive difference. Some measured values may be exactly equal to the decision criterion, producing a difference of zero. In the following example, we've taken some Arsenic measurements and calculated the difference between them and a decision criterion of 10mg/kg.

Arsenic	Criteria	Difference	Sign
12	10	2	+
28	10	18	+
8	10	-2	-
42	10	32	+
16	10	6	+
23	10	13	+
45	10	35	+
31	10	21	+

The sign test, true to its name, only cares about whether the sign of the result is positive or negative, as seen in the last column. How large or small the difference is becomes irrelevant. What the sign test does is test whether the number of +'s and –'s are equal. Because of this formulation, the sign test uses wording in the null and alternative hypotheses that is a bit unintuitive, so let's take some time with it. The null and alternative hypotheses (Gilbert 1987, p. 242) are written as follows.

- H0: the median of the population of all possible differences is zero.
- H1: the median of the population of all possible difference does not equal zero.

Let's stop for a moment and understand what is being said here with less formal language. Let's start with the "population of all possible differences." If you could take every single Arsenic measurement of the site, you would have the population of Arsenic values. If you then

subtracted them from the decision criterion (10mg/kg) and noted the number of +'s and –'s, you would move yourself from the population of measurements to the population of differences.

In the table above, we are subtracting measured values from the decision criterion and ending up with differences from which we record the number of +'s and –'s. So, we move ourselves from the sample of "measured values" to a sample of "differences."

Why do we care about the median of these differences? If the population (not the sample) of differences has exactly the same number of +'s as there are –'s, then we have central or median value of zero. Think of these as +1s and -1s, and if you have the same number, then the median of these would be halfway between or zero. Note that if you end up with any difference of exactly zero, it is thrown out (Gilbert 1987).

So what does this have to do with determining if our site is contaminated or not? Let's start by reformulating the hypothesis into an equivalent statement.

- H0: Measured values are just as likely to exceed the criteria as they are to be less than the criteria.
- H1: One of the following is true:
    - measured values are more likely to exceed the criteria.
    - measured values are more likely to be less than the criteria.

These are equivalent to the first formulation. In environmental assessment, we are more likely to be interested in H1A. We could then rewrite this as the following.

- H0: Measured values are just as likely to exceed the criteria as they are to be less than the criteria.
- H1: Measured values are more likely to exceed the criteria.

For the example we show above, let's write this one more time very specifically.

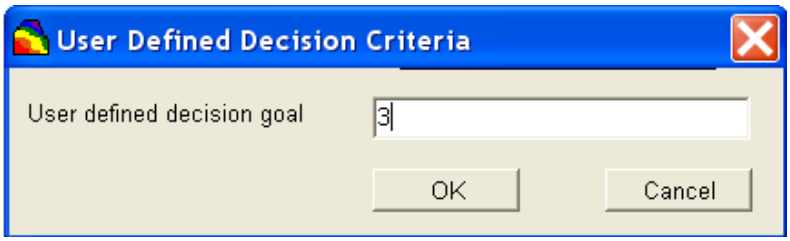
- H0: Arsenic values are just as likely to exceed 10mg/kg as they are to be less than 10mg/kg.
- H1: Arsenic values are more likely to exceed 10mg/kg.

The critical value used to conduct this one-sided test is provided as the number of +s that you must have to reject the null, and it comes from the binomial distribution. The sign test critical value tables can be found in introductory statistics books. In SADA, it is computed directly. Let's test the hypothesis now.

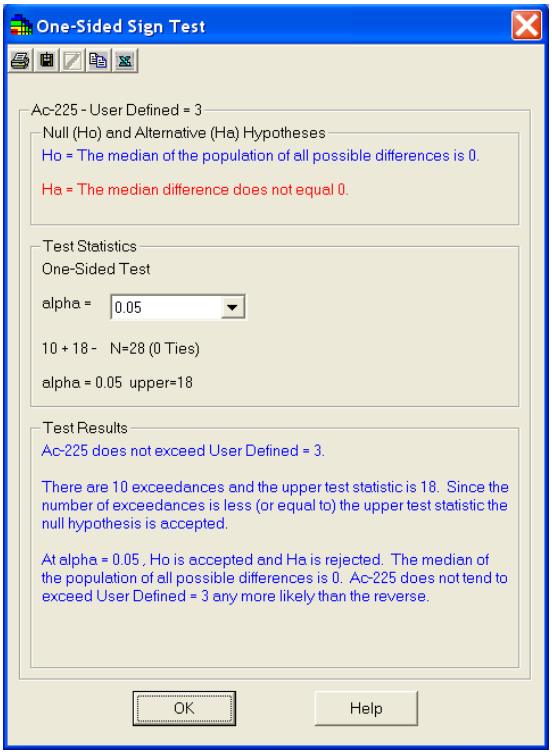
In our example, we want to be very sure that if we do reject the null hypothesis, we want there to be only a 5% chance we did the wrong thing. This is the type I error we talked about earlier. This value is referred to as alpha (so we have alpha = .05). The next thing we need is test statistic B and a critical value to compare it to for alpha = .05.

In our example, counting up the number of samples yields N=8 and the number of exceedances is B=7. Now, turning to the binomial distribution table (most statistics books have the table in the back) for a one-sided sign test with alpha = .05, N=8, and B = 7, we see the critical value is also 7. Therefore, since our test statistic equals the critical value, we are led to reject the null hypothesis. Since we computed the number of exceedances as our test statistic, this would in fact indicate the median is exceeding 10 mg/kg. This means that more often than not, Arsenic values exceed the decision criterion. When we apply our decision rule, we infer that the site is contaminated.

To use the sign test in our current file, select from the main menu *Statistics→Statistical Tests→ Sign Test vs Decision Criteria*. If you are working under *General* analysis, you will be presented with the following window asking for a decision criterion.



The *Human Health*, *Ecological*, and *Custom* analyses would present their usual windows for choosing decision criteria. These will be covered a little later. Enter a value of 3 here (pCi/g) and press *OK*. The Sign Test window is presented. At the top are the null and alternative hypotheses. *Note: The two-sided alternative is written there, but the one-sided formulation would have been more appropriate.* Look for updates as SADA releases continue. Nevertheless, we are doing a one-sided upper test where the HA is that the median exceeds zero.



So, at alpha = 0.05, with N=28, a decision criteria of 3pCi/g, and a test statistic of 18, we cannot conclude that the median of Ac-225 values exceeds 3 because we only have 10 exceedances.

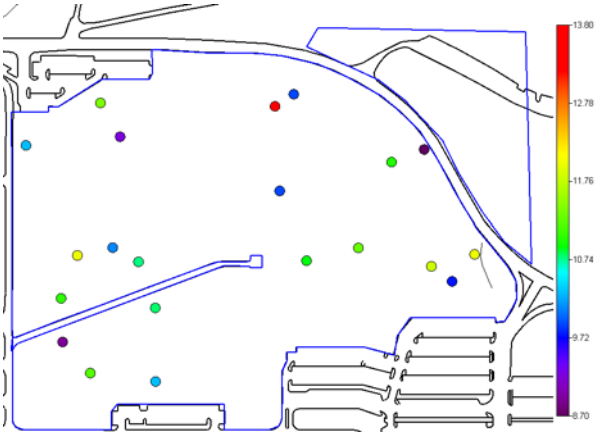
Wilcoxon Rank Sum Test (WRS)

The WRS is another non-parametric test that is implemented as either a one-sided or a two-sided test in SADA. In this case you compare one contaminant data set versus another data set (e.g., background).

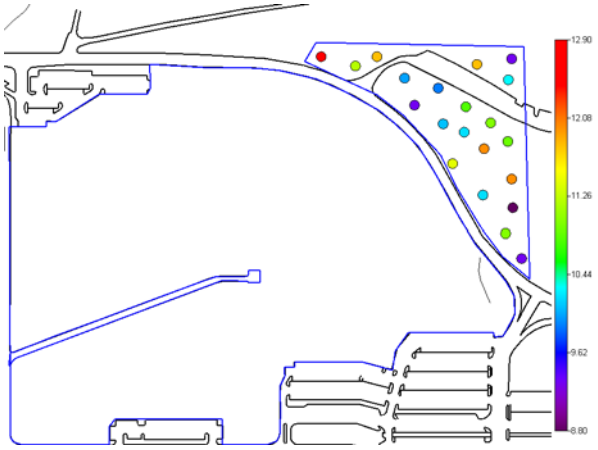
Specifically, the WRS test uses a sum-of-ranks comparison to determine if the two data sets have different means. The process is conducted as follows. First, combine data sets and order

from lowest to highest, each data value receives a rank (e.g., 1st, 2nd, 3rd, etc). Sum the ranks for the two different populations and compute the WRS test statistic (different forms depending on if there are ties). Compare to critical value for m and n sample sizes. Null hypothesis is either accepted or rejected.

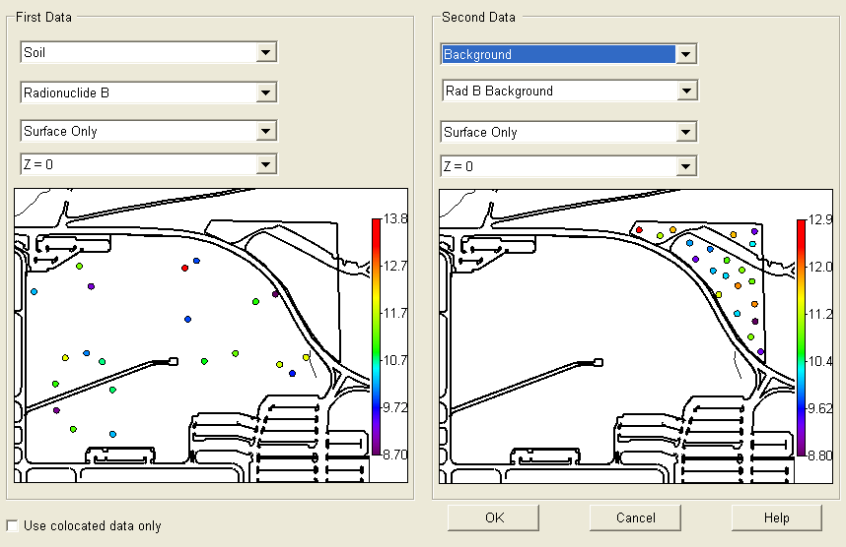
To see how this works, first open the file RadWithBackround.sda (you can save your current file or not). This SADA file contains some synthetic radiological data along with some background measurements. With *Soil* selected, choose *Radionuclide B*. You should see the following map:



To see the background data, select *Background* instead of *Soil*. The background data for Radionuclide B will display, since it is the first and only background dataset in this file.



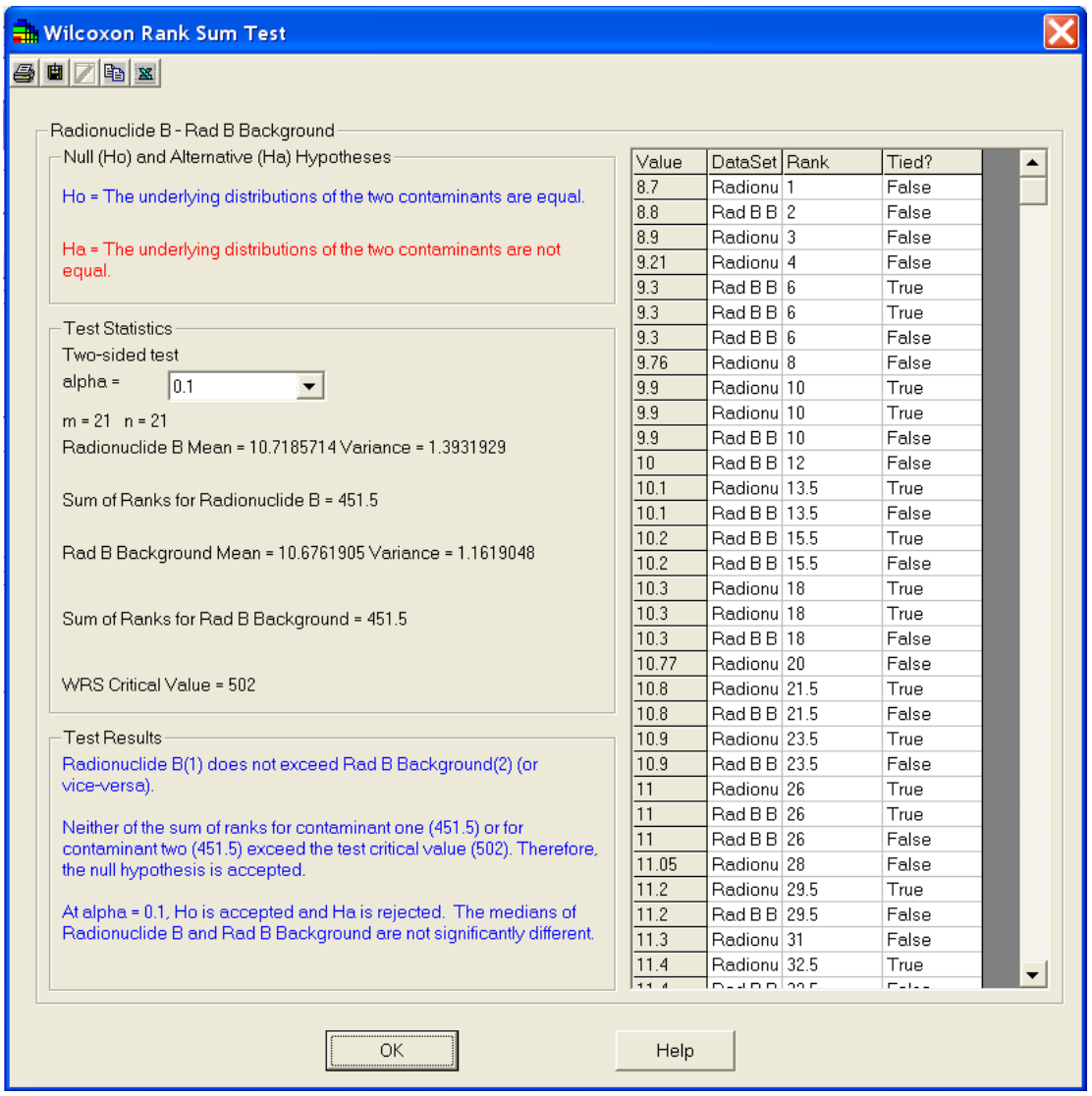
Next, we will perform the WRS test to see if the site data is greater than the background. To do this, select from the main menu *Statistics→Statistical Tests→Wilcoxon Rank Sum*, and the WRS test feature is initiated with the following window. Select in this window the two data sets you want to compare. In this case, select *Radionuclide B* from the left window and *Radionuclide B Background* from the right window. SADA shows a preview of each.



Two important points should be brought to your attention here. First, in the lower-left-hand corner is the option *Use collocated data only*. There are some situations where data occur concurrently and this may be appropriate. In this example, where we are comparing site and background data, we have no collocated data and so this is not appropriate.

Secondly, notice that you can select layering schemes as well. This is important. You may want to set up a particular layering scheme with polygons embedded in such a way as to include or exclude a set of data. For example, you may have background data from two locations and wish to exclude one while you use the other. In the other dataset selection, a different layering/polygon scheme may be appropriate. This permits you the flexibility to spatially zero in on two or more different locations when doing the WRS test. In this example, we will use the *Surface Only* layer with no polygons other than our site boundary polygons.

Press *OK*.



Value	DataSet	Rank	Tied?
8.7	Radionu	1	False
8.8	Rad B B	2	False
8.9	Radionu	3	False
9.21	Radionu	4	False
9.3	Rad B B	6	True
9.3	Rad B B	6	True
9.3	Rad B B	6	False
9.76	Radionu	8	False
9.9	Radionu	10	True
9.9	Radionu	10	True
9.9	Rad B B	10	False
10	Rad B B	12	False
10.1	Radionu	13.5	True
10.1	Rad B B	13.5	False
10.2	Rad B B	15.5	True
10.2	Rad B B	15.5	False
10.3	Radionu	18	True
10.3	Radionu	18	True
10.3	Rad B B	18	False
10.77	Radionu	20	False
10.8	Radionu	21.5	True
10.8	Rad B B	21.5	False
10.9	Radionu	23.5	True
10.9	Rad B B	23.5	False
11	Radionu	26	True
11	Rad B B	26	True
11	Rad B B	26	False
11.05	Radionu	28	False
11.2	Radionu	29.5	True
11.2	Rad B B	29.5	False
11.3	Radionu	31	False
11.4	Radionu	32.5	True

Here the null hypothesis is that the two groups are the same (site is no different than background). The alternative is that they are different. At the alpha value of .1, we accept the null hypothesis that the two populations are different and might conclude the site is clean. Had the null been rejected, SADA would have also conducted a one-sided test to determine if the site data in fact exceeds background.

Number of Samples

This topic is handled in the Chapter, “Overview of Sample Designs.”

MARSSIM Quick Calculations

This topic is handled in Chapter 40, “MARSSIM.”

3d Search Efficiency

Among SADA’s many sampling strategies is the 3d Hot Spot search design, where a 3d grid of candidate locations is positioned on the site and the probability of discovering a particular 3d

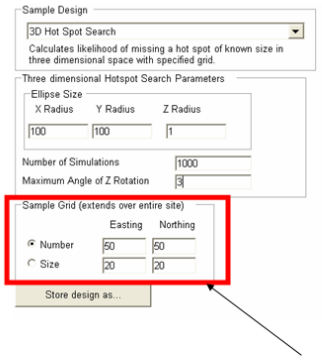
volume by sampling in those locations is calculated (see Chapter 39). In this feature, we generalize it to any sample design, gridded or not, and apply it to existing sample designs. In other words, to use this feature, you will need to import your data first. There are a couple of situations in which this feature can be useful.

First, a 3d search grid was implemented in the past, but when the samples were taken, adjustments had to be made due to physical obstructions or cost. This feature can analyze your resulting design and see how the probability of finding a 3d hotspot might be affected.

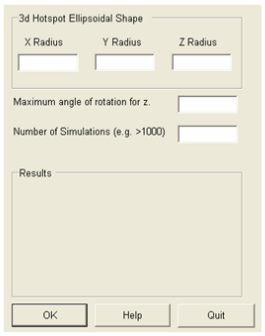
Second, the data you have may originate from a sample design that was not primarily interested in searching for a volume of a given size. Now the issue has arisen, however, and the interest is in evaluating how well it serves the 3d discovery objective.

The details of this feature are found in the Chapter 39 “Standard Initial Sample Designs” under the 3d Hot Spot Search discussion. You are encouraged to read that material first. We will point out here that the only difference in the interface is that under the sample design strategy, you will enter a candidate 3d grid which is then evaluated. Under this feature, the currently selected design is evaluated.

New 3D Sample Design Window

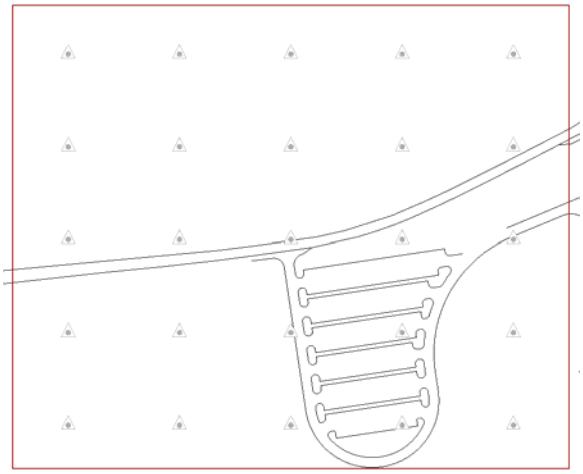


Existing 3D Sample Design Window

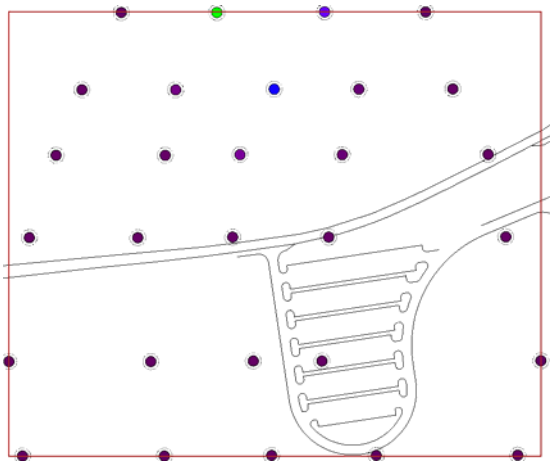


The initial sample design proposed and evaluated here is replaced by the actual sample locations in the 3d Search Efficiency Feature

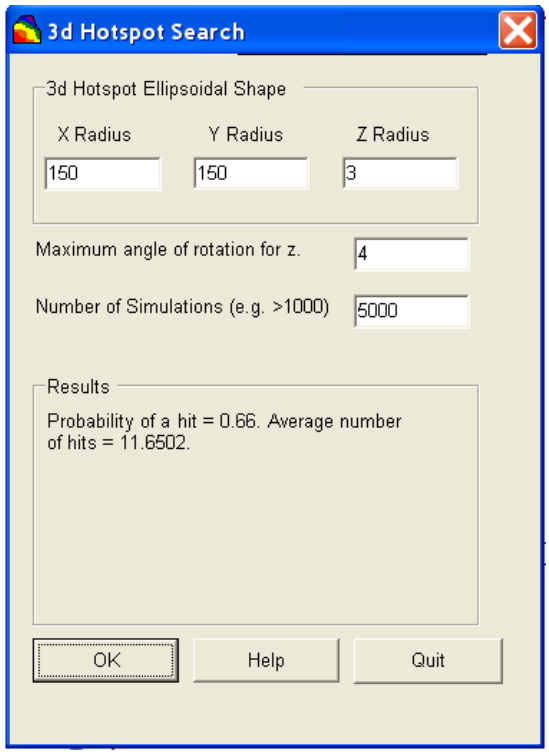
For a quick demonstration, open the file ThreeDimensional.sda. This is a 3d data set that was originally a 5 x 5 3d design with a 74% probability of discovering an elliptical 3d volume of dimensions 150 x 150 x 3ft (see Chapter 39 for definition of 3d volume geometry).



When the samples were collected, however, they could not adhere to this strict gridded pattern. As you can see by the Chlordane measurements, adjustments were made, resulting in the following map.



In this hypothetical example, two things are observed. First, this deviates from the strict sampling grid. Secondly, suppose that a screening of the values yielded no points above the decision criterion. The question becomes, could we have missed an elevated area with more likelihood than 26% (100-74=26%)? To find out, select from the *Statistics* menu *3d Hotspot Search* efficiency. Enter the parameters as you see them below and Press *OK*.



Here, we see in the *Results* summary that the probability of a hit (or probability of discovering an elevated ellipsoidal volume) is now about 66% (your value may vary slightly due to the nature of random realizations/simulations). This represents about an 8% (74-66=8%) reduction in reliability due to the adjustments made at sample time. The average number of hits is around 11.6. This means that on average, when a 3d ellipsoid is simulated, a hit is “discovered” or “hit” by about 11 existing samples each time. Press the *Quit* button to close the window.

Summary

SADA provides a basic set of statistical features. If you need other statistics, you can export your data out of SADA into a csv file, where it can then be imported into a statistics package. Finally, many of these statistical features apply to both gridded and point data. It is important, particularly in the case of some univariate statistics, to properly interpret gridded statistics. For example, SADA will report a UCL95 for a modeled dataset. This value is meaningless, however, as it depends on the number of values, which is arbitrarily set by the modeler. For an overview of environmental statistics, we find the book by Richard Gilbert, *Statistical Methods for Environmental Pollution Monitoring*, to be an excellent source.

References

Gilbert, R. 1987 Statistical Methods for Environmental Pollution Monitoring, Wiley & Sons, NY.

Chapter 13: How to Perform a Quick Spatial Data Screen

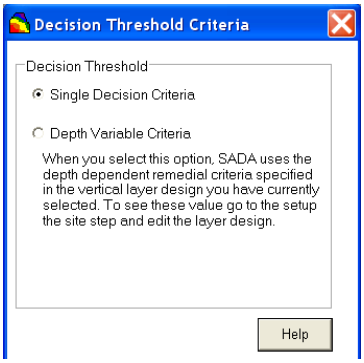
One of the earliest decision tools you can apply to your data is a simple data screen. In this tool, SADA uses a decision criterion to spatially identify those point values that exceed a screening criterion. There are different ways to arrive at a screening value. You could simply provide one just before the screen; screening values can be imported into SADA using the Custom Analysis feature; screening values in the form of preliminary remediation goals, or PRGs, can be calculated by SADA based on your calibration of the human health risk model; or screening values in the form of ecological benchmarks can be used after you've setup the ecological risk assessment. In this chapter, we will deal with the simplest scenario, where you will simply provide the value to SADA as the screening tool is initiated. The other methods have chapters dedicated to them later on in this user guide.

Open ThreeDimensional.sda (*File→Open*).

Regardless of how you arrive at a screening value, you will access the screening tool by choosing the interview *Draw a data screen map*. Select this interview now.

Anytime your selected analysis is General, you will be prompted for a screening value each time your run a screen. Make sure General is selected now.

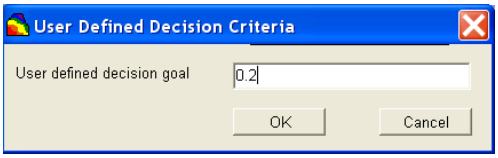
The steps for this interview are very straightforward. The first three steps deal with querying the data, setting up the site, and using GIS layers. These have been covered in previous chapters. What we are interested in here is the step *Set Decision Threshold Type*. Click on this step now.



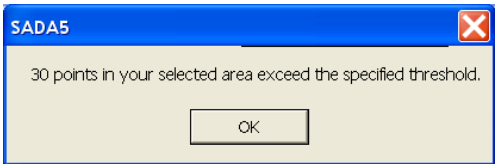
This step, or option, is available only for the General Analysis. On the parameter window, you can screen your Chlordane data set against a single value or against screening criteria that vary as a function of depth.

Single Decision Criteria

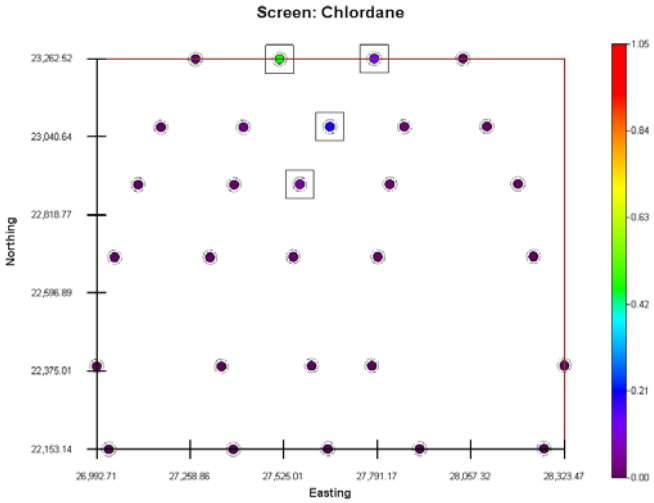
Let's choose *Single Decision Criteria*. Press the *Show The Results* button, and you will be prompted to enter a single screening value for Chlordane.



Enter a value of 0.2 and press *OK*. SADA informs you that you have 30 points exceeding this screening value.



Press *OK*. SADA draws a square box around each value that exceeds 0.2. Notice in the following image only four samples have boxes around them. This is because other samples that are exceeding .2 are obscured from view in this 2d layer view. Some of them may exist below this layer, while multiple exceedances may occur at the same location in this layer. For this reason, SADA always reports the total number of exceedances so you know there may be more below or above what you currently see.

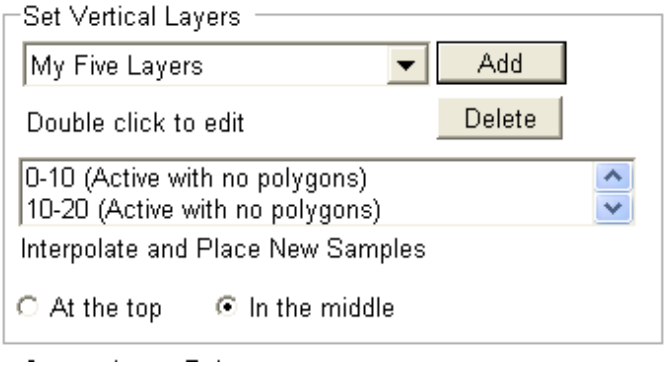


Sometimes it's difficult to see the boxes, particularly with a photograph located directly behind them. This is easy to change. From the main menu, choose *Graphics→Set Various Colors→Screening Box*. You'll be presented with a color palette. Select bright red and press *OK*. You should be able to see red squares now.

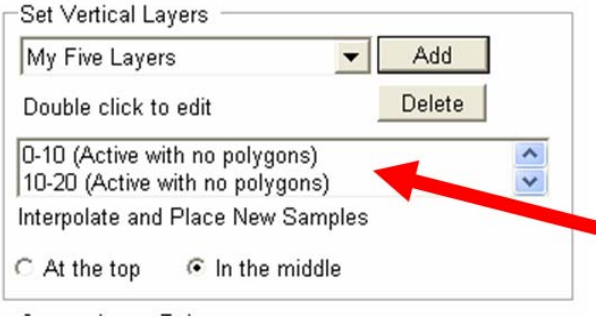
Depth Variable Criteria

In some situations, the screening criteria may vary with depth, becoming perhaps more relaxed when further from the surface. This can be handled by choosing the *Depth Variable Criteria* option. It is important to know that currently SADA assigns screening/decision criteria of this nature directly to the layering designs. So, you'll need to enter the screening criteria in the vertical layering editor.

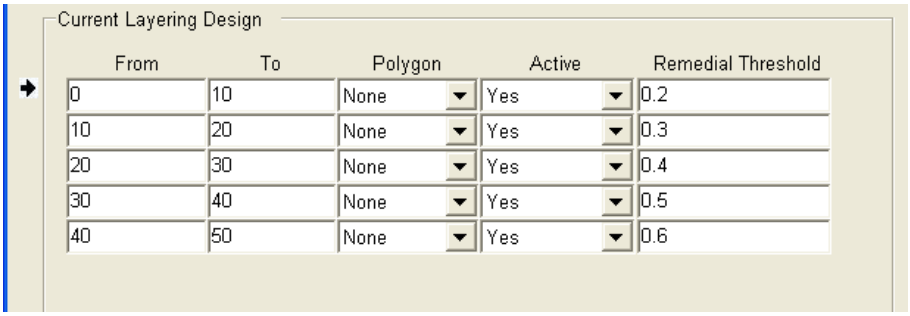
First select *Depth Variable Criteria*. Then click on *Setup the site*. Make sure you have the vertical layer design *My Five Layers* selected.



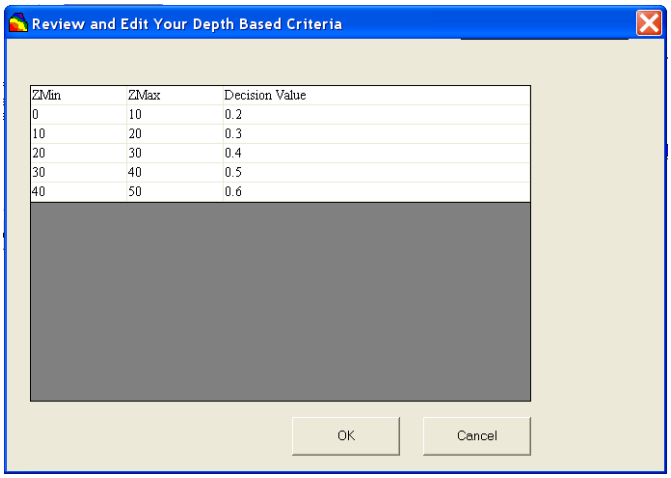
Double click the listing of individual layers to open the layer design window.



In the layer design window, enter the depth variable screening criteria in the remedial threshold column exactly as you see it in the following picture:



Press *OK*. Now press the *Show The Results* button, and SADA will allow you to look at and edit your depth variable screening criteria one more time before the screen is performed.



Press the *OK* button, and SADA reports now that 20 points exceed the screening criteria. This is less than 30, because the screening criteria increased above 0.2 as you increase depth; therefore, fewer samples in the subsurface exceeded the criteria assigned to their layer. Switch between the different layers to see which points exceeded their layer specific criteria.

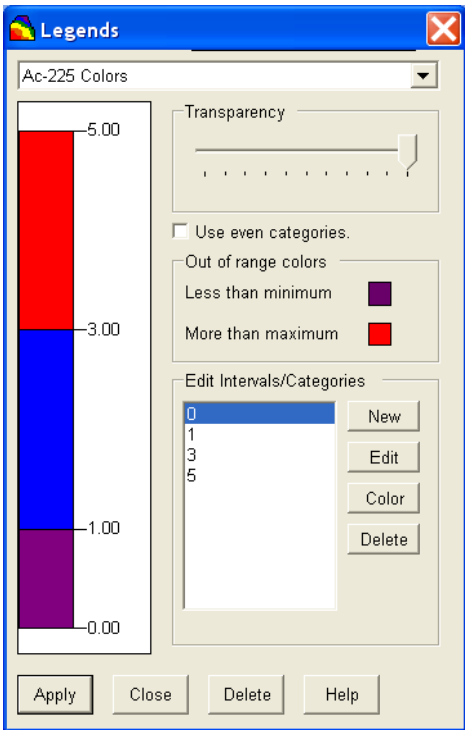
What about multiple contaminants?

Of course, if you have multiple contaminants with depth variable criteria, this can be a bit cumbersome but certainly doable. In future versions of SADA, we will improve this feature to work with multiple sets of depth variable criteria as well as with human health, custom analysis, and so forth. In the mean time, suppose you had two contaminants with depth-dependent criteria (Arsenic and Ac-225). You would first create a layer design for Arsenic and fill in the depth variable criteria. Then, you could create a second layer design named *Ac-225 layers* (perhaps based on Arsenic layers) and fill in the depth dependent criteria for Ac-225. This isn't too bad, except you have to remember to switch the vertical layers when you switch the contaminants.

Using the Legend Manager as a Screening Tool

The legend manager is covered in greater detail in Chapter 43. Here, we'll focus on how to use the legend manager to color ranges of numbers with different colors. This is essentially like a data screen tool itself.

Close any file you may have open without saving and open TwoDimensional.sda. From the main menu, select *Graphics→Legend Manager*. From the drop-down list of legends, select *(new interval)* and name the legend "Ac-225 Colors." Press *OK*, and select *Ac-225* from the list of legends. Modify the legend so that it looks like the image below. If you don't know how to do this, see the sections on the Legend Manager in Chapter 43.



This is really a two-part screen. All values above 3 will be assigned a red color. All values between 1 and 3 will be assigned blue. Purple is assigned for the lowest value. Press the *Apply* button and you can easily see the exceedances in the graphics window.

Entering by hand is the simplest ground-floor method of screening your data against set criteria. In the next chapter, we'll look at importing screening criteria from outside SADA. You can now close your Twodimensional.sda without saving it.

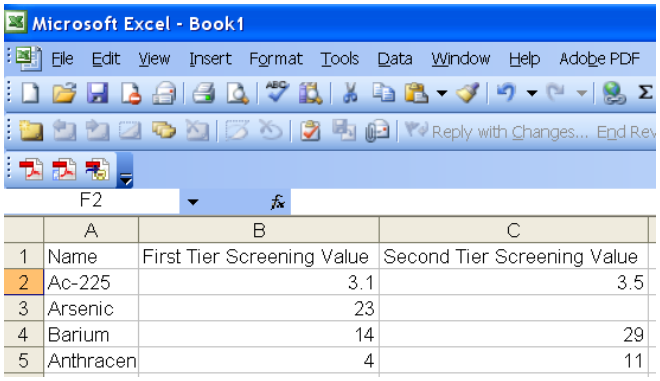
Part IV: Risk

Chapter 14: Importing Custom Screening and Remediation Criteria (Custom Analysis)

In the last chapter, we saw how to enter screening criteria by hand for spatial data screens. This can be monotonous if you do a lot of screening or have multiple contaminants. The best approach is to just import the criteria in from outside SADA. This is a custom analysis and requires only a brief setup on the user's part. Once imported into SADA, these values can be used not only to screen data but to determine areas of concern and support sample design.

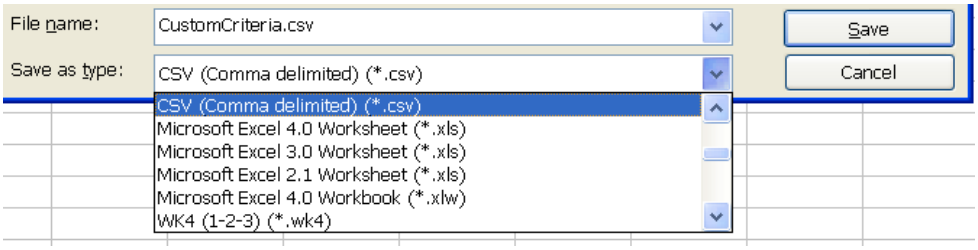
Preparing Custom Values

The first thing you'll need to do is create a table of custom values using a spreadsheet editor or perhaps an Access database. All SADA needs is a name column and one or more columns containing screening values. Columns can be in any order, and missing values are allowed in the screening columns. If you have a column of CAS numbers, this can speed up the import process a little but isn't necessary. The following is a simple table created in Excel and has been included in your installation files as CustomCriteria.xls. If you wish you can open it now.



	A	B	C
1	Name	First Tier Screening Value	Second Tier Screening Value
2	Ac-225	3.1	3.5
3	Arsenic	23	
4	Barium	14	29
5	Anthracen	4	11

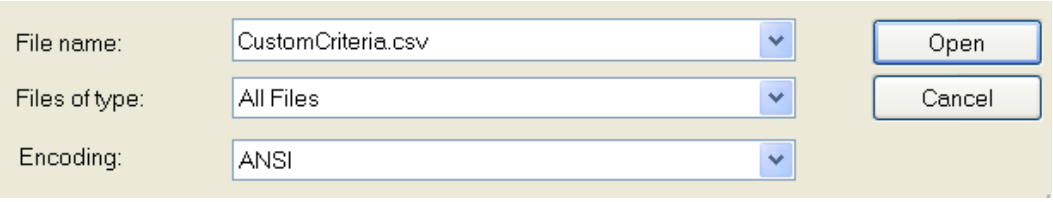
All you'll need to do in Excel is to save this as a comma separated value file (.csv). To do this, select, in Excel, File→Save As… and select Comma delimited (*.csv) from the *Save as type* at the bottom of the window.



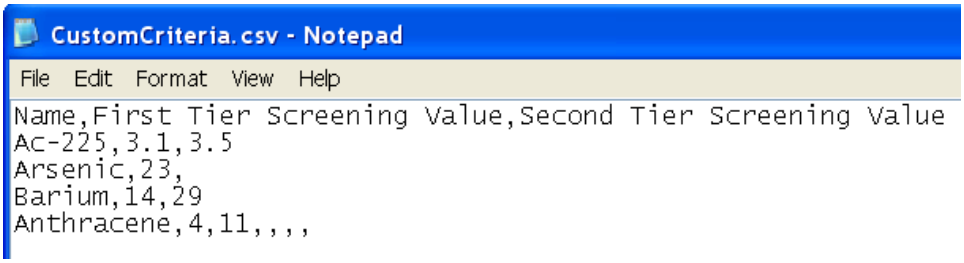
Enter CustomCriteria.csv as the name and press *Save*. Since CustomCriteria.csv is already installed with SADA, you maybe warned about an overwrite. If so, choose *No*. Otherwise, SADA will ask you some questions about losing features of the Workbook and so forth. Just press through this until you can finally save it.

Now let's look at the .csv file to make sure it is in good shape, using a simple text editor called Notepad. Sometimes, Excel-generated custom analysis files have the same problems as Excel-based data files, which are discussed in the Chapter 5, "Importing Sampled Data" (namely, the arbitrary and unpredictable placement of extraneous commas in the file).

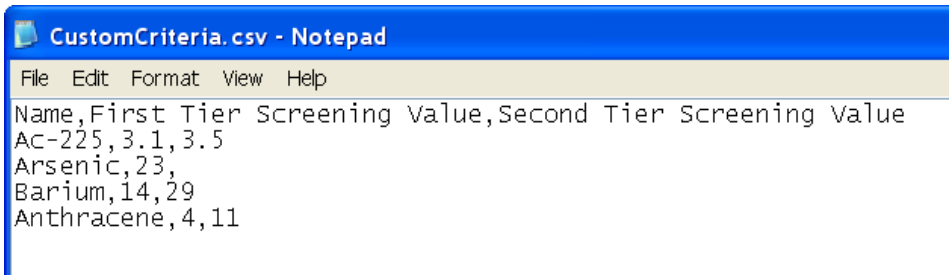
Let's start by opening Notepad. From the Windows Start button, select *All Programs→Accessories→Notepad*. Depending on how your desktop is set up, your Notepad may be located elsewhere. Once Notepad is open, choose *File→Open*, navigate to where SADA has been installed (likely C:\Program Files\SADA 5.0), and open CustomCriteria.csv. In the open window of Notepad, you'll need to make sure *Files of type* is set to *All Files.*



Select CustomCriteria.csv and press *Open*. Your file may look slightly different than the following image, which shows a common problem with .csv files and Excel. The last row for Anthracene has extraneous commas added. You'll need to clean this up by hand.

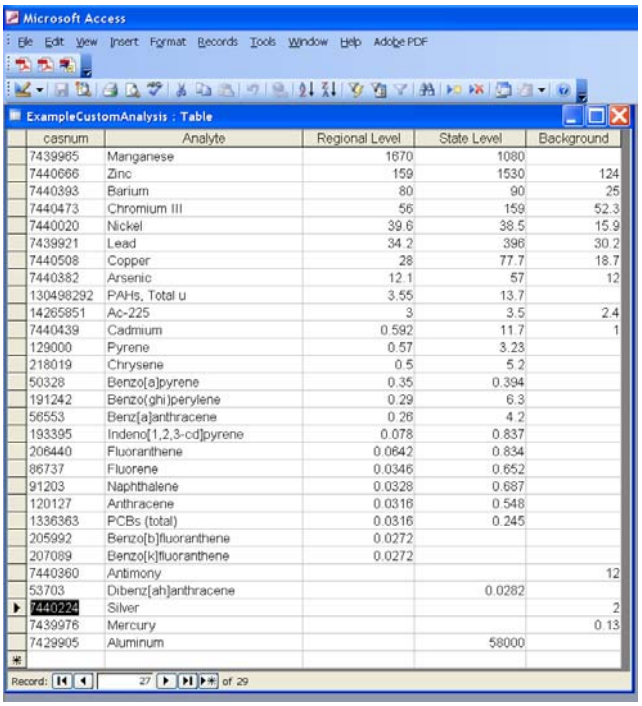


Simply delete any unnecessary commas and save the file again.



Then, close Notepad.

An example of what custom criteria looks like in a Microsoft Access database has been sent with your SADA installation. The file CustomCriteria.mdb contains some additional example criteria as well as a CASNUM field.



casnum	Analyte	Regional Level	State Level	Background
7439965	Manganese	1670	1080	
7440666	Zinc	159	1530	124
7440393	Barium	80	90	25
7440473	Chromium III	56	159	52.3
7440020	Nickel	39.6	38.5	15.9
7439921	Lead	34.2	398	30.2
7440508	Copper	28	77.7	18.7
7440382	Arsenic	12.1	57	12
130498292	PAHs, Total u	3.55	13.7	
14265851	Ac-225	3	3.5	2.4
7440439	Cadmium	0.592	11.7	1
129000	Pyrene	0.57	3.23	
218019	Chrysene	0.5	5.2	
50328	Benzo[a]pyrene	0.35	0.394	
191242	Benzo(ghi)perylene	0.29	6.3	
56553	Benz[a]anthracene	0.26	4.2	
193395	Indeno[1,2,3-cd]pyrene	0.078	0.837	
206440	Fluoranthene	0.0642	0.834	
86737	Fluorene	0.0346	0.652	
91203	Naphthalene	0.0328	0.687	
120127	Anthracene	0.0316	0.548	
1336363	PCBs (total)	0.0316	0.245	
205992	Benzo[b]fluoranthene	0.0272		
207089	Benzo[k]fluoranthene	0.0272		
7440360	Antimony			12
53703	Dibenz[ah]anthracene		0.0282	
7440224	Silver			2
7439976	Mercury			0.13
7429905	Aluminum		58000	

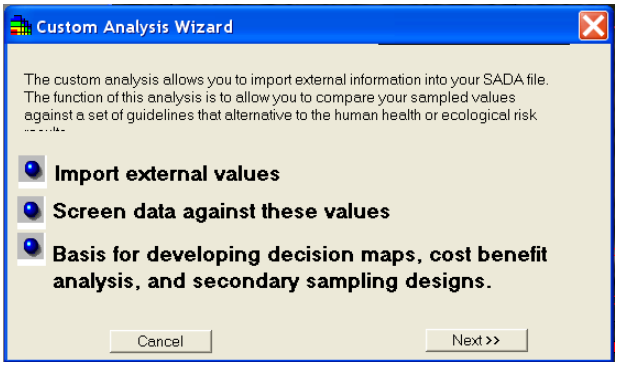
Record: 27 of 29

Both of these files can be used as templates for establishing your own custom criteria files. Close Excel and/or Access if you have them open. Let's import CustomCriteria.mdb now.

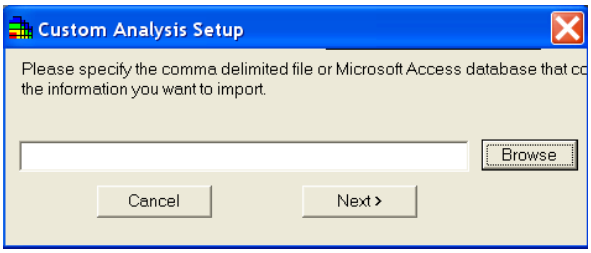
Importing Custom Values

The import process is relatively trivial. Open SADA and open the file Twodimensional.sda. Save the file as CustomExample.sda. To setup a custom analysis, from the main menu select *Setup→Custom….*

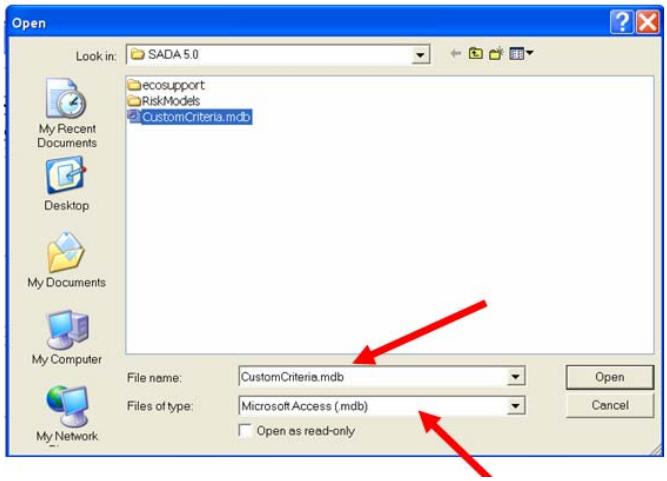
This initiates the custom setup processes.



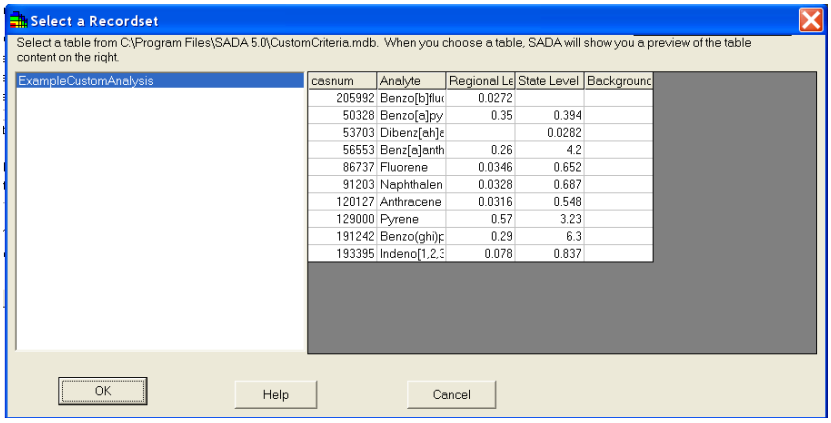
Press *Next* >>. The next window wants the name of the file containing your screening values.



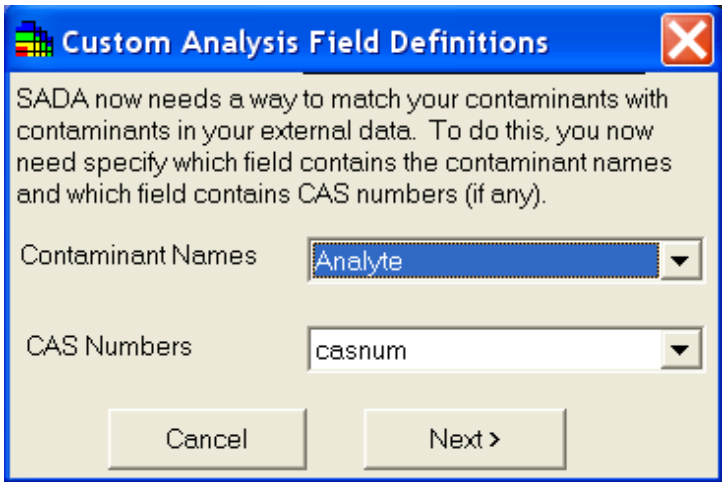
Press the *Browse* button, change the *Files of type* to *Microsoft Access (.mdb)*, and select CustomCriteria.mdb.



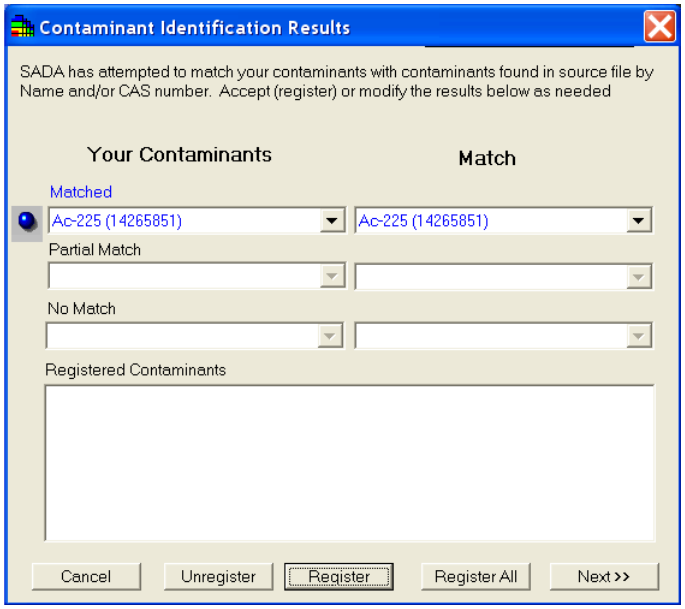
Press *Open* and *Next>>* on the last window. For Access files, you'll need to choose the proper table in your database. SADA shows you a snapshot of some of your records in each table. In this example, we only have one table (*ExampleCustomAnalysis*). Select this and press *OK*.



The next thing SADA needs is to know which column in your table contains the contaminant names and, if available, where the CASNUM column is found. Be very careful and make sure you've matched up the columns exactly like the following image:



Press *Next* > and SADA will present you with the contaminant matching form.

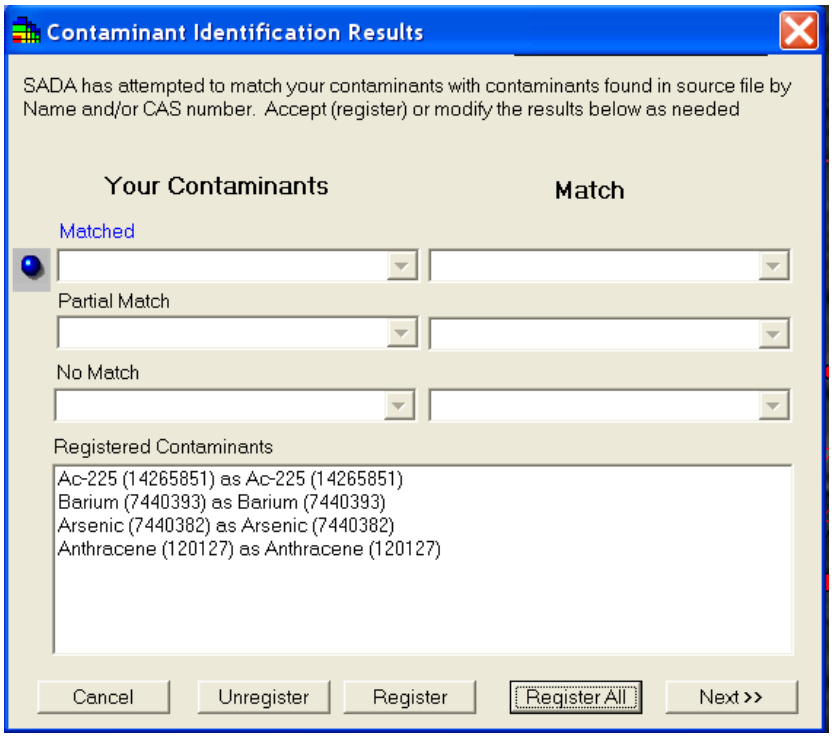


This form represents the contaminant matching efforts SADA has just done behind the scenes. In this form we have three categories:

- *Matched* – both the name and the CAS number match exactly between your data set(s) and the custom criteria you are trying to import. If your SADA file doesn't have any CAS numbers, that's OK. You just won't have any perfect matches.
- *Partial Match* – means either the names matched or the CAS numbers but not both. If your SADA file has no CAS numbers, then you may have some in this column where the names may still match.
- *No Match* – means that SADA could not find the contaminant name or CAS number anywhere in your custom analysis file.

In all cases, check what SADA has found and choose to *Register* or change the match up. To change the matchup, switch the selection on the right-hand side to the analyte you want to use and press *Register*. SADA will add the matched pair in the lower "Registered Contaminants" box. SADA will then automatically select the next contaminant on the left-hand side and show

you on the right-hand side what it is currently matched. You can continue previewing and registering contaminants in this manner until you have all the matches set.

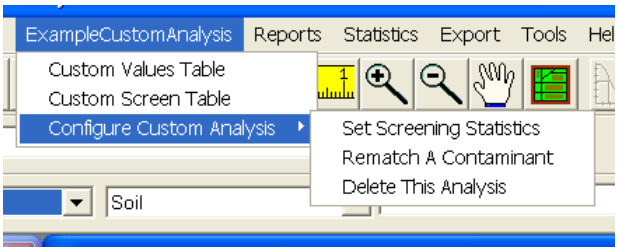


A potentially faster way is to highlight a particular row (e.g., click the left box on the *Matched* row) and select *Register All*. SADA will automatically dump all the auto-matches at once into the Registered contaminants Box. You can then more easily preview the matches and unregister any that are not correct.

Once all the contaminants have been registered/matched to analytes in your custom criteria database, press the *Next* button. SADA will reset itself. This is a good time to save your CustomCriteria.sda file. Select *File→Save* or press the *Save* button on the main toolbar.

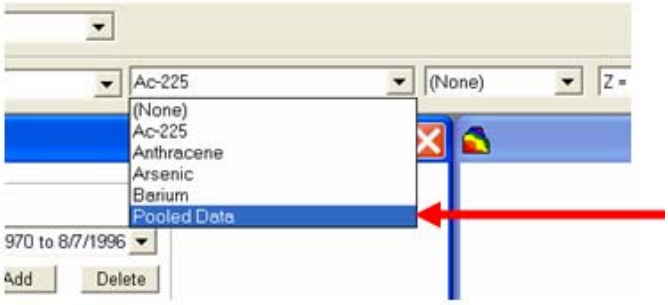
In order to see the effect of setting up custom criteria, you'll first need to select it from the analysis options. In the *General* analysis drop-down list, you will notice that a new analysis type with exactly the same name as your custom file name (or recordset) has been added, *ExampleCustomAnalysis*. Select this now.

SADA adjusts its interface accordingly. Some new interviews have been added, but perhaps the biggest change is the new menu item *ExampleCustomAnalysis*. Select this now and you can see features related to the display, use, and management of your custom criteria.



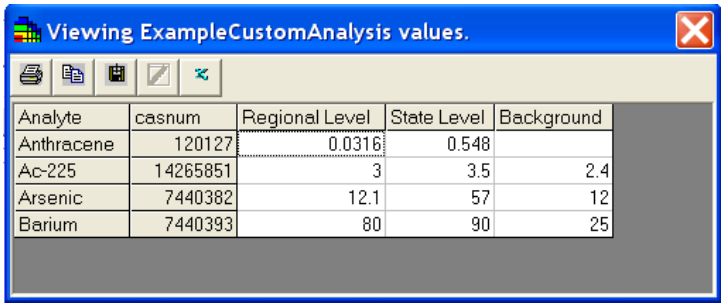
Select the option *Custom Values Table* to view your custom criteria. When the *Custom Values Table* appears, you can see and edit your custom values. Recall that any changes to the custom values inside of SADA will not affect the original source file or database.

Notice that you can only see custom values for *Ac-225*. This is because you currently have *Ac-225* selected. Close the *Custom Values Table* window and select *Pooled Data* from your data set drop-down list.



You'll notice that the sample points turn black and the interface may adjust slightly. This is normal. Each sample point represents a location where at least one analyte was measured. Since each location may have more than one analyte, SADA does not know which values to plot and so leaves the points blank. See Chapter 11, "Visualizing and Exploring Data," for more information on pooled data. For our purposes here, pooling the data allows the custom analysis table to show you all the custom criteria for soil at once.

Now select the *ExampleCustomAnalysis* menu and then the *Custom Values*. This time you should see all the custom criteria for each of the four soil contaminants. Pooled operations like these are done by data type (media type). If we had groundwater data in this file, you would have to see the pooled results for soil and groundwater separately.



Analyte	casnum	Regional Level	State Level	Background
Anthracene	120127	0.0316	0.548	
Ac-225	14265851	3	3.5	2.4
Arsenic	7440382	12.1	57	12
Barium	7440393	80	90	25

This window also provides some important features in the toolbar. Moving from left to right, we have

Print: prints the table out.

Copy: copies the contents of the table to the clipboard, where they can be copied to an outside software product, such as Word or Excel.

Save to disk: saves the table results to a comma delimited text file that can be opened elsewhere.

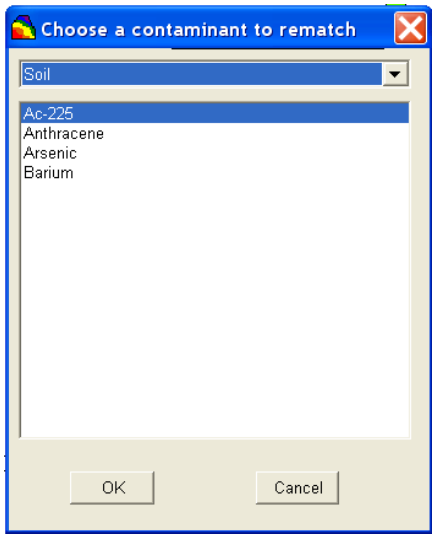
Autodocumentation: records the contents of the table to the currently opened autodocumentation report.

Export to Excel: possibly the most useful output format, where SADA opens Excel up and exports automatically to the first spreadsheet.

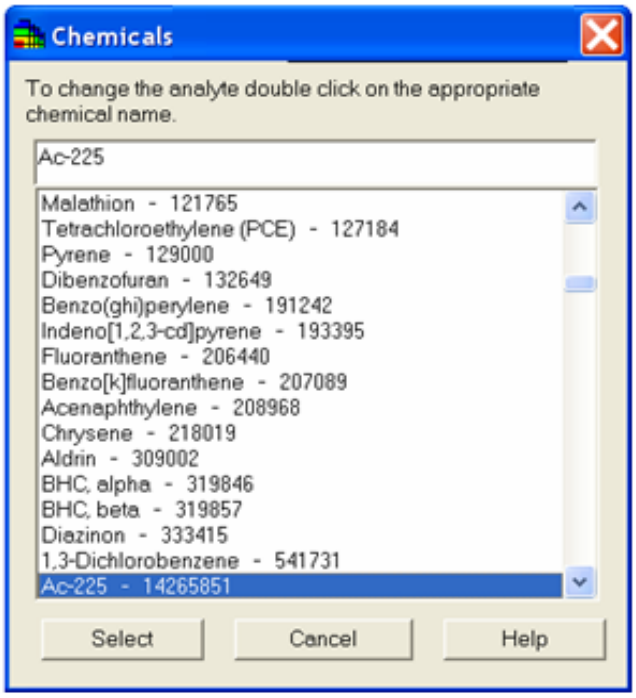
You can also perform custom screens against your imported criteria and configure your screening criteria. This is discussed in greater detail in the next chapter. Instead, we'll concentrate on two final issues related to management of your custom criteria.

First, select *Ac-225* as your current dataset. Then select from the main menu *ExampleCustomAnalysis→Configure Custom Analysis→Rematch a Contaminant*.

This feature allows you to re-register a single contaminant if needed. *Note: You may need to re-register if you incorrectly matched a contaminant the first time.*

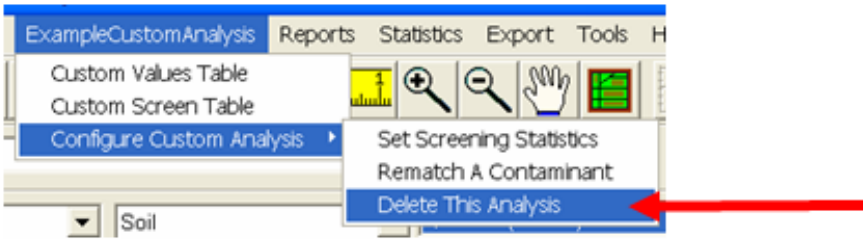


At the top of the window are your data types. In this case, only *Soil* is available. Select *Ac-225* from the list of soil analytes. SADA then looks at the full custom criteria imported during the setup and highlights the previously matched selection.



You can select a different analyte and choose *Select*. For now press *Cancel*.

If you want to re-access your database or file outside of SADA, you'll need to delete this analysis and set it back up again. From the main menu select *ExampleCustomAnalysis→Configure custom analysis→Delete this analysis*.



After prompting for confirmation, SADA will delete the analysis.

You are now ready to start using your custom criteria in your analysis. The next step is to read Chapter 15, "Screening Data Against Custom Criteria." The concepts in that chapter will be used repeatedly for custom criteria decision analysis further along.

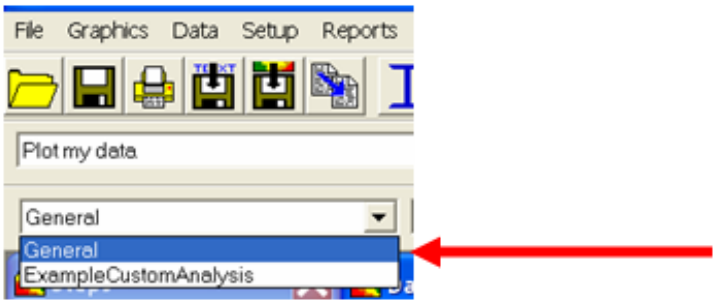
Chapter 15: Screening Data Against Custom Criteria (Custom Analysis)

In the last chapter, we saw how to create and then import custom screening criteria as a .csv (Excel). You can also create and import a Microsoft Access Database with your custom criteria. In this chapter, we will focus on how to screen your data against these custom criteria. This chapter assumes you know how to import custom criteria.

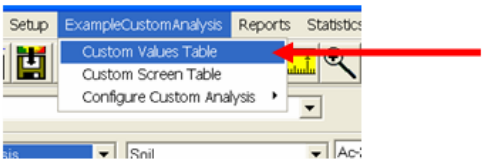
Producing a Tabular Screen

There are two ways to screen data against custom criteria: tabular and spatially. Let's begin by opening the file UsingCustomCriteria.sda. This is a two dimensional soil data set with the following analytes: Ac-225, Anthracene, Arsenic, and Barium. A set of custom criteria values was created (see Chapter 14) and imported into SADA. The SADA file containing these custom criteria was called CustomExample.sda; make sure this file is open and let's take a look at the custom criteria.

In the analysis drop-down list, select *ExampleCustomAnalysis*. SADA assumes the name of your file as the custom analysis type, so make sure to choose an appropriate name for your file.

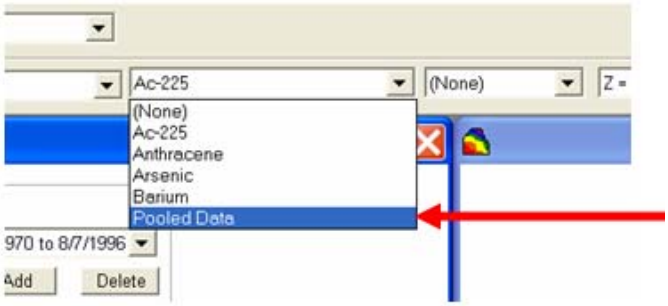


When you select *ExampleCustomAnalysis*, an extra menu item appears with the same name. Select this menu item and then select the submenu *Custom Values Table*.



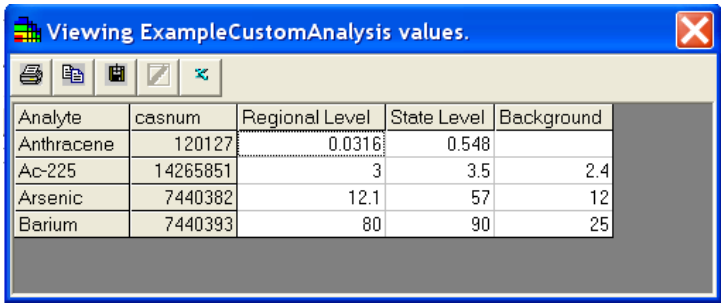
In the *Custom Values Table*, you can see and edit your custom values. Recall that any changes to the custom values inside of SADA will not affect the original source file or database.

Notice that you can only see custom values for Ac-225. This is because you currently have *Ac-225* selected. Close the *Custom Values Table* window and select *Pooled Data* from your data set drop-down list.



You'll notice that the sample points turn black and the interface may adjust slightly. This is normal. Each sample point represents a location where at least one analyte was measured. Since each location may have more than one analyte, SADA does not know which values to plot and so leaves the points blank. See Chapter 11, "Visualizing and Exploring Data," for more information on pooled data. For our purposes here, pooling the data allows the custom analysis table to show you all the custom criteria for soil at once.

Now select the *ExampleCustomAnalysis* menu and then the *Custom Values*. This time you should see all the custom criteria for each of the four soil contaminants. Pooled operations like these are done by data type (media type). If we had groundwater data in this file, you would have to see the pooled results for soil and groundwater separately.



Analyte	casnum	Regional Level	State Level	Background
Anthracene	120127	0.0316	0.548	
Ac-225	14265851	3	3.5	2.4
Arsenic	7440382	12.1	57	12
Barium	7440393	80	90	25

This window also provides some important features in the toolbar. Moving from left to right, we have:

Print: prints the table out.

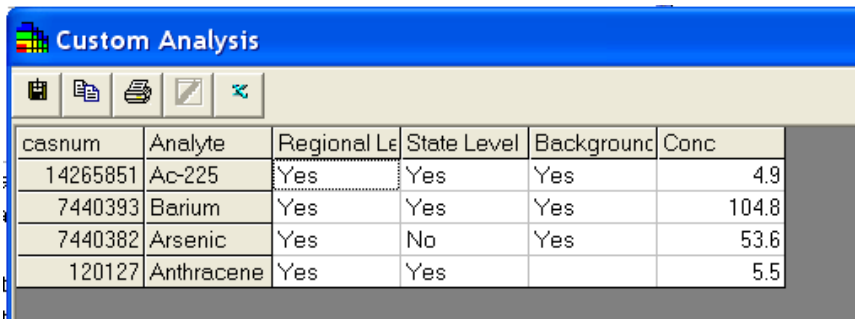
Copy: copies the contents of the table to the clipboard, where they can be copied to an outside software product, such as Eord or excel.

Save to disk: saves the table results to a comma delimited text file that can be opened elsewhere.

Autodocumentation: records the contents of the table to the currently opened autodocumentation report.

Export to Excel: possibly the most useful output format, where SADA opens Excel up and exports automatically to the first spreadsheet.

Close this window out. Now let's perform a tabular screen of your data against your custom criteria. Select, from the main menu once again, *ExampleCustomAnalysis* but then select *Custom Screen Table*. You will see a very similar window appear as follows:



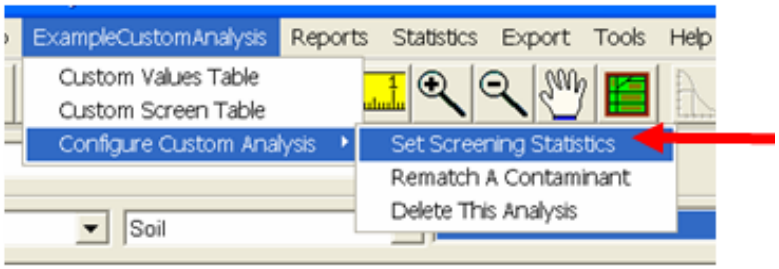
casnum	Analyte	Regional Le	State Level	Backgrounc	Conc
14265851	Ac-225	Yes	Yes	Yes	4.9
7440393	Barium	Yes	Yes	Yes	104.8
7440382	Arsenic	Yes	No	Yes	53.6
120127	Anthracene	Yes	Yes		5.5

Let's take a moment to discuss how to interpret this table. Each row contains an analyte. The first two columns show the Analyte name and CAS number (if available). The next three columns are the three custom criteria field names you assigned to your screening criteria: *Regional Level*, *State Level*, and *Background Level*. The fourth column shows the representative *Concentration (*we'll discuss this shortly).

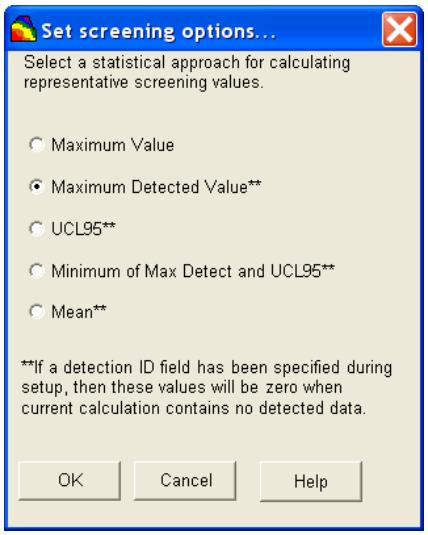
A *Yes* in any cell indicates that the screening criterion for that custom field for the analyte in the row was exceeded. For example, the data for Ac-225 exceed the criteria for the *Regional Level*. A value of *No* means that the criteria was not exceeded. A blank value means that no custom criteria were available in your file. In other words, there is no criterion for Anthracene for background.

When SADA screens your data against a criterion, it must first determine what value it will use. You likely have many data points and certainly have only one criteria value for any given field (e.g. Regional). SADA must select or calculate a representative value (in environmental lingo, it's called the representative concentration) to compare to the criteria. In the Ac-225/Regional cell above, SADA used a value of 4.9 for the representative value for Ac-225. In this case, 4.9 was the maximum detected value in the data set. You can control the value SADA selects.

Close the window and select from the main menu *ExampleCustomAnalysis* then *Configure Custom Analysis* and finally *Set Screening Statistics*.



A window appears with some common selections for determining the representative concentration.



There are five options:

- *Maximum Value*: The maximum concentration, detected or nondetected, for normal or lognormal distribution.
- *Maximum Detected Value*: The maximum detected concentration for normal or lognormal distribution.
- *UCL95*: The 95% upper confidence limit on the mean for normal or lognormal distribution.
- *Minimum of Max Detect and UCL95:* The minimum of the two calculations (see above).
- *Mean*: The average concentration over all values for normal or lognormal distribution.

Note that a common selection is *Maximum Detected Value* for environmental assessments. This is the default selection. Try selecting the *Mean* instead and then return to the screening table. You will notice that the concentration field has changed, as well as the screening results *Yes/No*).

The screening window has the same toolbar features as the custom viewing table. Using these toolbar buttons, you can export to Excel, print, save to file, and add to autodocumentation. Tabular screens are useful in highlighting contaminants of concern (to determine “the who”); however, they do not inform about the location or spatial distribution of any data points that may be exceeding the criteria. For this type of assessment, we will need to identify those contaminants of concern from the screening table and then individually apply spatial screens (to determine “the where”).

Spatial Data Screens

Tabular screens are very useful in quickly identifying which contaminants exceed a screening criterion. Spatial data screens quickly identify the spatial distribution of data points that are in excess. The approach is quite simple. SADA simply compares each individual point to the decision criteria; if it fails, the point is graphically enclosed in a box.

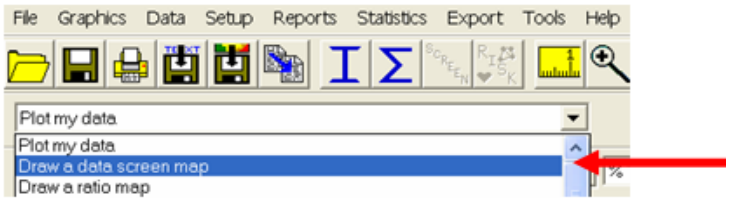
It is important to point out that there can be some disconnectivity between tabular and spatial screens. In a tabular screen, we are comparing the screening criteria to a singular representative concentration that you can control when configuring your screen. This

representative concentration has no meaning in a spatial data screen, where each point is individually compared to the decision criteria.

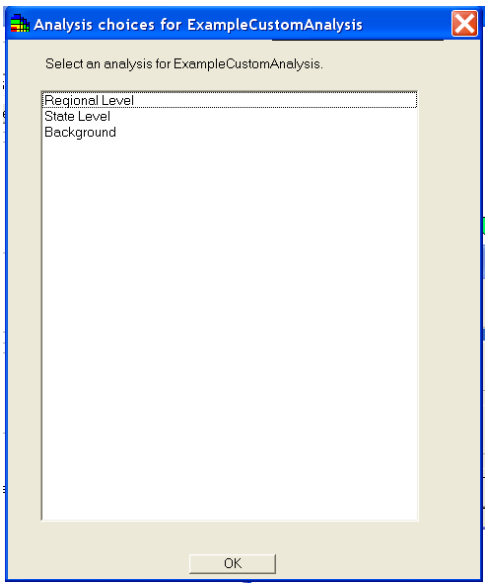
Suppose that the tabular screen's representative concentration is set to *Maximum Value*. In that case, one would expect that at least one data point in a spatial screen would be identified as in exceedance. Other selections, however, may not yield these results. For example, consider the UCL95. For small data sets, the UCL95 might be much greater than the largest measured value. In this case, you might see a *YES* in the tabular screen, but a spatial data screen shows no exceedances. This is actually correct. The converse may also be observed. Suppose the representative concentration is set to *Minimum of the Maximum Detect and the UCL95*. Furthermore, suppose that the *UCL95* is the smallest of the pair. It is possible that the *UCL95* does not exceed the criteria but the *Maximum Detected Value* does. In this case, at least one point in the spatial screen would be identified as in exceedance (at a minimum, the *Maximum Detected Value* itself).

In many cases, these differences are not problematic, as tabular screens are typically used simply to identify the contaminants of concern. Spatial data screens are then applied to these contaminants of concern, and any discrepancies are simply noted.

Let's try a spatial data screen using custom criteria. In particular, let's examine the Ac-225 data. We know from the tabular screen that this is a contaminant of concern when compared to the regional screening level. Make sure you still have *ExampleCustomAnalysis* selected in the analysis type box. Then, in the interview box, select *Draw a data screen map*.



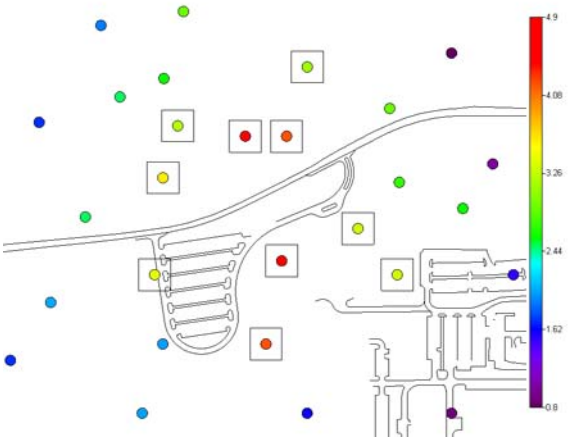
Press the button *Show Me The Results* at the bottom of the steps window. You will be presented with custom choices to screen against.



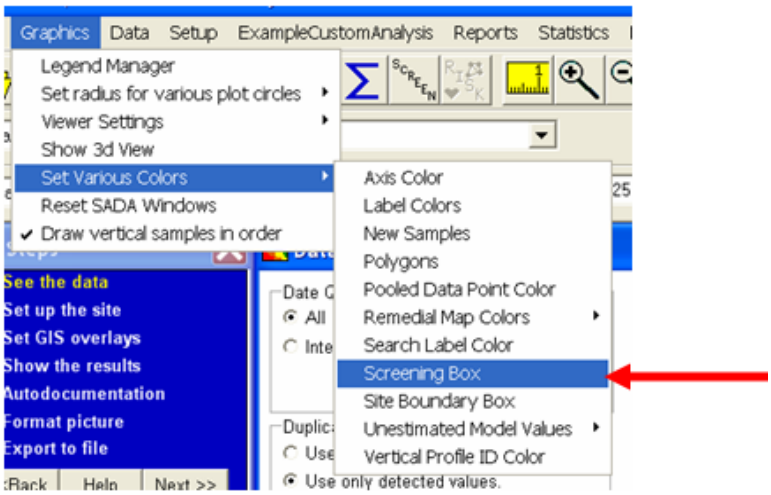
Select *Regional Level* and press *OK*. SADA reports that your custom value for *Regional Level* is 3.

Press *OK*. The next window reports that 10 individual points exceed this criterion.

The graphics viewer is then updated. Wherever a data point exceeds the criteria, a box is placed around the sample point.



The default color for these boxes is black; however, it may be difficult to see the boxes against dark backgrounds or photographs. To change the color of the exceedance boxes, choose from the main menu *Graphics menu → Set Various Colors→Screening Box*.



You can select a different color for the boxes and press *OK*.

Currently, the Data screen map cannot be shown in the 3d viewer. To see a Data screen map for 3d data in a true 3d environment, you can manipulate the legend manager to essentially perform the same screen using colors for a particular decision value. See Chapter 13, “How to Perform a Quick Spatial Data Screen,” for an explanation on how to do this.

Let’s place a copy of this Data screen map into a PowerPoint presentation. From the main toolbar select the copy button.

Open Microsoft PowerPoint and create a new blank presentation. In PowerPoint, select *Edit →Paste.*

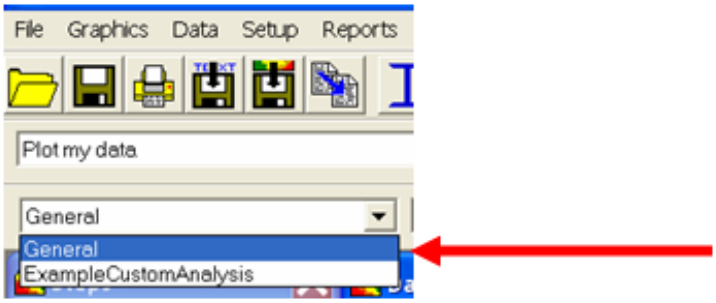
Chapter 16: Creating Ratio/Sum of Fraction Maps Using Custom Criteria (Custom Analysis)

In the last chapter, we saw how to screen raw data values against custom screening values imported with either Excel (.csv) or Access (.mdb). This chapter assumes you know how to import custom criteria.

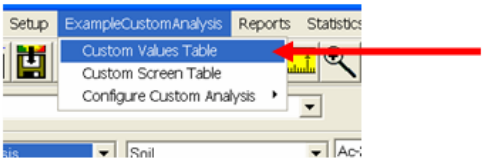
Producing a Ratio Map

Let's begin by opening the file UsingCustomCriteria.sda. This is a two dimensional soil data set with the following analytes: Ac-225, Anthracene, Arsenic, and Barium. A set of custom criteria values was created and imported into SADA (see Chapter 14, "Importing Custom Screening and Remediation Criteria"). The file containing these custom criteria was called *ExampleCustomAnalysis*. Let's take a look at those custom criteria first.

In the analysis drop down box, select *ExampleCustomAnalysis*. SADA assumes the name of your file as the custom analysis type, so make sure to choose an appropriate name for your file.

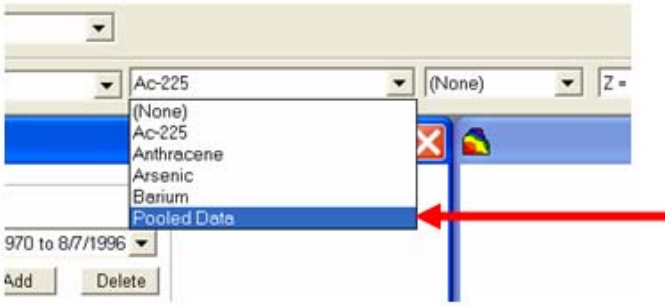


When you select *ExampleCustomAnalysis*, an extra menu item appears with the same name. Select this menu item and then select the submenu *Custom Values Table*.



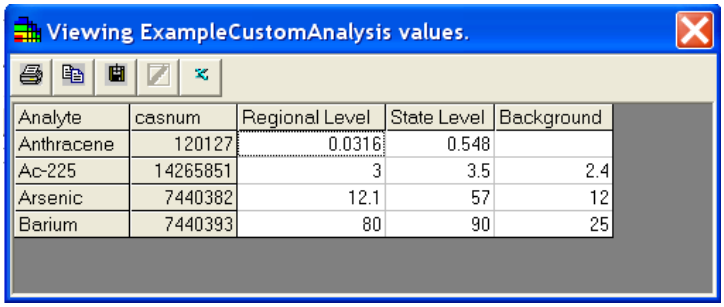
In the *Custom Values Table*, you can see and edit your custom values. Recall that any changes to the custom values inside of SADA will not affect the original source file or database.

Notice that you can only see custom values for *Ac-225*. This is because you currently have *Ac-225* selected. Close the *Custom Values Table* window and select *Pooled Data* from your data set drop-down list.



You'll notice that the sample points turn black and the interface may adjust slightly. This is normal. Each sample point represents a location where at least one analyte was measured. Since each location may have more than one analyte, SADA does not know which values to plot and so leaves the points blank. See Chapter 11, Visualizing and Exploring Your Data, for more information on pooled data. For our purposes here, pooling the data allows the custom analysis table to show you all the custom criteria for soil at once.

Now select the *ExampleCustomAnalysis* menu and then the *Custom Values*. This time you should see all the custom criteria for each of the four soil contaminants. Pooled operations like these are done by data type (media type). If we had groundwater data in this file, you would have to select the pooled results for soil and groundwater separately.



Analyte	casnum	Regional Level	State Level	Background
Anthracene	120127	0.0316	0.548	
Ac-225	14265851	3	3.5	2.4
Arsenic	7440382	12.1	57	12
Barium	7440393	80	90	25

This window also provides some important features in the toolbar. Moving from left to right, we have

Print: prints the table out.

Copy: copies the contents of the table to the clipboard, where they can be copied to an outside software product, such as Word or Excel.

Save to disk: saves the table results to a comma delimited text file that can be opened elsewhere.

Autodocumentation: records the contents of the table to the currently opened autodocumentation report.

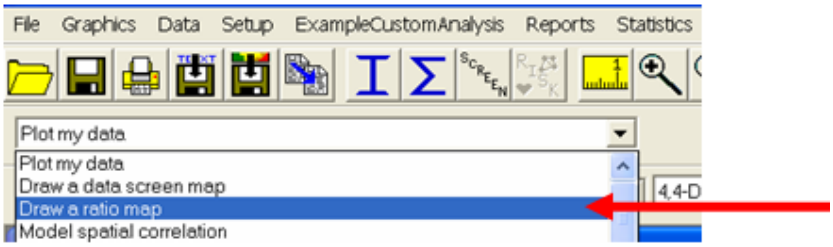
Export to Excel: possibly the most useful output format, where SADA opens Excel and exports automatically to the first spreadsheet.

Close this window. Now let's create a ratio map using your custom criteria.

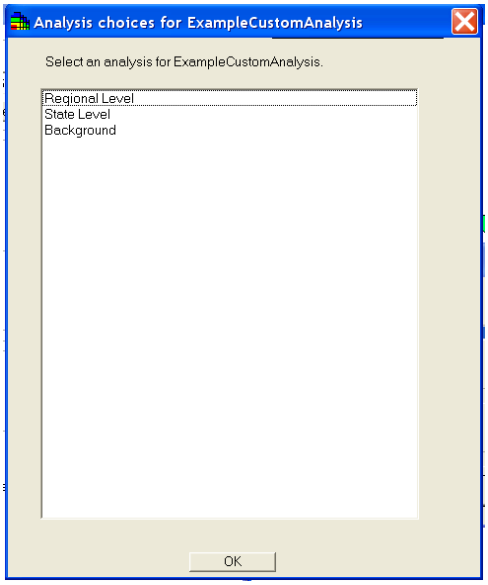
Spatial data screens quickly identify the spatial distribution of data points that are in excess. The approach is quite simple, as points are marked as either exceeding or not. Ratio maps extend this by showing the severity of the exceedance. SADA simply divides each individual point by the decision criteria and plots the resulting ratio value. If the ratio exceeds a value of one, this indicates that the concentration must be larger than the custom criteria. A ratio of 25 indicates

that the concentration is 25 times higher than the decision criteria. A ratio of 1.05 indicates that the concentration is only 5% higher than the decision criteria and a ratio of 0.50 indicates that the concentration is only half the decision criteria. A simple screening against the data values would not reveal the relationship between point values and custom criteria.

Let's try creating a ratio map now. In particular, let's examine the Ac-225 data. We know from the tabular screen that this is a contaminant of concern when compared to the regional screening level. Make sure you still have *ExampleCustomAnalysis* selected in the analysis type box. Then in the interview box, select *Draw a ratio map*.

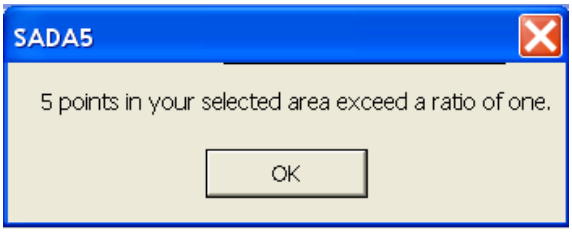


Press the button *Show Me The Results* at the bottom of the steps window. You will be presented with custom choices for screening against.



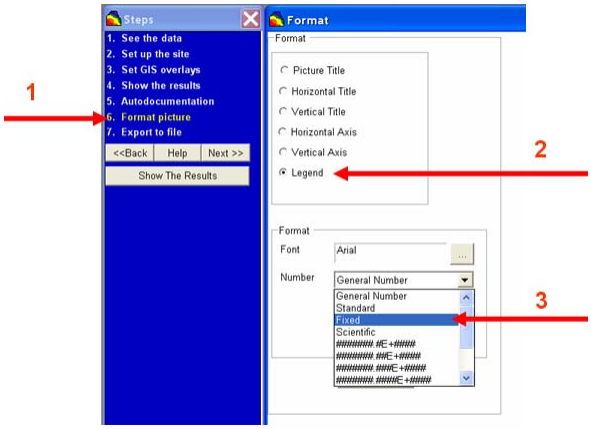
Select *State Level* and press *OK*. SADA reports that your custom value for State Level is 3.5.

Press *OK*. The next window reports that 5 individual points exceed this criterion (ratio > 1).



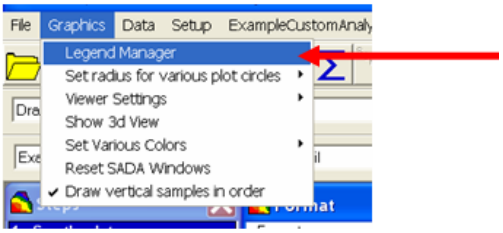
Press *OK* and the graphics viewer is then updated. The legend now refers to ratio values instead of concentration values.

Notice that the number formatting on this legend has a lot of digits to the right of the decimal. Let's clean this up. On the step *Format picture*. In the parameter window, select *Legend* and in the number format select *Fixed*. This will limit the number of decimals to 2.

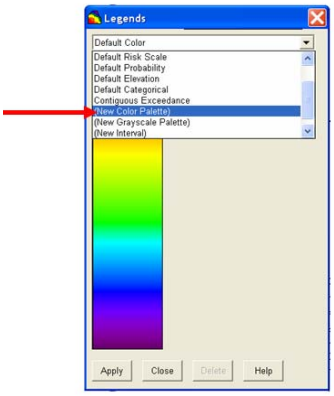


Press *Apply* and notice that your legend has changed accordingly. This change will remain in place for this legend for all ratio maps from now on. If you change your legend or create a new one, you will need to re-specify the *Number* format again.

As you look at the map, it may be somewhat difficult to determine the locations that are more severe than others. You can create a new legend that emphasizes those with the greatest exceedances. From the main menu select *Graphics→Legend Manager.*



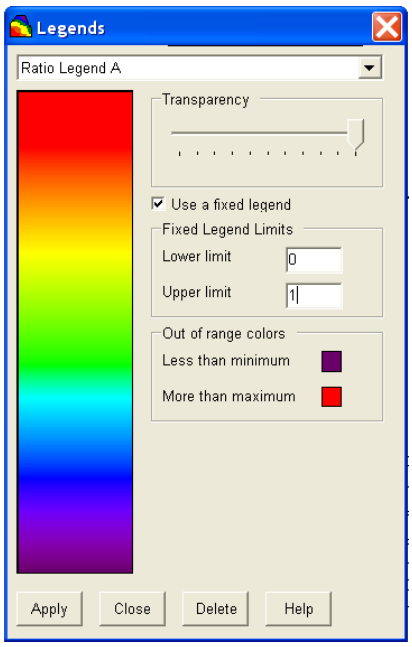
This brings up the *Legend Manager* window. Select from the list of legend options *(New Color Palette)*.



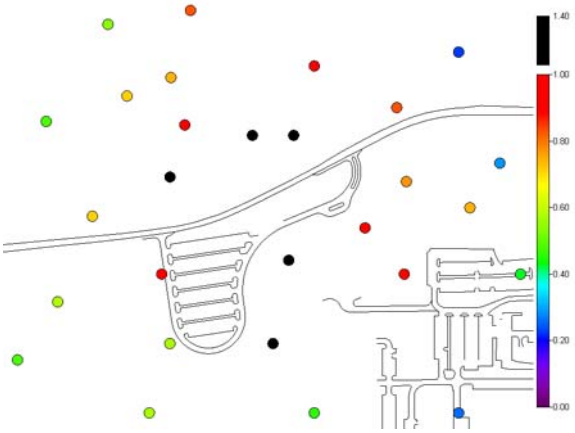
You will be asked for some information on your new legend. Give the new legend the name "Ratio Legend A" and in the *Based on* drop-down list choose the *Default Color* legend.

When you base a legend on an existing legend, SADA makes a copy of the base legend as a starting point. You can then customize the legend as you wish. Note that you cannot customize default legends. You will first need to create your own legend, just as we now did. Press the *OK* button.

SADA selects your newly created legend in the options box. Notice that there are now methods for customizing the legend. For ratio maps, it might be a good idea to highlight exceedance with a unique color. To do this, check the option *Use a fixed legend*. Below this checkbox, enter a value of zero for the lower limit and a value of 1 for the upper limit. The color band will now stretch between these two values. Below these upper and lower limit boxes are two *Out of range colors*. These colors will be used to highlight values that are not between 0 and 1 (or whatever values you choose for the lower and upper limits). We will change this next, but first, make sure your legend manager window looks like the following image.



Mouse-click on *More than maximum*, and change the color to black. Press *OK* on the color palette window. Press the *Apply* button and notice how clearly those that exceed a given criterion now stand out.

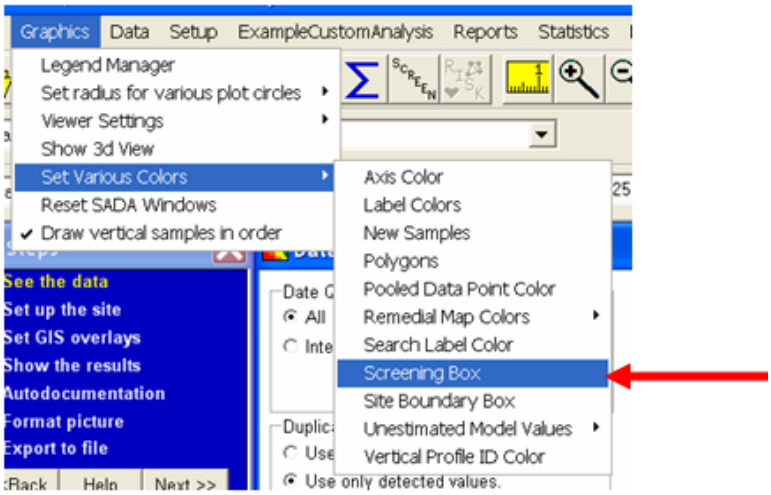


For fixed legends, SADA creates a break between the color band on the color used for all values above or below. If no values in your data set exist outside the range of your color band, then the legend segment will not be shown. For example, in the picture above, no ratio values were less than zero; therefore, no legend fragment was created below the color band.

This picture still does not necessarily show the severity with which points exceed the criteria very well. Let's return to the legend manager and create a new type of legend that will. First though, save your SADA file as UsingCustomCriteria_modified.sda.

From the main menu, select *Graphics→Legend Manager* again. In the drop-down list select *(New Interval).* In the new legend window, name your new legend "Ratio Legend B" and base it on *Default Interval*. If your new legend name window looks like the following image, press *OK* to close the window.

The default color for these boxes is black; however, it may be difficult to see the boxes against dark backgrounds or photographs. To change the color of the exceedance boxes, choose from the main menu *Graphics menu → Set Various Colors →Screening Box.*



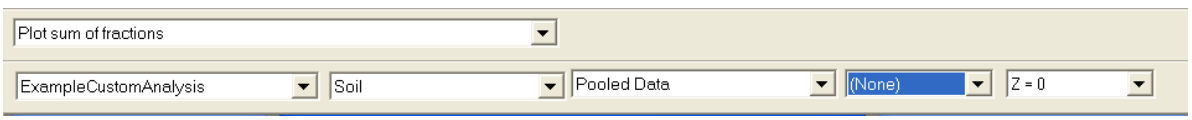
You can select a different color for the boxes and press *OK*.

Currently, the Data screen map cannot be shown in the 3d viewer. To see a Data screen map for 3d data in a true 3d environment, you can manipulate the legend manager to essentially perform the same screen using colors for a particular decision value.

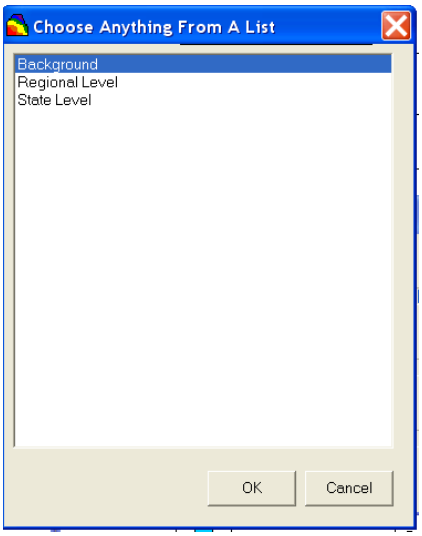
Creating a Sum of Fractions Point Map

In a sum of fractions map, a ratio map for each contaminant in the pooled data list is first created behind the scenes. These maps are then summed together to create a sum of fractions map. This type of map indicates how individual contaminants may be acting collectively with respect to a sum ratio of one. So, if a point on the map exhibits a sum of fractions value of 2, then the contaminants may be causing concern when considered together. This type of map is more typical in radiological assessments.

If you have already closed the file UsingCustomCriteria.sda, reopen it now. Select *ExampleCustomAnalysis* from the analysis drop-down list, *Soil* from the data type box, and *Pooled Data* in the data set box. In the interview box, select *Plot sum of fractions*. Make your SADA selections look like the following image.



Now in the steps window, press the *Show The Results* button and you are presented with custom criteria options.



Select *Background* and press *OK*. SADA responds that not all of your contaminants have a background value. You can either continue with the analysis or not. If you continue, SADA will create a sum of fractions map, excluding those contaminants with no background criteria. Select *Yes*. SADA reports that a total of 30 samples exceed a sum of fractions equal to one and highlights them with exceedance boxes.

Now you try. Repeat the above steps for *Regional Level* and *State Level*.

Re-enforcement

Let's place a copy of this data screen map into a PowerPoint presentation. From the main toolbar select the copy button.

Open Microsoft PowerPoint and create a new blank presentation. In PowerPoint, select *Edit →Paste.*

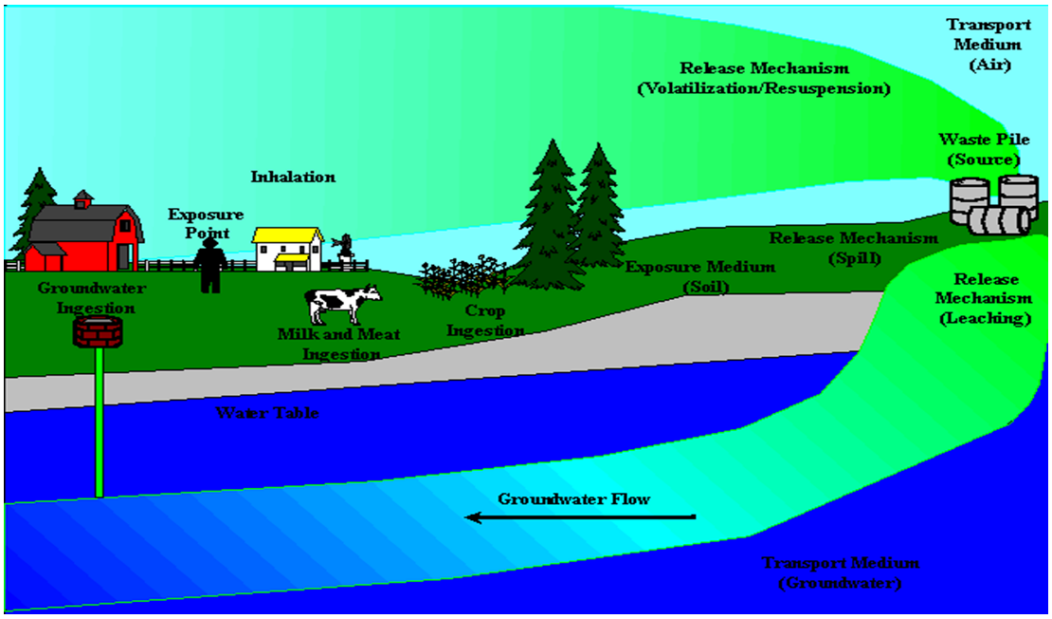
Chapter 17: Overview of Human Health Risk Assessment

Human health risk assessment refers to the estimation of carcinogenic risk and non-carcinogenic hazard caused by exposure to contaminated media. SADA's human health risk implementation is based on the EPA's Risk Assessment Guidance for Superfund. SADA provides a risk assessment module to calculate the risk of adverse health impacts on a population exposed to toxic chemicals found in groundwater, surface water, soil, and sediment. It also calculates risk-based screening values to quickly identify contaminants of concern. These results are integrated into the other modules, such as cost benefit analysis, geospatial analysis, and sampling schemes.

The risk of exposure to contaminants depends on three broad and interconnected factors: 1) the type of contaminant, 2) the severity of the contamination, and 3) the method of exposure. The type of contaminant is a critical component. Each contaminant has a unique set of toxicity values that vary with the method of exposure, with health effects depending on whether the contaminant is classified as a radionuclide or a carcinogenic or non-carcinogenic inorganic or organic chemical.

The level of exposure is also important. For exposed individuals, a positive relationship between level of exposure and health impact is estimated for carcinogenic contaminants. For non-carcinogenic contaminants, levels above a dose threshold are used to indicate the presence of a health hazard. Toxicity databases are packaged with SADA to assist in this calculation and determination. The presence of the contaminant in environmental media is usually summarized as a single value, the representative exposure concentration, to assist with this calculation. This concentration is generally a summary statistical, such as the mean value or some upper confidence limit (e.g., 95^{th}) on the mean. This value is then used in the exposure model to determine the potential risk for cancer or other negative health outcomes.

The route of exposure is also a key determining factor. Contaminants can have greater health consequences for some exposure routes than others. Typical exposure routes include ingestion, inhalation, and dermal contact. Influencing exposure is the type of land-use activity that will occur on the site. With all the input variables for exposure, toxicity, and routes of exposure, models can then be implemented to estimate potential or actual exposures along each exposure route for current or possible future land uses.



Current contaminant concentrations are often used for the on-site assessment of future exposure; however, modeled results that represent future contaminant concentrations can also be imported. The five land use scenarios considered in SADA include industrial, residential, recreational, excavation, and agricultural.

For residential land use, residents are expected to be in frequent, repeated contact with contaminated media. Residential exposures account for daily contact over a lifetime, including exposures for the receptor as child and adult, and typically produce the highest potential exposures and risk. Under the industrial land use scenario, workers are expected to be exposed routinely to contaminated media within a commercial area or industrial facility at the site. Estimate exposures are based on potential use of heavy equipment and related traffic in and around the contaminated soil and sediment. The recreational land use scenario addresses exposure to children and adults who spend limited time at or near the site while engaging in outdoor activities. For the excavation scenario, worker exposures to soil and sediment for a short period are considered appropriate. The exposure routes for soil and sediment for the excavation worker are similar to an industrial worker. The agricultural scenario assumes a resident is exposed to homegrown farm products. Exposure routes, in addition to the residential pathways, include consumption of site-grown vegetables, milk, and beef. In the software, users select exposure pathways used to calculate total risk, and separate calculations are conducted for surface soil, sediment, groundwater, and surface water. SADA's risk models follow the EPA guidance (1989 et seq.), and model input parameters can be modified to fit site-specific exposure conditions.

These three factors can be brought together to produce two important calculations: 1) the preliminary remediation goal (PRG) and 2) the calculation of carcinogenic risk and/or health index. A PRG sometimes is referred to as a Risk Based Goal (RBG). In this calculation, an acceptable concentration limit is calculated that would be protective of an exposed group under a specific land use/exposure scenario for a given risk limit. PRGs are typically used to screen site data to determine if any exceedances have occurred. If no exceedances have occurred, then the contaminant can sometimes be set aside and no longer considered. A lengthy list of detected contaminants at a site often can be reduced by such risk screening. SADA contains additional screens and statistical tests that consider background, sample detection frequency, bioavailability, and whether detected contaminants are essential nutrients that can be

conducted, including specific functions such as univariate statistics and non-parametric comparison tests (e.g., Wilcoxon Rank Sum test for comparing data to background).

The calculation of risk (and hazard) runs the same models forward for risk assessment purposes. Instead of specifying the risk level, the representative exposure concentration is summarized and combined with exposure parameters and model assumptions. The model then produces the risk level associated with the concentration for given land use/exposure scenario. Additional decision-support output includes comprehensive tables of forward calculations of exposure and risk that summarize over-all included contaminants at the site, used to support a full baseline risk assessment for the contaminants of concern.

If the site poses acceptable risk, no further evaluation and no remediation from a human health perspective may be warranted. If the site poses unacceptable human health risk, additional evaluation in the form of remedy development and evaluation would be appropriate. SADA produces tables of output that can be modified to support risk assessment documentation purposes. For identified contaminants of concern, it also can provide spatial data screens to visualize where exceedances are found, and risk can be mapped using the available interpolation functions. These latter features can be used to drive remedial design and secondary sampling strategies.

The following chapters will demonstrate how to setup the risk model in your SADA file, perform PRG screens, forward risk calculations, and how these are integrated with the spatial tools in SADA.

Chapter 18: Preparing for Human Health Risk

This chapter deals with preparing your data for import into SADA with the objective of conducting a human health risk analysis. Chapter 5 covered how to import data into SADA, and that same procedure will be used here; however, we will focus a little more on those fields required for risk and on some strategies you might use, depending on your situation. SADA can import various types of data, but point data is the only kind SADA can directly use in assessing risk.

What if I don't have any coordinates? What if I want to use exposure values calculated elsewhere? What if I have contoured data?

It is possible to calculate risk on modeled or gridded data types and even on data sets without spatial coordinates. In some cases, users may want to use exposure concentrations other than the four basic options in SADA: *Maximum Detect, Mean, UCL95*, and *Minimum of Max Detect and UCL95.* These special cases are covered a little later. For right now, we will use measured point values for a set of contaminants in the soil.

Working with Sampled Data with Coordinates

Recall from Chapter 5 that before SADA can import the data values, it needs the data in a particular format. Fortunately, the format is very easy to work with, and much of the environmental data comes in this type of configuration. Simply put, data needs to be in a tabular format with each row representing a single measurement value of a single contaminant (or attribute). Each column contains information about the measurement, such as the easting coordinate, the northing coordinate, depth, name of the contaminant, measured value at that location, and so forth. Many times, users look at this data with a spreadsheet program like Excel. Here is an example data set in Excel.

	A	B	C	D	E	F	G	H	I	J	K	L
1	Easting	Northing	Depth	Casnumber	Name	Value	Detected	Media	Date	SampleID	Region	
2	27596.25	21900	0	14265851	Ac-225	2	1	SO	9/26/1970	1	1	
3	28310.25	21900	0	14265851	Ac-225	1.6	1	SO	9/26/1970	2	1	
4	28935	21900	0	14265851	Ac-225	0.9	1	SO	9/26/1970	3	1	
5	27685.5	22200	0	14265851	Ac-225	2	1	SO	9/26/1970	4	1	
6	28131.75	22200	0	14265851	Ac-225	4.2	1	SO	9/26/1970	5	1	
7	29202.75	22500	0	14265851	Ac-225	1.5	1	SO	9/26/1970	6	1	
8	27150	23160	0	14265851	Ac-225	1.7	1	SO	9/26/1970	7	1	
9	27685.5	22920	0	14265851	Ac-225	3.6	1	SO	9/26/1970	8	1	
10	28042.5	23100	0	14265851	Ac-225	4.9	1	SO	9/26/1970	9	1	
11	28221	23100	0	14265851	Ac-225	4.2	1	SO	9/26/1970	10	1	
12	28667.25	23220	0	14265851	Ac-225	2.9	1	SO	9/26/1970	11	1	
13	29113.5	22980	0	14265851	Ac-225	1	1	SO	9/26/1970	12	1	
14	27417.75	23580	0	14265851	Ac-225	1.9	1	SO	9/26/1970	13	1	
15	27774.75	23640	0	14265851	Ac-225	2.9	1	SO	9/26/1970	14	1	
16	28310.25	23400	0	14265851	Ac-225	3.1	1	SO	3/22/1993	15	1	
17	28935	23460	0	14265851	Ac-225	0.8	1	SO	3/22/1993	16	1	
18	28200	22560	0	14265851	Ac-225	4.8	1	SO	3/22/1993	17	1	
19	28700	22500	0	14265851	Ac-225	3.3	1	SO	3/22/1993	18	1	
20	27200	22380	0	14265851	Ac-225	2.03	1	SO	3/22/1993	19	1	
21	28984	22787	0	14265851	Ac-225	2.6	1	SO	3/22/1993	4	1	
22	27350	22750	0	14265851	Ac-225	2.5	1	SO	3/22/1993	3	1	
23	27026	22129	0	14265851	Ac-225	1.7	1	SO	3/22/1993	2	1	
24	27690	23350	0	14265851	Ac-225	2.6	1	SO	3/22/1993	4	2	
25	27500	23270	0	14265851	Ac-225	2.5	1	SO	3/22/1993	44f	2	
26	28709	22900	0	14265851	Ac-225	2.7	1	SO	3/22/1993	43	2	
27	27650	22500	0	14265851	Ac-225	3.4	1	SO	3/22/1993	33	2	
28	28530	22700	0	14265851	Ac-225	3.3	1	SO	3/22/1993	11	2	
29	27750	23145	0	14265851	Ac-225	3.2	1	SO	3/22/1993	12	2	
30	27596.25	21900	0	7440393	Barium	42.7	1	SO	8/7/1996	13	2	
31	28310.25	21900	0	7440393	Barium	35.1	1	SO	8/7/1996	14	2	
32	28935	21900	0	7440393	Barium	18.4	1	SO	8/7/1996	15	2	
33	27685.5	22200	0	7440393	Barium	43.5	1	SO	8/7/1996	16	2	
34	28131.75	22200	0	7440393	Barium	87.9	1	SO	8/7/1996	17	2	

SADA can import measured values using two different kinds of files. A Comma Separated Value (.csv) file is the most commonly used file type. Anyone with a copy of Excel or a similar spreadsheet program can easily save their file as a .csv file and SADA can read it. Even if you don't have a spreadsheet program, you can use simple text editors (e.g., Notepad that comes with Windows) to create or format your data set. If you are using a .csv file, the first row must be the title row. Punctuation is not allowed in the column (field) names. The second way to import measured values is to use a Microsoft Access database. Let's examine each of the types of information now.

Easting and Northing Columns (required)

SADA must have these two columns. They tell where on the site your sample was collected. SADA does not require or enforce the use of any predefined coordinate system. There are a number of them available, such as UTMs, state plane, lat/long, and so forth. Many sites have their own localized system. In some cases, you'll see sample measurements relative to the corner of the site or a building. These are all OK. (Note to lat/long users: you'll need to make sure you use the negative of your latitude values or your data will plot backwards.)

It is important to note that these values must be in the same coordinate system as any GIS layer you plan on using. For example, the roads.dxf layer seen throughout these examples is in State Plane Coordinates for Tennessee. You would need to make sure that the easting and northing coordinates are using the same system or there will be big problems. Namely, your data will not

plot correctly on your GIS layer. Now, if you are not using GIS, then you have no worries with your coordinate system.

Blank values and non-numeric values are not permitted in either of these columns.

Depth Column (may or may not be required)

If the data are taken over depth, an additional column with the depth value is required. No empty values are allowed for any of these columns, and non-numerical values are not permitted for any coordinate or sample value. There are two important things to check about your depth column. First, make sure your depth is *depth below surface* and NOT elevation measurements. SADA operates on the concept of depth below surface. Elevation is used in SADA, but only as a 3d view feature. Second, make sure that your depth values are in the same units as your horizontal extents. For example, it is nonsensical to use lat/long for horizontal and meters for depth. Later on, when you do 3d geospatial analysis, this really has bizarre implications. If the horizontal measurements are in feet, make sure the depth is in feet and so forth. SADA is not going to do any conversion for you.

Sampled Values (required)

This field is where your measurement value is found. There are no units enforced by SADA because of the many and varied types of applications for SADA. You will need to make sure that if you are doing human health or ecological risk assessment in SADA you MUST have your measured units as follows.

- Soil, Sediment, and Biota: mg/kg for nonradionuclides, pCi/g for radionuclides
- Surface/Groundwater: mg/L for nonradionuclides, pCi/L for radionuclides

The sample values column cannot contain any blanks or non-numeric values. If a measurement value was a nondetect, you must enter the detection limit into the value column. For example, you cannot enter <5. You would instead enter a value of 5 and then not enter this in the detection field (see below). Values like NA are obviously not allowed either.

Name (required)

This column provides the name of the attribute (contaminant) you are measuring in each record. The field cannot contain any blank values. When you are working with environmental contaminants, it is common to work with contaminant names that contain commas. These are a serious problem for comma delimited files, as you might imagine. For example, if you have the contaminant name 1,2,3-Trichlorobenzene, it would be recognized as 3 separate columns: 1, 2, and finally 3-Trichlorobenzene. This will cause an error in import. Typically, this error states that SADA has reached the end of the file unexpectedly. To correct this, make sure your name field contains the name within quotation marks: “1,2,3-Trichlorobenzene.” Spreadsheet programs like Excel normally do a good job of taking care of this for you by placing quotations around the contaminant name. To be sure, however, you may want to open up your .csv file in a simple text editor, such as Notepad (shipped with Windows), to make sure that you are not running into this problem. Typically, SADA will complain about reaching the end of the file too soon in this situation.

It is also important to make sure that each contaminant is consistently named. For example, make sure you don’t have 1,2,3-Trichlorobenzene also spelled as Trichlorobenzene-1,2,3. SADA may see these as two different contaminants.

CAS Numbers (not required)

A column containing the CAS numbers can be useful when you are setting up different kinds of analyses, such as human health. It isn't required, though. If you do have CAS numbers and want to use them, make sure that they do not contain any dashes or spaces. This will render them virtually useless later as SADA tries to match them up to its various data bases. A simple search and replace in the CAS column using Excel can take care of this pretty quick.

Detection Qualifiers (not required)

Another field of some importance is the detection field. Instead of the normal Us and Js and so forth, SADA works with either a zero value (sample was below detection limit) or a 1 (sample was detected). As the user, you must decide whether a value should be considered detected or not before bringing it into SADA. This involves an area of the environmental process we are not involved in; however, we do provide some guidance for risk assessment in the tables below if you have no other instructions.

Flag	Meaning	Use for risk?
R	Rejected	No
B	Blanks contaminated	Treat as non-detect
J	Estimated	Yes, treat as detect
UJ	Estimated non-detect	Yes, treat as non-detect
K	Biased high	Yes, treat as detect
L	Biased low	Yes, treat as detect
U	Non-detect	Yes, treat as non-detect

Now for measured values that are considered non-detects, you should enter the detection limit of the instrument into the column containing measured values. Blank or missing values in the value column are not permitted.

Date Column (not required)

The date field is not required; but if you do use it, later you can query your data by date to see how things are evolving over time. This can be useful in monitoring situations. If you do use the date field you must use some combination of the m, mm, d, dd, yy, yyyy format. Examples include m/d/yy and mm/dd/yyyy. Dates that include any text (other than /) are not permitted. For example, September 26th, 2004 is not permitted. You would need to use 9/26/2004. In Europe and other areas that use day/month/year format, this should be ok as long as your data set and your operating system are consistent.

Media (required)

For human health, you MUST have a media column. Proper media identification qualifiers are as follows: Soil – SO, Sediment – SD, Groundwater – GW, Surface water – SW, Air – AIR, Biota – BIO, and Background – Background.

Beyond these columns, you can include any other information you like, but the total number of columns may not exceed 250.

Final Tips

Sometimes, the first attempt to import data fails. Chapter 5 covers some of the most common reasons this may occur. Please refer to this chapter now and make sure these conditions do not exist in your dataset before importing the data.

Re-enforcement

Create a SADA file to use for human health analysis. (Refer to Chapter 5 if you need help.) Do the following

1) Create a new SADA file.
2) Import the data file TwoDimensionalWithNonDetects_Duplicates.csv.
3) Match
    a. Name to Name
    b. Value to Value
    c. Easting to Easting
    d. Northing to Northing
    e. Depth to Depth
    f. CAS Num to Casnumber
    g. Detect Qualifier to Detected
    h. Media Id to Media
    i. Date to Date
4) Note: This file has both duplicates and non-detect values.
5) Snap the site boundaries to *All data sets*.
6) Import the GIS layers roads.dxf and water.dxf. Make water layer blue (Chapter 4).
7) Save as HumanHealthRisk.sda.

Preparing for Human Health Risk

If you have not already done so, perform the re-enforcement exercise above and bring up HumanHealthRisk.sda. We now have our data set and need to prepare it for conducting a

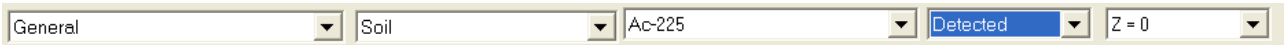
human health risk. The first thing we need to do is decide how to deal with duplicates and non-detects.

Dealing with Non-detects

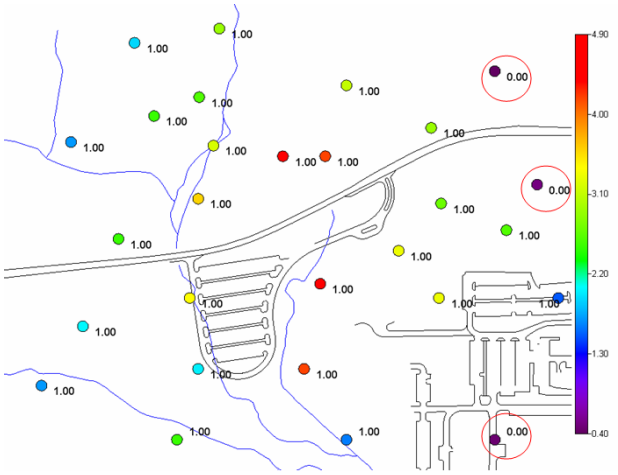
Recall that if a detection field is specified during import, SADA will assume a value of 0 if the sample is a nondetect and a value of 1 if the sample is detected. If you did not specify a detection field, SADA assumes all data are detected. For non-detects, you have three options: use full detection limit, use half the detection limit, or use zero. If a sample is designated as a detect (1), then the result in the value field is considered the actual measured value as you would expect. If a sample is designated as a non-detect (0), the result in the value field is considered the detection limit.

On virtually any interview for sampled data, you will find the step *See the data* in the steps window. For your current file, select *Plot my data* from the interview list and click on the *See the data*. In the associated parameters window, find the parameter block called Non-Detects. Within this block, you will see the three available options. The default is half the detection limit. This file has both duplicates and non-detect values. Let's use the label field to better understand what SADA is doing.

If you have not already done so, check that you have *Soil* and *Ac-225* selected as your media and data set. Select *Detected* as your label field (not *Plotted Value*).

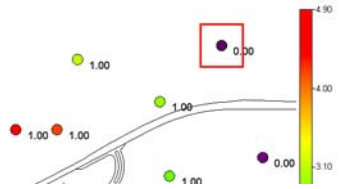


For Ac-225, you should see the following result (without the red circles).



Three red circles are added to the result on the right-hand side to indicate the location of three non-detected values for Ac-225. You can readily identify the non-detects because the detection field will have a zero value. If the data set is geographically dense, it may be difficult to visually search through the labels for your non-detect. Fortunately, a label search feature is available. From the main toolbar select *Tools→Search Labels* and the following window will appear:

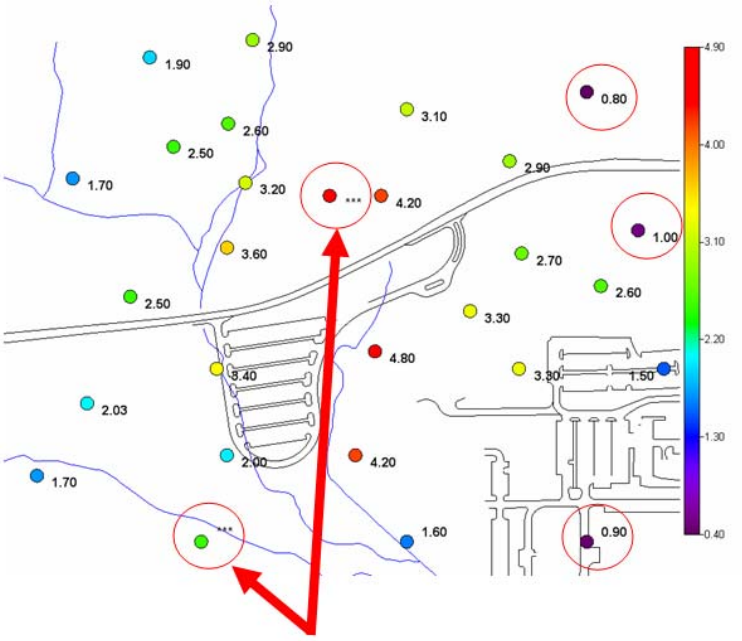
Enter a value of 0.00 in the text box and press *Find*. Note that you must be searching for EXACTLY the right string. For example, searching for 0 would result in no finds. SADA will enclose the first point it finds with this label in a large red box. Each time you press the Find button, SADA advances to the next string it finds. In this manner, you can quickly search the current layer for all labels with a certain text.



Press *Clear ID Circle* and then close the window by pressing the *X* button in the upper-right-hand corner.

This is a good opportunity to demonstrate the difference between the *Value Field* and the *Plotted Value* in the labels box. Recall that the *Value Field* will return the *Reported Value* for a data point. The *Plotted Value* option will label the point with the final result that SADA uses in the map. For detected values, these two values may vary. The result in the *Value Field* will be the detection limit for non-detected values. The *Plotted Result* however, will be half the detection limit (the default option we are currently using). Select *Value* in the labels box.

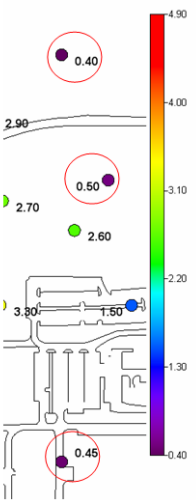
For *Ac-225*, you should see the following result (without arrows and circles). Each label is showing the actual value found in the data set for that sample point.



*** means duplicate

There are five import sample points to notice here, each contained in a red circle. The two circles on the left side show a trio of asterisks instead of a value. This indicates that these two points each have duplicate values. That is, more than one record exists in the field for each coordinate location for Ac-225. Therefore, SADA cannot produce a single label and instead draws asterisks to indicate this situation. This will be discussed shortly. Now notice the values

associated with the non-detect values (0.8, 1.0, and 0.9). Next, switch from *Value* to *Plotted Value*.



These values are the resolved values that SADA will use for all operations, including risk assessment.

Now switch the non-detect option from half to the full detection limit. Notice how the labels change to their full value. Switch the non-detect option from full detection to zero. The non-detected values are all zero now.

Make sure you have half the detection limit selected before continuing.

Dealing with Duplicates

Returning now to the issue of duplicates, click “See the data” and look for the information block called *Duplicate Data*. There are several options for resolving situations where you have duplicate values at one or more points. It is important to clearly define what is meant by a duplicate. The term duplicate has different meanings in different contexts. We use it in a very narrowly defined way.

If two or more points share the following properties they are considered duplicate values. In other words:

1) Each point shares exactly the same coordinate (x,y,z).
2) Each result is a measurement of the same contaminant.
3) Each result was sampled within the current date interval.

The first condition seems obvious, but it is not necessarily well understood. In some cases, true duplicate values (measurements of Ac-225 at the same location) will be reported with slightly different easting, northing, or depth measurements. You might intend that they be considered duplicates, but if they do not have exactly the same coordinates, they will be considered separate measurements.

The second condition is fairly straight forward. A measurement of Ac-225 at location (0,0,0) and a measurement of Barium at location (0,0,0) is never considered a duplicate value.

The third condition is very important. Most of the time, users do not use the *Data Query* feature located just above the duplicates parameters block. But the impact of this query on how SADA deals with duplicates is significant. Consider the following situation. The following two points

were duplicate values in real life. Note that the coordinates are exactly the same. Only the date differs by one day.

Attribute	Sample #1	Sample #2
Easting	100	100
Northing	200	200
Depth	5	5
Name	Ac-225	Ac-225
Value	33.5	38.2
Detected	1	1
Date Sampled	5/1/2008	5/2/2008

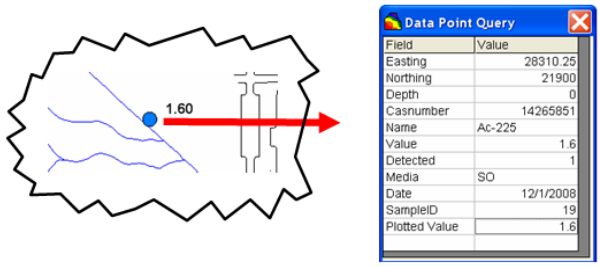
Here is how SADA would treat the duplicate values as a function of the time interval.

Date Interval	Duplicate Values?	Which one to use?
5/1/2008 – 6/1/2008	YES	Depends on the rule you choose.
5/1/2008-5/1/2008	NO	Only Sample #1 is used in the analysis. Since Sample #2 falls outside the date range, it is not considered in the analysis.
6/1/2008-7/1/2008	Doesn't matter	Neither is used because neither falls inside the date range.

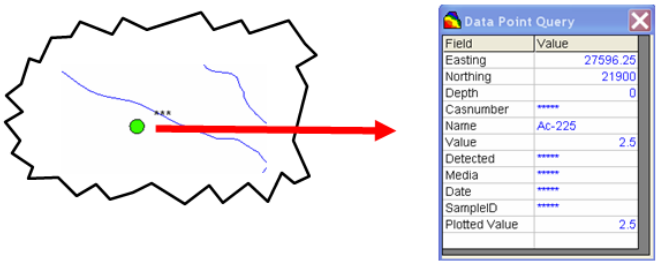
For monitoring data, there will be a lot of samples measured repeatedly at the same well. In doing a risk assessment, be aware that if you do not break up the date range into the sampling intervals (e.g., quarterly well measurements), SADA will assume that each well has a lot of duplicates, which is not truly the case. For monitoring situations, simply create multiple date ranges and perform a risk assessment on them individually.

An easy way to explore duplicate data is to use the label feature to label points with values from your *Value* field. This will produce the three asterisks like we saw earlier. Change the label box to *Value* now.

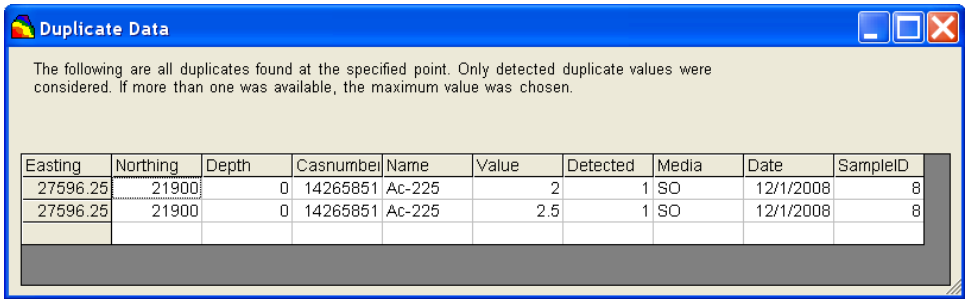
First click on the data point with value 1.60 on the south side of the site in about the middle. The point information box will appear and show the results for this point, taken directly from the sample record set.



Now, click on the neighbor to the left of this point with the three asterisks.



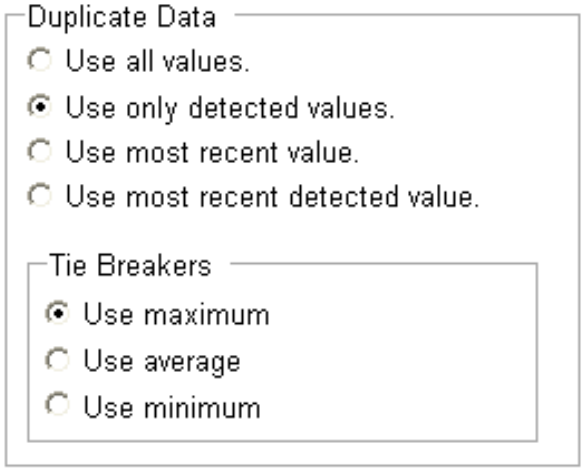
Notice now that the data point query box is populated with blue text and several fields contain asterisks. This indicates that because more than one Ac-225 value was found at this location, SADA can't show one single record. The "resolved" value of 2.5 is available though in the bottom line, next to *Plotted Value*. By resolved, we mean that the rules for duplicates were applied, and this was the final resolution. In a moment, we'll cover the rules for dealing with duplicates. Right now, click anywhere on the blue text and SADA will pull up those records identified as duplicate.



So, two Ac-225 samples were collected at this one point with two distinct measurement values of 2 and 2.5 pCi/g. In this case, the duplicate rule was to use the maximum detected value (2.5). Close this window and close the *Data Point Query* window.

Duplicate resolution rules are two-tiered. First, you must choose from a list of available data characteristics, such as *Use only detected value* or *Use all value*. This selection alone may pare down the number of duplicates you may have at any given point. In the event that two or more points meet this criteria (as happened above), however, a tiebreaker rule must be selected to decide or resolve the remaining points.

The duplicate rules are seen in the following image:



IMPORTANT! Keep in mind that some choices for resolving duplicates may entirely eliminate the point altogether. Suppose, for example, at point (x,y,z) we have two undetected measurements of Ac-225. If you have selected *Use only detected values*, both points will fail and no sample will be used at (x,y,z).

If you have a large dataset (1000s of points) and most are duplicates, it will really slow SADA down. If you are not using SADA as a monitoring tool and the slowdown is impeding your progress, you should deal with duplicates before bringing them into SADA.

Once you have explored the role of non-detects and duplicates in your dataset, you are ready to setup the human health risk model.

Remember that SADA is a 3d tool that can also work with 2d data. If your data set is 3d, you may have points below or above the particular layer you have selected. As a result, you may miss non-detects and duplicates on other layers as you explore your data. One option is to simply use the *Entire Data Layer* option. This results in one single think layer that covers the entire vertical extent of your data. The problem is that it might cause an untidy clustering of labels, making it difficult to see what is going on. *Tools→Search Labels* can help. Another option is to make sure you select relatively thin layers, so that ever label can be seen, but know that you must visit each layer to see all the data.

Setting up the Human Health Risk Model

To use the human health models in SADA, you must first “setup” the risk module. This is relatively easy to do. SADA comes with two Microsoft Access databases (.mdb) that contain relevant contaminant and exposure parameter values.

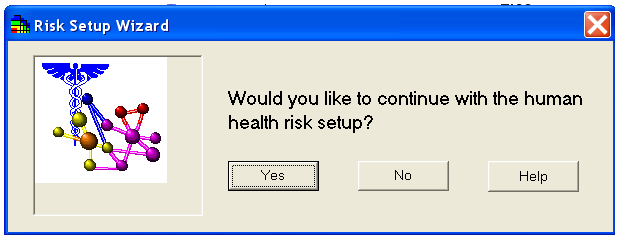
The ToxicologicalProfiles.mdb comes from Oak Ridge National Laboratory’s Risk Assessment Information System (RAIS - http://rais.ornl.gov/index.shtml). This toxicological database is maintained regularly, and recent versions of the database can be downloaded from the SADA home page in a SADA compatible format. The toxicity values were compiled using EPA's toxicity hierarchy, derived from values found in these EPA sources, or provided after contacting EPA. After a risk module has been setup, errors or small updates to toxicity values can be handled through the *Toxicological Links* menu. In addition to toxicity information, this file contains physical parameters used in human health exposure modeling. Parameters include

contaminant-specific Bioaccumulation Factors, Volatilization Factors, Particulate Emission Factors, Permeability Constants, Absorption Factors, Saturation Coefficients, and Radionuclide Half-Lives.

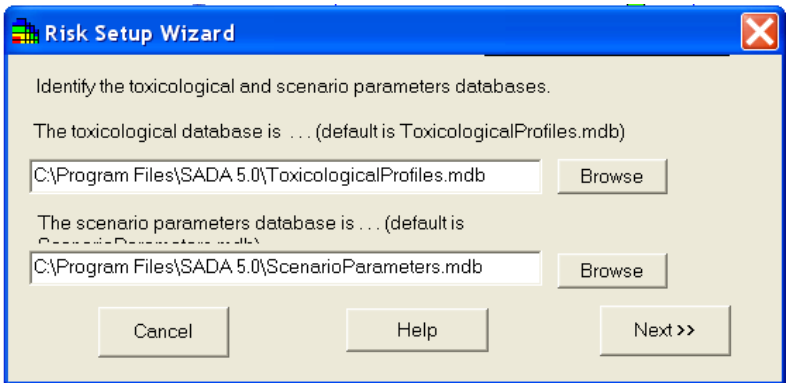
The scenario parameters database, ScenarioParameters.mdb, contains default parameters from EPA guidance and local guidance for the Southeastern US; the parameters can be customized as necessary to account for local exposure conditions.

When you set up the risk module, SADA will attempt to match the contaminants in your SADA file with those in the ToxicologicalProfiles.mdb file or equivalent. The matching will be based on both contaminant name and CAS number, if you provided this field during import. If you did not include a CAS number column during import, the setup will still work, but SADA may have less success matching contaminants.

To start the process, select from the main menu *Setup→Human Health Risk* and the startup window is displayed.



Press *Yes*, and you will be presented with an opportunity to choose the toxicological profiles database and the scenario parameters database you would like to use.

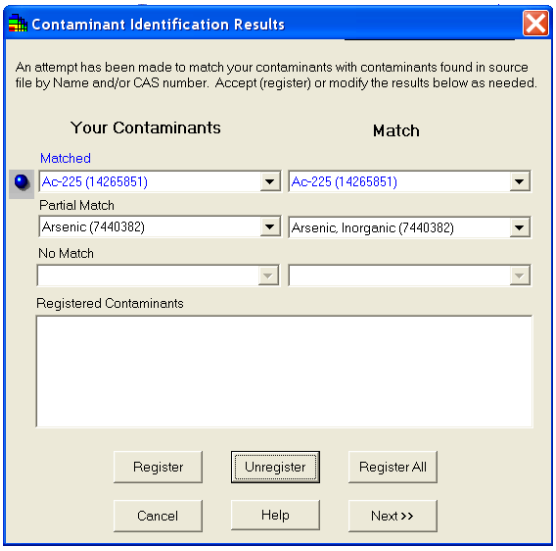


SADA provides these two databases during installation and on the website. Their content is imported into your SADA file during setup. From within SADA, you can modify the database with some user-friendly interfaces, but these changes are not made to the external database. Rather, they will only occur within your SADA file. Each time you create a new SADA file, you would need to make the modifications again. This is a problem when SADA will be applied to a number of sites on the same project.

In this case, you should create copies of the files and modify the content to suit your site-specific needs outside of SADA. You can change any of the field values, but you may not change the structure of the tables. Once you have calibrated the database for your project needs, select your customized copy of the database instead of the one sent with SADA. The changes will appear as you wish in each new SADA file.

Navigate to where the ToxicologicalProfiles.mdb is located (probably in the SADA directory location) and select it for import. Repeat this for the ScenarioParameters.mdb. Press *Next>>*

when you have selected these two files. The next window presents the matching results and allows you to modify and approve each match result.



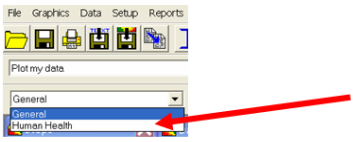
SADA now attempts to match each contaminant in your file with a contaminant found in the toxicological database. SADA searches by CAS number first (if available) and then by name. If the CAS number and name match exactly, SADA classifies it as *Matched*. If only the name or the CAS number match, then the classification is *Partial Match*. Finally, if no match is found for either, the classification is *No Match*. These three classifications are presented in the *Contaminant Identification Results* window.

On the left side of the window, your contaminants have been divided into the three categories. To view a resulting match for any contaminant, click on the down arrow and select your contaminant from the resulting drop down list. The corresponding selection on the right hand side will change to show SADA's match for your contaminant. If the match is acceptable, press the *Register* button. If all matches within a category are acceptable, press *Register All*. To unregister a matched pair(s), select the pair(s) in the *Registered Contaminants* box and press *Unregister*. Your contaminants will return to their original classification with their original match.

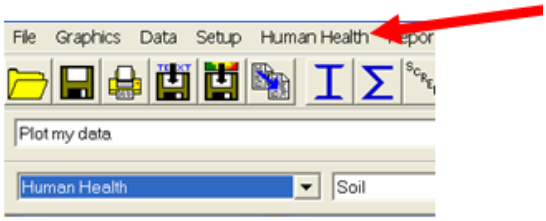
If no match is available for some of your contaminants, you may leave them as unregistered. Later, if the toxicological information becomes available, you may link these contaminants (or re-link registered contaminants) separately without setting up the entire risk module again.

Press *Next>>* to conclude setting up the risk module.

Once the module is complete, *Human Health* will appear in the analysis drop list on the second toolbar. Select this now.

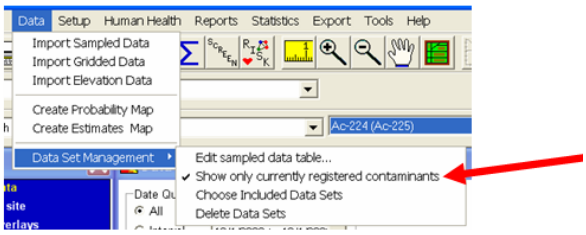


When this option is selected, a *Human Health* menu will be visible on the main menu bar.

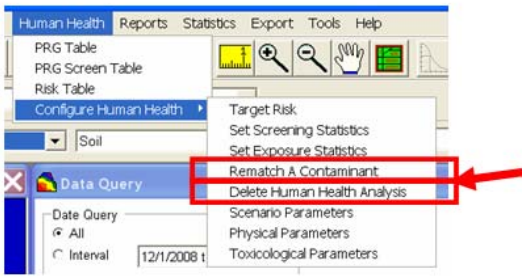


You may reset the risk module at any time; simply select *Human Health* under the *Setup* menu of the main window. The process is the same as before; however, SADA will give you the opportunity to reset or skip the toxicological and scenario parameter component identification.

Important Tip! If a contaminant is not registered during setup, it may no longer appear in the list of contaminants WHEN *Human Health* is selected as the analysis type. You should be aware of this convenience issue. If you want to see all contaminants when you select *Human Health*, whether they are registered or not, you can set this under the *Data Set Management* menu. Simply check or uncheck *Show only currently registered contaminants*.

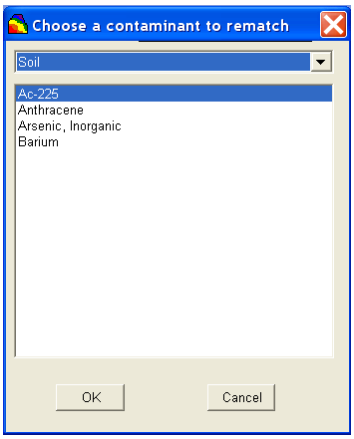


From the main menu select *Human Health→Configure Human Health*. These menu items provide you access points for calibrating your risk models, viewing tabular risk results, and if necessary, to rematch particular contaminants or delete the human health altogether.



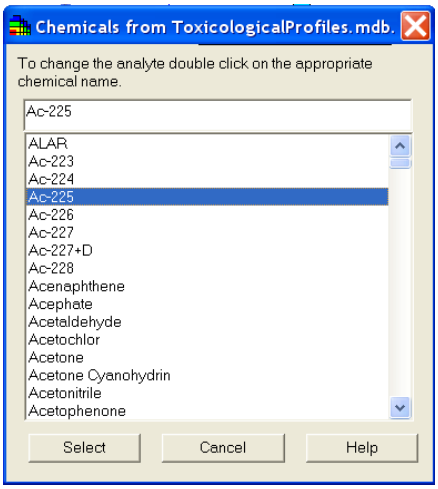
Rematching a Single Contaminant

Save your SADA file now. Select *Rematch A Contaminant* from the *Configure Human Health* drop-down list. In the next window, contaminants are organized by data type (media). Select *Ac-225* as the contaminant you wish to rematch and press *OK*.

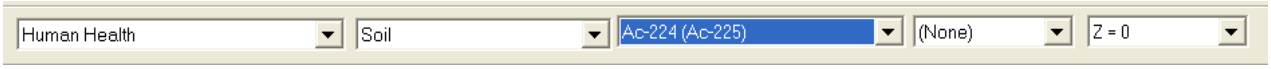


SADA will present you a open dialogue box. Navigate to where the ToxicologicalProfiles.mdb file is located, select it, and press *Open*.

You are now shown the complete list of contaminants found in this database. Notice that SADA has pre-selected *Ac-225* for you.



For the sake of experiencing this feature, select *Ac-224* instead of *Ac-225* and press *Select*. The window disappears. Select *Soil* and then *Ac-224* from the drop-down lists.



Notice now that SADA reports the original name in parentheses. This allows you to see how the data was matched during setup. If the names are identical, you won't see the parentheses.

Re-enforcement

Repeat the single contaminant match above and rematch Ac-225 to Ac-225 in the ToxicologicalProfiles.mdb. Save your file.

Deleting the Human Health Risk Module

You can completely delete the *Human Health* risk model by selecting *Human Health→Configure Human Health→Delete Human Health Analysis*. When prompted to confirm, answer *No*.

Working with non-spatial data, contoured data, and using exposure concentrations calculated outside SADA.

It is possible to calculate risk on modeled or gridded data types and even on data sets without spatial coordinates. In some cases, users may want to use exposure concentrations other than the four basic options in SADA: *Maximum Detect, Mean, UCL95*, and *Minimum of Max Detect and UCL95*.

Contoured Data

SADA certainly permits users to import raster or contoured data "as is," without any additional modification (See Chapter 7, "Importing Modeled Data"). You cannot perform *Human Health* risk on these data types as of now, however. You'll need to import the contoured results as point data, specifying each x,y,z value for each sample value. You will also need to add the *Media Type* and *Name* column. This will take some processing outside of SADA. In a future version, SADA will be able to conduct a risk analysis on contoured results without this extra effort. For now, this is the only method.

Non-Spatial Data

Sometimes samples are collected without noting the coordinates. If this happens, you can essentially "fake" the coordinates outside of SADA using Excel or something similar. You will need to be aware, though, that the coordinate system is fake, and any spatial trend you might see is purely artificial. You cannot responsibly conduct any type of spatial risk calculations. You can, though, fully use the traditional risk models (tabular PRGs/risk/screens).

Exposure Concentrations From Outside SADA

Normally, SADA calculates exposure concentrations internally, using normal or lognormal assumptions. If you wish to use another distribution or means of calculating the exposure concentration, you certainly can. Just as with the non-spatial data, you will simply fake the coordinate for each representative concentration for each contaminant. Each contaminant essentially has only one "sample" and that is the exposure concentration value itself. This can then be run through the risk models to produce a traditional or tabular risk assessment. As with non-spatial data, you cannot sensibly apply any spatial tools. You can establish two SADA files. In the first SADA file, you use the exposure concentrations with fake coordinates. In the second SADA file, you import the actual data values with true coordinates and then conduct spatial screens, remedial decision designs, and secondary sampling designs which do not rely on any particular underlying distribution for the exposure concentration.

Next Steps

In the next chapters, we will discuss how to calibrate your model and perform some traditional tabular risk assessments.

Chapter 19: Calibrating The Risk Model

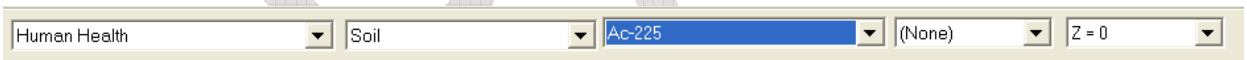
Once your SADA file has been setup for Human Health risk, the next step is to check that the risk model parameters are exactly what you need for your current assessment. The values that are sent with the ToxicologicalParameters.mdb and ScenarioParameters.mdb are primarily default EPA values for the southeastern US. These may not necessarily be the site specific values you'll need for your site.

There are two ways to customize you analysis. If you are going to do several small projects using the same customized set of parameter values, then you should directly modify the databases themselves using Microsoft Access. The fields are self explanatory and their values can be modified to suit your needs. Name this modified database something different before you begin, so you retain your copy of the original database. Be sure to not modify the table structures in any way. SADA will not behave predictably if structural changes are made, such as reordering fields, renaming fields, renaming tables, deleting tables and so forth. Once the database has been modified for your needs, you can use it to setup the Human Health risk module in each new SADA file you create.

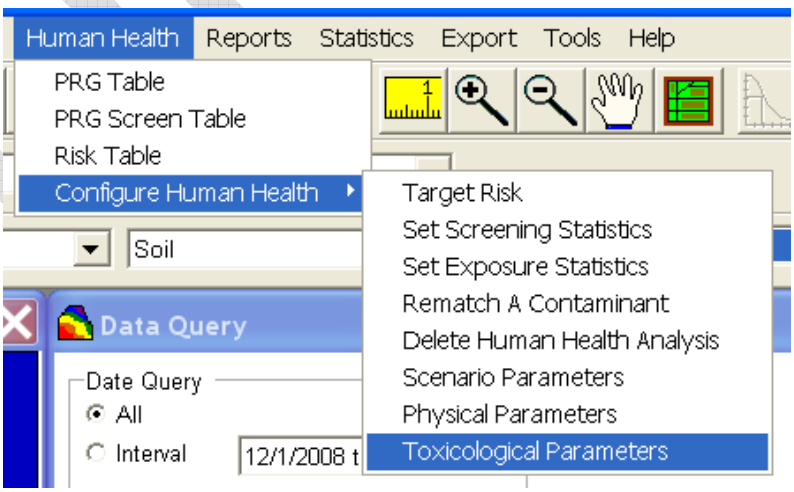
If you do not need to modify the database or if you feel uncomfortable dealing directly with the databases, you can calibrate the parameter values within each SADA file. There are some friendly windows in SADA that help you do this. Be aware, though, that the original database will not be modified during these changes – only the values within your SADA file. We will now cover how to modify or calibrate the model in SADA. Let's being with the toxicological profile data.

Setting toxicological parameters

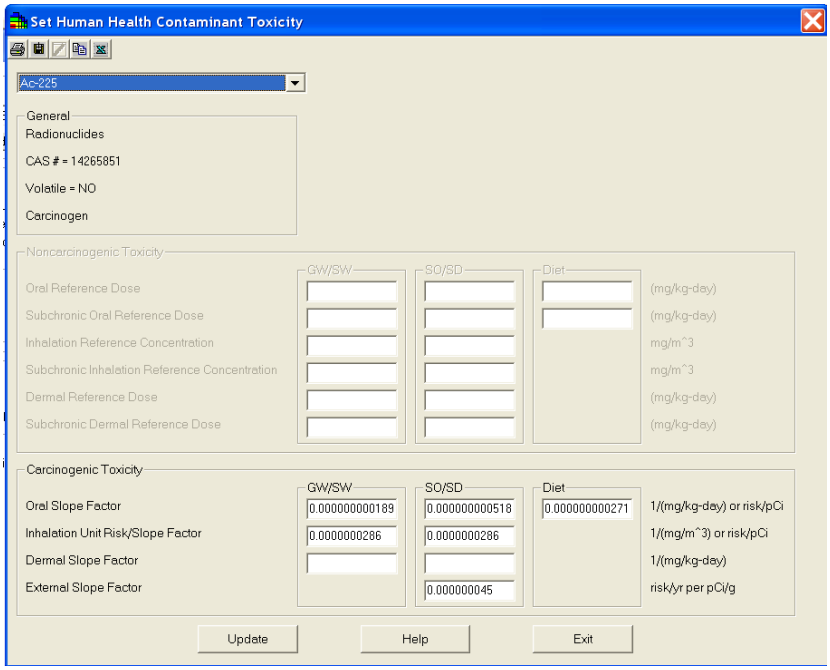
Begin by opening the file CalibratingHHRiskModel.sda. This is the same file you created in the last chapter, called HumanHealthRisk.sda. Select *Human Health* from the analysis drop-down list and make sure you have selected *Soil* and *Ac-225*.



From the main menu, select *Human Health→Configure Human Health→Toxicological Parameters.*



You will be presented with the toxicological parameters window. In this window, all the contaminants for the currently selected media type are presented. You can select contaminants individually by choosing them from the drop list in the upper-left-hand corner.

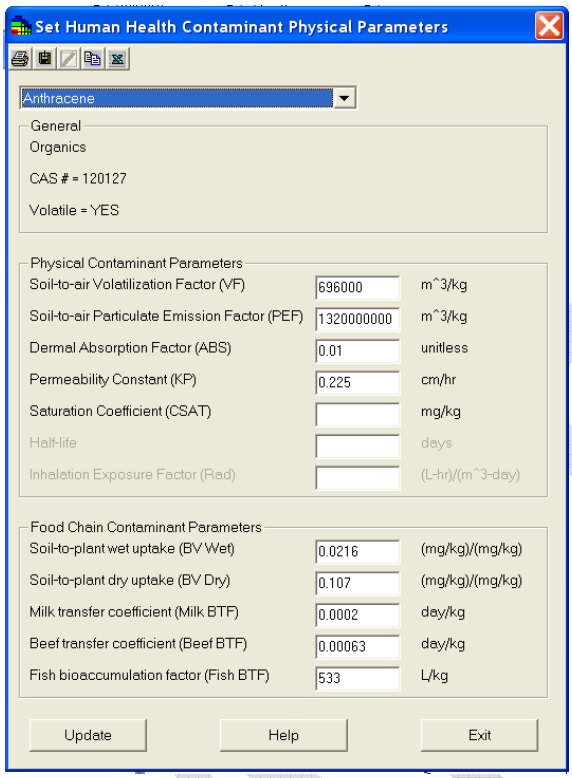


In the first information block, some basic information about the contaminant is provided. In the case of Ac-225, we are told it is a radionuclide and a carcinogen. Because it is not a non-carcinogen (some contaminants are both), the noncarcinogen parameter set is grayed out. Here, you can modify the carcinogenic toxicity/radiological values. Make changes as necessary and press the *Update* button before switching to another contaminant. Options for printing the results, saving the results to a text file, adding to autodocumentation, copying to the clipboard, and exporting directly to Excel are found in the toolbar buttons at the top. A complete explanation of the risk parameters are found in the *Help* file and are not repeated here.

Switch now to *Arsenic*, *Inorganic* and notice that both carcinogenic and non-carcinogenic parameters are now enabled.

Physical Parameters

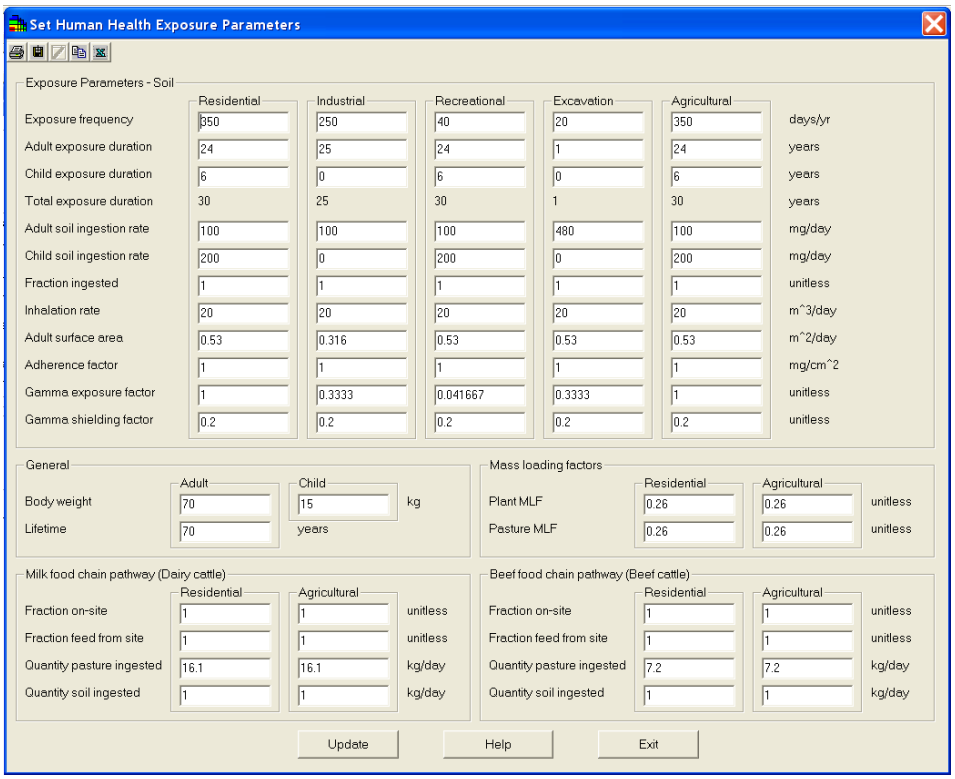
Now we will modify some contaminant-specific physical parameters that characterize how the contaminant behaves in the environment. These are important in exposure modeling. Select from the main menu *Health Risk→Configure Human Health→Physical Parameters*. You will be presented with the physical parameters window.



You can scroll through the different parameter values for each contaminant in soil by switching the selection in the drop list in the upper-left-hand corner. Options for printing the results, saving the results to a text file, adding to autodocumentation, copying to the clipboard, and exporting directly to Excel are found in the toolbar buttons at the top.

Scenario Parameters

Unlike the first two parameter sets, these parameters characterize the behavior of the exposure rather than the attributes of the contaminant. From the main menu select Human *Health→Configure Human Health→Scenario Parameters*. The scenario parameters window appears.

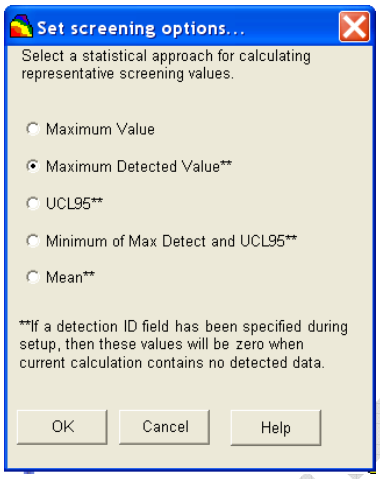


Here you can customize scenario or behavior-specific parameters for five different land uses: *Residential*, *Industrial*, *Recreational*, *Excavation*, and *Agricultural*. In addition, a number of other exposure-related scenario parameters are presented. A full explanation of these is beyond the scope of this user guide. If you do not see the land use you are interested in, simply choose one of the existing land uses and calibrate it to your specific land use. You will not be able to rename it, but the model will be calibrated to your exact situation. In the final report, you'll simply need to change the land use to its proper name.

Choosing the Representative Concentration

The representative concentration is a single concentration value that is statistically representative of the data set. There are four statistical values that can be used to calculate this value and can be established separately for exposure concentrations and screening concentrations.

Select from the main menu *Human Health Risk →Configure Human Health→Set Screening Statistics.* You will be presented with five options:

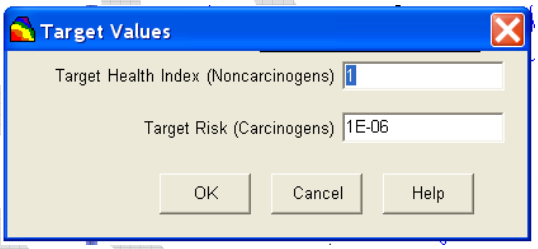


Choose from the *Maximum Value, Maximum Detected Value, UCL95, Minimum of Max Detect and UCL95,* and *Mean*. By convention, if no data are detected for a given contaminant, the representative concentration for the purposes of screening will be zero for all choices but the *Maximum Value*.

Select from the main menu *Human Health Risk →Configure Human Health→Set Exposure Statistics.* You will be presented with the same five options.

Setting The Target Risk/Index (for PRG calculations)

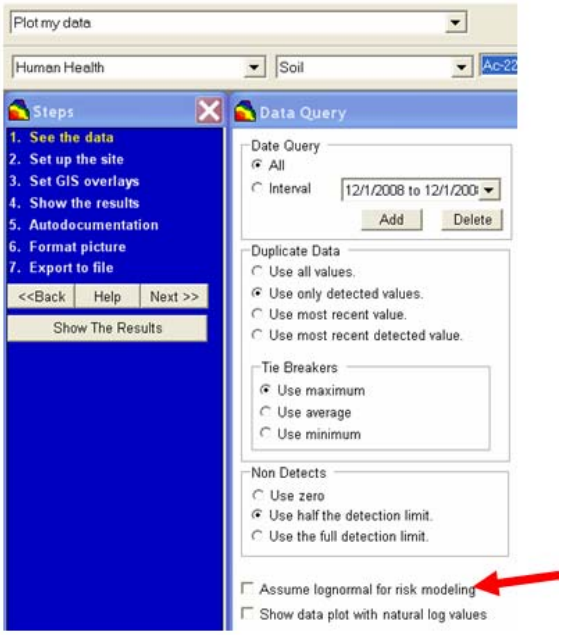
If you are calculating PRG values, you will need to set the *Target Health Index* and/or *Target Risk*. From these two values, the models may be run in reverse to determine the concentration that would correspond to this risk/index value. From the main menu select *Human Health→ Configure Human Health→ Target Risk.*



In this window, enter the *Target Health Index (Noncarcinogens)* and the *Target Risk (Carcinogens)*. We will leave these as 1 and 1E-06 for now. Press *OK*.

Setting the Distribution for Each Contaminant

For representative concentrations that require an assumption about the underlying contaminant distribution, SADA provides two options: normal and lognormal. You will need to specify this for each contaminant under the *See the data step*.



It is easy enough to declare *Assume lognormal for risk modeling* by just checking this option. If this option is blank, then a normal distribution is assumed. You should consult the histogram of data to guide you in making this choice.

Recent guidance from US EPA suggests using the PROUCL95 package to determine what the distribution for the data should be and what the resulting exposure concentration value would look like. If you are required to use this guidance, then you will need to use the PROUCL95 package outside of SADA to generate the exposure concentrations and import them as single "data" points, one for each contaminant. When you set the exposure statistics, select *Maximum Value*, and you will be using the representative exposure value directly in the model.

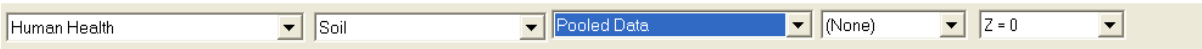
When the model has been calibrated and distribution choices made for each contaminant, you are ready to start conducting risk assessments.

Chapter 20: Tabular PRG, Screen, and Risk Calculations

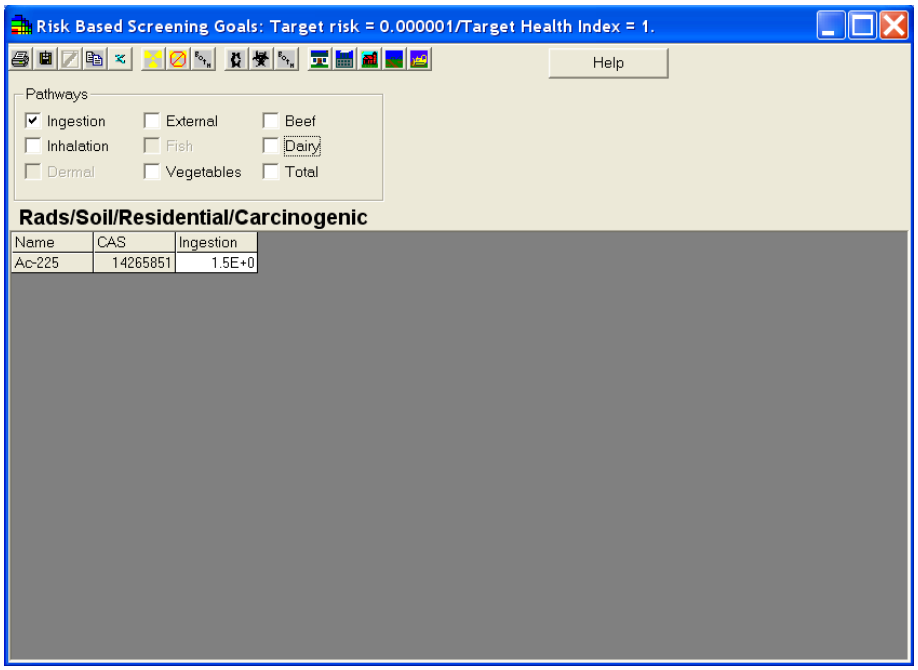
Now that you have set up and calibrated the risk model, you can begin doing some risk calculations. One of the first things risk assessors do is perform calculations of PRG values and then conduct a screening of their data to identify risk drivers. This allows users to pare down lengthy lists of contaminants to just those few causing considerable problems. If you have a list of PRGs or equivalent already available for your site outside of SADA, you will want to bring those in as a custom analysis instead of using this method (see Chapters 14-16).

Tabular PRGs

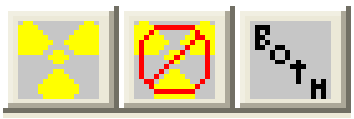
Let's begin by calculating PRG values with our newly calibrated model. You can calculate PRGs one at a time for each contaminant, or you can calculate them for all contaminants in a given media at once. Let's do the latter. Close out any SADA file you are working with and open TabularRiskResults.sda. Select *Human Health* as your analysis, *Soil* as the media, and then *Pooled Data*.



From the main menu select *Human Health→PRG Table*. The PRG table is presented and populated with relevant PRG values in units of mg/kg, mg/L, pCi/g and pCi/L, depending on whether the contaminant is a radionuclide or not and depending on the media type. Here we are working with soil, so units will be either mg/kg for non-rads and pCi/kg for rads.



The first question that may arise is why does the result show only Ac-225? SADA organizes the risk results by categories. In particular, the categories are: rad or nonrad, then carcinogen, noncarcinogen or both. You can elect to choose either rad, nonrad or both by selecting one of the following buttons respectively.



Following this selection you can choose to see either *Carcinogenic, Noncarcinogenic, or Both* PRGs by choosing one of these buttons respectively.

To see how these categories behave, select the *Rad* button first. Notice that the carcinogen/noncarcinogen/both trio of buttons becomes disabled. This is because under a radionuclide contaminant, only carcinogenic effects are considered. Therefore, the choice is not left up to the user. Since the only radionuclide is Ac-225, we see only one entry when the *Rad* button is selected.

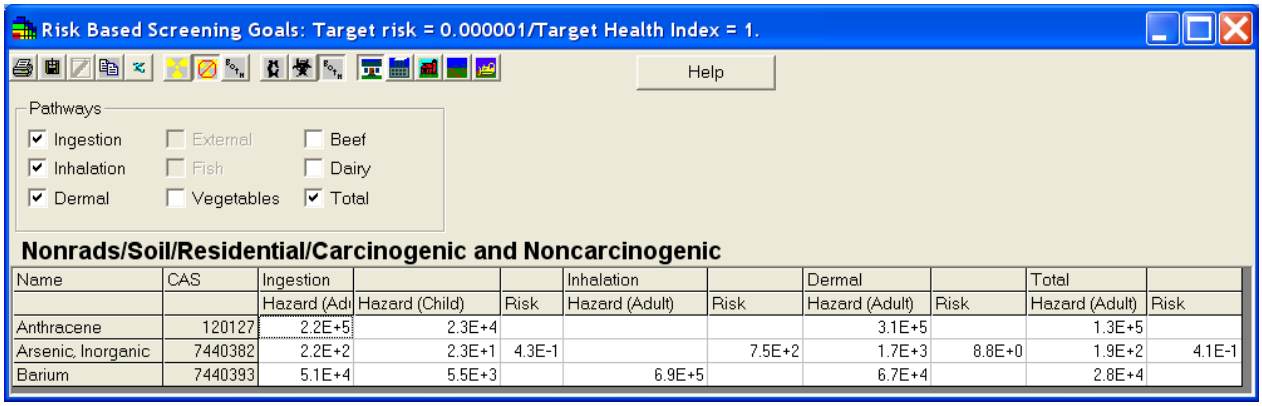
Now select the *NonRad* button. Notice that the carcinogen/noncarcinogen/both trio of buttons becomes enabled. Among our soil contaminants there is only one carcinogen – Arsenic. Press the noncarcinogenic button next. Now three of our soil contaminants appear – Barium, Arsenic, and Anthracene. Finally, view both carcinogenic and noncarcinogenic with the last button. The same three appear (Plus some additional columns are added, more on this shortly).

Now return to the *Rad/NonRad/Both* trio and select *Both*. When you are looking at *Both*, only carcinogenic results are considered because they are common to rads and nonrads. Notice that the *Carcinogen/Noncarcinogen/Both* trio of buttons is once again disabled. We have only two carcinogenic contaminants in our soil (Ac-225 which is a rad, and Arsenic which is a nonrad.) Select *NonRad* and then *Both*.

SADA also organizes based on land use. So, beginning with the first button to the right of the *Carcinogen/Noncarcinogen/Both* trio of buttons we see *Residential, Industrial, Agricultural, Recreational, and Agricultural* land use scenarios. Try switching between these results and notice the PRG values change.

Finally, we have options for which exposure pathways to include. These are found in the *Pathways* parameter block.

Some pathways are not permitted under certain conditions. For example, *Fish* is disabled because we are dealing with soil data. External is disabled because we are not using radiological contaminants (we have the *NonRads* button pressed). Try checking *Inhalation* and notice that an inhalation column is added. Now check the *Dermal* box. The last option is *Total*. If you check this box, then the total PRG is calculated, which includes all the exposure pathways that you have currently selected.



Name	CAS	Ingestion			Inhalation		Dermal		Total	
		Hazard (Adu	Hazard (Child)	Risk	Hazard (Adult)	Risk	Hazard (Adult)	Risk	Hazard (Adult)	Risk
Anthracene	120127	2.2E+5	2.3E+4				3.1E+5		1.3E+5	
Arsenic, Inorganic	7440382	2.2E+2	2.3E+1	4.3E-1		7.5E+2	1.7E+3	8.8E+0	1.9E+2	4.1E-1
Barium	7440393	5.1E+4	5.5E+3		6.9E+5		6.7E+4		2.8E+4	

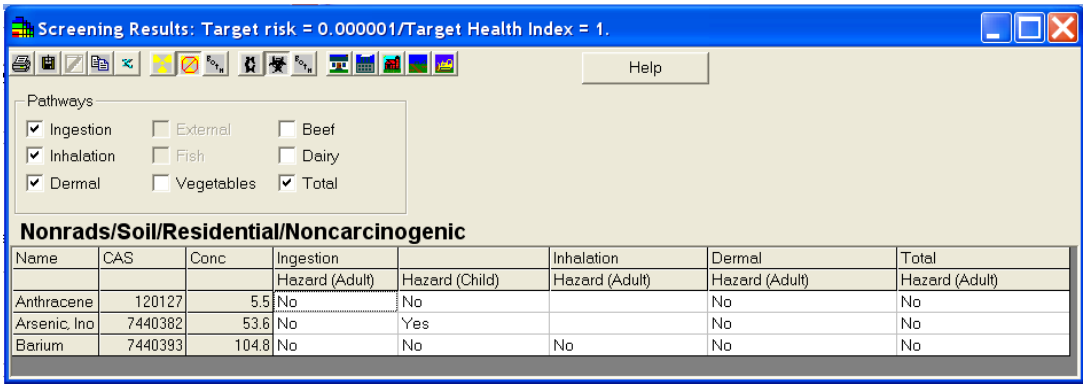
The table above reveals that we have a value of 8.8E+0 for Arsenic, Inorganic when exposed by dermal contact for carcinogenic risk. This means that in order to be protective of human dermal contact with the soil, the concentration in the soil will need to be less than 8.8mg/kg.

Some of the cells are blank. This is usually because there isn't enough information to calculate a value for every exposure pathway. This is sometimes limited by the toxicological information.

Note: Tthe risk for residential ingestion of arsenic is 0.043mg/kg. Now go back to the scenario parameters and change the exposure frequency from 350 days per year to only 50 days per year. Regenerate the PRG table and notice that the risk for residential ingestion of arsenic is now 3.0mg/kg, a considerable decrease. From a risk perspective, the limit can be raised because the receptor will only be living there 50 days out of each year rather than 350. Return to the scenario parameters page and change the value back to 350.

Tabular Screens

Once you are comfortable with using the PRG calculation feature, the tabular screening feature works exactly the same. Using the statistical choice for the representative screening concentration specified in the *Human Health→Configure Human Health→Set Screening Statistics* (see Chapter 19), SADA will now compare this value with each of the PRG values seen in the previous exercise. To perform a screen, select *Human Health→PRG Screen Table* or press the *Screen* button . The tabular screen window appears.



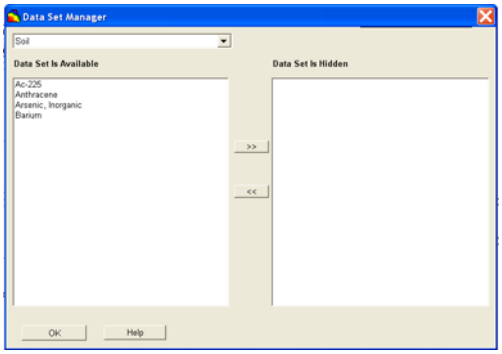
Name	CAS	Conc	Ingestion		Inhalation	Dermal	Total
			Hazard (Adult)	Hazard (Child)	Hazard (Adult)	Hazard (Adult)	Hazard (Adult)
Anthracene	120127	5.5	No	No		No	No
Arsenic, Ino	7440382	53.6	No	Yes		No	No
Barium	7440393	104.8	No	No	No	No	No

Instead of PRG values, the table is populated with *No*, *Yes*, or blanks. If a cell contains a *Yes*, then this contaminant is exceeding the acceptable screening criteria (PRG). A value of *No* means that it does not exceed the PRG value. A blank value means that the PRG was not available. This table is organized exactly the same as the PRG values table: Rad/NonRad/Both

or Carcinogen/Noncarcinogen/Both. To export these results to Excel, simply press the Excel button.

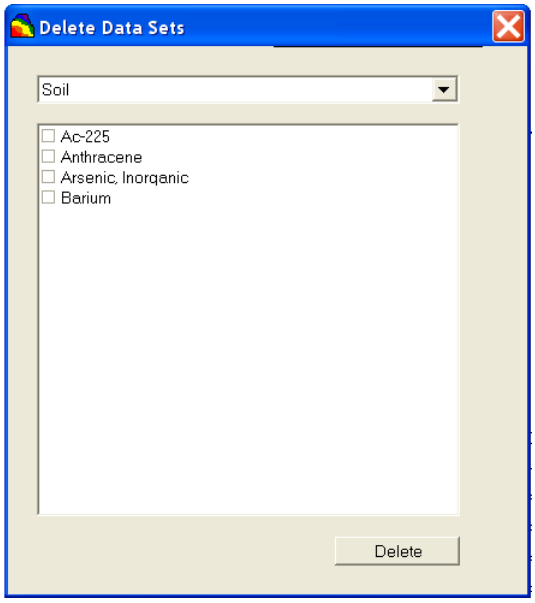
In some cases, project managers may exclude any contaminant that does not exceed the PRG. If this is the case, you can determine from this table which values do not exceed the PRG. Within SADA you can choose to *uninclude* or *delete* specific contaminants. Chapter 9 covers this in greater detail. For the sake of completeness, we will repeat some of those concepts here. In general, it is advisable to only uninclude the contaminants rather than delete them. If they are deleted, they cannot be recovered without re-importing them. If they are unincluded, they are simply hidden and can be re-included at any time. Close your PRG screen table.

To uninclude data, select from the main menu *Data→Data Set Management→Choose Included Data Sets.*



Using the arrows, move selected contaminants to and from the hidden group. If a contaminant is hidden, the data is not deleted nor are any of the modeling parameters. Rather, it is simply not included in the data set list boxes or in any output. Close this window.

To delete a contaminant entirely, select from the main menu *Data→Data Set Management→ Delete Data Sets.*



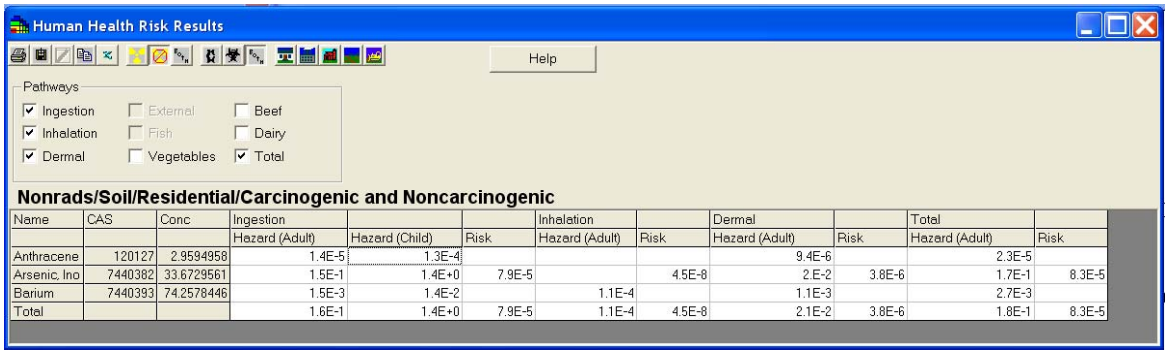
In this window, check the box next to the contaminants you wish to delete and press the *Delete* button. You will be asked to confirm deleting. Close this window without deleting any data sets.

Forward Risk

A major result in a risk assessment is the forward calculation of risk. In this case, the representative exposure concentration specified by the choice in *Human Health→Configure Human Health→Set Exposure Statistics* is used in the forward calculation of risk. The outcome will be a risk value for carcinogens and a health index value for noncarcinogens.

Note: As previously mentioned, if you are using an outside method for calculating exposure concentrations, you should import these exposure values into a separate SADA file as if they were "data" values. You will need to provide a fake coordinate for each concentration value (e.g., 0,0). You will then choose to use the Maximum Value for the representative concentration. This will obviously be equal to the exposure value you imported.

To execute a forward risk calculation, *select Human Health→Risk Table* or press the *Risk* toolbar button (). The risk table is generated and is organized in exactly the same format as both the PRG and screening table.



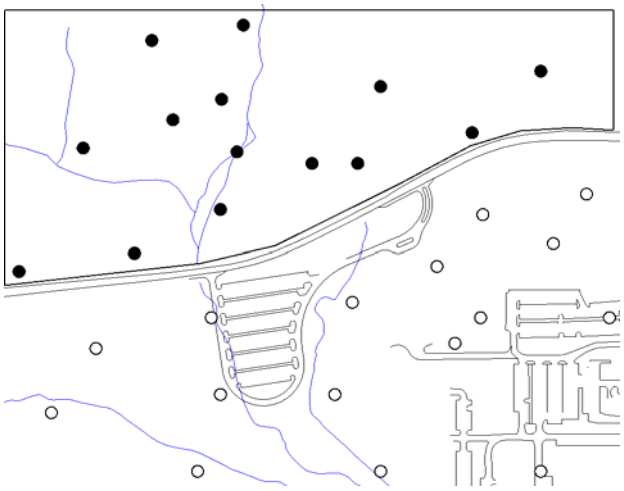
Name	CAS	Conc	Ingestion			Inhalation		Dermal		Total	
			Hazard (Adult)	Hazard (Child)	Risk	Hazard (Adult)	Risk	Hazard (Adult)	Risk	Hazard (Adult)	Risk
Anthracene	120127	2.9594958	1.4E-5	1.3E-4				9.4E-6		2.3E-5	
Arsenic, Ino	7440382	33.6729561	1.5E-1	1.4E+0	7.9E-5		4.5E-8	2.E-2	3.8E-6	1.7E-1	8.3E-5
Barium	7440393	74.2578446	1.5E-3	1.4E-2		1.1E-4		1.1E-3		2.7E-3	
Total			1.6E-1	1.4E+0	7.9E-5	1.1E-4	4.5E-8	2.1E-2	3.8E-6	1.8E-1	8.3E-5

These results can be exported directly to Excel by pressing the Excel button. There, they can be formatted as you wish. You can also print them (), copy them to the clip board(), save them to a comma delimited file(), or add them to your report document (), see the Autodocumentation section in Chapter 45).

Performing Risk Assessments on Sub-Areas of the site

Suppose you want to conduct a risk assessment only on a subset of the site. This is easy to do in SADA using simple polygon drawing elements. First, review the section on Working with Polygons in Chapter 4. Once you have reviewed how to create a polygon, continue reading.

Close out any tabular risk window you may have open. In the steps window, select *Setup the site* and notice the *Current layer polygons* parameter block in the parameters window. Add a new polygon called "North of the Road." Draw a polygon similar to the one below.



We will exclude all points south of the road. Now revisit the PRG Table, Screen Table, and Risk Table discussed above. Notice the changes in values? You can focus your risk assessment on any subarea of the site as easily as this.

Moving Toward Spatial Risk

Tabular results are useful in getting a handle on what contaminants may be driving risk, but these results do not indicate *where* risk exceedances may exist. Localized events may be driving the risk. Special attention to those areas, particularly in the remedial design phase, can more efficiently target the area of concern. The next chapters begin the discussion on this approach, which is continued later in the chapters on decision analysis (Chapters 35 and 36).

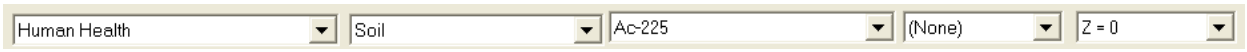
Chapter 21: Spatial Risk

In the previous chapter, you saw how to generate tabular risk results. These types of calculations are important classical approaches to risk assessment that help identify risk drivers. These same models can be integrated with spatial tools to show you *where* exceedances occurred, contour risk values, support the remedial area of concern, and even motivate secondary sampling strategies. This chapter will cover spatial data screens and point risk and introduce the more advanced principles of risk contouring, remedial design, and secondary sampling strategies. These principles are taken up in later chapters. SADA can also generate risk, remedial designs, and secondary sampling designs based on ecological risk and custom designs.

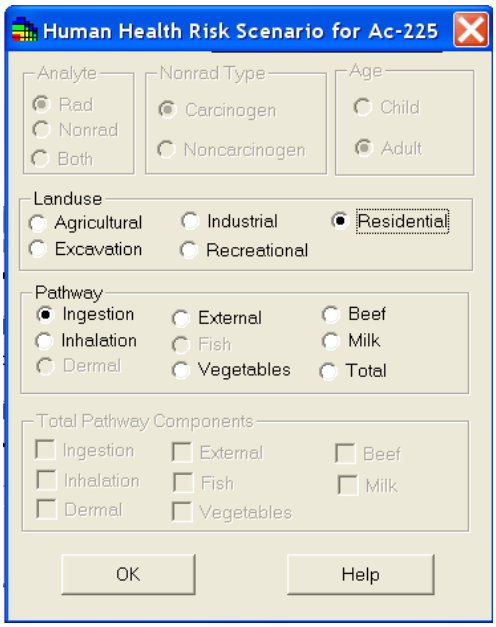
Spatial Screens

In Chapters 13 and 15, we demonstrated how to perform simple and informative data screens using user-defined or custom criteria values. The process is the same for human health modeling. In a human health analysis, the screening criteria (PRG) are compared to every individual sample point for a given contaminant. If the criterion is exceeded, the point is highlighted with a screening box. With this simple tool, it is easy to spatially determine where risk may be a greater concern.

Close out any SADA file you may have open and open the file SpatialRiskResults.sda. Select *Human Health* as the analysis, *Soil* as the media, and *Ac-225* as the contaminant.

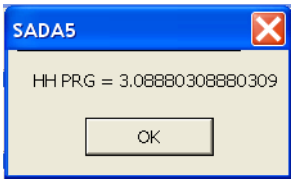


In the interview drop list, select *Draw a data screen map* and press the *Show The Results* button. You will be presented with a series of choices analogous to the options seen in the *PRG*, *Screen*, and *Risk* tables. You can choose between *Rad/Nonrad* and *Carcinogenic/Noncarcinogenic*. You can choose *Landuse* and *Pathway* of exposure.

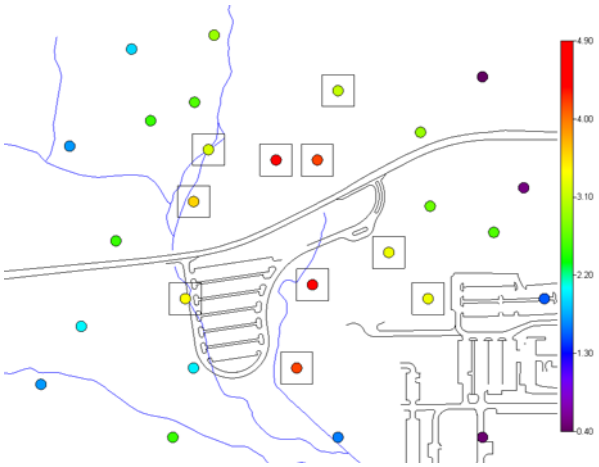


In this example, Ac-225 is a rad, so there is no option to change this. It is also a carcinogen and, for this type of contaminant, only adult exposure (which includes childhood) is included. Switch

the landuse to *Industrial* and the pathway to *Ingestion*. Press *OK* and SADA will report the PRG value for this selection. This will be exactly the same as the value seen for this combination in the tabular PRG table.



Press *OK* and SADA will report the number of exceedances and will highlight those values exceeding this particular PRG value. For Ac-225 with the selections *Industrial* and *Ingestion*, there are 10 values that exceed the PRG value.

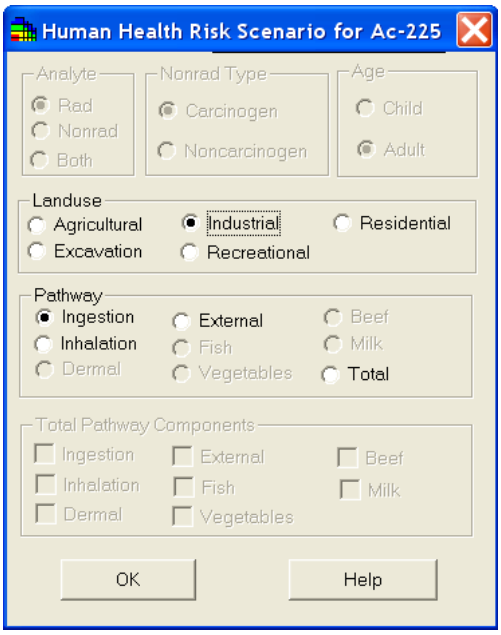


Reapply the screen with *Residential* and *Ingestion* selected. This time 24 points will exceed the criteria. At this time, there is no method for simultaneously screening all data spatially at once. The equivalent is to do a point risk map for pooled data and identify exceedances using the legend. This is covered in the next section.

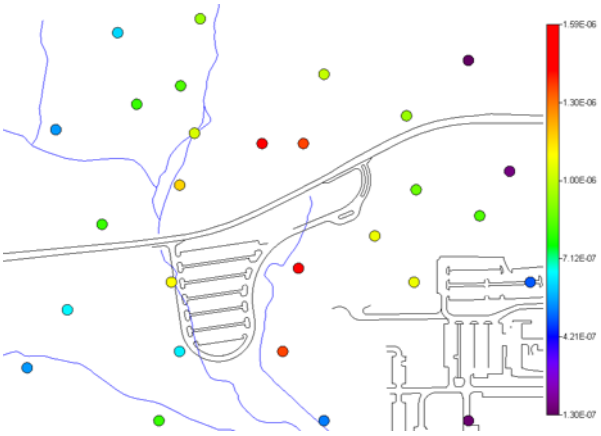
Point Risk

Data screens simply demonstrate where the exceedance occurred. To determine how severe the exceedance is, you can use the point risk feature to actually run each point through the risk model, producing either a risk or index result. It is also possible to pool your data and conduct pooled risk point maps. For groundwater well analysis, point risk is a very useful tool to determine ingestion risk at each well. These point risk maps can be generated for each sampling interval to generate “movies” of how risk varies over space and time.

We’ll begin with Ac-225. From the interview list select *Draw a point risk map*. Press *Show The Results* button. You will be presented with the *Analyte/Landuse/Pathway* options again. This time select *Industrial* and *Ingestion* and press *OK*.

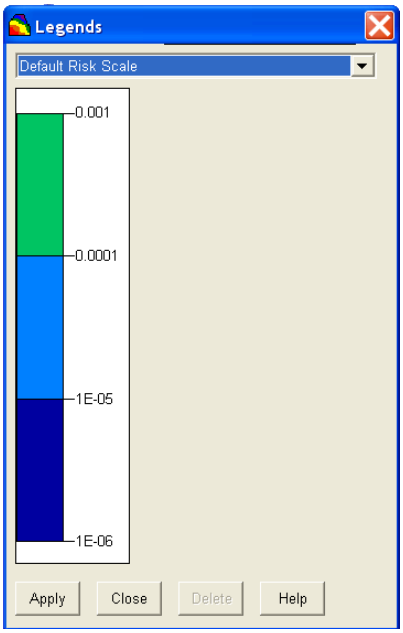


SADA will convert all concentration values to risk values for Ac-225.

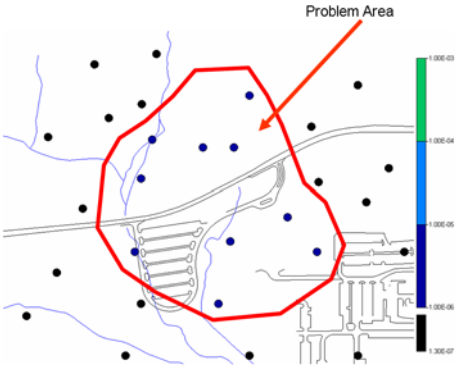


The distribution of colors will remain relatively the same as concentration plots. This is because, for the most part, risk models are linear in nature (except at the extremes). It can be difficult to determine from the continuous legend type where an exceedance of 1E-6, 1E-5, or 1E-4 might be found. To help us, we turn to the legend manager. From the main menu select *Graphics→Legend Manager*.

SADA comes with a predefined legend for risk called *Default Risk Scale*. This legend contains the traditional decision breaks for risk numbers (e.g. 1E-6, 1E-5, etc). If you need other risk values to be identified, you can create your own custom legend and modify it as required. If you need a custom legend, this is a good time to visit the sections on the legend manager in Chapter 43 .



Press the *Apply* button and then the *Close* button. The resulting map is dramatically changed, yet the numerical results have not changed. Notice that in the center of the map the same ten points that were highlighted with screening boxes are now colored dark blue. This indicates that the center part of the site may spatially be an important risk driver.

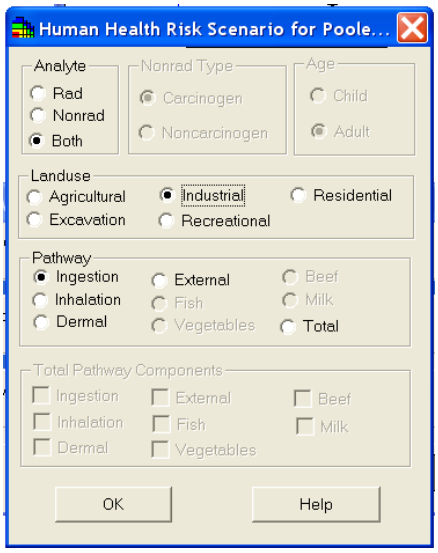


Pooled Point Risk Map

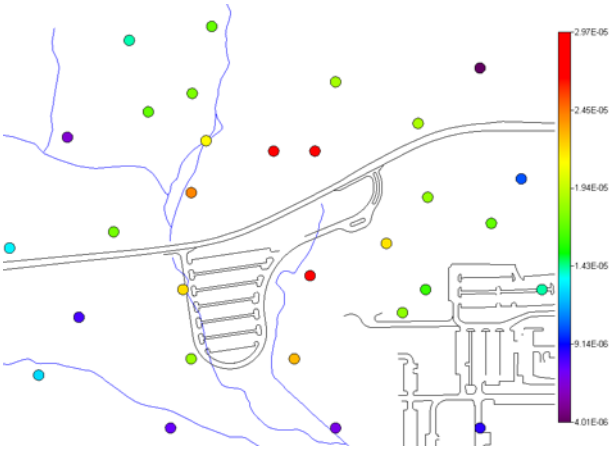
SADA can calculate the point risk at every location for each contaminant within the selected media and sum up the risk values. This shows the summed risk for each location and can also serve as an easy way to perform a cumulative screen.

In the data set drop-down list, select *Pooled Data* and press the *Show The Results* button.

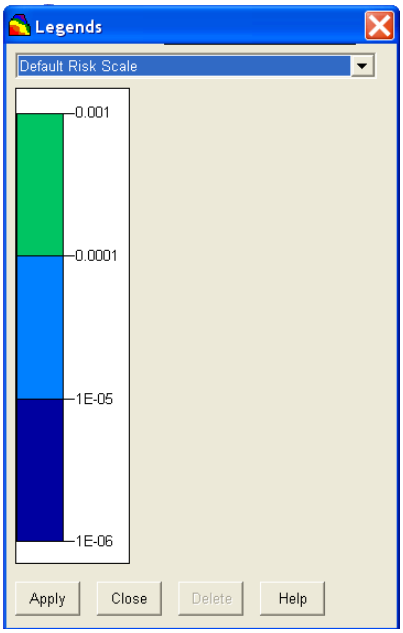
You will be presented with the risk options window again. This time many more options are available, as we have four very different contaminants involved.



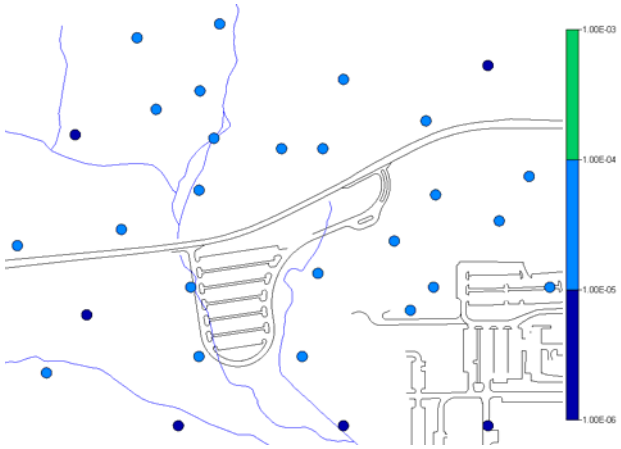
In this evaluation, we wish to map the carcinogenic point risk for both radionuclide and nonradionuclides alike. Select *Both* under *Analyte* and choose *Industrial* and *Ingestion*. Press *OK* and a cumulative point risk map is produced.



You may notice that the legend has reverted back to the continuous legend. This is because SADA stores legend selections by two identifiers: contaminant selection and picture type. Here, the picture type is still point risk map but the contaminant has changed from *Ac-225* to *Pooled Data*. This is easy to correct. Click on *Graphics→Legend Manager* and choose the *Default Risk* legend again.



Press *Apply* and *Close* the window.



When considered simultaneously, the complete set of soil contaminants demonstrate a higher cumulative risk than Ac-225 alone. In this mapping, we see that many points exceed a target risk level of 1E-6.

Contouring Risk

SADA is equipped with a reasonably comprehensive set of geospatial models that have a wide range of flexibility. Several later chapters (Chapters 28-31) are dedicated to the use of these models. Geospatial models perform many services, the most common of which is a contoured map. In SADA, you can contour the point risk values to create a continuous view of risk over space.

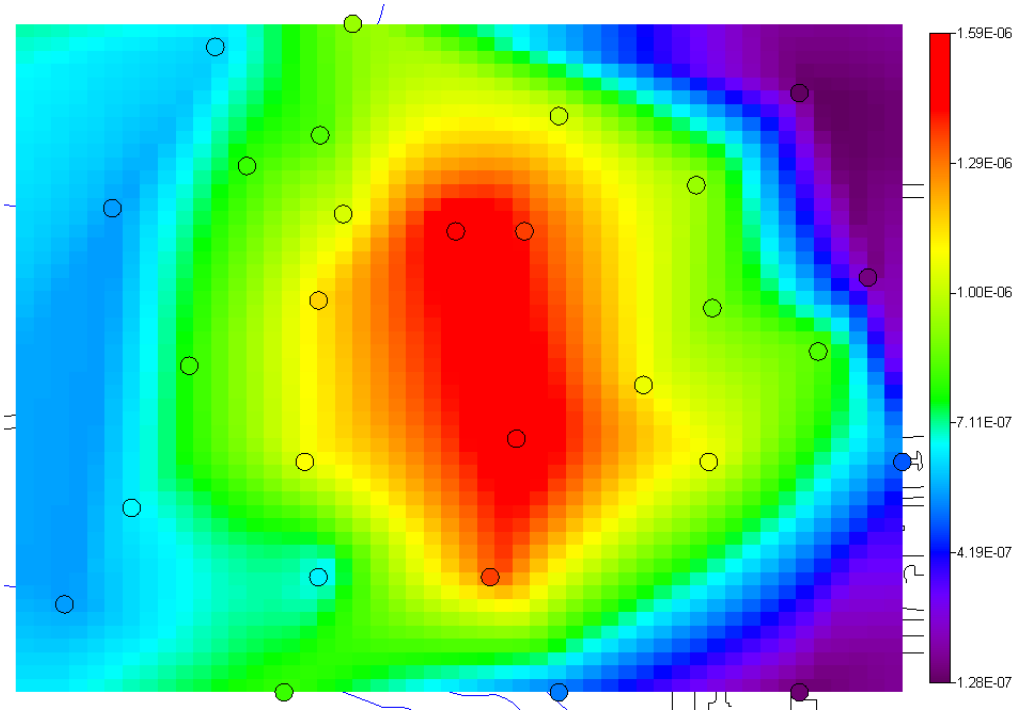
To demonstrate this, we will use the simplest of geospatial models: natural neighbor. This interpolant does not require any input from the user. So at this point in the guide, it is an ideal choice.

It is possible to produce cumulative risk maps as well. You must calibrate a geospatial model for each contaminant contributing to the cumulative risk map. In this example, we will only contour

the risk for *Ac-225* for *Industrial* and *Ingestion* exposure. From the data set drop-down list, select *Ac-225*. From the interview drop-down list select *Draw a contoured risk map*.

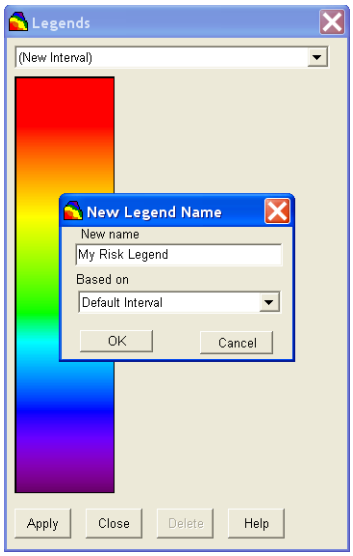
Click on the step Interpolation methods and choose *Natural Neighbor*.

Press the *Show The Results* button, choose the *Industrial* and *Ingestion* combination, and press *OK*. SADA presents the following result.

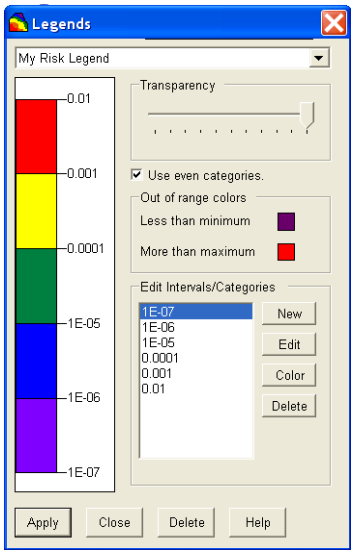


This is a contour of the point-wise risk over the entire site. Let's take a moment to improve the presentation of the map. From the main menu, select *Graphics→Legend Manager*. We will create a new legend for this map. If you are not familiar yet with the legend manager, you may wish to visit chapter 43.

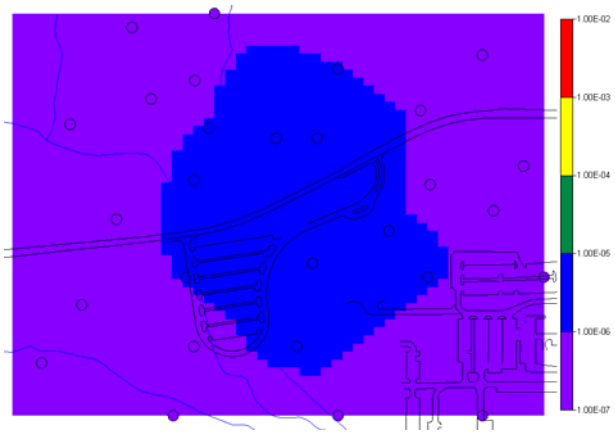
In the drop-down list, select (*New Interval)* and provide the name "My Risk Legend." Select *Default Interval* in the *Based on* drop-down. Press *OK*.



Using the editing features in the legend manager, create new intervals with breaks at 1E-7, 1E-6, 1E-5, 1E-4, 1E-3, and 1E-2. This will cause the scale in the lower right to bunch together considerably due to the order of magnitude differences between the values. Just below the *Transparency* block, check the box for *Use even categories*. Notice how the scale stretches out evenly now. Make the colors for the intervals purple, blue, green, yellow, and red as follows.



Press *Apply* and then *Close*. The map now demonstrates that all estimated values fall below 1E-5. Click on the step *Set GIS overlays*. Select *SADA Result* and use the down-arrow in that same window to move the *SADA Result* to the bottom. This will allow you to see the roads and water layers.

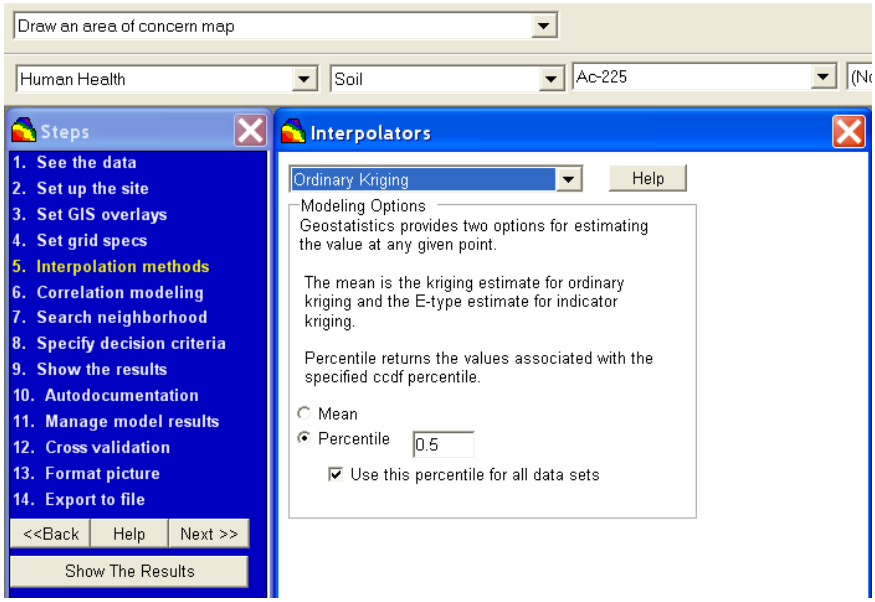


Using Risk in Decision Making

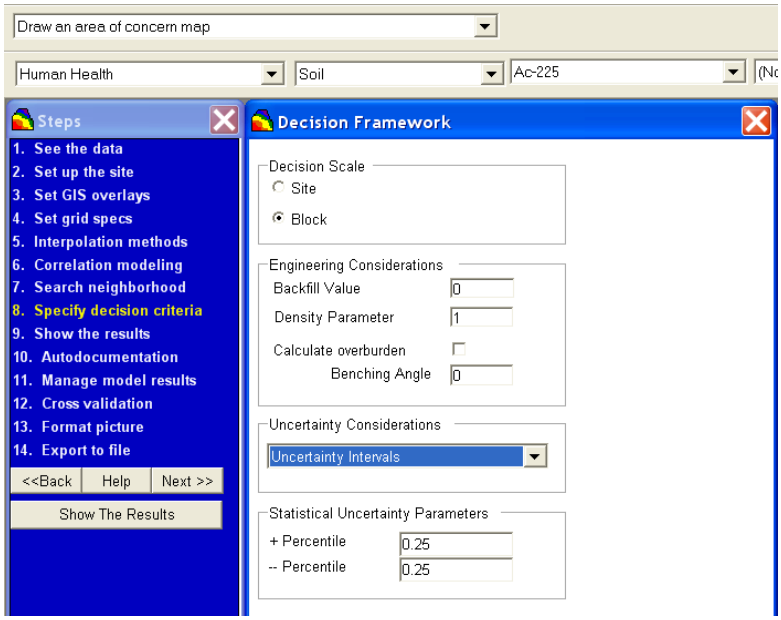
SADA provides a number of spatial decision analysis tools. Risk can play a role in this decision process. Later chapters are dedicated to various decision analysis tools that make use of human health, ecological, custom, and MARSSIM-like decision making features. We will present an example here to give a taste for the kinds of contributions risk can make in a spatial-decision framework.

One example of a decision result is the development of an area of concern (AOC). An area of concern is a spatial delineation of where concentration values may be too high and some action may be required. Risk assessors can estimate this area and also quantify the uncertainty about the location of the AOC boundary line.

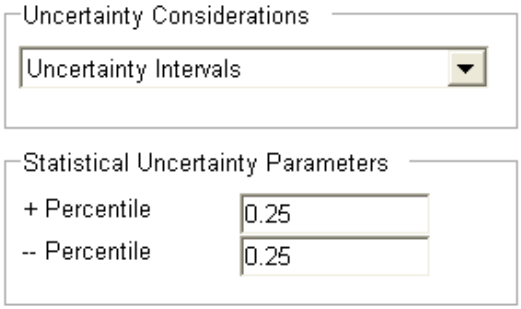
In the interview drop-down list, select *Draw an area of concern map* and click on the step *Interpolation methods*. From the drop list of available interpolators, choose *Ordinary Kriging*. This is a geostatistical model that allows one to quantify uncertainty in several ways. We will use this property of ordinary kriging to assist with quantifying the uncertainty about the AOC boundary line. The *Ordinary Kriging* model requires some calibration, and we have done that already for you. Make sure that your selections appear exactly as below.



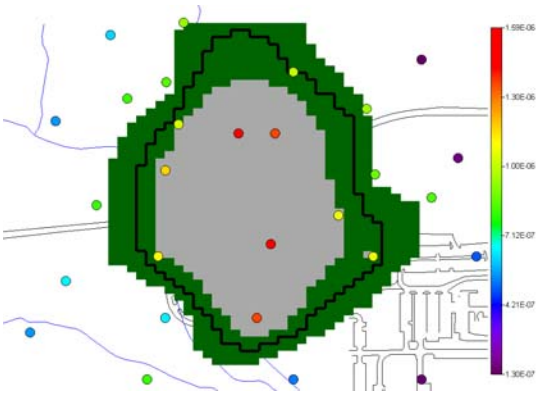
Click now on the step *Specify decision criteria*. There are a number of parameters here that deal with defining the area of concern. Focus your attention for now on the parameter block called *Uncertainty Considerations*. This section permits you to choose a method for understanding the AOC boundary uncertainty. In the drop-down list, select *Uncertainty Intervals*. Associated with this option are two parameters: +/- Percentile. For now, we'll accept the default values of 0.25 for each.



The important thing to understand at this point is that we are asking the code to specify the region where there is a 25%-75% (centered about 50%) point-wise probability that the true boundary line will occur.



Press the Show The Results button and select *Industrial* and *Ingestion* again. The following result appears.



In this map, we see several things. For a PRG value of roughly 3pCi/g, this is the area of concern. The thick black boundary line indicates the best estimate of the AOC boundary line. The green area represents the region where the actual boundary line could reasonably exist when considering our spatial uncertainties. The gray region is estimated to be very contaminated and will likely be within the AOC boundary, even when considering the uncertainty.

A more comprehensive discussion of this approach can be found in Chapter 35, Determining the Area of Concern Including Remedial Volume and Associated Uncertainty. Other methods of decision support, including secondary sampling designs, better isolate elevated samples or better refine the exact location of the AOC. Both of these are covered in later chapters.

Summary

SADA allows both tabular risk calculations and spatial risk modeling. The latter allows risk assessors to directly contribute to a decision process, such as developing an area of concern or locating another round of samples. For those circumstances where the log distribution or normal distribution are not adequate, users should import exposure unit concentrations into a separate SADA file, where they can still use the risk tools. Since spatial models rely on point-wise information, contouring and decision support tools, such as those presented here, can still be used on the actual geographically referenced data sets. If you do not have coordinates for your data values, you can fake the coordinates in order to get them into SADA. You can use them in the tabular risk features, but obviously you cannot use them in a spatial risk assessment.

Chapter 22: Overview of Ecological Risk Assessment

Ecological risk assessment is the process of gaining an understanding of the likelihood of adverse effects on ecological receptors occurring as a result of exposure to one or more stressors. It is often conducted in a multi-tiered/multi-stepped process, beginning with a preliminary conceptual model and site description that gathers existing information and forms the basis for a screening assessment. The purpose of the screening assessment is to focus on stressors most likely to be a concern (and eliminate those unlikely to be a concern from further consideration) and/or to focus on areas where receptors may experience unacceptable exposures. The results of the initial site description and screening are used to refine the conceptual model. If potential risks have been identified, the refined conceptual model points out the primary exposure pathways and any additional data needs necessary to evaluate potential risks. Data needs include information on the distribution and magnitude of stressors as well as effects of stressors on receptors. Ultimately, the information on exposures and effects are combined to characterize potential risks.

Generally, risk assessment involves a weight-of-evidence approach that takes into account results from multiple lines of evidence. For example, comparison of exposures to chemical effects data is one line of evidence used in evaluating potential risks to aquatic organisms in a stream; community survey data and site-specific surface water and sediment toxicity test results are additional lines of evidence that may or may not lead to the same conclusions indicated by the chemical toxicity data. The art of risk assessment is in the interpretation of different types of information to arrive at a final conclusion regarding conditions at the site.

Ecological risk assessment guidance from EPA includes an eight-step process. Step 1 of the process is a screening-level problem formulation that pulls together knowledge of the site history and ecology to develop a preliminary conceptual model. This preliminary conceptual model addresses the environmental setting and stressors known or suspected to exist at the site, contaminant fate and transport mechanisms, mechanisms of ecotoxicity and likely categories of receptors that could be affected, presence of complete exposure pathways, initial endpoints to screen for ecological risks, and conservative screening ecotoxicity values.

Step 2 of the process is a screening-level exposure estimate and risk calculation. Maximum exposure concentrations are compared to ecotoxicity screening values to help determine if risks are negligible or if a more detailed assessment is warranted. Ideally, this step eliminates at least some sites, contaminants, exposure pathways, or receptors from further evaluation.

Scientific/management decision points (SMDP) are built into the process beginning with the end of step 2. These are points where communication/discussion among the risk assessor, risk managers, and stakeholders takes place to decide whether the information available is adequate to make a risk management decision. Open communication between the risk assessment/risk management team and regulators/stakeholders smoothes the decision-making process.

Step 3 of the process is the baseline risk assessment problem formulation. This step uses the results of the screening assessment and additional site-specific information to refine the screening-level problem formulation and determine the scope and goals of the baseline assessment. It is only applied in situations where the screening-level assessment indicated a need for further evaluation. In this step, assessment endpoints are selected that result in a conceptual model that includes questions the site-investigation will address.

Step 4 of the process works from the conceptual model developed in step 3 to produce the study design and data quality objectives that will result in the information needed for the risk manager to incorporate ecological considerations into the site remedial process. Measurement endpoints are identified, and any additional data needs are addressed in the work plan.

Step 5 is field verification of the proposed sampling design, which confirms that the data needed can actually be collected (i.e., fish of the appropriate size are present and can be caught).

Steps 6 and 7 are the site investigation and analysis phase and the risk characterization phase. Step 6 specifies the assessment endpoints, risk questions, and testable hypotheses and characterizes exposures and ecological effects (the exposure assessment and effects assessment steps from previous guidance). Step 7 is the final phase of the risk assessment process and combines the exposure and effects assessments to characterize the likelihood of adverse effects on assessment endpoints. This is generally a weight-of-evidence approach, where multiple types of information on possible effects are considered to arrive at an overall conclusion.

Step 8 is risk management. In risk management, the results of the risk assessment (Steps 1-7) are integrated with other considerations to make and justify risk management decisions.

SADA provides a means of implementing components of the EPA process, including the screening assessment, development of a sampling plan, and exposure modeling for the baseline assessment. Just as importantly, SADA provides the user with a way to visualize results across a site, plotting data, exposure values, and risk numbers relative to features of the site and facilitating use of geospatial methods where applicable.

Ecological Capabilities in SADA

SADA was designed to address hazard assessments at environmental sites. It implements EPA methods for conducting ecological risk assessments. Chemical hazard assessment, or comparison of an environmental concentration to an estimated toxic threshold for a particular contaminant, is the most common method for examining effects of chemicals in the environment. SADA provides a media-based benchmark database for contaminant effects on ecological receptors, exposure models for 19 terrestrial species (birds and mammals), default exposure parameters for risk models (where available), tabular screening and risk results, and risk and dose mapping.

The primary ecological risk functions of SADA include:

- Setting Up Ecological Risk
- Ecological Risk Assessment Procedure
- Setting Physical Parameters
- Description of Ecological Benchmark Database
- Histograms of Benchmark Values
- Tables of Benchmark Values
- Setting Screening and Exposure Statistics
- Area Result Tables (Screens, Ratios)
- Map Result Values (Screens, Ratios)
- Rematching a Single Contaminant

- Checking Ecological Version
- Terrestrial Dose Modeling

Chapters 23-27 describe how to use these functions and assume you have already created a SADA file and imported data into it.

Chapter 23: Preparing for Ecological Risk

This chapter deals with preparing your data for import into SADA with the objective of conducting an ecological risk analysis. This chapter follows the same course of discussion as the chapter 18, Preparing for Human Health Risk, as preparation stages are quiet similar. Both will draw from the content of Chapter 5, which covered how to import data into SADA, and that same procedure will be used here. We will focus a little more on those fields required for ecological risk and on some strategies you might use, depending on your situation. SADA can import various types of data, but point data is the only kind SADA can directly use in assessing risk.

What if I don't have any coordinates? What if I want to use exposure values calculated elsewhere? What if I have contoured data?

It is possible to calculate risk on modeled or gridded data types and even on data sets without spatial coordinates. In some cases, users may want to use exposure concentrations other than the four basic options in SADA (max detect, mean, UCL95, and min of max detect or UCL95). This is also possible. Those special cases are covered a little later. For right now, we will be working with measured point values for a set of contaminants in the soil. In each of these cases, you will carry out your analysis in much the same way as you would with measured values at each point. You should continue reading the next section so you understand later what to do in your special circumstance.

Working with Sampled Data with Coordinates

Recall from Chapter 5 that before SADA can import the data values, it needs the data in a particular format. Fortunately, the format is very easy to work with, and much of the time environmental data comes in this type of configuration. Simply put, data needs to be in a tabular format with each row representing a single measurement value of a single contaminant (or attribute). Each column is information about the measurement, such as the easting coordinate, the northing coordinate, depth, name of the contaminant, measured value at that location, and so forth. Many times users look at this data with a spreadsheet program like Excel. Here is an example data set in Excel.

	A	B	C	D	E	F	G	H	I	J	K	L
1	Easting	Northing	Depth	Casnumber	Name	Value	Detected	Media	Date	SampleID	Region	
2	27596.25	21900	0	14265851	Ac-225	2	1	SO	9/26/1970	1	1	
3	28310.25	21900	0	14265851	Ac-225	1.6	1	SO	9/26/1970	2	1	
4	28935	21900	0	14265851	Ac-225	0.9	1	SO	9/26/1970	3	1	
5	27685.5	22200	0	14265851	Ac-225	2	1	SO	9/26/1970	4	1	
6	28131.75	22200	0	14265851	Ac-225	4.2	1	SO	9/26/1970	5	1	
7	29202.75	22500	0	14265851	Ac-225	1.5	1	SO	9/26/1970	6	1	
8	27150	23160	0	14265851	Ac-225	1.7	1	SO	9/26/1970	7	1	
9	27685.5	22920	0	14265851	Ac-225	3.6	1	SO	9/26/1970	8	1	
10	28042.5	23100	0	14265851	Ac-225	4.9	1	SO	9/26/1970	9	1	
11	28221	23100	0	14265851	Ac-225	4.2	1	SO	9/26/1970	10	1	
12	28667.25	23220	0	14265851	Ac-225	2.9	1	SO	9/26/1970	11	1	
13	29113.5	22980	0	14265851	Ac-225	1	1	SO	9/26/1970	12	1	
14	27417.75	23580	0	14265851	Ac-225	1.9	1	SO	9/26/1970	13	1	
15	27774.75	23640	0	14265851	Ac-225	2.9	1	SO	9/26/1970	14	1	
16	28310.25	23400	0	14265851	Ac-225	3.1	1	SO	3/22/1993	15	1	
17	28935	23460	0	14265851	Ac-225	0.8	1	SO	3/22/1993	16	1	
18	28200	22560	0	14265851	Ac-225	4.8	1	SO	3/22/1993	17	1	
19	28700	22500	0	14265851	Ac-225	3.3	1	SO	3/22/1993	18	1	
20	27200	22380	0	14265851	Ac-225	2.03	1	SO	3/22/1993	19	1	
21	28984	22787	0	14265851	Ac-225	2.6	1	SO	3/22/1993	4	1	
22	27350	22750	0	14265851	Ac-225	2.5	1	SO	3/22/1993	3	1	
23	27026	22129	0	14265851	Ac-225	1.7	1	SO	3/22/1993	2	1	
24	27690	23350	0	14265851	Ac-225	2.6	1	SO	3/22/1993	4	2	
25	27500	23270	0	14265851	Ac-225	2.5	1	SO	3/22/1993	44f	2	
26	28709	22900	0	14265851	Ac-225	2.7	1	SO	3/22/1993	43	2	
27	27650	22500	0	14265851	Ac-225	3.4	1	SO	3/22/1993	33	2	
28	28530	22700	0	14265851	Ac-225	3.3	1	SO	3/22/1993	11	2	
29	27750	23145	0	14265851	Ac-225	3.2	1	SO	3/22/1993	12	2	
30	27596.25	21900	0	7440393	Barium	42.7	1	SO	8/7/1996	13	2	
31	28310.25	21900	0	7440393	Barium	35.1	1	SO	8/7/1996	14	2	
32	28935	21900	0	7440393	Barium	18.4	1	SO	8/7/1996	15	2	
33	27685.5	22200	0	7440393	Barium	43.5	1	SO	8/7/1996	16	2	
34	28131.75	22200	0	7440393	Barium	87.9	1	SO	8/7/1996	17	2	

SADA can import measured values using two different kinds of files. Comma Separated Value or .csv files are the most commonly used file type. Anyone with a copy of Excel or a similar spreadsheet program can easily save their file as a .csv file. Even if you don't have a spreadsheet program, you can use simple text editors (e.g., Notepad that comes with Windows) to create or format your data set. If you are using a .csv file, the first row must be the title row. Punctuation is not allowed in the column (field) names. The second method to import measured values is to use a Microsoft Access database. Let's examine each of the types of information now.

Easting and Northing Columns (required)

SADA must have these two columns. They tell where on the site your sample was collected. SADA does not require or enforce the use of any predefined coordinate system. There are a number of them available, such as UTMs, state plane, lat/long and so forth. Many sites have their own localized system. In some cases, you'll see sample measurements relative to the corner of the site or a building. These are all OK. (Note to lat/long users: you'll need to make sure you use the negative of your latitude values or your data will plot backwards.)

What is important to note here is that these values must be in the same coordinate system as any GIS layer you plan on using. For example, the roads.dxf layer seen throughout these examples is in State Plane Coordinates for Tennessee. You would need to make sure that the easting and northing coordinates are using the same system, or there will be big problems.

Namely, your data will not plot correctly on your GIS layer. If you are not using GIS, then you have no worries with your coordinate system.

Blank values and non-numeric values are not permitted in either of these columns.

Depth Column (may or may not be required)

If the data are taken over depth, an additional column with the depth value is required. No empty values are allowed for any of these columns, and non-numerical values are not permitted for any coordinate or sample value. There are two important things to check about your depth column. First, make sure your depth is *depth below surface* and NOT elevation measurements. SADA operates on the concept of depth below surface. Elevation is used in SADA, but only as a 3d view feature. Second, make sure that your depth values are in the same units as your horizontal extents. So for example, it is nonsensical to use lat/long for horizontal and meters for depth. Later on, when you do 3d geospatial analysis, this really has bizarre implications. If the horizontal measurements are in feet, make sure the depth is in feet and so forth. SADA is not going to do any conversion for you.

Sampled Values (required)

This field is where your measurement value is found. There are no units enforced by SADA, because of the many and varied SADA applications. You will need to make sure that if you are doing ecological risk, you MUST have your measured units as follows.

- Soil, Sediment: mg/kg for nonradionuclides (no rad analysis currently available)
- Surface: mg/L for nonradionuclides, (no rad analysis currently available, no groundwater used)

The sample values column cannot contain any blanks or non-numeric values. If a measurement value was a nondetect, you must enter the detection limit into the value column. For example, you cannot enter <5. You would instead enter a value of 5 and then not enter this in the detection field (see below). Values like NA are obviously not allowed either.

Name (required)

This column provides the name of the attribute (contaminant) you are measuring in each record. The field cannot contain any blank values. When you are working with environmental contaminants, it is common to work with contaminant names that contain commas. These are a serious problem for comma delimited files, as you might imagine. For example, the contaminant name 1,2,3-Trichlorobenzene would be recognized as 3 separate columns: 1, 2, and finally 3-Trichlorobenzene. This will cause an error in import. Typically, this error states that SADA has reached the end of the file unexpectedly. This can be the cause of such an error. To correct this, make sure your name field contains the name within quotation marks: “1,2,3-Trichlorobenzene.” Spreadsheet programs like Excel normally do a good job of taking care of this for you, by placing quotations around the contaminant name; however, you may want to open up your .csv file in a simple text editor, such as Notepad (shipped with Windows), to make sure that you are not running into this problem. Typically, SADA will complain about reaching the end of the file too soon in this situation.

It is also important to make sure that each contaminant is consistently named. For example, make sure you don’t have 1,2,3-Trichlorobenzene also spelled as Trichlorobenzene-1,2,3. SADA may see these as two different contaminants.

CAS Numbers (not required)

A column containing the CAS numbers is very useful when you are setting up the ecological risk module. It isn't required, though. If you do have CAS numbers and want to use them, make sure that they do not contain any dashes or spaces. This will render them virtually useless as SADA tries to match them up to its toxicological profiles database (discussed shortly). A simple search and replace in the CAS column using Excel can take care of this pretty quick.

Detection Qualifiers (not required)

Another field of some importance is the detection field. Instead of the normal Us and Js and so forth, SADA works with either a zero value (sample was below detection limit) or a 1 (sample was detected). As the user, you must decide whether a value should be considered detected or not before bringing it into SADA. This involves an area of the environmental process we are not involved in; however, we do provide some guidance for risk assessment in the tables below if you have no other instructions.

Flag	Meaning	Use for risk?
R	Rejected	No
B	Blanks contaminated	Treat as non-detect
J	Estimated	Yes, treat as detect
UJ	Estimated non-detect	Yes, treat as non-detect
K	Biased high	Yes, treat as detect
L	Biased low	Yes, treat as detect
U	Non-detect	Yes, treat as non-detect

Now for measured values that are considered non-detects, you should enter the detection limit of the instrument into the column containing measured values. Blank or missing values in the value column are not permitted.

Date Column (not required)

The date field is not required. When used, however, you can query your data by date to see how things are evolving over time. This can be useful in monitoring situations. If you do use the date field, you must use some combination of the m, mm, d, dd, yy, yyyy format. Examples include m/d/yy and mm/dd/yyyy. Dates that include any text (other than /) are not permitted. For example, September 26th, 2004 is not permitted. You would need to use 9/26/2004. In Europe and other areas that use day/month/year format, this should be OK as long as your data set and your operating system are consistent.

Media Column (required)

For ecological risk, you MUST have a media column. Proper media identification qualifiers are as follows: Soil – SO, Sediment – SD, Groundwater – GW, Surface water – SW, Air – AIR, Biota – BIO, and Background – Background.

Beyond these columns, you can include any other information you like, but the total number of columns may not exceed 250.

Final Tips

Sometimes the first attempt to import data fails. Chapter 5 covers some of the most common reasons this may occur. Please refer to this chapter now and make sure these conditions do not exist in your dataset before importing the data.

Re-enforcement

Create a SADA file to use for ecological risk analysis. Refer to Chapter 5 if you need help. Do the following:

8) Create a new SADA file.
9) Import the data file TwoDimensionalWithNonDetects_Duplicates.csv.
10) Match
    a. Name to Name
    b. Value to Value
    c. Easting to Easting
    d. Northing to Northing
    e. Depth to Depth
    f. CAS Num to Casnumber
    g. Detect Qualifier to Detected
    h. Media Id to Media
    i. Date to Date
11) Note: This file has both duplicates and non-detect values.
12) Snap the site boundaries to *All data sets*.
13) Import the GIS layers roads.dxf and water.dxf. Make the water layer blue (Chapter 4).
14) Save as EcologicalRisk.sda

Preparing for Ecological risk

If you have not already done so, perform the re-enforcement exercise above and bring up EcologicalRisk.sda. We now have our data set and need to prepare it for conducting an ecological risk. The first thing we need to do is decide how to deal with duplicates and non-detects.

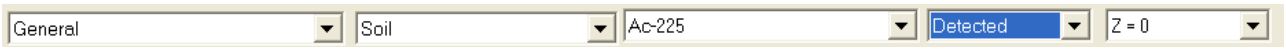
Dealing with Non-detects

Recall that, in your data set, if a detection field is specified during import, SADA will assume that a value of 0 means the sample is a non-detect and a value of 1 means the sample is detected. If you did not specify a detection field, SADA assumes all data are detected. For non-detects, you have three options: use full detection limit, use half the detection limit, or use zero. If a sample is designated as a detect (1), then the result in the value field is considered the actual measured value as you would expect. If a sample is designated as a non-detect (0), the result in the value field is considered the detection limit.

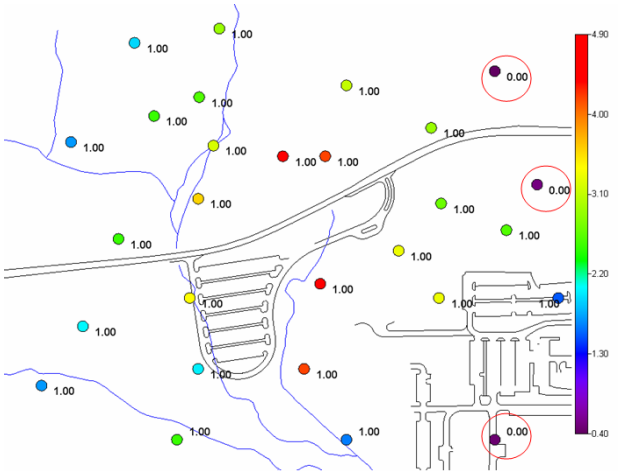
On virtually any interview for sample data, you will find *See the data* in the steps window. In your current file, select *Plot my data* from the interview list and click on *See the data*. In the associated parameters window, find the parameter block called *Non-Detects*. Within this block are the three available options. The default is *half the detection limit*. This file has both

duplicates and non-detect values. Let's use the label field to better understand what SADA is doing.

If you have not already done so, check that you have *Soil* and *Ac-225* selected as your media and data set. Select *Detected* as your label field (not *Plotted Value*). Ac-225 is actually a radionuclide and there are currently no rad analyses available for ecological risk. We use this dataset, however, to demonstrate how non-detects and duplicates are treated.

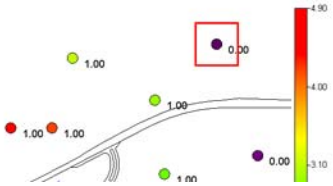


For Ac-225, you should see the following result (without the red circles).



Three red circles are added to the result on the right-hand side to indicate the location of three non-detected values for Ac-225. You can readily identify the non-detects because the detection field will have a zero value. If the data set is geographically dense, it may be difficult to visually search through the labels for your non-detect. Fortunately, a label search feature is available. From the main toolbar, select *Tools→Search Labels* and the following window will appear:

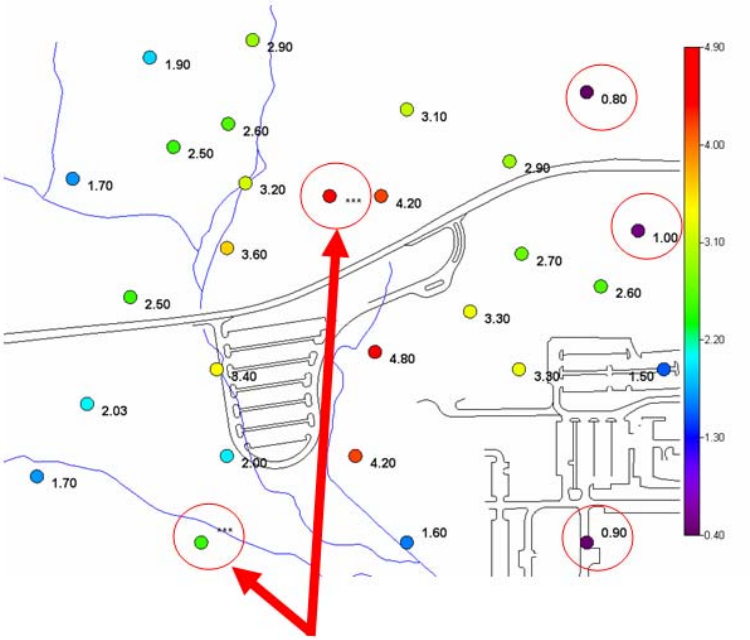
Enter a value of 0.00 in the text box and press *Find*. Note that you must be searching for EXACTLY the right string. For example, searching for 0 would result in no finds. SADA will enclose the first point it finds with this label in a large red box. Each time you press the *Find* button, SADA advances to the next string it finds. In this manner, you can quickly search the current layer for all labels with a certain text.



Press *Clear ID Circ* and then close the window by pressing the *X* button in the upper-right-hand corner.

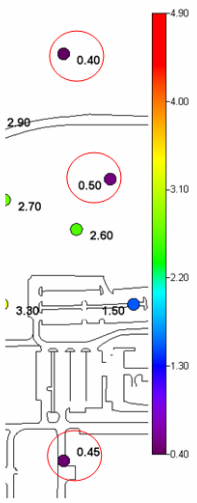
This is a good opportunity to demonstrate the difference between the *Value Field* and the *Plotted Value* in the labels box. Recall that the *Value Field* will return the *Reported Value* in the value column of your dataset. The *Plotted Value* option will label the point with the final result that SADA uses in the map. For detected values, these two values may vary. The result in the value field will be the detection limit for non-detected values. The *Plotted Result* however, will be half the detection limit (the default option we are currently using). Select *Value* in the labels box.

For *Ac-225*, you should see the following result (without arrows and circles). Each label is showing the actual value found in the data set for that sample point.



*** means duplicate

There are five sample points to notice here, each contained in a red circle. The two circles on the left side show a trio of asterisks instead of a value. This indicates that these two points each have duplicate values. That is, more than one record exists in the field for each coordinate location for Ac-225. Therefore, SADA cannot produce a single label and instead draws asterisks to indicate this situation. This will be discussed shortly. Notice for now the values associated with the non-detect values (0.8, 1.0, and 0.9). Now switch from *Value* to *Plotted Value*.



These values are the resolved values that SADA will use for all operations, including risk assessment.

Switch the non-detect option from half to the full detection limit. Notice how the labels change to their full value. Switch the non-detect option from full detection to zero. The non-detected values are all zero now.

Make sure you have half the detection limit selected before continuing.

Dealing with Duplicates

Returning now to the issue of duplicates, click on the *See the data* step and look for the information block called *Duplicate Data*. There are several options for resolving situations where you have duplicate values at one or more points. It is important to clearly define what is meant by a duplicate. The term duplicate has different meanings in different contexts. We use it in a very narrowly defined way.

If two or more points share the following properties they are considered duplicate values. In other words:

- Each point shares exactly the same coordinate (x,y,z).
- Each result is a measurement of the same contaminant.
- Each result was sampled within the current date interval.

The first condition seems obvious, but it is not necessarily well understood. In some cases, true duplicate values (measurements of Ac-225 at the same location) will be reported with slightly different easting, northing, or depth measurements. You might intend that they be considered duplicates, but if they do not have exactly the same coordinates, they will be considered separate measurements.

The second condition is fairly straight forward. A measurement of Ac-225 at location (0,0,0) and a measurement of Barium at location (0,0,0) is never considered a duplicate value.

The third condition is very important. Most of the time, users do not use the *Data Query* feature located just above the duplicates parameters block. The impact of this query on how SADA deals with duplicates, however, is significant. Consider the following situation where the two points were duplicate values in real life. Note that the coordinates are exactly the same. Only the date differs by one day.

Attribute	Sample #1	Sample #2
Easting	100	100
Northing	200	200
Depth	5	5
Name	Ac-225	Ac-225
Value	33.5	38.2
Detected	1	1
Date Sampled	5/1/2008	5/2/2008

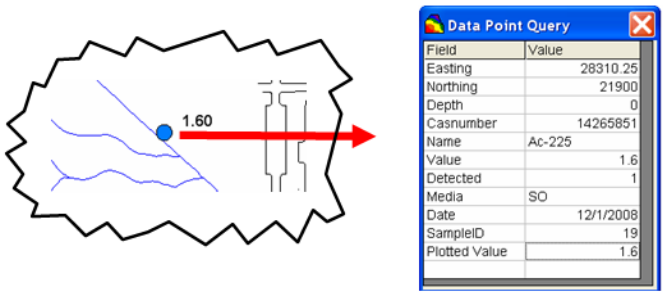
Here is how SADA would treat the duplicate values as a function of the time interval.

Date Interval	Duplicate Values?	Which one to use?
5/1/2008 – 6/1/2008	YES	Depends on the rule you choose.
5/1/2008-5/1/2008	NO	Only Sample #1 is used in the analysis. Since Sample #2 falls outside the date range, it is not considered in the analysis.
6/1/2008-7/1/2008	Doesn't matter	Neither is used, because neither falls in side the date range.

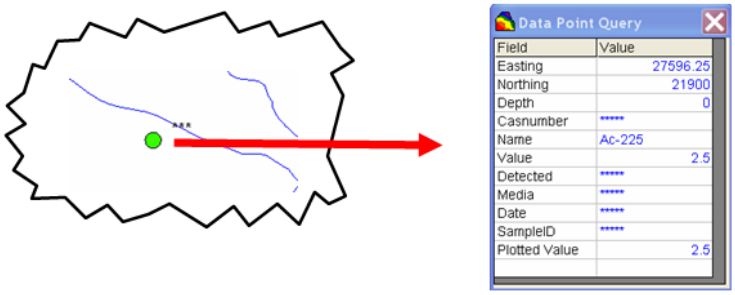
For monitoring data, there will be a lot of samples measured repeatedly at the same well. In doing a risk assessment, be aware that if you do not break up the date range into the sampling intervals (e.g., quarterly well measurements), SADA will assume that each well has a lot of duplicates, which is not truly the case. For monitoring situations, simply create multiple date ranges and perform a risk assessment on them individually.

An easy way to explore duplicate data is to use the label feature to label points with values from your *Value* field. This will produce the asterisks trio we saw earlier. Change the label box to *Value* now.

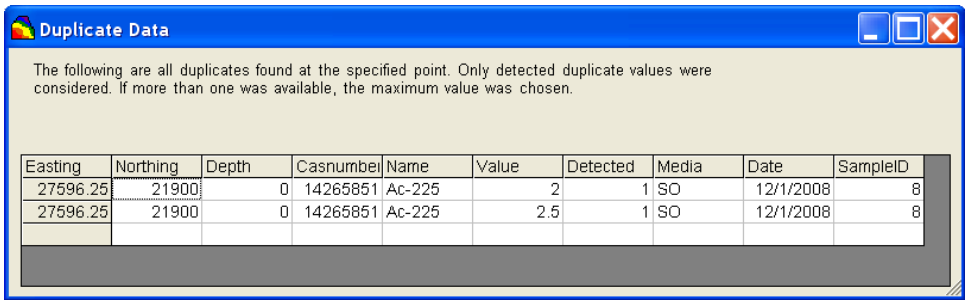
First, click on the data point with value 1.60 on the south side of the site in about the middle. The point information box will appear and show the results for this point taken directly from the sample record set.



Now click on the neighbor to the left of this point with the three asterisks.



Notice that the data point query box is populated with blue text and several fields contain asterisks. This indicates that because more than one Ac-225 value was found at this location, SADA can't show one single record. The "resolved" value of 2.5 is available, though, in the bottom line next to *Plotted Value*. By resolved, we mean that the rules for duplicates were applied and this was the final resolution. In a moment, we'll cover the rules for dealing with duplicates. Right now, click anywhere on the blue text and SADA will pull up those records identified as duplicate.

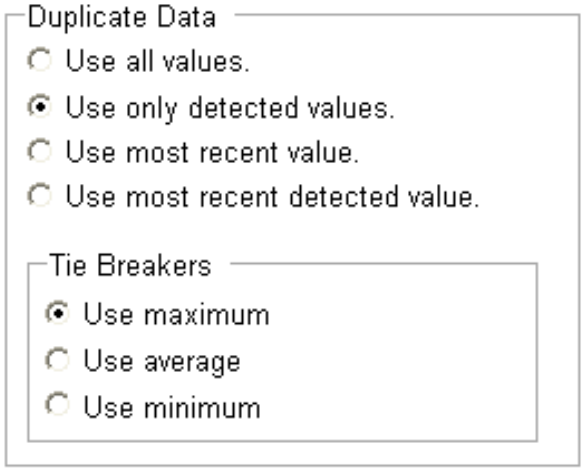


Easting	Northing	Depth	Casnumber	Name	Value	Detected	Media	Date	SampleID
27596.25	21900	0	14265851	Ac-225	2	1	SO	12/1/2008	8
27596.25	21900	0	14265851	Ac-225	2.5	1	SO	12/1/2008	8

So, two Ac-225 samples were collected at this one point with two distinct measurement values of 2 and 2.5 pCi/g. In this case, the duplicate rule was to use the maximum detected value (2.5). Close this window and close the *Data Point Query* window.

Duplicate resolution rules are two-tiered. First, you must choose from a list of available data characteristics, such as *Use only detected value* or *Use all value*. This selection alone may pare down the number of duplicates you may have at any given point. In the event that two or more points meet this criteria (as happened above), however, a tiebreaker rule must be selected to decide or resolve the remaining points.

The duplicate rules are seen in the following image:



IMPORTANT! Keep in mind that some choices for resolving duplicates may entirely eliminate the point altogether. Suppose for example, there are two undetected measurements of Ac-225 at point (x,y,z). If you have chosen *Use only detected values*, both points will fail and no sample will be used at (x,y,z).

If you have a large dataset (1000s of points) and most are duplicates, it will really slow SADA down. If you are not using SADA as a monitoring tool and the slow down is impeding your progress, you should deal with duplicates before bringing them into SADA.

Once you have explored the role of non-detects and duplicates in your dataset, you are ready to setup the ecological risk model.

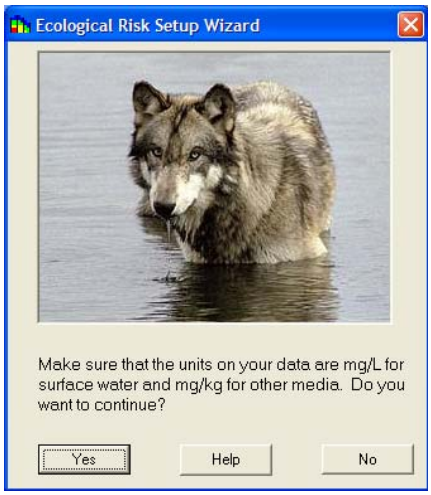
Remember that SADA is a 3d tool that can also work with 2d data. If your data set is 3d, you may have points below or above the particular layer you've selected. As a result, you may miss non-detects and duplicates on other layers as you explore your data. One option is to simply use the *Entire Data Layer* option. This option creates one single think layer that covers the entire vertical extent of your data. The problem is that it might cause an untidy clustering of labels, making it difficult to see what is going on. *Tools→Search Labels* can help. Another option is to make sure you select relatively thin layers, so that every label can be seen, but know that you must visit each layer to see all the data.

Setting up the Ecological risk Model

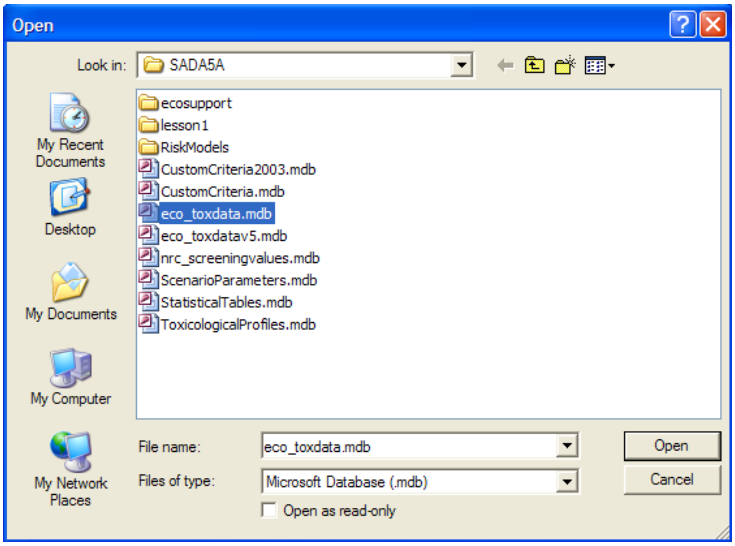
In order to use the ecological risk functions of SADA, you first have to set up the ecological risk component. The set up process identifies the source benchmarks database, matches contaminants in the site database to those in the benchmarks database, and adds ecological information to the SADA file.

To begin, select *Setup→Ecological Risk* from the drop-down menus.

Click *Yes* on the *Risk Setup Wizard*.

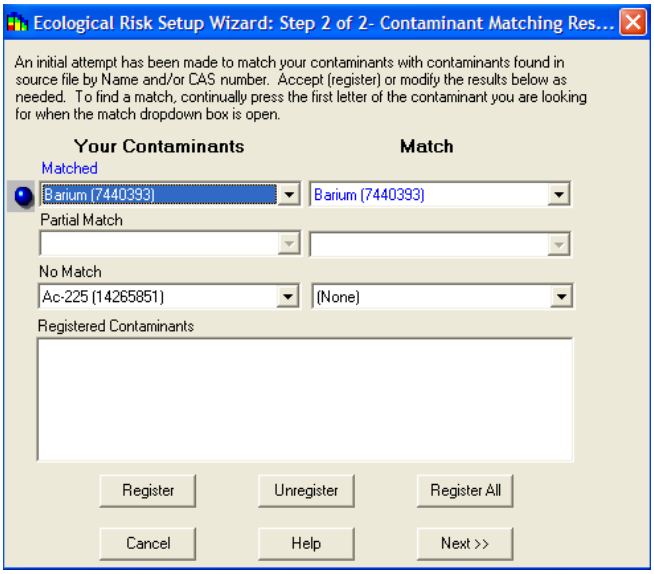


Browse for the file that contains the eco-toxicity benchmarks to be used for your site. The database provided with SADA is eco_toxdata.mdb. Select it, and click *Open>>*



Then click *Next* at the *Identify Source Databases* window. This sets the selected file as the source of benchmarks for your SADA file. Units for benchmarks in the benchmark database are mg/L for surface water and mg/kg for sediment and soil. It is imperative that units in your data match those in the benchmark database if you plan to use the benchmark screening capabilities of SADA.

SADA now needs to match the names of contaminants in your data file (left column) to those in the toxicity database (right column). You control this in the *Contaminant Matching* window. SADA attempts to match contaminants based on name and CAS number. If both match, SADA lists the contaminant as matched. If one or the other but not both match, SADA lists the contaminant as a *Partial Match*. If neither name nor CAS number match, SADA lists the contaminant in the *No Match* category.



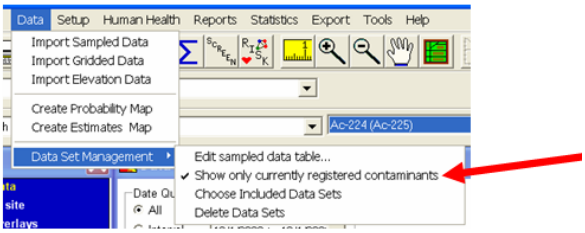
You must go through each category and register (or unregister) each contaminant. Generally the *Matched* category is good to go, and you can select the matched box and click on the *Register All* button to accept the matches. It is advisable to peruse the *Partial Matches* to be sure analytes in your database are matched with the appropriate contaminant in the benchmark database. If all are acceptable, you can *Register All* again, or you can use the *Register* button to accept the matches one at a time. For contaminants with no match, you need to select an analyte from your database (left column), then search the contaminant dropdown list in the right hand *Match* column to see if the benchmark database has a corresponding contaminant, perhaps with a different CAS number and different spelling. If it does, select it and click *Register*. In some cases, you might want to use a surrogate chemical to match with the one in your database.

After you have accepted matches for all analytes that you can (there are no matches for radionuclides because there are no eco-toxicity values for individual radionuclides in the benchmark database), click *Next* to save the matches to the SADA file. Unmatched analytes are not removed from your database, but they are unavailable for use on the ecological risk side of SADA.

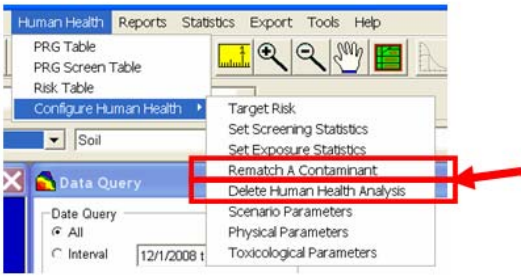
Once the module is complete, *Ecological Risk* analysis will appear in the analysis drop-down list on the second toolbar. Select this now. When this option is selected, an *Ecological Risk* menu will be visible on the main menu bar.

You may reset the Risk module at any time; simply select *Ecological* risk under the *Setup* menu of the main window. The process is the same as before; however, SADA will give you the opportunity to reset or skip the toxicological and scenario parameter component identification.

Important Tip! If a contaminant is not registered during setup, it may no longer appear in the list of contaminants if ecological risk is selected for the analysis. You should be aware of this convenience issue. If you want to see all contaminants when you select *Ecological* risk, whether they are registered or not, you can set this under *Data Set Management*. Simply check or uncheck *Show only currently registered contaminants* in the drop-down list.

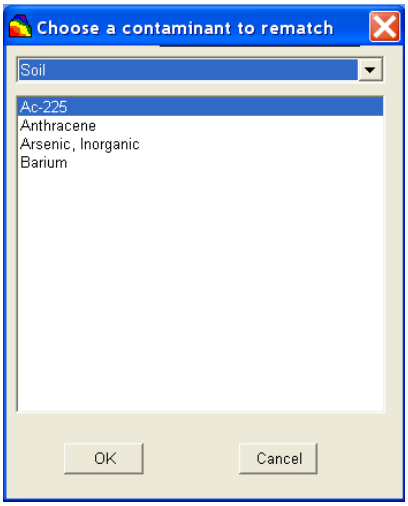


From the main menu, select *Ecological risk→Configure Ecological risk*. These menu items allow you access to calibrate your risk models, view tabular risk results, and, if necessary, rematch particular contaminants or delete the ecological risk altogether.



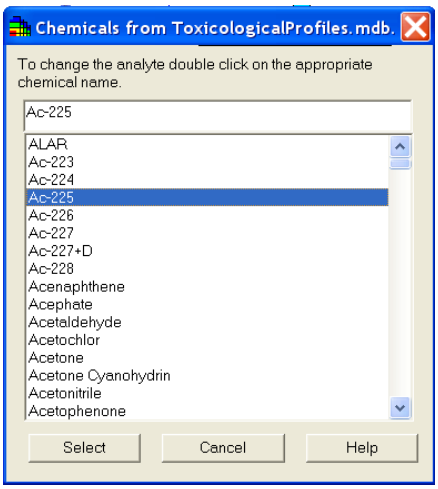
Rematching a Single Contaminant

Save your SADA file now. Select *Rematch a Contaminant* from the drop-down list. In the next window, contaminants are organized by data type (media). Select *Ac-225* as the contaminant you wish to rematch and press *OK.*

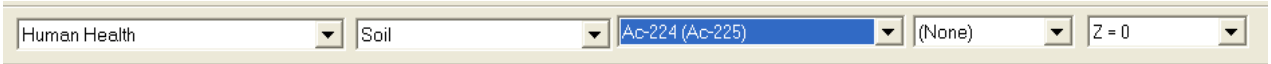


SADA will present an open dialogue box. Navigate to where the ToxicologicalProfiles.mdb file is located and select it. Press *Open*.

You are now shown the complete list of contaminants found in this database. Notice that SADA has pre-selected *Ac-225* for you.



For the sake of experiencing this feature, select *Ac-224* instead of *Ac-225* and press *Select*. The window disappears. Select *Soil* and then *Ac-224* from the drop-down lists.



Notice now that SADA reports the original name in parentheses. This allows you to see how the data was matched during setup. If the names are identical, you won't see the parentheses.

Re-enforcement

Repeat the single contaminant match above and rematch Ac-225 to Ac-225 in the ToxicologicalProfiles.mdb. Save your file.

Deleting the Ecological Risk Module

You can completely delete the ecological risk model by selecting *Ecological risk→Configure Ecological Risk→Delete Ecological Risk Analysis*. When prompted to confirm, answer *No*.

Working with non-spatial data, contoured data, and using exposure concentrations calculated outside SADA.

It is possible to calculate risk on modeled or gridded data types and even on data sets without spatial coordinates. In some cases, users may want to use exposure concentrations other than the four basic options in SADA: *Maximum detect, Mean, UCL95, or Minimum of max detect or UCL95.*

Contoured Data

SADA certainly permits users to import raster or contoured data "as is" without any additional modification (see Chapter 7, "Importing Modeled or Gridded Data"). You cannot perform ecological risk on these data types as of now; however, you'll need to import the contoured results as point data, specifying each x,y,z value for each sample value. You will need to add the Media Type and Name column as well. This will take some processing outside of SADA. In a future version, SADA will be able to conduct a risk analysis on contoured results without this extra effort. For now, this is the only method.

Non-Spatial Data

Sometimes samples are collected without noting the coordinates. If this happens, you can essentially “fake” the coordinates outside of SADA using Excel or something similar. You will need to be aware, though, that the coordinate system is fake, and any spatial trend you might see is purely artificial. You cannot responsibly conduct any type of spatial risk calculation. You can, though, fully use the traditional risk models (tabular PRGs/risk/screens).

Exposure Concentrations From Outside SADA

Normally, SADA calculates exposure concentrations internally, using normal or lognormal assumptions. If you wish to use another distribution or means of calculating the exposure concentration, you certainly can. Just as with the non-spatial data, you will simply fake the coordinate for each representative concentration for each contaminant. Each contaminant essentially has only one “sample” and that is the exposure concentration value itself. This can then be run through the risk models to produce a traditional or tabular risk assessment. As with non-spatial data, you cannot sensibly apply any spatial tools. You can establish two SADA files. In the first SADA file, you use the exposure concentrations with fake coordinates. In the second SADA file, you import the actual data values with true coordinates and then conduct spatial screens, remedial decision designs, and secondary sampling designs, which do not rely on any particular underlying distribution for the exposure concentration.

Next Steps

In the next chapters, we will discuss how to calculate traditional tabular and spatial risk assessments.

Chapter 24: Calibrating the Ecological Risk Model

Ecological Benchmarks

Ecological benchmarks are a central component in an ecological risk evaluation. Benchmarks are environmental effect concentrations derived from toxicity testing, extrapolated from other benchmarks, or from simulations of an assessment endpoint. SADA provides a number of benchmarks for surface water, sediment, soil, and biota (tissue concentrations). SADA developers have made a point of not independently deriving benchmarks, so benchmarks in the SADA database are all from published sources. Sources and citations are described in more detail in the SADA help file.

For surface water, the SADA ecological benchmark database includes:

- acute and chronic National Ambient Water Quality Criteria,
- Great Lakes Tier II secondary acute and chronic values,
- criteria from three EPA regions (4, 5, and 6),
- Canadian water quality guidelines, and
- Lowest chronic values, EC20s, and population EC25s from Suter and Tsao (1996).

Sediment values include no effects concentrations, threshold effects concentrations, and probable effects concentrations from sources such as:

- EPA's Great Lakes Assessment and Remediation of Contaminated Sediment program,
- Canadian sediment quality guidelines,
- consensus-based values (MacDonald et al. 2000),
- EPA region 4, 5, and 6 values,
- Florida Department of Environmental Protection,
- the National Oceanic and Atmospheric Administration,
- Ontario Ministry of the Environment, and
- the state of Washington.

Fewer benchmarks are available for soil, but the database includes values from:

- EPA's ecological soil screening level program,
- screening levels from EPA regions 4, 5, and 6,
- invertebrate, microbe, and plant values from Oak Ridge National Laboratory, and
- Dutch target and intervention levels.

While the benchmark database includes 424 chemicals, it is important to realize that not all sources have values for every chemical in the database.

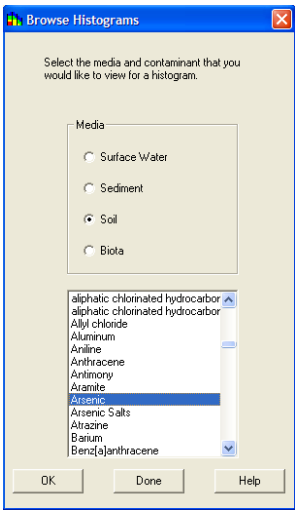
Some benchmarks (i.e., certain metals in surface water) are a function of environmental variables. SADA allows the user to enter site-specific pH, water hardness, and fraction of organic carbon and, where appropriate, calculates benchmark values associated with site-specific environmental variables.

Browsing Ecological Benchmarks – Benchmark Histograms

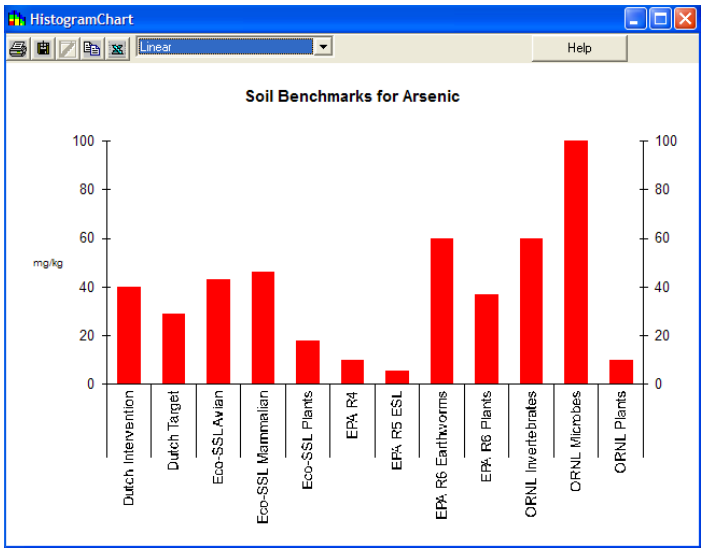
You can view benchmark values in the database in a couple of ways. One is as a histogram showing values from all benchmark sources for the selected contaminant and media type. You can browse values from any SADA file or from a master benchmark file.

If you do not already have EcologicalRisk.sda open, open it now in SADA. This file was created in the previous chapter. Select *Ecological* from the analysis list, then go to the *Ecological* drop-down menu and click on *Browse→Benchmark Histogram from….* Find and select the benchmark source database you want to view and click *Open*.

SADA responds with the *Browse Histograms* window.



Here, you select the data type, or media, (*Soil)* and data set (*Arsenic*) for which you want to see the benchmark histogram, then select *OK*. SADA then displays a histogram of available benchmarks for the selected media and contaminant.



You can also view a histogram of benchmarks simply by choosing a data type and data set and then selecting *Benchmark Histogram* from the *Ecological* drop-down menu. The difference is that the *Benchmark Histogram* views benchmarks currently loaded in your .sda file for contaminants that exist in your data set, while the *Browse Histogram From …* option allows you

to view benchmarks from other .sda files or from a master benchmark database not included in your .sda file and choose contaminants that are not in your data set.

Hint: If a source appears on the histogram but appears to have no value, try selecting a log scale representation from the drop-down list on the histogram window instead of the default linear plot.

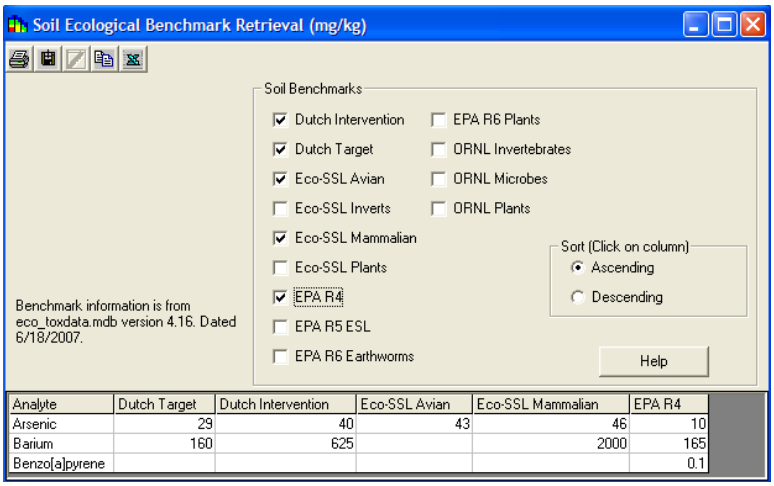
Browsing Ecological Benchmarks – Benchmark Tables

Another way to view benchmark values is in tabular form. This works the same way as for benchmark histograms. You can browse the benchmarks you've linked with your .sda file, benchmarks from another .sda file, or benchmarks from a master database external to SADA.

Select *Ecological* from the analysis list, then go to the *Ecological* drop-down list and click on Browse→Benchmark Table From…. Find and select the benchmark source database you want to view and click *Open*.

SADA responds with the *Browse Tables* window that looks very much like the *Browse Histograms* window. Select the media and contaminant(s) for which you want to see a tabular display of benchmarks. Note that you can select multiple contaminants by holding the control (CTRL) key while you click on contaminant names. When you've selected the media and contaminant(s) you want, click *OK*.

Whereas the *Benchmark Histogram* option displayed values for all sources for a single contaminant at this stage, in the tabular display you have to select which benchmark sources you want to view. Select them by clicking in the appropriate box, and SADA will display the corresponding benchmark values in table form in the *Ecological Benchmark Retrieval* window.



You can print the table, save it to a file, copy and paste it, or export it to Excel using the buttons in the upper left corner of the window.

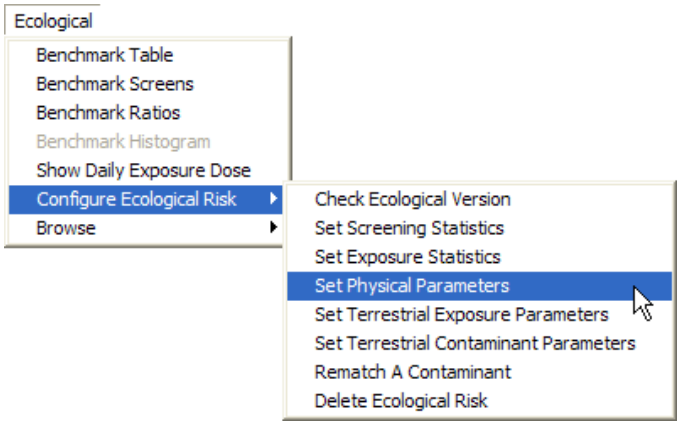
As with benchmark histograms, there is a shortcut for viewing a benchmark table for the analytes in your .sda file. Simply, select a media type, an individual contaminant (or pooled if you want to see benchmarks for all analytes in your file), and then select *Benchmark Table* from the *Ecological* drop-down list. Next, select the benchmarks to be viewed from the *Ecological Benchmark Retrieval* window. This route also gives you the option of specifying a statistic (i.e., mean, UCL95, max detect) to be displayed for the analyte(s) of interest. This is useful when trying to get a feel for where your data fall relative to available benchmarks.

Note: Screening benchmarks are meant to be conservative, so one approach is to use the most conservative value available for a given contaminant. It is inappropriate to browse the available benchmarks, looking for the least conservative. For the purposes of conducting an assessment, the best approach is to discuss available benchmark sources with risk assessment staff from involved regulatory agencies and other stakeholders to determine a preferred source, and then agree on a process to identify screening values for contaminants lacking values in the preferred source.

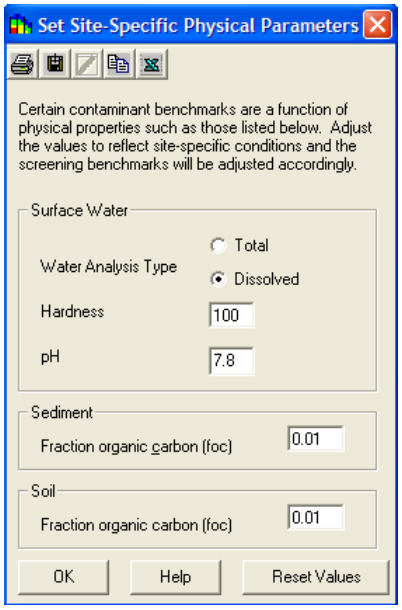
Setting Physical Parameters

As noted previously, some benchmarks are functions of environmental variables (water hardness or pH for surface water, fraction organic carbon for soil and sediment). Defaults are set at 100 mg CaCO3/L for hardness, 7.8 for pH, and 1% for fraction organic carbon. SADA allows you to use site-specific values for these variables and will recalculate benchmarks, where appropriate.

To set physical parameters, from the *Ecological* drop-down list select *Configure Ecological Risk→Set Physical Parameters*.



SADA responds with the *Set Site-specific Physical Parameters* window.



Enter site-specific values and click *OK*. The values are now set and will be used when viewing benchmarks and when conducting functions that use benchmarks in risk calculations.

Note that for surface water, there is also the option of expressing concentrations as *Total* or *Dissolved*. EPA has *Total* to *Dissolved* conversion factors for some metals, and for these, benchmarks will be recalculated depending on whether *Total* or *Dissolved* is selected. Values of analytes without conversion factors will not be changed.

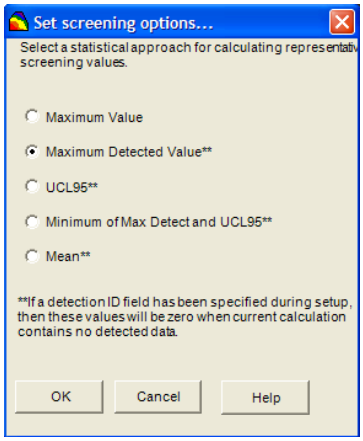
Values can be reset to the defaults by clicking the *Reset Values* button.

Setting Exposure Statistics

To conduct the benchmark screening, you need to select a contaminant(s), select a screening statistic (usually the maximum detected concentration), and a benchmark(s).

The simplest screen is a pass/fail comparison of the maximum concentration against a conservative screening benchmark. SADA allows you to specify the screening statistic (default is *Maximum Detected Value*) and choose a single benchmark source.

To conduct a benchmark screen for a given media and contaminant(s) combination, first check the screening statistic to be sure it is what you want. Select *Set Screening Statistics* from the *Ecological→Configure Ecological Risk* menu. At the *Set Screening Options* window, confirm that your desired statistic is selected, then click *OK*.



Similarly, for the calculation of benchmark ratios, you will need to specify the representative exposure statistic. The default exposure statistic is the lower of the maximum detect and the UCL95. If computing ratios for screening purposes, be sure to change the exposure statistic (not the screening statistic) to what you want used for the ratio screen. To change the exposure statistic, select *Ecological→Configure Ecological Risk→Set Exposure Statistics*.

By convention, if no data are detected for a given contaminant, the representative concentration for the purposes of screening will be zero for all choices but *Maximum Value*.

Save your EcologicalRisk.sda file. If you are continuing to the next chapter, keep it open. Otherwise, you may close it.

Chapter 25: Conducting Tabular Ecological Risk Assessments in SADA

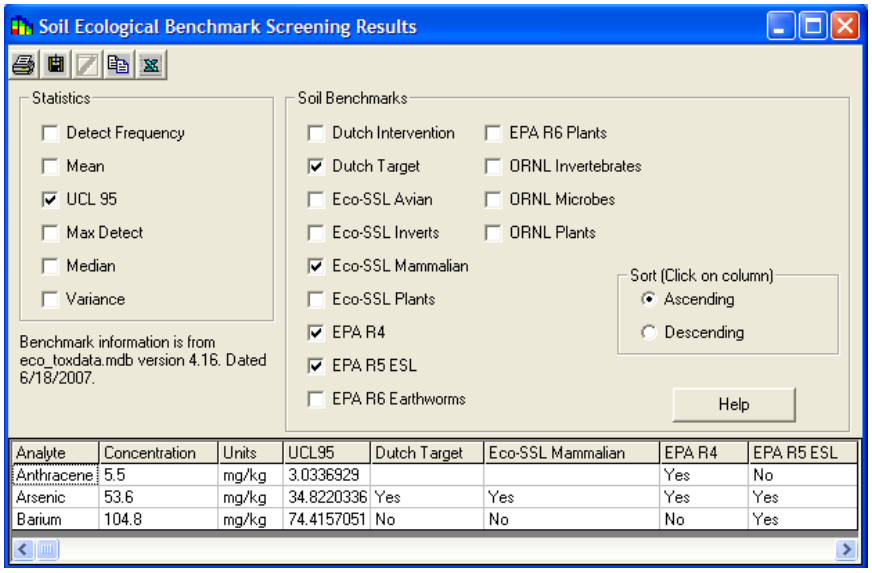
Assuming you have already explored your data (see Chapter 11, "Visualizing and Exploring Your Data"), you may want to screen observed contaminant concentrations against benchmark values. This is the screening step in ecological hazard identification, or Step 2 in the EPA ecological risk assessment process.

Benchmark Screening

To conduct the benchmark screening, you need to select the contaminant(s), the screening statistic (usually the maximum detected concentration), and the benchmark(s). In the previous chapter, we discussed how to set the screening statistics option. The simplest screen is a pass/fail comparison of the maximum concentration against a conservative screening benchmark. SADA allows you to specify the screening statistic (default is maximum detected value) and choose a single benchmark source.

If you do not already have EcologicalRisk.sda open (created in Chapter 24), open it now. From the main menu, select *Ecological→Benchmark Screens*. SADA will respond with the *Soil Ecological Benchmark Screening Results* window. This window includes columns for contaminants, screening exposure concentrations, units, additional selected statistics (i.e., mean, UCL95, max detect), and selected benchmarks.

You select the benchmarks to use by clicking the appropriate box. SADA responds by comparing the exposure concentration to the benchmark and reporting *Yes* if the concentration exceeds the benchmark and *No* if it does not.



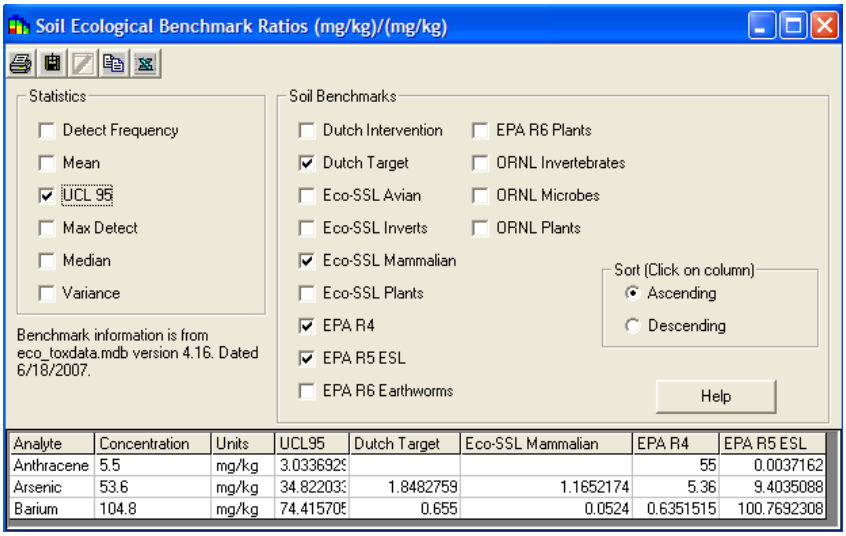
Analytes with maximum concentrations below a conservative benchmark do not need to be evaluated further in an ecological assessment. Those that failed the screen, having maximum concentrations above the benchmark as indicated by *Yes*, merit further investigation. If no analytes at a site exceed the screening benchmark, the entire site can be dropped from further investigation.

Tip: It is rare for a single benchmark source to have values for all analytes in a site data set. It may be desirable to create a custom analysis data set that draws from a number of benchmark sources to populate the benchmark database. This could involve selecting a primary source and then filling in the gaps using other sources. SADA allows users to select a benchmark hierarchy for single contaminants when plotting screening results but not for tabular results, which can contain multiple contaminants. For tabular screens, you need to create a custom analysis or edit an existing benchmark field in the ecotox database, so that it represents the screening values to be used for your site. Essentially, you are manually creating the benchmark hierarchy. Any time you tweak SADA like this, it is important to fully document what you have done, and it is best to reach agreement with regulators and stakeholders ahead of time.

A simple pass/fail screening analysis does not provide any information about the magnitude of exceedances. Often, it can be more illuminating to conduct the screen similarly but output ratios of maximum concentrations to benchmark values. The interpretation of the results is similar, but instead of *Yes/No*, it is >1/<1. Ratios less than 1 indicate the maximum concentration was below the benchmark value; those greater than 1 indicate the concentration exceeded the benchmark value.

Benchmark Ratios

Tabulation of the ratios is set up the same way as for *Benchmark Screens*. Select *Ecological→Benchmark Ratios*; then select the benchmarks to screen against. SADA responds with the *Ecological Benchmark Ratios* window, which displays numeric benchmark ratios instead of *Yes/No* values. These can be printed, saved to a file, copied, or exported to Excel.



Important Note: The default exposure statistic for *Benchmark Ratios* is the lower of the *Maximum Detect* and the *UCL95*. If computing ratios for screening purposes, be sure to change the exposure statistic (not the screening statistic) that is to be used for the ratio screen. To change the exposure statistic, select *Ecological→Configure Ecological Risk→Set Exposure Statistics.*

See Chapter 27, “Spatial Ecological Risk,” for more information on conducting a spatial ecological screening (i.e., mapping the screening results generated above).

Chapter 26: Ecological Risk: Terrestrial Dose Modeling

Terrestrial exposure methods and the ability to input toxicity reference values are available to estimate exposure for commonly used ecological risk assessment endpoints. In addition, the SADA GIS provides a platform for additional exposure/risk assessment methods that account for spatial dependence, including combining Interpolation methods with the exposure models to visualize areas of exposure at the site. These tools provide methods that can minimize remedial action decision errors, provide spatial designs for more efficient and cost effective remedial design under selective remediation conditions, and provide a rationale and context for additional sampling efforts at contaminated sites. The integral combination of spatial assessment approaches with screening and exposure assessment capabilities has the advantage of improving the quality of ecological risk-based decisions without overly complicating the assessment.

The ecological exposure and risk models follow USEPA guidance, and can be customized to fit site-specific exposure conditions for generating risk results and calculating Preliminary Remediation Goals (PRGs). Exposure and risk assessment calculations are based on an initial conceptual model for the site that includes the contaminated media, routes of contaminant transport, representative ecological receptors, and pathways of exposure for these receptors. Calculations can be limited to those contaminants that exceeded relevant screening values in the screening process. The next step for higher tier ecological risk assessments is determining the bioavailability of these chemicals to exposed receptors; this can be done through physiological or food chain models that model the bioaccumulation and biomagnification of the contaminants in different trophic levels of the food chain. The magnitude of exposures to the individual receptors can then be calculated using exposure models for the relevant exposure pathways. In the next step, a toxicity reference value (TRV) from a dose-response model is needed. This may call for inter-species extrapolation of the contaminant effects from a laboratory study to a species of interest at the contaminated site. Results of the exposure modeling are then compared to decision criterion (the TRV); the exposure models can then be modified for different exposure scenarios to determine the feasibility of available alternative remedial actions.

SADA provides terrestrial dose exposure models to estimate the daily doses of contaminants at a site. Modeling dose to wildlife receptors requires numerous chemical-specific and species-specific exposure parameters. The ecological dose modeling capabilities are located in the ecological risk module, and default exposure parameters are distributed with the software for over 20 terrestrial species (Table 1) that represent a variety of wildlife receptors, including characteristic herbivores, insectivores, and carnivores. These species can be parameterized individually for males, females, and juveniles, as well as USEPA Soil Screening Level (SSL) defaults that combine male and female parameters. Default values for males and females are based on available literature sources but can be user modified to reflect site-specific conditions.

Table 1. Terrestrial species with default exposure parameters

Common Name	Scientific name	Receptor Group
American kestrel	*Falco sparverius*	Avian carnivore
American robin	*Turdus migratorius*	Avian insectivore
American woodcock	*Scolopax minor*	Avian ground insectivore
Black-tailed jackrabbit	*Lepus californicus*	Mammalian herbivore
Burrowing owl	*Speotyto cunicularia*	Avian carnivore
Deer mouse	*Peromyscus maniculatus*	Mammalian omnivore

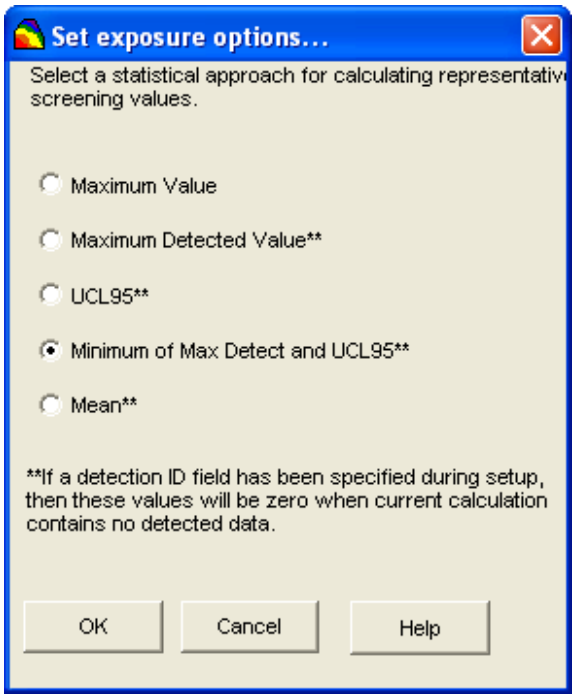
Common Name	Scientific name	Receptor Group
Eastern cottontail	*Sylvilagus floridanus*	Mammalian herbivore
Great Basin pocket mouse	*Perognathus parvus*	Mammalian granivore
Kit fox	*Vulpes macrotis*	Mammalian carnivore
Little brown bat	*Myotis lucifugus*	Mammalian insectivore
Long-tailed weasel	*Mustela frenata*	Mammalian carnivore
Meadow vole	*Microtus pennsylvanicus*	Mammalian herbivore
Mexican free-tailed bat	*Tadarida brasiliensis*	Mammalian insectivore
Mourning dove	*Zenaida macroura*	Avian granivore
Northern bobwhite	*Colinus virginianus*	Avian granivore
Prairie vole	*Microtus ochrogaster*	Mammalian herbivore
Red fox	*Vulpes vulpes*	Mammalian carnivore
Red-tailed hawk	Buteo jamaicensis	*Avian carnivore*
Short-tailed shrew	*Blarina brevicauda*	Mammalian insectivore

Routines are available in the exposure models for calculating daily intake rates for the selected receptors at each location in the site. The routines in the ecological risk module access the contaminant matching and data management functions, and model parameters and toxicity information are fully parameterized from USEPA guidance when available. SADA calculates dose (mg intake per kg body weight per day) from food ingestion, soil ingestion, dermal contact, and inhalation for terrestrial exposures, plus the total dose summed over all pathways selected. These results are directly comparable to TRVs for risk assessment, and are presented in tabular form commonly used to document risk assessment results.

Representative Exposure Concentration

An early decision when performing an exposure assessment is the selection of an appropriate statistic for the exposure concentration.

Select *Set Exposure Statistics* from the *Configure Ecological Risk* menu, select the desired statistic (the default of *Minimum of Max Detect and UCL95* is commonly used for risk assessments), then click *OK*.

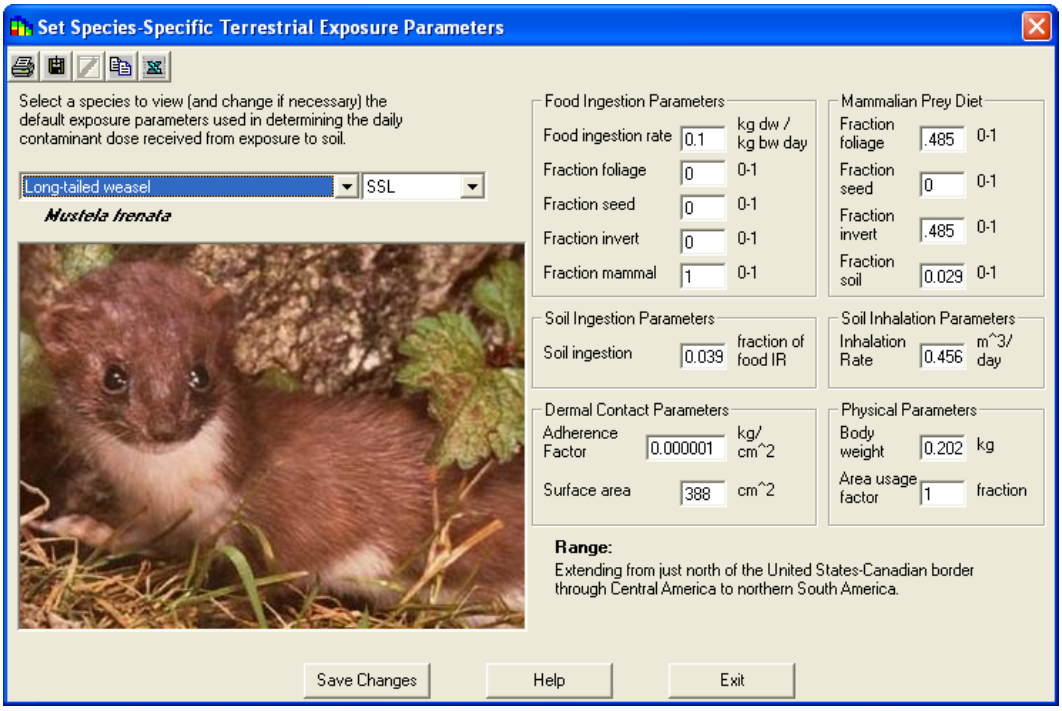


Species-specific Exposure Parameters

The exposure concentration is used with a number of species- and chemical-specific parameters to calculate the total exposure. SADA comes preloaded with default exposure parameters for a variety of wildlife receptors, including representative herbivores, insectivores, and carnivores. Defaults are provided for each parameter, but custom values can be entered in the *Set Species-specific Terrestrial Exposure Parameters* window. These parameters are used in modeling dose to each receptor from selected pathways.

To view (or change) current exposure parameters, go to the *Ecological* menu, select *Configure Ecological Risk*, and *Set Terrestrial Exposure Parameters*.

From the resulting window, select a species from the drop-down list, and select whether you want values for adult females, adult males, juveniles, or SSL. SSL follows EPA draft guidance for deriving ecological soil screening levels and combines data from adult males and adult females. While custom values can be entered for juveniles, SADA does not include default values for juveniles.



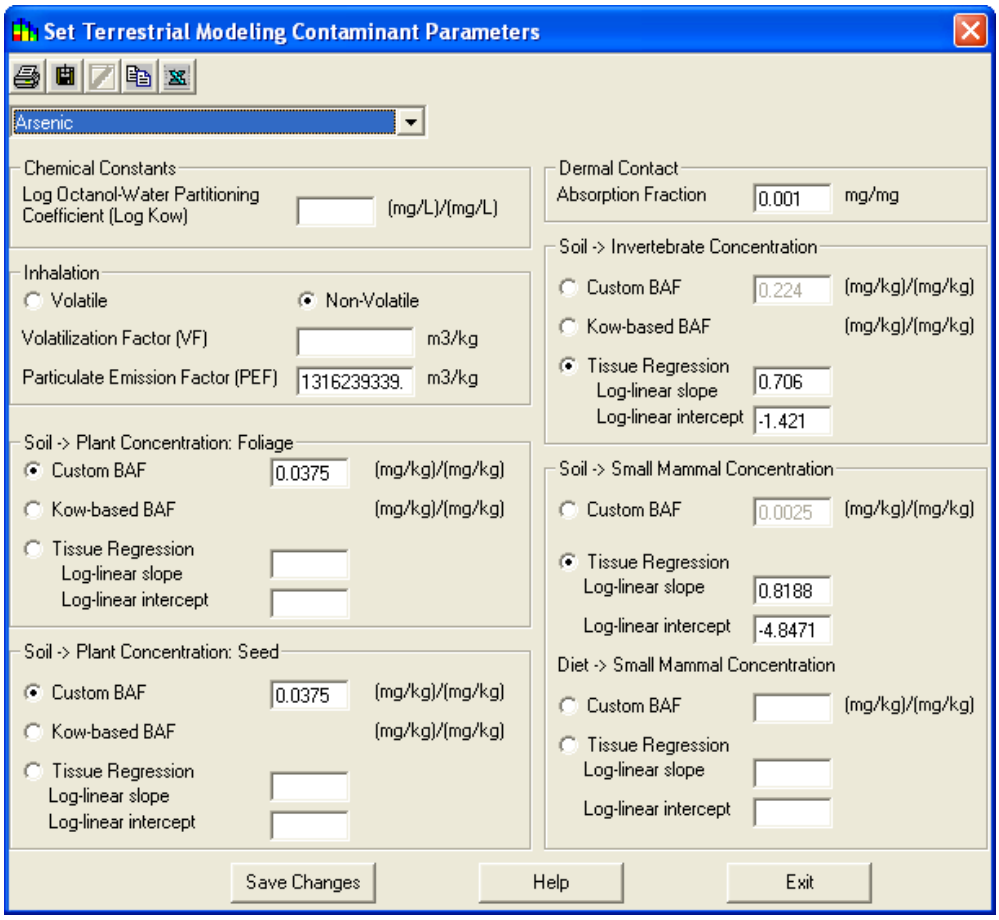
To change a parameter value, click on the box corresponding to that parameter, and enter the value. It is very important to use the appropriate units for all parameters if entering custom values.

Chemical Exposure Parameters

Many of the parameters used in modeling dose to wildlife receptors are chemical-specific rather than species-specific. For species-specific parameters, see *Set Terrestrial Exposure Parameters*. Custom values for contaminant-specific parameters can be entered in the *Set Terrestrial Modeling Contaminant Parameters* window. These parameters are used in modeling dose to each receptor from selected pathways.

To view (or change) current contaminant-specific exposure parameters, go to the *Ecological* menu, select *Configure Ecological Risk*, and *Set Terrestrial Modeling Contaminant Parameters*.

In the resulting window, select an analyte from the drop down list. To change a parameter value, click on the box corresponding to that parameter, and enter the value. When you've finished modifying the parameters for the species, click *Save Changes*, and the changes will be saved to your SADA file. Otherwise, changes will not be saved. It is very important to use the appropriate units for all parameters.

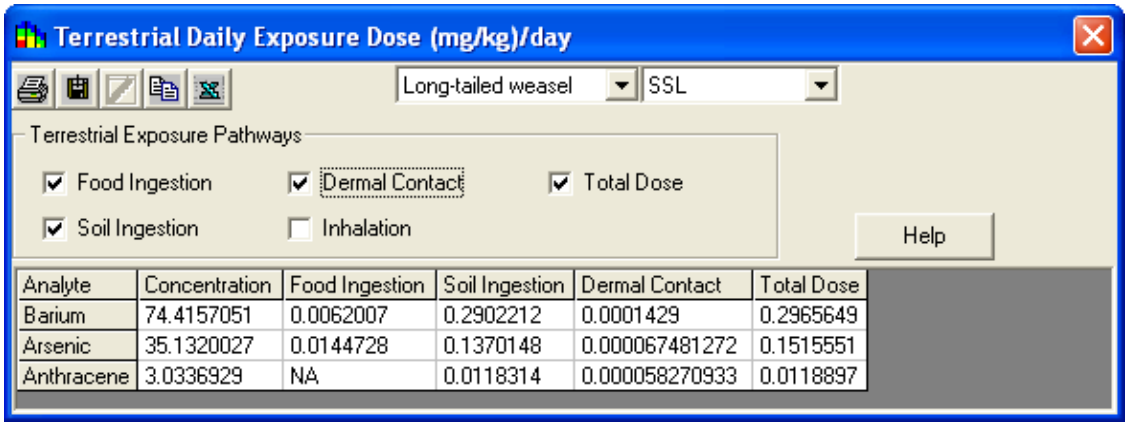


Terrestrial Daily Exposure Dose Table

SADA calculates dose (mg/kg BW d) from food ingestion, soil ingestion, dermal contact, and inhalation for terrestrial exposures as well as total dose summed over all pathways selected.

To select a pathway for dose estimation, select a soil contaminant, and from the *Ecological menu select Show Daily Exposure Dose*.

From the following form, select a receptor and set of parameters from the drop down list. Then, select the pathway(s) for which you want to calculate dose. The parameter sets are SSL, Female, Male, or Juvenile. Choosing SSL will use parameter values from EPA's draft Eco-SSL guidance, which represents data from males and females combined. Choosing Female or Male will restrict the parameter values to those for that gender. SADA does not currently provide parameter values for juveniles, but custom values can be entered via *Set Terrestrial Exposure Parameters.*



Analyte	Concentration	Food Ingestion	Soil Ingestion	Dermal Contact	Total Dose
Barium	74.4157051	0.0062007	0.2902212	0.0001429	0.2965649
Arsenic	35.1320027	0.0144728	0.1370148	0.000067481272	0.1515551
Anthracene	3.0336929	NA	0.0118314	0.000058270933	0.0118897

Modeling dose to wildlife receptors requires a number of chemical-specific and species-specific exposure parameters. To view (or change) the contaminant-specific or species-specific parameter values used in estimation of dose, close the *Daily Exposure Dose* window, and from the *Ecological* menu, click *Configure Ecological Risk*, and either *Set Terrestrial Contaminant Parameters* or *Set Terrestrial Exposure Parameters* discussed previously.

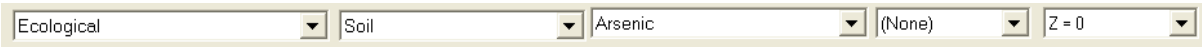
Chapter 27: Spatial Ecological Risk

In the previous chapter, you saw how to generate tabular benchmark, risk, and dose results. These types of calculations are important classical approaches that help identify risk drivers. As with human health risk, these models can be integrated with spatial tools to show you *where* exceedances occurred and what point risk and point dose maps look like. Even contoured risk and dose maps can be generated that support the remedial area of concern, and in some cases even motivate secondary sampling strategies. This chapter will cover spatial ecological screens, point risk, point dose and introduce the more advanced principles of risk/dose contouring, remedial design, and secondary sampling strategies. These principles are taken up in later chapters after the introduction to geospatial modeling is covered, but a glimpse of them now will be beneficial.

Spatial Screens

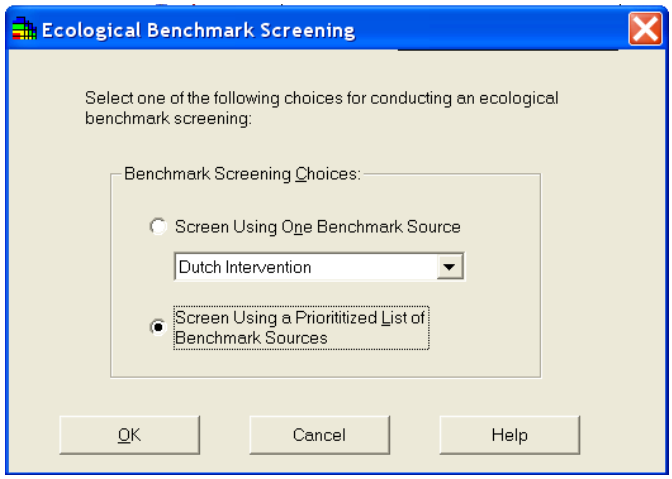
In Chapters 13 and 15, we demonstrated how to perform simple and informative data screens using user defined or custom criteria values. The process was the same for human health and you'll now see that ecological is no different. In an ecological analysis, a benchmark is compared to every individual sample point for a given contaminant. If the criterion is exceeded, the point is highlighted with a screening box. With this simple tool, it is easy to spatially determine where risk may be a greater concern.

Close out any SADA file you may have open and open the file SpatialEcologicalRisk.sda. Select *Ecological* as the analysis, *Soil* as the media, and *Arsenic* as the contaminant.

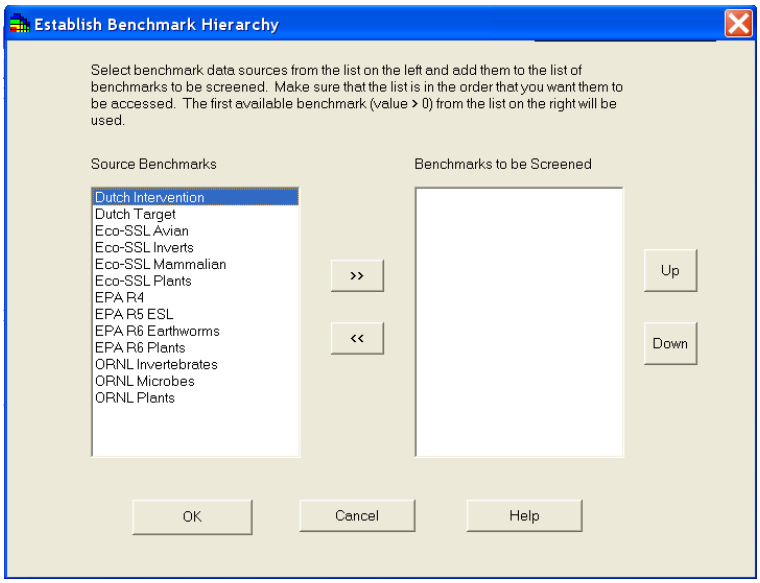


In the interview drop list, select *Draw a data screen map* and press the *Show The Results* button.

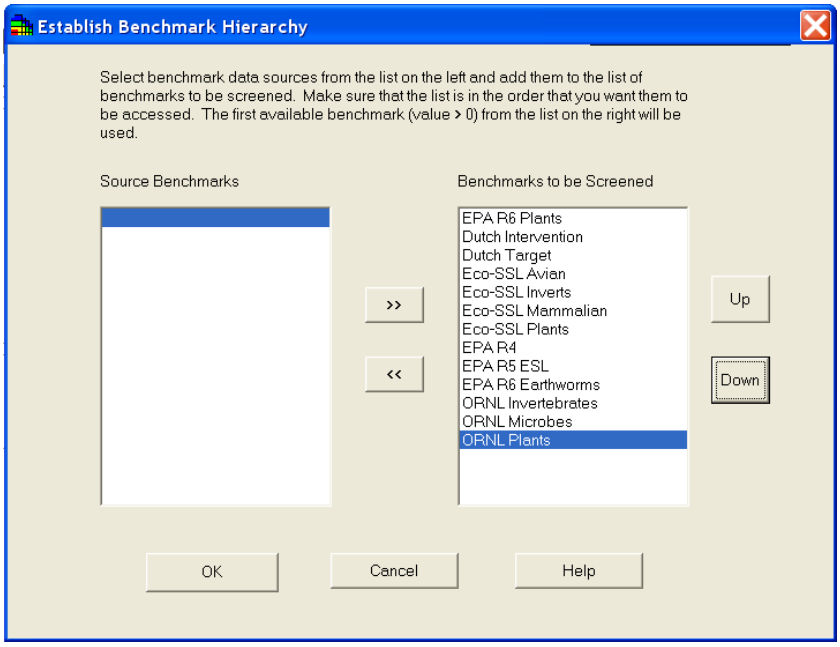
You will be presented with a window where you can choose a specific benchmark or establish a hierarchy of benchmark values. Many of the contaminants will not have benchmark values and so it is often necessary to specify a hierarchy of benchmarks to choose from. By hierarchy, we mean an order of preference. You'll put your favorite benchmark source on the top of the list; but if it has no value, SADA will next go to your alternate benchmark source. If there is still no value, the process continues. Let's take a look at this option now. Select the *Screen Using a Prioritized List of Benchmark Sources* option.



Press *OK* and you are presented with a new window where you can add and order your prioritized list.

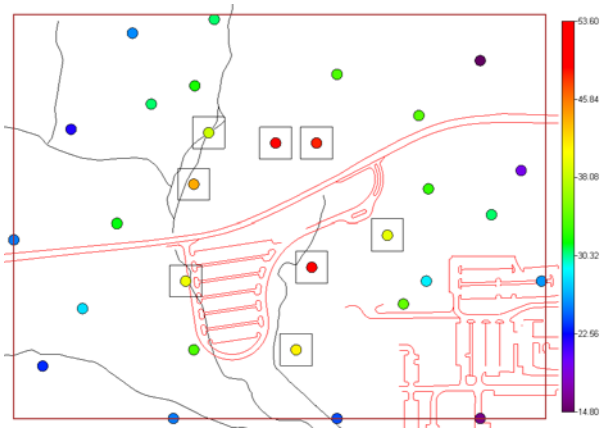


There are two lists here. On the left is the complete set of benchmark sources available in the ecological benchmark database. On the right are the selected and ordered *Source Benchmarks*. When you first access this window, the right will be empty. Using your mouse, select each benchmark you want to use on the left and use the arrow button to move it to the right. In this example, we'll move all of them to the right.

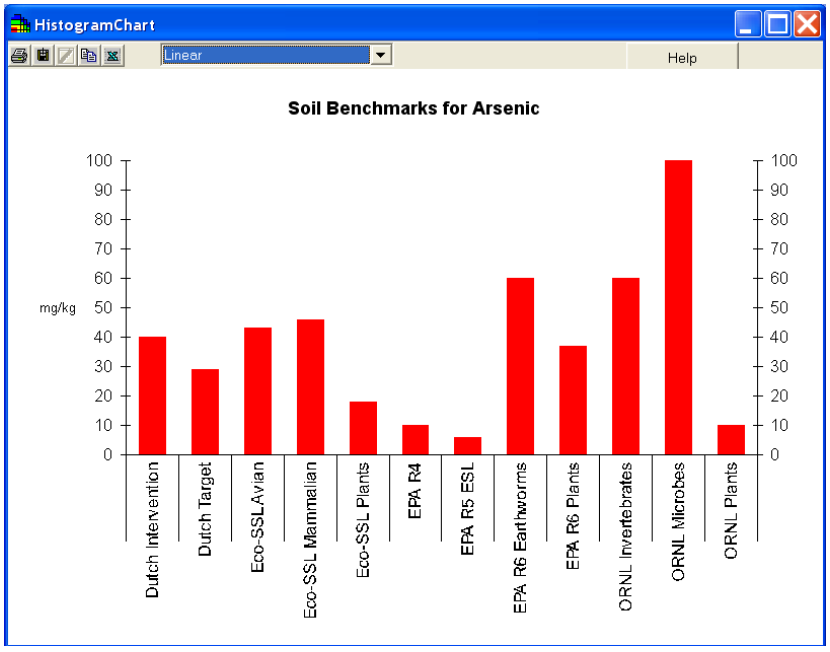


Next, move *EPA Region 6 Plants* to the top of the selected list by first selecting it and then repeatedly pressing the *Up* button until it reaches the top. If there is no value for *EPA Region 6 Plants,* SADA will check Dutch Intervention next. Press the *OK* button.

SADA reports that *EPA Region 6 Plants* did have a value of 37 (at the time this was written). Press *OK* and the screen is applied showing 8 points exceeding the criterion.



Recall that you can see all the benchmarks for a given contaminant by selecting *Ecological→Benchmark Histogram*. This may guide you in your selection of benchmark sources.

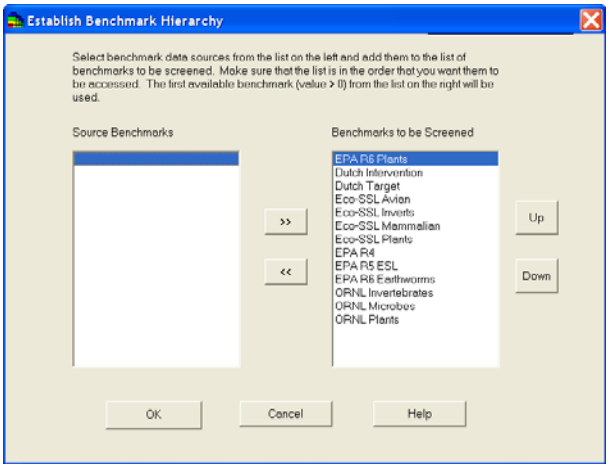


Point Risk

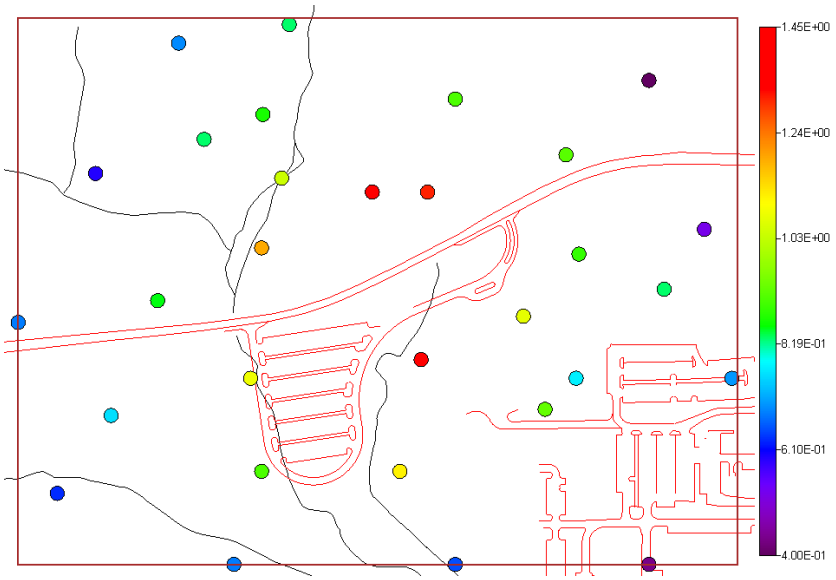
Data screens simply demonstrate where the exceedance occurred. To determine how severe the exceedance is, you can use the point risk feature to actually run each point through the risk model producing ratio value. Any ratio (risk) value greater than one exceeds the decision

criterion. A ratio shows severity. For example, a ratio value of 2 means that the concentration is twice the criteria.

From the interview list, select *Draw a point risk map*. Press the *Show The Results* button and you'll see the same options as with the screen. Let's choose a prioritized list again and pres *OK*. Accept the prioritized list we created before (with *EPA Region 6 Plants* our first choice) and press *OK* in this second window.

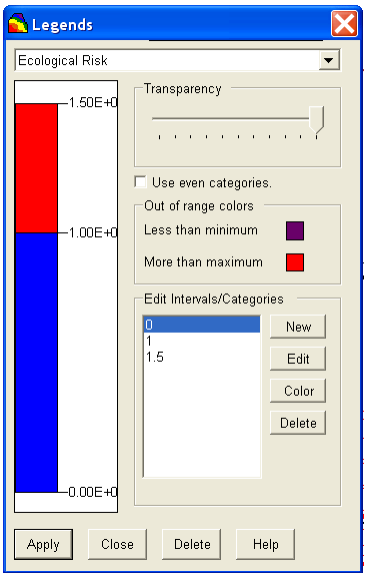


SADA will convert all concentration values to ratio (risk) values for the Arsenic/EPA R6 Plants scenario.

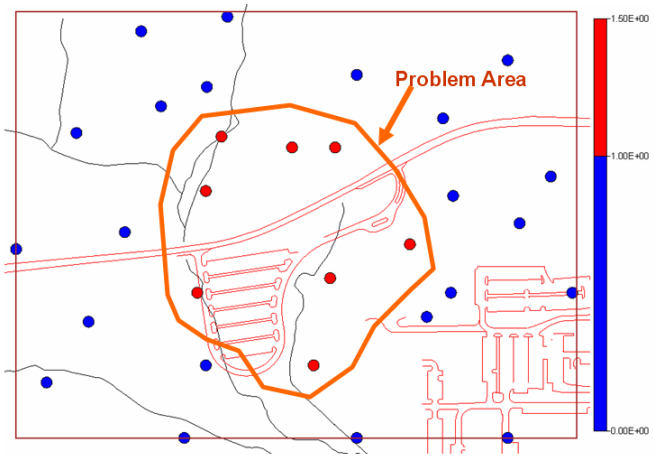


The distribution of colors will remain relatively the same as concentration plots. This is because a ratio is just a rescaling operation. It can be difficult to determine from the continuous legend type where a ratio greater than 1 might be found. To help us, we turn to the legend manager. From the main menu select *Graphics→Legend Manager*.

The legend manager is covered in detail in Chapter 43, but we will use it now to demonstrate how it can be used in this particular situation. In the drop-down list of available legends, select *Ecological Risk*. This is a custom legend we've created for you already. Any value greater than one will appear red; otherwise they will appear blue.



Press the *Apply* button and then the *Close* button. The resulting map is dramatically changed, yet the numerical results have not changed. Notice in the center of the map the same eight points highlighted with screening boxes are now colored red. This indicates that the center part of the site may spatially be an important risk driver.

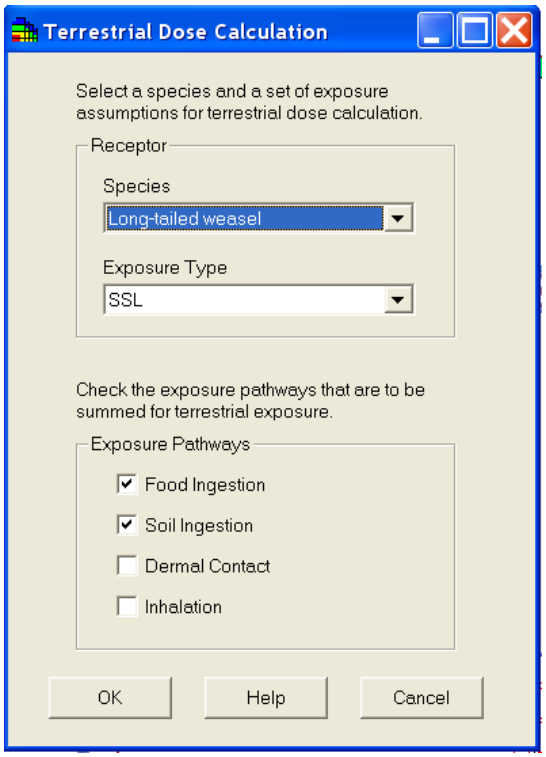


Point Dose Map

Recall from previous chapters that you can view dose estimates for a specific receptor/contaminant combination based on representative concentration value and a dose model. Those were tabular results and the code is fairly flexible in allowing you to parameterize these dose models (*Ecological→Configure Ecological Risk→Set Terrestrial Exposure Parameters and Ecological→Configure Ecological Risk→Set Terrestrial Contaminant Parameters*). You can also apply these dose models spatially in a couple of ways. One such way is to convert point values to dose values. While receptors may interact with a contaminant over a greater area than a point location, such a map can give a sense for where problem areas may exist.

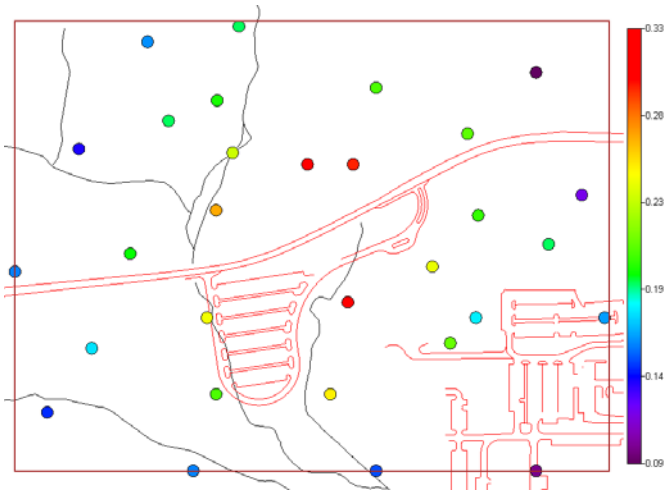
In the interview drop-down list, select *Draw an eco point dose map* and Press *Show the Results*. Using the dose modeling parameters you've specified (see Chapter 26), SADA now applies the

dose models to each point. First, you'll need to select a receptor (you'll have the opportunity to do so in the following window).



In the top parameter block are the *Species* and *Exposure Type* options. Let's choose the *Long-tailed weasel* and use the *SSL* option.

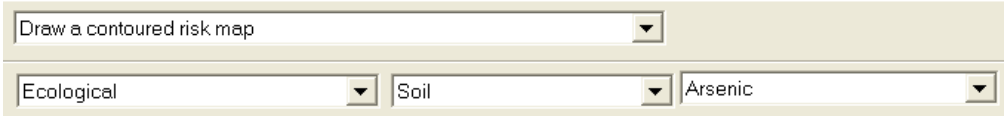
In the lower parameter block, we can choose what type of pathways we want to consider for the *Long-tailed weasel*. Let's include *Food Ingestion* and *Soil Ingestion* and press *OK*. SADA responds with the dose map.



Contouring Dose and Risk

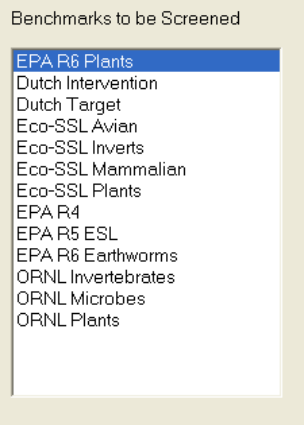
SADA is equipped with a reasonably comprehensive set of geospatial models that have a wide range of flexibility. Several later chapters are dedicated to the use of these models (Chapters 28-31). Geospatial models perform many services, the most common of which is a contoured map. In SADA, you can contour the point dose and risk values to create a continuous model over space.

To demonstrate this, we will use the simplest of geospatial models: natural neighbor. This interpolant does not require any input from the user. At this point in the guide, this makes it an ideal choice. Make sure you still have SpatialEcologicalRisk.sda open with the *Ecological* analysis selected, *Soil*, and *Arsenic*. From the interview drop-down list select *Draw a contoured risk map*.

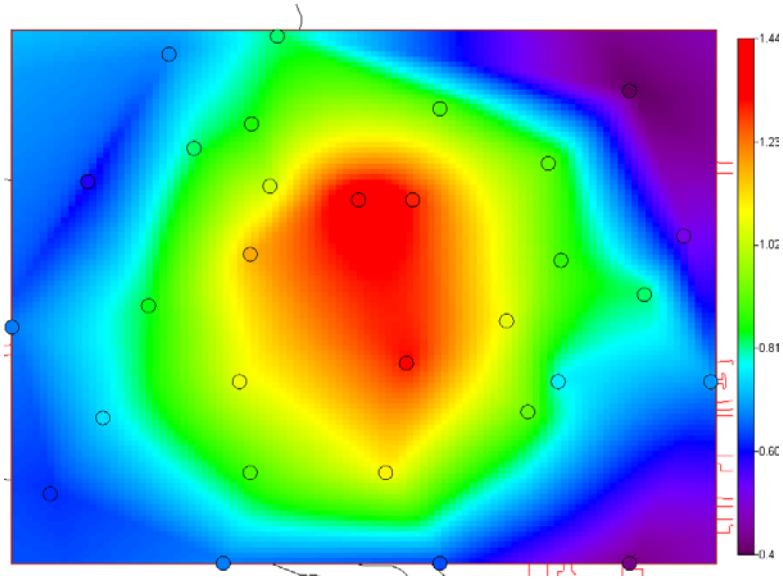


Click on the step *Interpolation methods* and choose *Natural Neighbor*.

Press the *Show The Results* button, choose the prioritized benchmark list and we'll continue to use the hierarchy we created earlier with *EPA Region 6 Plants* on top.

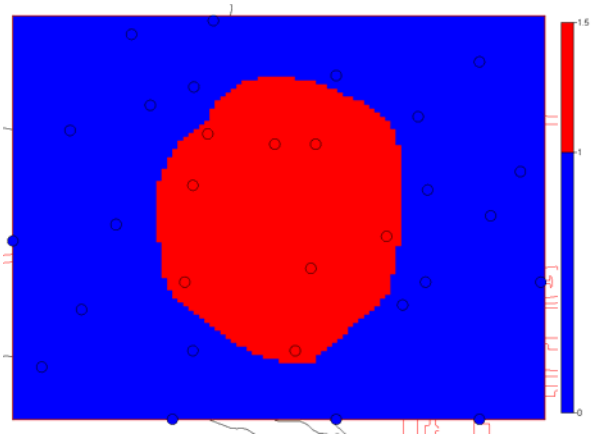


Pres the *OK* button on the hierarchy window, and the spatial risk map is produced.



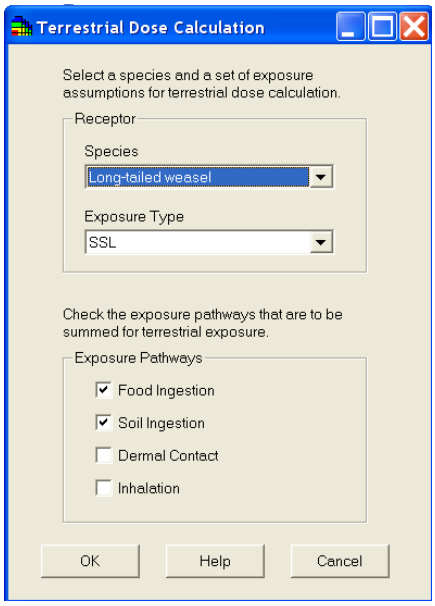
This is a contour of the point-wise risk over the entire site. Let's take a moment to improve the presentation of the map. From the main menu, select *Graphics→Legend Manager*. We will apply the *Ecological Risk Legend* again. If you are not familiar yet with the legend manager, you may wish to visit Chapter 43.

In the drop-down list, select *Ecological Risk* and Press *OK*.

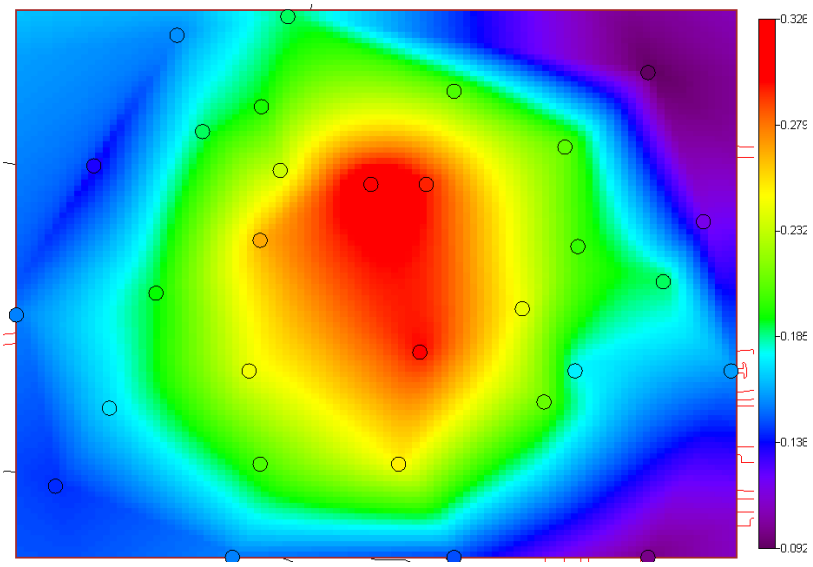


The resulting map shows a spatial delineation of where the problem may be found. This kind of result lends itself well to the discussion of geospatial decision making. We'll briefly cover this now and return to it after the geospatial modeling chapters have been presented.

A dose map may be created in much the same way. In the Interview drop-down list, select *Draw a contoured eco dose map*. Press *Show The Results*. First, we'll need to select a receptor and an exposure pathway, just like we did to create the point dose map. Under *Species* choose the *Long-tail weasel* and choose *SSL* for the *Exposure Type*. We'll again use the *Food Ingestion* and *Soil Ingestion* pathways. Make sure that your window looks like the following image, then press the *OK* button.



SADA responds with a contoured dose map.

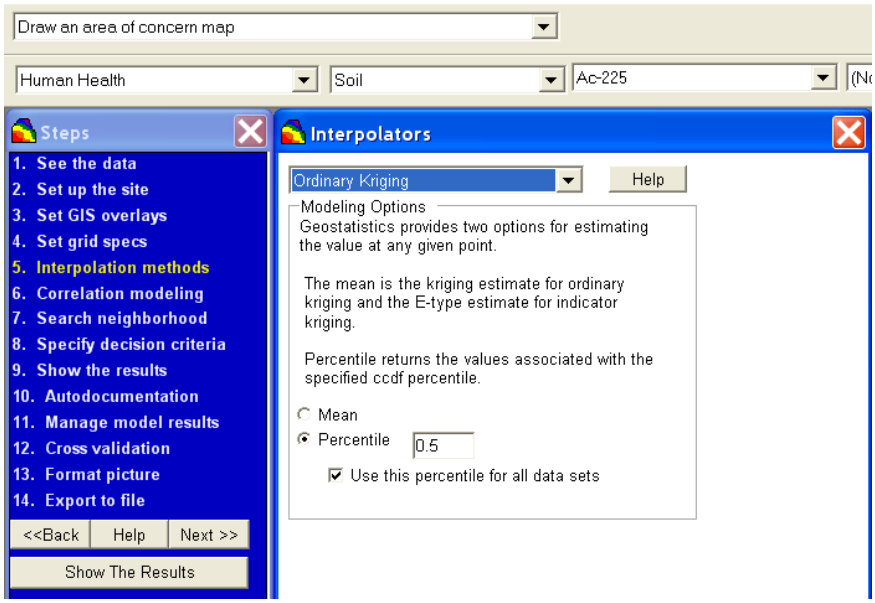


Using Risk in Decision Making

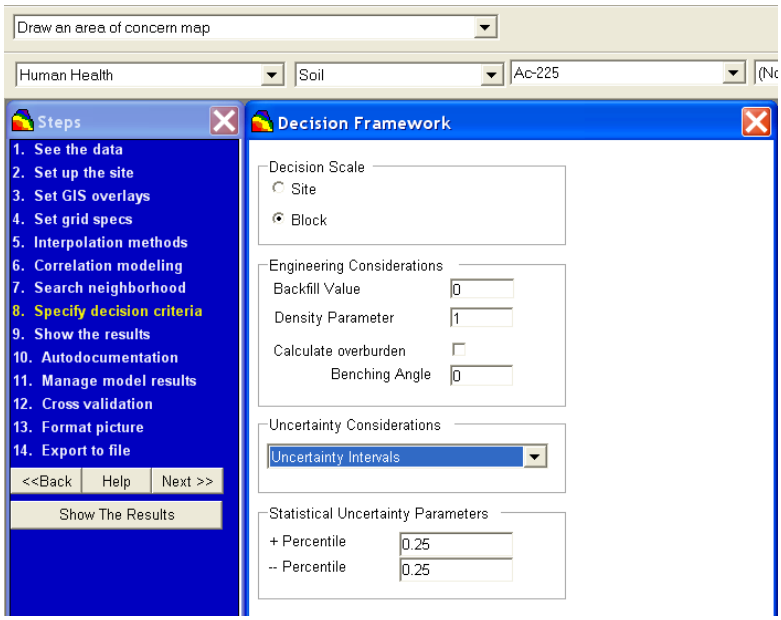
SADA provides a number of spatial decision analysis tools. Risk can play a role in this decision process. Later chapters (see Chapters 35 and 36) are dedicated to various decision analysis tools that make use of human health, ecological, custom, and MARSSIM-like decision making features. We will present an example here to give a taste for the kind of contribution risk can make in spatial-decision framework.

One example of a decision result is the development of an area of concern (AOC). An AOC is a spatial delineation of where concentration values may be too high and some action may be required. Risk assessors can estimate this area and also quantify the uncertainty about the location of the AOC boundary line.

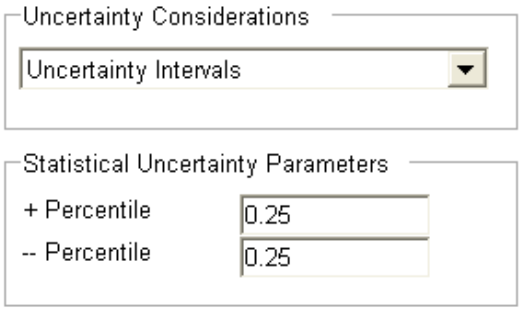
In the interview drop-down list, select *Draw an area of concern map* and click on the step *Interpolation methods*. From the drop-down list of available interpolators, choose *Ordinary Kriging*. This is a geostatistical model that allows one to quantify uncertainty in several ways. We will use this property of ordinary kriging to assist with quantifying the uncertainty about the AOC boundary line. The *Ordinary Kriging* model requires some calibration, and we have done that already for you. Make sure that your selections appear exactly as below:



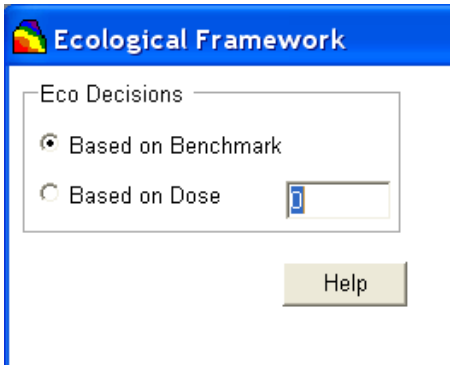
Click now on the step *Specify decision criteria*. There are a number of parameters located here that deal with defining the AOC. Focus your attention for now on the parameter block called *Uncertainty Considerations*. This section permits you to choose a method for understanding the AOC boundary uncertainty. The one you should select is called *Uncertainty Intervals*. Associated with this option are two parameters +/- Percentile. For now we'll accept the default values of 0.25 for each.



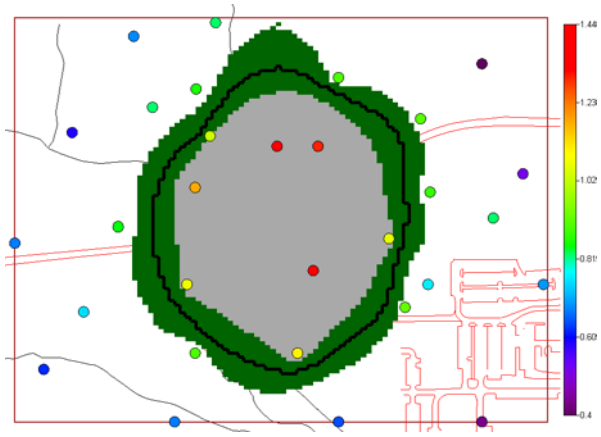
The important thing to understand at this point is that we are asking the code to specify the region where there is a 25%-75% (centered about 50%) point-wise probability that the true boundary line will occur.



For ecological decision analysis, you must also choose between a risk based (benchmark) or dose paradigm. Click on the step *Set Eco framework*. You can specify here whether to use risk or dose. If you choose dose, you'll need to enter a dose decision criterion. SADA will then ask for a receptor. If you choose risk, SADA will ask for a benchmark or benchmark hierarchy, just as before. Let's choose *Based on Benchmark*.



Press the *Show The Results* button, select prioritized list, and accept the hierarchy we created earlier (with *EPA Region 6 Plants* on top). The following result appears.



In this map, we see several things. For a benchmark value of roughly 37mg/kg, this is the area of concern. The thick black boundary line indicates the best estimate of the AOC boundary line. The green area represents the region where the actual boundary line could reasonably exist when considering our spatial uncertainties. The gray region is estimated to be very contaminated and will likely be within the AOC boundary even when considering the uncertainty.

A more comprehensive discussion of this approach can be found in Chapter 35. Other methods of decision support include secondary sampling designs to better isolate elevated samples or better refine the exact location of the AOC. Both of these are covered in later chapters.

Summary

SADA allows both tabular risk calculations and spatial risk modeling. The latter allows risk assessors to directly contribute to a decision process, such as developing an area of concern or locating another round of samples. If you do not have coordinates for your data values, you can fake the coordinates (create bogus coordinates outside of SADA) in order to get them into SADA. You can use them in the tabular risk features, but obviously you cannot use them in a spatial risk assessment.

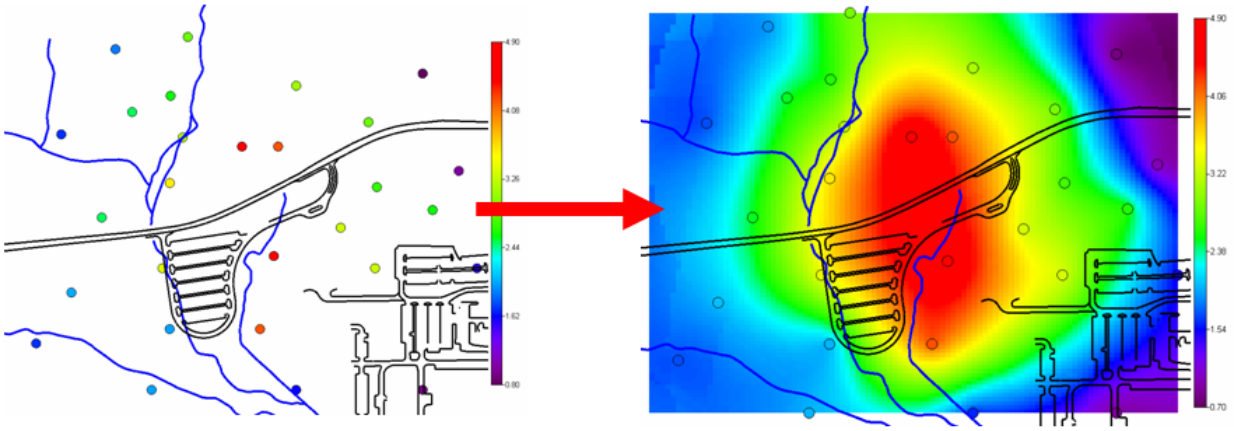
Part V: Spatial Modeling

Chapter 28: Geospatial Methods

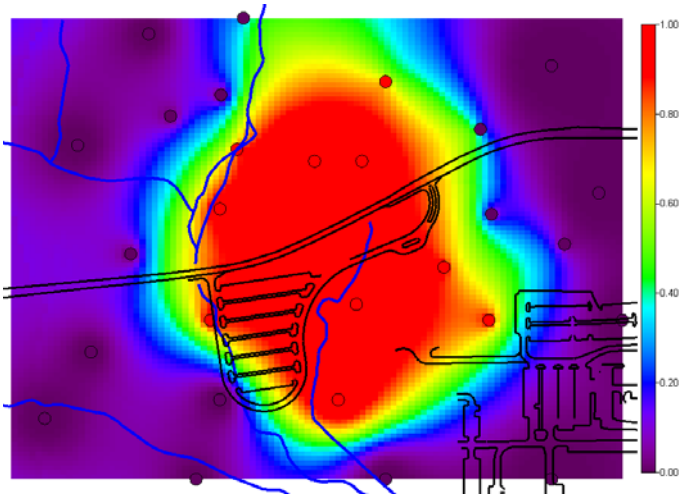
SADA is equipped with a suite of geospatial models that permit you to spatially analyze your point data, produce various interpolations, and quantify uncertainty about those outcomes. SADA's strength lies in the integration of these models with human health risk and decision frameworks that can directly inform and drive remedial strategies and sampling designs.

Overview

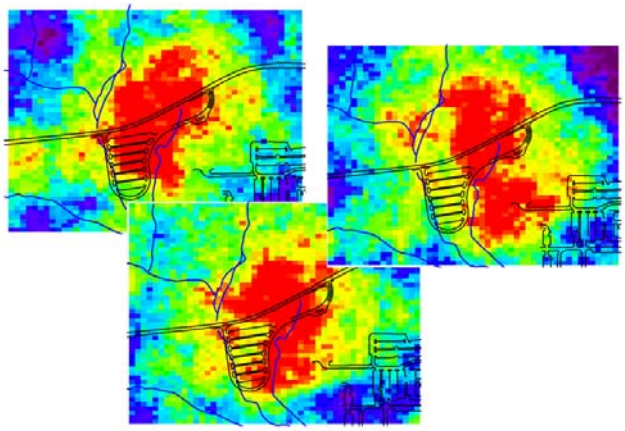
The term "geospatial methods" refers to a class of mathematical and statistical tools for characterizing how data vary in space. A common form of geospatial analysis is interpolation, where point data are transformed into a continuous map by estimating values at unsampled locations.



Of course, the estimated values at unsampled locations are only predictions. There is therefore some uncertainty about the exact value. Some geospatial methods can quantify this uncertainty in estimation. There are many different ways to express uncertainty. In the following image, we see uncertainty expressed as the probability that a regulatory limit of 3pCi/g has been exceeded.

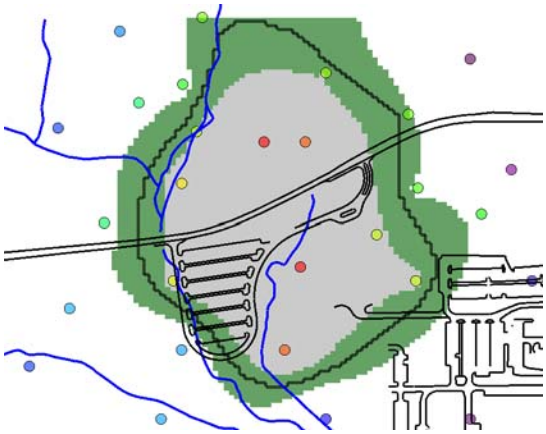


In the following image, uncertainty is expressed as alternate and equiprobable interpolations of concentration values across the site.

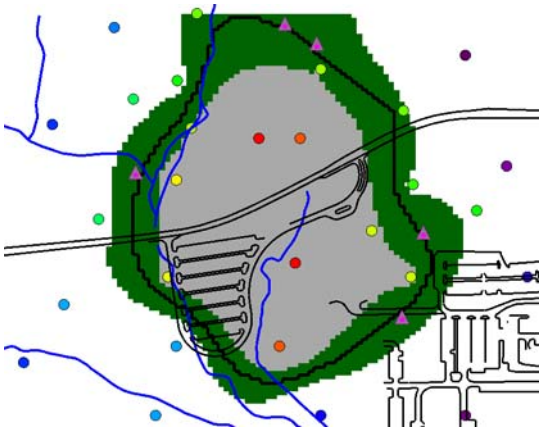


These alternate realizations can be used to visually observe uncertainty in modeling. They can also be used as multiple inputs into other models (e.g., groundwater modeling) to propogate uncertainty through the complete modeling effort.

Uncertainty in estimation can also be propagated through the decision models, creating areas of concern where uncertainty is found in the exact location of the boundary.



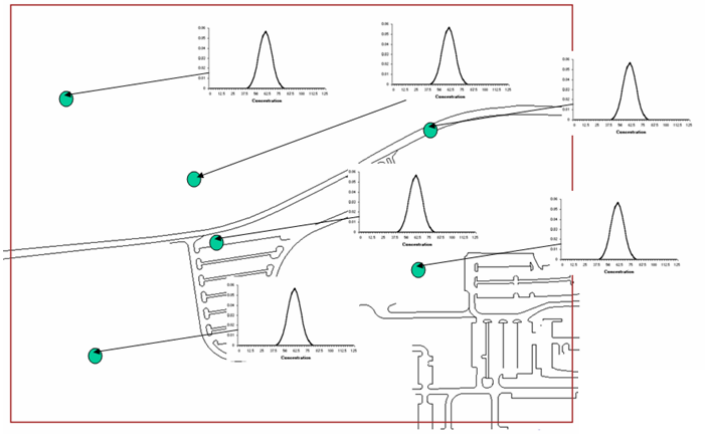
Uncertainty can also drive secondary sampling designs. In the following image, we see five new sample locations optimally positioned to better delineate the exact location of the boundary.



These are only a few representative outcomes of a geospatial analysis. We will provide discussions of the methods used to generate these kinds of results in SADA. The applications of spatial tools are too numerous to cover in this discussion, however. You are therefore

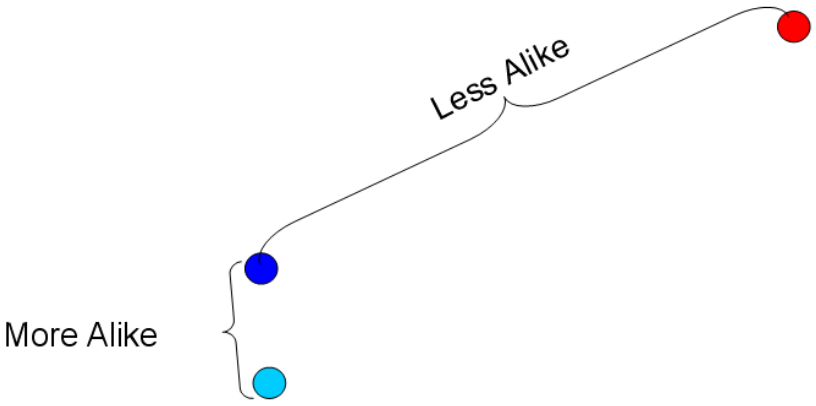
encouraged to supplement this guide by incorporating other sources. We recommend *Applied Geostatistics* by Isaaks and Srivastava or *Geostatistics for Natural Resources Evaluation* by Pierre Goovaerts as excellent sources for complete but tractable explanations of these methods.

In SADA, we divide the suite of geospatial tools into two groups: basic and advanced. Basic tools are deterministic models. These models include nearest neighbor, natural neighbor, and inverse distance. These models produce a single estimate for each unsampled location. Advanced methods refers to geostatistical methods, including correlation structure analysis, ordinary kriging, indicator kriging, cokriging (and variations), sequential gaussian simulation, sequential indicator simulation, and a variety of post-processing methods for synthesizing results. These methods produce a distribution of possible estimates for each unsampled point. This is demonstrated for a few unsampled locations in the following figure.

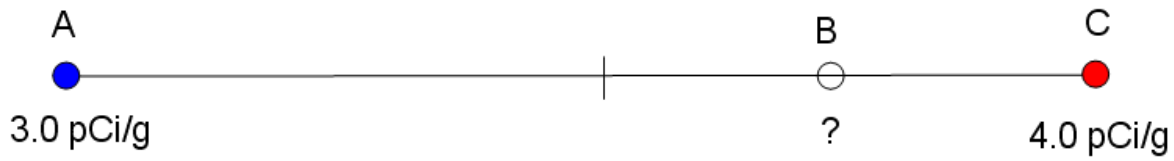


At their most fundamental level, all methods presented here have a similar approach at their core. That is, the estimation of values at unsampled locations is a function of the surrounding neighbors.

For the sake of discussion, in simplest terms, the estimation at any given point is a weighted average of the surrounding neighbors. An important distinction between these methods is how those weights are determined. In general, methods tend to assign more weight, and therefore more influence, to those data values closer to the unsampled point.



As a trivial example of what we mean by weighting neighbors, consider the following. Suppose two data points are separated by some arbitrary distance as in the following figure.

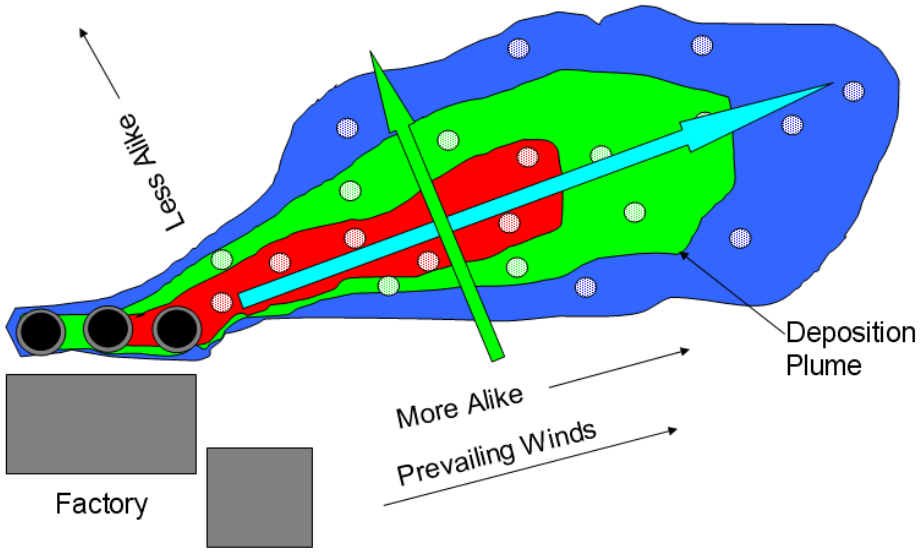


Suppose that we wish to estimate the value at the unsampled location B. One method would be to use an unweighted average, where we simply add the data up and divide by 2. This would yield a value of 3.5 pCi/g. Suppose instead, we wish to weight according to the proximity to the unsampled location. If the unsampled location is 75% of the distance from point A, then we will weight it by (1-0.75) = 0.25. Point C will also be weighted by its distance: (1-0.25) = 0.75. In this new estimation, sample C will receive more weight and therefore more influence in the estimation than sample A. Specifically,

B = 0.25 x 3.0 + 0.75 x 4.0 = 3.75pCi/g.

Isotropy/Anisotropy

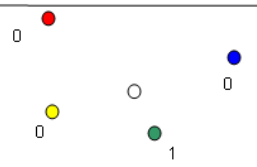
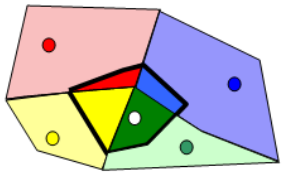
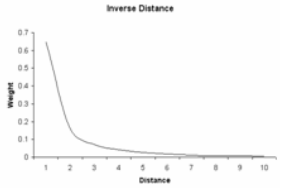
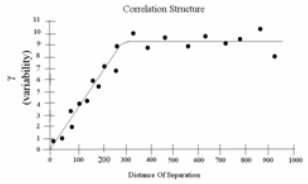
If the underlying spatial process depends only on separation distances, then it is referred to as an *isotropic* process. Under isotropy, data do not tend to be more alike in any one direction than any other. This may not always be true. Depending on the cause of contamination, data may be more alike in certain directions than in others. Consider the following air dispersion example.



In this example, deposition from three smokestacks tend to preferentially deposit contaminants in the direction of prevailing winds. One should expect then that data found along the N-E transect will be more alike than data found along the N-W transect for the same distance. Anisotropy is an important factor to detect and use in a geospatial model. This will directly affect the weight selection. Data found along the N-E transect would receive more weight than an equally distanced data point along the N-W transect.

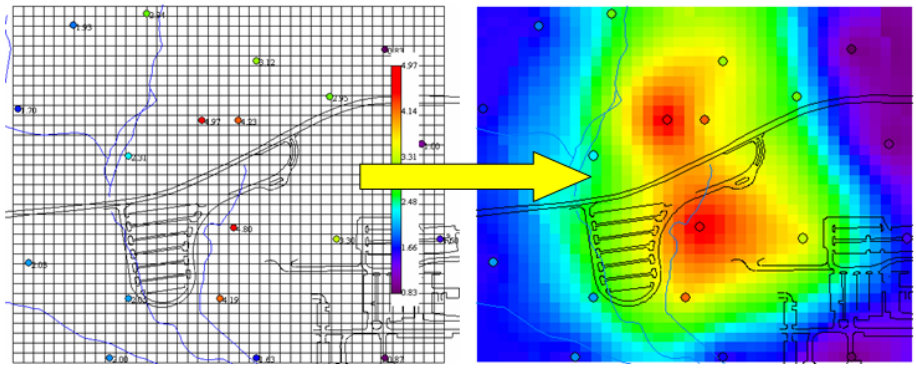
SADA has a number of geospatial models. In nearest neighbor, an unsampled point is equal to its nearest neighbor. In this formulation, all weight is given to the nearest neighbor and zero

weight is given to all other points. In natural neighbor, the weight depends on how much area each sample point is representing. In inverse distance, the weight given to any neighbor is simply based on the distance between the unsampled point and each neighbor. In geostatistical formulation, weights are also based on distance. Rather than simple geometric distance calculation, however, weights are based measures of variability over distance. The following table provides a summary with some visual representation of how weights are selected.

Nearest Neighbor	W_i = 1 if closest W_i = 0 otherwise	
Natural Neighbor	W_i = F(Overlap of Area of Influences)	
Inverse Distance	$w_i = \frac{h_i^{-p}}{\sum_{k=1}^{} h_k^{-p}}$	
Geostatistics	W_i = F(correlation structure) (correlation structure describes spatial variability and can be measured)	

Grids

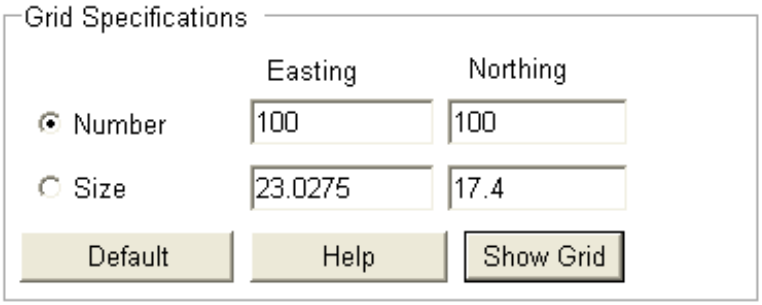
All geospatial methods presented here require the definition of a grid. A grid is a regular lattice overlaid across the site. The center of each grid cell will be the focus of interpolation. The entire cell is then associated with this interpolated value.



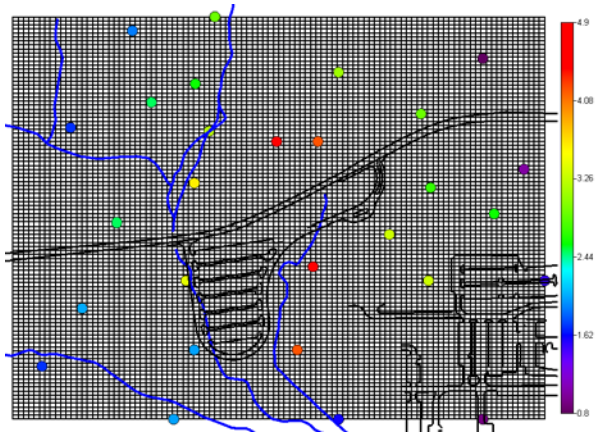
Selection of grid size will depend on the application. In some situations, each cell may represent a remedial unit, an exposure unit, or may only be selected to produce a high resolution image.

If you have not already done so, open SADA and open GeospatialOverview.sda. Select *Ac-225* as the data type and choose *Interpolate my data* as the interview. Now click on the *Set Grid Specs* step. Here is where you will specify the grid density used by all interpolation and simulation methods that follow.

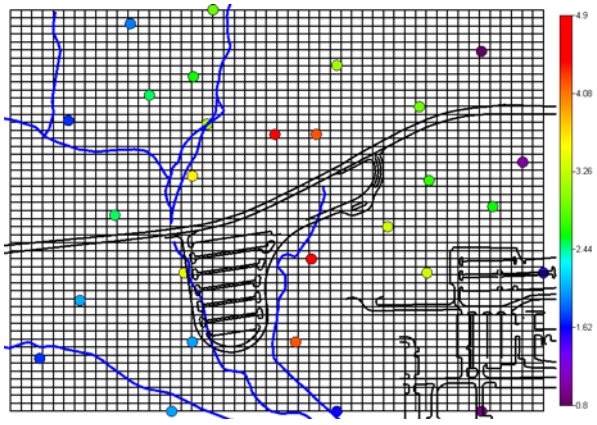
There are two ways to specify the grid: 1) by number of cells in each direction or 2) by cell size.



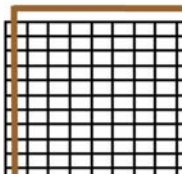
In the figure above, we are specifying that 100 cell blocks in each direction should be used. SADA then uses the length of the site boundary in each direction, divides this by 100 and then enters the size values into the size row. Select *Number* and enter 100 for each direction. Press *Show Grid*.



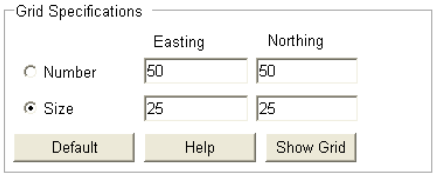
Try now changing the number in each direction to 50. Press *Show Grid* again. The resulting grid is much coarser. You can control how fine the resolution is by controlling the number of cells in each direction. There are practical limits to the number of cells SADA can reasonably control. Grid resolutions greater than 200x200 are difficult to visually distinguish from each other.



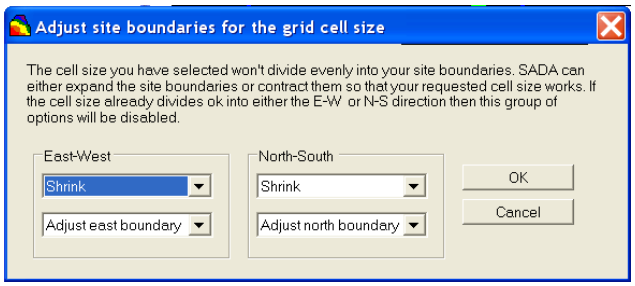
Grids can also be created with the *Size* option by entering a size for the easting and northing direction. There may be an additional complication, though; if the size you specify does not evenly divide into the site boundary extents, a decision must be made about how to resolve the conflict. Without a resolution, the last cells will either not reach the boundary or will slightly exceed this.



To resolve this, you can determine a cell size that will divide evenly into each direction. This might take some calculations on your part. Another solution is to slightly expand or contract the grid so that it fits perfectly. In the grid specification window, select *Size* and enter 25 into both *Easting* and *Northing* cells.

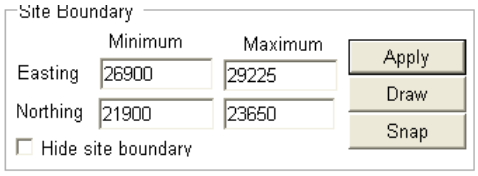


Press Show Grid and SADA presents the following window.

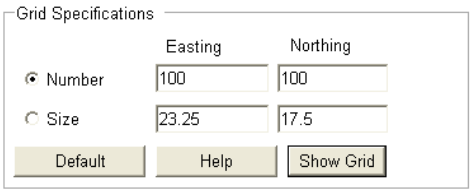


In this window, you are informed that the cell size you selected does not divide evenly into one or both directions. For each direction, you must choose whether to shrink or expand the site boundaries. Be advised that if you choose *Shrink*, you may inadvertently locate a sample outside the study area. Once you have decided, you must also choose which boundary to adjust. In this example, select *Expand* for both *East-West* and *North-South*. For the *East-West* direction, select *Adjust west boundary*. For the North-South direction select *Adjust north boundary*. Press *OK*.

Nothing dramatic seems to happen. Indeed, only a small adjustment is made, and you may not even visually notice it. You can check on the new site boundary extents by clicking on the *Set up the site* step.



You are now ready to explore some Interpolation methods. For the rest of the exercises, choose *Number* under *Grid Specifications* and use 100 x 100.



Press *Show Grid* and save the SADA file. In the next chapter, we will introduce basic geospatial methods.

Chapter 29: Basic Geospatial Methods

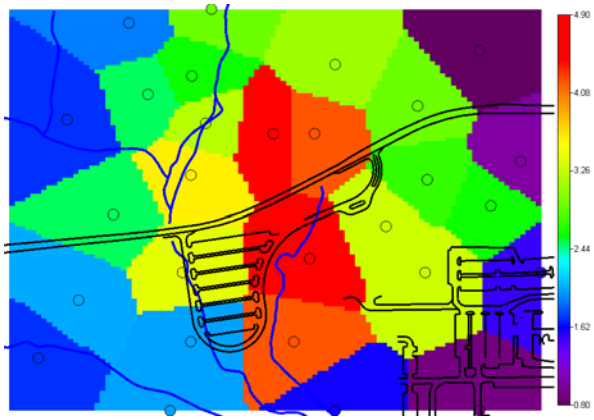
Basic methods are essentially non-geostatistical approaches to modeling. Each of these methods will produce a single outcome for a single set of inputs. These methods are, however, relatively easy to implement. Begin by opening the file named GeospatialOverview.sda.

Nearest Neighbor

Nearest neighbor is the simplest method found in SADA. In this approach, the estimation point is equal to the value of the nearest sample. From a weighted average view, this can be written in the following form:

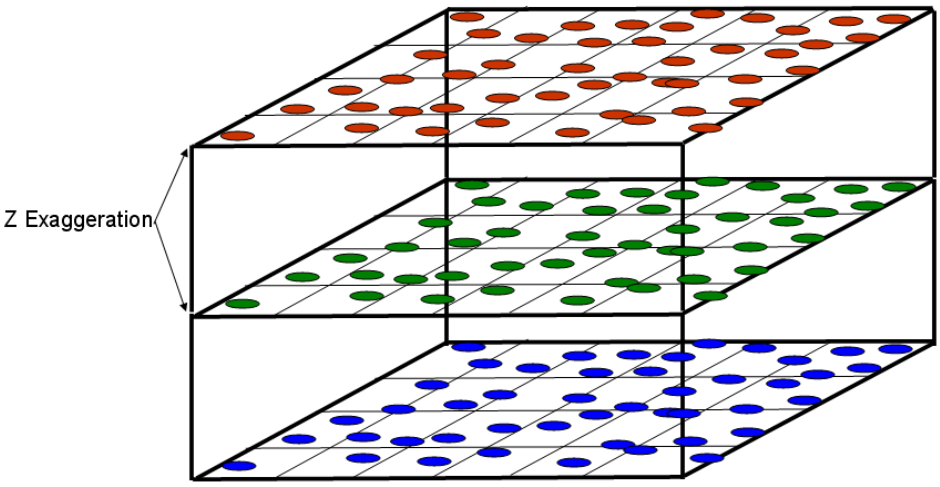
$$u_0 = \sum_{i=1}^{N} w_i u_i \; where \begin{cases} w_i = 1 if \; u_i \; is \; closest \; to \; u_0 \\ w_i = 0 otherwise \end{cases}$$

Nearest neighbor produces a jigsaw type contour map that may not be suitable for many spatial estimation goals. Each jigsaw piece is actually the *area of influence* for the data point it contains. The area of influence contains all points that are closer to the reference sample than any other point. This geometry plays important roles in a variety of applications, such as designating terrestrial habitats and creating point buffers in GIS. Area of influence also plays an important role in the Natural Neighbor method, which will be presented next.



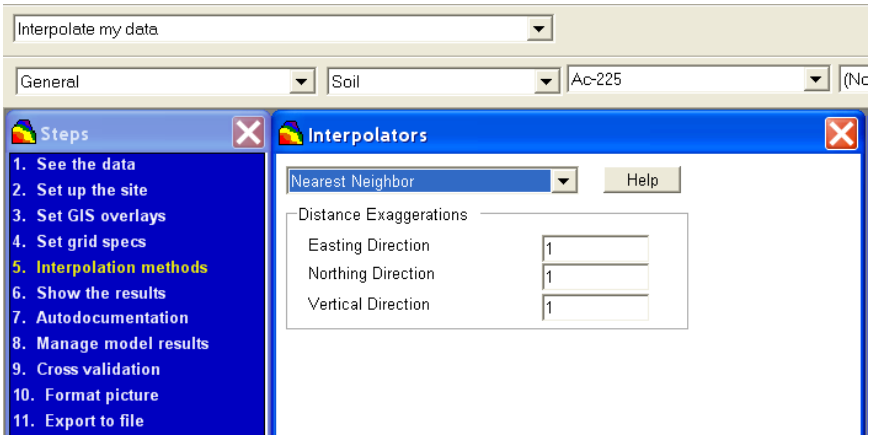
One useful application of nearest neighbor is the gridding of extremely dense data, such as geophysical results. Storage of point geophysical data in SADA is an extremely inefficient approach, from a data management perspective, and can result in dramatically reduced code performance. It is better to use an extremely dense grid of geophysical data rather than point data. Even though the number of grid nodes and the number of points may not differ significantly, the manner in which grids are stored yields much faster response times. The procedure is simple. First, create a temporary SADA file where you bring your geophysical data into SADA as point data, grid it, and then export it out as a gridded file. You can close this temporary file and create a new file that you will actually use. Import the gridded file as gridded data and SADA's response times should markedly improve.

Because subsurface data are often more heterogenous in the vertical direction than in the horizontal direction, you can specify a Z exaggeration that will increase vertical distances, allowing nearest horizontal neighbors to be selected in 3d grids. This may be particularly important for gridded geophysical data. Horizontal exaggerations are available, as well, if needed.

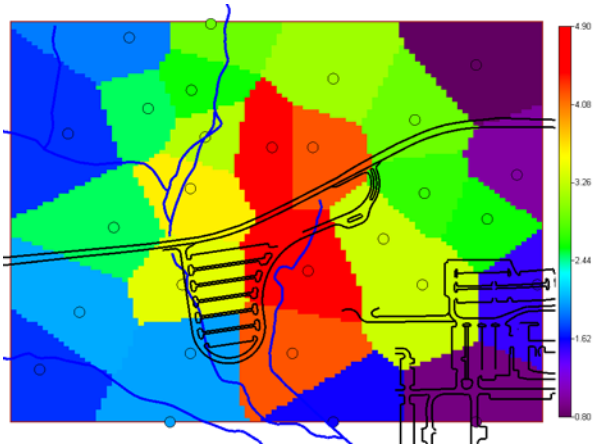


Nearest Neighbor in SADA

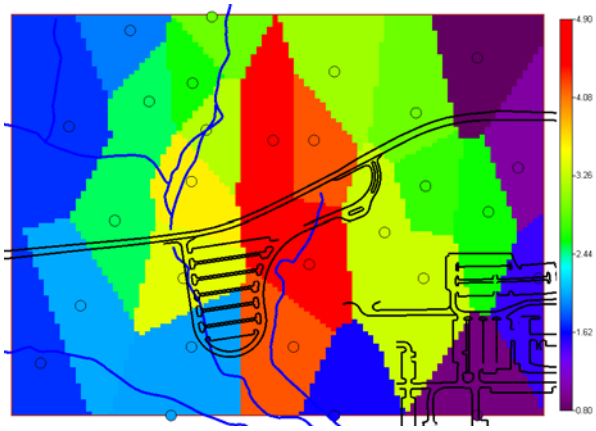
If you have not already done so, open the file GeospatialOverview.sda. Select *Ac-225* as the data type and choose *Interpolate my data* as the interview. In the steps window, select *Interpolation methods*.



Select *Nearest Neighbor* from the drop-down list and set all *Distance Exaggerations* equal to 1 (essentially specifying no exaggerations). The grid specs should be set to 100 x 100. Check the *Set grid specs* to confirm. Press *Show The Results*.



To see the effect of exaggerating one direction or another, change the *Easting Direction* parameter to 2. This will make the natural neighbor code use twice the real value for the easting direction distances. This essentially gives greater influence to samples found to the north and south of any given unestimated point.

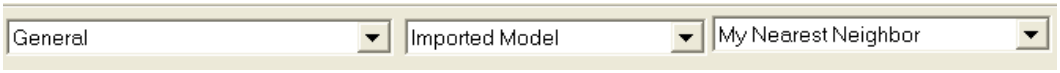


You can store this result and recall it at a later time. Click on step 8, *Manage model results*. In the parameters window, press the *Save Model Result* button. You will be asked to enter a name. Enter “My Nearest Neighbor” and press *OK*.

When you store a model result, SADA views this as the addition of a new data set. It will reset itself to the first data type/data set it encounters. This is an inconvenience, but no harm is done. Select *Soil* and *Ac-225* and then *Interpolate my data*, as before. In the *Interpolation methods* step, select *Stored Result*. You will see previously stored models in a new drop-down list.

If you press the *Show The Results* button now, SADA will recall the model and plot it. This is the recommended way to use SADA when performing dense modeling or simulation. Store the result and then it becomes more efficient to manage.

It also appears in another location. In the data type drop-down list, select *Imported Model*. This data type also stores any models created within SADA. You will see your result there in the data set window as well. Now you can access a variety of functions on your model, including map algebra.



Save the SADA file.

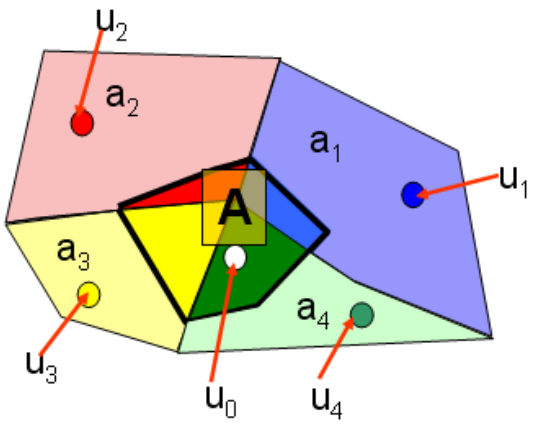
Natural Neighbor

In *Natural Neighbor*, it is not just the nearest neighbor that is allowed to influence the estimation. When data points are distributed in space, they inherently represent a certain area around them. Sample points in sparsely sampled regions must represent a larger region than those in more densely sampled areas. These regions are called *area of influence*, and it is possible to draw simple geometries that bound them.

In *Natural Neighbor*, the data areas of influence are calculated first and the area of influence for the point being estimated are overlaid. This creates regions of overlapping areas of influence. For any given sample point, the fraction of the overlap becomes the weight assigned to that sample point. Formally written, the equation is:

$$u_0 = \sum_{i=1}^{N} \frac{a_i}{A} u_i$$

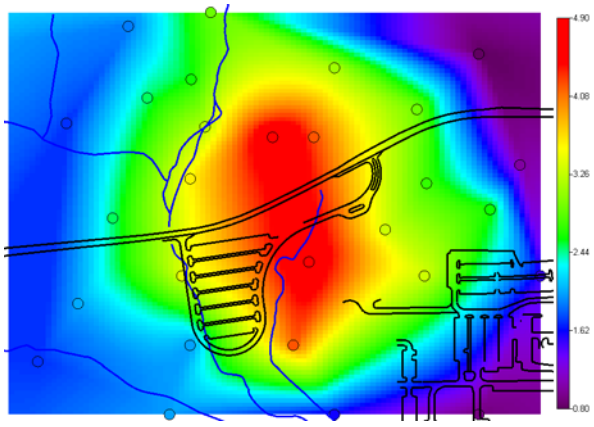
N is the number of intersecting areas of influence, A is the area of the area of influence for u_0, and a_i is the area of overlap between u0 and ui. See graph below:



This method was first presented in Sibson (1981) and further work can be found in Watson (1999).

Natural Neighbor in SADA

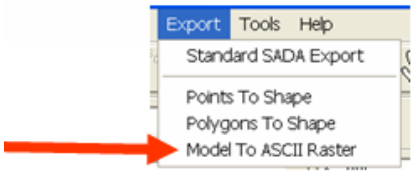
If you do not already have GeospatialOverview.sda open, open it now and select *Soil* and *Ac-225*. From the interview drop-down list, select *Interpolate my data* and click on the *Interpolation methods* step. In the parameter window, select *Natural Neighbor* and press *Show The Results*.



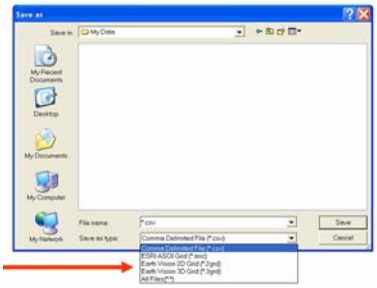
Natural Neighbor is currently limited to 2d applications. By 2d, we mean every sample has exactly the same z as recorded in SADA. So, soil surface data with depths of 0, 2”, and 6” would be considered 3d; to use natural neighbor in this situation, you would need to assign a single depth value for each point outside of SADA and then import the data.

You can also store this model in SADA by following the discussion found under Nearest Neighbor. Additionally, you can export this model (and all models) as an ASCII raster dataset that can be imported into another GIS package.

To export this result to an ASCII Raster, select from the main menu *Export→Model to ASCII Raster.*



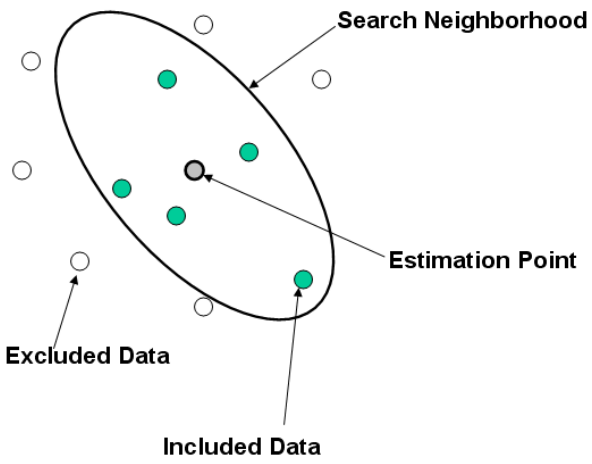
In the file name field, enter MyNaturalNeighbor and press *Save*. You can now open a GIS package, such as ESRI ArcGIS, and import the ASCII raster or convert it. SADA has other export options as well. These are discussed in a later chapter. To see them here, press the *Export to Text* button and look at the drop-down list under *File Types.*



Inverse Distance

The previous two methods were based on simple geometries: closest neighbor or area of influences of surrounding neighbors. In both cases, little is required by the user. Inverse distance requires more effort from the user, but in return it is a more flexible tool that can account for anisotropic conditions. Anisotropy occurs when data tend to be more alike in a particular direction than another.

Inverse distance, as well as most geostatistical methods, requires the specification of a *search window*. A search window is an elliptical (2d) or ellipsoidal (3d) geometry centered at the point of estimation. All points within the boundary of the ellipse will influence the estimation. All points outside the boundary will be excluded for the current estimation. The following figure shows a search window or *search neighborhood* (terms may be used interchangeably).



In this figure, all data outside the search window are not included in the estimation and are represented as hollow circles. All data within the window are included. Here, we see the specification of a search window that accounts for anisotropy in the N-W direction. This is accomplished by the use of an ellipse rather than a circle. By stretching the ellipse out in the N-W direction, data further along this transect will be included while data nearby along the N-E transect will be disregarded. If isotropic conditions prevail, one should use a circular search window. The specification of the search neighborhood is intuitive but at the same time very detailed. This discussion will also be necessary for geostatistical models.

The weighting scheme for Inverse Distance is as follows:

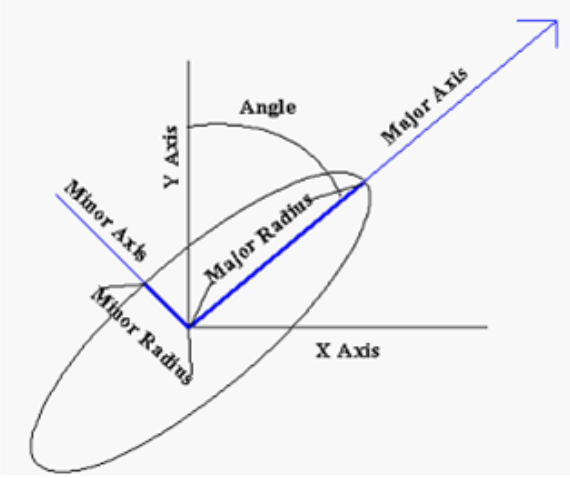
$$u_0 = \frac{\sum_{i=1}^{N(u_0)} \frac{1}{d_i^P} u_i}{\sum_{i=1}^{N(u_0)} d_i^P}$$

Where d_i is the distance between sample location u_i and estimation point u_0 and $N(u_0)$ is the number of samples used within the search neighborhood of u_0. P is the power parameter. High power values will result in dramatic weight reduction over shorter distances. P values of 3, 4, and on up to 7 or higher will result in highly localized estimation maps. Low power values will permit samples further away to have more influence. Most of the time, a power value of 2 is used. In fact, inverse distance is often referred to as inverse distance squared, for this reason. Now we need to form the search neighborhood.

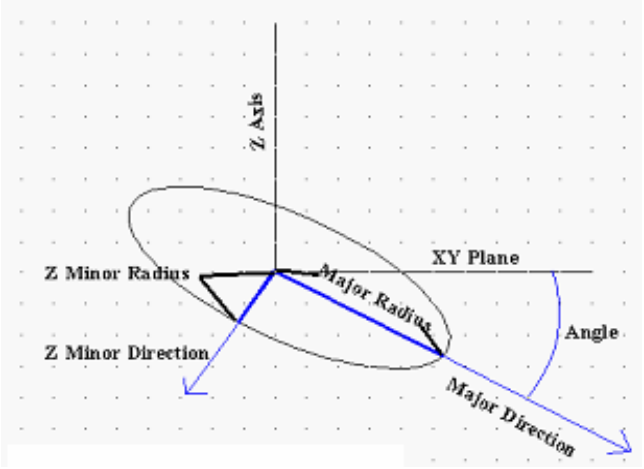
Specifying the Search Neighborhood

Defining the search neighborhood largely amounts to setting the parameters of an ellipse or ellipsoid. We will begin first with the two dimensional search window, an ellipse geometry.

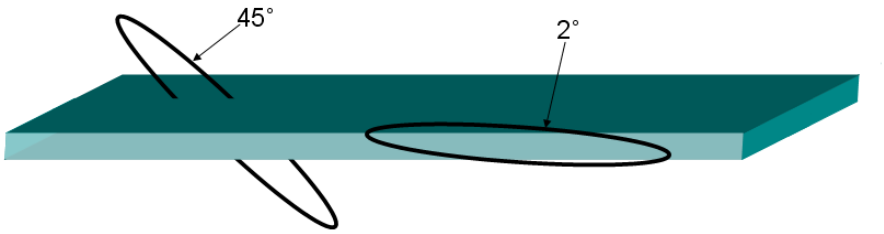
An ellipse is defined by two radii and a search angle. The *major radius* is the long radius of the ellipse and falls on the *major axis*. The *minor radius* is the short radius of the ellipse and falls on the *minor axis*. The *horizontal search angle* is the rotation of the major axis measured clockwise from due North.



For three-dimensional applications, an ellipsoid is required. Ellipsoids have three additional parameters that complete the 3d geometry. The *Vertical Search Angle* defines how the plane of the ellipse dips below the surface. The *Vertical Search Radius* defines the thickness of the ellipsoid at its center. The *Rotation Angle* defines how this 3d geometry is rotated about its major axis.



For 3d applications, search neighborhoods need to be carefully defined. Geographically speaking, most sites will be very wide and very thin. Consider, for example, a site that is 800m x 800m by 10 meters. The relative thickness is very small. If you select an angle of 45 degrees, it will likely shoot out of the site very quickly. Angle choices of 0.5 – 5 are usually more appropriate.



Now that the geometry is defined, additional rules can be defined about the minimum and maximum number of data to use within this search window.

The *minimum number of sampled data* parameter specifies the minimum number of data required before estimating the concentration. If the minimum is not met, SADA returns an unestimated value. These values become empty spaces in the plot.

The *maximum number of sampled data* parameter specifies the maximum number of data to use in estimating a point.

Helper data refers to secondary forms of information that can be used to infer concentration levels. Examples of helper data include geophysical data, field detection methods, or geophysical surveys. Inverse distance does not use this type of information, but in the interest of completeness, we will discuss these here. The cokriging and co-simulation routines discussed later will require them. For inverse distance, they are disabled.

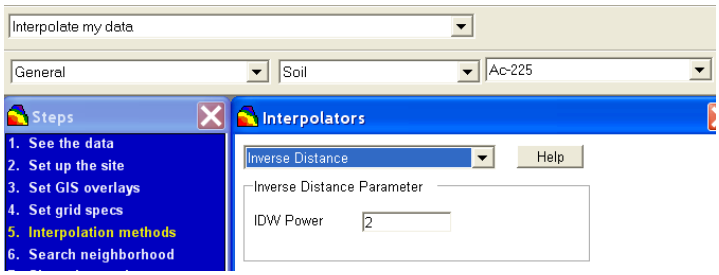
The *minimum number of helper data* is the minimum number of helper data required before estimating the concentration.

The *maximum number of helper data* is the maximum number of helper data to use in estimating a point.

In geostatistical simulations, previously simulated values can also be used as conditioning data. The *number of simulated nodes* parameter specifies the maximum number of simulated results to use in conditioning data. Inverse distance does not use this parameter and it will be grayed out.

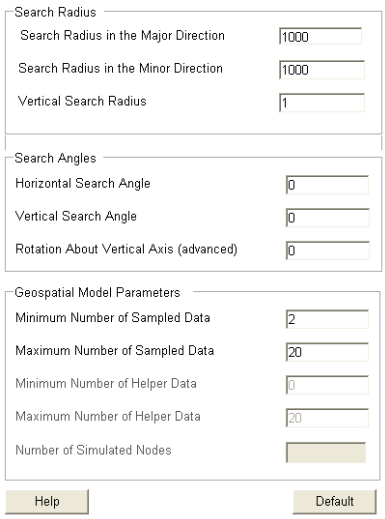
Inverse Distance in SADA

If you have not already done so, open up SADA and open up the file GeospatialOverview.sda. Select *Soil* and then *Ac-225*. Select *Interpolate my data* for the interview, click on the *Interpolation methods* step, and select *Inverse Distance*.



Directly under the *Inverse Distance* selection is the IDW Power parameter. Recall that high power values result in highly localized estimations. Low power values result in more regionally smooth estimation maps. For now, we'll leave the value as 2. Later we'll show the impact of increasing this power parameter on the estimation map.

Click on the step *Search neighborhood* to specify a search neighborhood.



SADA provides a default button in the lower right hand corner. This button will automatically fill in some parameters based on basic rule of thumb estimates. If you press this button, SADA will provide a message box about those default values; these default parameter values are by no means optimized for your data set. They should only be used to get you started.

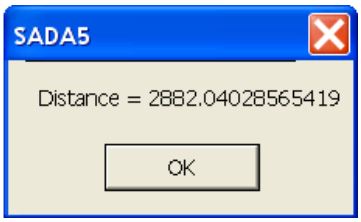
Let's begin by using the following values for the search neighborhood. The radii should be in whatever coordinate system you are using (m, ft, etc)

Parameter	**Value**
Major Search Radius	1000
Minor Search Radius	1000
Vertical Search Radius	1
Horizontal Search Angle	0
Vertical Search Angle	0
Rotation Angle	0
Min Number of Samples	2
Max Number of Samples	20

The obvious question is how to choose appropriate values. Unlike geostatistical modeling, which empirically estimates parameter values directly from the data, inverse distance does not provide a method for parameter estimation prior to modeling. There are ways to improve parameter values based on the behavior of the final estimation map, however.

One way to approach search radii is to measure the distance across your site. First, click on the *Set up the site* step. Make sure that the option for hiding the site boundary is unchecked. This will show you the boundary box. Press the measurement button  on the main toolbar. Most of SADA will become disabled. Click on the lower-left-hand corner of the boundary box and draw a

line to the upper-right-hand corner of the boundary box. SADA will report the distance between these two locations.



Note that due to variations in mouse control, you may have a slightly different number. This is OK, as we only need a rough estimate of this distance. As a rule of thumb, begin with a *Major Search Radius* of about ¼ to ½ of this distance, depending on how much data you have. In our case, we will start with 1000.

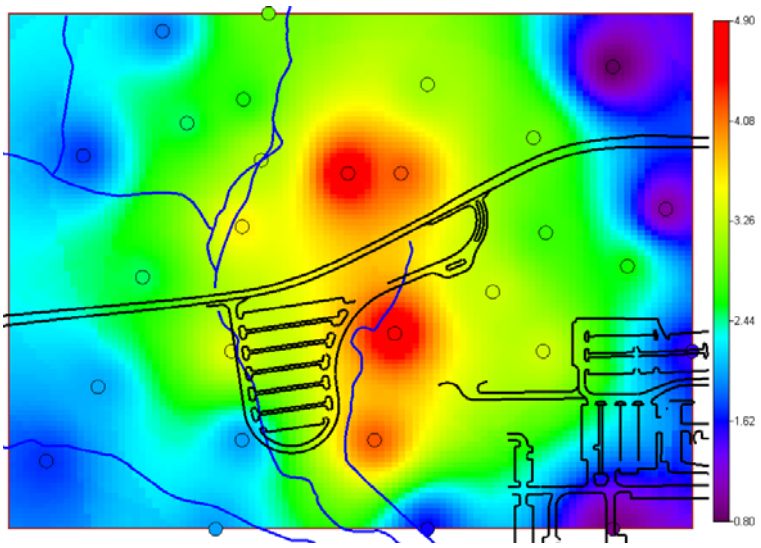
The choice of *Minor Search Radius* will depend on what you see in the data. If there is an observable trend (anisotropy), then you will want to choose a narrower search radius, so that data further away but along the trend can be used. At the same time, you will want to limit or eliminate data that are opposite or orthogonal to the trend direction. A smaller *Minor Search Radius* will accomplish this. For *Ac-225*, there is an obvious N-S trend in the data. If you trace the site from N-S along any portion of the site, the data vary less than if you trace the site from W-E. For demonstration purposes, let's ignore this trend and choose a *Minor Search Radius* of 1000 as well. Shortly, we will choose a more appropriate value and the effect will be made more apparent.

Since this is a 2d data set, the *Vertical Search Radius* isn't necessary. A value of 0 or 1 will suffice. For 3d data sets, the *Vertical Search Radius* will likely be very small compared to the horizontal radii. We will tackle this problem in a 3d example next.

The *Horizontal and Vertical Search Angles* aren't important yet, because we are choosing a circular search window in a 2d data set. Since the search window is circular, it doesn't matter which direction the major axis is pointing. For now, let's use a value of zero for each.

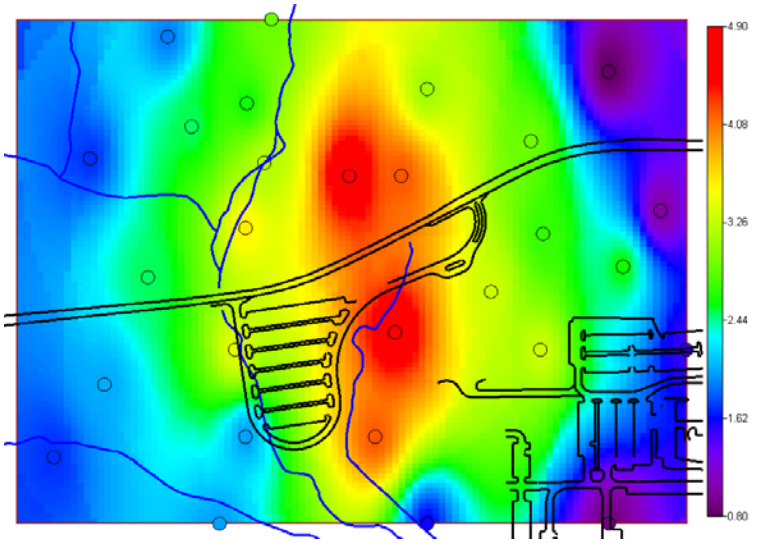
The *Min and Max Number of Samples* are set to 2 and 20, respectively. For the minimum number of data, for small data sets, you may be forced to use small values such as 1, 2, or 3 in order to render a complete image. Be advised, though, that if only one data point appears within a search window, the estimate is suspect and certainly unstable (see cross validation later). Unless you have extremely dense data sets, the maximum data to use should be very high. Computer speeds should not normally be a problem, permitting you to use a rich set of conditioning data whenever possible.

Return to the step *Search neighborhood* and press the *Show The Results* button.

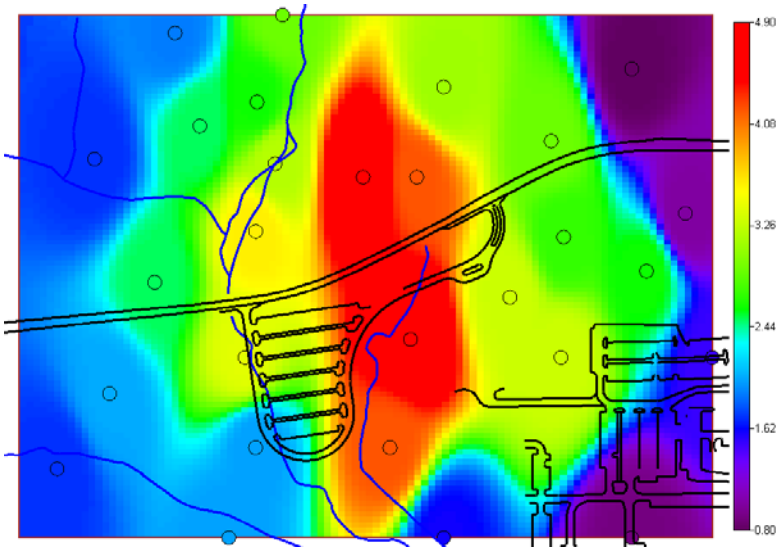


The resulting map reflects the circular search neighborhood and the failure to account for the N-S trend in the data. Notice that estimates, particularly along the boundary, show trends that are similar to the search window geometry.

Let's now acknowledge the N-S trend and choose to use a smaller *Minor Search Radius*. Set the *Minor Search Radius* to 600. Now, the *Horizontal Search Radius* does matter. The data appear to be in a N-S trend, so a *Horizontal Search Angle* of zero would be appropriate. Had the trend been in the E-W direction, a *Horizontal Search Angle* of 90 would be appropriate. Press the *Show The Results* button.



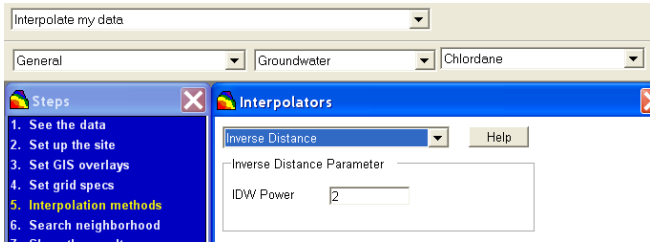
Inverse Distance now more accurately accounts for the observable trend. As a demonstration, click on the step *Interpolation methods* and use a value of 5 for the *IDW Power*. Press *Show the Results*. Notice how there is less smoothing and estimation becomes more localized. This can be very useful in data sets where you have relatively low values and one or two outliers.



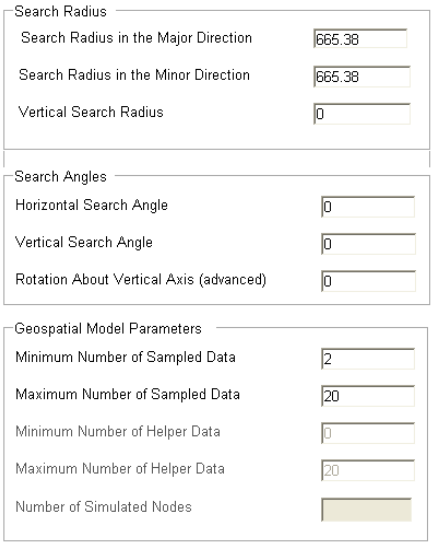
Save your file and we'll use inverse distance on a 3d data set.

Three Dimensional Modeling

Close any file you are currently working on and open GeospatialSubsurface.sda. Select *Groundwater*, *Chlordane*, and *Interpolate my data* as the interview. Let's first look at how the subsurface is divided up. Click on *Set up the site* and note that there is a 10 Layer Design for vertical layering that sets a new layer every 10 feet. Now, click on *Interpolation methods* and choose *Inverse Distance*. We will use a power value of 2 for now.

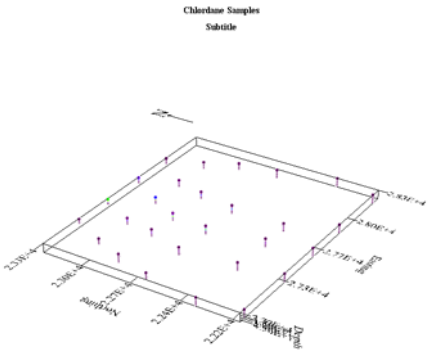


Click on *Search neighborhood* and press *Default* in the lower-right-hand corner. Click *Yes* to the message. The following parameters are entered for you:



SADA enters a value of zero for the *Vertical Search Radius*. Determining the vertical search radius really cannot be done by a rule-of-thumb approach. Usually, there is great heterogeneity along the vertical transect. It is important not use a large vertical radius that completely averages out the entire estimation. The choice for vertical search radius should be motivated by the observed vertical behavior.

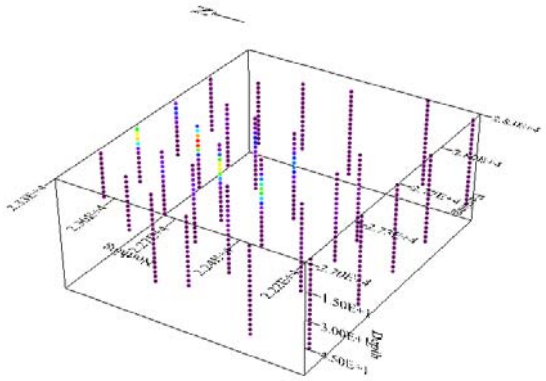
There are two ways to see how data are behaving vertically. First, turn on the 3d viewer. From the main menu, select *Graphics→Show 3d view*. This brings up the 3d viewer. The 3d viewer is a fairly involved subject, and you are encouraged to read Chapter 44, “The 3d Viewer.” For now, some basic explanations will help. Your 3d view should start out similar to the following (you may have a black background).



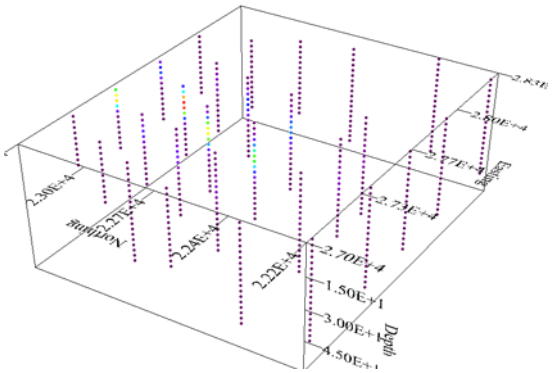
The first thing to note is how flat the site is. This is very common, as horizontal extents will be much larger than vertical extents. This circumstance makes it difficult to see what is going on in the boreholes. Click on the step *3D Viewer Controls*.

NOTE: If your 3d image disappears, simply mouse click anywhere in the image and move the mouse slightly. Some video cards will lose the image during refresh events.

In the parameter window, select the *Basic* tab and press *Bkgd Color*. Choose *White*. Click on the *Scaling* tab and change the *Z* parameter to 10 and hit *Enter*. This will expand the vertical scale by a factor of ten.



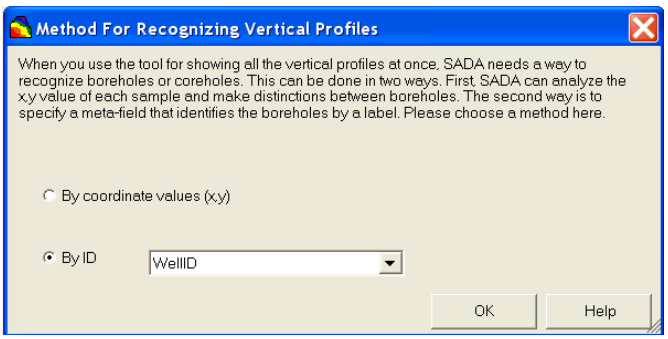
Return to the *Basic* tab, and press the *Zoom In* button until the data set appears closer to you.



Now right-mouse-click anywhere on the image and hold the button down while moving the mouse around. The image will rotate. Play with this feature and observe how your corehole data is behaving vertically, particularly in the area with higher values on the north side.

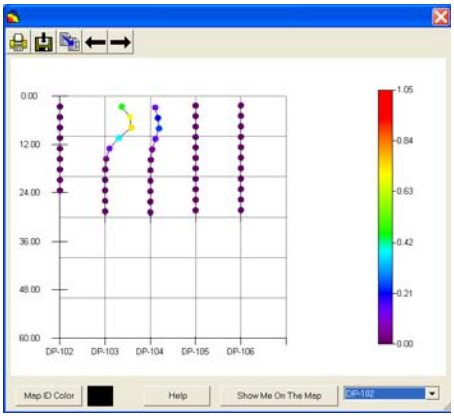
The 3d viewer is useful in getting a sense for how your data vary as a function of depth. You can also use the 2d viewer (used thus far) and change layers to show one layer at a time. This can make it more difficult to see certain trends. The 3d viewer can give you a general sense of spatial distribution, but determining a vertical range is not easily done here. For a more specific look at vertical variability, we will use the *Vertical Profiles* tool.

Select *Graphics→Show 3d Viewer* to turn off the viewer. Select *Tools→Show All Vertical Profiles→Set Vertical Profile ID Method*. SADA can identify profiles by one of two methods: 1) by comparing X,Y coordinates of each data point or 2) by using a well id field (or something similar).



If you select *By coordinate values (x,y)*, be advised that a data point will be associated with other data points as a single core or well only, if the x,y is exactly the same. If there is the slightest variation, SADA will recognize the point as a separate albeit very close well. In this example, we have a field that denotes the *WellID* for each point. Select *By ID*, choose *WellID* and press *OK*.

From the main menu select *Tools→Show All Vertical Profiles→Show profiles*. SADA responds with the following window.

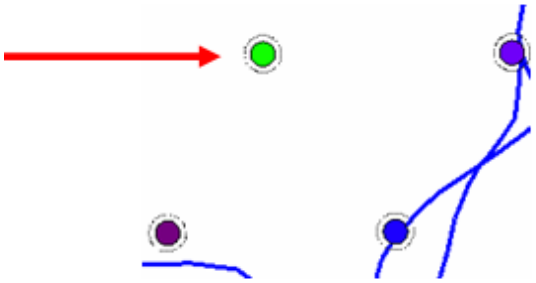


The vertical profile preview window has many features, which are covered in a later chapter. For now, we will concentrate only on determining the vertical search radius.

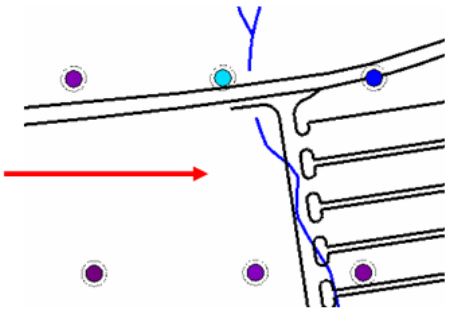
If you observe the first 5 wells, notice that in DP-103 there are significant changes over the first layer, or about 10 feet. Using the advance/rewind arrows ←→, move forward to see DP-107 to DP-111. Notice that when substantial variation does occur, it is over about a 10 feet interval. This holds true for most wells. As a rule of thumb, we may want to consider using about ¼ to ½ the distance observed in variation, or about 5 feet. Press the X button in the upper-right-hand window of the profile box. Enter a value of 5 for the *Vertical Search Radius*.

In this data, there appears to be a N-S trend, particularly in the middle of most of the top layers. The E-W transect of this trend appears to be about 500 feet across (use the measuring tool and measure distance across high value areas). We'll use this value for the minor search radius.

The N-S trend appears to increase in depth as it moves south. This is common. To measure this trend, we'll use the measuring tool again, but we'll need to change layers mid-measurement. For the current layer, select 0-10. Select the measuring tool button. Click on the green data value at the far north.

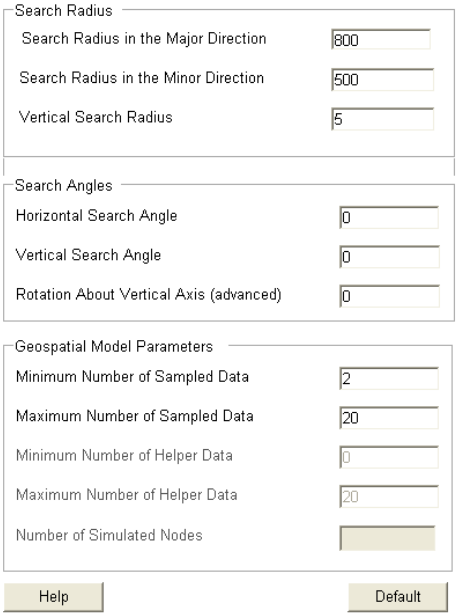


Now, with the measurement line following you, change the vertical layer to 20-30 and click about halfway between the light blue and purple data values, as in the following image.



This distance is the N-S distance measurement for the plume as it moves both south and down. The measurement is about 800ft. We'll use that for our major search radius.

To account for the vertical dip, we will use a *Vertical Search Radius* of about 5°. The *Horizontal Search Angle* should be set to zero, since the trend is north-south. *Rotation* will also be zero. We will set the *Major Direction* to 800, the *Minor Direction* to 500, and the *Vertical Search Radius* to 5. The *Minimum Number of Sampled Data* should be 2, and the *Maximum Number of Sampled Data* should be 20.



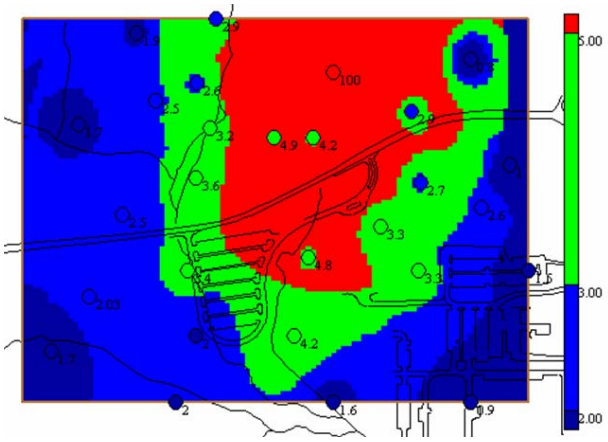
Save your file and press *Show The Results*. In the 2d viewer, you will see a reasonable first start at modeling the contamination with inverse distance. From this point forward, parameter values can be further refined to improve local matches between data points and nearby estimations, and methods, such as cross validation, can be used to identify stability in the model as well as inform parameter selection. Cross validation is covered later.

Dealing With Extreme Values

More often than not, your data set will be made up of many low valued samples along with one or two extreme value results. Sometimes, these extreme value results will be more than 2 or 3 orders of magnitude higher than the majority of the points. These scenarios are always difficult for any interpolation method. If all data validation checks that have been reapplied to the outlier

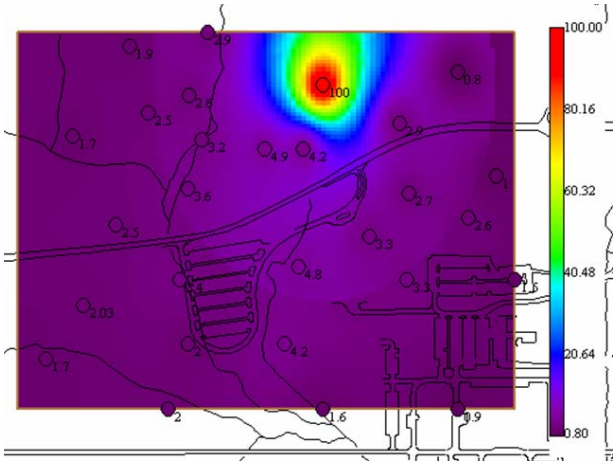
data indicate that the value is real, there are some things you can do with inverse distance. Natural Neighbor and Nearest Neighbor are not largely parameterized and therefore few options exist for handling this situation.

The most common effect is spatial smearing. In the image below, an extreme outlier on the north side of the site with a value of 100pCi/g is artificially inflating estimations well over halfway across the site. Most values are less than 5pCi/g.



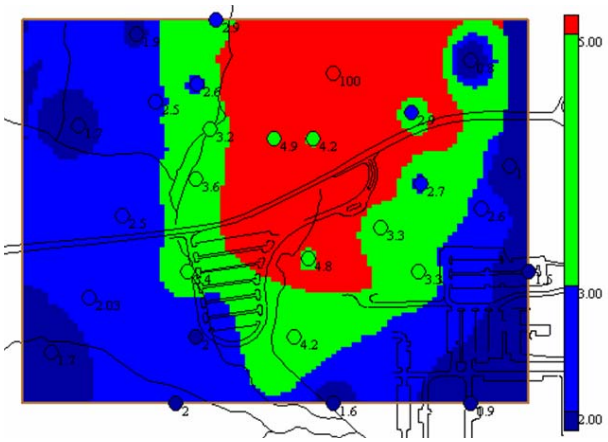
Inverse distance is supposed to geometrically reduce the level of influence that any value has over distance. As distance increases, weights assigned to the outlier will become geometrically smaller. So why doesn't this automatically take care of any outlier smearing? The reason is actually fairly simple. At great distances, weights will be very small, but a very small weight applied to a very large number may still yield a large number relative to the majority of data points.

Amazingly, smearing can often go unnoticed. The difficulty arises in the choice of legend. The following image is numerically equivalent to the results in the previous image.

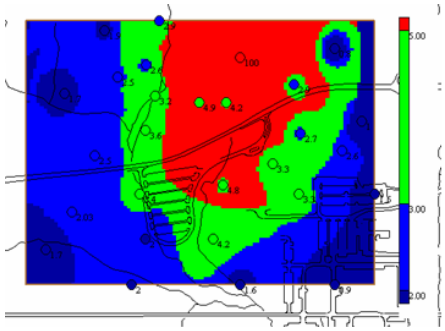


The choice of *Continuous* legend here (see Chapter 43, "Graphical Tools") is deceiving. Values range from 0.80pCi/g to 100pCi/g. The outlier value is essentially pushing down all other values into the dark purple region of the legend. Notice that values as high as 7 or 8 could exist in the purple region, even though the second highest data value is only 4.9. In the following image, a *Categorical* legend was selected that groups all values greater than 5 into a single color group (red). The rest are divided into green and blue. The value of 5 was selected because this was a decision threshold value (e.g. exposure criteria). It may be useful to select categories for your

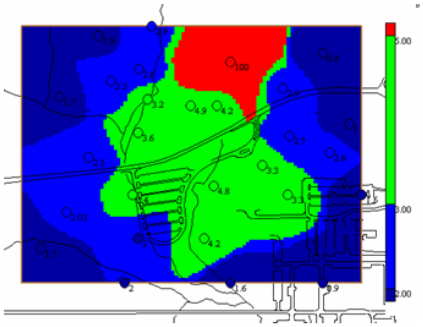
legend that correspond to large breaks between data values (for example in this case between 4.9 and 100).



There are two things that could be done to mitigate this artificial smearing. First, you could drastically shorten the major radius that runs N-S. This would eliminate the outlier from most node estimations. All areas of the site will suffer, however, due to a drop in conditioning data within each search window. Another alternative perhaps, in combination with adjusting the major radius, is to increase the power parameter. In the following image, a comparison is made between a power of 2 and 7. In the latter, the smearing has been contained within a reasonable area of the site.



Power = 2



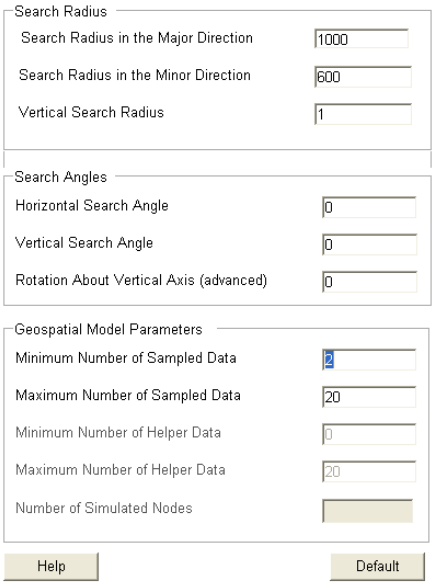
Power = 7

Cross Validation

Cross validation is the process of determining how well the modeling is reflecting reality. The truth is that we will almost never know how well it is performing, since we will likely never have samples at every point we estimate. If we did, we could simply compare them and evaluate the model performance. As an approximation to this approach, we can cross validate by removing each sample that we do have one at a time and allow the model to predict its value based on the remaining data. We can then compare this estimate with the real value and make comparative statements between the different models.

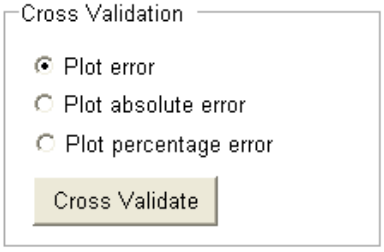
It is important to remember that cross validation provides evidence (rather than proof) of how a model is performing, since in each estimation we have N-1 samples rather than N. Even more than this slight reduction in sample size, cross validation can be particularly harsh on a model for those regions that have sparse data. In fact, cross validation often speaks more about the importance of specific data points than on a particular set of parameter choices.

Close out any file you have open and reopen GeospatialOverview.sda. Select *Interpolate my data*, *Soil*, and *Ac-225*. On the *Interpolation methods* step, make sure you have a value of 2 for the power parameter. On the *Search neighborhood* step make sure you have the following parameter values.

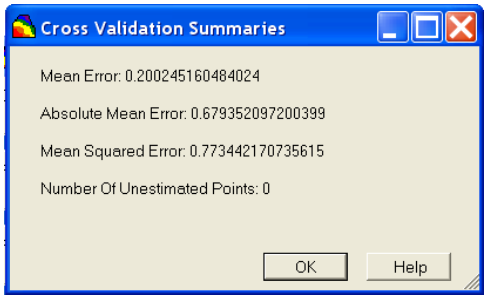


To cross validate, click on the *Cross validation* step. Three options are presented:

- *Plot error* (Simple subtraction of estimated and real values)
- *Plot absolute error* (Absolute value of the difference in real and estimated values)
- *Plot percentage error* (Absolute difference/Real value *100)



Select *Plot error* and press *Cross Validate*. SADA presents the following report.



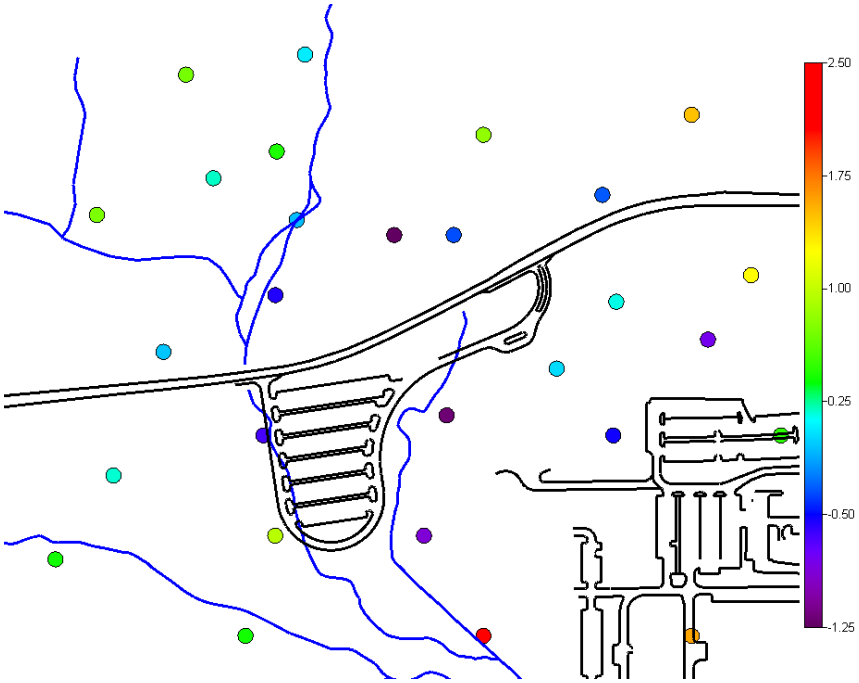
The mean error is the mean of simple errors. Ideally, this would be zero. For the *Plot error* option, this would mean that errors are perfectly balanced in the sense that points have as much over-estimation as under-estimation. In this example, there tends to be an over-estimation of about 0.2pCi/g spread among the data. The *Absolute Mean Error* is the mean of absolute

errors. Ideally, this would also be zero, indicating no errors in estimation. This will never happen. Here, the average error is 0.68pCi/g. The *Mean Squared Error* is similar to the *Absolute Mean Error*. Here, the *Average Squared Error* is 0.77pCi/g. Whether these are large or not depends on the range of your data values. If data values are ranging from 1-2pCi/g, these errors are huge. If data values are ranging from 1000-2000pCi/g, an *Absolute Mean Error* of 0.68pCi/g is outstanding.

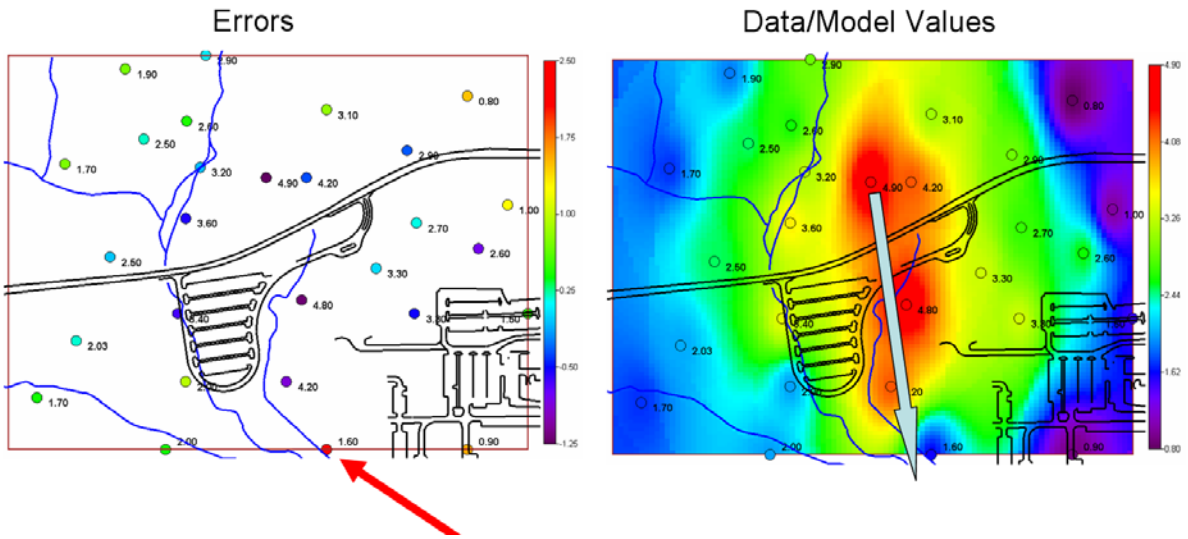
For sparse areas of the site, when a data point is removed for cross validation, there may be no other sample values within the search window. This would yield an unestimated value where the sample originally was found. This is normal, and one should not modify the search neighborhood of their site just to accommodate this aspect of cross validation.

In principle, one could vary certain parameters to produce lower average mean values. This would provide some indication of the "goodness of fit" of one model over another. While this is valuable, be aware that approaching optimization in this manner is an attempt to globally optimize the model at the expense of certain areas of the site. Those areas that are negatively affected may in fact be decision critical regions. Certainly, one would not want to reduce the reliability of the model in a critical area of the site simply to decrease the global error rate. In fact, much of the error may come from an area that, relative to a decision criterion, is unimportant.

For this reason, it is important to look at the plot of errors. Press *OK* to this report and SADA will show the spatial distribution of simple errors.



Notice that the legend now is errors in pCi/g (or whatever units your measurements are in). In this plot, the greatest errors will be in the purple and red range for over- and under- estimation, respectively. Notice that the high measurement values in the middle were consistently under-estimated. The most important thing to notice may be the bright red result in the middle of the site along the southern boundary. Cross validation yielded a very high error here. Does this mean that the model is performing poorly in this region? Not necessarily. What it means is that the data point is evidently very important, from a spatial perspective. Why is this? Let's look at both concentration and error plots side by side.



Looking at the Data/Model Values, you can see that a strong line of high values runs N-S along a transect of over 1200feet. The only data value that stands as evidence that the high value area may have come to an end is a single point (1.60 pCi/g). When this point is removed, the model can only assume that this strong trend is continuing in this large data gap. For this reason, the data point sees the greatest errors in cross validation. It is unlikely that any parameter choice will satisfactorily offset this situation. The truth is that spatially, although the data are uniformly distributed relative to the delineation of the high value area, this southern area is likely data poor. In fact, the model does send the high value region through the small gap to the left of the point even with the single data point there. In this case, we can conclude that this data point bears too much responsibility. More data should be collected to confirm that the trend has ended. This is possibly the real strength in spatial cross validation – the determination of individual data worth.

The other *Cross Validation* options yield similar interpretations. Try for yourself the *Plot absolute error* and *Plot percentage error*.

Basic Methods Summary

Basic methods, such as those presented here, use point geometries to estimate values at grid nodes. Simple algebraic functions of distance or area of influence are the basis for weighting conditioning data in the estimation of node values. Parameterization of these models (inverse distance) can be responsibly conducted, but there is an inevitable art to the process. Geostatistical methods provide tools to assist with the parameterization of models and expand the analysis from a deterministic framework to a powerful stochastic process where models of uncertainty exist that can be used in a variety of ways.

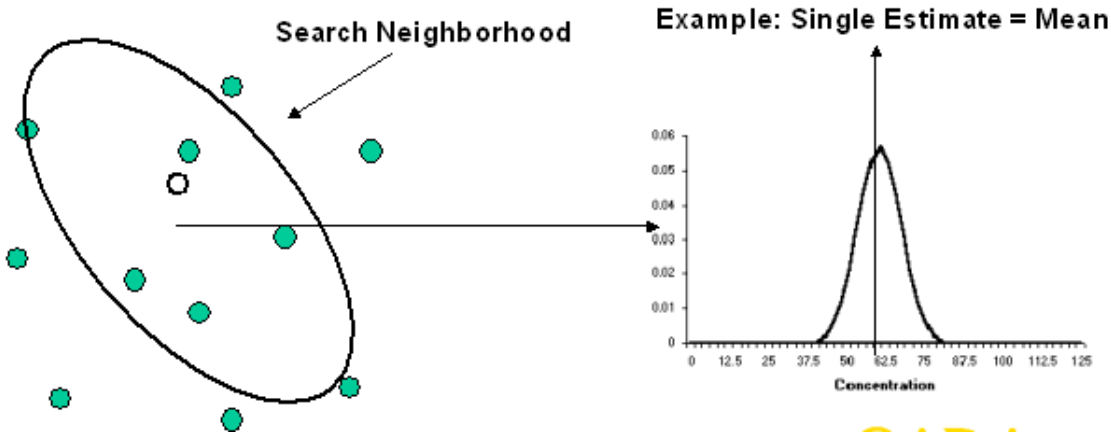
References

Watson, D. (1999). "The natural neighbor series manuals and source codes." Computers & Geosciences 25(4): 463-466.

Sibson, R., (1981). A brief description of natural neighbor interpolation. In: Barnett, V. (Ed.), Interpreting Multivariate Data. John Wiley & Sons, New York, pp. 21–36

Chapter 30: Advanced Geospatial Methods Part I: Overview and Correlation Modeling

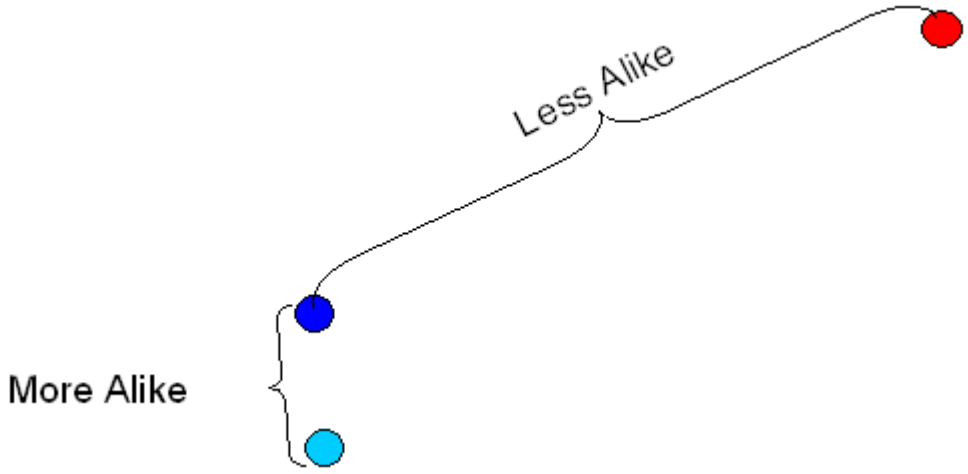
With basic spatial analysis tools, each interpolant produces a single estimate for each unsampled point. With a geostatistical approach, a distribution of possible values is constructed and used as a model of uncertainty for the actual, yet unknown, value at the unsampled location. From this distribution of possible outcomes (probability distribution), a central moment, such as the mean or media, is chosen as a single estimate for contouring purposes.



SADA provides three kriging (geostatistical) models: Ordinary, Indicator kriging, and Cokriging. Ordinary kriging is the most commonly used approach and can be used to construct probability distributions under the assumption of normality for the data. Indicator kriging is a non parametric approach that does not assume any specific shape for the probability distribution. Cokriging allows you to include other types of data in the analysis that may not be direct measurements of your contaminant of interest; this permits the inclusion of various kinds of field detection measurements, geophysics, and so forth.

Like the methods discussed in Chapter 29, “Basic Geospatial Methods,” these methods are based on a weighted combination of nearby samples. The development and expression of these weights, however, is more complex. It may be helpful at first to think of kriging as an advanced form of the inverse distance method. Recall that the inverse distance method weights sampled values by their distance from the unsampled location. Kriging approaches the problem in much the same way. Rather than distance (d), however, the weights are based on the amount of spatial correlation or spatial covariance that samples exhibit at varying distances. Another major difference between kriging and basic geospatial methods is that the geostatistical model accounts for the correlation among all the sample locations, assigning smaller weights to data that are clustered and provide redundant information about the unsampled location.

If data are spatially correlated, then on average, sample points that are close to each other are more alike than sample points further away. (More complex spatial correlations exist, but this type is the most common).

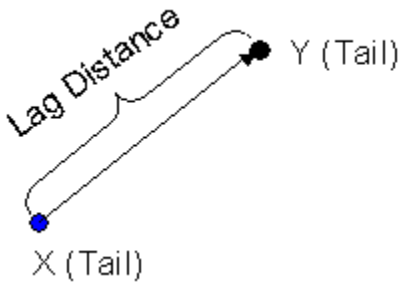


Concepts in Spatial Correlation

The cornerstone of any geostatistical analysis is the quantification of the degree to which data are more or less “alike” as a function of their geographical proximity. SADA uses the semi-variogram method, which returns a measure of variance for any given separation distance. This measure is defined as half of the average squared difference between values separated by distance h. The term h is referred to as the lag or lag distance.

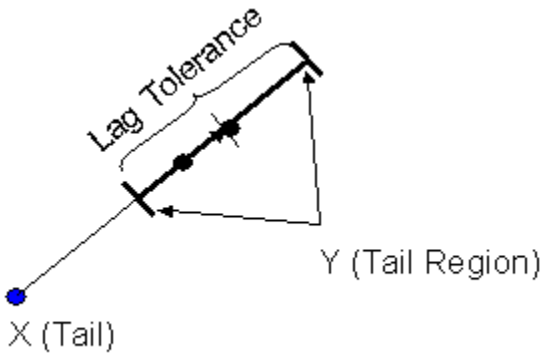
$$\gamma(\mathbf{h}) = \frac{1}{2N(\mathbf{h})} \sum_{i=1}^{N(\mathbf{h})} (x_i - y_i)^2$$

where N(h) is the number of pairs separated by distance h; x_i is the starting sample point (tail value); and y_i is the ending sample point (head value).

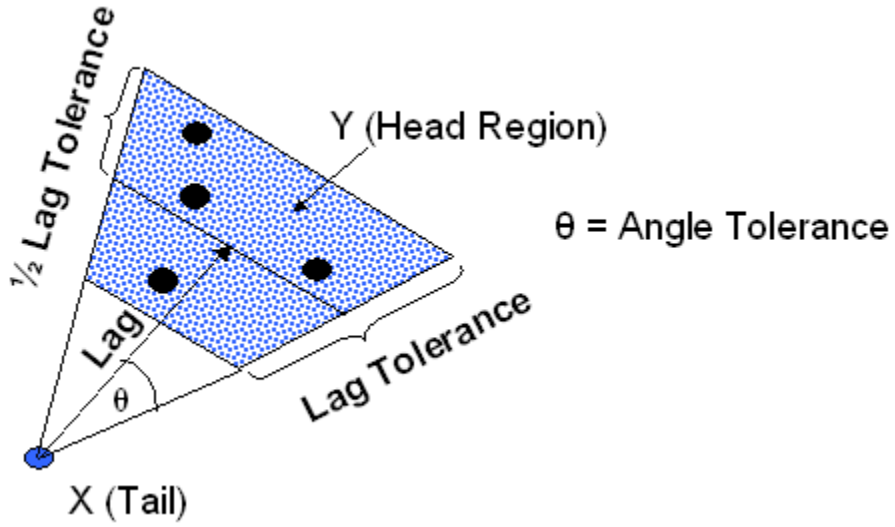


Except when observations are collected according to a systematic sampling design, you will never have a large number of sample points separated by exactly the same lag distance h. Therefore, a lag tolerance centered about the lag distance will permit the capture of more data points in the calculation of γ(h). In the figure below, all data points on the dark thick line area will be used.

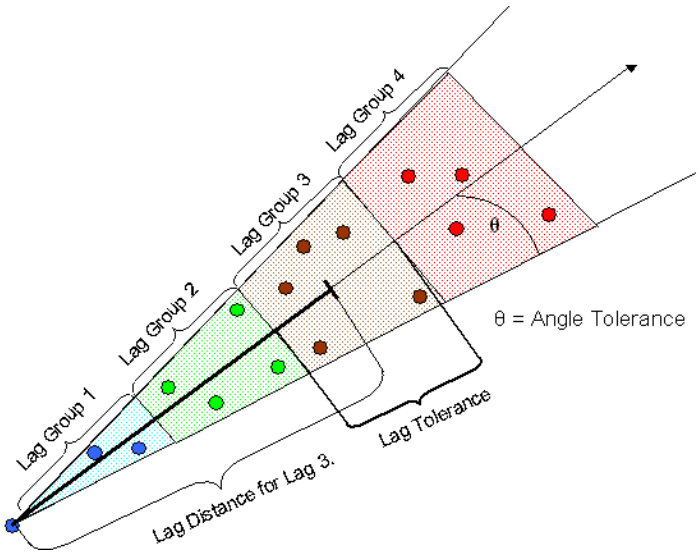
So, if we are interested in the variance of all data points separated by 10 feet and we permit a lag tolerance of 2 feet, we will actually be calculating the variance of all pairs of data between 9 and 11 feet apart.



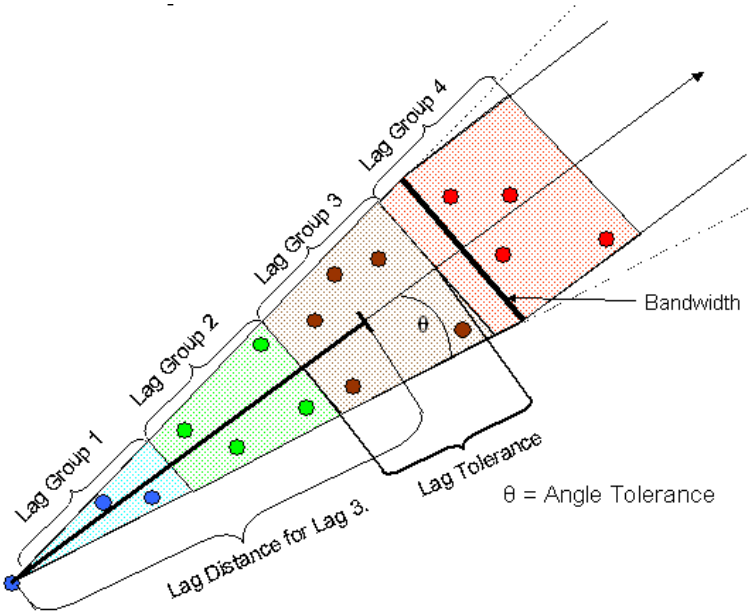
Although assigning a lag tolerance helps, most cases will never have enough samples separated by a lag - tol/2 to lag + tol/2 along a straight line to calculate the semivariogram value. Therefore, an angle tolerance, θ, is also introduced to expand the region and to include more points in the calculation of the semivariogram value for the specified lag distance. In the figure below, all data points within the blue shaded area will be used.



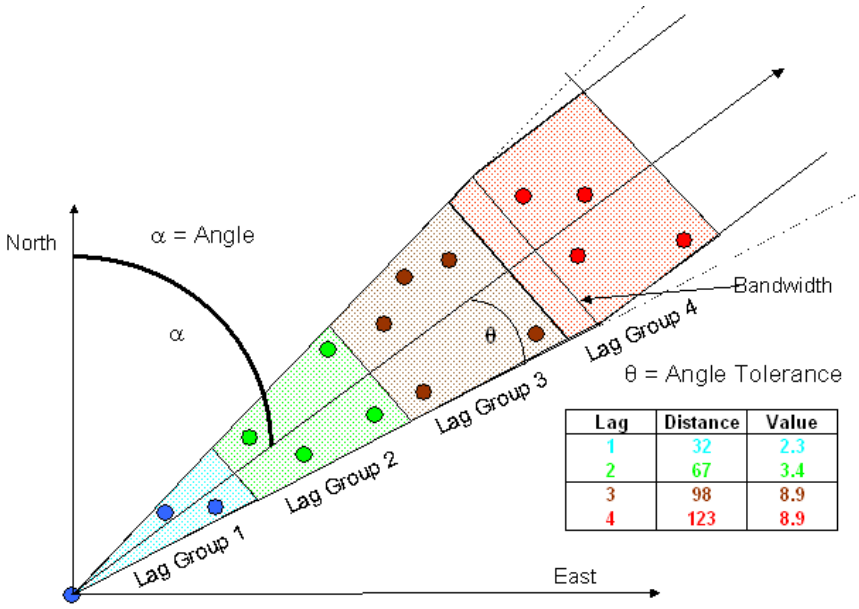
If we repeat this operation for a number of lag distances, we would generate a cone shaped object expanding outward from the point of interest. This cone would be partitioned by lag groups centered about our lag distances.



As the cone stretches farther out, it opens up increasingly wide, capturing more and more data points as it moves away. In practice, geostatisticians will often apply a constraint called the bandwidth. This bandwidth limits the expansion of the cone to a certain width. If you do not wish to constrain the cone's expansion, specify a very large bandwidth.

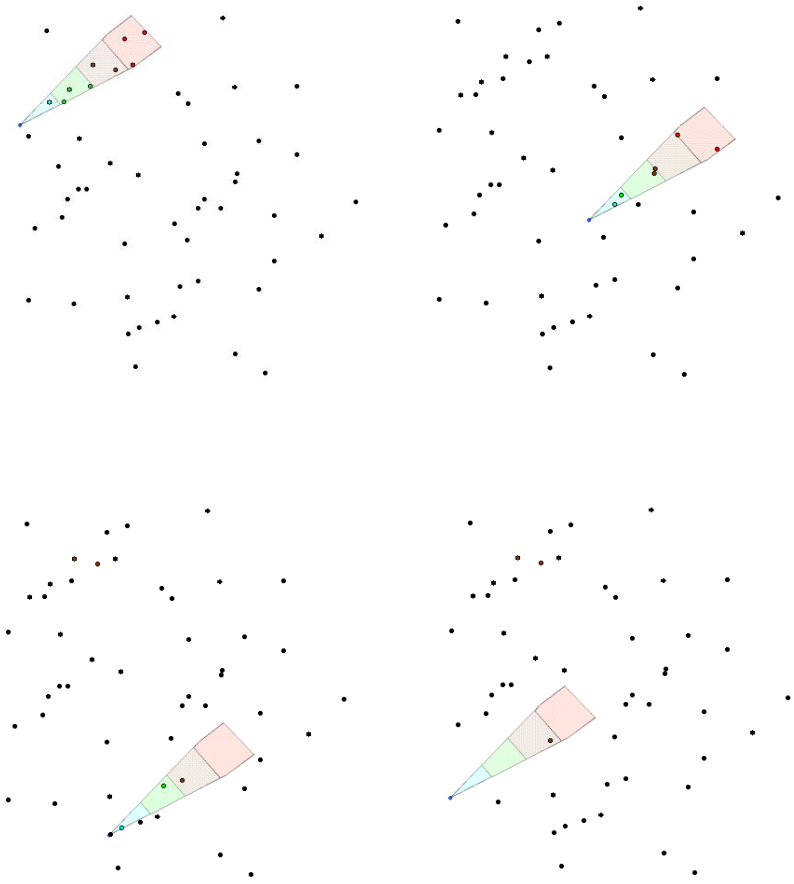


Our final parameter is the angle. The angle specifies in what direction you will be calculating the semi-variogram values. Now we are constraining our semi-variogram values to a certain direction.

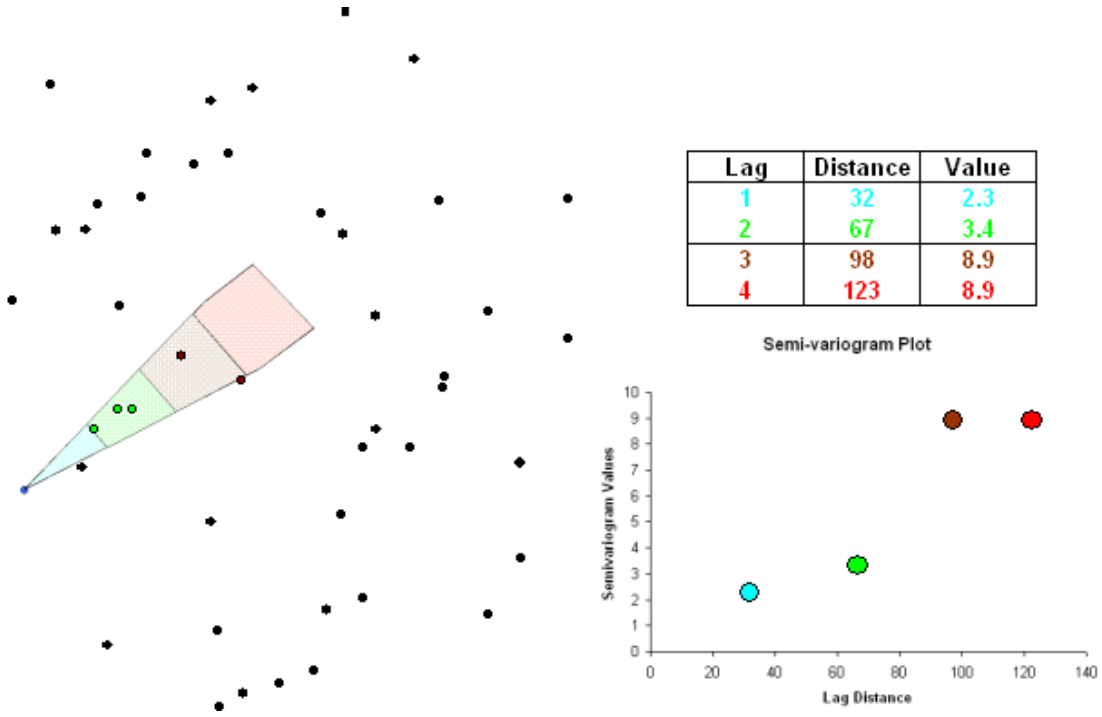


Lag	Distance	Value
1	32	2.3
2	67	3.4
3	98	8.9
4	123	8.9

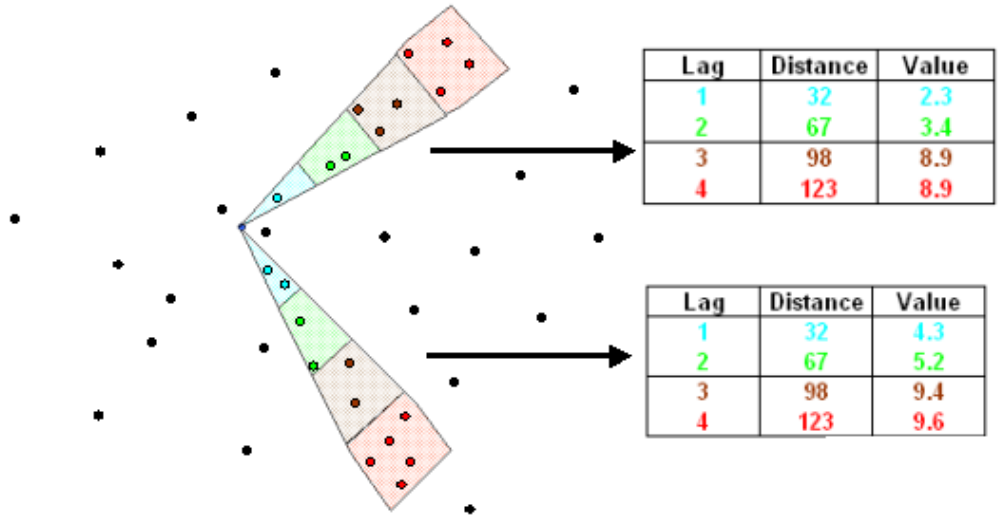
The semi-variogram cone is relocated at every sample location to build up the total set of pairs for each lag bin.



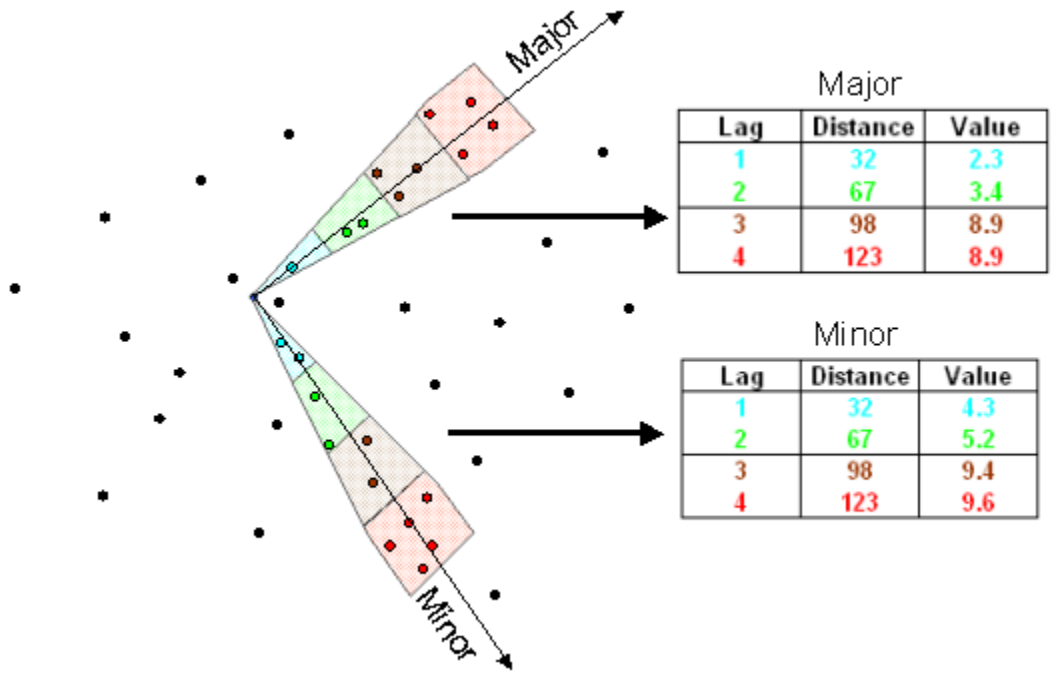
The semi-variogram values, averaged within each class of distances, are then plotted as a function of the lag distance.



Note that by specifying an angle tolerance θ, we are excluding all those data points located outside of the cone formed by the lines of azimuth α - θ degrees and α + θ degrees. In other words, we are exploring how data are correlated in a particular direction. The same procedure can be repeated for different classes of angles. If we find that the correlation changes depending on the direction being studied, the pattern of correlation is said to be anisotropic.



In fact, if anisotropic conditions exist, the direction of highest correlation is considered the major direction of anisotropy. The perpendicular direction is referred to as the minor direction of anisotropy. The major direction of correlation will exhibit semi-variogram values that increase at a slower rate than in any other direction (i.e. direction of lowest variability).



Major

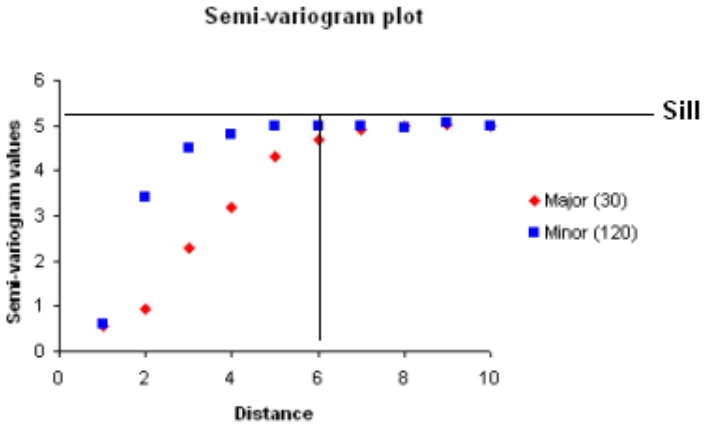
Lag	Distance	Value
1	32	2.3
2	67	3.4
3	98	8.9
4	123	8.9

Minor

Lag	Distance	Value
1	32	4.3
2	67	5.2
3	98	9.4
4	123	9.6

Theoretically, the semi-variogram values will continue to rise until they reach the sill value, which should be roughly equivalent to the variance of the data set. The distance at which the semivariogram plateaus is known as the range of autocorrelation. At that distance, the data are now far enough apart to be independent. A semi-variogram plot is useful in detecting the sill value and the range.

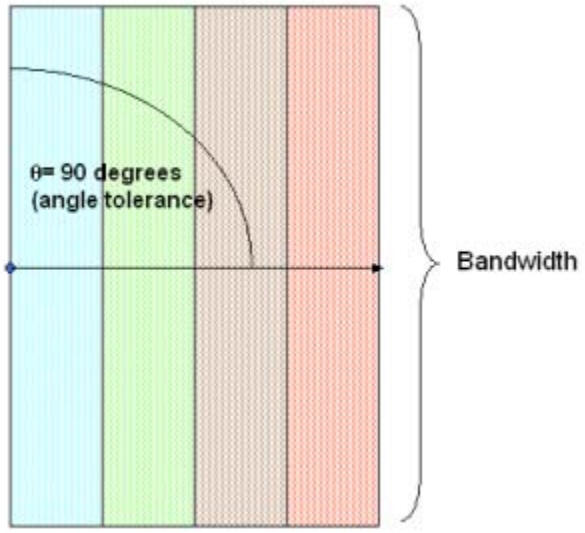
Lag	Major (30)	Minor (120)
1	0.56	0.6
2	0.95	3.4
3	2.3	4.5
4	3.2	4.8
5	4.3	5
6	4.7	5
7	4.9	5
8	5	4.95
9	5.01	5.05
10	4.99	5



In the above example, we see the *Major* direction at 30 degrees and the corresponding *Minor* direction at 120 degrees. The Semi-variogram reaches a *Sill* value of approximately 5 for a separation of around six feet.

Omni-directional Variograms

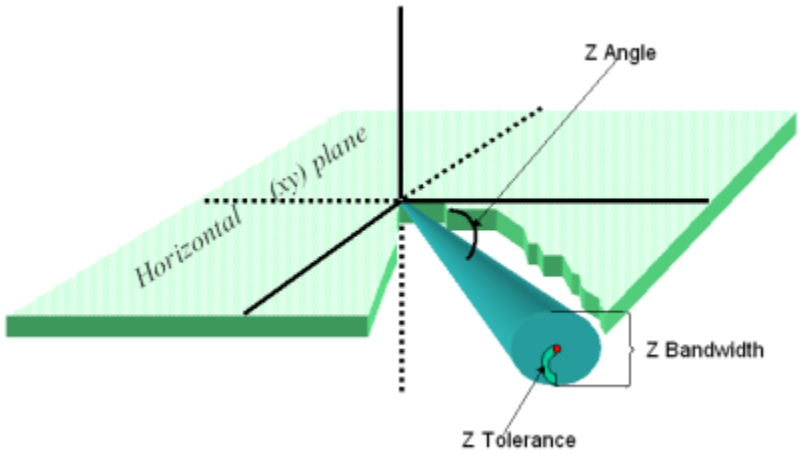
In order to calculate an omni-directional variogram (i.e., case of isotropic or direction-independent variability), simply set the angle tolerance to 90 degrees and make the bandwidth significantly larger than the site. This setting will force the cone to include the entire spectrum of data points.



Three Dimensional Variography

Three-dimensional semi-variogram calculation proceeds similarly as in the two-dimensional case. In addition to the previously defined parameters, a z angle (dip), z tolerance, and z bandwidth must be specified.

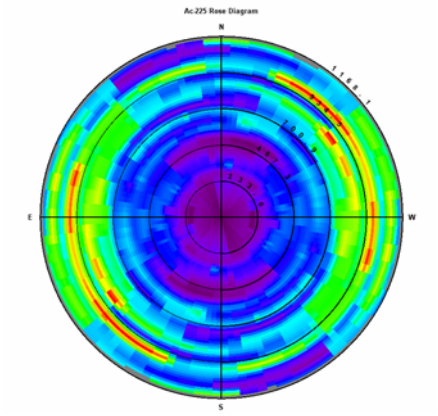
- Z Angle (Dip) – The angle below the horizontal plane that the cone should dip.
- Z Tolerance – The tolerance on this dip angle.
- Z Bandwidth – The maximum distance the vertical component of the cone is permitted to go.



Variogram Maps (Rose Diagrams)

Rather than considering only one direction at a time, users can view semi-variogram values in all directions at once. Here, the semi-variogram is calculated with an increasing, incremental horizontal angle, and the results are plotted on a disk where the color represents the value of the semi-variogram. The center of the disk corresponds to the zero distance point. For three dimensional data, you can specify a range of z dip angles, and SADA will compute this disk for incremental changes in the dip angle as well.

Rose maps are particularly useful to identify the major direction of anisotropy that appears as the direction with low “trough like” values. In the example below, there is a trough of low values running N-S, which indicates a lower variability (i.e., smaller variogram values) along that direction. The use of colors makes interpretation of variogram results much easier.



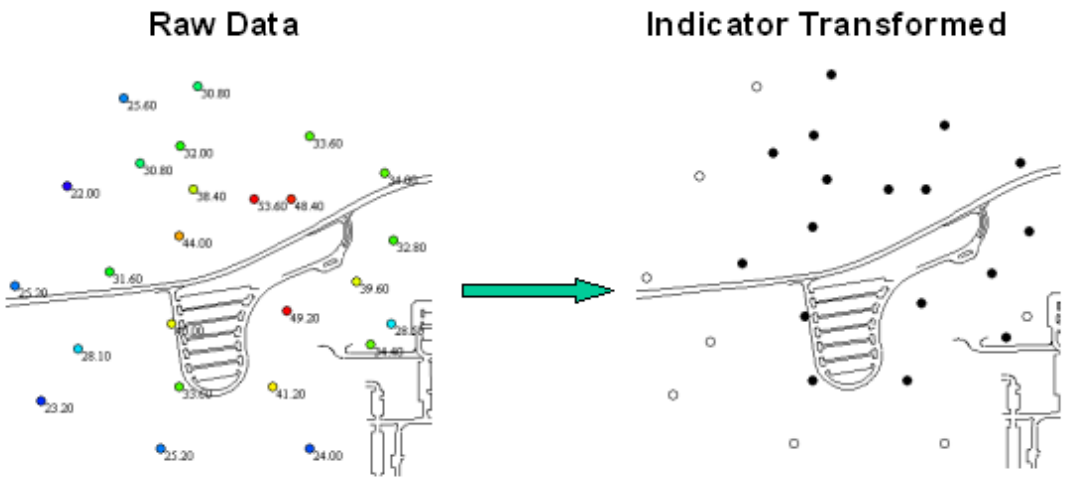
Indicator Variograms

To this point, all variograms have been calculated on the actual observations. This type of variography is necessary for ordinary kriging and cokriging. Indicator kriging and sequential indicator simulation require variography to be conducted on the indicator transformed values. When performing an indicator transform, you need a threshold value t. All sample values greater than t are changed into 1, and all values less or equal to t are coded as 0. We can write the indicator transform of the data as follows:

$$I(u_i) = \begin{cases} 0 \; if \; u_i > t \\ 1 \; if \; u_i \leq t \end{cases} \quad for \; all \; i$$

If we plot the transformed data values, they would look something like this:

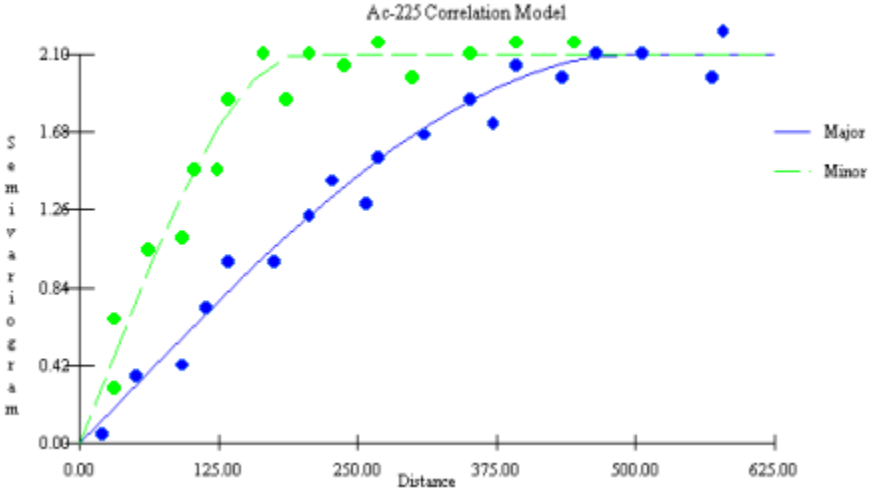
The following example shows Arsenic data that have been indicator transformed at a threshold of 30 mg/kg:



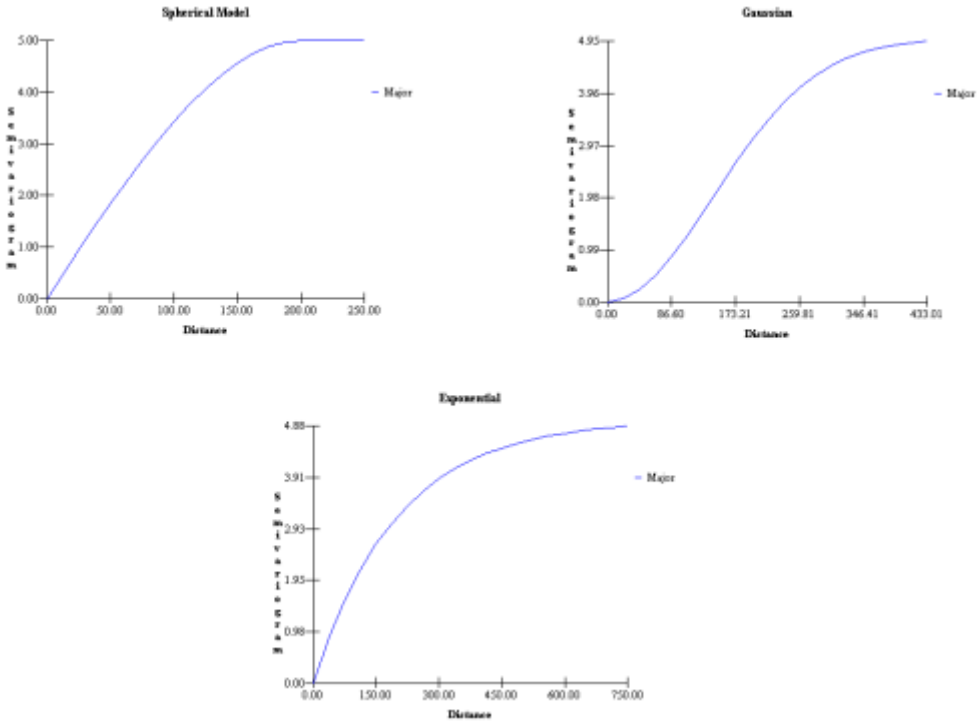
Calculations of the variogram for the indicator transformed data then proceed exactly as in the case of the untransformed data.

Spatial Variogram Modeling

For kriging or simulation procedures, one needs to know the semi-variogram values for any separation distance h and azimuth α. At this point, we only have these values for a few discrete lag distances. Therefore, we need to fit a model to the experimental semi-variogram values so that a semi-variogram value can be derived analytically for any vector *h*.



SADA provides three standard semi-variogram models that allow a great deal of flexibility in capturing the shape of the experimental curves: Spherical, Exponential, and Gaussian.



These correlation models actually are elliptical or ellipsoidal in 2d and 3d space, respectively. Typical plots only show model behavior along the major and minor axes. Recall that these two axes are determined through the use of the rose diagram or variogram map. While correlation models are defined for any angle between 0 and 360º, it is sufficient for you to fit the model along the major and minor axis (checking the quality of the fit in two additional perpendicular directions is strongly recommended, though).The specification of these models in three dimensions amounts to the specification of an ellipsoidal geometry defined by three axes. All three semi-variogram models require the same set of parameters (e.g., range, sill), and most of these parameters were encountered while specifying a search neighborhood. For a review of how to construct a search window please see Chapter 29, “Basic Geospatial Methods.”

Ellipsoidal (Domain) Geometries

- Major Range – distance to sill or correlation length along the major anisotropic axis.
- Minor Range – distance to sill or correlation length along the minor anisotropic axis.
- Angle – the horizontal angle of anisotropy
- Z Angle(3d) – the angle of anisotropy in the Z plane (equal to the Dip parameter in experimental variography).
- Z Range(3d) – a value describing how anisotropy behaves in the z minor direction, relative to major axis.
- Rotation(3d) – how the anisotropic ellipsoid is rotated about its major axis.

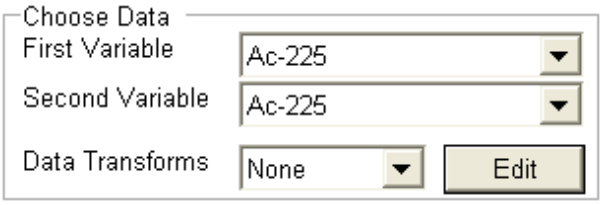
Model Properties

- Contribution – The model’s contribution to the sill (maximal model value)
- Nugget – value of the variogram for a zero distance (i.e., intercept of the vertical axis) (white noise)

SADA also has a variogram fit function to assist the user with assigning these values.

If you have not already done so, open SpatialCorrelation.sda in SADA. This file contains 2D soil data (Ac-225) and 3d groundwater data (Chlordane). We'll use these two data sets to start getting familiar with semi-variogram calculations.

Make sure you have selected *Soil* and *Ac-225*. Then in the interview drop-down list, select *Model spatial correlation*. We will use this file to demonstrate how to assess and model spatial correlation for all forms of kriging and simulation available in SADA. Each method requires different transforms of the data and/or multiple structures. Click on the step *Correlation modeling*. The parameter window associated with this step appears involved, and to some degree it is. Let's start by looking at the top of the parameter window at the information block called *Choose Data*.



Here, you see a *First* and *Second Variable*. Recall that in a semi-variogram construction there is a head and tail variable. This is where you specify these two variables. At the moment, they are both set to *Ac-225*. This means that we are only interested in examining the correlation structure of Ac-225 measurements. If you select either drop-down list, you'll notice that *Ac-225* is the only option. It is also possible to specify another variable and examine the spatial correlation between Ac-225 and a "helper" data set, such as field survey data set. Helper data, or secondary data, is necessary for cokriging and cosimulation techniques, as is the correlation structure between helper and primary data (e.g., Ac-225). In the section below on correlation modeling for cokriging, we'll show you how to add "helper" data and change the second variable to this helper information.

The Data Transforms parameter contains three methods for transforming the data: *None*, *Normal Score*, and *Unit Transform*. None means that raw data values will be used only.

When performing a sequential Gaussian simulation (SGS), the data must be normally distributed. Unfortunately most data sets are not normally distributed; however, a normal score transformation can convert any distribution into a normal distribution. The simulations are run in this space and then back transformed into real space. The following explanation of the normal score transform is based on Deutsch and Journel (1992).

Each data value will be denoted by z_i (z_1, z_2, …, z_n). The cumulative probability associated with each of these observations will be c_i. So, c_5 = prob($Z < z_5$) and so forth. The normal score transform of z_i noted as y_i, is calculated as

$$y_i = G^{-1}\left(\frac{c_i + c_{i+1}}{2}\right)$$

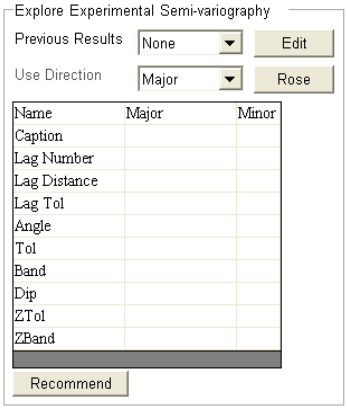
where G(y) is the standard normal cumulative distribution function. Then, $y_c = G^{-1}(C)$ is the corresponding standard normal c-quantile. SADA uses the *nscore* program developed by Deutsch and Journel (1992, p. 211) to perform the normal score transform.

The *Unit Transform* option simply rescales the data so that there is a mean of zero and a variance of 1.

$$y_i = \frac{z_i - \bar{z}}{\sigma^2}$$

This drop-down list can also include any indicator cutoff (threshold) value you have defined. The indicator thresholds are defined using the *Edit* button. We'll visit this feature shortly when we talk about indicator estimation and simulation.

The information block below the *Choose Data* block is called *Explore Experimental Semi-variography*. *Previous Results* has a drop-down list of semi-variogram evaluations we may have done previously and saved. The *Edit* button just to the right of this drop-down list allows you to manage your saved results. At the moment, there is nothing to choose from. Later, we'll show you how to save your results for both semi-variogram plots and rose maps alike.



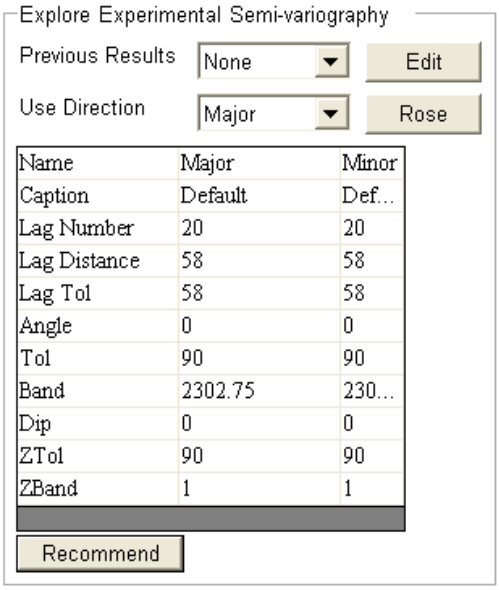
Next is the Use Direction option. This drop-down list allows you to choose from: *Not Used*, *Major*, *Minor*, or *Both*. The options in this list box are a bit misleading at first glance. They do not refer directly to the major or minor direction of anisotropy, but rather to the columns in the parameter table just below them. In that parameter table, there is a major and a minor column. Typically, investigators will parameterize the search window for the major direction in the major column. Likewise, the minor column holds parameters typically associated with the minor direction. You do not have to restrict yourself to this idea, however. In fact, when you begin you will likely not know in advance the major or minor direction. It is better to think of this as the ability to parameterize two search directions at once. The *Use Direction* drop-down list allows you to tell SADA which column of parameters you want to use at the moment. Sometimes you don't wish to use either, because you only want to plot the model and not compute the point values (see below).

Next to the *Use Direction* drop-down list is the *Rose* button. This will generate a rose map using one of the columns. You cannot generate a rose map for *Not Used* or *Both*; you must select either *Major* or *Minor* to base the *Rose* map on.

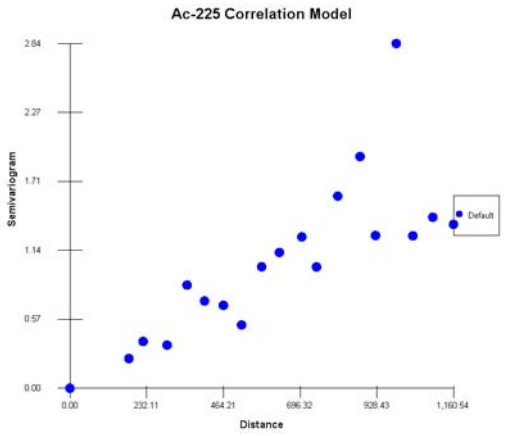
Let's Practice

For practice, let's go through a couple of examples of semi-variogram calculations. The first step is to explore semi-variography values to detect the existence of anisotropy. Most of your time will likely be spent on this step. The first thing we need to do is populate the block of information called *Explore Experimental Semi-variography*. At the moment it is empty and perhaps even a bit intimidating. If you have a good handle on how the data are patterned across your site, you can make educated guesses for the parameters. From that point, you can use actual variography results to further refine your parameter selection.

Another way to begin, especially if you are relatively new to this, is to use the *Recommend* button. The *Recommend* button uses some basic rules of thumb to get you started. You should under no circumstances believe that these recommendations are optimal in any sense. Rather, they serve as a good starting point. Push the *Recommend* button for *Explore Experimental Semi-variography* now and say *Yes* to the confirmation message box that appears. You should have something like this:



At the bottom of the parameters window there is a *Show Me* button. Clicking this will produce a semi-variogram plot using the parameters in the selected columns (major and/or minor). Go ahead and push the *Show Me* button now. SADA produces a standard semi-variogram plot.



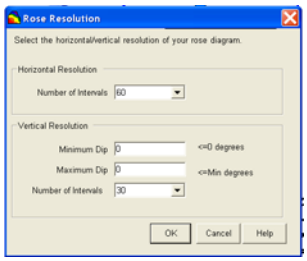
From this plot, you can see how data are changing as a function of their separation distance. This is the central concept in geostatistical modeling and will drive the selection of model parameters. Now, this is an omni-directional semi-variogram, which means that the cone geometry discussed earlier is so wide that for any given point it includes data in all possible directions. Some spatial patterns are isotropic and so an omni-directional variogram is completely appropriate.

Let's modify the parameters to see if there is a dramatic change when we narrow the focus of the semi-variogram cone and look in a particular direction. To do that, we will need to change only two parameters: *Tol* and *Band*. For 3d applications, you would also need to change the

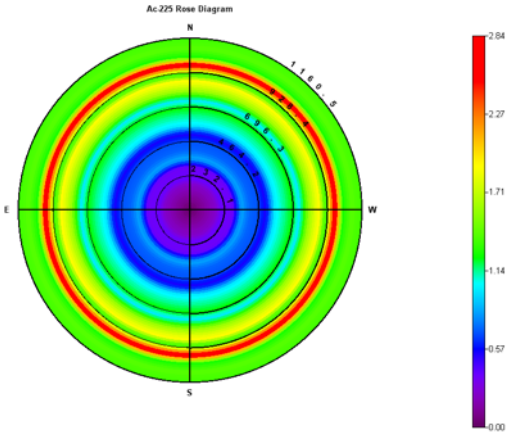
ZTol and *ZBand*. *Tol* is the angular tolerance that controls how wide our cone is. Right now it is 90 degrees, which means that regardless of the horizontal angle (*Angle*), the cone is so wide (i.e., 180º) that is includes all directions (recall that "behind" each point is exactly the same result as in "front" of each result). Change the *Tol* to 45 and the *Band* to 1000. Press *Show Me* again.

You can see that there is a difference, but it is difficult to determine what changed. The distribution of points changed and the y-axis also changed. A better way to evaluate directional variograms is with the rose map. The rose will incrementally increase the *Angle* parameter from 0 to 180, calculate the semi-variogram values, and perform a linear interpolation between points. This will greatly facilitate the visualization of spatial patterns. One way to approach this is to set your tolerance (*Tol*) back to 90 degrees (do this now). This will create an omni-directional rose map that will look like a wagon wheel. Using this as a base, we can slowly tighten the spread of the variography cone and see how the rose map changes. Sites with anisotropic variability will move away from a wagon wheel look and will start showing troughs of low semi-variogram values in the direction of anisotropy.

We will use the *Major* column as our set of parameters; so make sure you have *Major* selected for *Use Direction*. Press the *Rose* button.



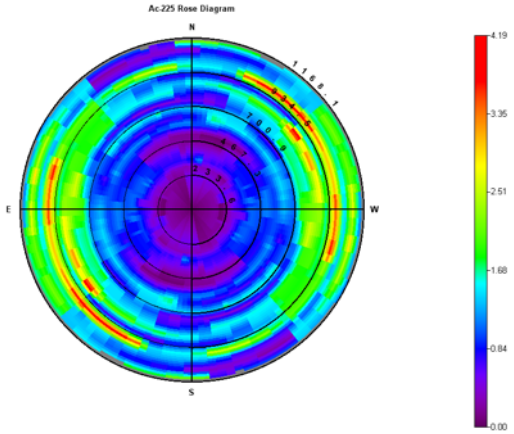
This window is asking you to select an angular resolution (i.e., discretization level) in both the horizontal and vertical directions. The default is 60 intervals between 0 and 180 degrees: the semi-variogram is recalculated every 3 degrees (note that 180-360 is the mirror result of 0-180). We are not currently dealing with 3d data, but if we were, we could specify a minimum and maximum angle to define a range of reasonable vertical angles. This range is then divided into a number of intervals as well. We'll keep the default values and press *OK*. An isotropic or omni-directional rose diagram is produced.



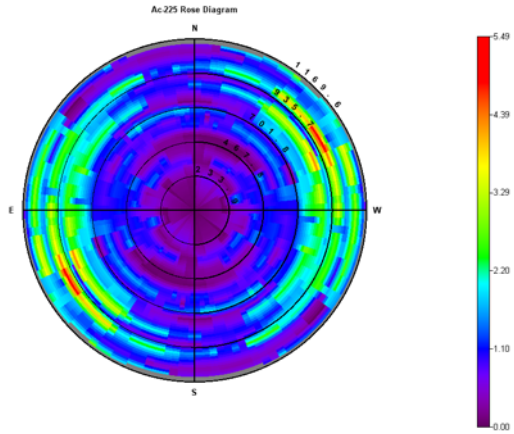
If the variability is truly isotropic (direction doesn't matter), then this wheel-like pattern will continue to hold, even as we vary parameters under the major column. If anisotropy exists, we will see a trough of low values open up in the major direction of anisotropy, which is the

direction of maximum continuity or minimum variability. So, the values are not as important at this point as the relative behavior of the rose map.

Let's demonstrate this by changing the *Tol* parameter under the *Major* column to 45 (degrees). This will narrow the field of view for the model as it sweeps around the site. With this narrowed view, if anisotropy exits, it will show up in the scan. If not, a wheel-like result will persist even though the values may be different. Press the *Rose* button again and accept the interval defaults.



A trough of low values appears to be opening up as the wagon wheel geometry begins to fade. Let's narrow the angular tolerance (*Tol*) even further to 30. Press *Rose* and accept the defaults.



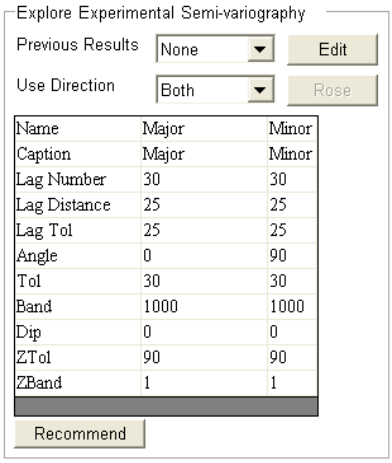
By now, a fairly pronounced trough of low values has opened up along the N-S transect, indicating lower variability along that direction. This trend is exactly what we would expect even by just visually inspecting the location map itself (note that the data map showed a trend running N-S). Therefore, the azimuth angle of 0 degrees is entirely appropriate as the major direction of anisotropy. We will keep the *Tol* at 30 degrees and the *Band* at 1000. At the bottom of the parameter window, press the *Show Me* button; this will show you the exact semi-variogram values in the N-S direction (assuming you have *Angle* = 0 and *Tol* = 30).

Let's increase the number of lags and shorten the lag distance and tolerance. This will give us more points to visualize. Change the *Lag Number* from 20 to 30 and shorten the *Lag Distance* and *Lag Tol* parameters to 25.

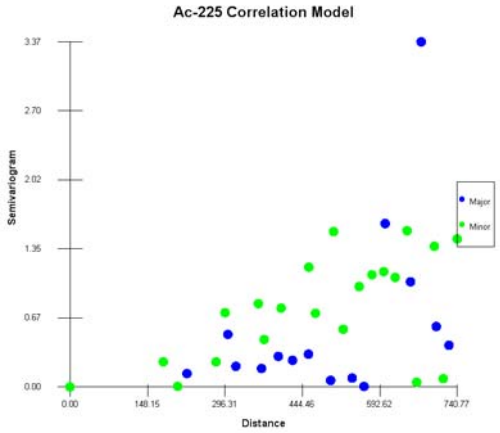
Name	Major
Caption	Default
Lag Number	30
Lag Distance	25
Lag Tol	25
Angle	0
Tol	30
Band	1000
Dip	0
ZTol	90
ZBand	1

In addition, let's parameterize the *Minor* direction. This is necessary only if anisotropy exists. If it does not exist, you would use only the *Major* direction and keep the *Tol* parameter at 90.

Enter exactly the same values for the *Minor* column as for the *Major* column with the following exception: the *Angle* parameter should be 90. This is because the minor and major axes are perpendicular to each other. Therefore, if the major direction of anisotropy is 0, then the minor must be 90. If the major direction were 45, then the minor direction would be 135 and so forth. In the *Caption* row, change the value to *Major* and *Minor* for the *Major* and *Minor* columns, respectively. One last thing, you will need to change your *Use Direction* from *Major* to *Both*. Your semi-variogram parameters should look as follows:

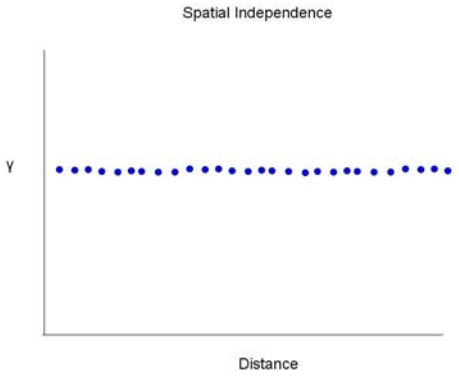


Now at the bottom of the parameters window, press the *Show Me* button. The following result appears:



Save your file. The green dots are semi-variogram values calculated along the E-W transect (90 degrees). The blue dots are the N-S semi-variogram values. Notice how the green dots are almost always above the blue dots. This is because data in the E-W direction are more variable than the N-S data separated by the same distance. The next task is to model these values; we will get to that shortly. First, we will discuss semi-variography requirements for each geostatistical model. For each case, you will approach the problem in a manner similar to what we've just described. Of course, situations vary a lot and rarely will you get well-behaved semi-variography results for small environmental data sets. You may need to use judgment and a great deal of patience in your evaluation.

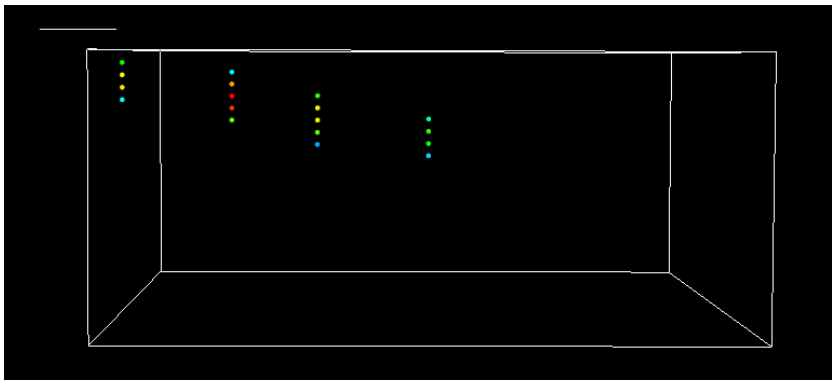
If you cannot get a well-behaved semi-variogram, you still need to be careful about assuming there is no spatial auto-correlation, since variability is often spatially structured. In the figure above, both the green and blue dots are bouncing around quite a bit. There is, however, still a trend of increasing variability with distance. If you ignore this feature and assume no correlation among the data, you are implying that the semi-variogram values behave like the following figure. While in fact, this type of result is evidence of the lack of spatial correlation.



3D Variography

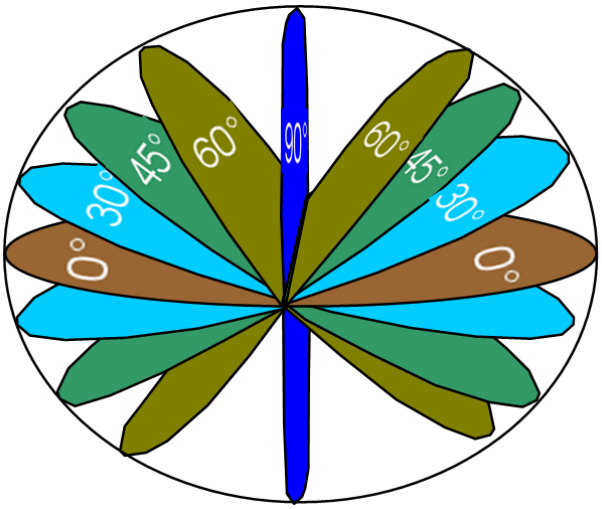
We will demonstrate how to derive and peruse a 3d variography evaluation in SADA. Switch from *Soil* to *Groundwater* and select *Chlordane* for the contaminant. The site has been set up with a vertical layer design of 10 foot layers from 0 to 52 feet deep (check the *Set up the site* step). If you use the vertical layer drop-down list on the main tool bar, you can see that a trend of high values is running not only from North to South but dropping vertically as well. The following image was taken from the 3d Viewer (see Chapter 44, "The 3D Viewer"), where low value points have been hidden to show the trend in the upper values. The shot is due east, so

you are looking at the transect of high N-S values that are dropping vertically as they move south.



There is clearly a correlation structure for high valued samples. We want to capture that information in our correlation assessment. Note that the type of correlation structure may vary across the range of sample values. For example, if you remove the high values and look at only low-valued samples, you can see there isn't much directionality in their behavior. In fact, the spatial variability looks direction-independent with little difference in values outside the plume. Both of these behaviors can be captured using the indicator approach. We'll show you how to do this shortly. Right now, we'll focus just on how to get multiple rose maps that show horizontal variation as well as vertical variation.

If you could see a three dimensional variography result, it would look like a sphere comprised of many, many points that vary in color and density, like a tight cluster of stars in a circular or ellipsoidal shape. SADA will allow you to see cross sections of this cluster one at a time. If you could see the cross sections in 3d space, each rose transect would have the following geometry.



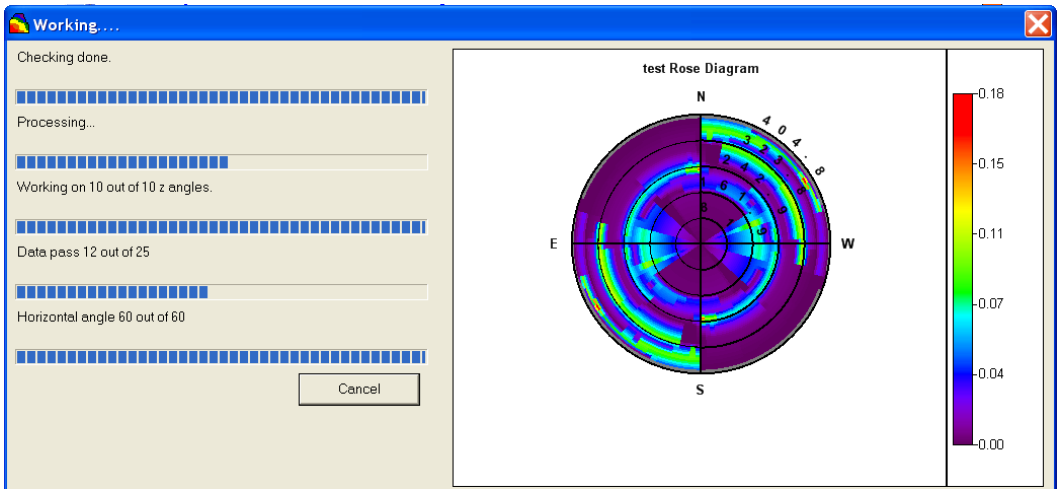
Of course, each transect would not be a single color but painted with a rose map. Be aware that when you are doing 3d variography, the parameters of horizontal angle, horizontal angle tolerance, and even the lag and lag distances now pertain to the direction of the cross section. For 2d data sets, this is in fact the horizontal direction. But in 3d, it is horizontal, relative to the dip angle.

SADA will generate these transects automatically and allow you to scroll through them.

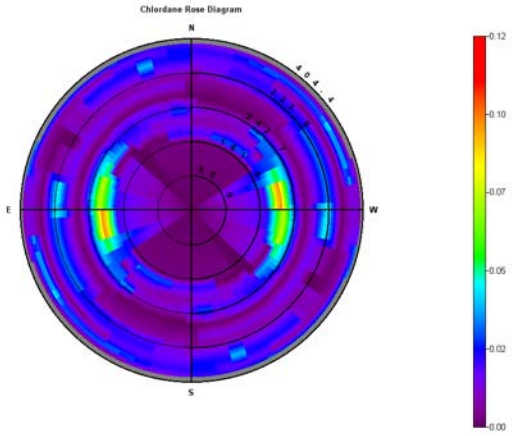
For the *Chlordane* data set, use the following parameters. The data have a smaller spatial range than the Ac-225 data, so we'll need to modify the default parameters some. Make sure you have *Chlordane* selected for both the first and second variables, no transform, major direction only and that the parameter values are the same as in the image below.

Now press the *Rose* button. Accept the default horizontal parameters; however, for the vertical direction, select a *Minimum Dip* of 0° and a *Maximum Dip* of -5°. *Set number of intervals* to 10, so that SADA will calculate 10 transects (one at 0°, another at 0.55°, another at 1.11° and so forth). Recall that we want to use a small maximum interval of 5° because with such a vertically flat site, many large dip values will cause ellipsoidal geometries to jump out of the site quickly.

You will notice that now a larger progress window appears. Three dimensional variography can take considerably more time, and SADA is showing you the progress as calculations are being done. SADA will look at a single transect and calculate the variography data a little at a time so that you can see progress unfolding. For each new transect, the rose image will start out crude and become finer as the total number of samples is used.

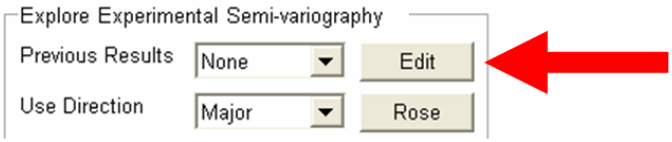


When SADA is finished you will be presented with the first rose transect.

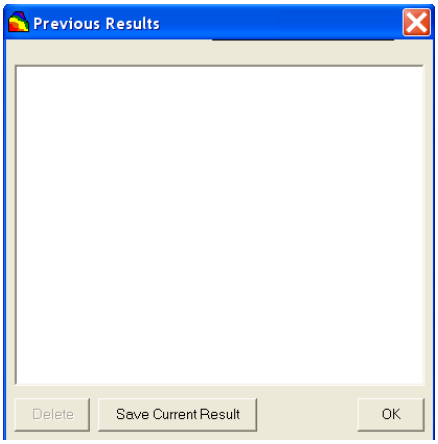


Notice that the epth layer drop-down list is now a transect selector.

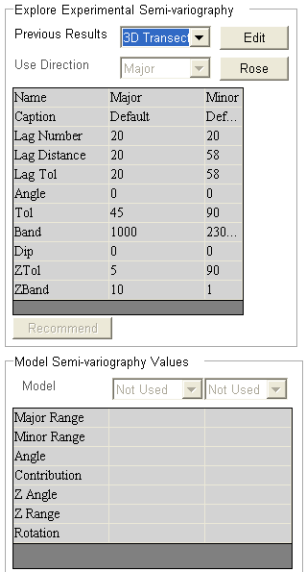
You can now switch between transects and observe variography results along each vertical dip transect. Before you begin exploring this feature, let's save these results. Variography generation can take some time and you don't want to have to repeat the calculations each time you want to see the results. The *Rose* button runs the calculations from scratch every time you press it. To avoid this, let's save the results by pressing the *Edit* button next to the *Previous Results* drop-down list.



You will be presented with the Rose/Correlation Result manager. This manager will contain a list of all the results you have previously saved. You can delete and save new results here. The list is empty as no previous results have been saved at the moment. Press *Save Current Result*.



Enter "3D Transects" and press *OK*. Press *OK* to the Rose manager as well. In the *Correlation Modeling* parameter window select *3D Transects* from the *Previous Results* drop-down list.



Notice that the parameters we used in this calculation are entered into the *Major/Minor* columns, but everything is disabled. This makes sense because you cannot modify the parameters of a

stored result. If you want to make parameter changes and rerun the model, select *None* in the *Previous Results* drop-down list again.

As you change the Z dip in the main tool bar, you will notice that for each dip value, there is a relatively constant N-S trough. This is what we would expect with the presence of the high value samples. Results are not markedly different among transects, although the trough intensifies as you get around -1 to -2 degrees. This is also expected.

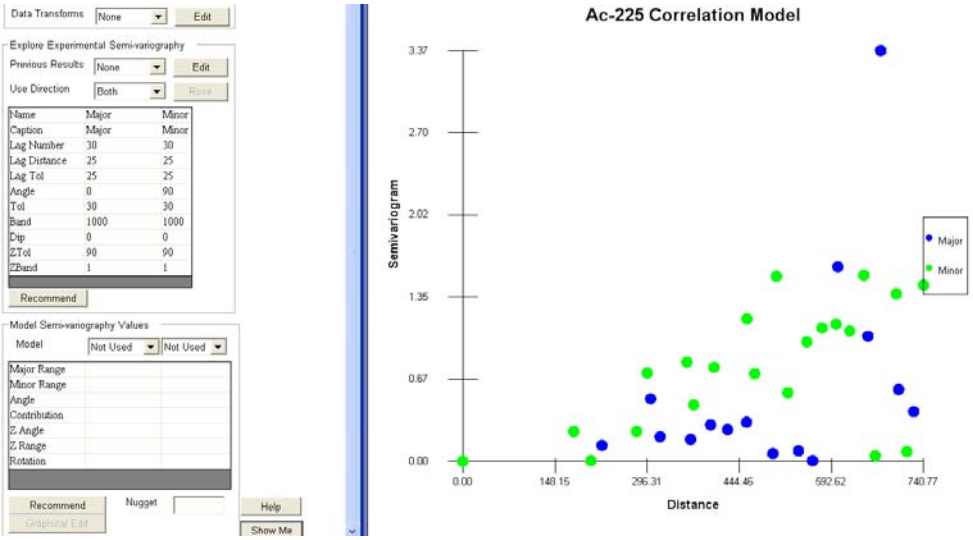
Based on the visualization of the data in 3d, there are likely two different correlation structures going on here. Low values are uniformly scattered in an almost isotropic pattern, while the high values are dipping along an N-S transect. To really capture this contrasted spatial pattern, we will need to use an indicator-based formulation. This is discussed in the section on correlation modeling in preparing for indicator kriging below.

These two applications are only an introduction to correlation analysis. Data sets will widely vary, and some patience will be required to accurately capture patterns observed in the data.

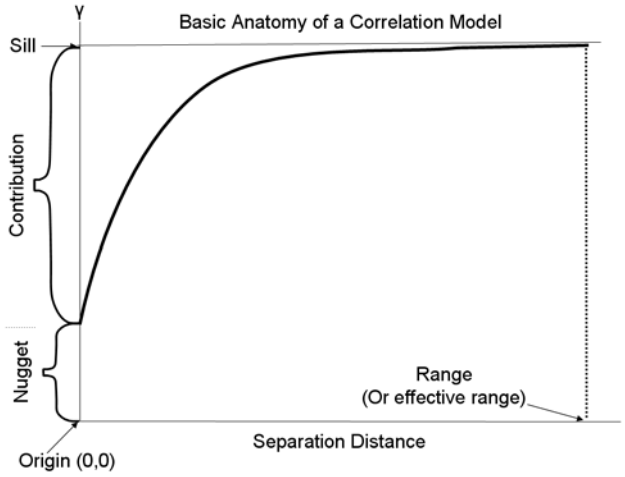
Correlation Modeling in SADA

For kriging or simulation procedures, one needs to know the semi-variogram values for any separation distance *h* and azimuth α. At this point, we only have these values for a few discrete lag distances. Therefore, we need to fit a model to the experimental semi-variogram values so that a semi-variogram value can be derived analytically for any vector *h*. Let's do this for the Ac-225 model.

Switch back to *Ac-225*, select *Model spatial correlation* as the interview and click on the *Correlation modeling* step. Make sure your parameters are as follows, then press the *Show Me* button in the lower-right of the parameter window.



We need to fit a model through these points. There are three possible correlation models in SADA. Each one has a particular shape, and basic models can be combined to create additional shapes. Each model has the same basic anatomy defined by sill, nugget, contribution, and range parameters. Please refer to the following picture as we talk about each parameter.



Sill

The sill is the highest value that the model will reach or asymptotically approach. The sill is equal to the nugget plus the contribution.

Nugget (n_0)

The nugget parameter describes how the model behaves as it approaches the origin of the graph. For example, a nugget of zero means the curve of the model passes through (0,0). A nugget of one means that the model passes through (0,1). The interpretation of the nugget is that for some reason as data become closer and closer (asymptotically closer), their variation does not approach zero (Goovaerts, 1997). This feature is used to account for things like sampling errors and/or sudden changes in an attribute (think of a gold nugget buried amid sand).

Contribution (c)

This defines how high model values increase relative to the nugget value.

Range (a)

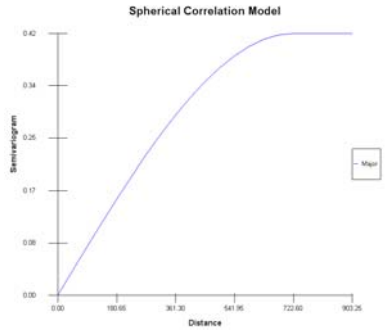
This is the point at which the model reaches its sill value. Some models never reach their sill value, but rather only asymptotically approach it. In this case, the range is sometimes referred to as the effective range or that range where it is essentially the same as the sill.

SADA provides three models based on Deutsch (1995): Spherical, Exponential, and Gaussian.

Spherical

This model rises from the nugget value to its sill (n_0 + c) value at range (a) and then flattens out immediately.

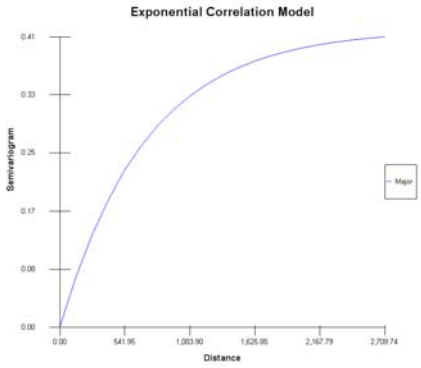
$$g(h) = nugget + c \cdot \begin{cases} 1.5 \cdot \frac{h}{a} - 0.5 \cdot \left(\frac{h}{a}\right)^2 & \text{if } h \le a \\ 1 & \text{otherwise} \end{cases}$$



Exponential

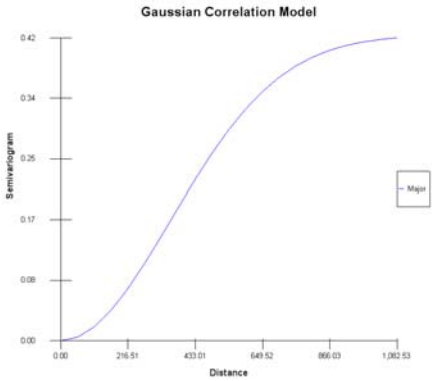
This model rises from its nugget value and asymptotically approaches the sill at an effective range of 3 times *a*.

$$g(h) = nugget + c \cdot \left[1 - \exp(-\frac{h}{a})\right]$$



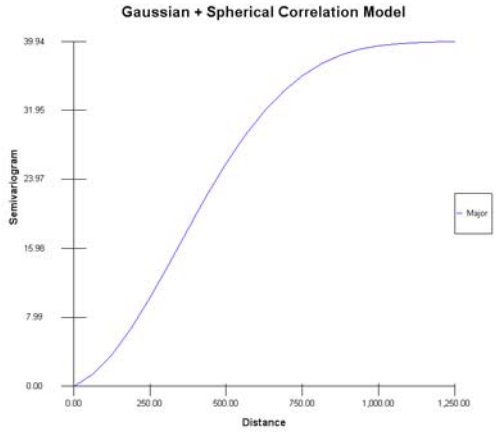
Gaussian

This model rises from its nugget value and forms an “s” like shape as it asymptotically approaches its sill at an effective range of $3\sqrt{a}$. This model is known to cause some instability in the kriging solutions. It should be used only with great caution and always in combination with a non-zero nugget effect.



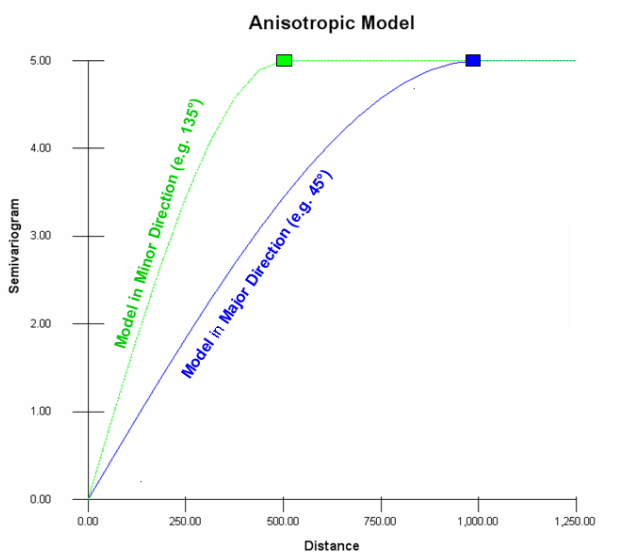
$$g(h) = nugget + c \cdot \left[1 - \exp(-\frac{h^2}{a^2})\right]$$

It is possible to nest these models to create different shapes. For example, the following figure combines a Gaussian model with range 500 and contribution of 30 with a Spherical model of range 1000 and contribution of 10. A nugget of zero is used.



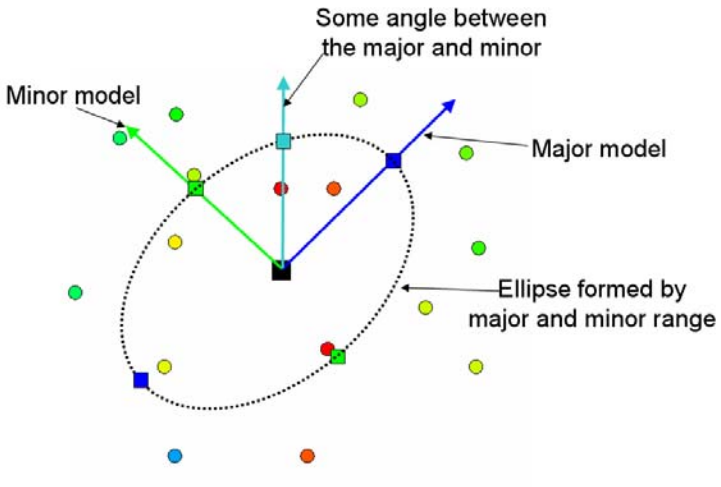
Notice how the shape of the function begins with a slight “s” hook near the origin and then resembles a spherical shape more at the larger distances. Typically, a single model is usually sufficient.

If there is anisotropy, you must also specify a minor direction model. The only parameters that distinguish a major model from a minor model are their ranges. The two models must be of the same type (spherical, exponential, or Gaussian) and have the same sill, nugget, and contribution. In the following image, a spherical model is used with a major range of 1000 and a minor range of 500.

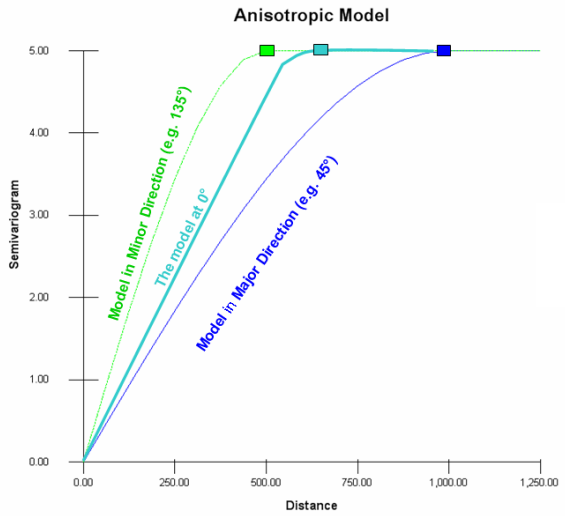


You will often see semi-variogram models shown in this manner: a minor direction model and a major direction model. These models are used to decide what weights will be assigned to data points in the neighborhood of the node we wish to estimate. So clearly, if a data point lies exactly in the major direction (relative to the node), then the major model will be used. If a data points lies exactly in the minor direction (relative to the node), then the minor model will be used. What about all the directions in between? Most data in fact will lie somewhere off one of these directions. In the figure above, the range/sill point of both models is denoted by

rectangles. Let's show where these would be found relative to some arbitrary node we wish to estimate or simulate.

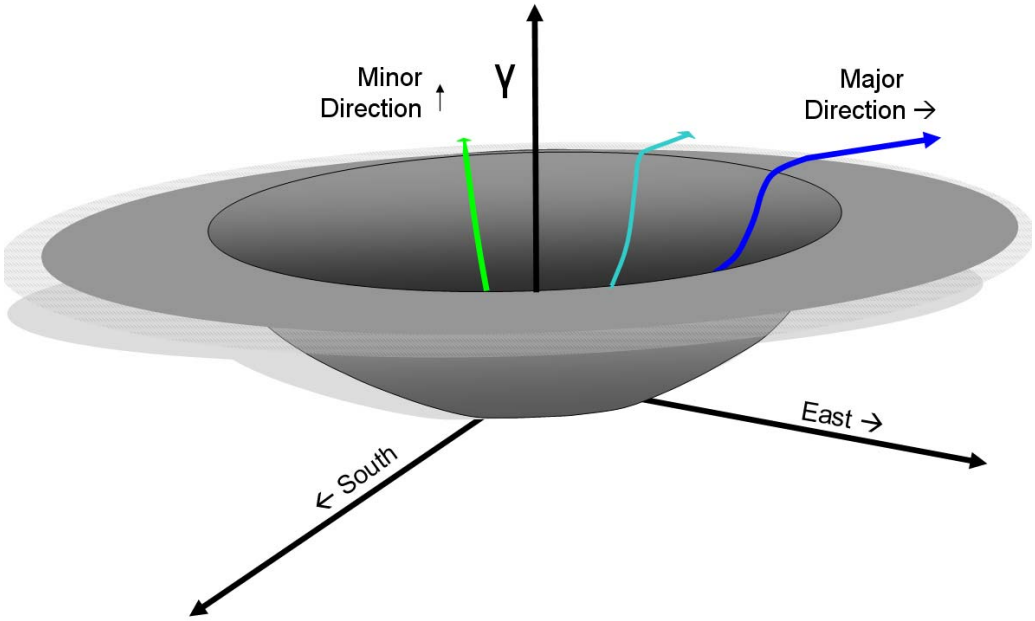


Suppose you were to stand a z-axis up on its end at the center of the node we wish to estimate (black square) and then draw the major correlation model along the major direction and draw the minor model along the minor direction. The image above is what you might see if you were to look straight down from above the site. The major and minor models would just look like straight lines because you're seeing them from above. Now the major range and minor range rectangles seen in the former image are now drawn in their proper places once again in the graph above. Since correlation models are mirror, or reciprocal, in nature, you can also draw them on both sides of the node. Notice how these can form the vertices of an ellipse. This ellipse will pass exactly through the range/sill point on the major model and exactly through the range/sill point on the minor model. It will also exactly pass through range/sill point of any angle found between them. In the image above, the light blue line is some angle (e.g., 0°) between the minor and major angles. This is how any angle between the major and minor direction is handled. In fact, one could also plot the model along this angle in a standard graph along with the major and minor directions.

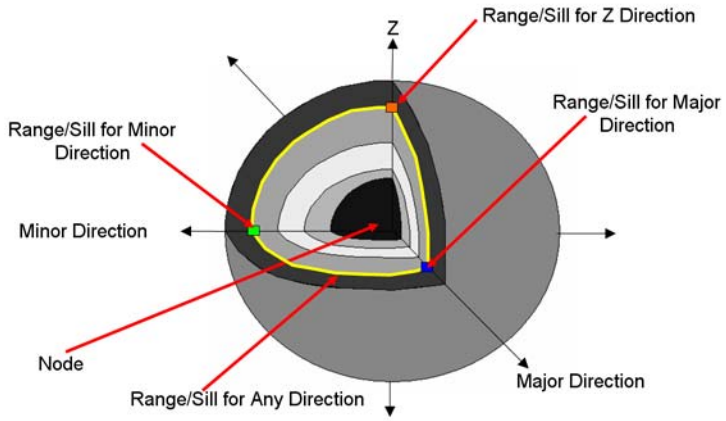


The following image is a 3d representation of this same model viewed from an off angle. If you could see every single model between the major and minor direction, it would form a 3d object

not unlike a bowl. The interior of the bowl would follow the shape of the model. So, it would appear Gaussian, exponential, or spherical. Outside the range, the surface would flatten out and continue indefinitely. In the following image, the surface is colored in gray, but the three specific directions are shown in the image.

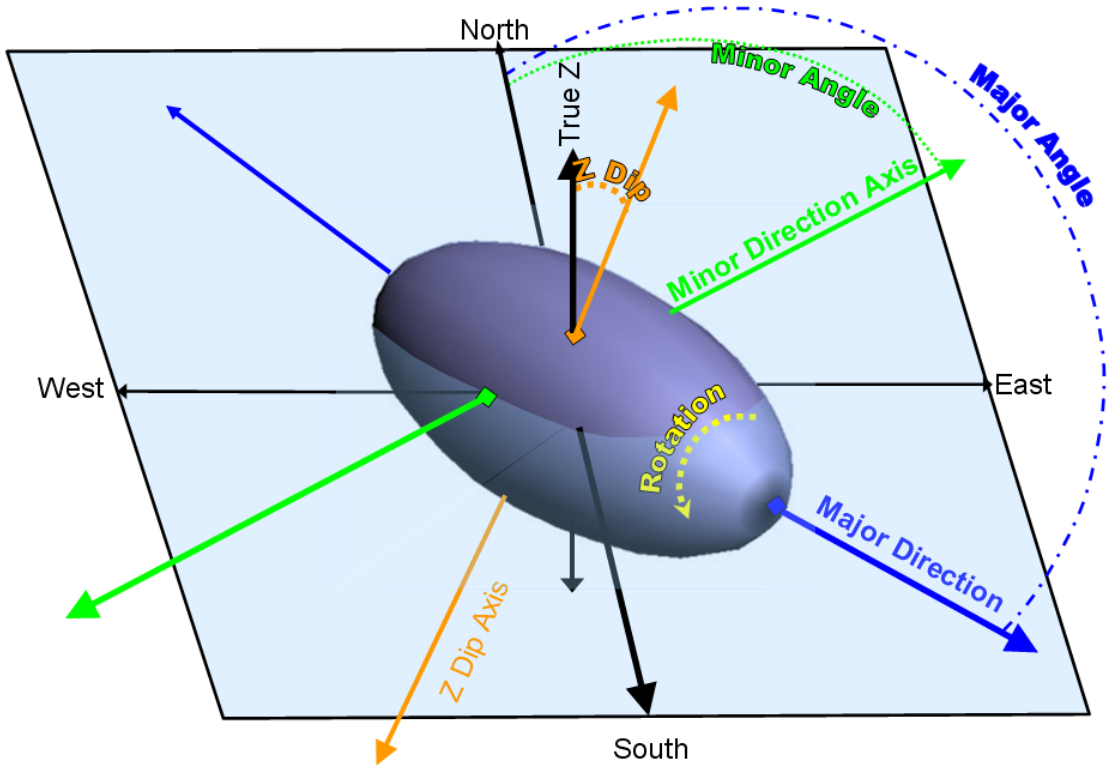


Correlation modeling in 3d follows the same type of thinking. It is a little more difficult to visualize. Since we are talking true 3d, we can't use the z-axis for our semi-variogram axis anymore. This will have to be the true z-axis (depth below surface). Instead, the model values will have to be represented by color gradations. In 3d, we have a correlation volume rather than a planar geometry, as seen above. It would look like a sphere or ellipsoid with the unestimated node at its center. If we could cut away a section of the sphere, we might recognize some familiar items.



At the center is the unestimated node. The blue and green range/sill rectangles for the major and minor directions are shown just inside the outer hull. They are now joined by the z range/sill shown in orange. The ellipse range/sill that we showed earlier is now a range/sill isosurface. In the image above, it's the yellow peel or line you see in the transect. It's a bit complicated, but spend some time with it and you'll get it.

It can get a bit more complicated in 3d. There is also a z angle and a rotation parameter as well. Z angle can be important. Rotation has less use in practice, but there are applications where it may be useful. In the previous image, we saw a depiction of an isotropic 3d correlation model. We know it is isotropic by looking at it because it's a sphere. To show the effect of these new parameters in 3d, we'll need to use an ellipsoid. We'll remove the cutaway to simplify the picture and see what the ellipsoid will look like when we apply z angle and rotation.



**Base Ellipsoid shape taken from Wikipedia Ellipsoid Topic. Picture freely available for use under GNU.*

(http://commons.wikimedia.org/wiki/File:Ellipsoid_3d.jpg)

So, in 3d space the correlation model geometry is an ellipsoid. We have elected to trim the ellipsoid back to the range/sill surface for all directions. The range/sill location for the major, minor, and z directions are noted again by rectangles of corresponding color. They are by definition on the surface of the ellipsoid if we are trimming the ellipsoid just to the model range in each direction.

The ellipsoid is tipped forward into the horizontal plane through the node center (noted by light blue area) by the Zdip parameter (in negative degrees). The rotation angle is how this ellipsoidal shape is rotated along is major axis. Just as in search neighborhoods, great care must be made in specifying the z angle. The z angle will normally take very small values due to the "flatness" of most environmental sites (see previous discussion on search neighborhoods). The z dip should be inferred by a careful analysis of semi-variogram values.

Correlation modeling in SADA

In SADA, these parameters are found in the information block at the bottom called *Model Semi-variography Values*. There are three columns. The first column just describes the parameters that are needed by the models. These are the parameters that define an ellipsoidal geometry for

the correlation model. As mentioned earlier, there are only 3 possible models in SADA: spherical, exponential, and Gaussian.

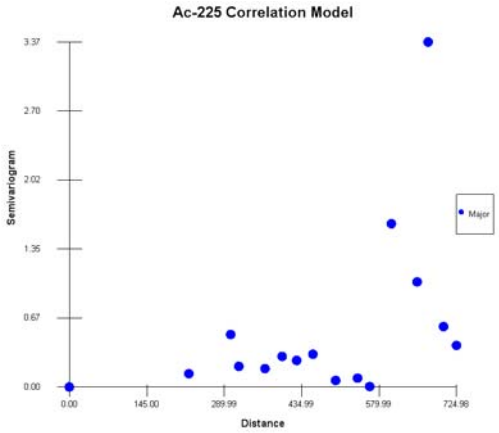
If you click on the first *Model* drop-downlist, you will see these three options. If you click on the second *Model* drop-down list, you will see these three options repeated. You have access to these two models because SADA allows you to use nested models. As previously mentioned, a nested model is really just the sum of two individual models. This allows you to create different shapes that might fit your data better. Most of the time, one model is sufficient.

Because of the alignment of each model under the *Major* and *Minor* columns in the variography block above it, it is easy to make the mistake that the first *Model* is associated with the *Major* column and the second *Model* is associated with the *Minor* column. This is NOT the case. Both the major and minor directions are fitted with one model only. This model can be a nested model (by using both model columns), but still there is only one model.

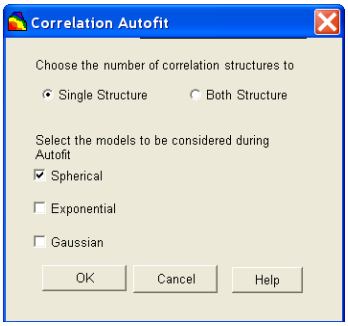
So, how do you take one model and fit it through two sets of points? Recall that the semi-variogram shown above is presenting the variography calculations along only two directions: 0 deg (North-South) and 90 deg (E-W). Each of the three models will have a major and a minor range parameter. This is the only way they can be distinguished from each other. It is a limiting factor but mathematically necessary for reasons outside the scope of this guide. The best way to understand how this is done is to work through an example.

From a practical view, the plot of semi-variogram values for both the major (blue) and minor (green) directions is a messy result. It can be difficult to determine how to choose a model that will fit through these points. Methods, such as minimizing the least squares, are not recommended for a variety of reasons discussed shortly. As a result, many analysts fit the model visually. SADA does provide a *minimize by least square* option, although it is recommended only as a starting point. To make the process more tractable, let's start with the major (blue) values.

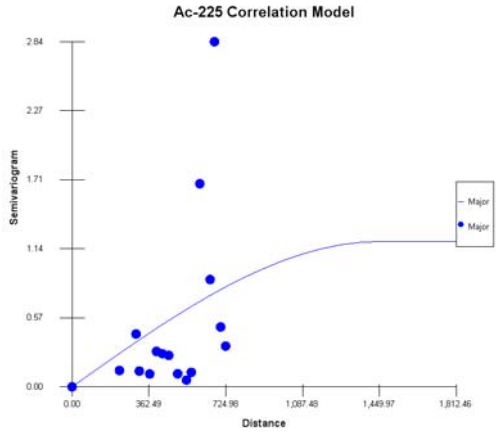
In the information block *Explore Experimental Variography*, Select *Major* in the *Use Direction* drop-down list and press the *Show Me* button at the bottom of the parameters window. You should get the following result.



We will now use the *Recommend* button to auto-fit a line through these points. In the *Model Semi-variography Values* block, push the *Recommend* button. The autofit window will appear.



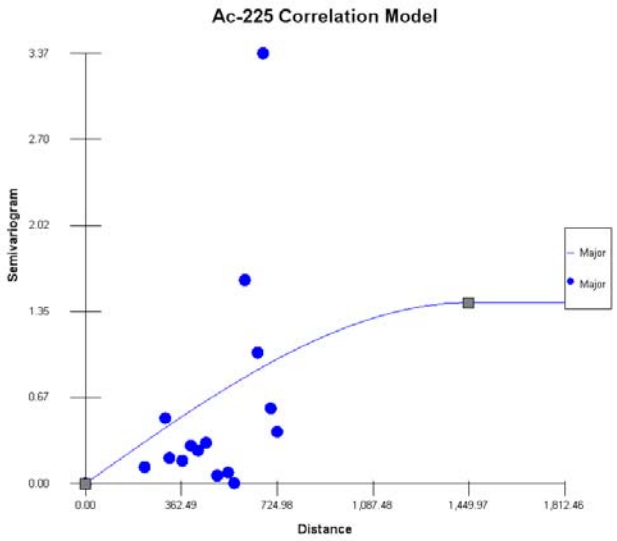
For this example, we'll use only a single structure (not two nested models) and we'll only look at the spherical model. If you select more than one basic model, SADA will try each model and determine which one is best according to a least-squares criterion. For now choose only the *Spherical* model. Press *OK* and SADA will display a least squares fit line through the data.



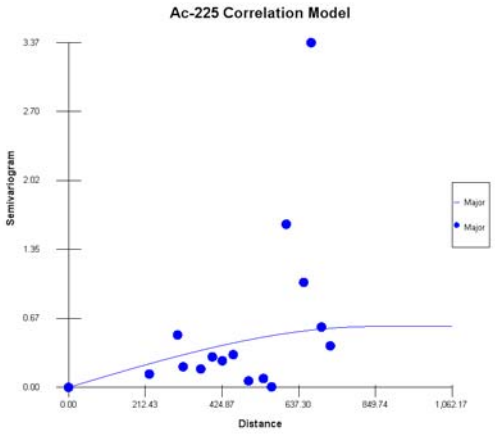
This is not necessarily the optimal answer for two reasons. First, outliers can negatively affect the total fit as the model stretches to accommodate them. Second, from a geostatistical perspective, it is better to concentrate on fitting your model at short distances rather than worrying about the total fit. Those data points that are closest to any estimation node will weigh more heavily in the estimation and therefore should be more accurately fitted.

You can play around with the model parameters by manually entering different parameter values. Now note that the standard xy type graphs for correlation models do not show the impact of angle choices or z range. These would only be seen in a 3d rendering, as we showed earlier. You'll need to examine the choice for angles and z range by using rose maps.

Another way to proceed is to graphically edit the model. You can only do this when you have a single basic model. With the correlation model up, press the *Graphical Edit* button. Two gray rectangles will appear, one at the origin and one near or at the range/sill point.

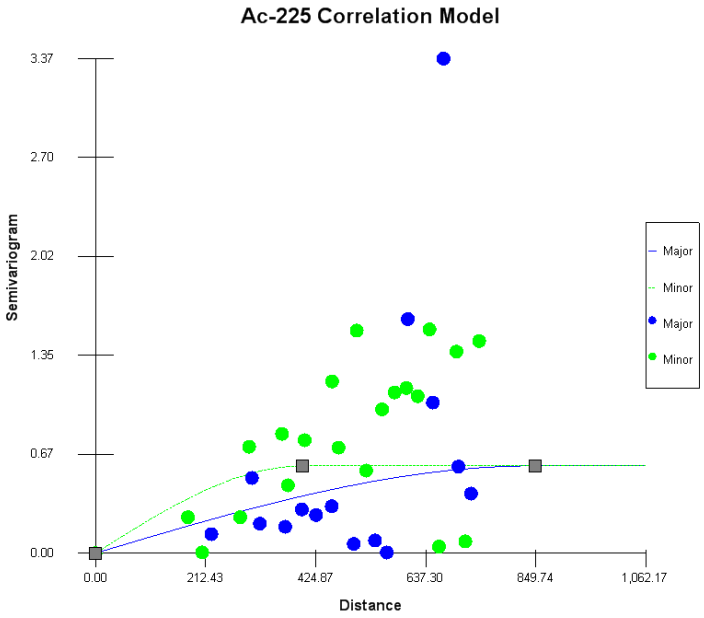


These edit boxes are moved by left mouse clicking, holding the button down, and sliding the mouse around. You can move the left-most gray rectangle only on the y axis. This is the nugget parameter. You can move the right-most gray rectangle up and down or left and right; however, it cannot move below the nugget as this is nonsensical (contributions can never be negative since they represent variances). SADA will simply stop moving the rectangle if you try and move below the nugget. It is possible to edit both the major and minor ranges at once, but it can be trickier. Spend some time moving these points around, and when you're done press the *Graphical Edit* button again. The following image shows a correlation model modified in this manner.

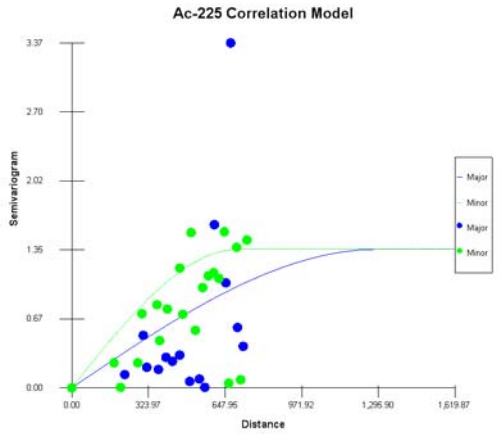


If you do not have anisotropic variability, you are done. If you do, you can now turn on the green dots and attempt to fit these as well. In the *Explore Experimental Variography* block, switch the *Use Direction* drop-down list to *Both*. Press the *Show Me* button again, and the green dots will reappear.

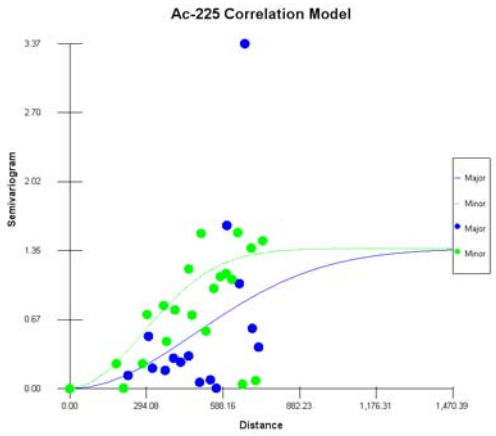
If you were to press the *Graphical Edit* button again (don't do it now), you would still get only two graphical edit boxes. In reality, there are three graphical edit boxes: one for the nugget, one for the major range, and one for the minor range. Because the major and minor ranges are currently the same, these two edit boxes are sitting on top of each other. Let's fix this first. For your spherical model, change the minor range to roughly half the major range and press the *Show Me* button again. You should now see three boxes



It is clear that in our image, we may have to increase the sill for both models in order to accommodate the minor direction (green dots). As you work with all three edit points, you will notice that you also cannot make the major range smaller than the minor range and vice versa. If the minor range runs into the major range during editing, they will simply collapse into a single edit box again. You'll have to manually enter a new minor range into the parameter window to separate them. Go ahead and play with the editing now. The following image shows the results of such a manual edit.



Now try to fit other types of models, such as the exponential model or the Gaussian model. You can practice using the autofit routine and graphical edits as well. The following image shows that a Gaussian variogram model may fit the data best. Unfortunately, the Gaussian model is notorious for causing instabilities in the computations when used without a nugget effect.



Each geostatistical algorithm will require a slightly different variation on the correlation model. We will now talk about correlation needs under each approach.

Correlation Modeling for Ordinary Kriging

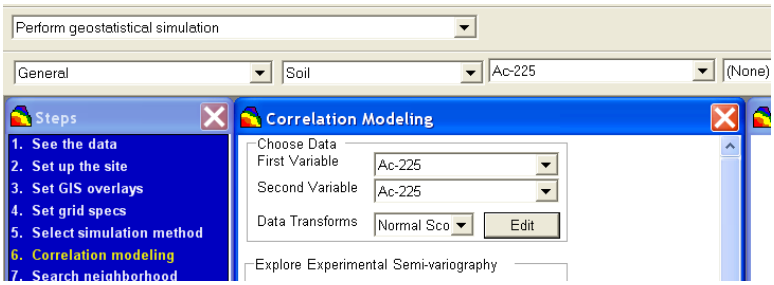
Ordinary kriging requires a single spatial correlation model for the untransformed data. The correlation modeling step is also available under the *Interpolate my data* interview. If ordinary kriging is selected as the interpolant, SADA will restrict the variable options to only the contaminant currently selected and the transform option will be restricted to *None*. Switch back to *Soil* and *Ac-225*. Select *Interpolate my data* as the interview. In the *Interpolation methods* Step, choose *Ordinary Kriging*. Click on the *Correlation model* parameters step again. Notice that the parameters we entered earlier are still there. This is because we were working with untransformed data, which is exactly what ordinary kriging needs. Also, click on the data transforms drop list and notice that the only option is *None*. Because we are doing ordinary kriging, SADA restricts our choice to this option only.

If you worked through the previous section on practicing with correlation modeling in SADA, you have done the work you need to do in preparation for ordinary kriging of Ac-225. If you are actually interested in doing ordinary kriging now, you can proceed to the section below on ordinary kriging.

Correlation Modeling For Sequential Gaussian Simulation

Sequential Gaussian simulation (SGS) creates equi-probable realizations of the spatial distribution of contaminants across the site. We will not discuss SGS at this point other than to say it requires a correlation model for the normal score transformed data. The normal score transform was discussed earlier in this chapter and converts any data set into a normally distributed result. SGS requires normally distributed data, and since most environmental data sets are not normally distributed, this is an important step.

With *Soil/Ac-225* selected in the file SpatialCorrelation.sda, switch the interview from *Interpolate my data* to *Perform geostatistical simulation*. Click on the step *Select simulation method* and choose *Sequential Gaussian Simulation*. Click on the *Correlation modeling* step.



To calculate the semi-variography for normal score transformed data is very easy. In the *Choose Data* block, select *Normal Score* for the *Data Transforms*. If you are in the *Perform geostatistical simulation* interview with *Sequential Gaussian simulation* selected, then this is the only option. You will now perform the correlation assessment activities as you normally would. SADA will do the transformations of the data behind the scenes for you.

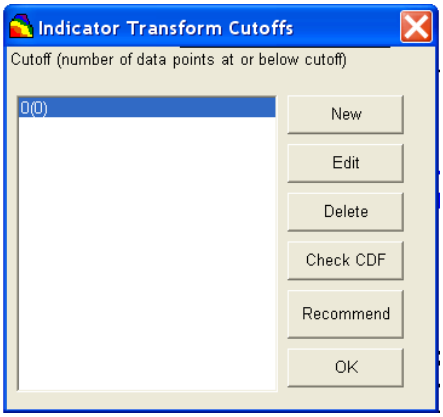
Correlation Modeling for Indicator Kriging/Indicator Simulation

When defining the point-wise probability distributions (the distribution of possible concentration values at an unestimated node), ordinary kriging and sequential Gaussian simulation assume a normal distribution model. If your data is not normally distributed, then you may consider using indicator kriging or sequential indicator simulation (keep in mind the normal score transform feature converts any distribution into a normal distribution, permitting you to use sequential Gaussian simulation-see above).

Rather than assume any distribution, the probability distribution or, more accurately, the cumulative distribution function is generated numerically at each point. Recall that a cumulative distribution function specifies the probability that a response will be below any given threshold. Indicator methods accept a set of thresholds (1pCi/g, 3pCi/g, etc.) and determine the probability that the concentration is less than each of these thresholds individually. Therefore, you will need to specify the set of thresholds and provide a correlation model for each.

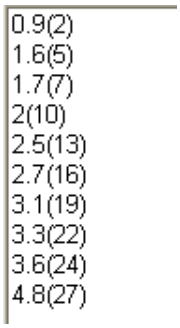
The choice of thresholds may be motivated by different objectives. It is usually important to get a good representation of the range of possible values. This can be done by setting indicator thresholds to a set of percentiles (e.g., 10^{th}, 20^{th}, 90^{th}). Another important threshold may be the decision criterion under consideration. If you have a decision criterion in mind and don't specify it as an indicator threshold, then the code will be forced to estimate the CDF at your criterion by interpolating between the thresholds just above and just below. SADA can help with these.

If you have not already opened SpatialCorrelation.sda, do so now. Select *Soil* and *Ac-225* and select *Model spatial correlation* as the interview. Click on the step *Correlation Modeling*. In the *Choose Data* parameter block, press the *Edit* button.

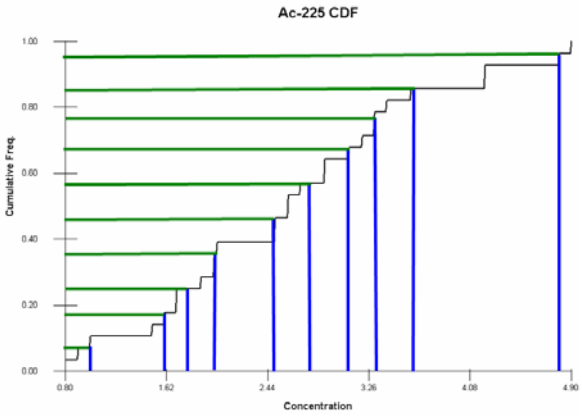


This is the *Indicator Transform Cutoffs* manager. Sometimes the word "cutoff" will be used, but it is synonymous in this context with threshold. The first thing we'll point out is the *Recommend* button. This button will analyze your data and produce ten indicator thresholds that roughly correspond to the 10th, 20th, etc. percentiles of your data set. Press this button now and SADA will warn that any previous thresholds will be eliminated. Say *Yes*.

SADA selected the following thresholds:

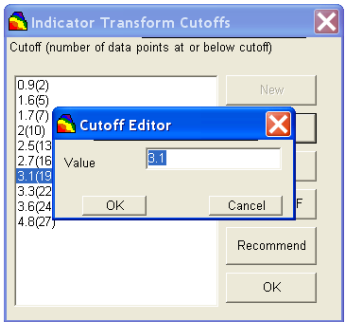


The first number is the indicator threshold. The second number, in parentheses, is the number of data falling below this value. So, you can see that this number increases with increasing threshold values. If you plot these results on a cumulative distribution function, you will see a fairly even spacing among CDF values and irregular spacing among concentration values. This is normal, as the goal is to have about 10% between each CDF value, and concentration selections are adjusted to accommodate.



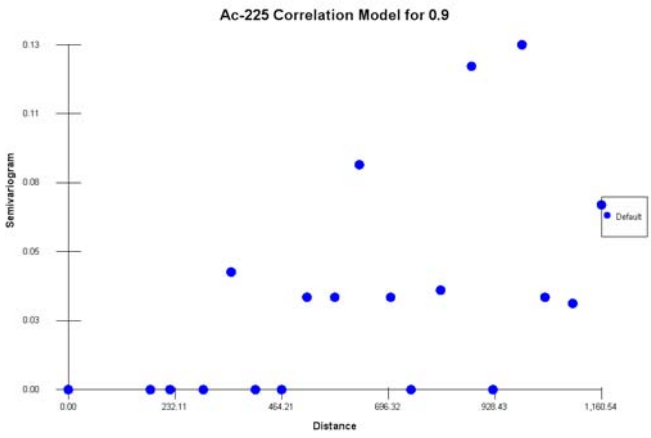
Suppose now that we have a particular decision threshold in mind (for example, 3.0 pCi/g). SADA currently allows a maximum of 10 threshold values. This means that you will need to give up one of the 10 currently selected. There are a couple of ways to make this choice. First, you simply find a value that is very close and adjust it slightly. If the closest threshold isn't too close, you may need to slightly modify several thresholds in one direction or another. The important thing is to really calibrate the model right around your decision threshold. So, lower threshold values may not be as important as the high concentration thresholds and one may be dropped.

Let's adjust the 3.1 cutoff to 3. From the list of numbers, select the 3.1 cutoff and press the *Edit* button.



Enter 3.0 and press *OK*.

While assessing the correlation structure for each of these thresholds, you may notice that in the very-low and very-upper ends, variogram values are very erratic, or at least variography values are not well suited for some of the models. Consider the lower end of threshold values. When the indicator transforms are applied at the 0.9 threshold, only 2 out of 28 samples will be coded as 0, while the remaining 26 samples will be valued at 1. This means that for many data pairs, the two observations will be 1 and their difference will be 0. When a 0 indicator is encountered, the semi-variogram suddenly bounces higher. Consider the following semi-variogram calculation for 0.9 pCi/g.

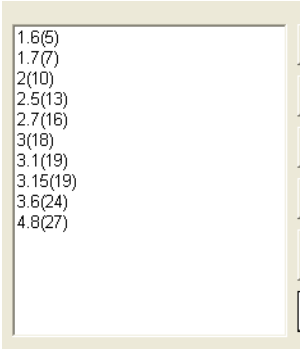


A number of the results fall on the zero value line, which means that for that lag distance, no sample less than 0.9pCi/g was encountered. If the decision criterion is a very low value, you may have difficulty using indicator kriging to model your contaminant and corresponding decision. You may need to consider a normal score transform and use ordinary kriging instead.

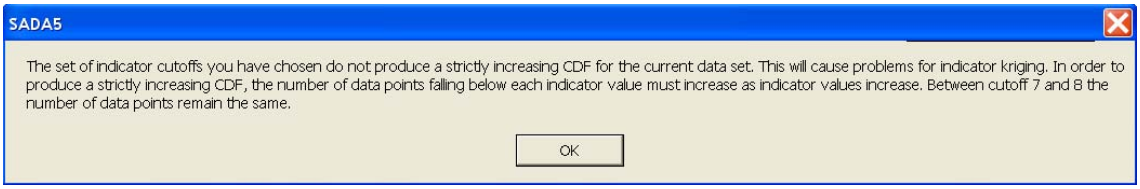
As a demonstration, let's eliminate the 0.9pCi/g threshold due to the semi-variogram problems and also because it is considerably lower than our decision rule. Select 0.9pCi/g and press the *Delete* button. Answer *Yes* to the confirmation question.

We can add a new threshold by simply pressing the *New* button. Let's add another cutoff near the 3.0 pCi/g. Press *New* and enter 3.1 again. Press *OK*.

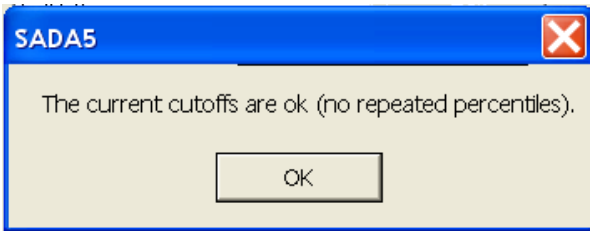
When performing indicator kriging or simulation, your global CDF must be monotonic increasing. This simply means that every threshold value you select must have more data values below it than the previous threshold value did. So for example, the following selections violate this rule.



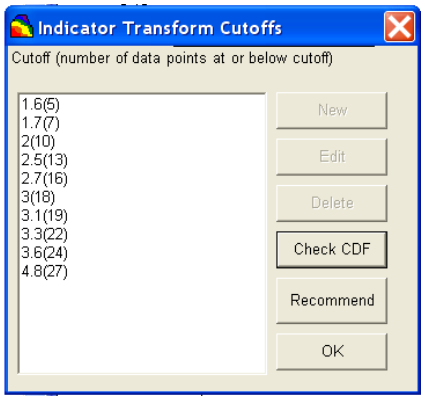
The problem here is that the thresholds 3.1 and 3.15 both exceed the same number of observations: 19. SADA can quickly check this constraint for you when you press the *Check CDF* button. If you have a problem such as the one above, SADA will warn you.



In this case, you will need to manually adjust the thresholds so that no two numbers in parenthesis (the CDF) are the same. When the threshold selection is acceptable, you get an ok message.



Make sure you have 9 cutoff threshold values that look like this and Press *OK*.



Back in the Parameter window, in the *Choose Data* parameter block, select the drop-down list button next to *Data Transforms*. Notice that all your cutoff criteria are now loaded in this list. The only time you'll see this choice in the list will be when you have the interview *Model correlation* selected or anytime you have *Indicator Kriging* or *Sequential Indicator Simulation* selected. Otherwise, SADA will always restrict the data transforms to those that make sense in the current context. For example, if you had selected *Ordinary Kriging* under the *Interpolate my data* interview, all these cutoffs would no longer appear. It doesn't mean they have been lost. SADA is just hiding them in that context, since they are not relevant to ordinary kriging.

Select 1.6 and press *Recommend* under the *Explore Experimental Semi-variography* block and *Recommend* under the *Modeling Semi-variography Values* block (choose *Spherical*). Press *Show Me* in the lower-right-hand corner of the parameter window.

Repeat this process for 1.7, 2, 2.5 and so forth. What you are doing is establishing a correlation model for each of these indicator thresholds. This is required by indicator kriging and sequential indicator simulation alike. You will learn more about these interpolators/simulators in the next chapter. Save your file.

Correlation Modeling for Cokriging

Cokriging allows you to support your geospatial model with other data types that are correlated with the contaminant you are trying to model. For example, gamma count data might provide relevant information for the estimation of Cs-137 in the soil. Depending on the flavor of cokriging you choose, you may need to estimate the correlation model for combinations of your contaminant and your auxiliary data. Suppose you are interested in modeling Contaminant "A" and wish to supplement with Field Detection Device "B." You may be required to estimate the spatial correlation between A and B. In this case, the "head" in the variogram discussion will be an A measurement and the "tail" will be a B measurement. This allows you to see how Contaminant A and Field Device B spatially co-vary as the separation distance increases. If you also have geology result "C," the number of correlation models can get out of hand. In a fully specified cokriging model, you would need the following correlation models: A/A, A/B, A/C, B/B, B/C, and C/C. There are some simplifying options, and when you select cokriging as your interpolant, you will be asked to choose among some easier options. Suppose for now we only have Contaminant A and Field Device B.

Unrestricted coregionalization model (very hard to implement)

In this option, you must model the correlation model for A, for B, and for A/B. Here, you are free to model each variogram separately; however, there is no guarantee that the cokriging system always has a solution, in which case a warning message is issued.

Linear model of coregionalization (hard to implement)

Here, both A, B, and A/B semivariograms are modeled using the same type of model, the same range, and the same anisotropy parameters. The sill is permitted to vary, but a constraint called the Cauchy-Schwarz (CS) inequality must be met for each basic model.

$$\left|Sill_{1,2}\right| \leq \sqrt{Sill_1 x\ Sill_2}$$

Suppose we had the following models for Contaminant A with Contaminant B.

$\gamma_A(\mathbf{h})$ = 25 + 20 × Exp(range=2km)

$\gamma_B(\mathbf{h})$ = 9 + 29 × Exp(range=2km)

$\gamma_{A,B}(\mathbf{h})$ = 2+ 18 × Exp(range=2km)

We would be ok, because

$\left|2\right| \leq \sqrt{25x9}$ (nugget model) and $\left|18\right| \leq \sqrt{29x20}$ (exponential model) (Goovaerts, 1997).

Intrinsic Model of Coregionalization (easy to implement)

This model is more constraining than the linear model of coregionalization in that one more condition is imposed: the contribution of each basic model must be the same across all variogram models. For example in the following intrinsic model, the contributions of each basic model, while satisfying the CS inequalities, represent 50% of the total sill:

γ_A(h) = 20 + 20 x Exp(range=2km)

γ_B(h) = 9 + 9 x Exp(range=2km)

$\gamma_{A,B}$(h) = 5+ 5 x Exp(range=2km)

With $\left|5\right| \leq \sqrt{20x9}$

Another way to think about the intrinsic model is as a set of variogram models that are all proportional to the same function.

The model above can be expressed as a rescaling of the following model:

0.5 + 0.5 × Exp(range=2km).

The rescaling constant would be 40(γ_A(h)), 18(γ_B(h)), and 10($\gamma_{A,B}$(h)) (Goovaerts, 1997).

Markov Model Of Coregionalization (very easy to implement)

The Markov Model (MM) is the most straightforward model of coregionalization: only one direct variogram needs to be modeled and the other variograms are derived through a proportional relationship. Two Markov models are available: MM1 and MM2.

Markov Model (1) states that the cross variogram is proportional to the variogram of the primary variable:

$$\gamma_{12}(\mathbf{h}) = \sqrt{\frac{C_{22}(0)}{C_{11}(0)}}\rho_{12}(0)\gamma_{11}(\mathbf{h})$$

Where $C_{11}(0)$ and $C_{22}(0)$ is the variance of the primary and secondary variables respectively, while $\rho_{12}(0)$ is their correlation coefficient. This model is used in the framework of collocated

cokriging, which does not require knowledge of the variogram of the secondary variable since only one secondary datum is used for interpolation.

Markov-Bayes

This is a special application of the Markov model for probability mapping using both hard and soft (prior probabilities) indicators. Rather than ordinary cokriging, an indicator cokriging approach is used, whereby hard data are first converted to 0s or 1s, depending on whether they exceed a specified criterion. The cokriging method is then applied to these 0s and 1s, along with the prior probability map. This results in an updated probability map that contains the influences of both the hard and soft data. In SADA, one must first create a prior probability map. (This is done by creating a user defined map and then choosing the interview *Update my probability map*.)

The rescaling for the Markov-Bayes model is slightly different than a traditional Markov approximation. Let Y represent the soft prior-probability map data. Let I represent the sample data you wish to interpolate, where values have been transformed to zero if the measured value is less than or equal to the decision criterion, and 1 otherwise. In the *Correlation modeling* step, you develop the correlation model, γ_I, for the indicator transformed data set. Then, borrowing from Goovaerts (Geostatistics for Natural Resources, 1997), we have that

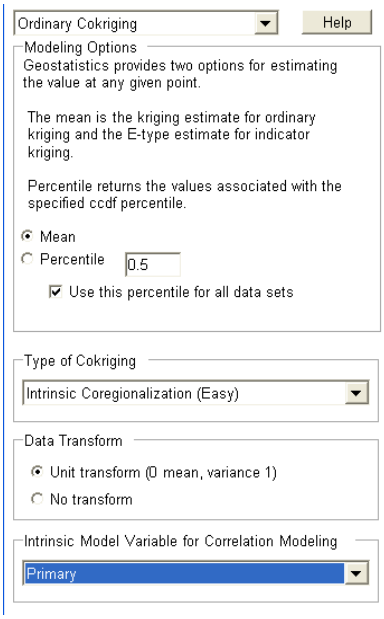
$$\gamma_Y(0) = |B|\,\gamma_I(0)$$
$$\gamma_Y(h) = B^2\gamma_I(h) \quad \forall\, h > 0$$

Where the coefficient B is defined as the difference between two conditional expectations:

$$m^1 = E[Y(u) \mid I(u) = 1] \in [0,1]$$
$$m^0 = E[Y(u) \mid I(u) = 0] \in [0,1]$$
$$B = m^1 - m^0$$

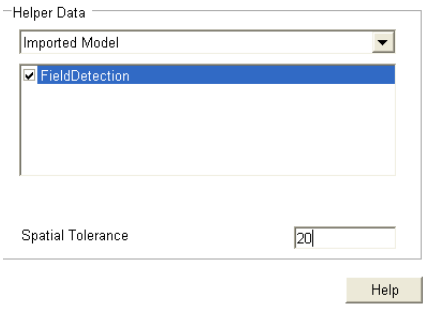
Try it Yourself

While much of this seems fairly intimidating, most of the effort has been worked out for you in the interface. For example, let's demonstrate how easy it is to build an intrinsic model of coregionalization. Later on, you'll revisit this example. Open up the file SpatialCorrelation.sda. Select *Soil* and *Ac-225*. In the interview list, select *Interpolate my data*. In the step *Interpolation methods*, select *Ordinary Cokriging* and make sure that you have *Intrinisic Coregionalization (Easy)* selected. We'll worry about what the rest of this means later.



With intrinsic coregionalization, you only have to model one correlation structure. You can choose to model either the Ac-225 data or *FieldDetection* results. Ac-225 is our primary variable because that is the contaminant we've selected.

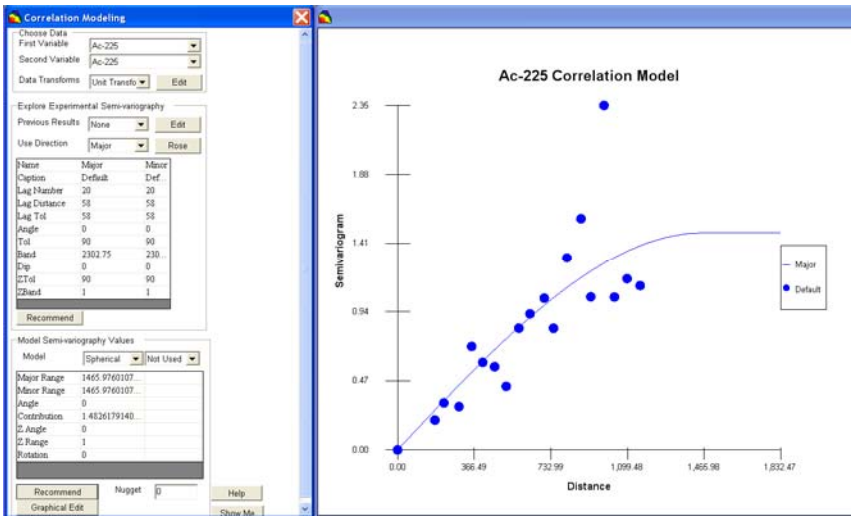
Click on the step *Choose helper data* and make sure the parameter window looks like this:



This is where we tell SADA which secondary dataset we want to "help" us better model our Ac-225 data. Make double sure the *Spatial Tolerance* is 20.

Click on the *Correlation modeling* step. You can see that the only option is to perform correlation modeling on Ac-225. This is because we chose to only model our primary (Ac-225) variography. Notice that *Unit Transform* is the only option under *Data Transforms*. This is also caused by our previous step choice *Interpolation methods*. Don't worry about this right now. It will be covered again later. The point here is that all you have to do is the same thing we did in the previous example, and SADA will take care of calibrating the other correlation structures.

In the interest of practice, press both *Recommend* buttons on the *Correlation Modeling* block: variogram first and *Model* second (choose *Spherical*). You should get a reasonable correlation model on the first try.



Keep in mind that in practice, you would need to look at anisotropy and the rose diagram before moving forward with actual interpolation.

Summary

This chapter presented the central concept of correlation structures as well as how to compute and model them. This step is likely the most time consuming step you will encounter in the geospatial modeling processes. The next chapter introduces the different geospatial models and explains how correlation models that were developed in this chapter are used. If you wish to further your knowledge of spatial correlation structures, please refer to the books listed in the references section.

References

Deutsch, C. V. and A. G. Journel (1992). GSLIB: Geostatistical Software Library and User's Guide. New York, Oxford University Press.

Isaaks, E., Srivastava, R.M. (1989). An Introduction to Applied Geostatistics. New York, Oxford University Press.

Goovaerts, P. (1997). Geostatistics for Natural Resource Evaluation. New York, Oxford University Press

Chapter 31: Advanced Geospatial Methods Part II: Geospatial Modeling, Uncertainty Analysis, and Simulation

Spatial correlation assessment and modeling is a central requirement to all the geostatistical tools that will be covered in this chapter. If you have not read Chapter 30 or struggle with the material, we recommend you return to that chapter now as well as look at materials outside this user guide. More thorough treatment of the subject matter can be found in Isaaks and Srivastava (1989), Goovaerts (1997), and Deutsch and Journel (1992).

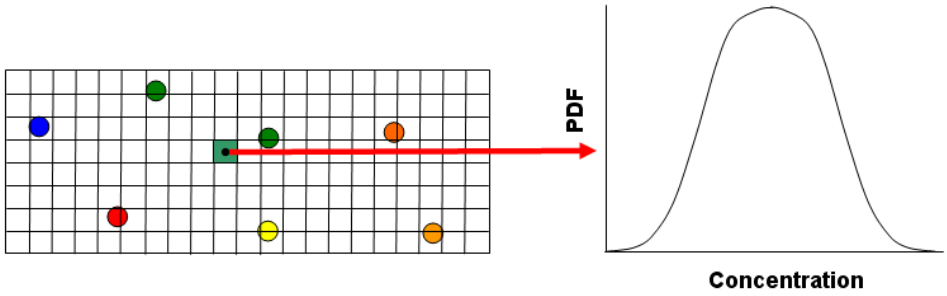
This chapter is divided into two parts. Part I covers the geospatial estimation methods: ordinary kriging, indicator kriging, and cokriging. These methods are useful in creating contour maps and modeling local uncertainty. Part II covers geostatistical simulation. Simulation routines produce multiple equi-probable realizations of what contamination across the site might look like. It can be useful to think of these as the Monte-Carlo versions of their kriging counterparts. These are useful not only for assessing point or local uncertainty but joint or spatial uncertainties as well.

Part I: Estimation and Local Uncertainty

Estimation refers to contouring the site using the data at hand and one of the Interpolation methods listed below. We've already seen estimation maps in Chapter 29, "Basic Geospatial Methods." By local uncertainty, we mean one of two things:

1) Uncertainty about the estimation value.
2) Whether the true value exceeds the decision threshold.

We can report these uncertainties in various ways, but the main idea is that it is only the uncertainty at each estimation node expressed as a cumulative density function or probability density function.



So for example, it is not possible to estimate the uncertainty about whether a group of estimation points will all simultaneously exceed the decision threshold. This must be handled with geostatistical simulation. Local estimation and uncertainty, however, are extremely powerful tools, as we shall soon see.

First let's take a moment to introduce some new concepts and a couple of notation adjustments. Recall from Chapter 29, "Basic Geospatial Methods," formulas were written for the estimation of an unsampled point as a variable on the left side of the equality, equal to some complicated summation on the right-hand side. Recall for example, inverse distance's formula:

$$u_0 = \frac{\sum_{i=1}^{N(u_0)} \frac{1}{d_i^P} u_i}{\sum_{i=1}^{N(u_0)} d_i^P}$$

Where u_0 is *the* estimated value at some unsampled location on the site. This notation and line of discussion will not work within a geostatistical context.

In geostatistics, there isn't a single estimated value at each point. Rather, there is a distribution of possible values at each unsampled point. Essentially, geostatistical models assume that each unsampled location is a *random variable*. A random variable is some variable, Z, that can take on any number of values drawn from a probability distribution, such as the normal distribution.

In the spatial modeling case, we don't have just one random variable, but rather every single unsampled location is represented by a separate random variable. Suppose we have a 10x10 grid defined across the site. Each node of the grid will have a separate (x,y,z) coordinate. We'll call this location $u_{(x,y,z)}$. So, the random variable associated with the location will be called Z, and to let you know that this particular random variable is located at $u_{(x,y,z)}$, we'll introduce the notation $Z(u_{(x,y,z)})$.

This notation can be cumbersome. Therefore, some references will instead refer to the location of the node in the sequence of grid nodes (e.g., 1st node, 2nd, node, … , Nth node). So instead, we have $Z(u_i)$. In some cases, such as this text, we drop the "i" as well, indicating any arbitrary node u: Z(u).

While there are many nodes and every node has a random variable, these random variables cannot simply take on any value they like regardless of what the neighboring data points and nodes are doing. Instead, they form a *random field*. A random field is a set of interdependent random variables. The correlation model joins these random variables by making the values they can assume dependent on each other.

Ordinary kriging

Ordinary kriging, like all other interpolation schemes used in SADA, assigns a weight to each nearby point in the estimation of an unsampled location. In particular, we have:

$$Z^*(u) = \sum_{\alpha=1}^{n} \lambda_\alpha(u) Z(u_\alpha)$$

Here, $Z(u_\alpha)$ is the value at some data location u_α . The u_αs are the n data locations, the $\lambda_\alpha(u)$s are the ordinary kriging weights and Z*(u) is the estimated value. Here the weights $\lambda_\alpha(u)$ are determined by the system:

$$\sum_{\beta=1}^{n} \lambda_\beta(u) C(u_\beta - u_\alpha) + \mu(u) = C(u - u_\alpha), \alpha = 1,...., n$$

The function C is the covariance model. In SADA v5, this covariance model is derived from the more commonly used semi-variogram model covered in the previous chapter. In other words, this is where the connection is made between the correlation model and the weights assigned to each data point. For ordinary kriging, it turns out that the weights must sum to one.

$$\sum_{\beta=1}^{n} \lambda_\beta(u) = 1$$

So, the first step in ordinary kriging is to specify a reasonable semi-variogram model (provides C) and applicable search neighborhood (provides n). These two factors produce the weights (λ,), which in turn produce the ordinary kriging estimate Z*(u) at location u.

So what about the random variable Z(u)? How is this different than $Z^*(u)$? First, ordinary kriging provides an estimate for the error variance at each location called the *kriging variance*. In particular:

$$\sigma^2_{ok}(u) = C(0) - \sum_{\alpha=1}^{n(u)} \lambda_\alpha(u) C(u_\alpha - u) - \mu(u)$$

The term $\sigma^2_{ok}(u)$ can be used as a model of local variance. The specification of the distribution of values is accomplished by using Z*(u) as the mean and $\sigma^2_{ok}(u)$ as the variance of the local normal distribution of possible values at u. This approach is commonly used in practice and is the current implementation in SADA. This assumption of normality has implications regarding the data set as a whole, namely, the assumption that the data are normally distributed.

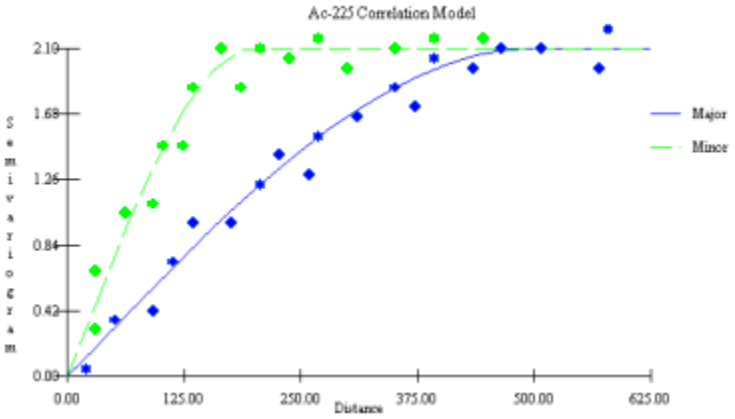
There is a great deal of discussion in the literature about the meaning and use of the kriging variance as a model of uncertainty. The fundamental problem with the kriging variance is that it is only a function of the distance between data values or, more accurately, their spatial configuration. The kriging variance does not depend on the actual data values themselves. Therefore, two estimates at different locations $Z(u_1)$ and $Z(u_2)$ may have exactly the same kriging variance $\sigma_{ok}(u_1) = \sigma_{ok}(u_2)$, even though data in the neighborhood u_1 may be much more variable than in u_2 (Goovaerts,1997).

Practical Recipe for Ordinary Kriging

The basic steps for producing an ordinary kriging map are as follows:

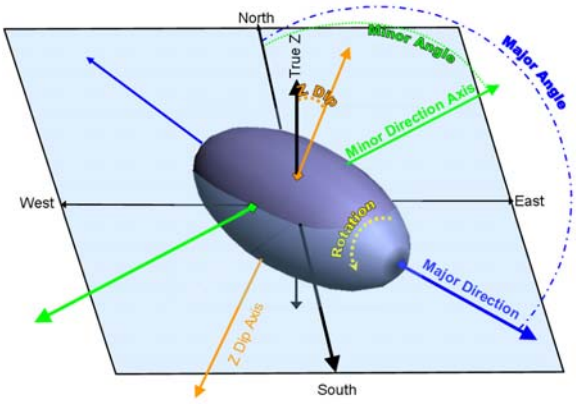
1) Develop a correlation model (previously discussed).

   This is largely the work of Chapter 30. When you select *Ordinary Kriging* in SADA, you will be restricted to evaluating the correlation model for the raw untransformed data. If you wish to apply a normal score transform, you will to do this outside of SADA in version 5 (Note SADA does do a normal score transform on the fly during sequential Gaussian simulation and this will be made available for ordinary kriging in a future version as well).

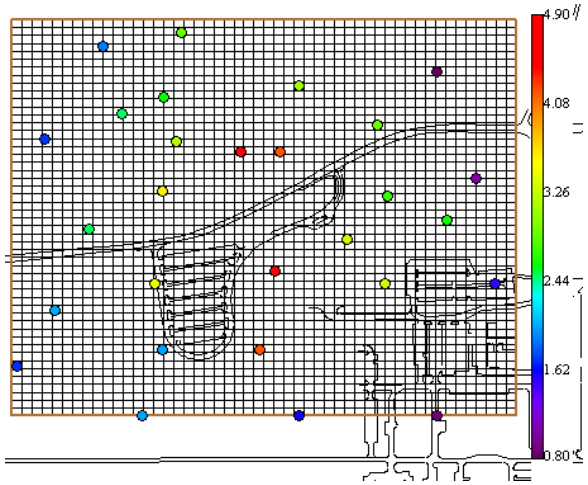


2) Define a search neighborhood

This is done in exactly the same way as with the inverse distance model (Chapter 29). Here, one can calibrate the search neighborhood based on the result of the correlation model. So for example, if the correlation range in the major direction at a 45 degree angle is 500ft and 300ft in the minor direction, this might well inform the choice for the search neighborhood. You should not absolutely adhere to this, however, as special circumstances may prevail. For example, you may have insufficient data in some areas of the site to restrict the search neighborhood. You may be struggling with the smearing effects of an outlier (see Chapter 29, "Basic Geospatial Methods"). These types of circumstances may warrant either a wider or smaller neighborhood, respectively.

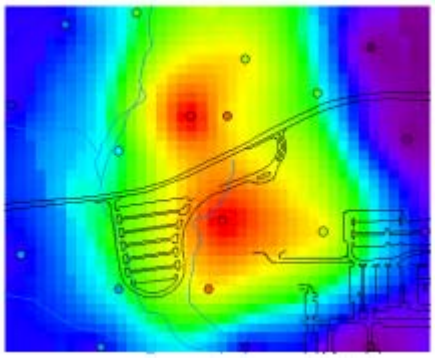


3) Create a grid across the site (previously discussed).



4) Apply kriging algorithm.

This is completely handled by SADA. Once the correlation model is available and a search neighborhood is specified, pressing the *Show The Results* button will produce the kriging map. You can produce two types of maps from the ccdf: mean and percentile. Under the assumption of local normality, the mean and the 50th percentile will be equal. See the section below on mapping percentiles as a means of understanding uncertainty.



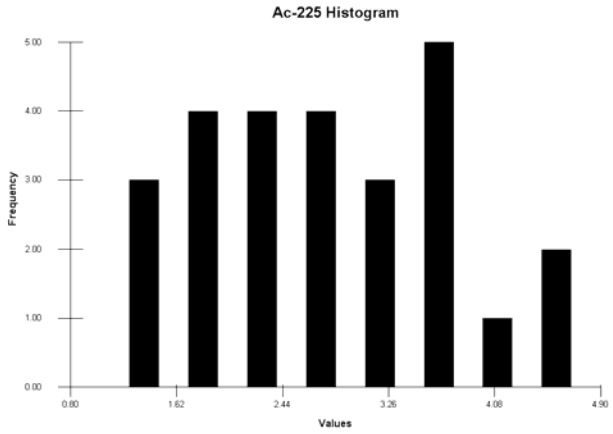
You can also view the ordinary kriging variance map. Simply select the interview *Variance map* and press the *Show The Results*.

Ordinary Kriging Example

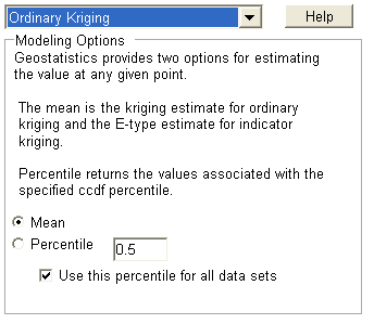
If you have PowerPoint, Word, or another Windows software package you normally use, open it now. Open Geospatial.sda, select *Ac-225*, and select *Interpolate my data* from the interview list. Let's get a copy of the data we are about to interpolate into PowerPoint (or other software). Press this button in SADA:

Switch to PowerPoint and select *Edit→Paste* or CTRL V. Switch back to SADA.

The first thing you will want to do is some basic data exploration. This includes looking at the histogram of Ac-225 values. For ordinary kriging, particularly when using the model of uncertainty, you'll need to check the normality of the data. *Choose Statistics→Show Histogram*.



This histogram does not look particularly normal. In practice, a better choice might be to use indicator kriging, which does not assume any distribution. Switch to PowerPoint and select *Edit→Paste* or CTRL V. Switch back to SADA. In the *Interpolation methods* step, select *Ordinary Kriging*.



With Ordinary Kriging selected, you will need to choose a value from the ccdf to plot in the map. You can select either the mean value or a percentile value. In this case, we'll select the *Mean* value. This is the same as the 50th percentile and is the traditional kriging estimate used.

Next, let's establish a correlation model for the Ac-225 data. If you need reinforcement, now would be a good time to revisit Chapter 30. If you are comfortable with these concepts, continue.

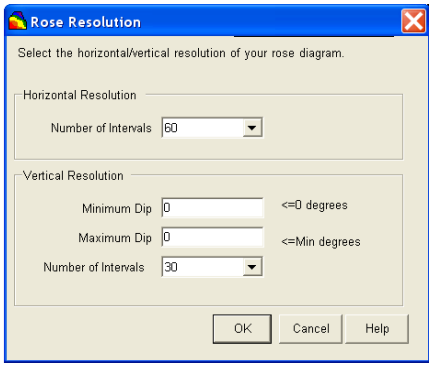
Click on the *Correlation modeling* step. In the *Explore Semi-variography* parameter block, let SADA fill in some default values by pressing the *Recommend* button.

A good way to begin is to look at the rose map. Visual inspection of the data indicates an N-S trend, and we want to measure and use this information in the modeling process. For the *Major* direction column, change the Angle Tolerance (*Tol*) to 50°.

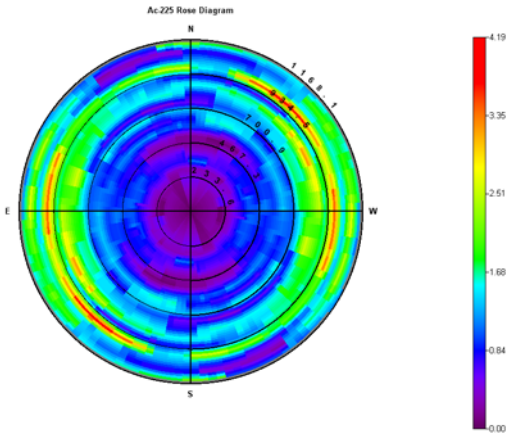
Name	Major	Minor
Caption	Default	Def...
Lag Number	20	20
Lag Distance	58	58
Lag Tol	58	58
Angle	0	0
Tol	45	90
Band	2302.75	230...
Dip	0	0
ZTol	90	90
ZBand	1	1

Recommend

Press the *Rose* button and accept the defaults in the resolution window that follows.

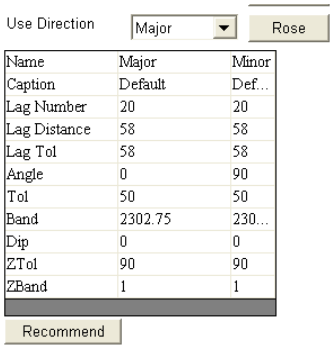


Press *OK*. The resulting rose plot of semi-variography values shows the trend we noticed running approximately N-S.

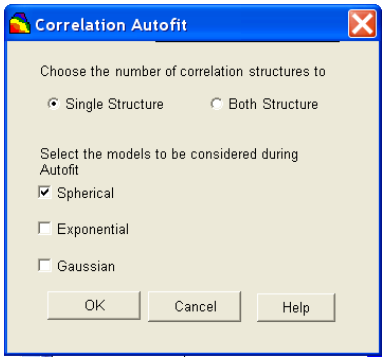


Let's copy this to PowerPoint now. Press the copy button again  . Switch to PowerPoint and paste the result in a new slide then return to SADA.

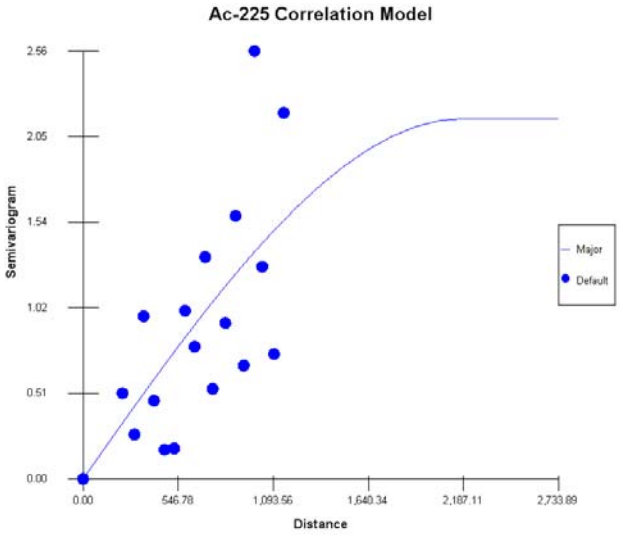
Let's take a look at particular semi-variograms in the N-S and E-W directions. In the *Explore Experimental Variography* block, change the *Angle* in the *Minor* direction column to 90° and the angle tolerance (*Tol*) to 50°. Leave the *Use Direction* as *Major*. We'll model the *Major* direction first.



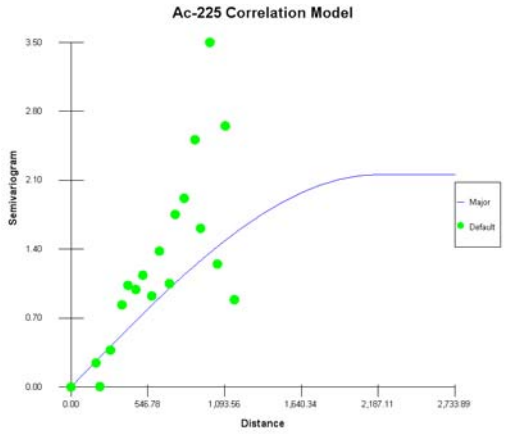
In the *Model Semi-variography Value* block, we'll start by allowing SADA to attempt to fit the data using a least squares approach. Press the *Recommend* button and select *Single Structure* and *Spherical* in the window that appears.



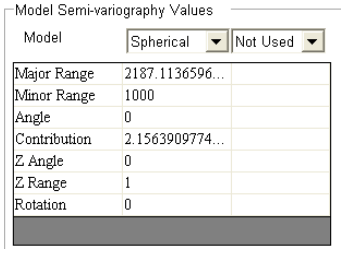
Press *OK*. The resulting model isn't too bad. Let's see how well we can get it to also model the minor direction.



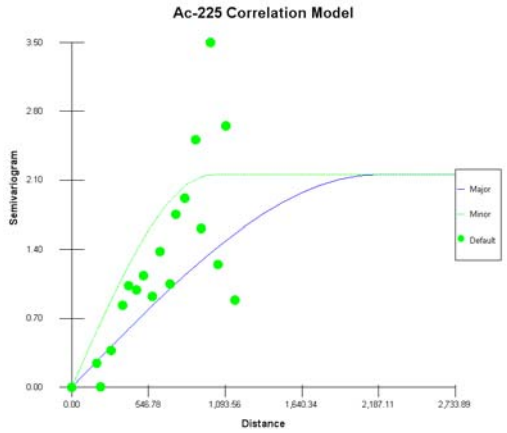
In the *Explore Experimental Variography* parameter block, select *Minor* for the direction. Press the *Show Me* button in the lower-left-hand corner of the parameter window. You should see the following result:



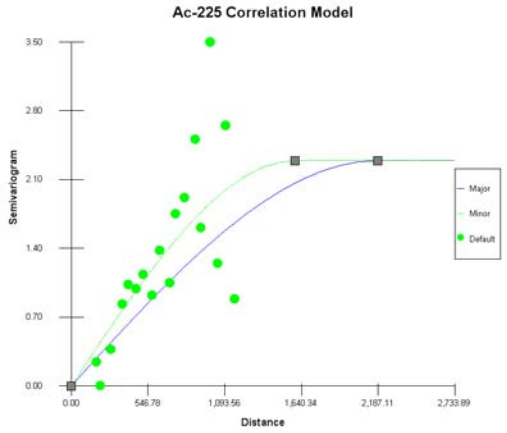
The only flexibility in modeling both the major and minor direction simultaneously is in the choice for the *Minor Range* and *Major Range*. Let's change the *Minor Range* to 1000 now.



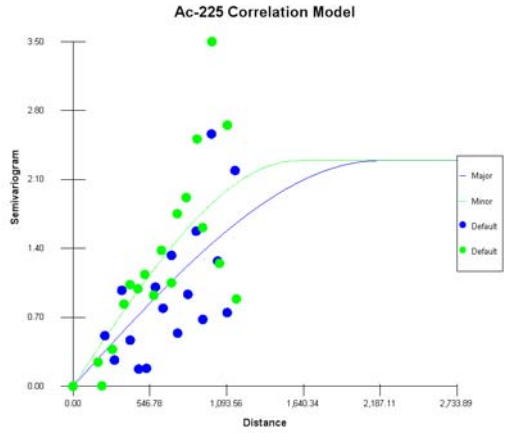
Press *Show Me* in the lower-left-hand corner of the parameter window. This will separate the major and minor line. Now we can graphically edit them.



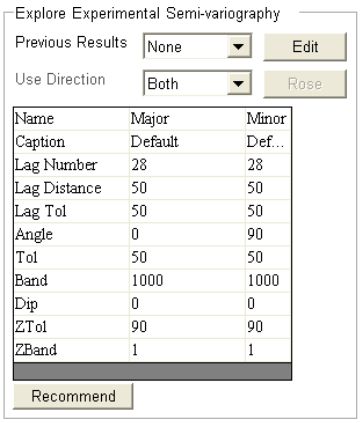
Press the *Graphical Edit* button. Using the edit boxes, edit the correlation model so that it better accommodates the minor direction. When you are done, press the *Graphical Edit* button again to get back out of edit mode.

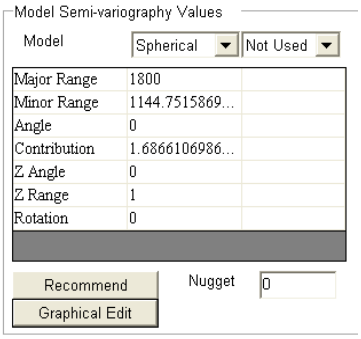


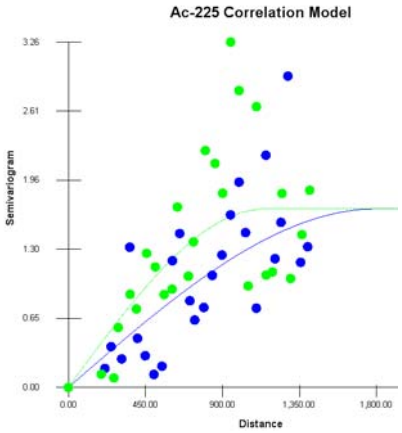
Now turn both directions back on and see how reasonable the fit is.



In this example, we are breezing through the assessment very fast. In practice, you would want to experiment with the experimental variography lag parameters, model parameters, and so forth to get a better handle on the correlation structure. To speed things along, below are some parameters that deliver a clearer picture of correlation structure after working and exploring the data, rather than relying on automated procedures.

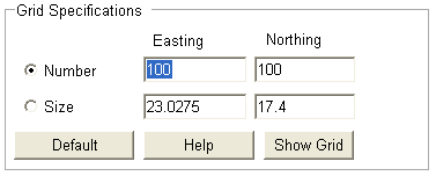






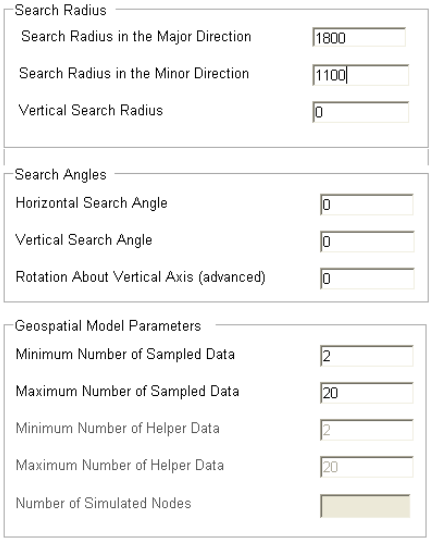
Make sure your parameters match these and press *Show Me*. Copy this correlation model into your PowerPoint presentation. Save your SADA file now.

Let's set up the grid specs next. Click on the *Grid Specs* step and choose *Number*. Enter a 100x100 grid.



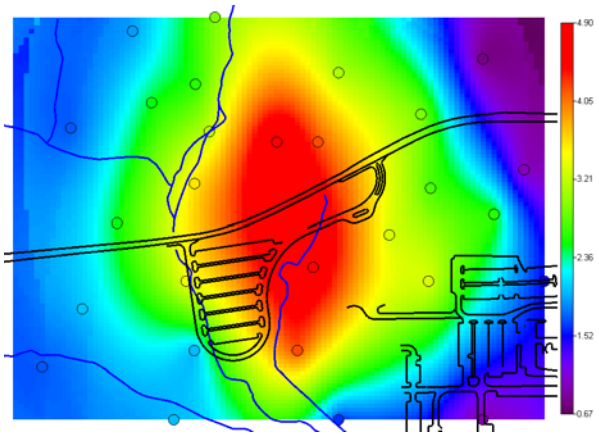
Next, we'll specify the search neighborhood. If you feel uncertain about how to define a search neighborhood, please review the search neighborhood discussion in Chapter 29. As a guide,

we'll use the same major and minor search radius as our correlation model indicated (1800 and 1100, respectively). Click on the *Search neighborhood* step and enter the following parameters.



Note that the horizontal search angle of zero corresponds to an N-S direction, so we don't need to change it. Since this is 2d interpolation, we also don't need to consider the 3d parameters.

Now we're ready to apply the kriging algorithm. In the steps window, you can press the *Show the Results* button or click on the *Show the results* step. In this case, let's click on the step. SADA will show a progress bar and produce the results.

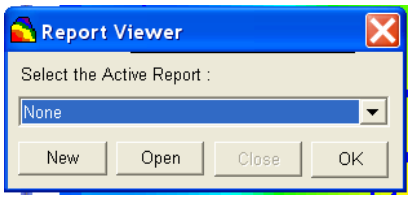


Copy this picture into your PowerPoint slide show now. Notice in the parameter window of the *Show the results* step that SADA has completely documented the parameters associated with this kriging result (you may have to refresh the log).

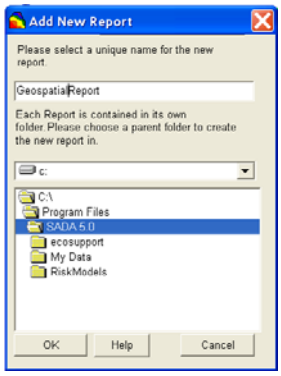
If we want to add this to a report, we'll need to use the auto-documentation feature in the *Autodocumentation* step. Click on this step now. In this step, you can add your documentation to one or more reports. A report is little more than an HTML document that can be opened by Word, web browsers, and others. It is a good way to quickly add documentation to a report you may be working on.

In the parameter window, you'll see that we don't have any reports set. That's OK; SADA will lead us through this. You'll also see a list of the kinds of things you can document about the result. Check all of them now and press the *Apply* button.

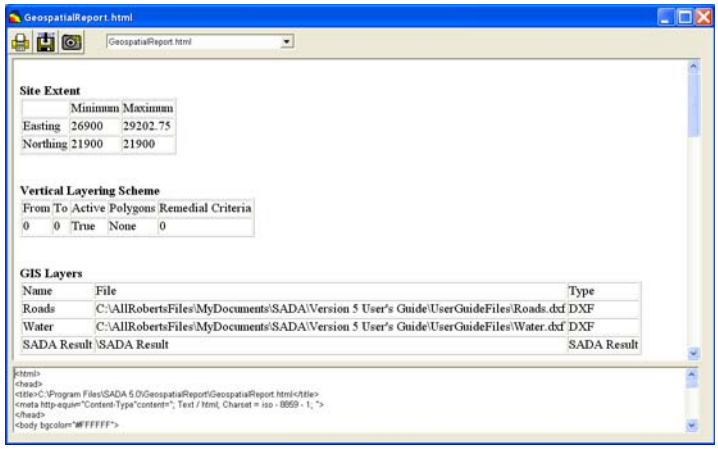
SADA warns that you don't have a report open. Answer *Yes* that you would like to create one now. In the *Report Viewer* window that appears, press the *New* button.



The *Add New Report* manager window appears. Navigate to a location where you would like the report to reside and enter the name "GeospatialReport" at the top.



Press *OK*. Press *OK* again and SADA indicates the report was updated. To see your report, from the main menu select *Reports→Bring Report Viewer To Front*. The report viewer is displayed.



At the top of the viewer, you may print, save, refresh, and change reports. In the middle, you see how the report will look, and at the bottom you can edit the report. Editing the report requires some HTML knowledge. A better way is to open this report up in Word when you are finished and save it as a Word document file. Then you can edit from there as you like. Close the window by pressing the *X* button (not the SADA X button). Open up Word and try opening this html file there. Note that you may need to change the *Files of type* in Word's Open dialogue box to *all Files* in order to see the file, as normally only .doc files are seen.

You may also want to store this model and apply transformations or simple map algebra functions and comparisons to it later. Click on the *Manage model results* step. In the parameter window, click on the *Save Model Result* button. Enter "Ac-225 Ordinary kriging" into the name window that appears and press *OK*. You can find your model in a couple of locations. First,

switch back to the *Interpolate my data* interview and click on the *Interpolation methods* step. You will have *Ordinary kriging* selected. Click on the drop-down list and notice that you also have a *Use stored result* option. When this is selected, you will see all the stored and/or imported in a drop-down list below.

The first one in the list is *Ac-225 Ordinary kriging*. The *FieldDetection* item is a model we imported from outside of SADA. So, if you have another contouring package you prefer to SADA, you can use that, import the result, and use it instead. When you select *Use stored result*, notice also that the steps list has changed; some steps such as the *Correlation model* and *Search neighborhood* are no longer required. Press the *Show The Results* button. You can use this stored result in decision models, risk models, sample designs and so forth.

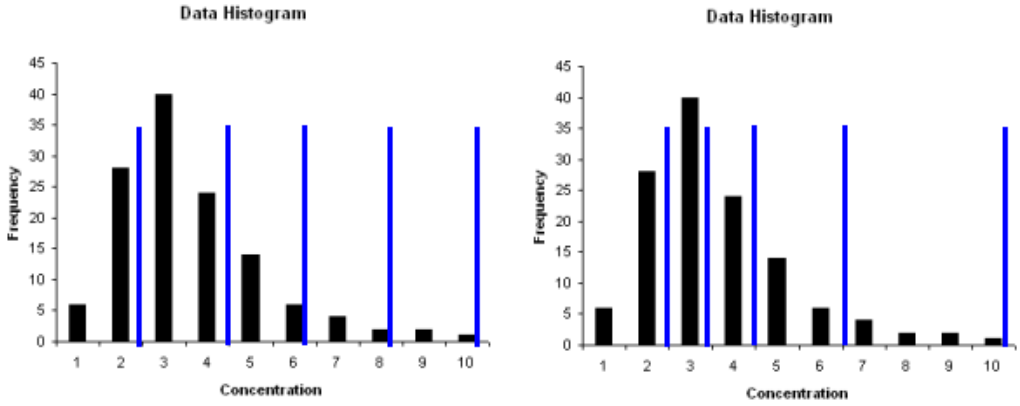
You can format this result. For more information on this topic, refer to the *Help* file. You can also export the result in a variety of formats with the *Export to file* step. Refer to the *Help* file for more details. Save your SADA file.

Indicator kriging

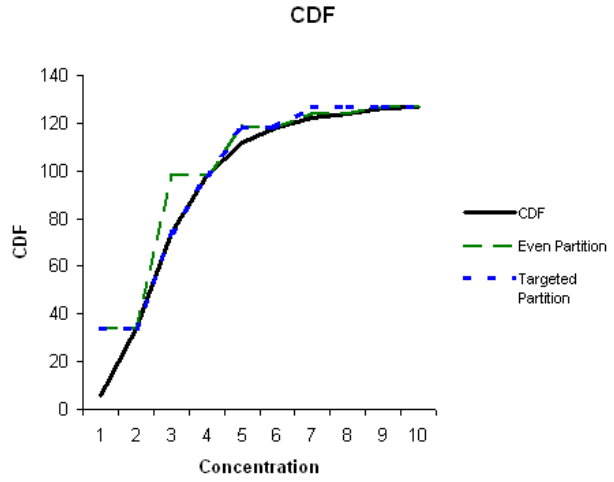
Indicator kriging (IK) provides a non-parametric alternative to ordinary kriging. Rather than assume a normal distribution at each Z(u) and use the estimate $Z^*(u)$ and σ^2_{ok} as the mean and variance of the normal distribution, IK builds the conditional cumulative distribution function (ccdf) at each point for each cutoff z_k:

$$[i^*(u; z_k) = prob^*\{Z(u) \leq z_k \mid n\}$$

where (n) represents data in the search neighborhood of location u. This value is calculated as a series of cutoff values z_k, k = 1,….,K. In practice, the user will observe the range of data values, say [0, 23.2], and discretize this range into K partitions called indicator thresholds. The objective here is to approximate the cdf of the data in your selection of k partition values. The user may make even partitions of the data or uneven partitions to better reflect the trend in the cdf curve. For example, consider the following histograms. Each histogram has been partitioned into 4 sections. In the first image, the partitions are even. In the second image, there are more partitions in the area of greatest change in the histogram curve.



Comparing the partitions against the CDF of the data, we see that the targeted partition better approximates the cdf of the data set.



One important rule is that the CDF value (or number of data less than a threshold value) must increase for each higher threshold value. Suppose that we select four threshold values of 1mg/kg, 2mg/kg, 3mg/kg, and 4mg/kg. In the following table using an example data set (not shown here), we'll determine the number of data that are less than each threshold value.

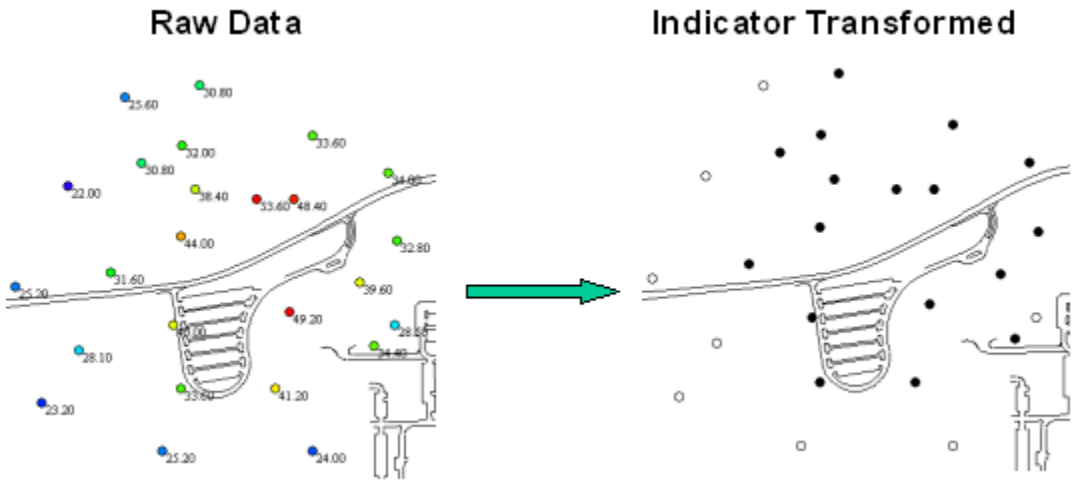
Threshold Value	Number Of Data Less Than Threshold
1	3
2	**5**
3	**5**
4	7

This set of threshold values is not valid because for both threshold values of 2mg/kg and 3mg/kg, the number of data below each is still 5. The rule says that there must be more than 5 for the 3mg/kg threshold. If not, then another value needs to be selected.

For continuous applications, one must transform the data values for each threshold z_k (e.g., 1mg/kg, 2mg/kg etc) according to:

$$i(u;z_k)=\begin{cases}1\ if\ \ z(u)\le z_k\\ 0\ if\ z(u)>z_k\end{cases}$$

Here we see a graphical plot of this type of transform for a specific threshold z_k for each data point u_i.

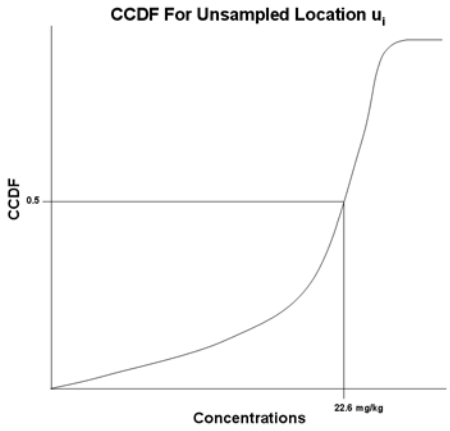


SADA performs transform behind the scenes for the user. The next step is to develop semi-variogram models that describe the spatial behavior of the transformed data at each threshold z_k.

SADA then calculates for each node the *conditional cumulative distribution function* (ccdf) at each threshold value z_k. The term *conditional* refers to the fact that the distribution was conditioned by nearby data. So, we have at every node a ccdf. It is not easy to visualize a map of ccdfs. Furthermore, the objective is often to estimate a concentration value and/or quantify the uncertainty about the unknown value. There are two ways to handle this:

Percentile

You can specify a percentile of the ccdf to use as the concentration estimate, for example the 50th percentile.



E-Type Estimate

For contouring maps, SADA calculates the e-type estimate as the estimate at the unsampled location. The e-type estimate is defined as (Deutsch and Journel, 1992):

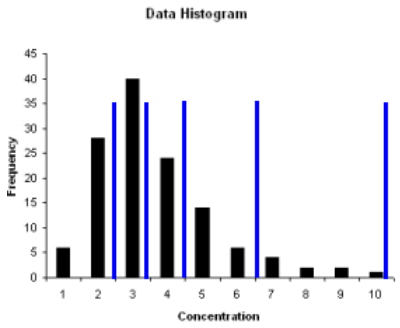
$$[z(u)]_E^* = \int_{-\infty}^{+\infty} z\, dF(u; z \mid n)) \approx \sum_{k=1}^{K+1} z_k' [F(u; z_k \mid (n)) - F(u; z_{k-1} \mid (n))$$

where z_k, k = 1, ….,K are the K cutoffs, and z_0 = zmin, z_{K+1} = zmax, are the minimum and maximum of the z range, entered as input parameters by user. Z' is the interpolation procedure. SADA V5 uses simple linear interpolation.

Practical Recipe for Indicator Kriging

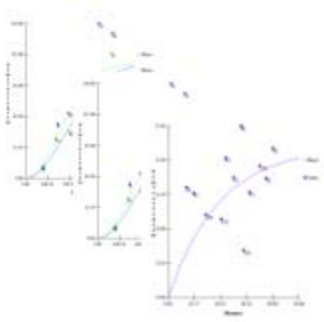
1. Determine set of partition values.

   Choose the threshold values so that you well represent the features of your distribution. More threshold values should be located in areas of the distribution where the function is changing more rapidly. Also, consider using the decision threshold (if one is available) as one of the indicator thresholds. This will increase the accuracy of your model at that range.



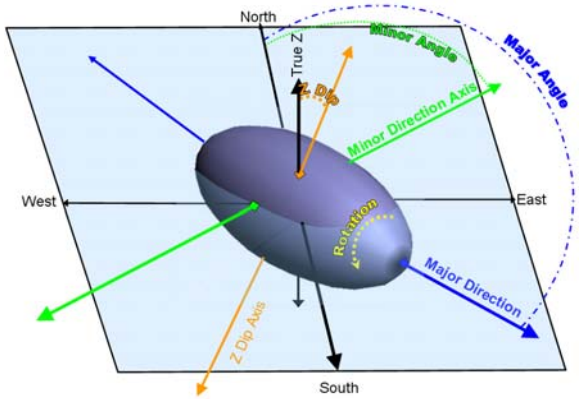
2. Develop semi-variogram models for each partition value.

   Be sure to review the section in Chapter 30 that deals with how to set up correlation models for indicator kriging/simulation. A correlation model must be available for the indicator transformed data for every indicator threshold you specify.

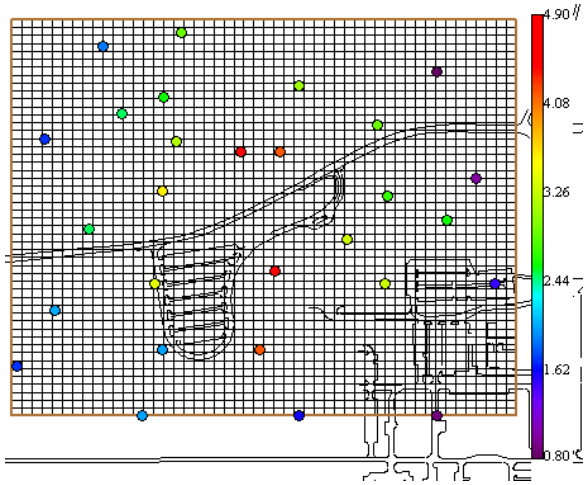


3. Define reasonable search neighborhood.

   This is done in exactly the same way as with the inverse distance model (Chapter 29). Here, one can calibrate the search neighborhood based on the result of the correlation model. Unlike ordinary kriging, you have multiple correlation models (one for each indicator threshold). So, you may have several ranges to select from. Begin by trying the largest range. If you have anisotropic conditions and with varying degrees of anisotropy for each cutoff, consider using an isotropic variogram or perhaps choose the correlation structure ranges near or at your decision criteria. You should not absolutely adhere to this approach, however, as special circumstances may prevail. For example, you may have insufficient data in some areas of the site to restrict the search neighborhood. These types of circumstances may warrant either a wider or smaller neighborhood, respectively.

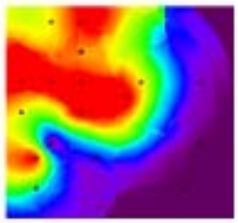


4. Create a grid across the site (previously discussed).



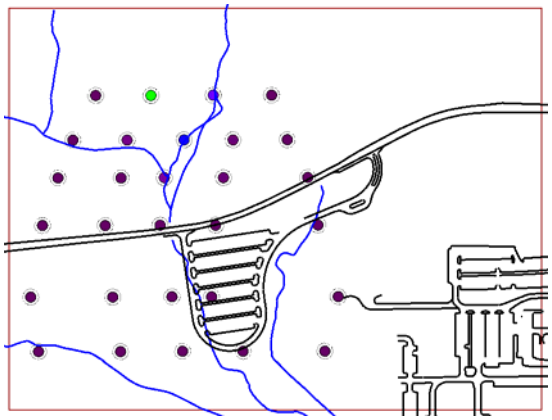
5. Apply kriging algorithm.

   This is completely handled by SADA. Once the correlation models are available and a search neighborhood is specified, pressing the *Show The Results* button will produce the indicator kriging map. Depending on your choice, either the E-Type estimate or a specified percentile based value will be produced.

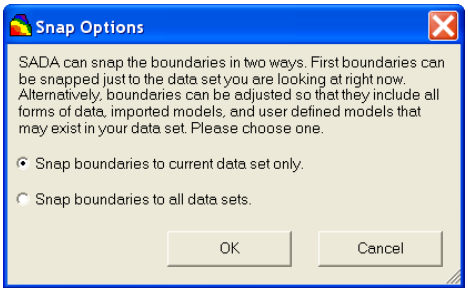


Indicator Kriging Example

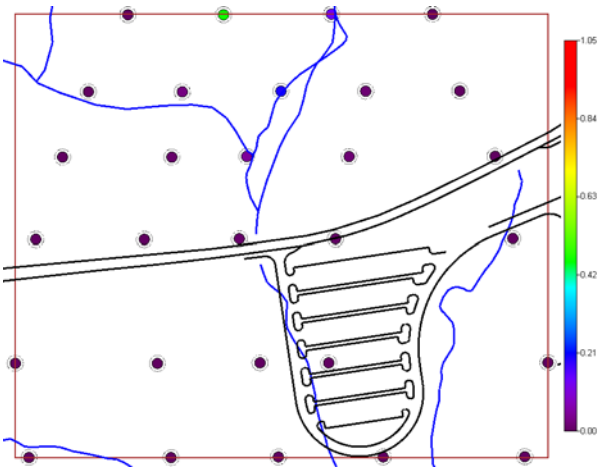
Open up the SADA file Geospatial.sda. Switch to *Groundwater*. SADA will likely give you a warning about data falling outside the currently selected vertical layer design. This is because we were looking at a 2d data set (Ac-225) in a previous assessment and the first data set under groundwater is a 3d Chlordane data set with samples deeper than 50ft. Say *OK* to the warning. Click on the *Setup the site* step. In the *Set Vertical Layers* parameter block, switch to 10 Foot Layers. You should see more data in the 2d viewer now, as the first layer is 0-10ft. Now in that same parameter window, notice that we are hiding the site extent boundary. Uncheck the *Hide Site Boundary Option* and notice that the boundary extends to far north, east, and south.



Let's snap the boundaries to our current data set. Press the *Snap* button in the *Site Boundary* parameter block. SADA will ask if you want to snap to the current data set or to all data sets. Choose the *Snap boundaries to current data set only* option and press *OK*.



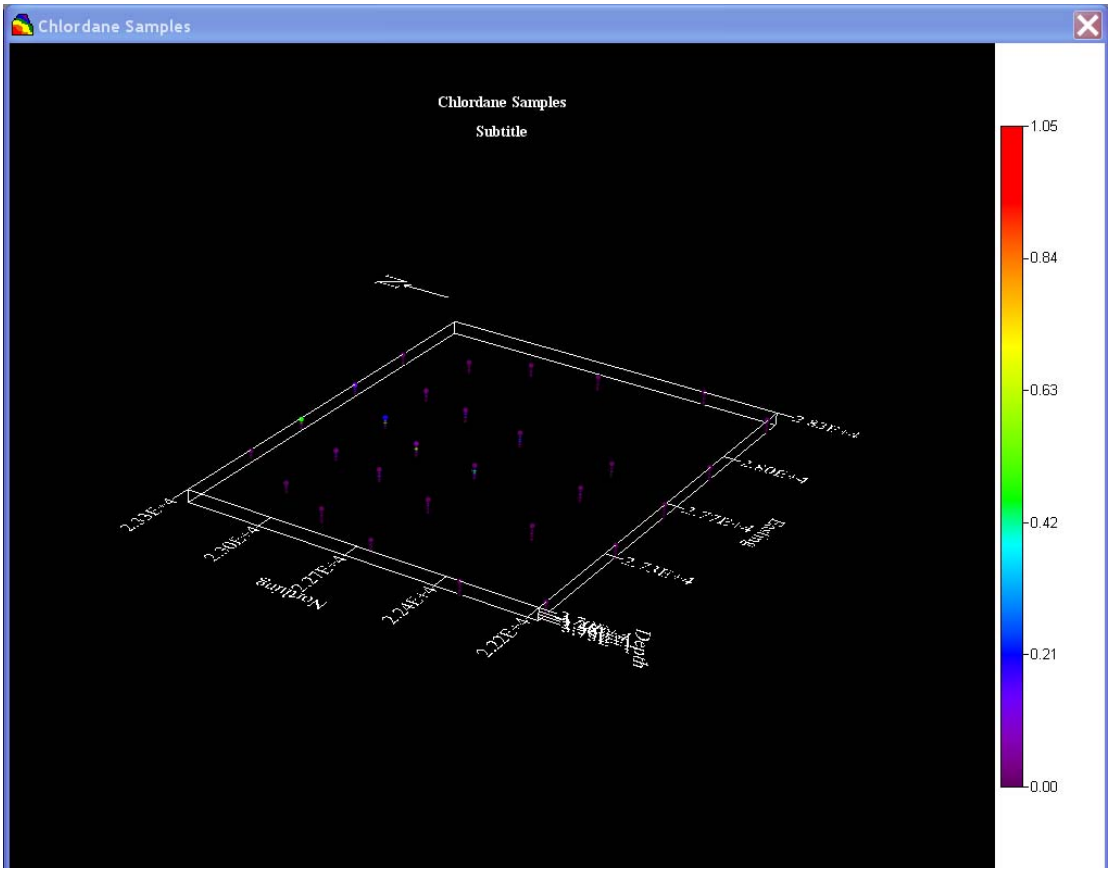
SADA respond by placing a tighter boundary around the data set. For more options on setting boundaries, see Chapter 4.



Save your SADA file.

Let's begin with some data exploration. Using the data level drop-down list, explore data by the layer. Notice that higher concentrations are found in the north, near the center of the site. This band of higher values extends south but drops in depth at the same time.

To see this more clearly, turn on the 3d viewer. It is well worth noting that 3d viewers take longer to render than 2d viewers, and SADA is no different. We recommend that for most of your analysis you leave the 3d viewer turned off. In addition, many tools are not available when the 3d viewer is on. The 2d viewer should be your primary base of operation. To turn the 3d viewer on, choose *Graphics→Show 3d View*.

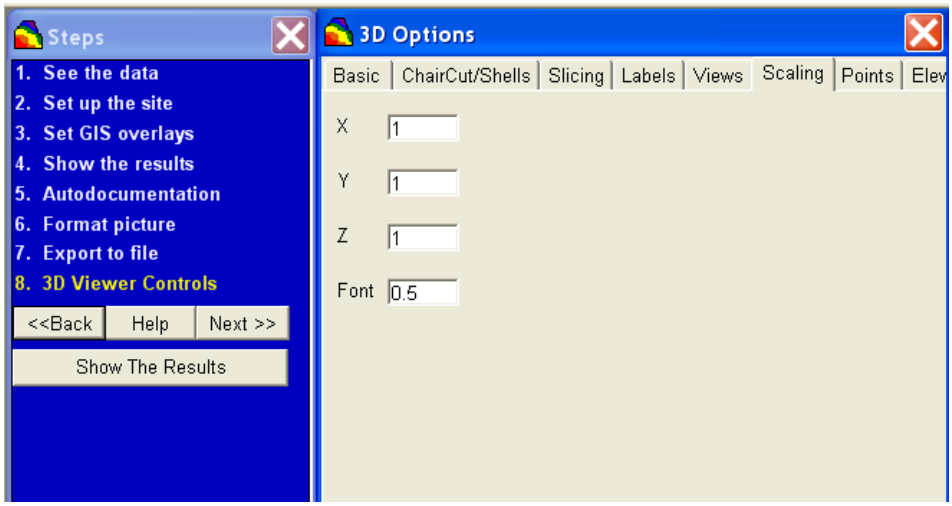


IMPORTANT TIP!!!

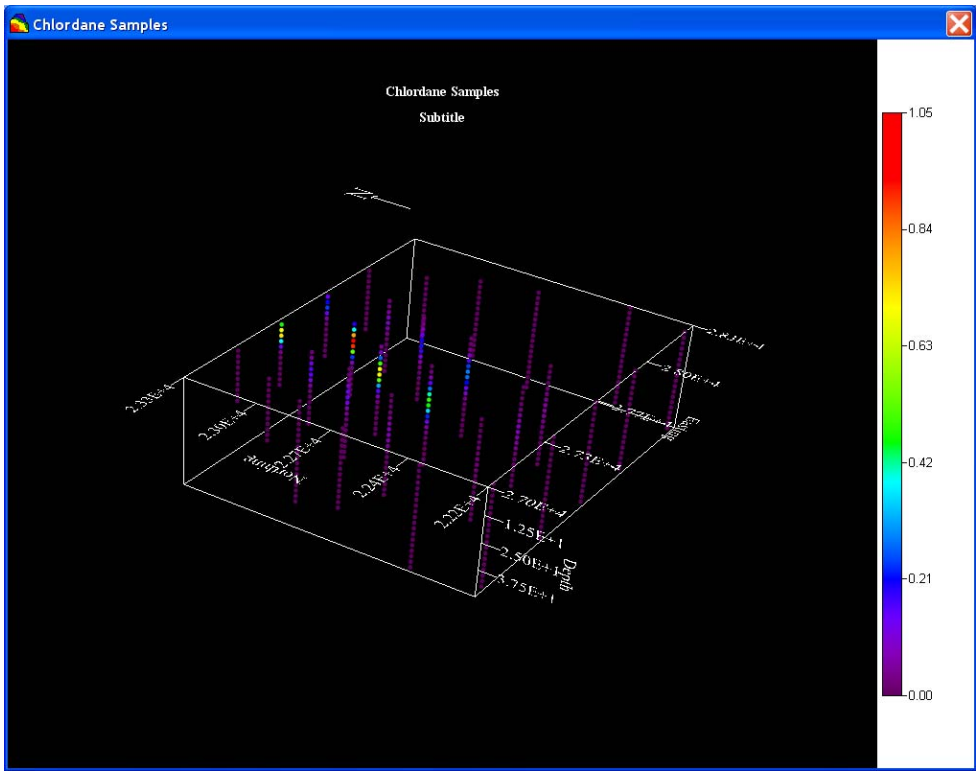
Depending on the video card you have, the 3d viewer may act temperamental. If you don't see this view or see "remnants" of the last window you had open in its place, just click anywhere in the 3d viewer to refresh it.

Most of the time for environmental data, the vertical extent is considerably smaller than the horizontal extent. For this reason, when you view the results in 3d, your site looks "pancaked" or mashed flat. So, the first thing you'll want to do is exaggerate the z direction.

Click on the step *3d Viewer Controls* (this is a new step that popped up when you turned the 3d viewer on). This parameter window has quite a few tabs and can be intimidating at first. Most of the time, you'll use only a few of the features. Scroll to the right in the window and click on the *Scaling* tab. You should have the following view.



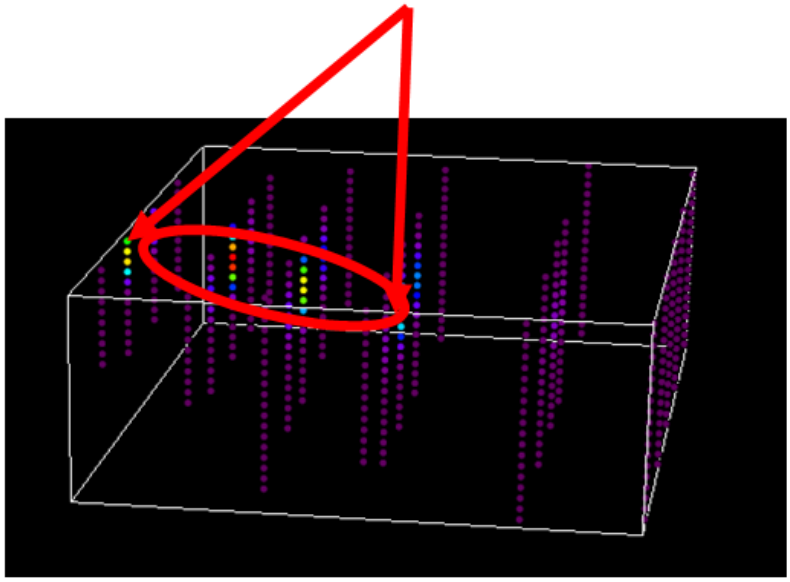
In the *Z* parameter box, enter a value of 10 and press the Enter key on your keyboard. You should see the following change in your 3d view. If you don't see a change, just mouse-click anywhere in the 3d view:



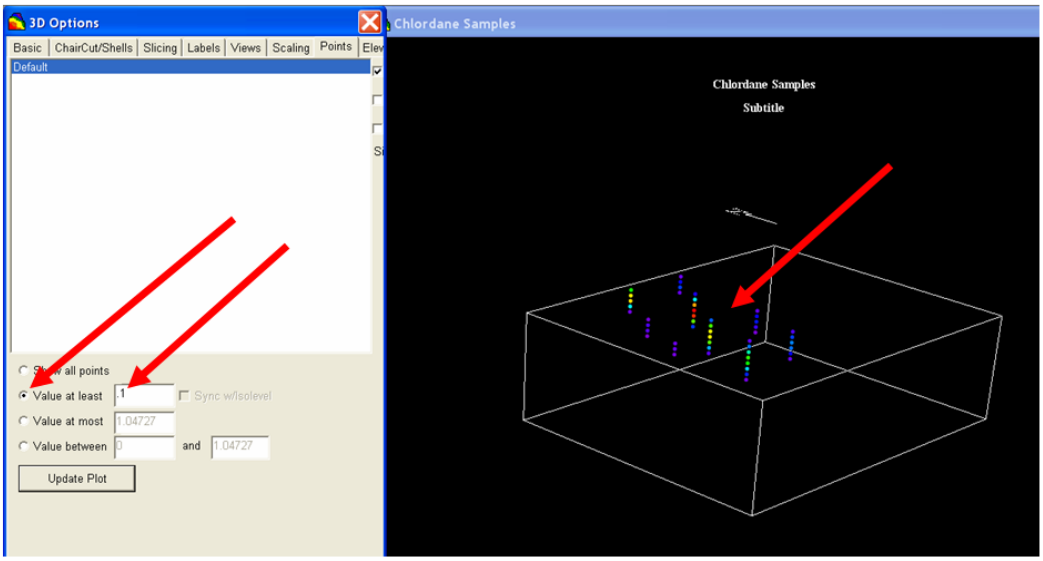
Rotating the view is easy. Just left mouse click and hold anywhere in the 3d viewer area. Now drag your mouse and you'll see the image begin rotating. The best way to learn how to rotate is to just play around with moving your mouse in different directions.

If the axis labels are annoying, you can get rid of them by right-mouse-clicking on the axis they belong too. They turn off and on accordingly.

You can now see the location of the elevated zone in the subsurface very easily by rotating the 3d results into different positions.



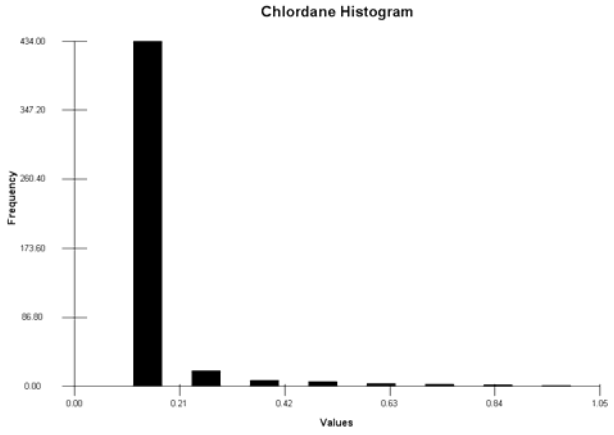
You can now turn off all the lower value points so you see your results better. To the right of the Scaling tab is the Points tab. Click on this tab. Select the *Value at least* option and enter a value of 0.1. This will eliminate any measured values below a value of 0.1 mg/L. (Recall that SADA doesn't track units, these are selected by the user. The exception is when doing risk assessment.)



Now you can clearly see the elevated values in the 3d viewer. Unfortunately, we don't have a GIS layer at this point for reference. If you did, the 3d viewer would draw it on the top of the 3d cube. Let's turn the 3d viewer off by choosing *Graphics→Show 3d Viewer*. The menu item acts like a toggle and will turn the 3d viewer off and on.

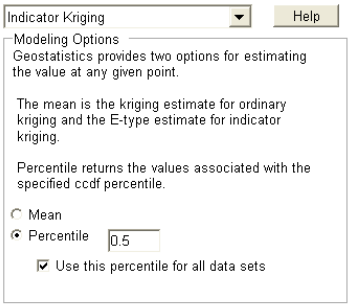
You can also look at the vertical profiles. Please refer to Chapter 11 for this feature.

Now we'll take a look at the data histogram. *Select Statistics→Show Histogram*.

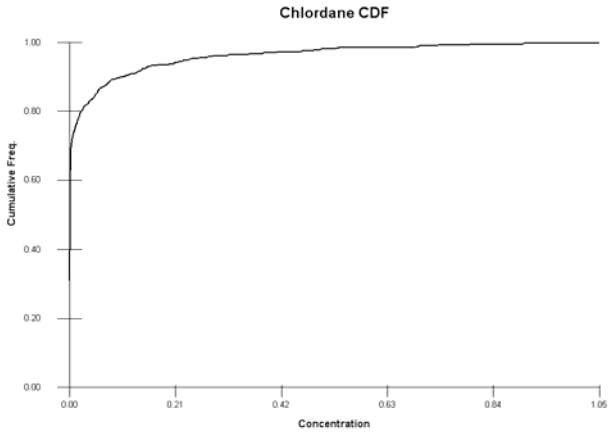


Here, we have a highly skewed distribution often encountered in environmental data. Data transformations such as the normal score transform or lognormal transform may be helpful. Indicator kriging is well suited to deal with this kind of data. Return to and deselect *Statistics→Show Histogram*.

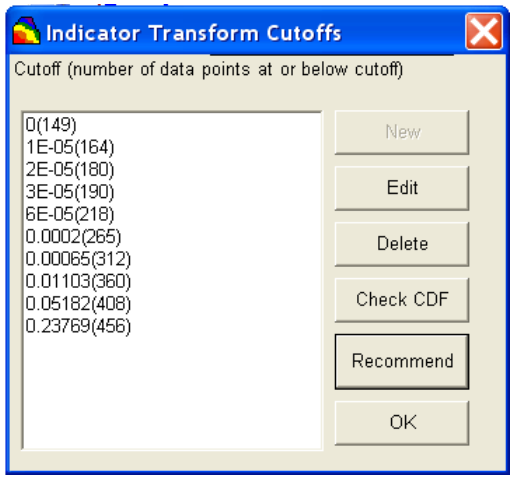
Choose the interview *Interpolate my data*. Click on the *Interpolation methods* step and choose *Indicator Kriging* from the list of available interpolants. Let's choose *Percentile* 0.5 instead of the E-type estimate (*Mean*).



Now we'll need to specify some indicator thresholds that will well partition the cumulative distribution function. Given the skew of the data, one can anticipate a larger number of threshold values in the lower range of data. Most of the data appears to be less than .1 mg/kg.



Click on the *Correlation modeling* step. In the *Data Transform* parameter block, press the *Edit* button next to the *Data Transforms* drop-down list. In the indicator threshold manager, press the *Recommend* button. The following selections are made.



Notice that there are many zero values, 149 to be specific. It is not really possible for SADA to divide into equal deciles due to this fact. SADA provides roughly equal percentile intervals for those values greater than zero. Press the *OK* button.

We will now need to specify a correlation model for each of the indicator transforms. In the *Correlation modeling* parameter window, make sure that the IC=0 transform is selected. In the

interest of time, we'll simply accept the recommendations of SADA for both experimental variography parameters and modeling parameters (for spherical). Switch now to the second indicator threshold and repeat the process. Repeat this for every indicator threshold value. Save your SADA file.

Now, of course, a more in-depth analysis of correlation structure should be performed. In particular, you would want to capture the way that the high values dip in an N-S transect. This would be accomplished by examining the anisotropic structure in the upper thresholds values. Press *Show The Results*. You can peruse the results by switching between layers. The smearing in the second layer is particularly noticeable. This is simply the first cut at modeling the data and emphasizes the danger in simply accepting default values for every parameter. At this point, you should return to the correlation modeling structure to improve the experimental variography and modeling alike.

You may also want to store this model and apply transformations or simple map algebra functions and comparisons to it later. Click on the *Manage model results* step. In the parameter window, click on the *Save Model Result* button. Enter "Chlordane Indicator kriging" into the name window that appears and press *OK*. You can find your model in a couple of locations. First, switch back to the *Interpolate my data* interview and click on the *Interpolation methods* step. You will have *Indicator kriging* selected. Click on the drop-down list and notice that you also have a *Use stored result* option. When this is selected, you will see all the stored and/or imported data in a drop-down list below.

The first one in the list is *Chlordane Indicator kriging*. The *FieldDetection* item is a model we imported from outside of SADA. So if you have another contouring package you prefer to SADA, you can use that, import the result, and use it instead. When you select *Use stored result*, notice also that the steps list has changed; steps such as *Correlation model* and *Search neighborhood* are no longer required. Press the *Show The Results* button. You can use this stored result in decision models, risk models, sample designs and so forth.

Cokriging

In some cases, secondary forms of data may shed some light on the behavior of contamination on your site. These secondary types of data are not necessarily direct measurements of your contaminant but are field measurements of conditions relevant to your contaminant. For example, on a site contaminated with radionuclides, a walkover survey records gamma counts at a large number of locations or walking transect across the site. For a lead contaminated site, you may have a large number of XRF data that provide insight into the presence of lead. Typically, these secondary forms of data are cheaper to collect and therefore may be more abundant than lab qa/qc measured values.

From a modeling standpoint, a dense sampling of secondary data may more accurately recreate the heterogeneity seen in real environmental scenarios (see for example Goovaerts, 2000).

From a decision standpoint, it may be useful to use this information to support spatial estimation or even reduce uncertainty in key areas of the site. The secondary data must have some measurable relationship with the primary data, and if so, then cokriging methods can explicitly make use of this data in producing contour maps and assessing spatial uncertainty.

In cokriging, we simply add the secondary variables to the kriging equation. In the case of a single secondary variable W, the ordinary cokriging estimator of Z(u) is written (taken from Deutsch and Journel, 1992, p.69):

$$Z^*_{COK}(u) = \sum_{\alpha_1=1}^{n_1} \lambda_{\alpha_1}(u) Z(u_{\alpha_1}) + \sum_{\alpha_2=1}^{n_2} \lambda_{\alpha_2}(u) W(u_{\alpha_2})$$

So, the $\lambda_{\alpha 1}$'s are the weights assigned to nearby primary data $Z(_{\alpha 1})$ (e.g., lab measured pCi/g values), just as in ordinary kriging. The $\lambda_{\alpha 2}$'s are the weights assigned to nearby secondary data $W(u_{\alpha 2})$ (e.g., gamma count results).

Cokriging requires a correlation model for Z, a correlation model for W, and a cross correlation model between Z and W. If you have more than one type of secondary data, things can really get out of hand fast. Suppose you had another secondary data called M. Then, you would need a correlation model for M and cross correlation models between Z and M and between W and M. In practice, this can make the process tedious at best and intimidating at worst. Fortunately, there are some reasonable assumptions that can make this requirement a little easier. For example, one can assume that a correlation model for Z and for W must be similar and only different by some scaling factor. If they were very different, then W probably wouldn't be very useful in predicting Z, because they clearly would have different spatial distributions. This is a good time to return to the section "Correlation Modeling for Cokriging" in Chapter 30, which is dedicated to correlation modeling and covers the options for cokriging.

There are actually three forms of cokriging: ordinary cokriging, standardized cokriging, and simple cokriging (Deutsch and Journel, 1992, p.70). SADA uses ordinary cokriging. In this formulation, the additional constraints are placed on the weights.

$$\sum_{\alpha_1} \lambda_{\alpha_1} = 1 \text{ and } \sum_{\alpha_2} \lambda_{\alpha_2} = 0$$

The next version of SADA may contain standardized cokriging. You are encouraged to return to the website to get updates to this user guide as new versions are released.

One final topic is data transformations. Sometimes, the magnitude difference between primary data (e.g., pCi/g) and secondary data (e.g., count) can be substantial enough that the solution becomes affected. You can tell this kind of problem is occurring when you get estimates that are far above (or below) the range of your primary data values. In that case, you may consider a data transform. A common approach is to transform the data so that each has a mean of zero and a variance of one. This puts each data set on the same scale. Once the estimation is complete, the results are back transformed into real measurement space. In the example we'll do in a moment, you'll see where to select this option.

Practical Recipe for Cokriging

The following steps provide some initial guidance on how to perform cokriging.

1. Identify the secondary data set you wish to use.
2. Assess how well correlated the collocated measurements are. This actually refers to the correlation strength between the primary and secondary data when measured at the same location. This is the normal type of correlation assessment, and no spatial correlation is considered at the moment. A common measure of strength is the correlation coefficient or a simple scatter plot of collocated primary and secondary data.

   There are three measures of correlation strength in SADA: Spearman Rank, Pearson R, and traditional scatter plot.

Spearman Rank

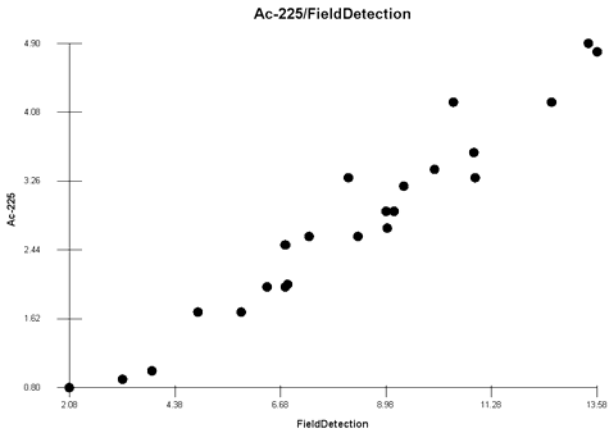
The spearman rank correlation coefficient is a non-parametric measure of correlation between two variables without making any assumptions about the frequency distribution of the variables. Furthermore, it does not require the assumption that the relationship between the variables is linear. The statistic ρ will always be -1<=ρ<=1, where a value of 1 is high positive correlation, -1 is high negative correlation, and 0 is no correlation. Statistical tests for significance are available but are not currently calculated in SADA.

Pearson R

Although similar statistically to the Spearman Rank, the data are assumed to be normally distributed and their scatter plot yields a roughly linear relationship. A negative or positive correlation indicates that the helper data may be of some use. Now in some cases, your primary data set may be too sparse to make any real use of this type of helper assessment. In others though, there may be a physical reason why one data set indicates the value of another. Gamma walk over surveys and surface contamination by radiological substances are practical examples. You should not be necessarily discouraged if there is not a one to one correspondence (i.e., perfectly clear linear relationship).

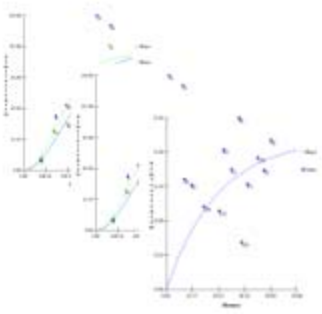
Scatterplot

In a traditional scatterplot, collocated values (measured at same location) of primary and secondary variables are created. For cokriging, you hope to see a roughly linear relationship between the two.



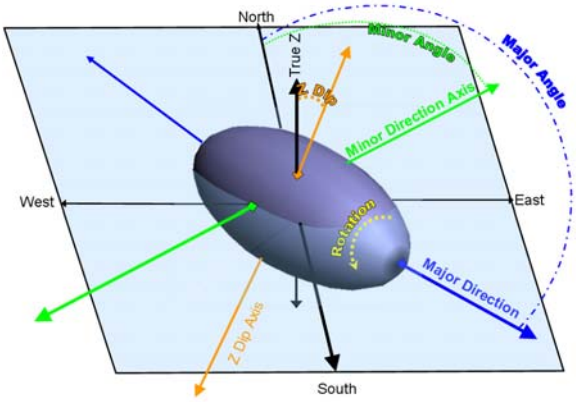
In practice, data may not be exactly collocated. Depending on the physical circumstances, you may need to relax exact collocation by specifying a small radius that, when centered on primary data values, may include a secondary value.

3. Select a correlation model scheme (see Chapter 30):
   a. Linear model of coregionalization (hard to implement)
   b. Intrinsic Model of coregionalization (easy to implement)
   c. Markov Model Of coregionalization (very easy to implement)
4. Depending on the correlation model scheme, measure and specify spatial correlation models for any required combinations of primary and secondary data.

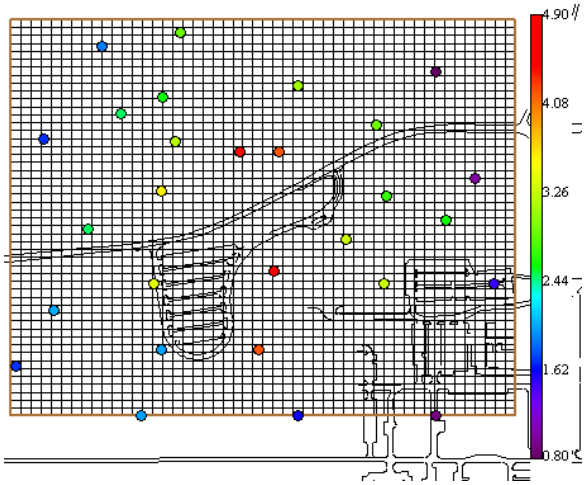


5. Define a search neighborhood

   This is done in exactly the same way as with the ordinary kriging model. Here, one calibrates the search neighborhood based on the result of the correlation model evaluation. Unlike ordinary kriging, you might have multiple correlation models, but due to the constraints on each of the models (see Chapter 30), they will all have the same geometries (range, angle, etc). You should not absolutely adhere to this approach, however, as special circumstances may prevail. For example, you may have insufficient data in some areas of the site to restrict the search neighborhood. These types of circumstances may warrant either a wider or smaller neighborhood, respectively.

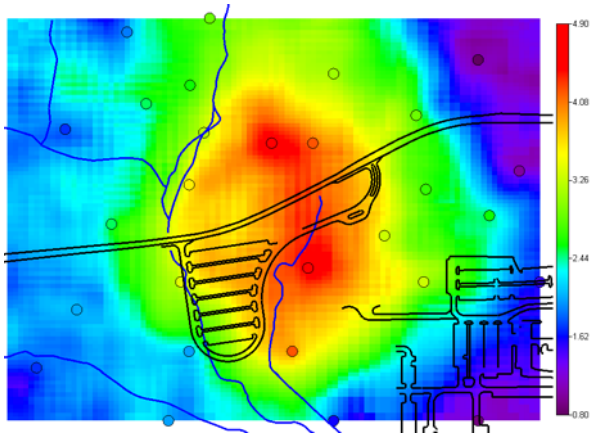


6. Create a grid across the site (previously discussed).



7. Apply kriging algorithm.

This is completely handled by SADA. Once the correlation models are available and a search neighborhood is specified, pressing the *Show The Results* button will produce the cokriging map.

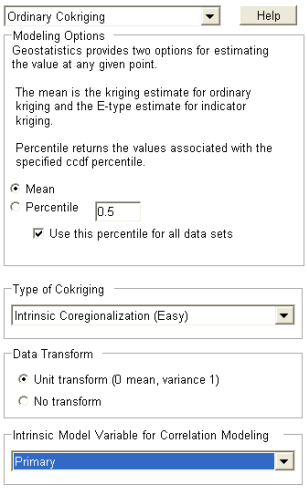


You can produce two types of maps from the ccdf: mean and percentile. Under the assumption of local normality, the mean and the 50th percentile will be equal. See the section below on mapping percentiles as a means of understanding uncertainty.

Cokriging Example

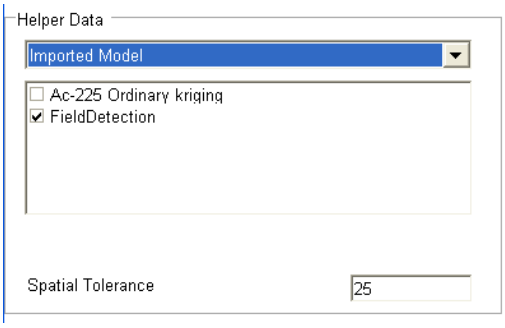
Open up Geospatial.sda. If you just finished the indicator kriging example, SADA will complain about the site boundaries we were using with Chlordane. This time we will be working with Ac-225. Therefore, if the warning message appears and you have the option to keep or adjust boundaries, choose the latter. Then choose *Snap boundaries to current data set only*. Save your SADA file.

Select *Interpolate my data* from the interview drop-down list. In the *Interpolation methods* step, choose *Ordinary Cokriging* from the list of available interpolants. For this example, we'll choose to interpolate using the mean of the distribution, and we'll use the simplest cokriging format, the intrinsic model of coregionalization for correlation structures. For a review of what the different options for correlation modeling under cokriging are, please revisit Chapter 30. This option only models the correlation structure for either the primary or secondary data sets. We will choose the *Unit transform* option. Make sure your parameter window looks like the following image:

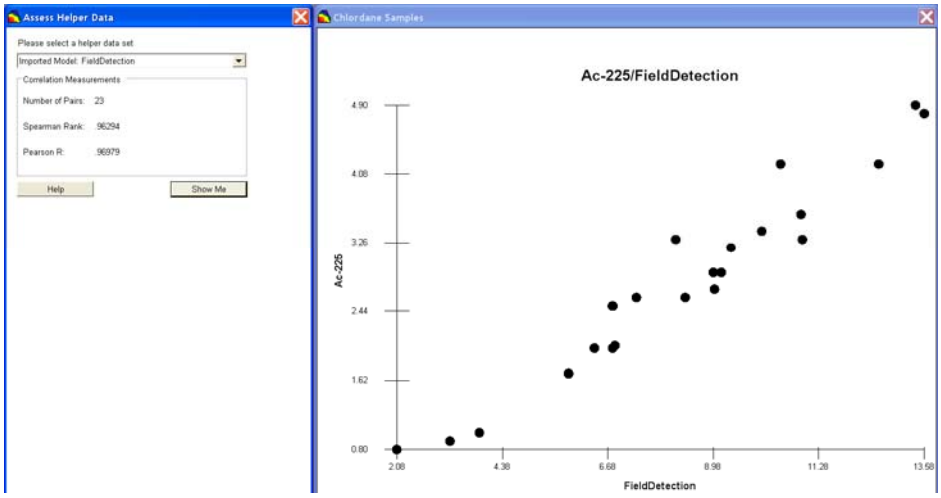


NOTE: If you did the 3d example in indicator kriging, you may need to set the vertical layering scheme back to *surface*. Do this now in the *Setup the site step*.

Now let's pick our secondary data set (helper data). Click on the Choose Helper Data step. In the parameter window, select the model called FieldDetection. This is mock field detection gridded across the site in a previous application. Since the grid nodes are not likely to fall exactly on Ac-225 sample points, select a spatial tolerance of 25.

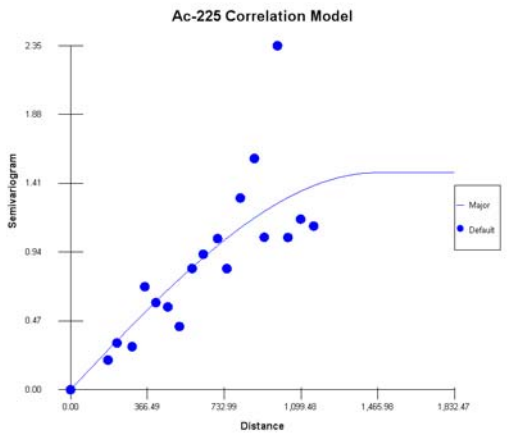


The next step is to assess the correlation between FieldDetection and Ac-225. Click on the *Assess helper data* step. Select the *FieldDetection* helper data set and press the *Show Me* button in the parameter window.

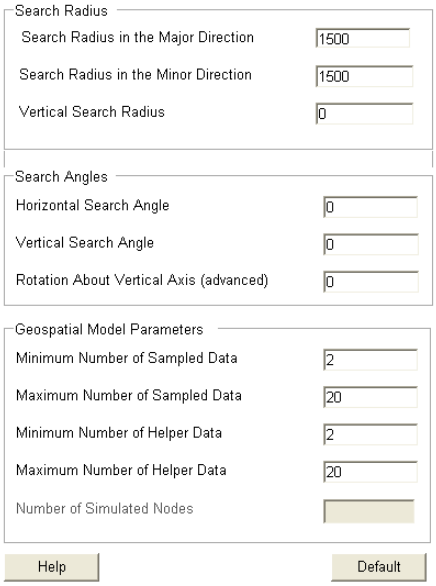


We see two results. First, in the parameter window itself there is a summary result that reports the *Number of Pairs*, *Spearman Rank,* and *Pearson R* values. The latter are measures of correlation strength. In the graphics window is a scatter plot of collocated *FieldDetection* and *Ac-225* data. This kind of correlation strength may not be encountered in practice.

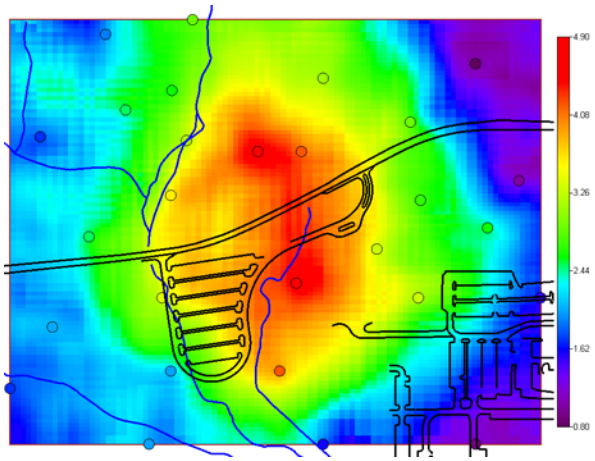
We'll need to assess the correlation model for the unit transform of the Ac-225 data. Click on the *Correlation modeling* step and accept the recommendations for both the experimental variography and modeling (spherical) parameters. Press *Show Me* in the parameter window.



A more thorough analysis using the Rose map would capture the anisotropic conditions clearly at play in this site. We did this very type of thing in the ordinary kriging example. If interested, return to the ordinary kriging example and follow exactly the same steps. In the meantime, we'll move on to the *Search neighborhood* step and enter a major and minor search radius informed by the current correlation model. Make sure your parameter window looks like the following:



Now set the grid specs. Click on the *Set grid specs* step and set the grid *Number* to 100x100. After this is all complete, press the *Show Me The Results* button in the steps window.



This result is informed by both the Ac-225 data and the field detection data through the correlation between the two. You may want to export, store, or document this result. Please look at the ordinary kriging example for how to accomplish this.

Uncertainty

Within the kriging paradigm, there are three ways to quantify: variance maps, percentile maps, and probability maps.

Variance Maps

This is essentially a map of the kriging variances for ordinary kriging and cokriging. This map is useful in understanding spatial data gaps. With a fully parameterized ordinary kriging or cokriging model specified, switch to the interview *Draw a variance map* and press *Show The Results*.

Percentile Maps

You can plot different percentiles of the ccdf distribution. Normally, kriging maps are created by plotting the mean or E-type estimate (indicator methods). It is also possible to plot percentiles to get a sense for how variable the estimates are as a function of location. The chapter on decision analysis (Chapter 35) makes use of these principle in combination with a decision criteria to create bands of uncertainty regarding the area of concern.

Probability Maps

Since we are dealing with a distribution at each location, it is possible to specify a decision criterion and then calculate the probability that the decision criterion is exceeded. This is also treated in Chapter 35, “Decision Analysis.”

Geostatistical simulation provides an alternative approach to uncertainty that is more flexible and more powerful, particularly in the estimation of spatial uncertainty (uncertainty over areas instead of at points). This will now be discussed.

Part II: Simulation

Simulation allows you to create equiprobable maps of your contaminant. Rather than choosing a single estimation value, such as the mean or a percentile, simulation randomly draws values from the CCDF at each node. The random draws are done such that the correlation structure (correlation model) and data histogram is reproduced each time, however. Simulation maps avoid the smoothing effect that is typically observed on interpolated maps. Instead of a single

map of interpolated values, the output of the simulation algorithm is a set of multiple equiprobable representations (realizations) of the spatial distribution of attribute values.

Simulated maps can serve the following purposes:

- Measure of local uncertainty. At each grid node, the single interpolated value computed using kriging or nearest neighbors is now replaced by a set of simulated values. The larger the differences among simulated values (e.g., measured by the variance or inter-quartile range), the larger the uncertainty at that location. One can also compute the probability of exceeding a particular threshold simply by counting the proportion of simulated values that are larger than that threshold value. This replaces the model of uncertainty in ordinary kriging and cokriging, which relies on the kriging variance.
- Measure of spatial uncertainty. The probability that a threshold is exceeded at every point does not provide a measure of uncertainty about whether these locations *jointly* exceed that threshold. In addition to a measure of "local" uncertainty, one thus needs to assess the "spatial" uncertainty (the uncertainty attached to the spatial distribution of values at many locations simultaneously). The quantification of spatial uncertainty is particularly important for the detection of clusters, such as cluster of high pollutant concentrations or string of high permeability values that may represent a flow path.

  The spatial uncertainty can be assessed visually by looking at a series of realizations; areas that remain stable over all realizations (low uncertainty) are distinguished from those where large fluctuations occur between realizations (high uncertainty). The uncertainty can also be quantified numerically by counting the number of realizations where a specific threshold is exceeded simultaneously over the locations of interest.
- Propagation of uncertainty. The set of simulated maps can be used to propagate the uncertainty through spatial operators or transfer functions. For example, a set of simulated permeability maps can be fed into a flow simulator, yielding a distribution of response values, such as travel times to the water table.
- Change of support. Decision-making is frequently performed over supports (e.g., remediation units or flow simulation cells) that are much larger than the measurement supports (e.g., soil core), hence some aggregation or upscaling must be conducted. This change of support is easily conducted within the framework of stochastic simulation. For the example of remediation units, simulated block values are computed by a simple average of simulated point concentrations that fall within that unit/block. The operation is repeated for each realization, yielding a set of simulated block values that quantify the uncertainty attached to the block concentration. The approach is very flexible in that any type of averaging (e.g., non-linear) can be applied, and once the grid of point values has been simulated, probability distributions for various block sizes and shapes can be derived at little computational cost.

Simulation of spatial phenomena can be accomplished using a growing variety of techniques that differ in the underlying random function model (multi-Gaussian or non parametric), the amount and type of information that can be accounted for, and the computer requirements. Sequential simulation is one of the most commonly used algorithms. The generic procedure is as follows:

1. Define a random path (i.e. using a random number generator), visiting each node of the simulation grid only once.
2. At each node, determine the probability distribution that models the uncertainty at that location.

3. Randomly draw a simulated value from that distribution, and add it to the set of observed values.
4. Proceed to the next location along the random path, and repeat the two previous steps.
5. Loop until all grid nodes are simulated.

Other realizations are generated by repeating the procedure using a different random path and set of random numbers to draw from the probability distributions. SADA handles this procedure internally.

Two major classes of sequential simulation algorithms can be distinguished, depending on whether the series of probability distributions (conditional CDF, CCDF) at step 2 are modeled using the multi-Gaussian (sGs = sequential Gaussian simulation) or the indicator (sis = sequential indicator simulation) formalism.

Sequential Gaussian Simulation

Sequential Gaussian simulation relies on the assumption that each CCDF is Gaussian; hence, it is fully characterized by its mean and variance, which correspond to the simple kriging estimate and variance. The approach typically requires a prior normal score transform of data to ensure that at least the univariate distribution (histogram) is normal. Once all values have been simulated in the normal space, they are back-transformed so as to reproduce the histogram of the original variable.

If you can establish an ordinary kriging model, then with little additional effort, you can setup a full sequential Gaussian simulation. The recipe for both are similar, but sequential Gaussian simulation requires a handful of additional parameter values above and beyond those required by its ordinary kriging counterpart.

Number of Simulations

This is simply the number of simulations you wish to run. It is recommended that until the model has been fine tuned (with search neighborhoods, etc.), one should use a small value for this parameter (e.g., 10). Once the simulation is fine tuned, large values can be used (e.g., 1000).

Minimum Permitted Simulation Value

Since we are assuming a Gaussian model, there is no theoretical limit to lower values that can be selected. In practice, though, there may be practical constraints. For example, we would not expect to find negative contaminant concentrations. So this constrains or truncates the results on the left side of the distribution. One could set this value to something reasonable, such as the lowest measured value in the data set, half that value, or zero. Of course, it depends on the application.

Maximum Permitted Simulation Value

In the same way, there is no theoretical limit to the upper values that can be selected. In practice, though, there may be practical constraints. One could set this value to something reasonable, such as the highest measured value in the data set or the maximum possible value one could see (e.g., 100% contaminant).

Lower Tail Option

This is used when back transforming from the normal score transformed values to the real world values. It specifies the interpolation in the lower tail of the distribution:

Linear: Performs a linear interpolation to the lower limit specified by the *Minimum Permitted Simulation Value.*

Power: Power model interpolation, with parm = ltpar, to the lower limit specified by the *Minimum Permitted Simulation Value*.

Tabulated: Interpolation is based on internally tabulated values.

Upper Tail Option

This is used when back transforming from the normal score transformed values to the real world values. It specifies the interpolation in the upper tail of the distribution:

Linear: Performs a linear interpolation to the lower limit specified by the *Minimum Permitted Simulation Value.*

Power: Power model interpolation, with parm = ltpar, to the lower limit specified by the *Minimum Permitted Simulation Value*.

Tabulated: Interpolation is based on internally tabulated values.

Hyperbolic: Hyperbolic interpolation with parm = ltpar to the upper limit specified by the *Maximum Permitted Simulation Value.*

Minimum Data Value

This is the lowest measured value to consider in simulation. In other words, it is the lower trimming limit on your data values, which is useful in eliminating outliers. SADA will use the lowest data value as the minimum value or you can enter a custom value.

Maximum Data Value

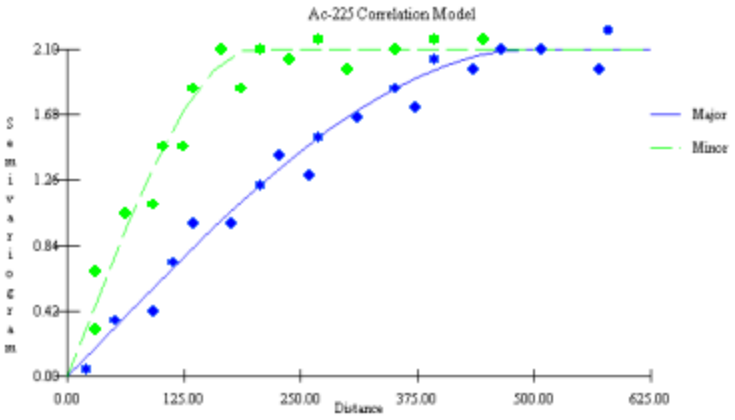
This is the lowest measured value to consider in simulation. In other words, it is the lower trimming limit on your data values, which is useful in eliminating outliers. SADA will use the highest data value as the maximum value or you can enter a custom value.

Practical Recipe for Sequential Gaussian Simulation (SGS)

The steps for SGS are almost identical to those of ordinary kriging. There are only a couple of additional steps in specifying the maximum number of previously simulated nodes to use in the simulation of each node and the number of simulations to run, and the correlation model is based on the normal score transform.

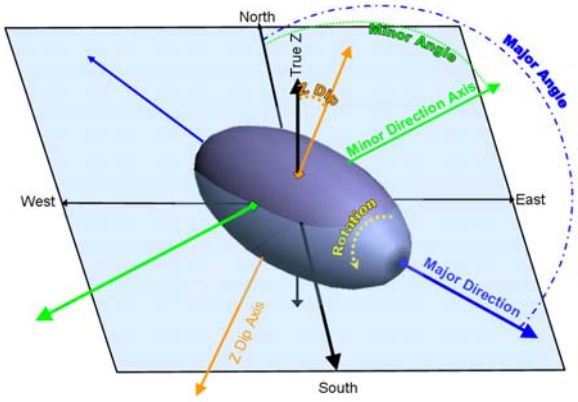
1. *Set Sequential Gaussian Simulation specific parameters*. This step includes the number of simulations, lower tail option, upper tail option, minimum permitted simulation, maximum permitted simulation, minimum sample value to consider, and maximum sample value to consider.
2. Develop a correlation model for the normal score transformed data (see Chapter 30).

   When you select SGS in SADA, you will be restricted to evaluating the correlation model for the normal score transformed data. You will follow exactly the same type analysis as if the data were untransformed.

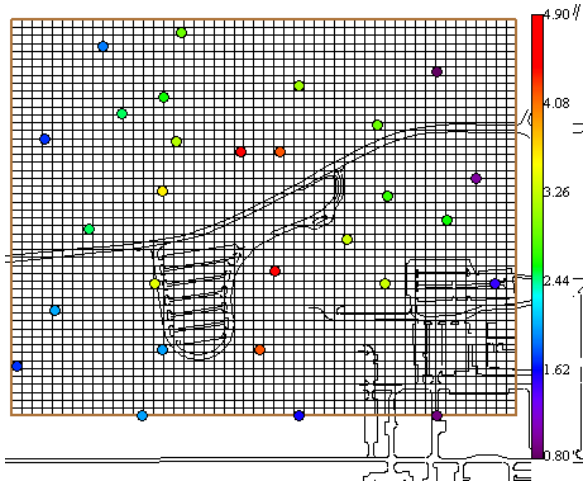


3. Define a search neighborhood

   This is done in exactly the same way as with the ordinary kriging example. Here, one can calibrate the search neighborhood based on the result of the correlation model. So for example, if the correlation range in the major direction at a 45 degree angle is 500ft and 300ft in the minor direction, this might well inform the choice for the search neighborhood. You should not absolutely adhere to this, however, as special circumstances may prevail. For example, you may have insufficient data in some areas of the site to restrict the search neighborhood. These types of circumstances may warrant either a wider or smaller neighborhood, respectively. You will also need to specify the number of previously simulated nodes to use within each search neighborhood. In simulation, we have an extra parameter: Number of simulated nodes. During sequential simulation, nodes are simulated one at a time. During the simulation of any given node, any previously simulated node may also be used as a conditioning value.

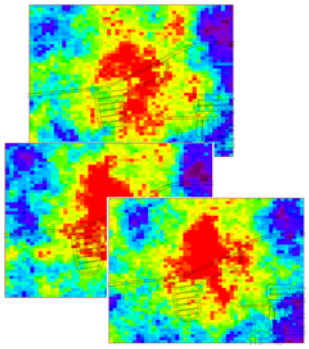


4. Create a grid across the site.



5. Apply simulation algorithm

   This is completely handled by SADA. Specify the number of simulations you want to run, the normal score correlation model, and a search neighborhood. Press the *Show The Results* button, and the simulations will be produced. Keep in mind that it may take some time to produce these results, especially for 3d simulations. Be patient.



6. Store the results immediately

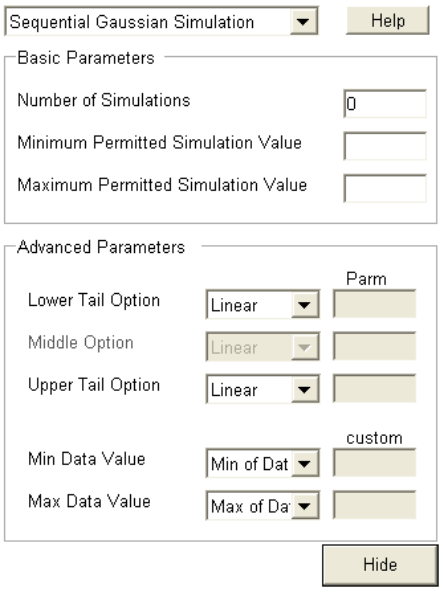
   This is probably the most important step you can take. As soon as the simulation is run, click on the step *Manage model results*, enter a name into the simulation name text box, and press the *Store* button. Save your SADA file immediately. If you are not satisfied with the run, you can always delete the simulation set later.

7. Post-process results

   There are a variety of way for you to view and post process your simulations to get estimates and maps of uncertainty. Since these are the same for sequential indicator simulation, they will both be covered in the later section on post-processing simulation sets.

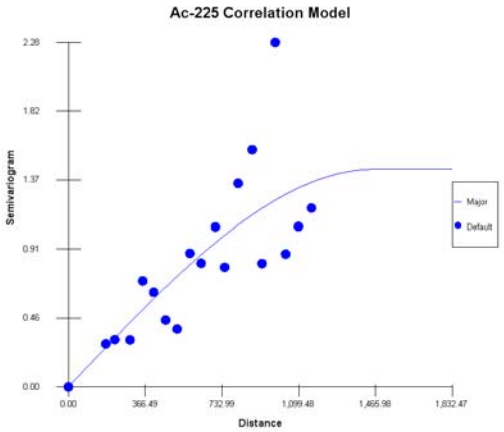
A Sequential Gaussian Simulation Example

Open up Geospatial.sda and select Soil and *Ac-225*. Switch to the interview *Perform geostatistical simulation*. In the *Setup the site* step, confirm that the vertical layering scheme is set to *Surface Only*. In the *Select simulation method* step, choose *Sequential Gaussian Simulation*. Three parameter choices appear along with an *Advanced* button. Press the *Advanced* button now.



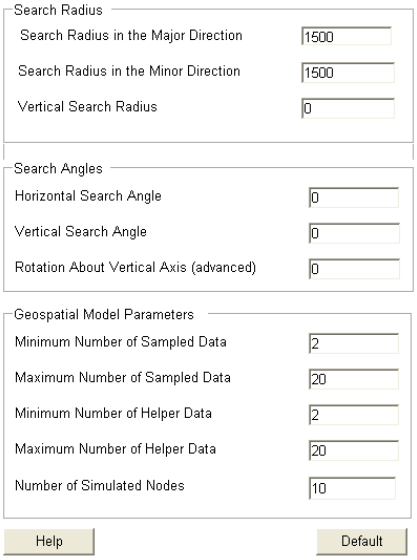
Here is where you will find the Sequential Gaussian Simulation parameters. Let's begin by specifying only 3 simulations. Let's set the *Minimum Permitted Simulation Value* to zero (you could set it to background) and the *Maximum Permitted Simulation Value* to 5.0pC/g. This is just slightly greater than the Max Data Value.

In the *Advanced Parameters* box, we will accept the default values for now. Next, we'll need to model the spatial correlation under the normal score transform. Click on the *Correlation modeling* step. Notice that the only data transform available is the Normal score. At this point, we would do a comprehensive semi-variogram analysis using the rose map to explore for anisotropy and so forth. Please revisit Chapter 30 for a more detailed discussion. In the interest of simplicity, we'll blindly accept SADA's recommendations for semi-variogram parameters and correlation model parameters (spherical). Press the *Recommend* button in the *Explore Experimental Semi-variography* parameter block first. Then press the *Recommend* button in the *Model Semi-variography Values* parameter block (choose only *Spherical* when prompted). You should see the following result:

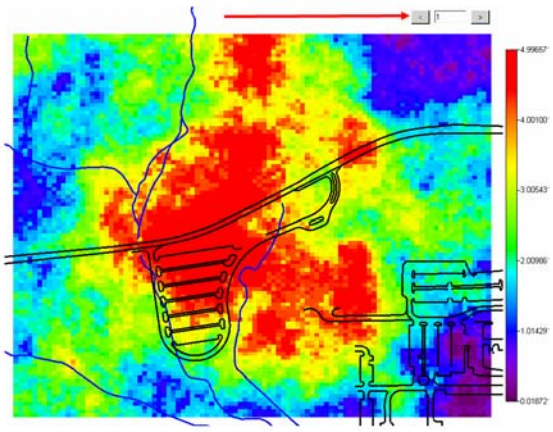


Next, you'll need to choose the search neighborhood parameters. We'll choose the search ranges based on the correlation model results. In simulation, we have an extra parameter: *Number of Simulated Nodes*. During sequential simulation, nodes are simulated one at a time. During the simulation of any given node, any previously simulated node may also be used as a conditioning value. Therefore, you'll need to specify the maximum number of such nodes to use

when encountered within the search window. We'll use 10. Make sure that your parameters look like those in the following image.

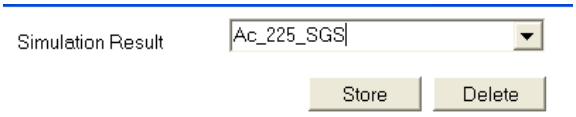


We are now ready to simulate. Click on the *Show the results* step. Click on the *Show The Results* button in the parameter window. SADA will show a progress window as it calculates each simulation. When finished, you should see a result like the following image (your result may be different due to the nature of random number draws).



In the upper-right-hand corner, there is a text box centered between two arrow buttons. This is how you can navigate through your simulation results. Press the forward arrow (on the right). Now, you are looking at the second simulation. Press it again to see the 3rd simulation. You can now reverse through the sequence by pressing the back arrow (on the left). You can also enter a simulation number, such as 1, and press Enter on your keyboard to jump directly to a specific simulation.

Each of the simulations represents a distinct way that contamination could be dispersed across your site based on the data you have and the correlation model you're using. The first thing to do is save these results. Large simulation runs can take a long time, especially for three-dimensional applications. Click on the *Manage model results* step. In the parameter window, enter Ac_225_SGS and press *Store*.



Keep in mind that you will also need to save your SADA file to make the save permanent. Now, Ac_225_SGS will appear in the drop list. Check the drop-down list to see it. If the result is unsatisfactory, you can delete it by pressing *Delete* (don't do this now).

To use this simulation in the future, click on the *Select simulation method* step and select *Use stored result* in the parameter window. Select Ac-225_SGS and press *Show The Results*. The simulation will immediately appear in the graphics window.

Simulations are made useful through the use of post processing techniques to acquire different measures of uncertainty from the results. Since sequential indicator simulations are post processed in exactly the same way, this discussion will be paused momentarily to visit this non-parametric alternative.

Save your SADA file now.

Sequential Indicator Simulation

Sequential indicator simulation (SIS) is the simulation counterpart to indicator kriging and requires little effort beyond that required for indicator kriging. Like indicator kriging, it does not assume any particular shape or analytical expression for the conditional distributions. Instead, the values of the CCDFs are computed for a series of threshold values using indicator kriging. Then, the complete function is obtained by interpolation/extrapolation of the estimated probabilities. This approach requires the coding of each observation into a series of indicator data (one for each threshold), followed by semivariogram modeling and kriging of indicators for each threshold.

The indicator approach appears much more demanding than the multi-Gaussian approach, both in terms of semivariogram modeling and computer requirements. This additional complexity is balanced by the possibility of modeling spatial correlation patterns specific to different classes of attribute values through indicator semivariograms. In particular, the connectivity of extreme values can be accounted for, whereas the Gaussian model does not allow for any significant spatial correlation of very large or very small values, a property known as destructuration effect. A potential pitfall for the indicator approach is the interpolation or extrapolation of the estimated probabilities to derive a continuous ccdf model. Characteristics of the ccdf, such as the mean or variance, may overly depend on the modeling of the upper and lower tails of the distribution

Practical Recipe for Sequential Indicator Simulation (SIS)

If you can establish an indicator kriging model, then with little additional effort, you can setup a full sequential indicator simulation. The recipe for both are similar, but sequential indicator simulation requires a handful of additional parameter values above and beyond those required by its ordinary kriging counterpart.

Number of Simulations

This is simply the number of simulations you wish to run. It is recommended that until the model has been fine tuned (with search neighborhoods, etc.) one should use a small value for this parameter (e.g., 10). Once the simulation is fine tuned, large values can be used (e.g., 1000).

Lower Tail Option

This is used when back transforming the values. It specifies the interpolation in the lower tail of the distribution:

Linear: Performs a linear interpolation to the lower limit specified by the *Minimum Permitted Simulation Value*.

Power: Power model interpolation, with parm = ltpar, to the lower limit specified by the *Minimum Permitted Simulation Value*.

Tabulated: Interpolation is based on internally tabulated values.

Upper Tail Option

This is used when back transforming the values. It specifies the interpolation in the upper tail of the distribution:

Linear: Performs a linear interpolation to the lower limit specified by the *Minimum Permitted Simulation Value*.

Power: Power model interpolation, with parm = ltpar, to the lower limit specified by the *Minimum Permitted Simulation Value*.

Tabulated: Interpolation is based on internally tabulated values.

Hyperbolic: Hyperbolic interpolation, with parm = ltpar, to the upper limit specified by the *Maximum Permitted Simulation Value.*

Minimum Data Value

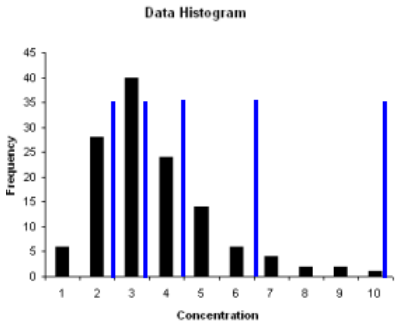
This is the lowest measured value to consider in simulation. In other words, it is the lower trimming limit on your data values, which is useful in eliminating outliers. SADA will use the lowest data value as the minimum value or you can enter a custom value.

Maximum Data Value

This is the highest measured value to consider in simulation. In other words, it is the higher trimming limit on your data values, which is useful in eliminating outliers. SADA will use the highest data value as the maximum value or you can enter a custom value.

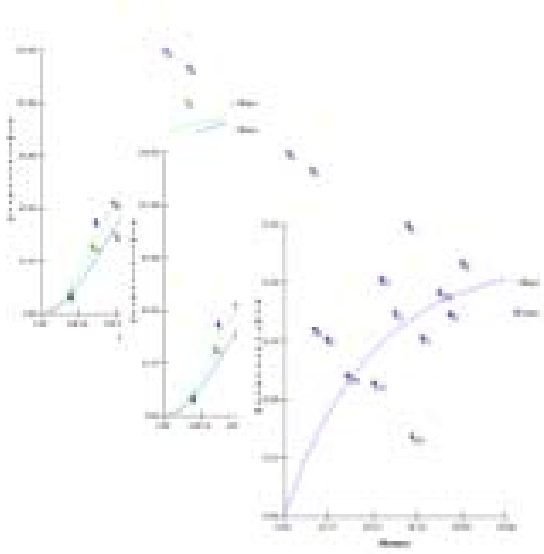
1. Determine set of partition values.

   Choose the threshold values so that you well represent the features of your distribution. More threshold values should be located in areas of the distribution where the function is changing more rapidly. Also, consider using the decision threshold (if one is available) as one of the indicator thresholds. This will increase the accuracy of your model in at that range.



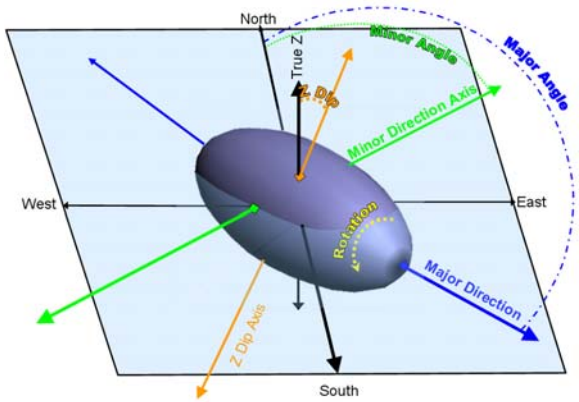
2. *Set Indicator Gaussian Simulation specific parameters*. This step includes the number of simulations, lower tail option, upper tail option, minimum permitted simulation, maximum permitted simulation, minimum sample value to consider, and maximum sample value to consider.
3. Develop semi-variogram models for each partition value.

   Be sure to review the section in Chapter 30 that deals with how to set up correlation models for indicator kriging/simulation. A correlation model must be available for the indicator transformed data for every indicator threshold you specify.

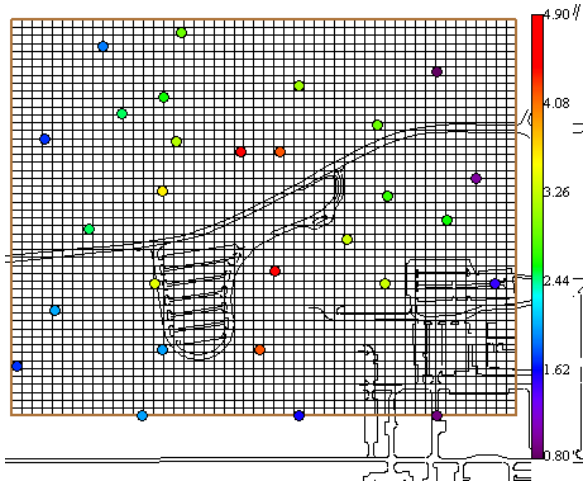


4. Define reasonable search neighborhood.

   This is done in exactly the same way as with the indicator kriging model. Since there are multiple correlation models (one for each indicator threshold), you may have several ranges to select from. Begin by trying the largest range. If you have anisotropic conditions and varying degrees of anisotropy for each cutoff, consider using an isotropic variogram, or perhaps choose the correlation structure ranges near or at your decision criteria. You should not absolutely adhere to this approach, however, as special circumstances may prevail. For example, you may have insufficient data in some areas of the site to restrict the search neighborhood. These types of circumstances may warrant either a wider or smaller neighborhood, respectively. In simulation, there is an extra parameter: Number of simulated nodes. During sequential simulation, nodes are simulated one at a time. During the simulation of any given node, any previously simulated node may also be used as a conditioning value.

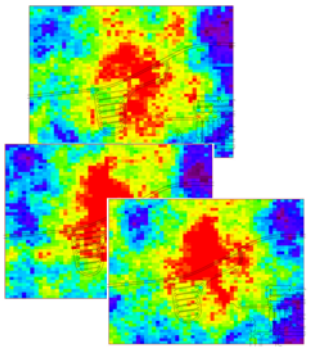


5. Create a grid across the site (previously discussed).



6. Apply simulation algorithm.

   This is completely handled by SADA. Specify the number of simulations you want to run, the normal score correlation model, and a search neighborhood. Press the *Show The Results* button, and the simulations will be produced. Keep in mind that it may take some time to produce these results, especially for 3d simulations. Be patient.



7. Store the results immediately.

   This is probably the most important step you can take. As soon as the simulation is run, click on the step *Manage model results*, enter a name into the simulation name text box,

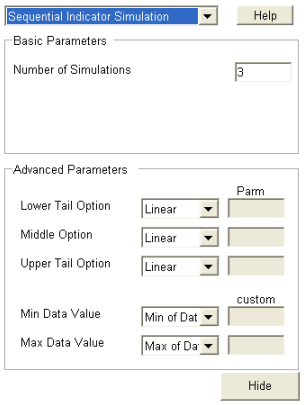
and press the *Store* button. Save your SADA file immediately. If you are not satisfied with the run, you can always delete the simulation set later.

8. Post-process results

   There are a variety of ways for you to view and post process your simulations to get estimates and maps of uncertainty. Since these are the same for Sequential Gaussian Simulation, they will both be covered in the later section on post-processing simulation sets.

A Sequential Indicator Simulation Example

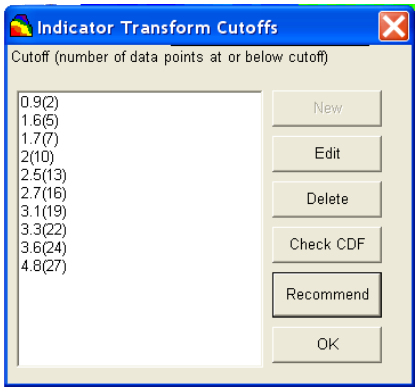
If you have not opened Geospatial.sda yet, do so now and select Soil and Ac-225. In the interview drop-down list, select *Perform Geostatistical Simulation*. Click on the step *Select Simulation Method* and select *Sequential Indicator Simulation*. If the *Advanced* button appears, press it now to expand the view to include additional parameters.



Here is where some indicator simulation specific parameters are presented regarding the tail and data trimming limits discussed above. Also, the *Number of Simulations* is required here. From the previous exercise, 3 have been selected. We'll use 3 again and accept the default values for the *Advanced Parameters*.

Next, we'll setup our indicator thresholds and correlation models for each of them. Click on the *Correlation modeling* step. In the *Choose Data* parameter block, click on the *Edit* button next to the *Data Transforms* drop-down list. You may have some values in here from a previous exercise. We'll overwrite these (if you have them) by asking SADA to recommend threshold values that will represent the cumulative distribution function for Ac-225.

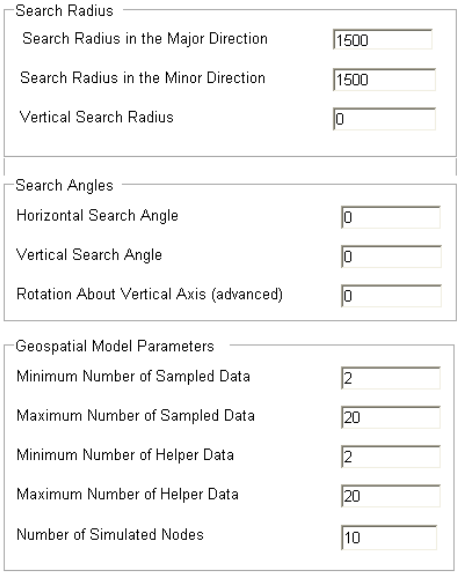
Press the *Recommend* button in the threshold manager and when prompted answer *Yes*. SADA calculates the following threshold values.



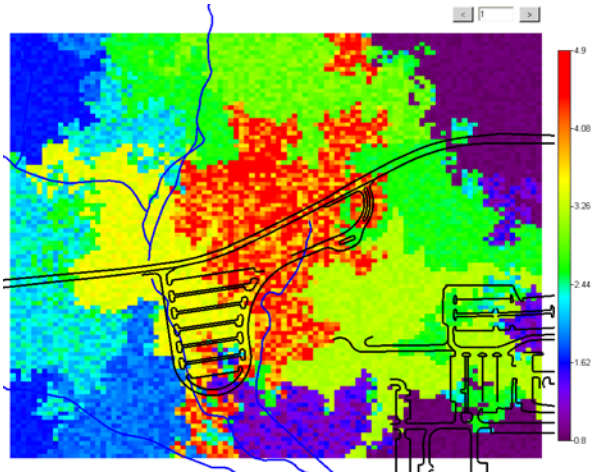
The selection of threshold values should be well considered. You may have a particular range of values you are interested in or a particular threshold value you wish to have special emphasis. The *Recommend* button in SADA simply assigns threshold values taken at approximately every decile (see Chapter 30). Press *OK*.

Now, we'll need to establish a correlation model for each indicator transform. Select the first one in the *Data Transform* drop-down list, with a value of 0.9. For each of these transforms, you should do a comprehensive semi-variogram analysis (see Chapter 30). In the interest of simplicity, we will accept the recommended parameter values for both the semi-variography as well as the models. With 0.9 selected, press both *Recommend* buttons in the parameter window. For the correlation model recommendation, choose *Spherical*. Repeat this for each indicator threshold. When you are done, save your SADA file.

Moving on to the *Search neighborhood* step, in the interest of brevity, we'll use the same search neighborhood values as in the previous example. In practice, your search neighborhood values would need to be based somewhat on the correlation ranges you discovered in the semi-variogram modeling process. Make sure your parameter values look like the following.

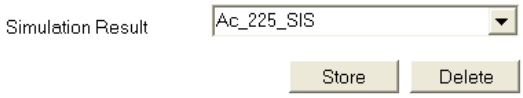


Make sure the grid specs are still 100 x 100, and press the *Show The Results* button.



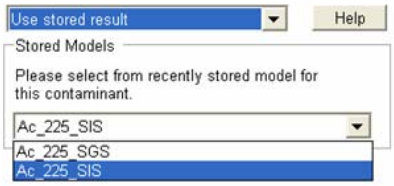
In the upper-right-hand corner, there is a text box centered between two arrows. This is how you can navigate through your simulation results. Press the forward arrow (on the right). Now you are looking at the second simulation. Press it again to see the third simulation. You can now reverse through the sequence by pressing the back arrow (on the left). You can also enter a simulation number, such as 1, and press Enter on your keyboard to jump directly to a specific simulation.

Each of the simulations represents a distinct way that contamination could be dispersed across your site based on the data you have and the correlation model you're using. The first thing to do is save these results. Large simulation runs can take an especially long time, especially for three dimensional applications. Click on the *Manage model results* step. In the parameter window, enter Ac_225_SIS and press *Store*.



Keep in mind that you will also need to save your SADA file to make the save permanent. Now Ac_225_SIS will appear in the drop list. Check the drop-down list to see it. If the result is unsatisfactory, you can delete it by pressing *Delete* (don't do this now).

To use this simulation in the future, click on the *Select simulation method* step and select *Use stored result* in the parameter window. Select *Ac-225_SIS* and press *Show The Results*. The simulation will immediately appear in the graphics window.



Simulations are made useful through the use of post-processing techniques to acquire different measures of uncertainty from the results. Save your SADA file now, and we'll discuss post processing.

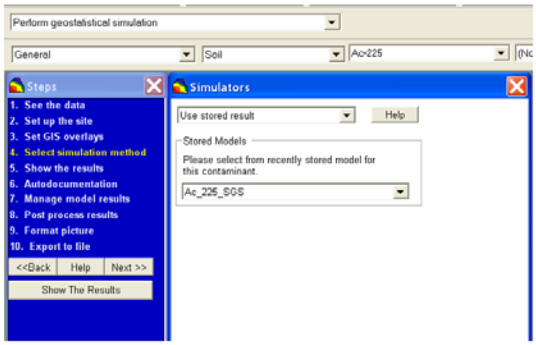
Post Processing Simulation Results

The most import thing to do before you post-process any simulation is to first store your simulation run. It's very easy to do something unintentional and lose your entire run. If you followed the examples above for indicator and Gaussian simulation, then your results have been stored.

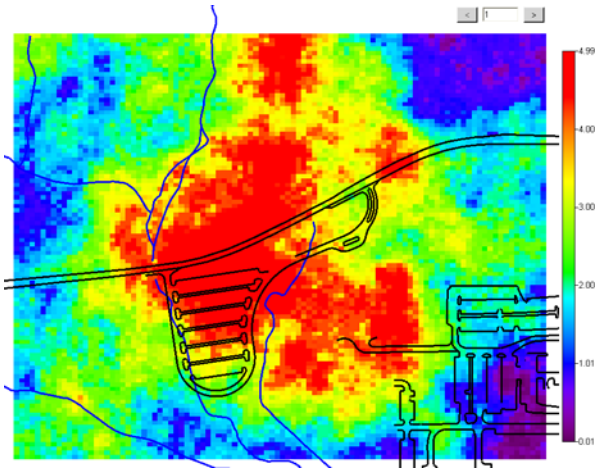
If you have not completed the examples above, please return and complete the example *Sequential Gaussian Simulation Application* now. We'll be using this as a base simulation.

Open Geospatial.sda, choose *Soil*, *Ac-225*, and the interview *Perform geostatistical simulation*. Click on the step *Select simulation method* and choose *Stored Result*. Select *Ac_225_SGS* from the drop-down list of stored results. Press *Show The Results* to recall the run.

If you don't see the *Ac_225_SGS* option, you might not have completed the Gaussian simulation example or failed to save your SADA file afterwards. If you don't see this result, you'll need to repeat the sequential Gaussian example (it's fast) and then return here.



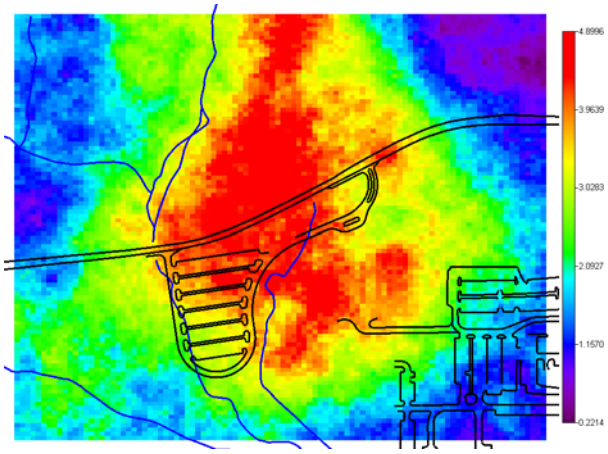
In the graphics window, you should see the first simulation (It may look different than the following):



Now, we are ready to post-process these results. Click on the *Post process results* step and we'll discuss the options seen there.

Create a Local Estimation Map

As with kriging, you can create a location estimation map using simulation. Here, every simulation is added together and divided by the number of simulations. Essentially, the simulations are averaged together. For sequential Gaussian simulations, this average should be based on a large number of simulations. Press the *Create Contour Map* button now and the average is calculated.

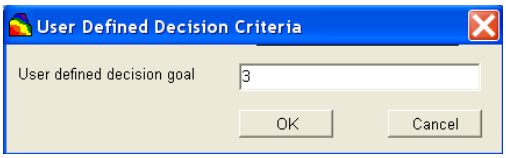


The result is fairly heterogeneous because we only have three realizations (simulations). You can store this map as a model and use it on most decision framework or sampling designs that SADA provides. Let's store it now.

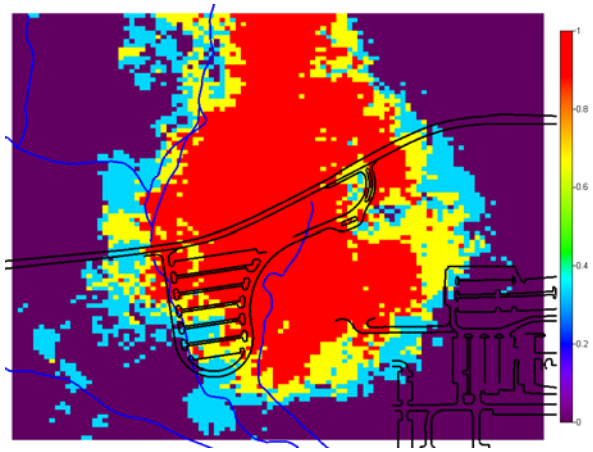
Click on the *Manage model results*. Enter the name *Ac_225_SGS_Average* and press *Store*. Save your SADA file.

Create a Local Probability Map

This is the first substantial encounter with a geospatial decision model. If a decision criterion is available (e.g., human health risk PRG, ecological benchmark, or any custom value), you can post process the results to determine the probability that the value will be exceeded at any given point. If you created a local estimation map, then the result in the graphics window is not technically a simulation run but rather a single map. If you now press *Create Probability Map*, SADA will complain. What you will need to do is press the *Restore Simulation* button near the bottom of the parameter window first. Do that now, and then press *Create Probability Map*. SADA will respond by asking you to enter a custom value. This is because we are under the *General* analysis. Enter 3 and press *OK*.

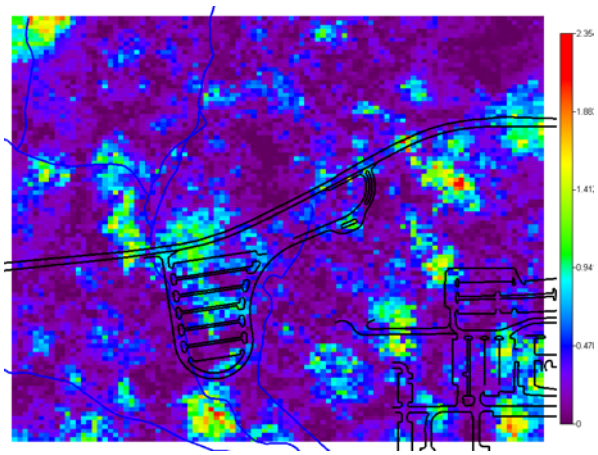


The resulting probability map is fairly jagged looking, because we only have three simulations, and therefore only the following probability values are possible: 0, 1/3, 2/3, or 1. When you have many, many simulations, the map will appear more reasonable. You can store this result if you like. If you do, be sure to save your SADA file.



Create a Local Variance Map

Let's begin by pressing the *Restore Simulation* button near the bottom of the parameter window. The local variance map is calculated at each node. The calculation simply uses every simulated value at the node and calculates its variance. The result is a map showing regional variations in variability. This kind of result can be useful in secondary sampling design. If you wish to use this in a secondary sample design, you will first need to store the result as a model in SADA (i.e., click on *Manage model results* and store).

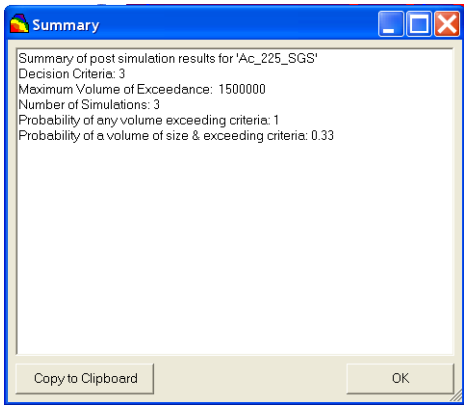


Contiguous Area Exceedances

Given a decision criterion and a maximum volume limit, this calculates the probability that you have a volume of a given size or greater that exceeds the decision criterion anywhere on your site. This is one of the strengths of simulation: the ability to estimate the probability that collections of grid cells (rather than individual cells) simultaneously exceed the decision criteria. In particular, this feature evaluates the probability that a contiguous volume/area of the site exceeds the value. SADA uses a *queen* contiguity rule. In a queen contiguity rule, all blocks that touch left, right, top, bottom, above or below (3d), and on any diagonal are considered contiguous. So, if two blocks exceed the criterion and are diagonal from each other, they are considered together to be a contiguous exceedance of the decision criteria.

All you need to do is enter the maximum contiguous volume that can exceed this criterion. This would obviously be specific to your application. Examples include a radiologically driven value or an ore volume value. Enter 1,500,000 and press *Generate Summaries*. Enter a value of 3.0 and press *OK*. It may take a while, but SADA will process each simulation and look for any contiguous area that exceeds 3.0.

SADA will generate a summary of its findings that looks like this.



There are a number of results reported here.

Decision Criteria: In this context, you cannot have depth-specific criteria. This approach has not been developed yet. You must use a single decision criterion. In this case, the criterion is 4.

Maximum Volume of Exceedance: This is the largest amount of contiguous volume that is permitted to exceed the limit. So, you might have a number of locations on the site that exceed 3, but if they are small enough the criteria is passed.

Number of Simulations: This is the number of simulations you ran (in this case only 3). In reality, the more you can do the better. Simulations of less than 100 are very small in number.

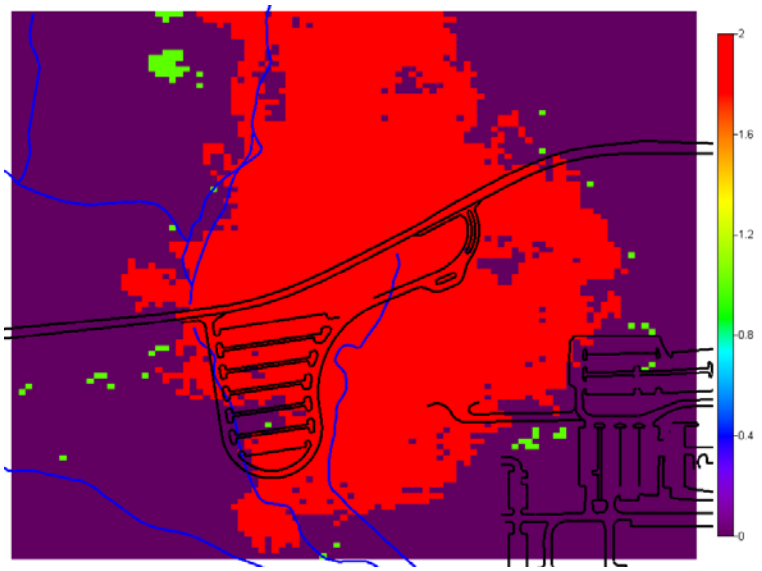
Probability of any volume exceeding criteria: This parameter reports how many times out of N simulations that any cell exceeded the criterion. In this case, we have a value of 1, meaning that all 3 simulations exceeded the criterion (no matter how small or large the volume of exceedance might have been).

Probability of a volume of size & exceeding criteria: This is the percentage of simulations in which a volume at or larger than the size of your maximum volume of exceedance parameter exceeded the criterion everywhere. In this case, we see that about 1/3 of the simulations showed an area larger than 1,500,000 square (cubic for 3d) feet exceeding the criterion of 3.

Press *OK* to see the visual results. SADA shows the results of each simulation post processing. Essentially, it creates another simulation set that contains the following:

- All those areas that exceeded the decision criteria and were too large (shown as category = 2).
- All those areas that exceeded the decision criteria and were small enough (shown as category = 1).
- All those areas that did not exceed the decision criteria anywhere (shown as category = 0).

The following image presents the post-processing results for a single simulation – simulation #2. You can see that we no longer have a continuous simulation of concentration values, but rather a processed set where the red area represents a contiguous area greater than 1,500,000 in size that exceeds a criterion of 3. The green areas also exceeded 3 but were not large enough to be included as problem areas.



Note that Version 5.0.78 will sometimes experience an error when post-processing simulation values. This is because a recursive algorithm is being used to search through the node space. Sometimes, this recursive algorithm becomes too complex and the code fails. If you experience this, you may need to reduce the grid resolution (e.g., 50x50).

Summary

Advanced geospatial methods provide a powerful set of tools for building a case for spatial model calibration and quantification of uncertainty. Later, we will see how this translates into the decision model. First, we will spend some additional time in the next chapter working with models, including storing them, exporting them, and applying post interpolation functions.

References

Deutsch, C. V. and A. G. Journel (1992). GSLIB: Geostatistical Software Library and User's Guide. New York, Oxford University Press.

Isaaks, E., Srivastava, R.M. (1989). An Introduction to Applied Geostatistics. New York, Oxford University Press.

Goovaerts, P. (1997). Geostatistics for Natural Resource Evaluation. New York, Oxford University Press.

Goovaerts, P. (2000). "Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall." Journal of Hydrology 228(1-2): 113-129.

Chapter 32: Working With Models

With SADA, you can perform simple mathematical operations on any imported or stored map. This creates a new map that can be stored as well. In this way, you can process your results through functions of your own. SADA provides a set of basic map functions for individual maps.

If you have two or more imported or stored maps, you can allow one map to operate on another with functions as simple as subtracting maps (good for telling the difference between models) to using one model to cut and replace the contents of another.

This step is available under the interview *Simple map arithmetic*, which is available whenever an imported or stored model is visible in the graphics viewer.

Select *Simple map arithmetic* from the interview box and click on the *Select map function* step.

Map functions are organized into two categories: single map and double map operations. The top drop-down list presents the single map functions. These functions operate on whatever map is currently available in the viewer. Some of these operations require a parameter value, which you enter in the *Parameter* box. The parameter SADA needs is typically intuitive but will be explained here. Let's start by clicking on the drop-down list for the *Current Map Functions*.

Current Map Functions

With *Current Map Functions*, it is possible to construct fairly complex custom models and extend your spatial models further. We now provide a brief description of the elemental algebraic functions found here and provide a quick example.

Add

Adds the parameter value to the current map. To try this, enter a value of 3 in the *Parameter* box and press *Calculate* (top box). Your map should be increased unilaterally by 3.

Subtract

Subtracts the parameter value from the current map. To try this, enter a value of 3 in the *Parameter* box and press *Calculate* (top box). Your map should be decreased unilaterally by 3.

Multiply by

Multiplies the map by the parameter value. To try this, enter a value of 3 in the *Parameter* box and press *Calculate* (top box). Your map should be three times its original value.

Divide by

Divides the current map by the parameter value (parameter value cannot be zero here). To try this, enter a value of 3 in the *Parameter* window and press *Calculate* (top box). Your map values should be three times smaller than before.

Take Natural Log

Takes the natural log of the map.

Antilog (natural)

Takes the antilog (e) of the current map.

Take Base 10 Log

Takes the log base 10 of the map.

Antilog (base 10)

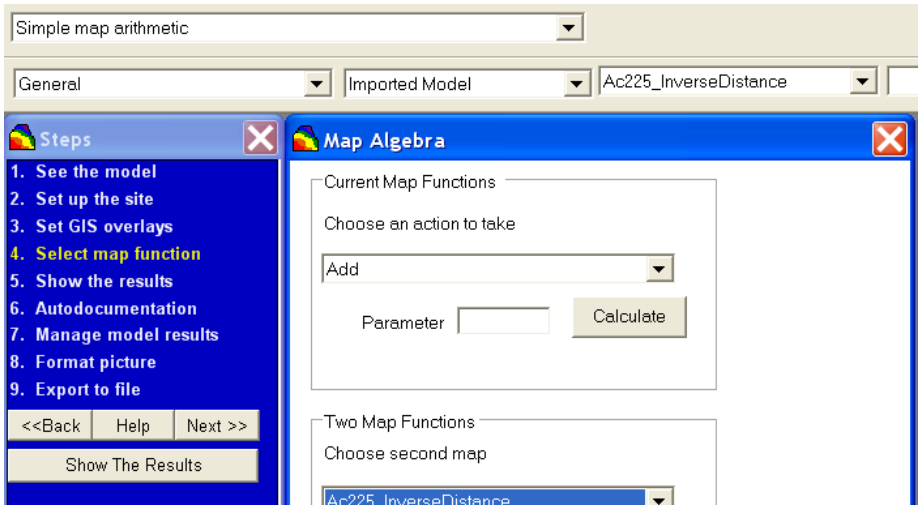
Takes the antilog (base 10) of the current map.

Raise to power

Uses the parameter value as the power to which map values are raised (be wary of overflows here).

A Quick Example

Open up the SimpleMapFunctions.sda file and select *Imported Model*. The Ac225_InverseDistance model you see was created earlier and stored here. Now, select the interview *Simple map arithmetic* and click on *Select map function*.



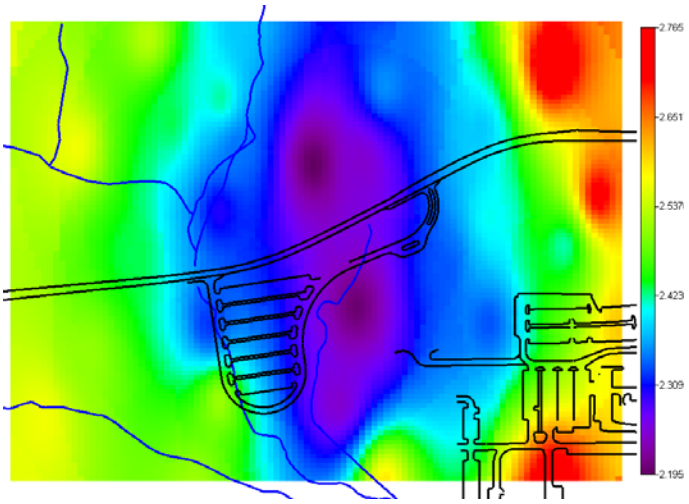
Let's extend this model by applying the following function to it:

$$f(u) = 2 + e^{-u/3}$$

where u is each modeled value. Applying this function to the map is much like using a calculator. You should adhere to the order of operations and apply each piece individually.

1. Select *Divide* in the drop-down list, enter -3 into the *Parameter* box and press *Calculate*.
2. Select *Anti-log (natural)* in the drop-down list and press *Calculate*.
3. Select *Add* in the drop-down list, enter 2 into the *Parameter* box and press *Calculate*.

You should see the following result. If you make a mistake along the way, just click on *See the model* to reset.



For practice, let's store this model. Click on *Manage model results*. Click on *Save model results*, enter "MyFunction" for the name, and press *OK.* When SADA returns control to you, select *Imported Model* and *MyFunction*. You should see the same result. Save your SADA file now to truly save your extended model.

2/2.5d Smoothing

This is a moving window average function that smoothes data in each layer. This differs from 3d smoothing in that averaging only occurs within each layer. That is, no cell values in layers below or above are included in the current cell's estimation. In particular, each cell is visited and is equal to the average of itself and its nearest N neighbors. You will enter the value of N in the parameter box. So, for example, if you enter a value of 1, then the new cell value is the average of the current value and each cell that touches it on all sides. If you enter a value of 2, then the new cell value is the average of the current value and each neighbor within two blocks distance. In the following figure, all cells inside the heavy black outline will be used to estimate the value of the blue cell.

To try this, let's further modify *MyFunction*. Select *2/2.5d* smoothing in the *Select map function* step under the *Simple map arithmetic* interview. Enter a value of 2 into the *Parameter* box. Press *Calculate* (top box), and you will see a smoothing of the values. Click on the *See the model step* momentarily to bring the original model back up, and then return to this step. Now enter a value of 4 in the *Parameter* box and press *Calculate*. The model will get smoother as more neighbors are included. Taken to the extreme, if you make N big enough to include all neighbors for each block, your map will simply equal the block average in every cell.

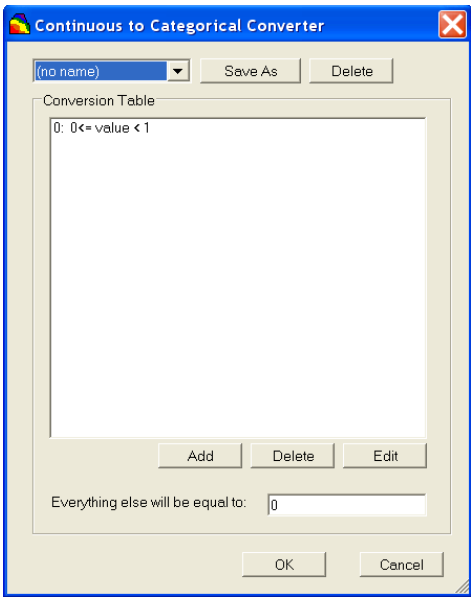
3d Smoothing

This is exactly the same as 2/2.5d smoothing except neighbors above and below each cell can participate.

Convert to Categories

This will let you convert any map into a set of categories. To do this, you must set up some conversion rules. In other words, you explain what will become what. For example, one simple conversion rule could say that all values between 0 and 5 get converted to 1 in the conversion process. To see how this works, let's try it now. Make sure you have *MyFunction* selected and press *See the model* to restore the original values.

Under the *Simple map arithmetic* interview, select *Select map function*, select *convert to categories* and press *Calculate*. You will be presented with this window.



At the top is a drop-down list with only one entry: (no name). SADA allows you to create a conversion rule set and save it so you can recall it later. At the moment, there are not any previously created conversion rules, so we are presented with only the default. Any time you wish to save your conversion rules, press *Save As* and give it a relevant name. Similarly, if you want to delete a rule set press *Delete*.

One or more conversion rules can be created in your conversion rule set. The box just below the *Save As* and *Delete* buttons displays the current conversion rule (created by SADA as a default). It means that any value greater than or equal to 0 and less than or equal to 1 will be converted to a value of 0.

Suppose we want to change that rule. Select the rule in the list box and then press the *Edit* button. Change the interval to 2 to 2.5 and the conversion value to 1.

Press *OK* and you will see your change updated in the *Conversion Table*. Now let's add another one. Click the *Add* button. Again you are presented with the same window. This time enter a value of 2.5 and 3 for the range values and a conversion value of 2. Press *OK*. Now your conversion table (rule set) contains two rules:

2<=value<2.5 = 1

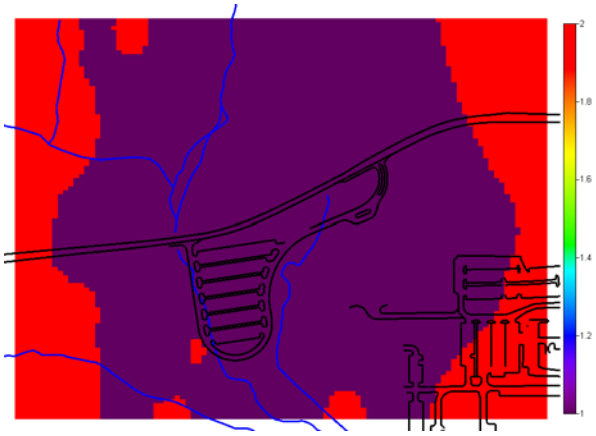
2.5<=value<3 = 2

You can easily delete a conversion rule by selecting it from the list and pressing *Delete*. Let's keep these two for now.

Below the *Conversion Table* is a "catch all" rule: *Everything else will be qual to:* This basically tells SADA that if you encounter any values in my model that are not accounted for by the conversion rules I've selected, use this value in the conversion process. Let's leave it as a zero for now.

Before we continue, let's save this rule set. Press *Save As…* and enter the name "MyRules." Press *OK* and you will now see it in the drop-down list of rule sets. Remember, nothing is really saved until you save your SADA file.

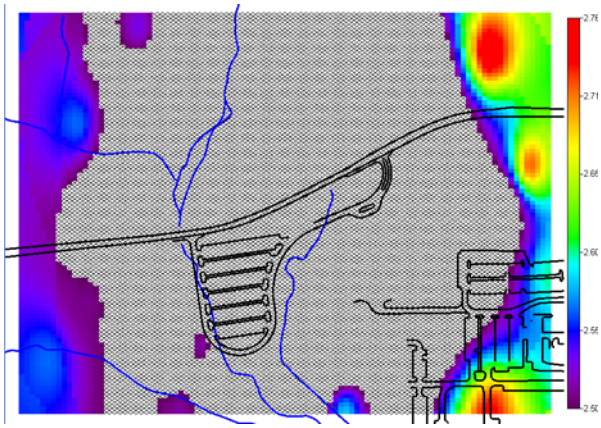
Now press the *OK* button and SADA will convert your map.



Depending on what legend you are using, the result may not look like what you would expect. You may want to work with the legend manager to arrive at a good color scheme.

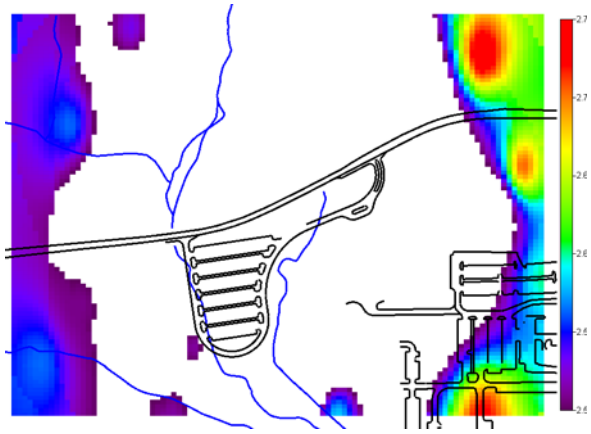
Cutaway Using Value Interval

This function allows you to cut away any values found in a certain range. First, click on the *See the model* step. Then come back to this step. Choose *cutaway using value interval* as your function. You will need to enter a lower and upper cutting threshold. Any value in your model/data that has a value in this range will be erased. Enter a value of 2 for the minimum and a value of 2.5 for the maximum in the lower *Threshold Cutting* parameter block. Press *Calculate* (top box).



SADA will look through your model results and erase any cell values that fall in this range. Press *OK* and look at the results in the graphics viewer. Everywhere there is an x, the cell has a missing value because it has been erased. Depending on what legend you have selected, your colors may look different.

Now is a good time to apply a graphics feature often used in SADA. Suppose we don't want to see the hash marks associated with missing values. You can control how SADA represents missing values by choosing *Graphics→Set Various Colors→ Unestimated Model Values→.* Some menu options appear for dealing with unestimated values. Select *Show with Hash Mark→No.*



SADA responds by eliminating any hash representations.

Double Map Functions

We now move to the double map functions found in the window under the heading *Two Map Functions*. The first model is the one that is selected and displayed in the graphics window. In the first list box, you'll choose the map you want to use as your second map for the operation. These include any model imported or stored in SADA. A key thing to know here is that before you can use a second map to operate on your primary map, they both must have exactly the same layering and grid system.

We will now briefly describe what these two map operations do. First let's refresh our viewer by clicking on the *See the model* step and returning to this step.

Add

Adds the currently selected second map to the map in your graphics viewer. Let's try adding MyContaminantPlume to itself. Select *Add* and press the *Calculate* button (lower box). You can see that the model has doubled, because it was added to itself.

Subtract

Subtracts the currently selected second map from the map in your graphics viewer.

Multiply

Each cell value in your second map is multiplied by the exact same cell in your current map to produce a final result.

Divide

Each cell value in your first map is divided by the exact same cell in your second map to produce a final result. Values of zero in the second map yield un-estimated values in the final result.

Cutaway with 2nd map.

This is analogous to cutting away by using value interval in the single map functions previously discussed. In this case, SADA compares each cell in the SECOND map to the threshold value range. If that value is in this range, then the cell in the FIRST map is erased. This principle applies as well to the Cutaway/Combine with 2nd map, and an example is used to demonstrate how it works.

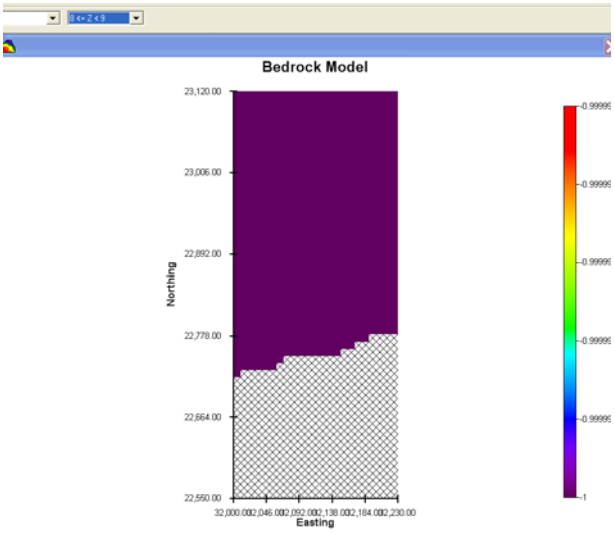
Cutaway/Combine with 2nd map

This feature works the same way as the *Cutaway with 2nd map* operation. The only difference is the following. If a cell value in the 2nd map falls inside the cutting threshold range, the cell value in the first map is replaced by the cell value in the 2nd map. The previous operation *Cutaway with 2nd map* would have simply erased the cell value in the first map.

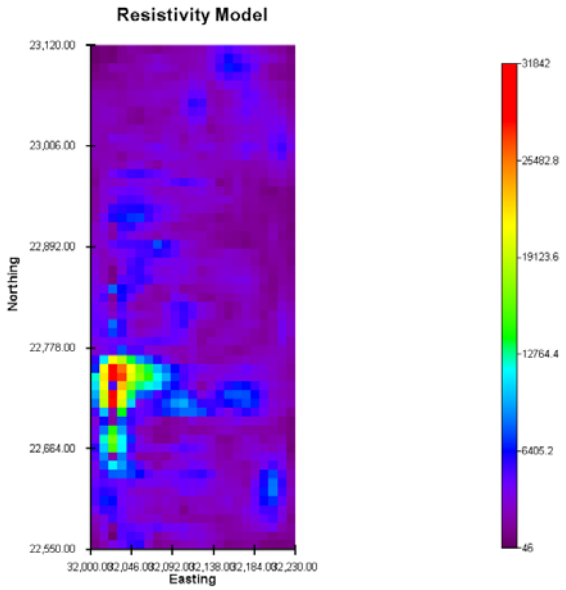
Let's see how this works. Open the file Resistivity_Bedrock.sda. This file contains two imported models: *Bedrock* and *Electrical Resistivity*. We will combine these two models into a single model showing both bedrock geology and geophysical resistivity measurements. Before we imported these two models, we had to make sure they had exactly the same boundaries, grid definition, and layering system. If they are not identical, the feature will not work.

Furthermore, we needed to somehow to distinguish between *Bedrock* and *Resistivity* when the two are merged. For the bedrock model, a value of -1 for any cell means that it is part of the bedrock. Resistivity measurements ranged from 0 to 31,000+.

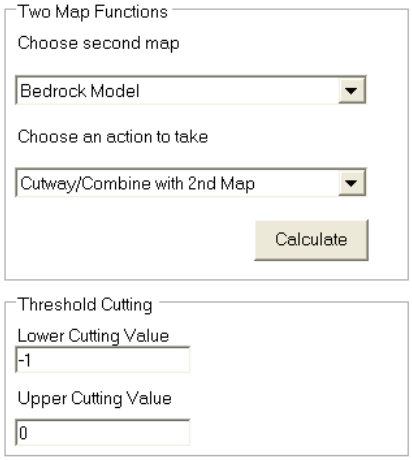
First, select the *Bedrock Model*. Notice that most cells have missing values. This is because they are not part of the bedrock. They are empty space, so to speak, ready to be filled in by the geophysical results. If you continue down through the layers, you will see a group of purple blocks. These are the bedrock.



Now switch to the *Resistivity Model* and take time to peruse through the layers. We will need to cut away the lower part of the *Resistivity Model* that overlaps with the *Bedrock Model*. We'll actually use the *Bedrock Model* to accomplish this.

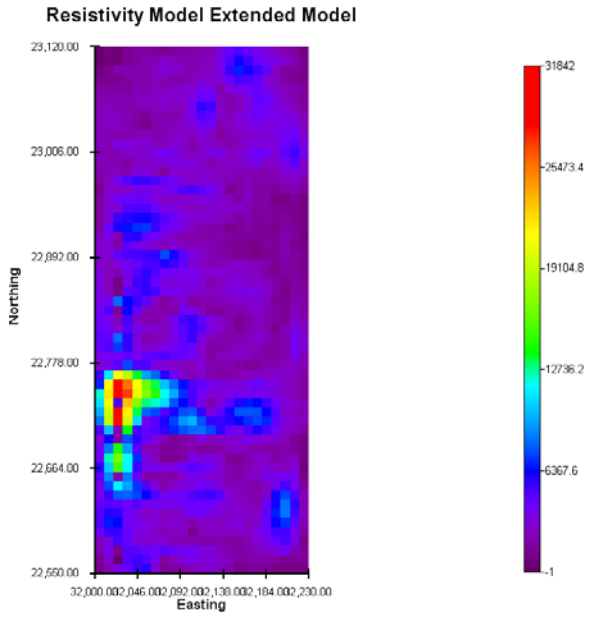


Make sure you have selected the model *Resistivity Model*. Click on *See the model* to refresh it. Then re-click the *Select map function* step. In the *Two Map Functions* box, select the *Bedrock Model* as the second map. In the list of actions, select *Cutaway/Combine with 2nd Map*.



The *Bedrock Model* defines the areas of 3d space that are bedrock and those that are some other soil type. If a cell has a value of -1, it means that the cell is part of the bedrock. If a cell has a missing value, it means that the cell lies above the bedrock. What we want to do is cut away the part of our resistivity model that is actually in the bedrock.

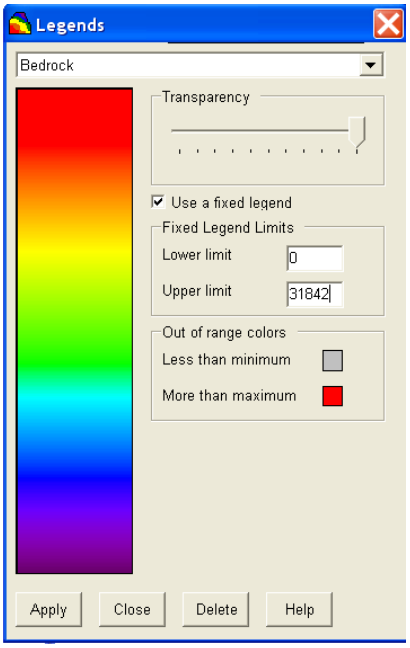
To do this, enter a value of -1 for the *Lower Cutting Value* and a value of 0 for the Upper Cutting Value. What this means is that wherever the second map has a value of -1 (bedrock) we will replace the cell value in the *Resistivity Model* with the value for bedrock. Note that when you do this kind of map cutting, you want the second map to contain values that would not appear in your second model. For example, if the *Resistivity Model* contained cells with a value of -1 before we did this operation, it might not be clear afterwards if something was bedrock or was really a resistivity value of -1. Press *Calculate*. SADA responds with a *Resistivity Extended Model*. This means that the resistivity values have been affected by a post-modeling operation (such as combining with bedrock).



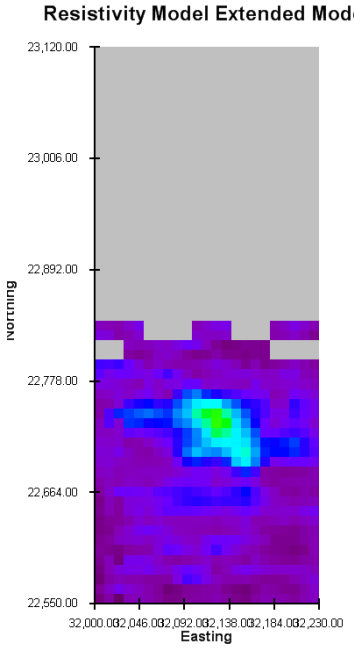
Notice that the legend now runs down to -1. This means our merge was successful. Let's first save the model as R_Bedrock. Click on *Manage model results* and click *Save model result*. Enter "R_Bedrock" into the *Model Name* and press *OK*. After SADA returns control to you, select *R_Bedrock.*

In this application, we want to distinguish bedrock (-1) from the rest of the model. To do that, we'll use a fixed continuous legend. Details about creating and managing legends are left to a later chapter (Chapter 43); however, we will briefly show some details here. From the main menu, select *Graphics→Legend Manager*. In the legend manager window that appears, select *New Color Palette* and enter the name "Bedrock" into the resulting window. Press *OK*.

In the legend manager, select *Bedrock* and make it a fixed legend with a lower limit of zero and color of gray for anything less than zero. This will make our bedrock values look gray. Make sure your legend manager looks like the following image, then press *Apply* and *Close*.



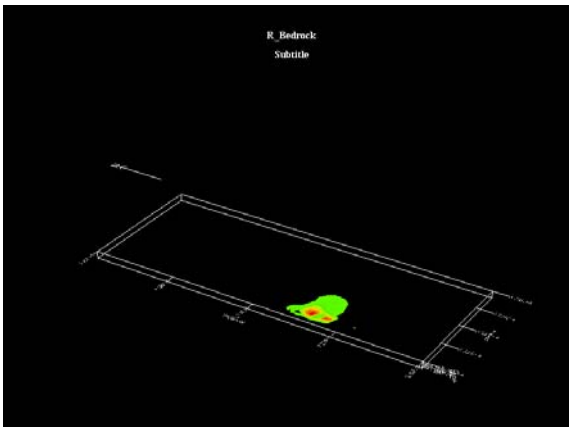
Switch now to Layer 7<z<8, and you can see the bedrock (gray values) combined with resistivity values.



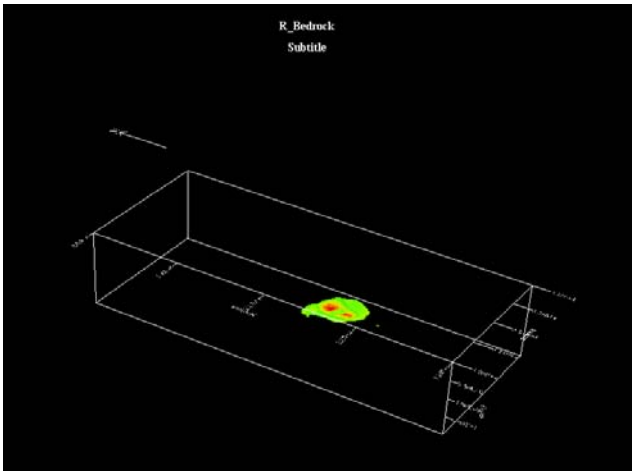
The ability to combine maps relies on the fact that there are two legend values outside the range of your data that can be used as flags for up to two other types of information. A typical example might be geology, as we presented here. You may also, however, be combining maps of like measurements, and so the full color range is appropriate for the final map as well.

Let's continue the discussion by using the 3d viewer. Select *Graphics→Show 3d view*. Details of the 3d viewer are left to a later chapter (Chapter 44); however, we will cover a few features we need right now. You will need to be patient with 3d rendering. It can take some time, particularly with 3d models. You should see something like the following image. If it does not appear after

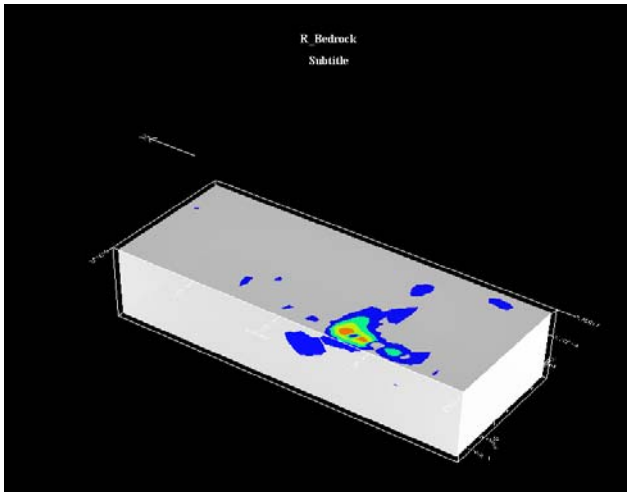
some time, place your mouse in the center of the graphics window, click the left-mouse button, and while holding the button down move the mouse around. The graphic should appear (this is often due to a video card anomaly).



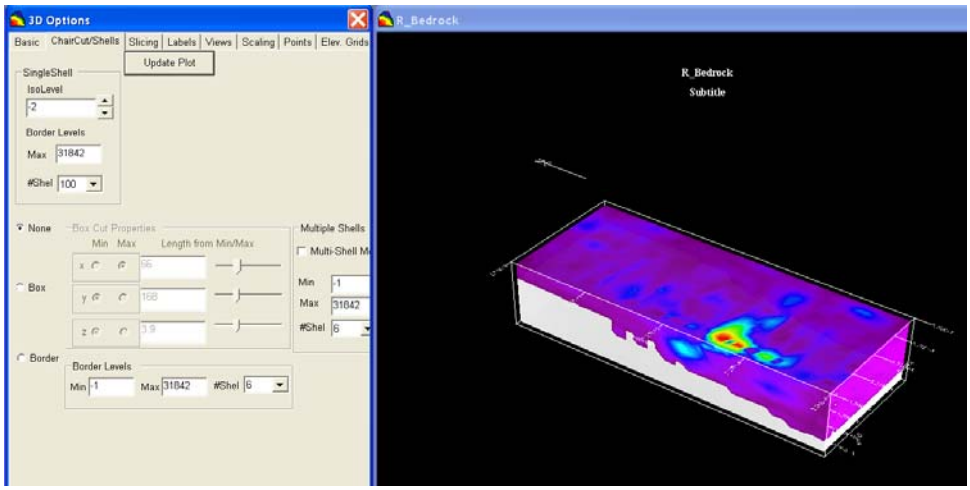
Let's first increase the vertical scaling so we can see the depth better. Click on the *3d viewer controls* step which just appeared in your steps window. In the parameter window, click on the Scaling tab, enter a value of 10 for the z scaling and press Enter on your keyboard.



Now we will set the isosurface value to a number well below any of our R_Bedrock model values. This will ensure that the entire volume is rendered. Click on the *ChairCut/Shells* tab and enter a value of -2 into the *IsoLevel* text box and press *Update Plot*.

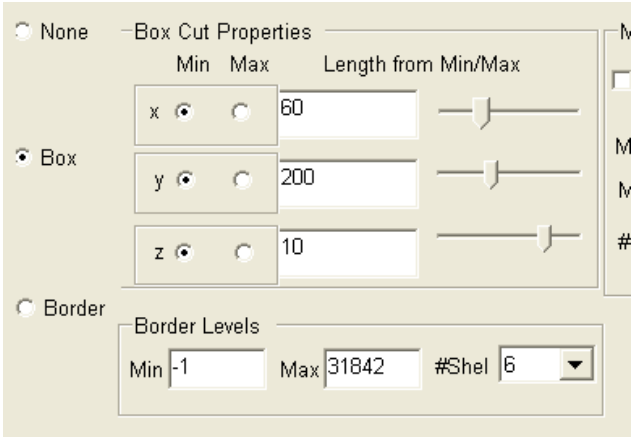


We see a lot of gray in the result, which seems to indicate that bedrock is everywhere. This is simply because our isosurface shell resolution is too big. Let's increase the number of shells from 6 to 100 and press *Update Plot* again.

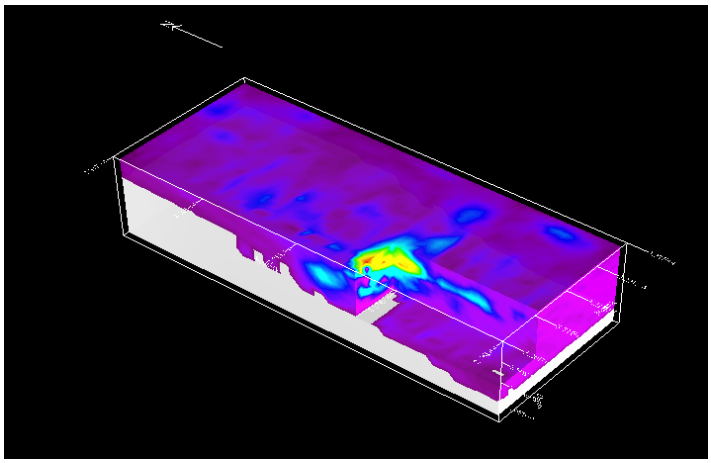


Now we can see the bedrock very clearly. Let's do a quick chair cut stemming from the origin.

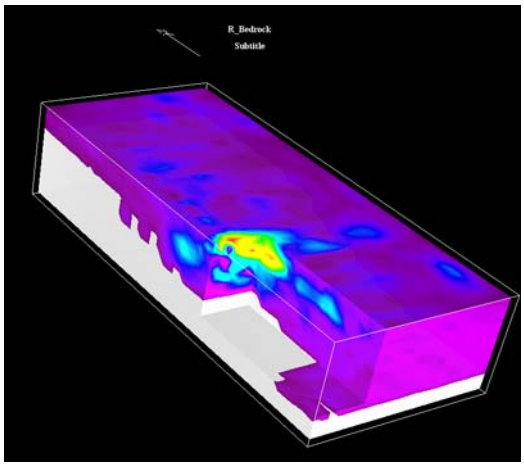
In the *ChairCut/Shells* tab, select the *Box* option. For x,y, and z directions, select *Min*, and enter values of 60, 200, 10, respectively.



Press *Update Plot*. You can see a cutaway in the lower-left-hand corner of the site that reveals the inside of the volume.



Use your zoom feature (*Basic* tab) and rotation feature (mouse-click, hold, and move) to see the result better.



Chapter 33: User Models

In some situations, you may want to manually create a 2d or 3d model by hand. For example, you may want to spatially delineate a geological representation of the subsurface or estimate plume boundary based on professional judgment. You may also want to express prior knowledge about the site before an initial sampling design. In particular, you may want to spatially assign areas of greater concern. This kind of professional, expertise-driven construct is sometimes done as a normal course of any environmental investigation. Later, in the MARSSIM chapter (Chapter 40), you'll divide areas of the site into subclass regions according to the likelihood they may be contaminated.

SADA permits you to express spatially relevant information or *user models* in two and three dimensions by providing some basic drawing tools. You can then use these manually created models directly or use some of the model algebra methods discussed in the previous chapter. You can also use these results to drive target initial sample designs (Chapter 41). These types of designs distribute initial samples not according to some statistical endpoint but rather to a targeted end point driven by search or cost objectives.

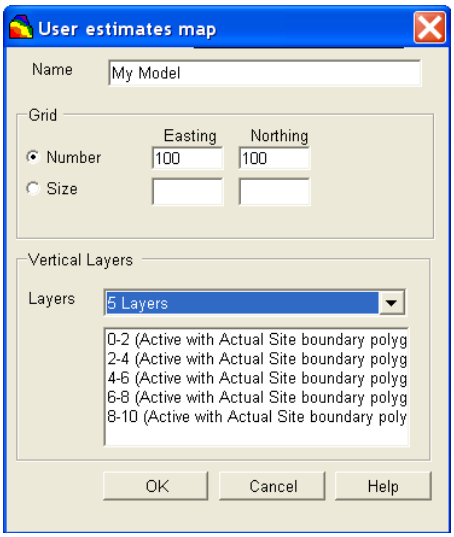
There are two kinds of user models: standard and probabilistic. A standard user model allows you to distribute any kind of values in model space. You can "paint" in 2d and 3d space any type of values you wish. These values may represent concentration values or perhaps categorical values such as soil type. The possibilities are endless. A probabilistic user model spatially expresses the probability of something being true or false. This could be the probability that a decision criterion would be exceeded. It could also be the probability that a contaminant is present. Again the possibilities are endless.

We will show you how to create a user model and provide an example of how to continue with the model using map algebra methods previously discussed. The chapter on targeted initial designs (Chapter 41) will take up this discussion again and demonstrate how to distribute samples according to less statistical or classical objectives. Both user models and standard models are created in exactly the same way. The majority of the discussion will be found in the standard user model section.

Standard User Model

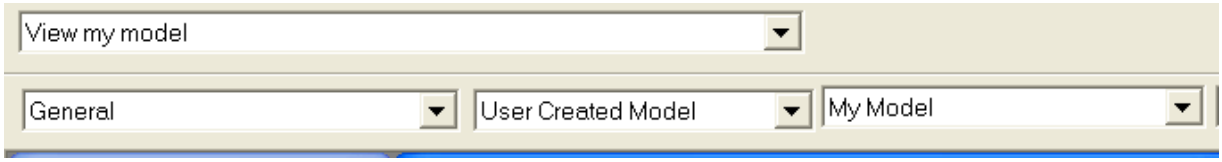
Open the file UserModels.sda. This SADA file has no data in it, but the GIS layers have been imported and the site boundaries have been drawn (under *Setup the site*). We will be drawing a 3d model, so we've created a vertical layering scheme called "5 Layers". Each layer in this layering design is two-feet thick.

To create a default user model, select *Data→Create Estimates Map*. This begins the process by which a user model will be applied to the currently set site boundaries. SADA needs you to do two things at this point. First, provide a name for your model (enter "My Model").

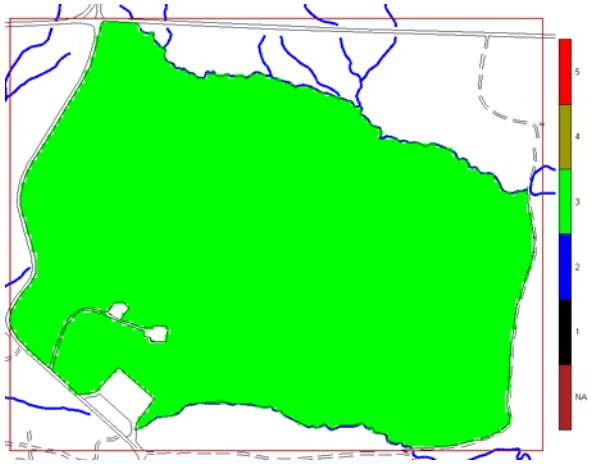


Next, we'll set the grid resolution. Let's enter 100 x 100 grid cells by selecting the *Number* option and entering 100 and 100 for both *Easting* and *Northing*. This will divide the site boundary box into 100 x 100 cells. Cell width and height may be different, depending on whether the site boundary box is a square or rectangle. Finally, choose the layering design you want to use. In this case, select *5 Layers* and press the *OK* button. SADA will produce a model with a default value of 3.

First, notice where the model is placed. A user model is another data type called *User Created Model*. These kinds of models are different than imported models because they can be edited.



You can see from the resulting image that a polygon boundary has been established within the boundary box. *Hint: if you don't see your brown site boundary box, click on the Set GIS overlays step and deselect Hide site boundary box*

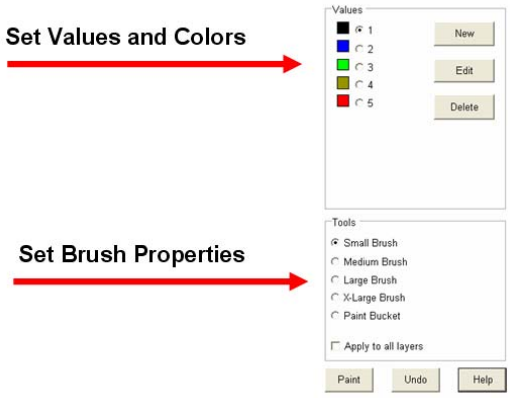


Note: if you don't see the same legend as the image above, select *Graphics→Legend Manager*, select *My Model*, and press *Apply*. Save your file.

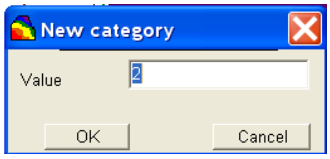
The structure of this model is permanently set in the sense that you can no longer change the grid resolution or vertical layering design. Click on *Set up the site* and note that the layering design box is now uneditable. We will now modify this model.

Customizing model values

All model editing features are found under *Edit my model*. Click on this step now.



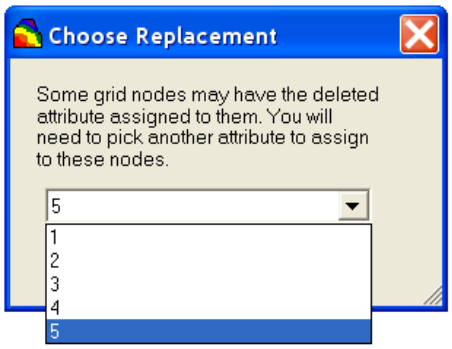
At the top of the parameter window is where you will set the values and colors you will "paint." When you use your painting tools, you are actually adding values to the model and not just color. The default values for standard estimates are just 1, 2, 3, 4, and 5. You can change these to anything you like. Let's customize these values now. Click on the 2 value (in blue) and click the *Edit* button to the right.



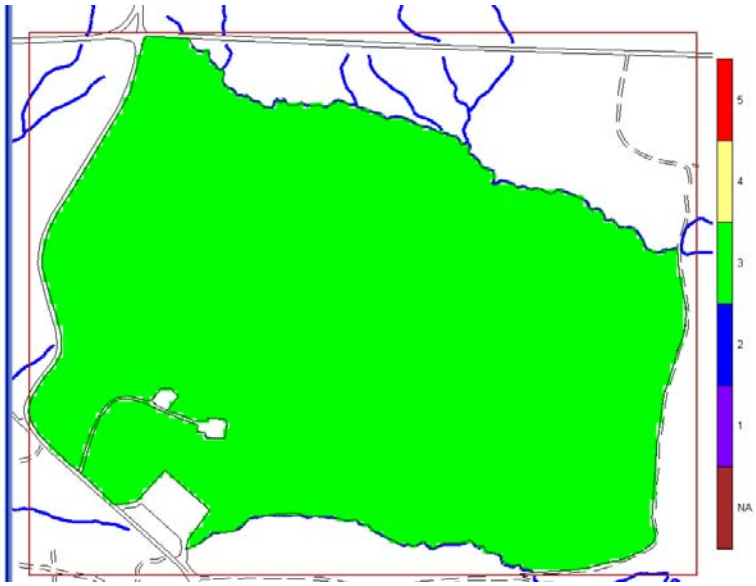
Let's enter a value of 6 here and press *OK*. IMPORTANT! If you have previously "painted" any 2's into your model, they all just got updated to 6's. SADA requires that all values in the model have a corresponding "paint" value in the parameter window. Notice that the values rearrange themselves in the paint selection ordering. Also recall that SADA searched the grid for all 2's and replaced them with 6's. In this case, it wouldn't have found any because we did not add any yet.

Now, let's change the color for the value 4. Click on the brown box just to the left of 4. The color palette will appear. Select a yellow color. Save your file.

Now, let's add a new value to the list. Press the *New* button and enter a value of 2 once again so we have an even 1-6 range. Then, delete the 6 value to restore us to where we began. Click on the 6 value option and press *Delete*. SADA will ask you for confirmation. Say Yes. Before SADA can delete a value, it must know what value you wish to use as a replacement. Every model cell must have a value. Select the value 5 from the drop-down list and press *OK*. SADA will search for any cells with a value of 6 and replace them with a value of 5. Then the 6 option will be removed from the list.

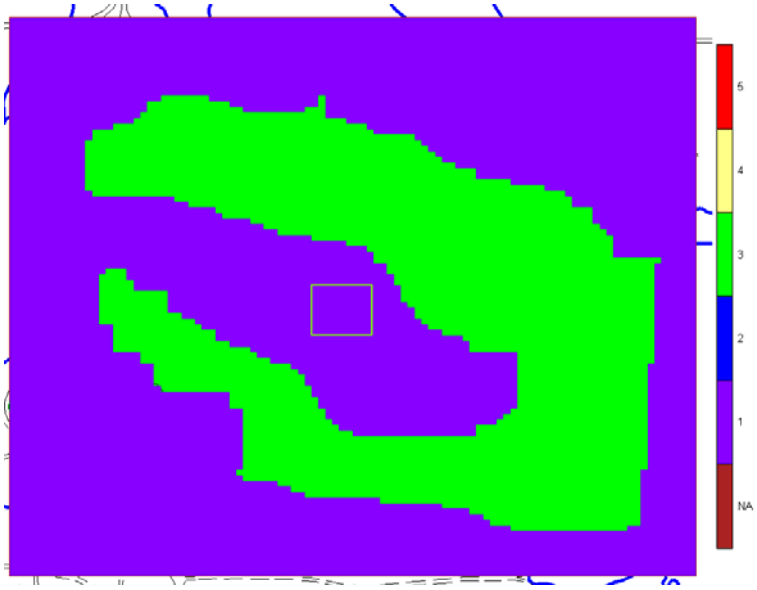


If you want to add a missing value, then you can use -1E+20 as a paint value. SADA uses this value throughout its analysis as a missing value. As a final practice, let's change the color for the 1 value from black to purple. Your screen should look like this.

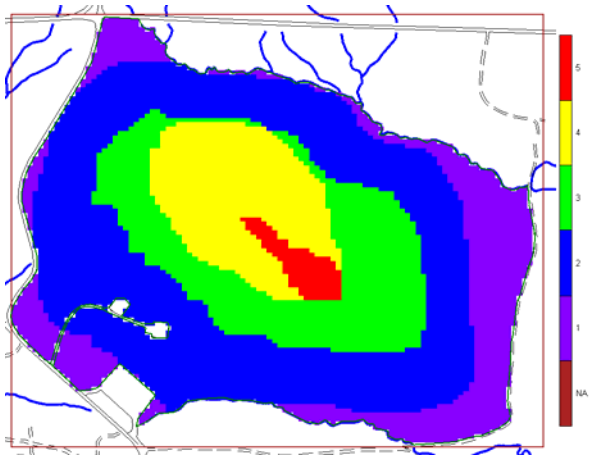


Let's move on to setting the brush size. In the lower parameter window you'll see the *Tools* parameter block. This includes the paint brush sizes as well as the paint bucket. These operate like most Windows paint tools. We'll try each of them out. At the bottom is an *Apply to all layers* option. If this is selected, then any painting done will apply to all layers directly above and below, so long as it remains selected. If you deselect it, painting only affects the layer you are currently on. Let's not select this option now. That way we'll paint on layers individually.

Click on the *X-Large Brush* option, select *1* (purple) in the top parameter block, and press *Paint*. Move your mouse over to the graphics window and a target window the size of your brush will appear. You can paint anywhere you like inside the site boundary box, but if you paint in an exclusionary polygon or outside an inclusionary polygon, SADA will apply these rules when painting is complete. To paint, left-mouse-click and hold while dragging your mouse around. In the image below, we've painted a *1* value around the edge of the site (outside the polygon boundary) and some in the middle.



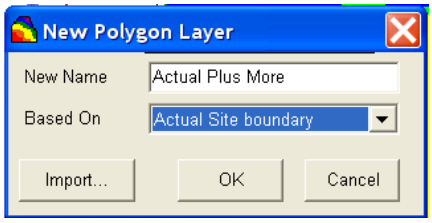
When you are done with 1, switch to *2*; choose a smaller brush, and paint some more. Play around with painting on this layer. When you are done, press the *Done* button. SADA will apply any polygon rules necessary and restore the interface.



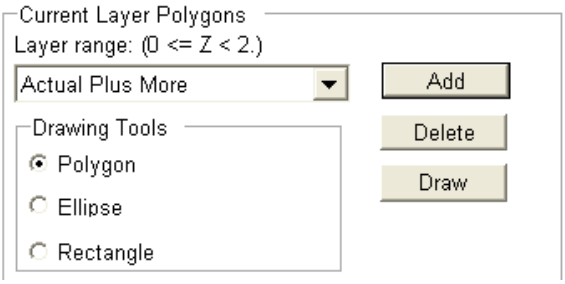
Save your SADA file and take some time to scroll through the layers. Notice that on deeper layers, the default values of 3 remain intact. This time, select the *Apply to all layers* option and paint a small portion of your site in any value you wish. When you are finished, press the *Done* button and notice that now changes are found through every layer.

Let's work with the paint bucket now. This feature works hand in hand with polygons by filling the interior of polygons you've created. Let's start by creating a copy of the current polygon design. We'll then add some smaller polygon features to it and fill them with the paint bucket or *autofill* them.

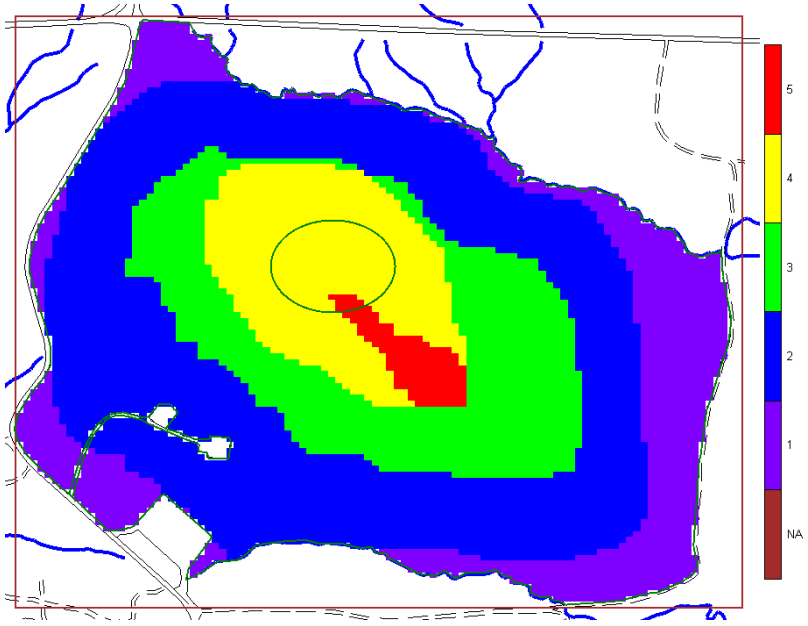
Select the *Set up the site* step. In the *Current Layer Polygons* parameter block, you'll see we've selected *Actual Site boundary*. Click the *Add* button next to it. Let's create a new polygon design called "Actual Plus More" that will be an exact copy of the *Actual Site boundary* design. In the window that appears, enter "Actual Plus More" into the *New Name* box and select *Actual Site boundary* from the *Based On* drop-down list.



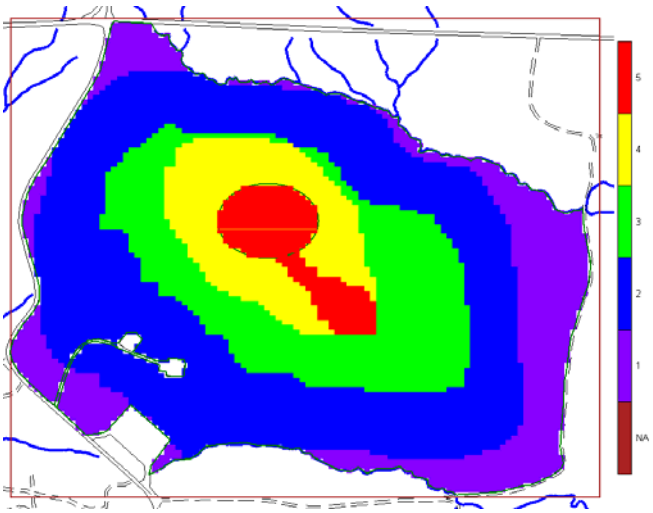
Press the *OK* button. When SADA asks if you want to apply it to all layers, say *YES*. In the *Current Layer Polygons* box you should see *Actual Plus More* selected.



Suppose now that we want to paint an elliptical area near the center of the site. Let's first draw that area by clicking *Draw* and selecting *Ellipse* as the drawing tool (polygon methods were discussed in Chapter 4). Move your mouse over the center of the site, left-mouse-click and drag to open the ellipse up. Then return and press the *Done* button (formerly *Draw*). Your result may differ from the following image and that's *OK*.



Now we'll use the paint bucket. In the *Values* palette on the step *Edit my model*, select *5*, then select the *Paint Bucket* option under the *Tools* palette, and press *Paint*. Now simply click in the center of the ellipse you just drew.



You could add more small polygon features and autofill them as you like. When you are done, simply return (if you wish) to the *Actual Site boundary* polygon.

SADA does provide you with an *Undo* button. Anytime you make a mistake, simply press Undo. You can step backwards many times. It is advisable, though, that you save your SADA file often.

Probabilistic User Model

The approach for setting up and customizing a probabilistic user model is exactly the same as a standard user model with only one exception. In a probabilistic user model, you can only paint values into your model Value between 0 and 1. As you might guess from the name, this type of model is intended to paint values into the model that spatially indicate the probability of some event occurring (or not). Examples include: the probability that the decision criterion has been exceeded, the probability that the soil type is clay, or the probability that the water table reaches that point. Many other applications are also possible. This type of model can be used later in the Bayesian Ellipgrid model or as a secondary form of information in cokriging.

Take this time to practice setting up and customizing the probabilistic user model on your own. Use the previous discussion as your guide. To recap, you will find the following steps useful.

1. Set up the model (*Data→Create Probability Map*) provide a name, grid resolution, and layer design.
2. Switch to your default model (switch data type to *User Created Model*)
3. Set up your paint values (*Edit my model* step)
4. Choose a paint tool (*Edit my model* step)
5. Decide if you want to paint all layers at once (*Edit my model* step)
6. Customize your model (*Paint*)
7. Save your File (*File→Save*)

Of course, in your own application, you will begin first by setting up the site, defining the site boundary box, and applying any polygons, ellipses, or rectangles to more accurately set your boundaries. These have already been completed here.

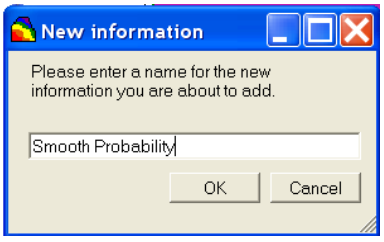
Post Processing User Models

Recall from Chapter 32, “Working with Models,” that you can apply a number of post-processing algorithms and manipulations to a static copy model. We can do that with user models as well. We’ll demonstrate how to prepare your user model for post-processing and provide a simple example you might find useful in your applications.

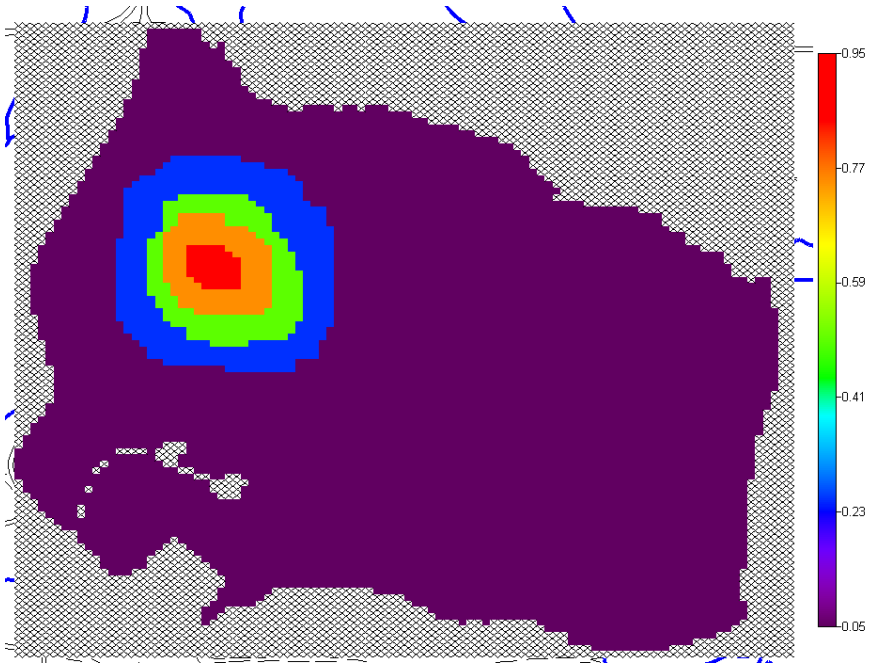
Open the file PostUseModels.sda (you can save your previous work if you like). This file has the same site boundary as the last file and a user created model called Prob > Decision. We’ll use this three dimensional probabilistic model of a hypothetical source term (groundwater source) location to demonstrate. If you scan through the 5 layers, you can see that previous geophysical measurements indicate a possible source term location. This could have been based on an interpretation of the geophysical results, for example.

Because user models only allow up to 10 different paint values, many of them will look coarse, particularly along categorical boundaries. This may be an asset in some situations (such as defining a survey unit) or a hindrance (as in producing a reasonable prior model of a more continuous nature). As a demonstration of one type of post-processing that can be done, we will show you how to smooth your custom values to create a more continuous model.

We will first need to store a static copy of the model, much like we did with the geospatial models generated in previous chapters. Click on the step *Manage model results* and press *Save model result* in the parameter window. In the new name window enter “Smooth Probability” and press *OK*.



Then, select *Imported Model* from the drop-down list of available data types. You will see a result that looks something like what you did but not exactly. Don’t worry; numerically it is exactly the same. The reason it looks different is SADA has made a break with the user model legend you’ve selected and also dropped the polygons that define the boundary momentarily. This is because the now-static model is a new source of data in its own right and has not been fully customized (from SADA’s perspective). This is only a minor inconvenience that we can correct quickly.

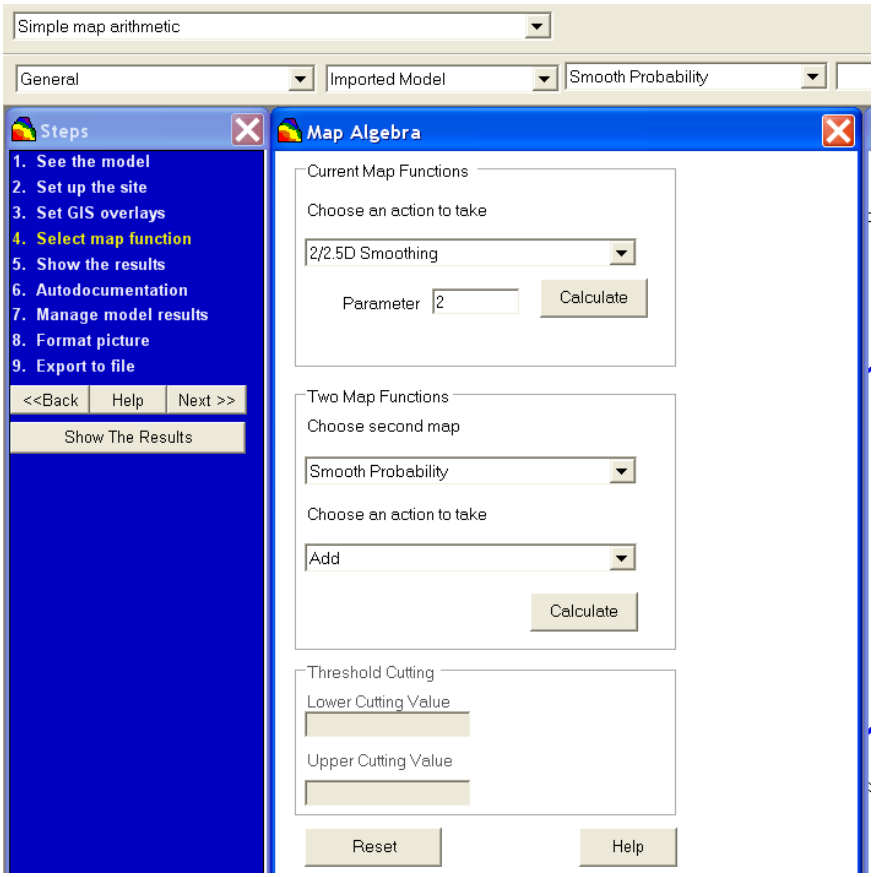


First, let's turn the polygon layer back on. Click on the *Set up the site* step and in the current polygon layer parameter block, switch to *Actual Site boundary* (answer *Yes* to the *apply to all layers* question).

Let's fix the legend next. Click on *Graphics→Legend Manager* and choose *Prob>Decision*. Press *Apply*. The static version of your model should now look exactly the same as before. We want to be careful here though. The legend *Prob>Decision* is dynamically connected to the *User Model* and may change in the future should you make adjustments. If you want to keep this exact legend, then you'll need to create a new legend based on this *Prob>Decision* legend. If you like, you can do this now (recall how from previous chapters or Chapter 43).

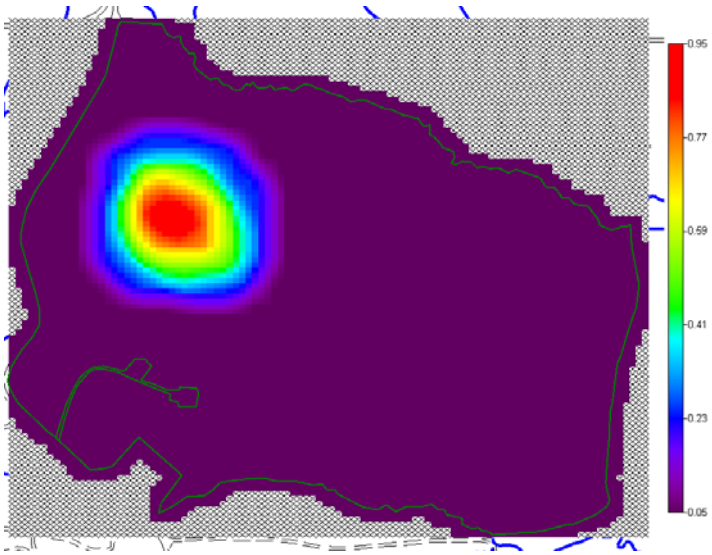
We will not use this approach here, however. In this situation, we wish to smooth the values to create a more continuous model. We want to then adopt a continuous and not a categorical legend. Moreover, we want to adopt a fixed continuous legend than extends from 0 to 1. We'll do that shortly. First let's smooth our model. (You can revisit Chapter 32 to review the rich set of post-processing features there.)

In the interview box, select *Simple map arithmetic* and click on *Select map functions*.



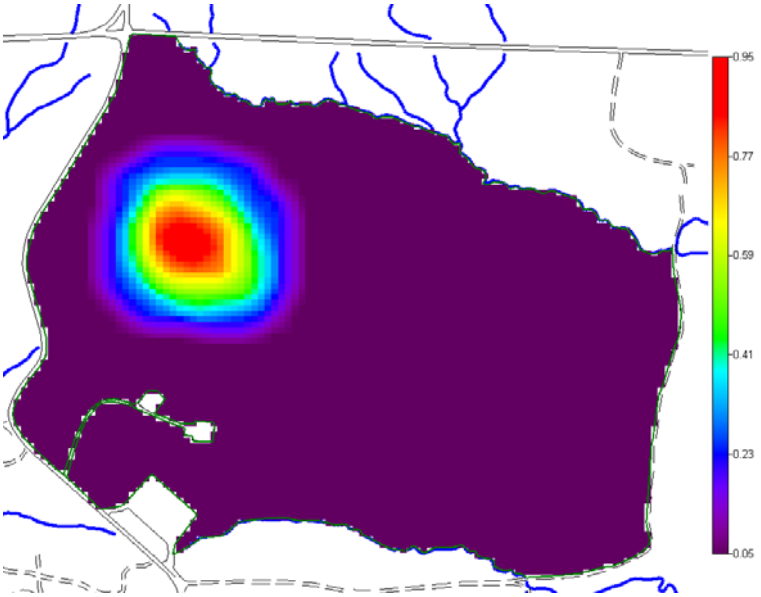
In the top parameter block, choose *2/2.5D Smoothing*. For the *Parameter*, choose a value of *2*. Recall from Chapter 32, "Working with Models," that 2/2.5D smoothing only smoothes by the layer (not between layers) and a parameter value of 2 means each cell value is smoothed by using the next two neighbors on all sides of it. You can use the 3d smoothing feature here as well, but be advised that the smoothing parameter will apply in all three directions. For course layering systems like this (only 5 layers), sometimes you can "average out" vertically. Let's apply the 2/2.5D smoothing now by pressing *Calculate* in the top parameter window.

SADA responds with the following image:

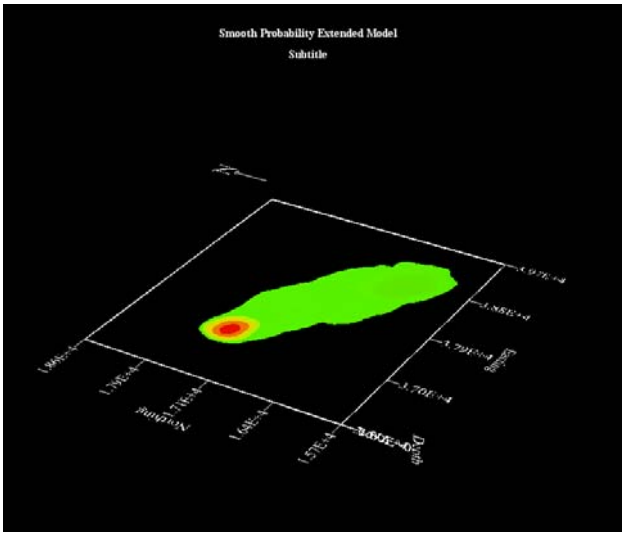


Let's now scan through the different layers and you can see that we have captured the source term as it dives down and to the southeast. Let's save this model as well. At the moment, it's only a temporary result and will be lost otherwise. Click on *Manage model results* and choose *Save Model result* button. Enter the name "Smoothed Model" and press *OK*.

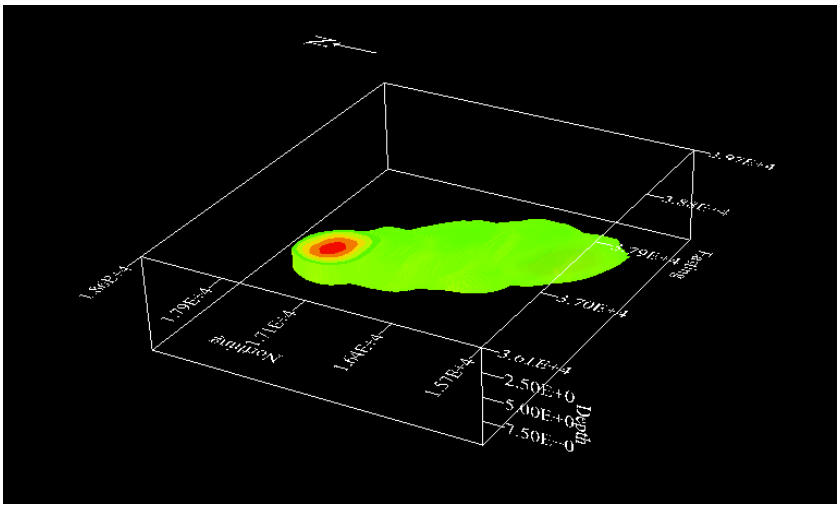
Next, select *Imported Model* from the drop-down list of data types, and select *Smoothed Model* from your data set list. Add the polygon layer *Actual Site Boundaries* to this new data set. You should see the following image.



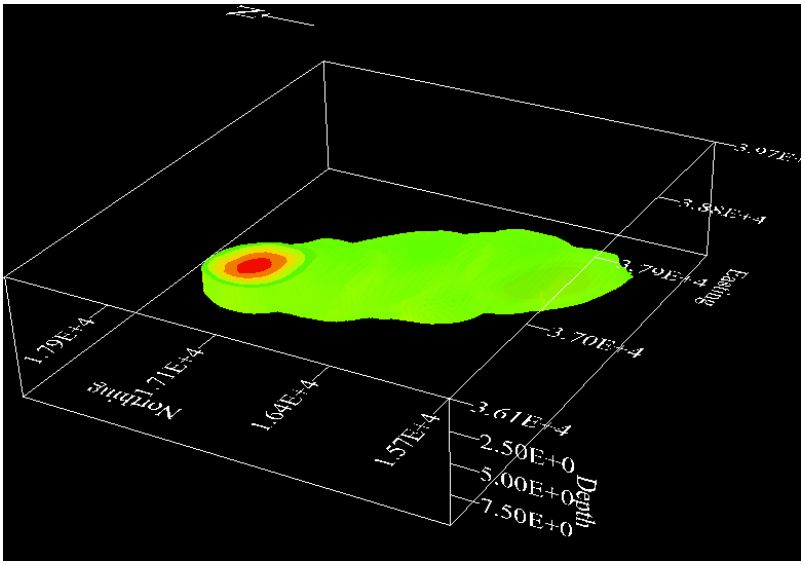
Save your SADA file (*File→Save*). Now let's look at it in 3d. The 3d viewer is discussed in its own chapter (Chapter 44) later, but we can use some of its features right here. Select *Graphics→Show 3d View*. 3d models take some time to generate, so be patient. Also, some video cards have difficulty with *OpenGL* (in our experience), so you may need to "wake up" the view by left-mouse-clicking in the graphics window, holding the button down, and slightly moving your mouse. You will be presented with the following image:



The ratio of the site depth to the horizontal extent is very, very small. For this reason, the true rendering looks very flat. This is not helpful. Let's change the z scale. Click on the *3D Viewer Controls* step that was added when you turned on the 3d viewer. In the parameter window, click on the *Scaling* tab and enter a value of 100 into the Z parameter and press Enter on your keyboard.

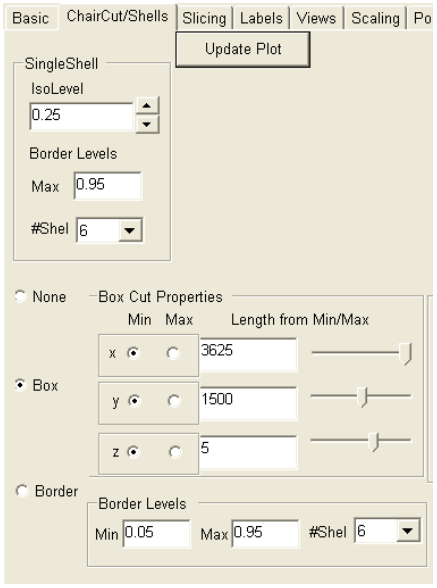


We can zoom in some by clicking on the *Basic* tab and pressing the *Zoom In* button repeatedly.

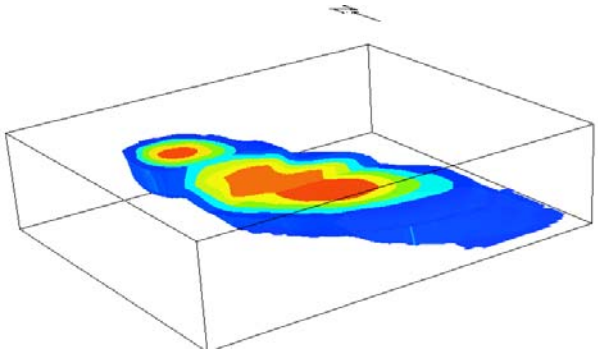


You can rotate this view by left-mouse-clicking anywhere on the graphics screen, holding the mouse button down, and moving the mouse around. You can get rid of the axis tick marks by right-mouse-clicking on each axis.

What we are looking at here is the volume for all probabilities greater than 0.5. Let's change that to 0.25 and perform a chair cut (see chapter 44) and make the background color white. Click on the *Basic* tab. Click on the *Bkgd color* tab and choose *white*. Click on the *ChairCut/Shells* tab and make sure your parameters look like this.



Press *Update Plot* and then rotate your volume around to see inside the prior model.



When you are done, save your SADA file and close it.

Converting Any Model to an Editable Model

Some applications may require the addition of professional knowledge to the model. This may require some editing of numerically derived values. At the time this section was written, SADA allows you to edit a numerically based model by first converting it to a user-defined model. There is a limitation. The numerically based model can have no more than 10 distinct values. This is an unlikely situation. In order to get only 10 distinct values, you will need to apply some of the post-processing tools (e.g., convert to categories) to the static model first and then convert it to an editable model. To convert any static model (found under *Imported model* data type), simply choose the *Manage model results* step and press the *Convert to Editable model*. If you have more than 10 different model values, SADA will warn you. Once it has been stored as an editable model (now located under *User Created Models* data type), you can edit it.

Another possibility exists if you need to retain your continuous values. You can create a user-defined model separately. In this user-defined model, you will spatially paint in the model values you need to add to the continuous model. When you are done, store this user-defined model as we did in the previous discussion. Now you have two models: 1) your user defined model stored as a static model and 2) the numerically generated model. Use the *Simple map arithmetic* interview and the *Two Map Functions* list to merge the two together. Please visit Chapter 32 for details on different two-map functions available to you.

Updating a User Model with Real Data

Probabilistic user models offer a highly specialized feature in SADA. It is possible, with a probabilistic model, to update the map of probabilities when real data become available. This is done using the Markov-Bayes geostatistical model (Goovaerts 1997). Like all geostatistical models, a correlation model is required for the update. In this case, a correlation model is derived for the update data. The exact nature of the correlation model was discussed in the chapter on Advanced Geospatial Methods (Chapter 30). A portion of that discussion is repeated here.

Let 1 represent the sample data you wish to interpolate, where values have been transformed to zero if the measured value is less than or equal to the decision criteria, and 1 otherwise. So the data you wish to interpolate is actually point data or measured locations. In the *Correlation modeling* step, you develop the correlation model, γ_I, for the indicator transformed data set. Then borrowing from the book *Geostatistics for Natural Resources* (Goovaerts 1997), we have that

$$\gamma_Y(0) = |B|\gamma_I(0)$$

$$\gamma_Y(h) = B^2\gamma_I(h) \quad \forall h > 0$$

where Y is the soft prior model data (i.e., the probabilistic user model). Behind the scenes, we calculate B as follows. B is defined as the difference between two conditional expectations:

$$m^1 = E[Y(u) \mid I(u) = 1] \in [0,1]$$

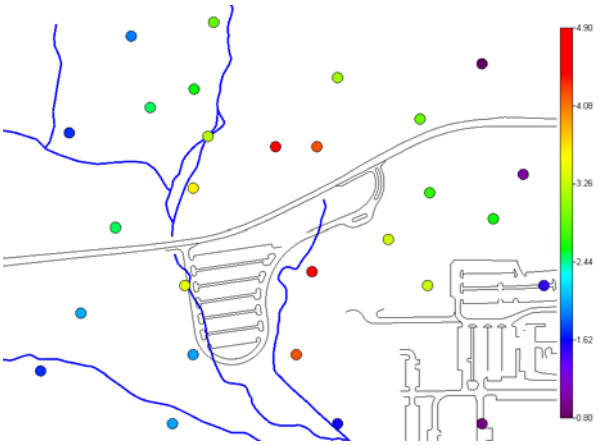
$$m^0 = E[Y(u) \mid I(u) = 0] \in [0,1]$$

$$B = m^1 - m^0$$

While all this may seem fairly intimidating, what it means to you is really fairly simple. All you have to do is establish an indicator correlation model for your point data and the rest is just like an ordinary kriging model from your perspective. Let's try one.

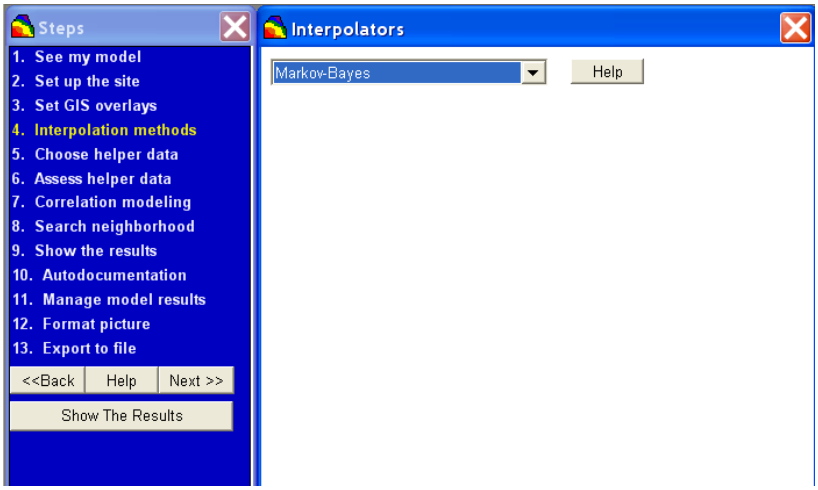
Open the file UpdateMap_MarkovBayes.sda. In this file, we play out the following scenario. First, you created a prior probability map that reflected some knowledge you had about the site gained from one or more lines of evidence. In this case, the map was drawn as the probability that the contamination might exceed 3pCi/g, a locally established screening value. This prior map is found under the data type *User Created Model* and is simply called *Prior Map*. If you select these now, you should see the following:

Based on this map (or not), samples of Ac-225 were collected. These can be seen by changing the data type to *Soil* and the dataset to *Ac-225*. If you do this now you should see the following image:

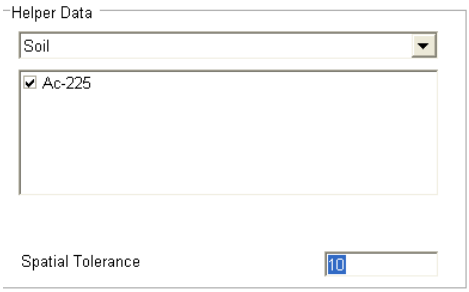


What we want to do now is update our “state of knowledge” represented by the *Prior Map* with more information provided to us in the Ac-225 measurements. In this sense, it is a Bayesian-type update of the prior map (although Bayes theorem is not explicitly used). What we will get in the end represents our new state of knowledge that now includes hard measurement values. This kind of approach is novel and may have tremendous applications in areas where data collection is necessarily sparse due to cost or accessibility. Let’s now update the *Prior Map* with this new dataset.

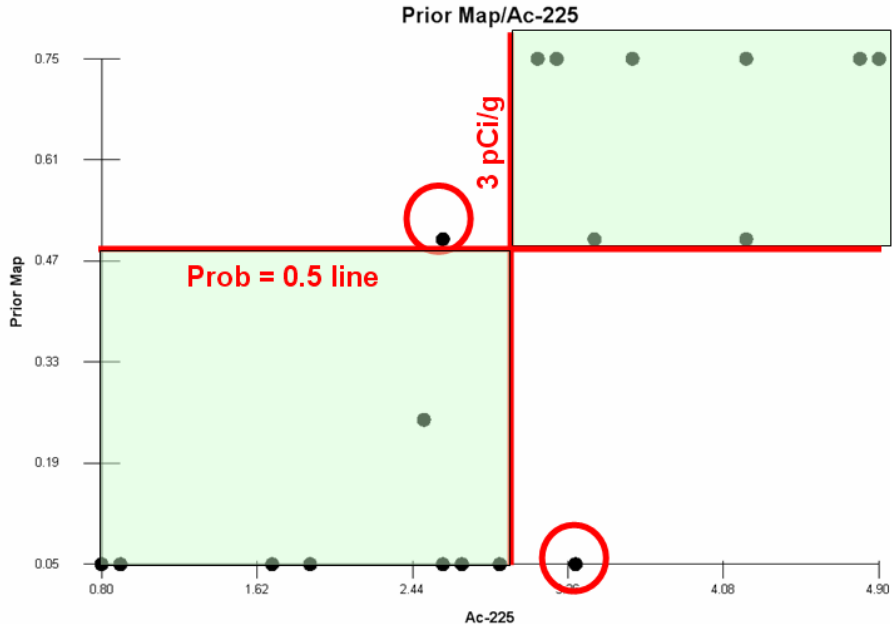
Switch back to the *Prior Map*. Select *User Created Model* as the data type and *Prior Map* as the dataset. In the interview box, select *Update my probability map*. If you’ve studied the chapters on geospatial modeling, the steps that show up for map updating should look familiar. Click on the step *Interpolation methods* and notice that the only available method now is in fact *Markov-Bayes.*



We will next need to choose the dataset we want to use to update the *Prior Map*. This is done by clicking on the *Choose helper data* step. In the parameter block at the bottom of the parameter window, we’ve selected *Soil* and *Ac-225* as our helper data set. We’ve also provided 10 feet of slack, as it is unlikely that nodes from the *Prior Map* and data from the *Ac-225* will exactly coincide.

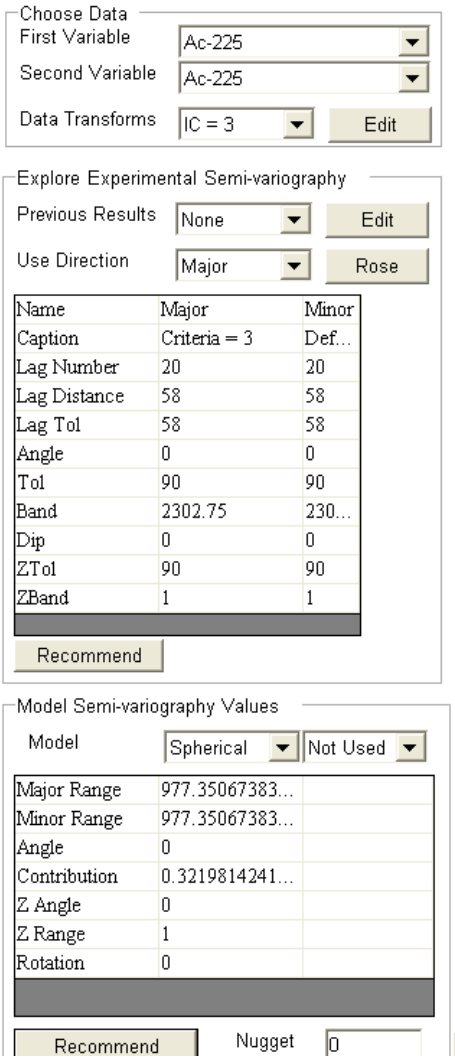


Next, click on *Assess helper data*. Under usual cokriging is where you would calculate Spearman Rank and Pearson's R and see a scatter plot of primary and helper data sets. In this case, the meaning of these two correlation values becomes ambiguous. As the time this book was written, SADA was not first converting the data into 1s and 0s based on a value of 3pCi/g before displaying the plot. It is doing this conversion in the background in all other cases. What would be ideal would be a scatter plot of Ac-225 data as 0s and 1s against the probability values; however, the result is still quite useable. Press the *Show Me* button in the parameter window. SADA produces a scatter plot, which has been further annotated in the following graph.

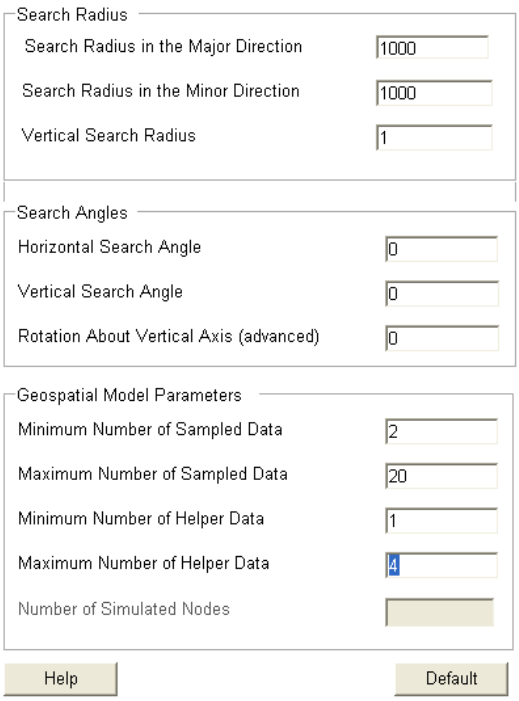


In this graph, Ac-225 is on the horizontal axis and a red line represents our decision criteria of 3pCi/g plotted vertically from X=3. On the Y axis are our prior map values and a line extending from the probability = 0.5 line. This forms a set of quadrants. Ideally, if our prior probability map is good in any sense of the word, then you would like to see all your scatter plot points falling into one of the green quadrants. This means that when we said the probability was greater than 50% of exceeding, it always was. When we said the probability was less than 50% of exceeding, it never exceeded. Here, we have only 2 out of 28 points that are out of line with this. This is a fairly good situation. At some point, if the prior map appears obviously incorrect, it may have to be abandoned and a strictly data-driven geostatistical model should be used. At what point the decision to abandon the prior map happens is unclear but will likely be site-specific and based on the professional judgment of the investigators. In such as case, it should raise alarms as to why our site understanding was so poor and what it means to the investigation now.

Next, we'll need to establish the correlation model for Ac-225 transformed to 0s and 1s based on the decision criterion 3pCi/g. This is done by clicking on the *Correlation modeling* step. Notice that even though *Prior Map* is the data set selected, *Ac-225* appears in the first and second variable sets. This is what you would expect with Markov-Bayes. We went ahead and set up an indicator cutoff of 3 for you (see how to do this in Chapter 31) and developed the correlation model for you as well.

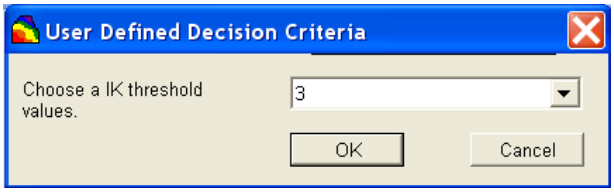


Next, we'll need to set the search neighborhood. Recall from the previous discussions on geospatial modeling that this can be based on the correlation structure. We did this for you already. Click on the *Search neighborhood* step.

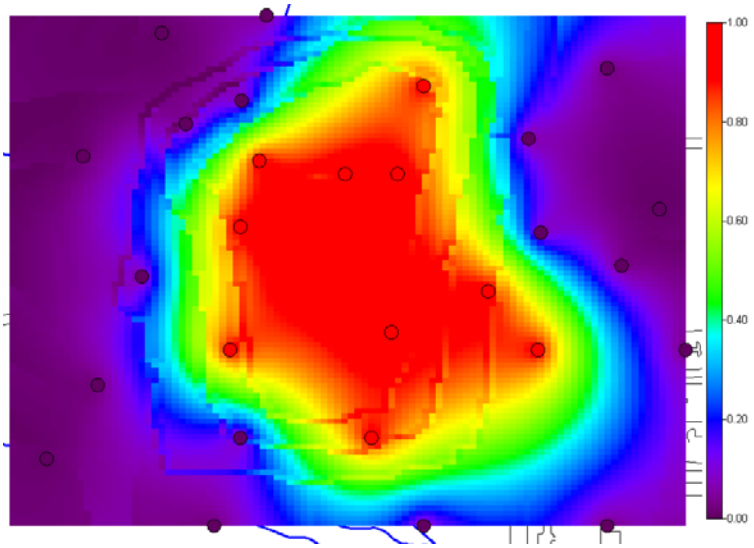


Defining the search neighborhood follows the same principles outlined in the advanced geospatial modeling chapters, so we won't repeat that here.

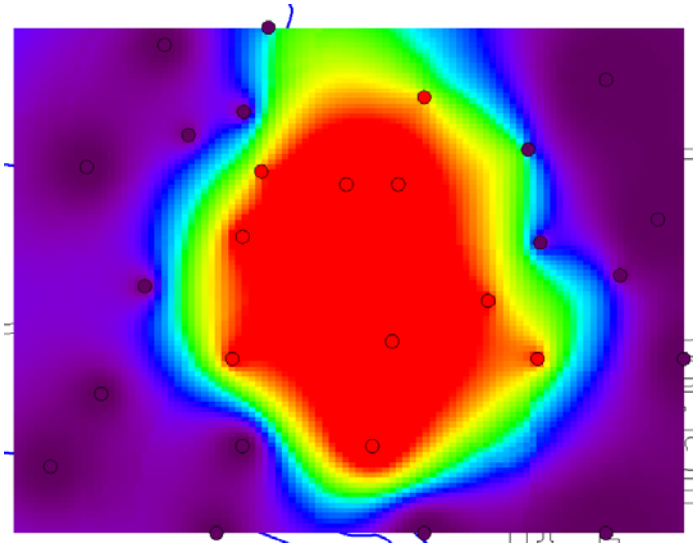
We're ready now to update the map. Click on the *Show the results* step and click on the *Show The Results* button. SADA will ask you to specify which indicator threshold value you want to use. In this case, select 3 and press *OK*.



Depending on what indicator threshold modeling activities you've done, you may see more than just one option here. Recall that the prior map was created with 3pCi/g in mind and so we need to stick with that indicator transform value here. SADA produces the following result.



The result is somewhat (but not remarkably) different than simply applying an ordinary kriging algorithm (below) to just the Ac-225 data to produce a local probability map (see Chapter 35).



There are differences, however, as the Markov-Bayes model is capturing the process knowledge coded into the prior map. Secondly, 28 fairly-well correlated samples is a good situation. If the data set had been 6 samples, then the model would have relied more on the prior map.

At the time this book was written, SADA can apply Markov-Bayes only to *User Created Models*. There is no reason it could not be applied to any probability map (user created or otherwise). We will be correcting this situation in an upcoming release.

Chapter 34: Spatial Risk

As mentioned in the overview of human health risk, Chapter 17, risk is based on exposure scenarios. An exposure unit is a geographically defined area where exposure will occur. Exposure is usually statistically summarized by a representative concentration that reflects the exposure experienced within the unit over time. This unit is based on some geographically defined area or a predefined, reasonable size (e.g., $1/8^{th}$ acre for residential). The tabular risk models discussed in previous chapters allow you to spatially define exposure units and automatically aggregate data found within the units into the representative concentration. This results in a single risk value for any given scenario.

In this chapter, we discuss contouring risk over space. SADA will convert any geospatial estimation map into a geospatial risk map. This conversion is performed by using each modeled block concentration value as the exposure concentration in the risk models. From a risk standpoint, individual block risk may not provide useful information on its own, especially for small block sizes. The scenario parameters often assume the receptor will be exposed to the contaminant over a number of years. This translates into an assumption that the individual will be exposed to a single block for a number of years and/or a large portion of the day. For small block sizes (e.g., 1ft x 1ft), this is not reasonable; however, for block sizes that are equivalent to the exposure area, block sizes can give an accurate perspective on the potential areas of concern at the site.

The spatial risk map does, however, provide useful information from a decision standpoint, as it identifies the geographic areas that are driving the risk when exposed over the entire unit. With the risk map visible, these geographic risk drivers are more easily identified. This has important implications when deciding on a remedial course of action. Removal of higher risk areas may sufficiently reduce exposure for certain activities. There are a couple of decision-analysis frameworks discussed shortly that address this very issue.

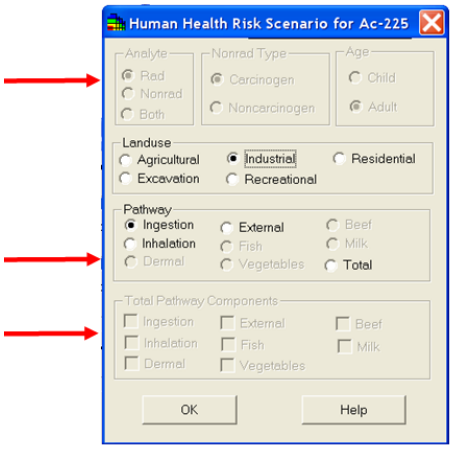
In this chapter, we will simply discuss how to convert an estimation map into a risk map and leave the discussion regarding decision analysis until later.

Human Health Risk Maps

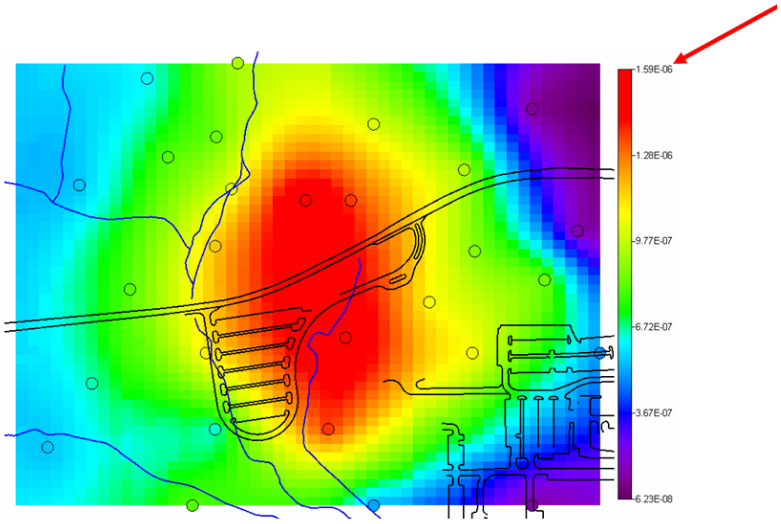
This discussion assumes you have read Chapters 17-21 on human health risk in SADA. We will begin where the tabular risk assessment and spatial risk screening tools left off. Open up the file SpatialRiskModeling.sda. Change the analysis from *General* to *Human Health*. Make sure you have selected *Soil* and *Ac-225*.

In a spatial risk model, you have complete control over the calibration of the toxicity and scenario parameters; however, the exposure concentration statistics have no meaning here. Each modeled block value will become an individual exposure concentration. This is necessary to identify the geographic risk drivers contributing to the total risk experienced on the site.

Switch the interview to *Draw a contoured risk map*. You will notice that the same steps appear as when you are simply contouring the site. The risk component is introduced after you press the *Show The Results* button. SADA will then ask for a specific scenario. An ordinary kriging model has been established already for Ac-225. In practice, you would need to explore the spatial variation in your data and derive your own geospatial model (Chapters 29-32). In the interest of simplicity, we have already taken care of this part. All you will need to do now is press the *Show The Results* button. SADA then presents you with a set of valid risk choices given your current contaminant and media type. Notice that all irrelevant options or no-choice options have been grayed out.



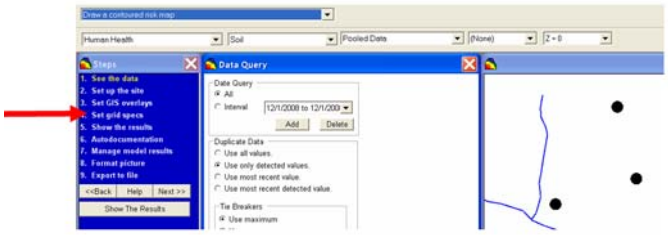
From the remaining enabled options, you can customize your own scenario. In this case, we'll just choose an industrial ingestion scenario. Select *Industrial* as the *Landuse* and press the *OK* button. SADA produces a risk map.



In most situations, the conversion of concentration or activity to risk will yield a similar looking map. This is because at most concentration ranges, risk is roughly a linear transformation. Notice, however, the legend has changed to reflect the application of the industrial ingestion risk model.

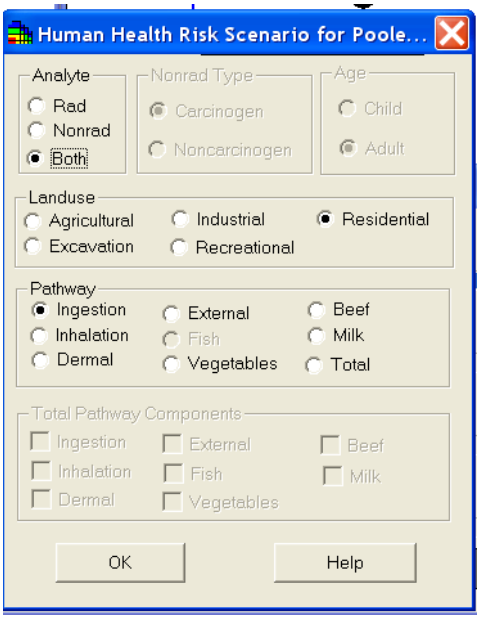
It is also possible to create a pooled risk map. In a pooled risk map, SADA will convert all relevant contaminants (depends on your scenario and toxicological choices) into risk and then sum them up. You will need a geospatial model established for each included contaminant. In this example, a geospatial model has been established for every contaminant found in the soil. In practice, you would need to do this yourself.

Choose *Pooled Data* from the list of available contaminants and make sure you still have the interview *Draw a contoured risk map* selected.

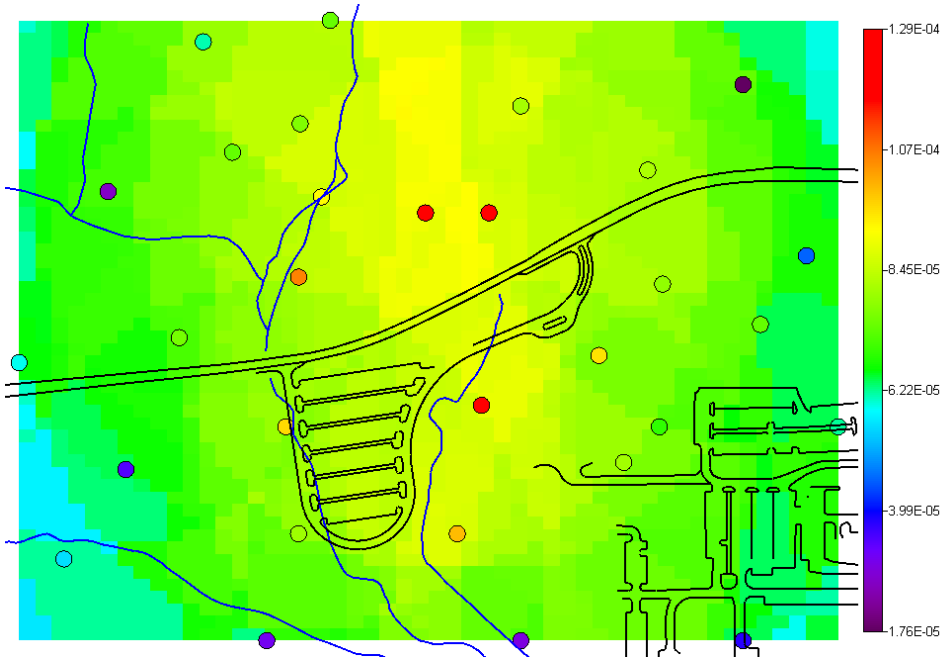


Notice that in the list of steps, there are no longer any steps specific to geospatial modeling other than *Set grid specs*. For pooled risk maps, these parameters must be calibrated for each contaminant individually. You would first need to visit each contaminant and establish a geospatial model and then return to this location. Fortunately, this has been taken care of for this example.

Press *Show The Results* and a risk/scenario options window appears. This time, many of the options are enabled. You will first need to choose the class of analytes you want to work with: *Rad*, *Nonrad*, or *Both*. Depending on your selection, the rest of the window adjusts itself, disabling and enabling various options accordingly. With *Both* selected, you must choose *Carcinogen*, as a *Noncarcinogen* evaluation would not make sense when you are including radionuclides in the analysis. If you switch to *Nonrads*, the *Carcinogen/Noncarcarcinogen* options become enabled. Spend some time clicking on the various options. When you are finished, make sure your options look like the following and press *OK*.

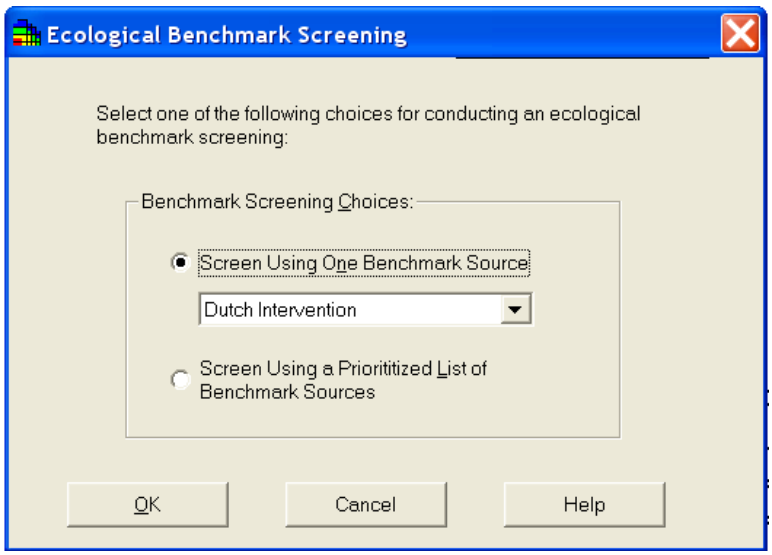


SADA responds with a summed risk map.



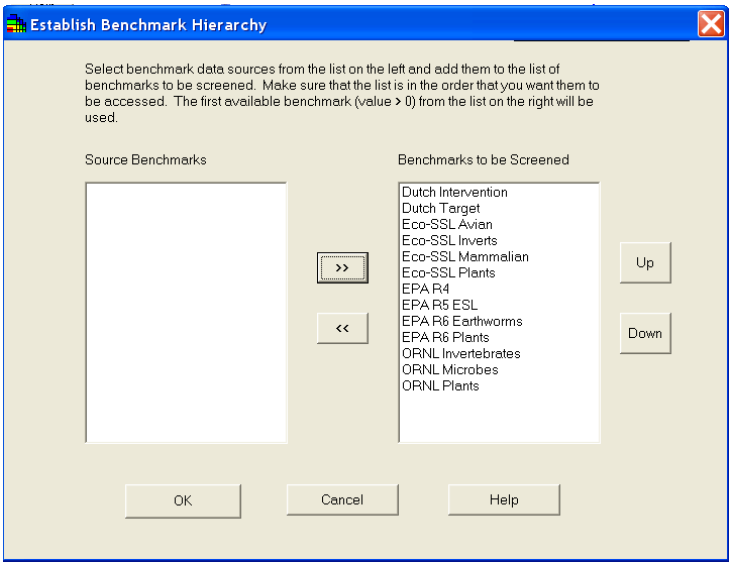
Ecological Risk Maps

The same kind of mapping can be conducted for ecological risk. This discussion assumes you have read the chapters pertaining to ecological risk assessment (Chapters 22-25). Switch to *Arsenic*, *Inorganic* and then switch the analysis type to *Ecological*. The first step in creating an ecological risk map is to calibrate the underlying geospatial model. In this case, for *Arsenic*, this has been done already. If you click on the *Interpolation methods* step, you can see that *Ordinary Kriging* has been selected. Press *Show The Results* and a benchmark selection window is presented.

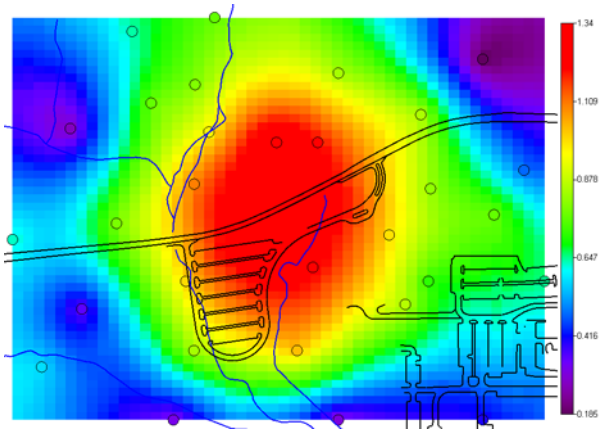


You may recognize this from the ecological point risk or ecological Data screen map. Select *Screen Using a Prioritized List of Benchmark Sources* and press *OK*.

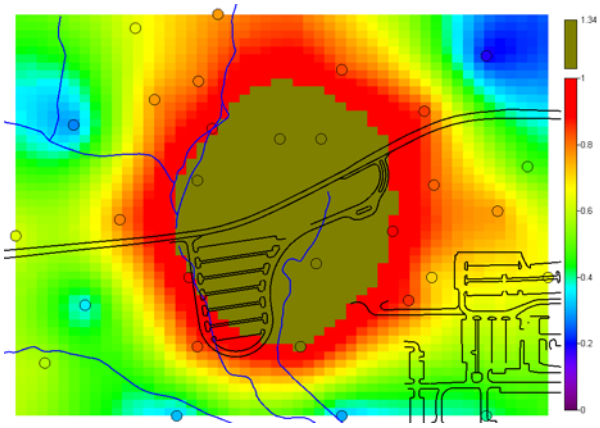
In the next window, you will create a prioritized list of benchmarks to choose from. Make sure your window looks like the following and press *OK*.



SADA will respond by converting geospatial model values into a risk ratio map.



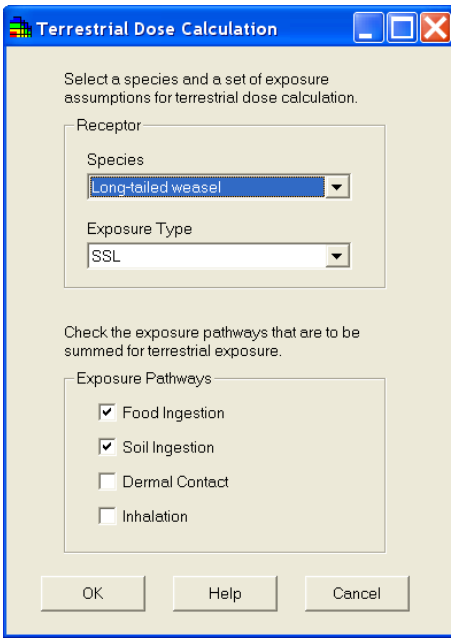
Try your hand at creating a new fixed continuous legend that is fixed from 0 to 1 (hint: *Graphics→Legend Manager*). Anything greater than 1 will receive a brown color. See if you can get this picture.



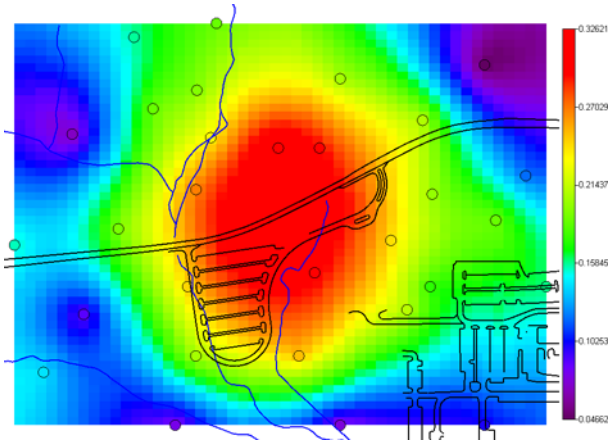
If you switch now to *Pooled Data* under *Ecological* analysis, you will notice that virtually all interviews including Draw a contoured risk map are now missing. This is because under *Ecological* risk, there is no equivalent to pooled risk.

Ecological Dose

The same kind of mapping can be conducted for ecological dose. This discussion assumes you have read the chapters pertaining to Ecological dose assessment (Chapters 26-27). Switch back to *Arsenic* and choose the interview Draw a contoured eco dose map. As previously mentioned, the first step in creating a dose map from scratch is to establish a geospatial model. This has already been done for you in this example (ordinary kriging). SADA will then convert these model values into dose values, allowing you to see the areas of the site that are driving the dose. Press *Show The Results* and you will be presented with receptor and exposure options.



You may recognize this from earlier chapters on point dose. If you are not familiar with these terms or options, please refer to the introductory chapters on ecological risk. Select *Long-tailed weasel*, *SSL*, *Food Ingestion*, and *Soil Ingestion* and press *OK*.



The dose map for this receptor is produced. Recall that you can control the behavior of the *Long-tail weasel* (and other receptors) by going to *Ecological→Configure Ecological Risk→Set Terrestrial Exposure Parameters*. Other parameters are found in *Ecological→Configure*

Ecological Risk→Set Terrestrial Contaminant Parameters. Please return to the introductory ecological risk chapters for more information on these screens.

We are now ready to move on to the decision analysis frameworks, which can integrate risk, geospatial models, and uncertainty together to quantify how each may affect a final decision outcome.

Part VI: Decision Analysis

Chapter 35: Decision Analysis

There is no practical limit to the number of ways geospatial and visualization tools can support a decision process for almost any type of application. You are certainly not limited to the set of formal decision tools provided here. There are a set of themes that continually arise, particularly in the context of environmental pollution, that have been encapsulated here. These tools are written in the context of environmental contamination assessment, but certainly can apply to a variety of situations.

The decision-support tools found here range from the very simple to the more involved. Many are based on the idea of a decision threshold limit or some value that represents an action/no action decision. A simple example would be a screening limit for mercury in the soil. Data points that exceed this value create concern and perhaps further investigation into their occurrence.

In certain cases, it may also be possible to ascertain the level of uncertainty associated with a decision and explicitly delineate how this uncertainty translates into a spatial context. For example, one may want to remediate an area that exceeds 5 mg/kg; however, the spatial distribution of samples exceeding 5 mg/kg may lead to uncertainty as exactly what the boundaries of remediation should be.

Decision thresholds

Decision thresholds are a core principle in all of SADA's formal decision frameworks. In SADA, there are four ways to establish a decision value, such as a screening value or an action/no action value. Each method is really associated with the four types of analyses that SADA provides. Under General analysis, you can enter your decision rule by hand. If you have setup a custom analysis, then SADA will allow you to choose from your set of imported decision criteria. The *Human Health* analysis will calculate the threshold value for you based on your selections for *Pathway*, *Landuse*, and so forth. The *Ecological* risk analysis allows you to choose from a set of benchmark values. Regardless of how the value is queried or calculated, it can be used in a number of decision tools discussed below.

Simple Spatial Data Screens

SADA can show you spatially where sampled values may have exceeded a decision threshold value. This has already been covered in Chapters 15, 21, and 27.

Probability Map

Geospatial models, such as kriging and simulation, allow you to quantify uncertainty by drawing probability maps. Probability maps display the probability of exceeding the decision criterion at each unsampled location. This is the continuous analogue to the simple spatial data screens applied to point data.

Area of Concern Maps

Based on the decision threshold, an appropriately calibrated spatial model, and the data at hand, you can determine where the boundaries of the area of concern should be. SADA can also calculate volume and mass and include overburden. You can also view uncertainty bands around your area of concern. We will take up the discussion of this framework in detail in this chapter.

Cost vs. Decision Threshold

This tool essentially runs the area of concern tool for a number of decision threshold values and calculates the volumes associated with contaminated media, overburden, and so forth. We will take up the discussion of this framework in detail in this chapter.

Secondary Sample Design

Sample designs, such as Threshold Radial and Area of Concern boundary, also make use of the decision threshold to spatially distribute new samples. We will be discussing these types of decision-support features in the sections on sampling designs.

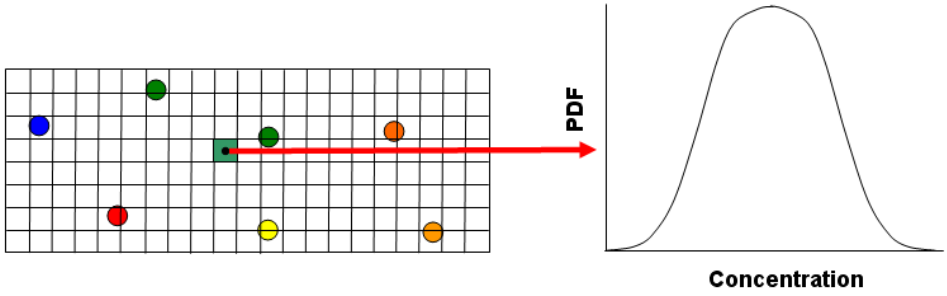
Contiguous Areas of Elevation

Geostatistical simulation allows you to calculate the probability that a contiguous volume of media exceeds a specific criterion. This has already been covered in Chapter 31.

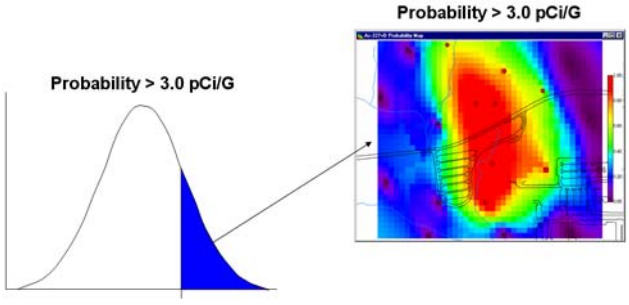
In this chapter we will discuss probability maps and area of concern maps.

Probability Maps

Simple spatial screening maps compare measured values against a decision threshold and highlight those points that exceed the criterion. At unsampled locations, we cannot determine with complete accuracy whether or not the measured value would exceed the criterion. What we can do is model our uncertainty about exceedance through the use of geostatistical modeling. Recall from Chapter 31 that geostatistics creates a CCDF (conditional cumulative distribution function) or PDF (probability distribution function) at each unsampled location.



It is therefore fairly simple to use this distribution as a model of uncertainty about whether an exceedance will occur. The following figure shows that the area under the PDF above the decision threshold (e.g., 3.0pCi/g) is the probability that the value is exceeded.

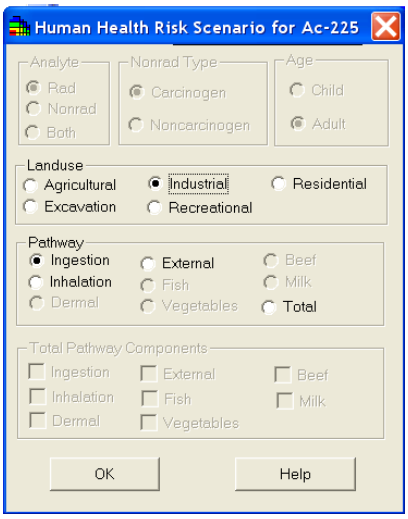


Calculating this probability at every unsampled node produces a probability map.

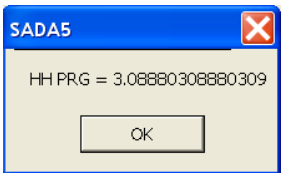
Open the SADA file DecisionAnalysis.sda, choose *Soil* for the media, and *Ac-225* for the contaminant. In the analysis drop-down list, select *Human Health*. In the interview drop-down list, select *Draw a probability map*. We will demonstrate how to draw a probability map using a human health risk model. The steps will be exactly the same for general, ecological, and custom analyses alike. The only difference will be that SADA will ask for a user-defined value, a benchmark criterion, and a custom value, respectively. (Visit Chapters 15, 21, and 27 for a refresher on how to set decision criteria within each analysis context.)

The first step in creating a probability map is to establish a geostatistical model. Chapter 31 and 32 devote a great deal of time to this subject. In the interest of time, we have already established a model for you. If you wish to change the model, you may do so by visiting the *Interpolation methods* step. Be sure to have Chapter 31 and 32 handy if you explore on your own. Before that, let's do a quick probability map based on an *Industrial* and *Ingestion* scenario for Ac-225.

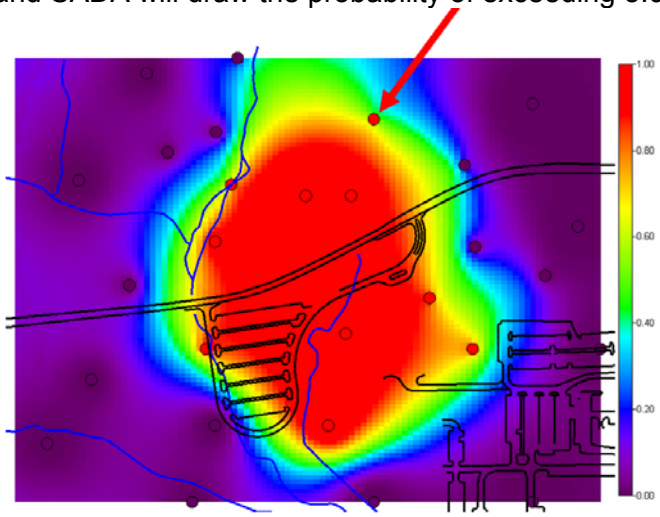
Press *Show The Results* and SADA responds (under a *Human Health* analysis) with a scenario options box that you've likely seen before.



Select *Industrial* and *Ingestion* and press *OK*. SADA pauses for a moment to tell you the decision criterion from this choice (based on risk models and parameters you've used to parameterize them (see Chapters 18-21).

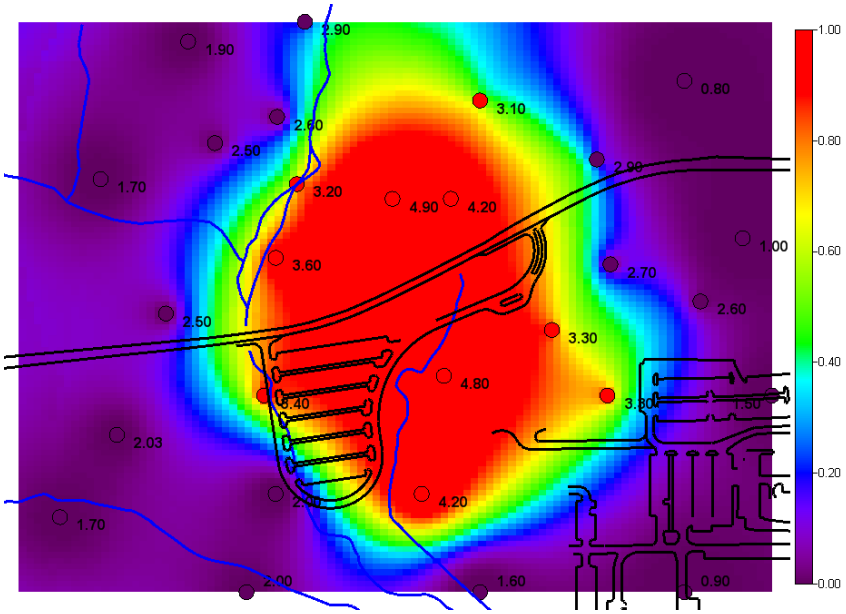


Press OK and SADA will draw the probability of exceeding 3.0888 pCi/g.



One of the criteria you always like to see when doing contour maps is a good matching between measured values and modeled values in their immediate vicinity. This is not necessarily true in probability maps. In a probability map, all data values exceeding the criterion are converted to 1. All values not exceeding the criterion are converted to 0.

Consider the red data value highlighted by the bright red arrow. Here we have a red value residing in a yellow/green area. Is this bad? Not necessarily. If you ask to see the original value next to each point (*metadata* drop-down list: value) you will notice that the value of this red point is 3.10. This is very close to the decision criterion and therefore the model is quite uncertain whether the values in the immediate vicinity exceed the criterion or not.



Uncertain areas such as these are good springboards to start talking about secondary sampling designs. In this case, a secondary design may be selected to place samples in the green areas (50/50) where the current sampling strategy has not shed much light on whether contamination exists or not. Secondary sampling designs are handled later (Chapter 38) and some of these decision principles are reiterated there.

Area of Concern Maps

An area of concern (AOC) is a geographic region in which failure to pass a decision criterion (user defined, human health, ecological, custom) can be attributed. You can create AOC maps by manually drawing the AOC yourself or using one of SADA's decision frameworks.

Manually Creating an AOC

In this approach, you simply create one or more polygons around the area you wish to include in the AOC (see Chapter 4, working with polygons section). Once you've created your polygons, you can access volume information by choosing *Tools→Area of Polygons*.

AOC Decision Frameworks (geospatial model)

Areas within the AOC can contribute to a decision criterion failure at two different scales: *block scale* and *site scale*. At the block scale level, if an individual cell value exceeds the decision criterion it is included in the AOC. At the site scale level, all blocks (grid cells) are sorted from highest to lowest modeled values. Beginning with the most contaminated block, SADA simulates the remediation of individual blocks from most to least contaminated until the average of all blocks no longer exceeds the decision criterion. Before we get into the details of each decision scale, some practical considerations need to be discussed.

Decision Result IDs

In the model, SADA identifies what class the block belongs to with simple integer IDs.

Block is contaminated	1
Overburden (including benching angle volume)	2
Block is not contaminated	3
Block might not be clean	4
Block might not be contaminated	5

So, if you export the model or setup a legend later on, you'll want to keep these numbers handy.

Backfill Concentration

For soil type applications, remedial designs may include *in situ* remediation and replacement or just removal and media replacement obtained from elsewhere. An important consideration for these applications is the amount of contaminant in the backfill. Not all backfill may be completely devoid of contamination. Now for the block scale, this factor is irrelevant and not even used. Obviously, as long as the backfill concentration is less than the decision criterion, the approach is valid. If it isn't, the remedial action will not satisfy the decision criterion anyway. For the site scale, it can matter. In fact, the backfill may actually exceed the criterion. It is possible that some blocks may be so contaminated that replacement with backfill that is slightly higher than the decision criterion may yield a substantive reduction in the site-wide average. Nonzero backfill values can lead to larger AOCs, as more must be remediated because the remediation is not 100% effective.

The value you enter for this parameter depends on the analysis you are doing and the measurement units at hand. If you are doing *General*, *Ecological*, or *Custom*, then you need to

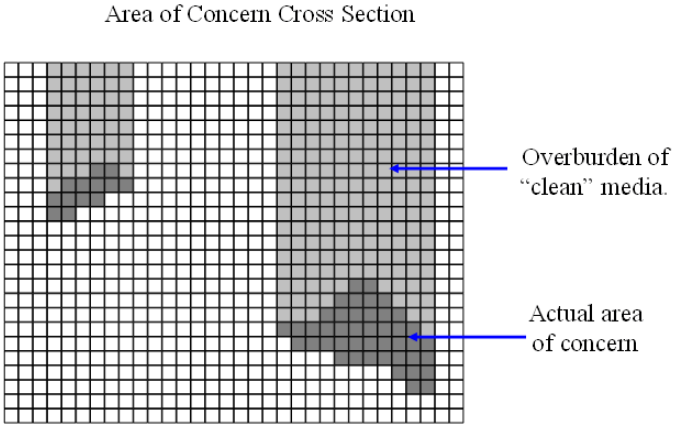
enter a backfill value in the same units as your data. **Important! If you are doing *Human Health* risk, you will need to enter the risk or health index value found in the backfill!**

Density Parameter

The density parameter is used to estimate the mass associated with a given remedial area. The mass is calculated by the product of the volume to be remediated and the entered density value for the media. This calculated mass value can be used to better estimate costs of remediation when mass is a cost driver.

Calculating Overburden

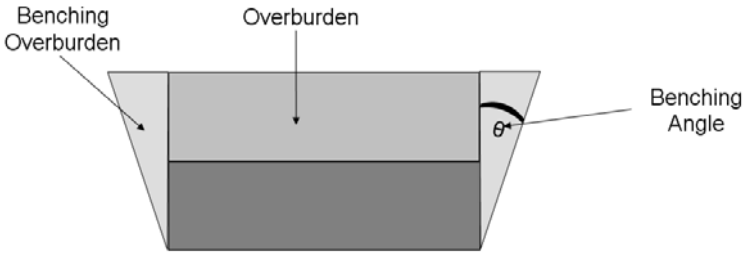
For three dimensional applications, after the area of concern has been identified using either block or site scale, SADA evaluates the "overburden" associated with this area. The overburden is any part of the site not included in the area of concern but lying vertically above the area of concern. The idea is to determine in an excavation-type scenario how much media would have to be excavated, including both clean and unclean zones. The implications for this are that any backfill concentration will not be applied to the overburden portion of the remedial zone. From a practical standpoint, this implies that any "clean" media will be removed and set aside while dirty media is removed and remediated. Afterwards, both "clean" media and any backfill are used to replace the affected volume. In order to be included in the overburden, however, a region must be found directly over the area of concern.



Whether the overburden is used or not depends on the characterization need. For example, if the purpose of the effort is to delineate the plume or to estimate a source term for calibrating a groundwater model, then the overburden should not be included. However, if the media is soil or sediment and an excavation scenario is required, SADA can provide two solutions. First, turn the overburden off to determine the volume of contaminated media that may be shipped or remediated. This will help ascertain the cost of dealing with the affected zone once it has been excavated. Secondly, turn the overburden on to determine the total volume that would be involved in vertically reaching and removing the affected area. Note that if you are using uncertainty banding, the overburden applies only to the decision criterion at the centered map (50^{th} percentile).

Benching Angle

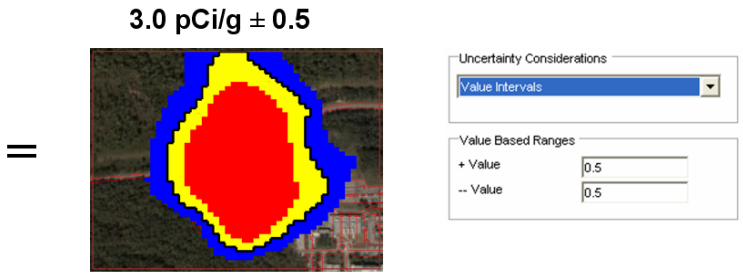
The benching angle refers to the amount of layback an excavated pit will require for safety concerns. The volume generated by the layback is treated as overburden. It will be set aside and will require no remedial action.



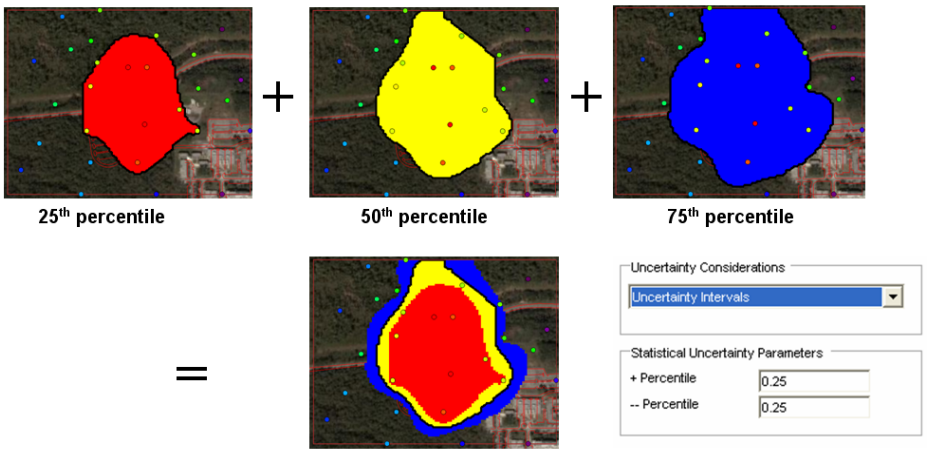
Note that if you are using uncertainty banding, the benching angle applies only to the decision criterion at the centered map (50th percentile).

Uncertainty Considerations

SADA allows you to put bands of uncertainty on your area of concern. This is available for both deterministic (basic) and geostatistical models (advanced). For deterministic methods, such as natural neighbor, an *ad hoc* method is used. Here, we will simply specify a range for the decision criterion itself (for example, 3pCi/g +/- 0.5pCi/g) and determine spatial intervals associated with these. Consider a decision criterion of 3pCi/g with an range of +/- 0.5 pCi/g. SADA first determines the AOCs separately for 2.5, 3, and 3.5 and then overlays these three results together.



Geostatistical methods afford a more rigorous delineation of uncertainty. With these methods, you specify a single decision criterion but apply the AOC framework to three different percentile maps. When you select your geostatistical model, you must choose to use the percentile option instead of the mean or E-type estimate. Typically, a value of 0.5 would be reasonable. Then you can specify a percentile range (e.g. +/- 0.25) and SADA would produce the percentile maps at 0.25, 0.50, and 0.75 and apply the decision framework to each of them.



Note that regardless of the model you choose, if you store the model, SADA loses the geostatistical elements such as kriging variance or CCDF specific information. Only the resolved map is stored. At the time of this writing, plans are in place to save all aspects of the model in future versions. For stored models, you can only use the value intervals.

Block Scale

The parameters and features we just presented are all applied after the selection of a decision scale. The block scale approach is the most conservative. Every single block must be less than the decision criterion. In this framework, the backfill concentration does not apply.

Site Scale

This scale may be more intuitive to risk assessors. In this approach, each cell block is essentially the remedial unit and the site is the exposure unit. Cell values are sorted from most to least contaminated and remediation is simulated (with backfill concentration considered) until the site wide (or exposure unit) area meets the decision criterion.

An AOC Example

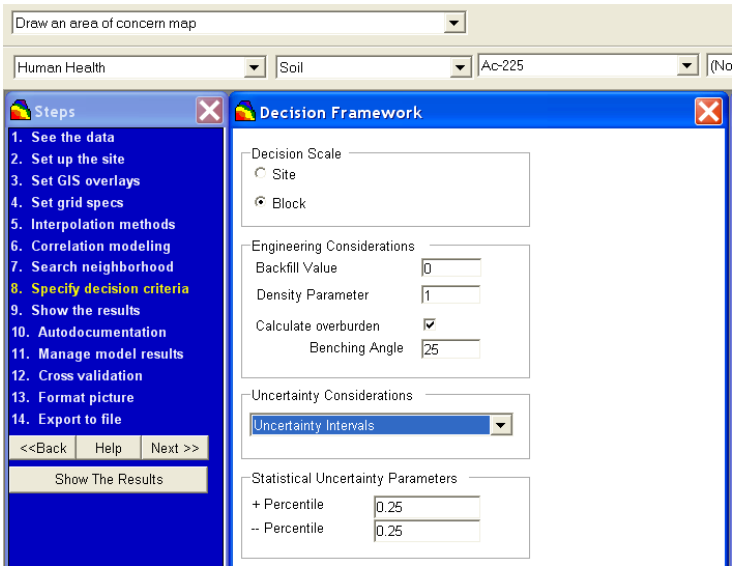
Let's open up DevelopingAnAOC.sda. This is the same data set as the Chlordane example we've been looking at; however, we've modified it some to make it more relevant to the current tutorial. The contaminant is now *Ac-225* and the media is *Soil*.

The first step is to setup a geospatial model. If you intend to use uncertainty bands by percentiles, then you need to choose a geostatistical model. Otherwise, value intervals will work with either deterministic (basic) or geostatistical (advanced) models. In this example, we've taken the liberty to set up an ordinary kriging model for you. If you are interested in calibrating your own model, please refer to Chapters 29-32.

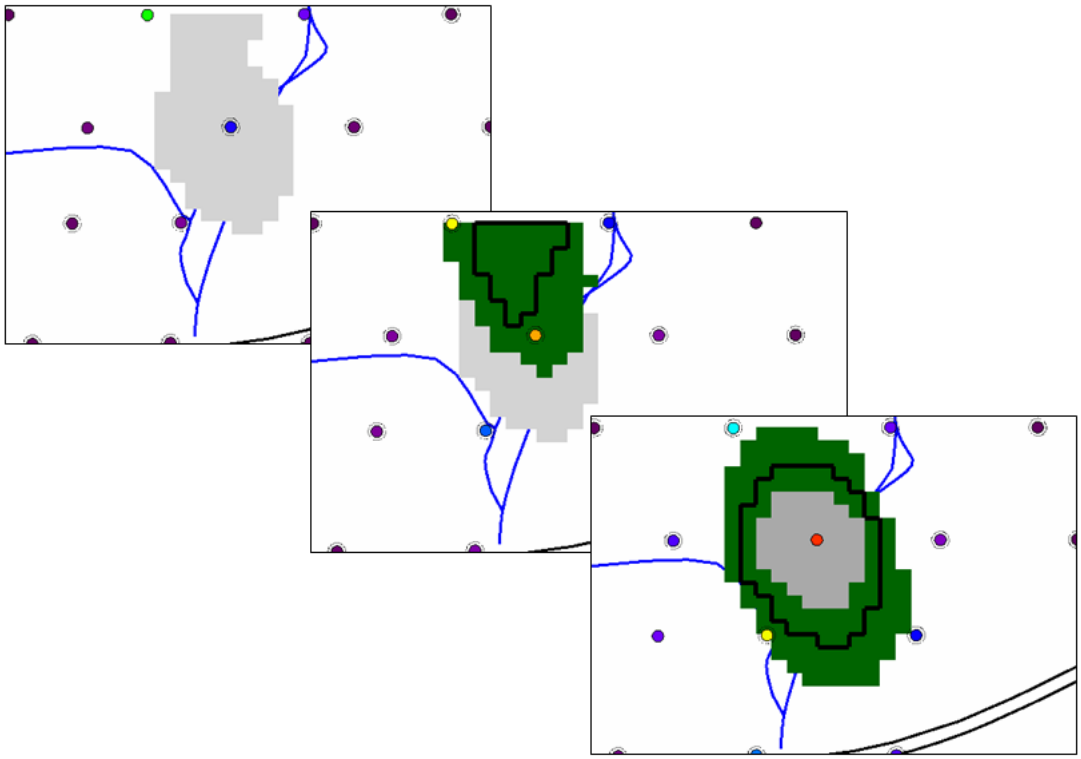
Switch to the interview *Develop an area of concern*. Notice the steps associated with setting up a geospatial model are present. You could calibrate the model here, although it may be more advisable to calibrate the model under the *Interpolate my data* interview. Feel free to peruse the model selection and corresponding correlation models and search parameters. These were selected again on quick recommendations by SADA (not advisable in practice).

Let's use a human health risk criterion in this example. You could just as easily setup an ecological model or import your own custom criteria and use those values at this point. If you are interested in this, please visit those chapters that cover this material. For now, switch to *Human Health* risk.

Now we'll specify the decision criterion. Under *General* analysis, an additional step is visible in the steps window called *Set decision threshold type*. Under *General* analysis, you can specify decision criteria as a function of depth (see Chapter 13). Under *Human Health*, *Ecological*, and *Custom* analyses, you can only have one decision criterion. You could compute the different decision criteria under these analysis types and enter them as depth-specific values. It would only take a little extra effort. Click on the *Specify decision criteria* step.



Let's choose the block scale first. With *Block* selected, *Backfill Value* remains enabled but isn't used. For the *Density Parameter*, we'll just use a value of 1. Let's check the box next to Calculate Overburden and set a *Benching Angle* of 25°. Let's also use the *Uncertainty Intervals* with percentile ranges of ± 0.25. Your decision criteria parameters should match the above image. Press the *Show The Results* button. Select *Industrial* and *Ingestion* and press *OK*.



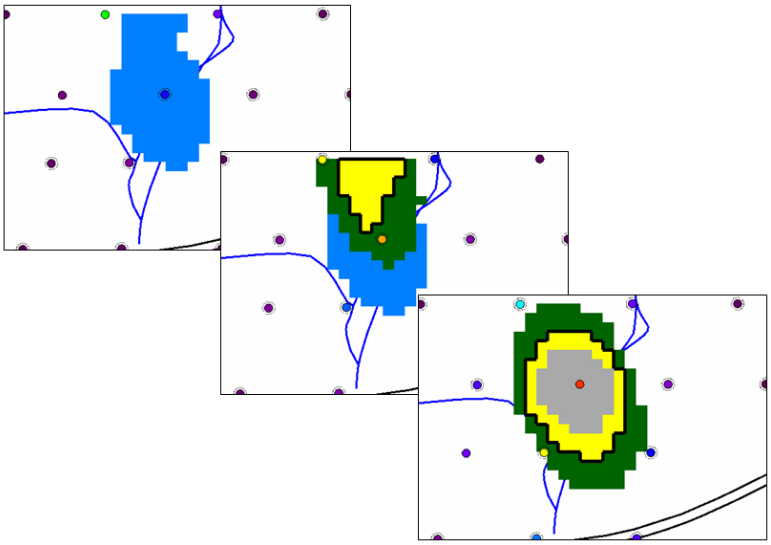
If you scroll through the top three layers, you will see different slices of the 3d AOC. At the top level, we have nothing but gray. This is all overburden. This area would need to be removed to get to the contamination underneath. The second layer sees the first intersection with the AOC. The thick black line designates the boundary of the AOC at this depth. The green area is an area where there is some uncertainty as to whether this should be included or not in the AOC. The gray region is also part of the overburden due to lower contamination. In the third layer, we see another section of the 3d AOC. Here, we again have the heavy black line designating the AOC boundary. The dark-gray area in the center should definitely be included in the AOC, regardless of any geospatial uncertainty. The green areas indicate uncertainty about the exact location of the boundary line.

The green area inside the boundary line is that region that is currently included in the AOC but may not in fact actually be contaminated. The green area outside the boundary is the area that is currently not included in the AOC but in fact should be. Let's change some color themes to better distinguish between these.

If you select *Graphics→Set Various Colors→Remedial Map Colors* you can see four options for coloring the AOC.

- Block is contaminated: this item will set the color for the area that is definitely part of the AOC given our model values (area currently colored dark gray).
- Block might not be contaminated: this item will set the color for the area just inside the AOC that may be misclassified (currently colored green).
- Block might not be clean: this item will set the color for the area just outside the AOC that may be misclassified as clean (currently colored green).
- Block is part of overburden: this sets the color of the overburden volume.

Let's change the *Block might not be contaminated* value to yellow. Keep the *Block might not be clean* as green and set the *Block is part of overburden* to light blue.



To see the volumes associated with this result, click on the *Show the results* step. In the log window, scroll down until you see the *remedial summary table*. This table contains results on both number of blocks and volume totals.

Description	Subtotal	Total
Total Number Of Unaffected Blocks		24704
Uncontaminated	24569	
Might be contaminated	135	
Total Blocks To Remove		296
Contaminated	50	
Might not be contaminated	79	
Overburden Blocks	167	
Total Unaffected Volume		72942327.22
Uncontaminated	72543719.14	
Might be contaminated	398608.09	
Total Volume To Remove		873985.14
Contaminanted	147632.62	
Might not be contaminated	233259.55	
Overburden Volume	493092.97	
Total Contaminated Mass		380892.17

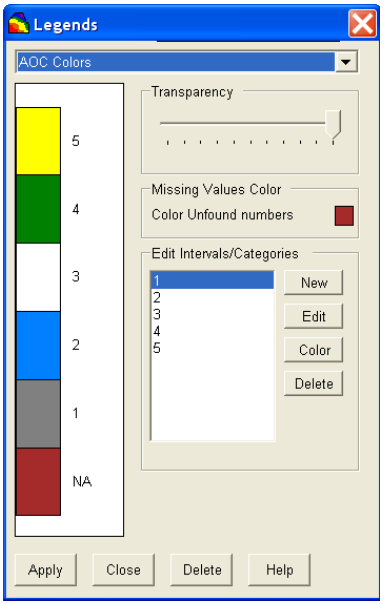
To export this result you will need to use the *Autodocumentation* feature.

Let's save this result. Click on the step *Manage model results*. Press the *Save Model Result* button and enter *Block Scale Decision* and press *OK*. Save your SADA file.

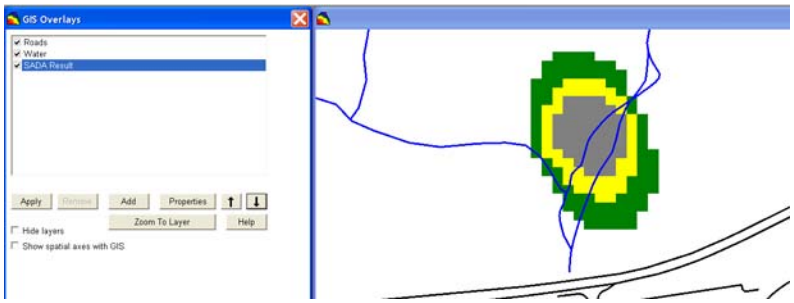
Change the media from *Soil* to *Imported Model*. Your result is saved here. Notice that the colors are quite different. SADA does not carry the remedial design colors with the store but you can

create a new legend that does the same thing. Recall from above that model values are recorded now as ID values that indicate to what class each cell belongs. Let's quickly setup a new legend that will handle our stored results appropriately. The details of the legend manager are given in Chapter 43. For right now, we'll just show how to get the legend we need for this application:

Select *Graphics→Legend*. In the drop-down list, select *New Categorical Scale* and enter AOC colors in the window that appears. Press *OK* to create the new legend. In the *Legend Manager* window, select the *AOC Colors* from the drop-down list and configure it to look like the following (refer to chapter 43 if you need help):

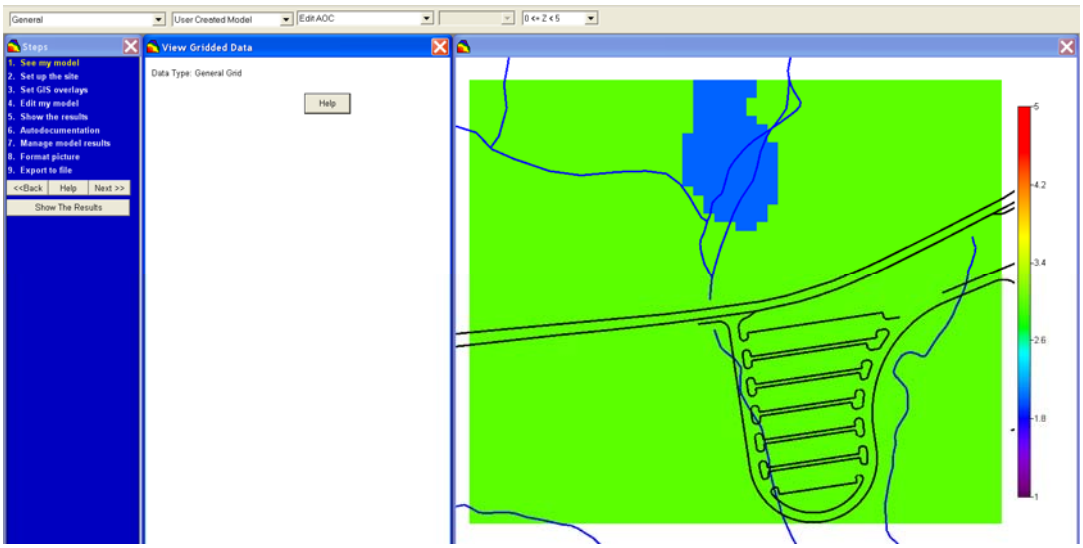


Press the *Close* button and the previous color scheme is restored. To finish the example, click on the step *Set GIS overlays* and move the *SADA Result* to the bottom of the list.



Save your SADA file now.

You can also store this result as an editable model. If a model result contains ten or fewer values, you can create an editable copy that you can further modify or extend. Let's do that quickly. Click on the *Manage model results* step again. This time click on the *Convert to Editable Model*. Enter "Edit AOC" in the box and press *OK*. Then, select *User Created Model* from the list of data types. SADA automatically switches to the interview *View my model* for this data type.



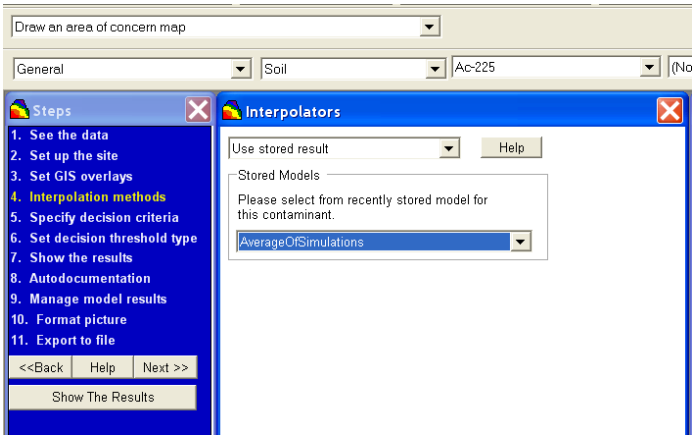
Select *Graphics→Legend Manager* and choose *Edit AOC* as the legend type. Click on the step *Edit my model*. Here you can change the color schemes associated with the editable version of the model. You can also extend the model using the graphics tools found here. For more information, please see Chapter 7.

Performing a site scale analysis is done exactly the same way. Select *Soil*, *Ac-225*, and the interview *Draw an area of concern* and set the analysis to *Human Health*. In the *Specify secision criteria* step, change to *site scale* and press *Show The Results*. We'll once again use the *Industrial* and *Ingestion* scenario, yielding a decision criterion of about 3pCi/g. It will take some additional time to compute due to the sorting part of the feature. When it's done, SADA will report that no remediation is required. This is because the average of all modeled values is less than 3pCi/g.

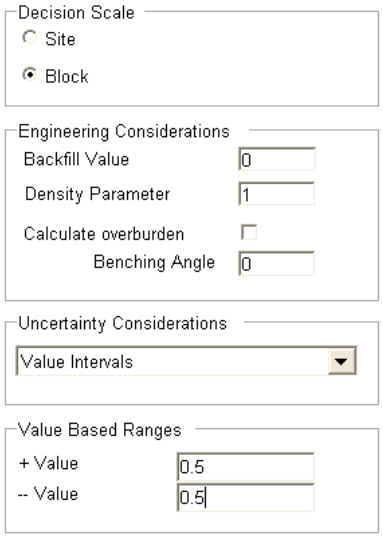
Areas of Concern from Geostatistical Simulations

The same type of results can be produced from post-processed simulations. We can either use the average of simulations map or the probability of exceedance map. Earlier, we produced a set of sequential Gaussian simulations and post-processed them to create an average-of-simulations map and a probability of exceedance map. We stored these results in the file DevelopAnAOC_Simulation.sda file, which you should open now. Select Soil, Ac-225, and choose the interview *Draw an area of concern map*.

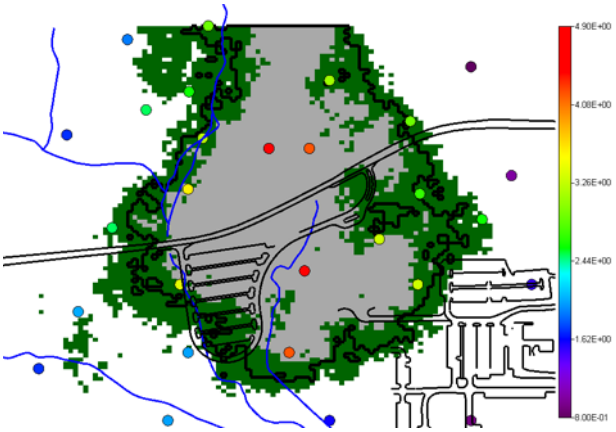
In the *Interpolation methods* step, choose *Use Stored Result* and select the model *AverageOfSimulations* in the drop-down list.



Click on the step *Specify decision criteria*. Since this is a 2d example, we won't need to worry about overburden or benching angles. We'll choose *Block* scale first. We also want to use *Value Intervals* for the *Uncertainty Considerations*. Since we are using a stored model result of concentration values, we no longer have any other measure of uncertainty.

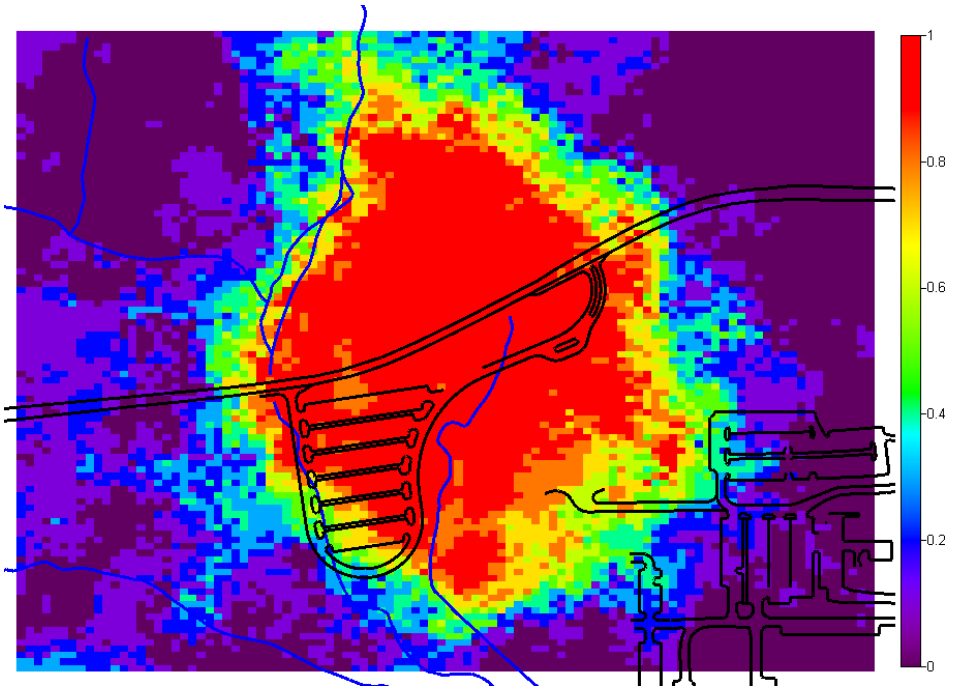


Press *Show The Results* and enter a value of 3 for the decision criterion.

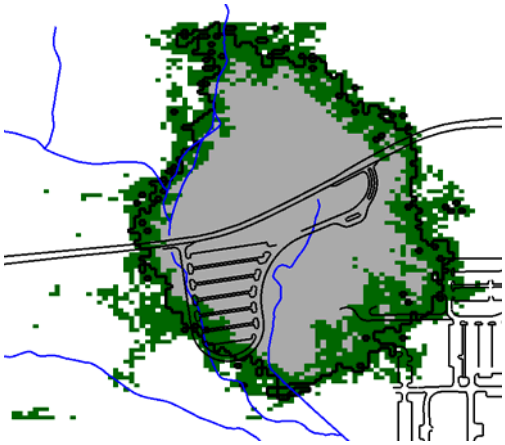


You can see from the results that simulations tend to produce more heterogeneous results.

We can also create an area of concern based on a probability map. In this demonstration, we'll show another way that you can apply decision frameworks. Instead of *Soil*, select *Imported Model* and select the *ProbabilityFromSimulation* model. Now we are working directly on the model instead of through the Ac-225 sample set. The only difference here is that the data points won't show. This is actually what we want, because the data are in pCi/g. Our *ProbabilityFromSimulation* model is unitless and runs at most from 0 to 1.



Switch to the interview *Draw an area of concern map*. SADA recognizes we are operating on a stored model and removes any questions regarding the type of geospatial model to use. If you click on the *Specify decision criteria* step, you will notice our previous values are missing. This is because decision criteria are data-set specific and technically this is a different data set than the Ac-225 soil data. So, we'll just quickly enter our values. Remember now that we are working in probability space, so we'll want to choose a decision criterion not in pCi/g but rather as a probability value. So, we'll choose 0.5. That is, we'll cleanup any area that has a greater than 0.5 chance of exceeding the criterion. We know that the criterion is 3.0pCi/g, but that has already been entered into the map result during post-processing. We need to be careful and select *Value Intervals* under the *Uncertainty Considerations*. Since this is a stored model, SADA can only do value intervals. We'll use a value of 0.25 for both + and -. This means that we are interested in the AOC at the 0.25, 0.5, and 0.75 probability of exceedance levels. Keep in mind that this is different than looking at exceedances at different percentiles. Press *Show The Results* and enter a value of 0.5 for the decision criterion.



The interpretation of this result is that the gray region has a greater than 75% chance of exceeding the criterion (0.5 or equivalently 3pCi/g). The interior green area is greater than a 50% chance but less than 75%. The outer green area has a less than 50% chance of exceeding the criterion.

Summary

The two examples provided here show the simplest way to apply decision analysis. Each of these decision frameworks apply to a single decision criterion (with a possibility of applying a +/- range). SADA also allows you to vary the decision criteria and evaluate many, many areas of concern at the same time. This will allow you to see the relationship between cost or volume and decision criteria. We'll cover that in the next chapter.

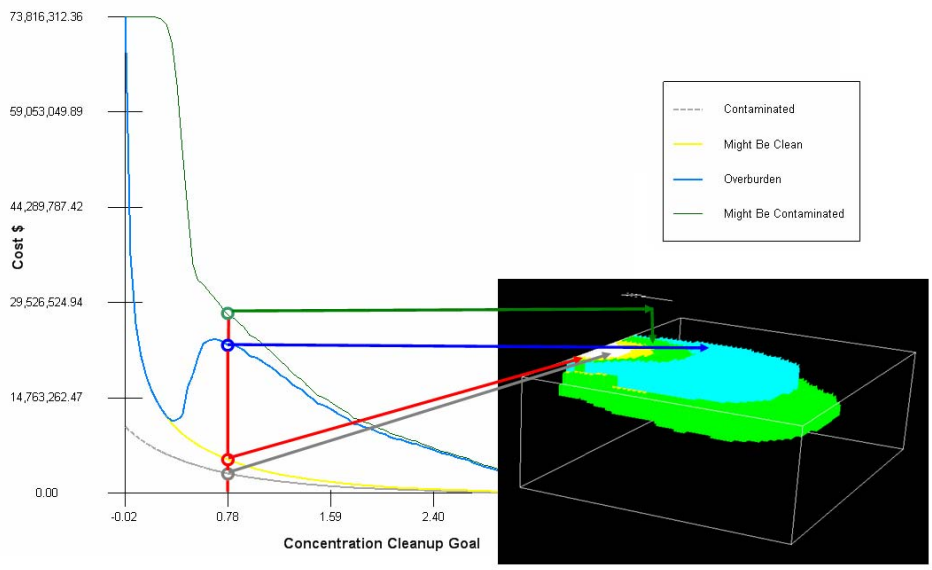
Chapter 36: Cost Benefit Analysis

We continue the decision-analysis discussion from the last chapter with an exploration of cost-benefit analysis regarding the area of concern and the decision criteria. An important component to decision-analysis software is cost/benefit analysis. This may be translated as payoff per level of effort. In a cost/benefit analysis, the level of effort or cost required to achieve a desired goal is modeled or estimated. Often, if the goal is numerically definable and ranges over a set of values, cost/benefit curves incrementally show how cost increases, decreases, or fluctuates across an increasing range of goal values through an XY graph. SADA provides cost/benefit curves for a range of remedial action goals. These goals may range over the minimum and maximum sample values or may range over the corresponding human health risk values, ecological benchmark ratios, or custom analysis values.

It is important to remember that typical XY lines in SADA are often calculated exactly for specific values of X. These points of exact calculation are usually distributed regularly along the X axis. Between these points, a simple linear interpolation is used to approximate values. Typically, the approximated value and actual value are very close; however, if an exact calculation is required for any given x or y, this value should be computed directly. For example, if one is querying a cost curve and must know the exact cost associated with a specific decision threshold, the *Area of concern* tool should be used instead. It will calculate the exact volume at the point of interest.

Recall that with geostatistical interpolants, this tool will strictly use the results of the interpolation routine. That is, if one plots a 10th percentile map, then the decision criteria is applied at about (or centered at about when developing an uncertainty model) the 10th percentile map. If a mean map is produced, then the decision criterion is applied to the mean map.

Let's take a look at a cost-benefit result before we demonstrate how to derive it.



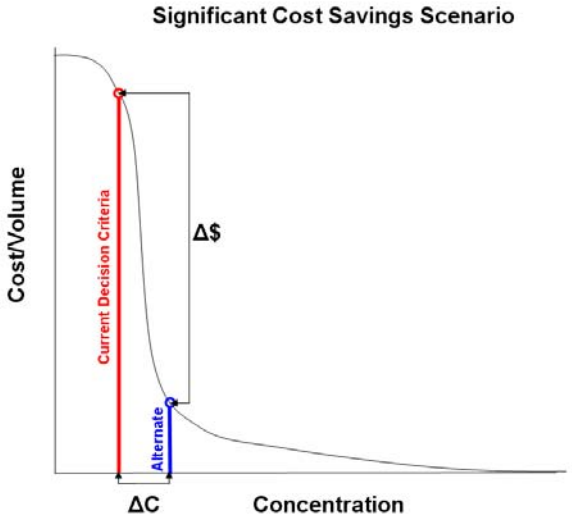
The image above shows a cost-benefit curve where the volume and/or cost of remediation is determined for over 100 evenly spaced values between the lowest and highest modeled values. For each value, the entire area of concern, including any uncertainty or overburden, is computed. Between values, a simple linear interpolation is used as approximation in order to create continuous cost-benefit curves. What this means is that for each point on the graph, there is a complete AOC model available (behind the scenes). For example, consider a

concentration cleanup criterion of 0.78pCi/g. This was selected because it's conveniently found at an X-Axis tick mark. If you draw a vertical line at this criterion value, notice that it will intersect four lines, each color coded so that they match the corresponding AOC map from which they are taken. The gray line corresponds to that portion of the AOC that is certainly a problem even when considering uncertainty. The yellow line represents the volume associated with the part of the AOC that might actually be clean, and cleanup may be potentially wasting resources. The blue line is the overburden (including any benching angles) that would need to be removed before accessing the contaminated media. The green line is the volume outside our designated AOC that might actually be contaminated when one considers uncertainty. The results are shown cumulatively in this graph. For example, the intersection of our 0.78pCi/g line with the blue line is the cumulative volume of the gray area + the yellow area + blue area at that decision criterion. It is also possible to show noncumulative results, as we'll see momentarily. In the case of 0.78pCi/g, we show the actual 3d AOC along with the plot. Lines are drawn from each line to its corresponding volume on the plot. The cost axis is trivially calculated as

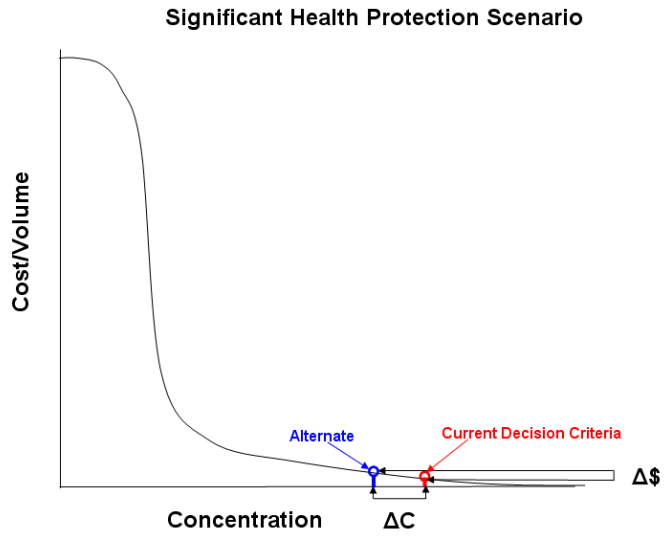
$$Cost = volume \times \frac{\text{cost}}{\text{unit volume}}$$

So a cost per unit volume of one would simply plot the volume.

Two important scenarios to consider are "significant cost savings" or "significantly increased health protection" in the vicinity of the decision criterion. Suppose that a decision criterion has been selected, resulting in a fairly large AOC volume. By plotting a cost-benefit curve, it may be possible to identify how small negotiations in the decision criterion may yield large project savings with little impact to environmental protection. The following graph shows that if the decision criterion is found along a steep portion of the curve, then small changes in the criterion (an alternate criterion) may yield big cost savings. This forms the connection between risk, site geography, and cost.

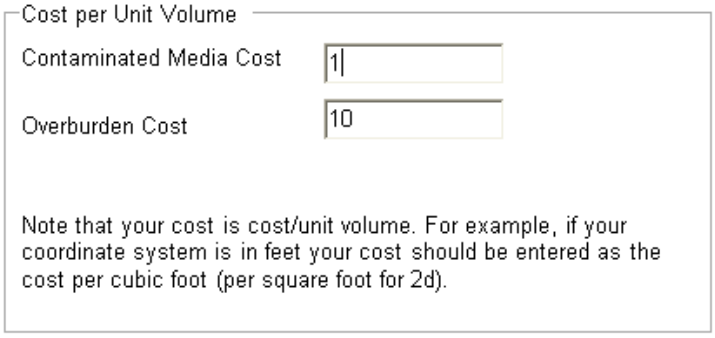


Conversely, for the same change in the ΔC for a decision criterion found in the upper end of the concentration range, you might be able to significantly increase the level of environmental protection at little cost.

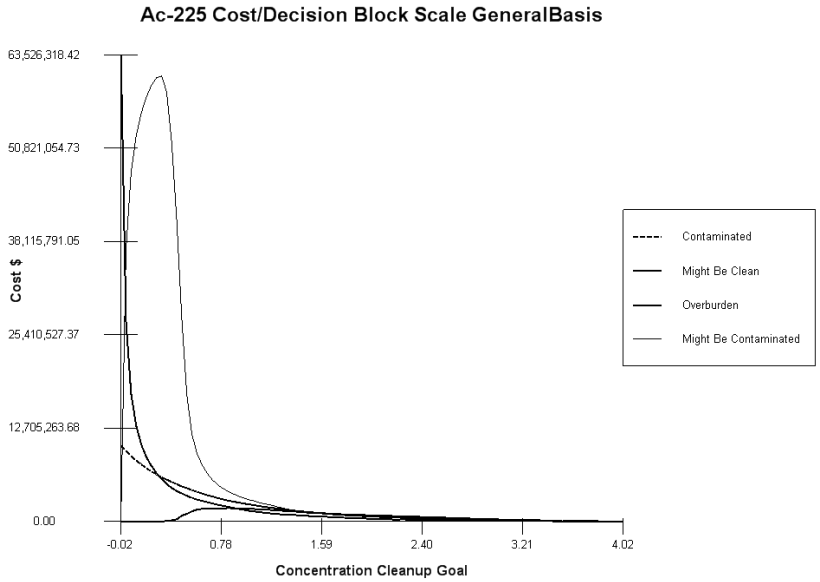


A Cost-Benefit Example

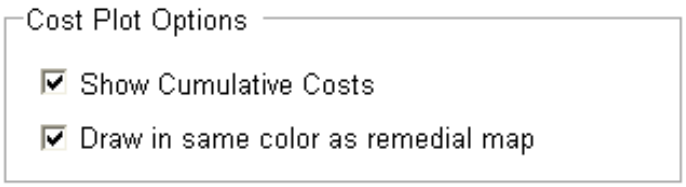
To demonstrate the cost-benefit analysis feature, open the file CostBenefit.sda. Select the interview *Calculate cost vs cleanup*. Make sure *Soil* and *Ac-225* are both selected. The first step in establishing a cost-benefit analysis is to select an appropriate geospatial model. In the interest of brevity, we have already calibrated an ordinary kriging model for you. The next requirement is to calibrate your decision parameters (e.g., parameters under the *Set decision criteria* step discussed in the previous chapter). Click on the *Set decision criteria* step now to view how this part has already been initialized for you. Now click on the *Set cost information* step.



In the first parameter box, we entered a value of $1/unit volume as the cost of dealing with *Contaminated Media Cost*. This produces a simple volume value. The *Overburden Cost* entry is the cost of dealing with uncontaminated media. In this example, we absurdly entered a value of $10/unit volume, which is 10 times the cost of dealing with contaminated media. We did this to exaggerate the overburden in the cost-benefit analysis discussion above. Sometimes, the overburden cost is so small that it is difficult to distinguish from the other lines. Let's change this now to a value of $1/unit volume. Press the *Show The Results* button.



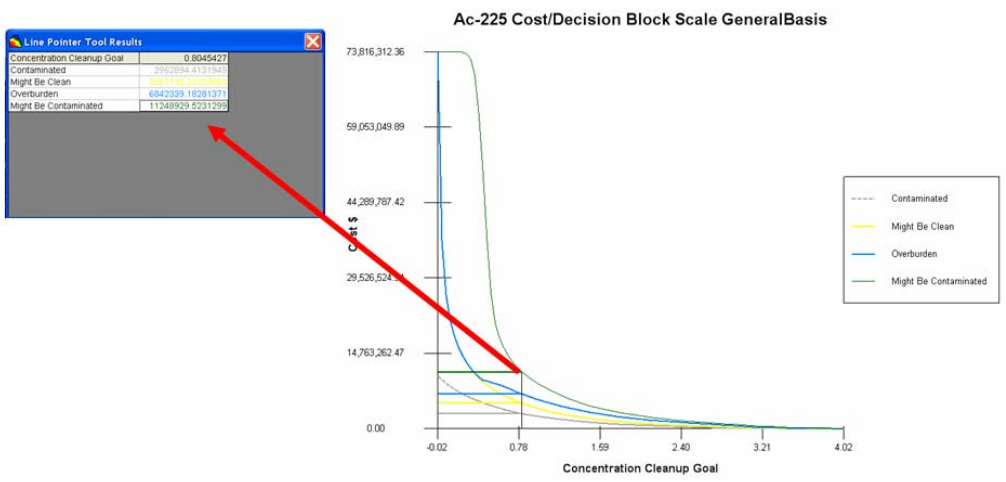
The first thing you may notice is that the result looks nothing like the result presented in the earlier discussion. This result is not reported as cumulative costs. So, each line you see is the actual cost line associated with that part of the AOC. Also, the lines are all drawn in black. This is the default setting for the cost-benefit curve. It is technically accurate but perhaps a bit difficult to interpret with various lines crossing each other. Let's first convert this to cumulative costs and draw the lines in the same colors as we've set for AOC maps (*Graphics→Set Various Colors→Remedial Map Colors*). Click on the *Format picture* step and check both *Cost Plot Options*.



Now the plot is reported in cumulative costs per decision criteria and can be a little more tractable to interpret. There are two tools that can further aid in understanding the cost-benefit result.

Line Pointer Tool

The Line Pointer is a tool that will allow you to graphically scan a particular XY line graph, such as a cost-benefit curve. By moving your mouse over the graph, the criteria/cost values will be reported in the status bar. You can access the tool by pressing the Line Pointer tool button (). Then, simply move your mouse over the cost-benefit graph. You will see each line value reported along with the corresponding X (concentration) value.

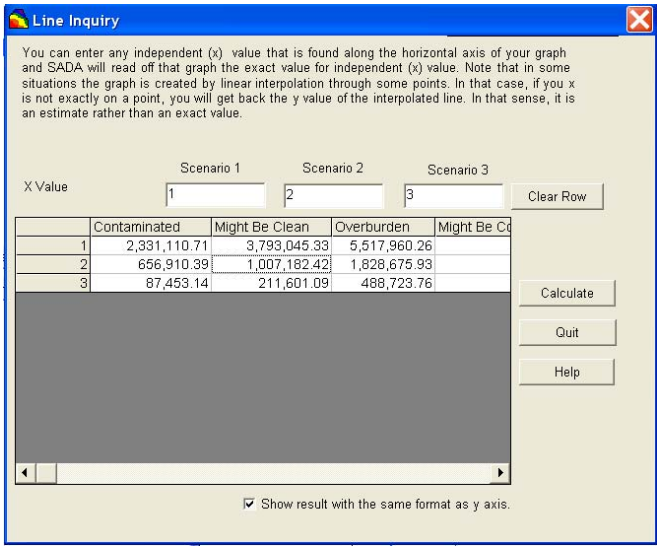


The text will be colored the same as the line. If the text is too hard to read, you might want to change the line color (*Graphics→Set Various Colors→Remedial Map Colors*). When you are done, press the Line Pointer tool button again.

It is important to remember that typical XY lines in SADA are often calculated exactly for specific values of X. These points of exact calculation are usually distributed regularly along the X axis. Between these points, a simple linear interpolation is used to approximate values. Typically, the approximated value and actual value are very close; however, if an exact calculation is required for any given x or y, this value should be computed directly. For example, if one is querying a cost curve and must know the exact cost associated with a specific decision threshold, the Line Query tool should be used instead. It will calculate the exact volume at the point of interest.

Line Query Tool

This tool allows you to query the graph for specific X values. Press the Line Query tool button () and you are presented with the *Line Query* window.



You can query up to 3 X values at a time. The associated volume/cost line values are reported when you press the *Calculate* button. Enter 1, 2, and 3 into scenario 1, 2, and 3, respectively.

Press the *Calculate* button. The results can be seen in the table. If you press the *Clear Row* button, then all X scenario values are erased. The checkbox near the bottom of the tool window allows you to report values in the table in the same format as the Y axis numerical format. Unfortunately, at this time, there is no way to export this table out automatically. Press the *Quit* button to terminate the tool.

Copy to PowerPoint (or other packages)

This is a good time to refresh our memory on different methods of export in SADA. Let's start by pressing the Copy to Clipboard button (). This will copy whatever image is in the graphics window to the clipboard. Open Microsoft PowerPoint and open a new slide show. Select *Edit→Paste*.

Export to Excel (or another package)

Suppose you would like to bring this result into another graphing package to improve the look and feel. Press the Export to Text button (), enter cost_benefit into the save box, and press *Save*. This will export the lines as columns in a comma separated format. Open Microsoft Excel and select *File→Open*. In the open dialogue window, change the *Files of type* to *Text file* (.csv, etc.). Navigate to the location where you saved the cost_benefit file and select it now.

Microsoft Excel - cost_benefit.csv

File Edit View Insert Format Tools Data Window S-PLUS Help Adobe PDF

A1 Cleanup Goal

	A	B	C	D	E	F	G
1	Cleanup Goal	Contaminated Volume Cost	Might Be Clean Cost	Might Not Be Clean Cost	Overburden Cost		
2	-0.023958	10287041.29	63526318.42	2952.652494	0		
3	0.016853697	9569546.734	27188024.17	37058741.46	0		
4	0.057665394	8896341.966	17370454.62	47549515.77	0		
5	0.098477091	8291048.204	13033008.11	52492256.05	0		
6	0.139288788	7768428.713	10467153.09	55580730.55	0		
7	0.180100485	7260572.484	8751661.993	57786361.97	0		
8	0.220912182	6791100.737	7564695.691	59354220.44	0		
9	0.261723879	6407255.913	6572604.453	60408317.38	0		
10	0.302535576	6014553.131	5778340.932	60644529.58	5905.304989		
11	0.343347273	5689761.357	5087420.248	58521572.44	41337.13492		
12	0.38415897	5379732.845	4552990.146	51597602.34	144679.9722		
13	0.424970667	5063799.028	4121902.882	40888331.74	286407.292		
14	0.465782364	4741959.906	3803016.413	27489194.72	788358.216		
15	0.506594061	4520510.969	3481177.291	17160816.3	1033428.373		
16	0.547405758	4266582.854	3180006.736	11846041.81	1364125.452		
17	0.588217455	4018560.045	2973321.062	9221133.74	1615100.914		
18	0.629029152	3788253.15	2793209.26	7688707.095	1706633.142		
19	0.669840848	3563851.561	2616050.11	6554888.538	1765686.192		
20	0.710652545	3392597.716	2435938.308	5669092.789	1795212.717		

Summary

The results of a cost-benefit analysis may indicate that the cost associated with remediation under such uncertainty may be too high. It may be more beneficial to first take some additional samples at key locations to reduce the uncertainty and produce an AOC that is less cost prohibitive. We take this discussion up in the next chapter.

Part VII: Sample Design

Chapter 37: Overview of Sample Designs

SADA provides several sample design strategies. These designs can be broadly categorized as either initial or secondary approaches. Initial sample designs are used to create first round sampling designs. These can be exploratory designs, regulatory specified approaches, or even targeted type designs if some prior knowledge is known. Secondary sample designs are used when sampled data and/or modeled results already exist and you wish to further refine your understanding, model, or decision regarding the site. Most sample designs in SADA are available for 2d and 3d data. Before we get into any specific design strategies, we need to cover some prerequisites.

Determining Number of Samples

The method for determining the number of samples will depend on the chosen sample design. For a number of classic sample designs that seek to control Type I or Type II errors in some future hypothesis test, SADA provides a couple of non-parametric ways to determine the number of new samples (sign test, and Wilcoxon rank sum test). This number of samples is then used in whatever design was selected to distribute them. Other designs, such as those that search space in some efficient manner, will produce both number and location of samples as the simultaneous outcome of the design. Still other geographically targeted designs may depend on geospatial models and the number of samples is difficult to compute a priori. These may require the minimization or maximization of some external cost-benefit function. From a more practical vantage, sampling budgets will often dictate the number of samples that can be afforded. In these cases, it's the objective of the code to place these as optimally as possible. SADA provides two classical approaches for determining the number of samples in support of a future hypothesis test about whether the mean or median will exceed a decision criterion.

Number of Samples for Sign Test

Suppose we want to sample the site for the purpose of determining whether or not the median value of the site exceeds our decision criterion or not. After we've collected the data, we anticipate that we will use the sign test to determine if the site is clean or contaminated. The following hypothesis determines the number of samples you will need in order to test the hypothesis at some point in the future. We state the hypothesis in two ways that are equivalent in order to better understand what is going on.

$$H_0 : \text{The site is clean}$$
$$H_1 : \text{The site is contaminated}$$

While the null and alternative hypotheses are easy to understand when written this way, we need to be a bit more specific when testing in the context of the sign test. The way the sign test works is by first transforming all data values into 0 (if value is less than or equal to the decision criterion) or 1 (if the value is greater than the decision criterion). Under this framework, the site would be considered clean if the median of all these transformations is equal to zero. The site would be considered contaminated otherwise. This is intuitively a one-sided test, simply by the way we've defined the transformations. With this in mind, we rewrite the hypothesis in the context of the sign test.

$$H_0 : \text{The median of the population of all transformed values is zero}$$
$$H_1 : \text{The median of the population of all transformed values is not zero}$$

At first glance, this might appear to be a two-sided hypothesis test. But recall that if the median is not zero, then the median must be greater than zero and therefore the site is contaminated. Because of the way we defined our transform (0 = data is at or below threshold, 1 = data is above threshold), we can never have a median less than zero.

When the data is collected, SADA will transform the values into 0s and 1s and count the number of 1s for the test statistic beta. A critical value for a given alpha level or p-value is derived based on the binomial distribution.

Then the null hypothesis is either accepted or not accepted. If the null is accepted, we infer that the site is clean. If the null is not accepted, we infer that the site is contaminated. Whether we accept it or not depends of course on the Type I/Type II error (alpha) we are willing to accept.

With all this in mind, we wish to estimate the number of samples we'll need to take. This will obviously require that we make some considerable assumptions about the data, which we have not yet collected. The non-parametric sign test will require the following information from you.

Sign Test Parameters

Decision Criterion

This is the value for which you will later test whether the data you collected exceeds or not.

Lower Bound of The Gray Region (LBGR)

If you insisted on being able to reject the hypothesis under any sampling outcome, then the sampling requirements might be very large. To mitigate this, you can specify a range just below your decision criterion for which you are willing to accept uncertainty about whether the decision criterion is exceeded or not. If your data fall below this lower range, the site is clean. If the data fall above your decision criteria, the site is contaminated. If the data fall between these two, you may not be able to tell if the site is contaminated or not. You will either have to assume it is contaminated and run the risk of unnecessarily cleaning your site, or you may take more data to distinguish whether they are above or below the criterion. Either way, this is something to worry about after the collection is done. Right now, it allows you to control the number of samples

Sigma

This is the standard deviation of the data you will collect. This is possibly the most difficult parameter to really specify. If you have previous experience with the data, it could be helpful. If you are wrong about this value, then after the sampling is done, you may find you have too much or too little statistical power. For example, if the real sigma is smaller, you might end up with a really small p-value (e.g., 0.00000001). If the real sigma is larger, you might end up with a larger p-value than you hoped for (e.g., 0.15).

Alpha

This is your Type I error rate, which is the risk you accept of rejecting the null hypothesis when you should accept it.

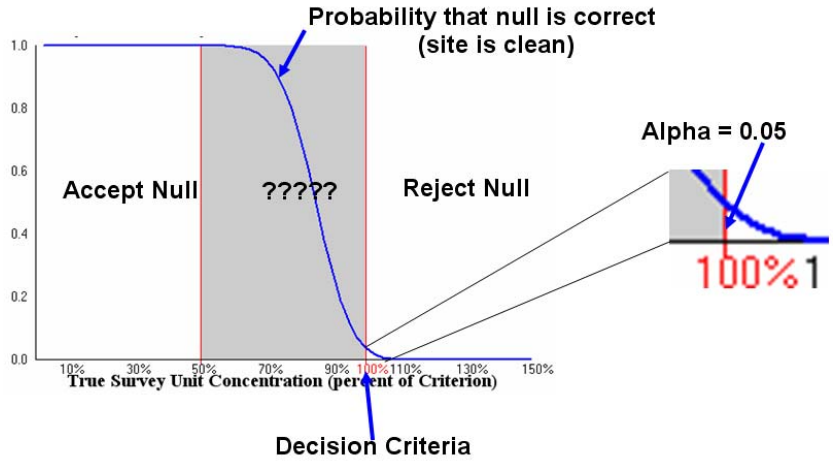
Beta

This is your Type II error rate, which is the risk you accept of accepting the null hypothesis when you should reject it.

The specification of these parameters will create a power curve for the particular decision criterion you've selected. In this power curve, rather than present the criterion as an actual measured value, the X-Axis is a range of values expressed as a percentage of your criterion.

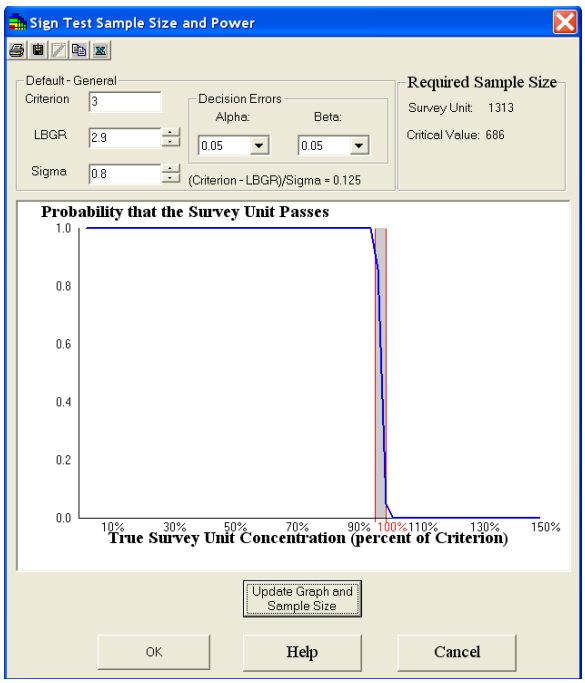
So, 100% is the criterion itself. If the criterion is 3pCi/g, then the 50% tick mark is actually 1.5pCi/g.

Suppose we set a decision criterion of 3.0pCi/g, a sigma of 0.8, a LBGR of 1.5pCi/g and alpha and beta values of 0.05. The following power curve is produced, suggesting 16 samples should be taken.



The blue line represents the probability that the null hypothesis is correct. In this case, if we collect the data and the median value is 1.5pCi/g or less, we know we can accept the null hypothesis. If we collect the data and the median value is greater than 3.0pCi/g, we can reject the null hypothesis. Note that the blue line does not cross the 100% mark (decision criterion) at probability of zero. This means that there is some chance of incorrectly rejecting the null hypothesis when the median is above the decision criterion (specifically equal to alpha). If the median falls into the gray region, then we really can't say with much confidence whether the median will exceed the criterion. In this case, more samples would need to be taken or you would need to accept some higher risk.

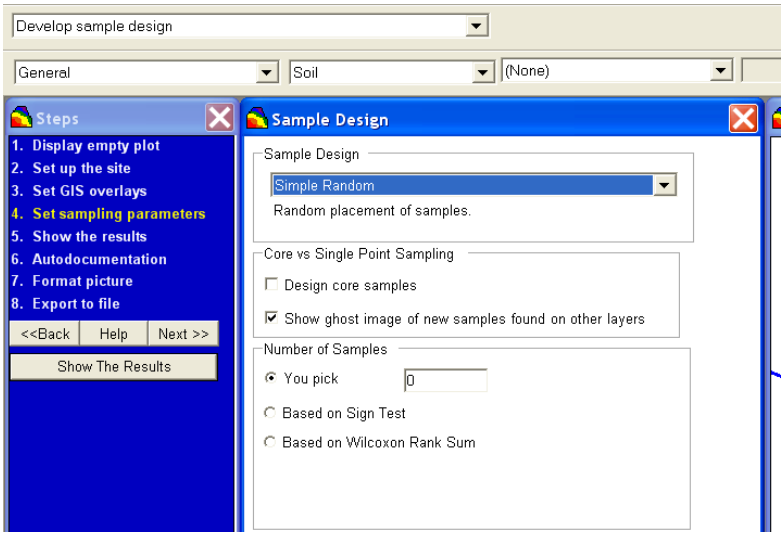
As a demonstration of the value of the LBGR, the following graph shows the same curve with a LBGR of 2.9pCi/g. This curve requires 1313 samples.



The LBGR is really a balance between cost and willingness to accept some chance that your data set may not be sufficient to test the hypothesis at the specified alpha level.

Sign Test Example

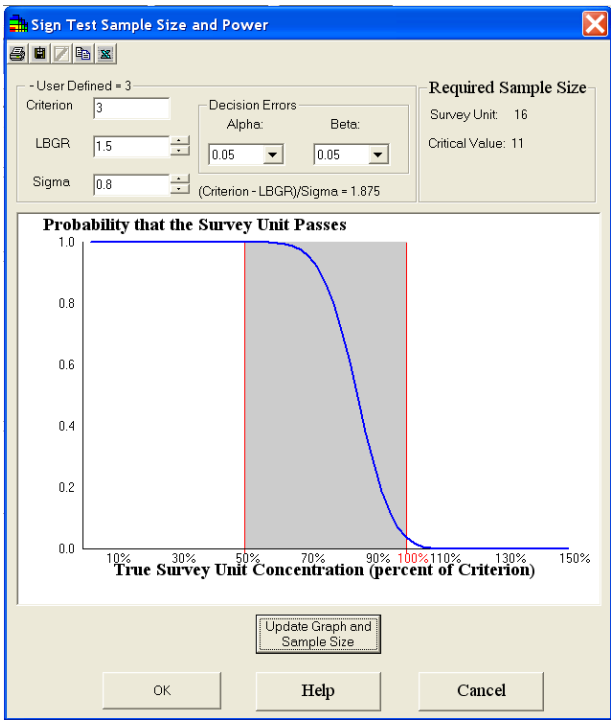
Open up the file SampleDesigns.sda. In the drop-down list of contaminant data sets, select None. In the interview drop down list, select *Develop sample design*. Click on *Set sampling parameters*. At the top of the parameter window is a drop-down list of possible sampling designs. We'll get to each of these shortly, including why we selected None as the contaminant. Right now, we'll just select a sample design that allows you to see where the choice for Sign Test is found. Select *Simple Random*.



Below your *Simple Random* selection is a parameter block called *Number of Samples*. This is normally where you would specify what method you wish to use for determining the number of samples. Not all sample designs have this parameter block, as the calculation of number and

location of samples is done at the same time. Here, select *Based on Sign Test* and press *Show The Results.*

You will next be asked for the decision *Criterion*. Let's enter a value of 3 and press *OK*. The sign test power curve is presented.



Notice in the upper-right-hand corner is the *Required Sample Size* information block. This contains the number of samples required in your survey unit (your site). It also shows the number of data that must exceed the criterion (11) in order to declare the site contaminated at the alpha levels you've specified. Change the LGBR to 1.5 and press *Update Graph and Sample Size*. Spend some time playing around with each of the parameters to see how the *Survey Unit* and *Critical Value* is affected. Note that the toolbar at the top of the window has the usual export features. When you are done, press *OK* and SADA will distribute your samples randomly throughout the site.

Note that you can also access this feature independently of spatial distribution. Select *Statistics→Number of Samples→Sign Test*.

Number of Samples for Wilcoxon Rank Sum Test

The Wilcoxon Rank Sum (WRS) test is another non-parametric approach for testing whether the site is contaminated or not. An important difference from the Sign Test is that WRS tests not against a single decision criterion but against another distribution. This second distribution is usually background within an environmental context. So, WRS tests to see whether or not the set of data collected in the contaminated area is different than the background data set.

The way the test works is to combine site data and background data sets into a single array of numbers and sort them. Each of the sorted numbers is then given a rank according to their position in the ordering. The first value receives a rank of 1, the second value a rank of 2, and so forth.

Suppose we have measurements of a contaminant both on site and in the background. For simplicity, 10 samples are available for each.

A		B	
Value	Location	Value	Location
21	Site	8	Background
15	Site	7	Background
16	Site	13	Background
12	Site	5	Background
18	Site	11	Background
16	Site	14	Background
15	Site	6	Background
12	Site	4	Background
19	Site	8	Background
16	Site	9	Background

Now they are combined and sorted.

Sorted Values

Value	Location	Rank
4	Background	1
5	Background	2
6	Background	3
7	Background	4
8	Background	5.5
8	Background	5.5
9	Background	7
11	Background	8
12	Site	9.5
12	Site	9.5
13	Background	11
14	Background	12
15	Site	13.5
15	Site	13.5
16	Site	16
16	Site	16
16	Site	16
18	Site	18
19	Site	19
21	Site	20

Now, we compute sum of the ranks for Background and for Site separately.

Background Sum Of Ranks = 59

Site Sum of Ranks = 151

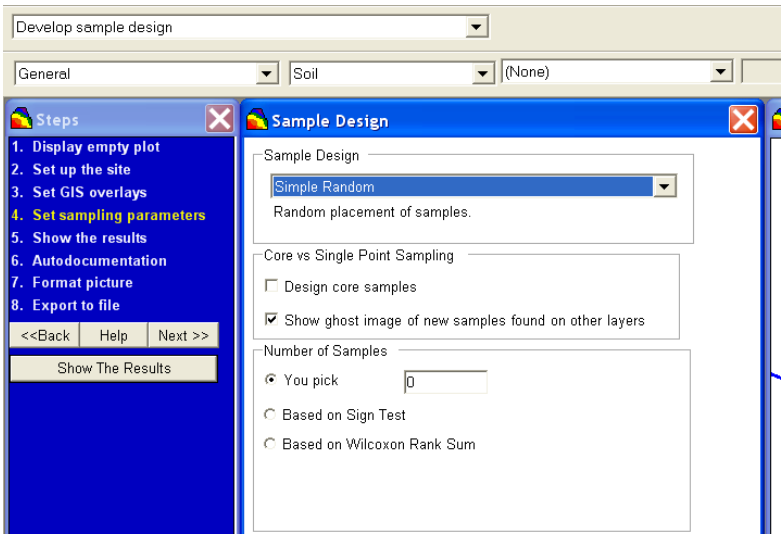
The WRS upper tail test statistic is 127 (retrieved from WRS critical values table) at $\alpha = .0\ 5$ with 10 samples each. Therefore, the Site value of 151 exceeds 127 and we can infer that the site distribution is greater than the background distribution. In other words, the site is contaminated relative to background.

The parameters for calculating the number of samples is exactly the same as the sign test. The difference is in the meaning of the decision criterion. In the case of the WRS test, you will be asked for a decision criterion. This is the number you expect to be the median of the

background data set. Furthermore, the number of samples is calculated based on the approximation that all background data values will be equal to this value.

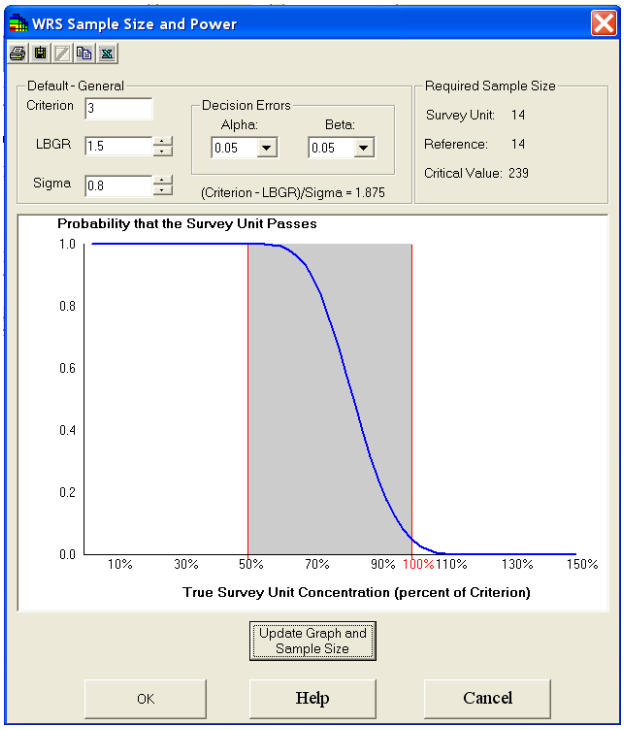
WRS Example

Open up the file SampleDesigns.sda. In the drop-down list of contaminant data sets, select *None*. In the interview drop-down list, select *Develop sample design*. Click on *Set sampling parameters*. At the top of the parameter window is a drop down list of possible sampling designs. We'll get to each of these shortly, including why we selected *None* as the contaminant. Right now, we'll just select a sample design that allows you to see where the choice for WRS is found. Select *Simple Random*.



Below your *Simple Random* selection, you will see a parameter block called *Number of Samples*. This is normally where you would specify what method you wish to use for determining the number of samples. Not all sample designs have this parameter block, as the calculation of number and location of samples is done at the same time. Here, select *Based on Wilcoxon Rank Sum* and press *Show The Results*.

You will next be asked for the decision *Criterion* (background in this case). Let's enter a value of 3 and press *OK*. The WRS test power curve is presented.



Notice in the upper-right-hand corner is the *Required Sample Size* information block. This contains the number of samples required in your *Survey Unit* (your site) and in your *Reference* area (e.g., background). It also shows the *Critical Value* to reject the null hypothesis. In this case, the sum of ranks must exceed 239 at the alpha level you've specified. Change the LGBR to 1.5 and press *Update Graph and Sample Size*. Spend some time playing around with each of the parameters to see how the number of samples and critical value is affected. At the top of the window is a toolbar with the usual export features. When you are done, press *OK* and SADA will distribute your samples randomly throughout the site.

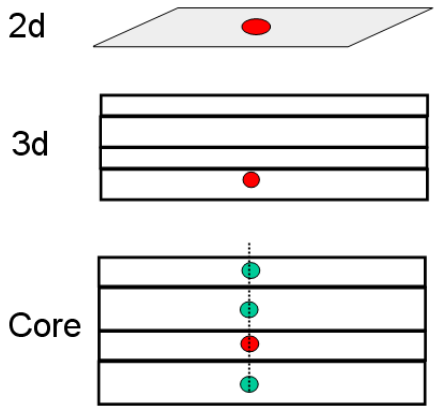
Note that you can also access this feature independently of spatial distribution. Select *Statistics→Number of Samples→Wilcoxon Rank Sum*.

You Pick

In this option, there is some other type of driver for the number of samples. Perhaps an external algorithm was used to determine the number. Perhaps there are economic limitations. In this option, you simply specify how many samples you wish to take, and they are spatially distributed according to the sample objective.

Sample Placement: 2d, 3d, and Core

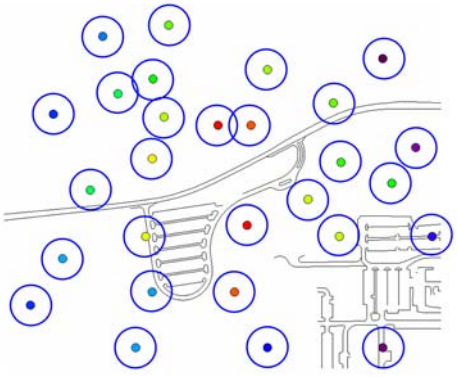
Whether initial or secondary sampling, SADA recognizes three methods for placement of samples in space. In a 2d placement, we are talking about placement only on the vertical layer currently selected (usually the surface). In a 3d placement, any layer becomes a candidate for a new sample. This works well when thinking about placing a groundwater well that will terminate/sample at some depth. In a *Core* placement, every layer directly above or below a sample placement is also included in the sample design. This works well when you are intending to bring up a soil core and sample at various intervals along its length. Placement within a layer has two options for all sample designs: middle of the layer or top of the layer.



Minimum Distance Constraint

In some sampling designs, it is possible for samples to inherently cluster together in order to meet some optimization criteria. For example, in a high-value design (secondary), the objective is to sample in those areas of high-concentration estimates. Algorithmically, samples are placed at those nodes that have the highest modeled values. This seems to be the right thing to do, but in reality the highest valued nodes are sometimes adjacent to each other and often near the highest sampled value. In many cases, the objective of the user is really two-fold: 1) sample where the model values are high and 2) spread new samples throughout the high value region to provide good spatial coverage. By specifying a minimum distance constraint, you can create “no sample zones” around existing samples and potential new sample locations alike. While the sampling scheme seeks to place samples in the highest areas, it is also met with a repulsive constraint that forces samples to spread out.

The minimum distance constraint is a circle defined by a simple radius. Anything inside the circle is off limits to new sample locations.



Tie Breakers

There are times when multiple new sample locations fulfill the goal of the sample design. In this case, a decision must be made so that only one of the new sample locations is chosen. SADA has three types of tie breaker methods.

Random

A random number generator is used to select one of the sample locations. The seed is used to feed the random number generator. Using a seed allows the user to recreate a resulting sample design by entering the same value.

Maximize spatial coverage

The value that will maximize the spatial coverage of the site is chosen as the new sample.

Closest to center of site

The value that is closest to the center of the site is chosen as the new sample.

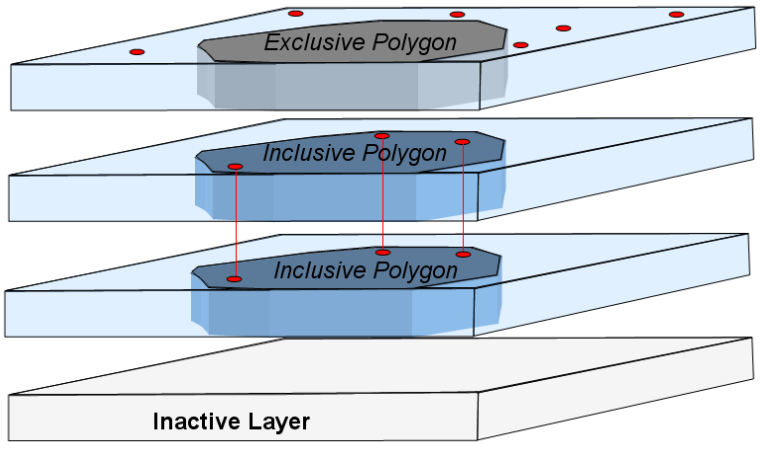
Ghost Samples (visualization)

A final option common to most designs is the ability to see the location of samples that may be positioned at depths above or below the vertical layer you are on. If you elect to see these samples, they will appear as hashed images (ghosts) instead of solid images.

Polygons and Vertical Layers

Polygons, ellipses, and rectangles all affect the allowable space in which samples may be positioned. If the area inside one of these objects is not included, then no samples will be positioned there.

If a vertical layer has been turned off, then no samples will be placed there. By using a combination of active and inactive layers along with inclusive and exclusive polygons, it is possible to fairly well control the space in which samples may be placed.



We will now turn our attention to describing each sample design separately. One might think that we would start with initial sample designs; however, in order to continue the decision analysis discussion in an unbroken fashion, we will first turn to secondary sampling designs. Secondary designs are based on the decision frameworks covered in the last couple of chapters. In addition, understanding how they work will prepare you for some of the targeted initial designs that are discussed afterwards.

Chapter 38: Secondary Sampling Designs

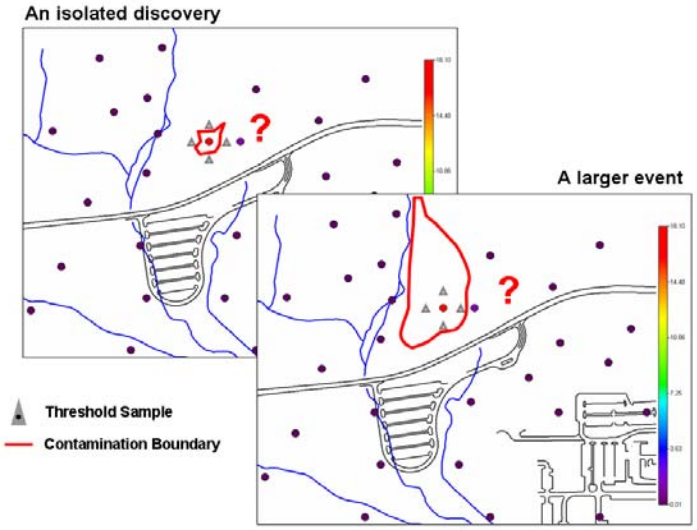
Secondary sampling designs are applied after some data or other information has already been obtained. Generally, the objective of any secondary sampling design is to further refine the model or the decision in some very specific way. Secondary designs can be either point (sample) or model (geospatial model) based. Before reading this chapter, please make sure you've read the previous chapter "Overview of Sampling Designs," Chapter 37. A number of important concepts are covered there that will only briefly be discussed here. We will now present each of these sampling strategies. *Note: If you are looking for initial sample designs, you must have (None) selected in the data set box.* Please open the file SamplingDesigns.sda and we'll begin.

Judgmental Design

This design can be classified as either initial or secondary and relies completely on the user to place samples where they wish. Often there are sampling objectives that cannot easily be automated and certain occasions warrant a strict placement based on professional judgment. We are not going to discuss the details of the judgmental design here. Rather they are covered in the chapter on initial designs.

Threshold Radial

Threshold radial (also known as Adaptive Cluster Sampling) is a straightforward secondary sample design that places samples in a radial pattern around existing data points that exceed a decision threshold. The user has control over the pattern of the surrounding new sample points. They can be circular or rectangular. Threshold radial can be useful in situations where you have a lot of very low or undetected samples and one or two very high measurements. You can encircle these points with new samples to help determine whether the result is an isolated event or the edge of a larger contamination event.

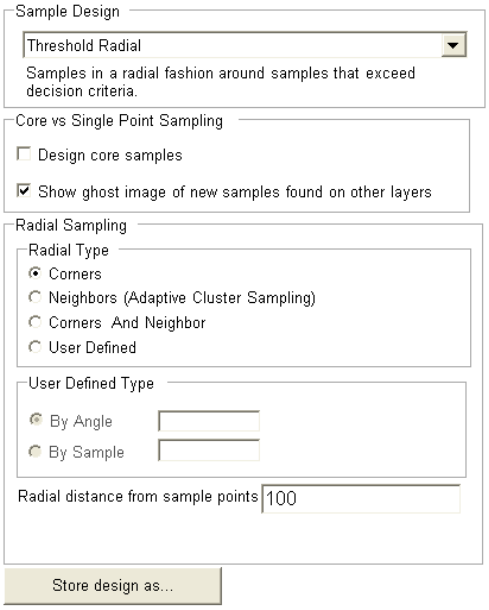


When you execute the threshold design, SADA will ask you for a decision criterion. Those measured values that exceed this criterion will be the target of the sampling design. Normally, this decision criterion is a single value, typically the screening value (see chapters on performing data screens); however, recall that with *General* analysis it is possible to specify depth-specific decision criteria. In the example that follows, we will use a two-dimensional

(surface) example, so depth-specific criteria isn't necessary. Please see the Chapter 13, "How to Perform a Quick Spatial Data Screen," for more information about where and how to set up and use depth-specific criteria. The following configurations are possible:

Name	Appearance
Corners	
Neighbors	
Corners + Neighbors	
User Defined	

For the file SampleDesigns.sda, select *Develop sample design* as the interview and make sure you have *Soil* and *Ac-225* selected. Click on the *Set sampling parameters* step. In the drop-down list under *Sample Design* at the top of the parameters window, select *Threshold Radial*.



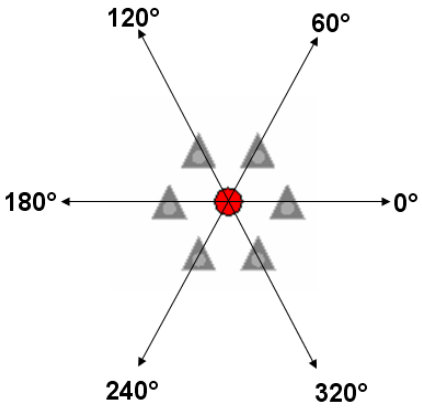
This is a 2d example, so we will not elect to design core samples in the *Core vs. Single Point Sampling* parameter block. Under *Radial Sampling*, select *Corners*. Enter a value of 100 into the *Radial distance from sample points* parameter box. This is the distance of new samples from existing sample(s) in exceedance of the decision criterion. If you wish to use depth-specific criteria (we don't right now), you can select this option under *Set decision threshold type*, found in the next step. Press *Show The Results*.

Since we are under the *General* analysis, SADA will now ask for the decision criterion to be manually entered. Had we been under the *Human Health*, *Ecological*, or *Custom* analysis settings, we would receive a window with options for selecting appropriate scenarios for each of these. If you have read the chapters on screening point data, you should be familiar with these

by now. Enter a value of 4.8 and press *OK*. SADA presents the *Corners* results and reports that only one value exceeded our decision criterion. New samples appear as triangles in the image below.

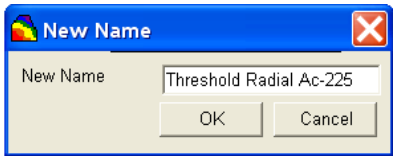
Now try selecting *Neighbors* and *Corners And Neighbors* on your own.

When you use *User Defined Type* arrangements, you can specify new samples in two ways: *By Number* and *By Angle*. If you choose *By Angle*, SADA will place a new sample each X degrees where X is the specified angle increment. If you choose by number, then SADA will first divide 360 by the number of samples you wish, and then place each sample according to this increment. In the following image, we've specified an angular increment of 60°, resulting in 6 new samples.



Select *User Defined* in the *Radial Sampling Block*, *By Angle* in the *User Defined Type*, and enter 60 for the angle increment. Press *Show The Results* and enter 4.8 as the decision criterion. SADA reports that 1 sample exceeded the criterion and 6 new samples are generated.

You can now store this result by pressing *Store design as…* and entering "Threshold Radial Ac-225" as the *New Name*.



To see this result, select *Stored Results* from the drop-down list of available *Sample Designs*. *Threshold Radial Ac-225* should be the only one. Press *Show The Results* (or *Show Me* button in the parameter window) and SADA generates the plot again.

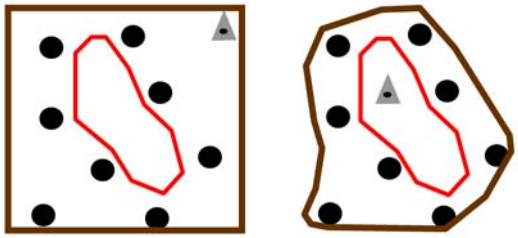
Adaptive Fill Design

In this sample design, samples are placed in the largest spatial gaps among data points. Unlike *Threshold Radial*, this design gives no regard to the measured values, only their relative positions. A set of new sample candidates is defined by a grid (much like a spatial model) that overlays the data points and acknowledges site boundaries, polygons, and whether layers are active or not. From this set of N candidates, the first winning location is simply that value which has the maximum distance to its closest neighbor. The design searches for the second location among the remaining candidates by comparing with the N+1 locations. If there are ties among the two locations, then the tie breaker method is used (see Chapter 37). The process repeats until one of the following becomes true:

- The total number of samples has been located
- There are no remaining candidates
- No remaining candidate satisfies the minimum distance constraint.

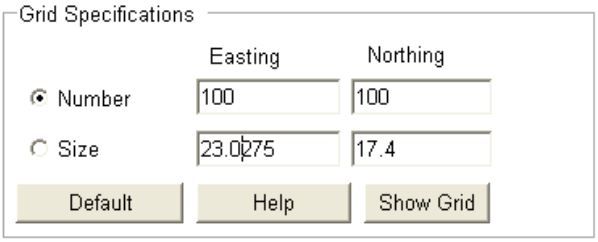
The number of samples is strictly determined by the user. The sign test and WRS test have no bearing here, as the objective is to fill spatial data gaps and not to test any hypothesis about the median.

This design requires that special care be taken in defining the study area. If the study area is arbitrarily chosen, then spatial data gaps map appear artificially along the boundaries. Consider the following example.

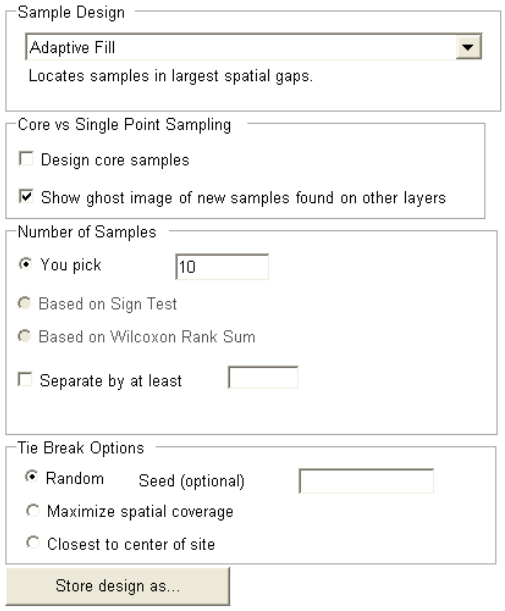


On the left we have a set of data points collected such that an area in the center has been under-sampled (red boundary). We want to use the adaptive fill method to place one more sample in this area to create optimal spatial coverage. However, because of our rigid rectangular definition of the site boundary (thick brown line), SADA places the new sample (gray triangle) in the upper-right-hand corner. If this truly is the study area, then the placement of the new sample is correct; however, if a more reasonable boundary (such as the one on the right) is in place, the adaptive fill sample will be placed appropriately in the interior spatial gap.

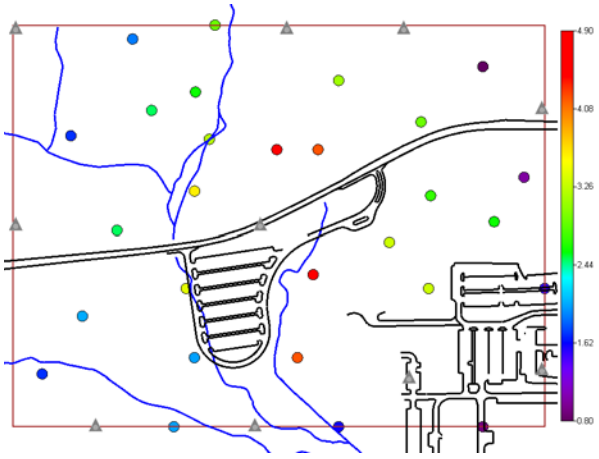
In the *Set sampling parameters* step, choose *Adaptive Fill* from the drop-down list of available design strategies. Click on the *Set Grid specs* step. This will define our underlying grid of candidate points. We'll use 100x100. If your grid is too coarse, your results may be less than optimal.



Click back on the *Set sampling parameters* step. We'll place 10 new samples and we won't require a minimum distance constraint. If there are ties, we'll opt to decide by random draw. We'll leave the *Random Seed* blank. This means that if there are ties, then the tie may be decided each time we reapply the design. Note that it is difficult to determine if ties are occurring just by applying the design. Your parameter window should look like this.



Press the Show The Results button.



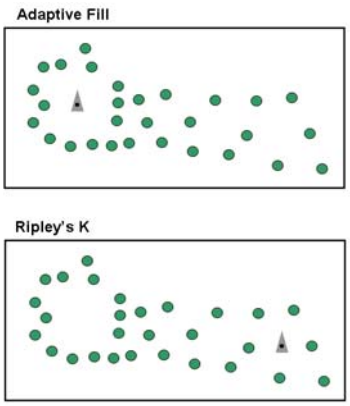
The results seem to favor the area along the site boundary. To see the site boundary, click on the *Set up the site* step and de-check *Hide site boundary*. This is consistent with what can happen with arbitrarily selected, rectangular site boundaries. If you wish to see the new sample

a little better, you may try to apply different colors (*Graphics→Set Various Colors→New Samples*).

Let's export these results to .csv file now. Click on the Export to Text button (). Enter the name "AdaptiveFill" into the file name box. Click *Save*. Open Excel and navigate to this file and open it. Now let's copy this graphical result to Microsoft PowerPoint. Open PowerPoint now. Return to SADA and click on the Copy to Clipboard button (). Return to PowerPoint and choose *Edit→Paste*.

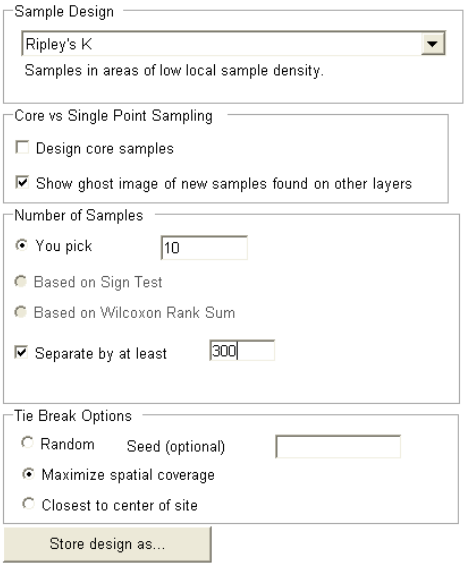
Ripley's K

This sampling design is based on the Ripley's K map. The Ripley's K statistic is a measure of neighborhood sampling density and is discussed in Chapter 42, "Local Index of Spatial Association Tools (LISA)." If you have not read that chapter yet, you are encouraged to do so now. Ripley's K value is evaluated at a node in the grid by specifying a simple search neighborhood (anisotropic neighborhoods not yet available) about that node and assessing the number of data points found there. Repeating this for each node creates a complete map. The Ripley's K design locates samples in those areas with the lowest sampling density. This is really in principle an extension of the objective in *Adaptive Fill*; but rather than base the answer on who is the furthest from their closest neighbor, the location is based on that node with the lowest sampling density in the nearby vicinity. In the following image, *Adaptive Fill* would be drawn to the slightly-larger empty hole found on the left side. Ripley's K would avoid this area with a higher sampling density and concentrate in a less densely sampled area on the right.



Ripley's K may require use of the minimum distance constraint, as nodes of low sampling density will likely be clustered together. Previous versions of SADA simulated the placement of Ripley K values to offset this problem. This approach will be reinstalled into SADA in the near future. In the meantime, we must rely on the minimum distance constraint.

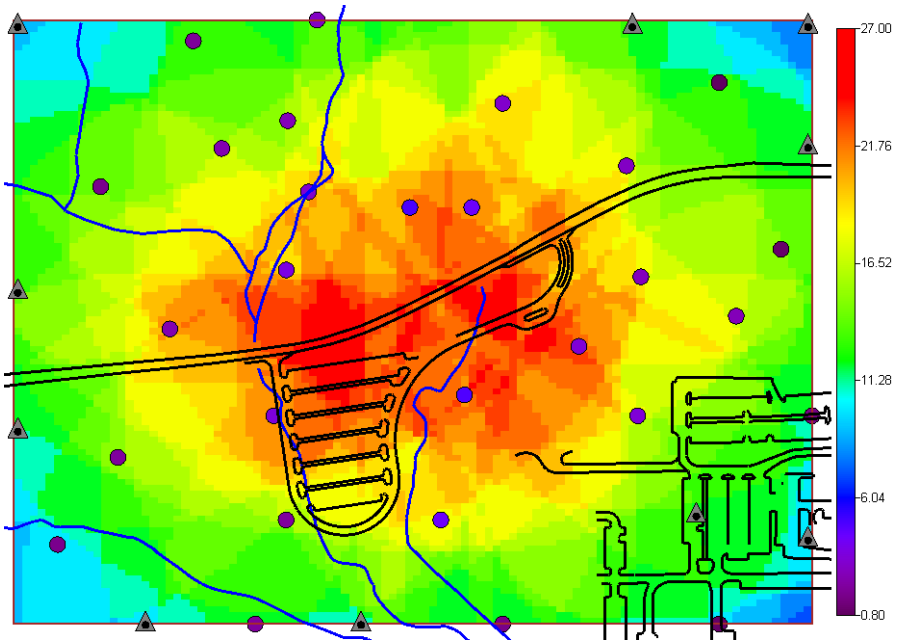
Click on the *Set sampling parameters* step and select *Ripley's K*. We'll place 10 samples and we'll force them to be a minimum of 300 feet apart. If there is a tie, we'll choose the location that maximizes spatial coverage (this method is really a one-time application of *Adaptive Fill*, where the sample in the largest data gap will be selected). Your parameter window should look like the following image:



We'll also need to specify the LISA search radius. Click on the *Set LISA parameters* step. Enter 1000 into the *Search Radius* box.

Note: At the time this text was written, a bug appears in the interface at this point (Version 5.0.78). You have in this window the option to choose between Ripley's K, Moran's I, and Geary's C. This is extraneous information, and when the Ripley's K sample design is selected, these options should not be presented. This will be corrected in a future version. Please return to the website and check for updates.

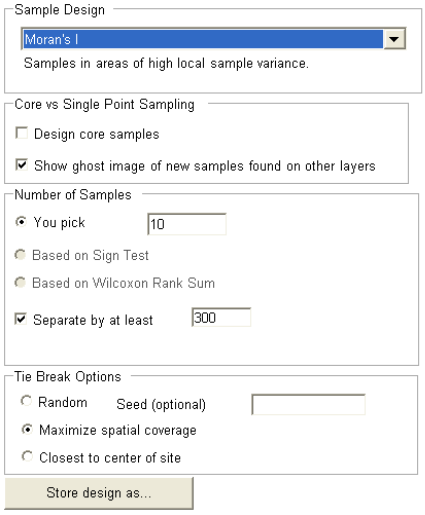
We'll also need to make sure we've specified the grid. From the previous exercise, you should still have a grid definition of 100x100. Check that now. Press *Show The Results*. SADA places the 10 new samples over the top of the underlying Ripley's K map from which these locations are based.



Moran's I

This sample design places samples in those areas of high local sample variance as defined by the Moran's I map. The Moran's I map is discussed in Chapter 42, "Local Index of Spatial Association Tools," and should be read before continuing with this section. The idea behind this sample design is to collect more data in those locations where greater heterogeneity exists.

Click on the *Set sampling parameters* step and select *Moran's I*. We'll place 10 samples and we'll force them to be a minimum of 300 feet apart. If there is a tie, we'll choose the location that maximizes spatial coverage (this method is really a one time application of Adaptive Fill, where the sample in the largest data gap will be selected). Your parameter window should look like the following image:

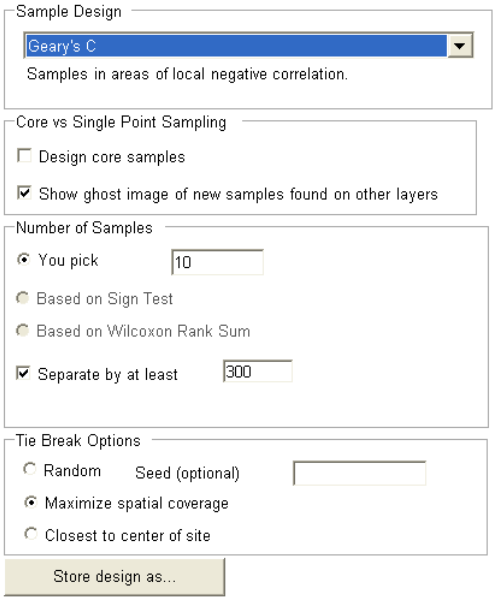


From the previous result, you should have the LISA search radius set to 1000 and a grid specification of 100x100. Press *Show The Results*. The new sample locations are plotted along with the *Moran's I* map on which each location was based.

Geary's C

This sample design places samples in those areas with greater (in magnitude) negative correlation among samples found in the search neighborhood. The Geary's C map is discussed in Chapter 42, "Local Index of Spatial Association Tools," and should be read before continuing with this section. The idea behind this sample design is to collect more data in those locations where greater heterogeneity exists. The difference between this approach and Moran's I is that heterogeneity is measured not by local variance but by local correlation. The more negative the correlation among data within the neighborhood, the more they are unalike.

Click on the *Set sampling parameters* step and select *Moran's I*. We'll place 10 samples and we'll force them to be a minimum of 300 feet apart. If there is a tie, we'll choose the location that maximizes spatial coverage (this method is really a one-time application of *Adaptive Fill*, where the sample in the largest data gap will be selected). Your parameter window should look like the following image:

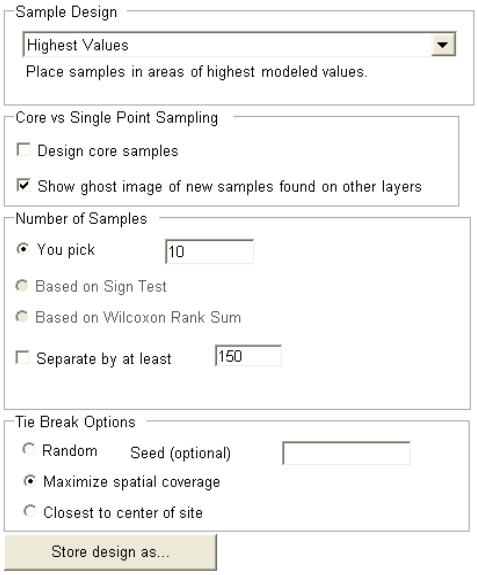


From the previous result, you should have the LISA search radius set to 1000 and grid specs at 100x100. Press *Show The Results*. The new sample locations are plotted along with the *Geary's C* map on which each location was based.

High Value

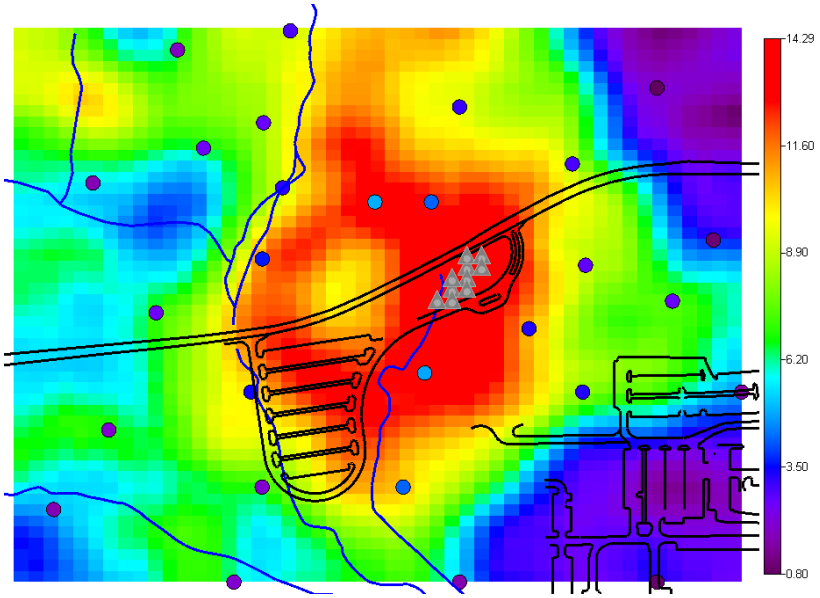
This sample design places new samples at nodes with the highest modeled values. This design can be applied to real-time models (calculated just before applying the design) and stored models alike. You will most likely need to use the minimum distance constraint, as high value samples will cluster around the highest sampled values. Typically, when one wants to sample in the high region, they do not want to locate samples strictly at the highest valued nodes. Rather, they also want some spatial coverage to spread samples throughout a high value area. Let's demonstrate with an example.

If you have not already done so, open SampleDesigns.sda. With *Soil* and *Ac-225* selected, choose the interview *Develop a sample design*. Click on the *Set sampling parameters* step and choose *Highest Values* as the *Sample Design* from the drop-down list of available methods. We will place 10 new samples (*You pick*) and choose to separate them by a minimum of 150 feet. For demonstration purposes, however, deselect at this time the *Separate by at least* option. This will demonstrate what happens if you elect not to use this secondary constraint. Your parameter window should look like the following image:

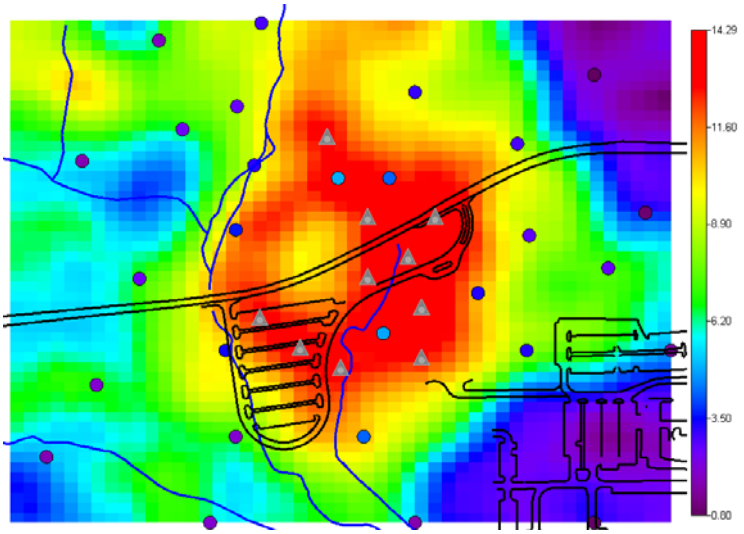


Notice that a number of steps that you've seen before when doing geospatial modeling appear. The next steps are *Specify Grid Parameters*, *Choose an interpolation method*, and so forth. In order to simplify the example, we'll choose a stored model so that we can dispense with calibrating a geospatial model as part of this exercise. While you certainly could calibrate a geospatial model under this interview, it is not advised. Largely, because it will simply slow you down. If you try to calibrate here, you will also have to wait while SADA attempts to place the new samples each time. Rather, select the interview *Interpolate my data* and calibrate the model there first. Those same parameters will then appear here as well. An alternative is to store your model or import a model from outside SADA. This is the approach we will use now.

Click on the *Interpolation methods* step. From the drop-down list select *User Stored* result. *Field Detection* is a geospatial model that we've previously imported for you. Notice that when this imported model is used, the steps list adjusts and no longer asks for calibration parameters from you. Press *Show The Results*.



You can see that strict adherence to choosing the highest value nodes results in severe clustering in one local area where several nodes are among the highest. Now recheck the *Separate by at least* option in the *Set sampling parameters* step. Press *Show The Results* once again.

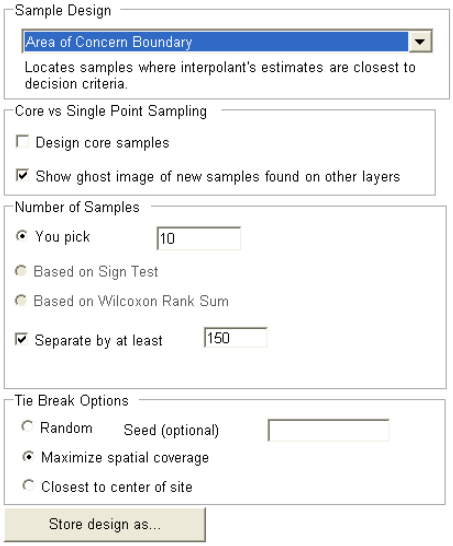


The result is a more reasonable distribution of new sample locations throughout the high value region, providing a balance between strict mathematical optimization and practical sample design.

Area of Concern

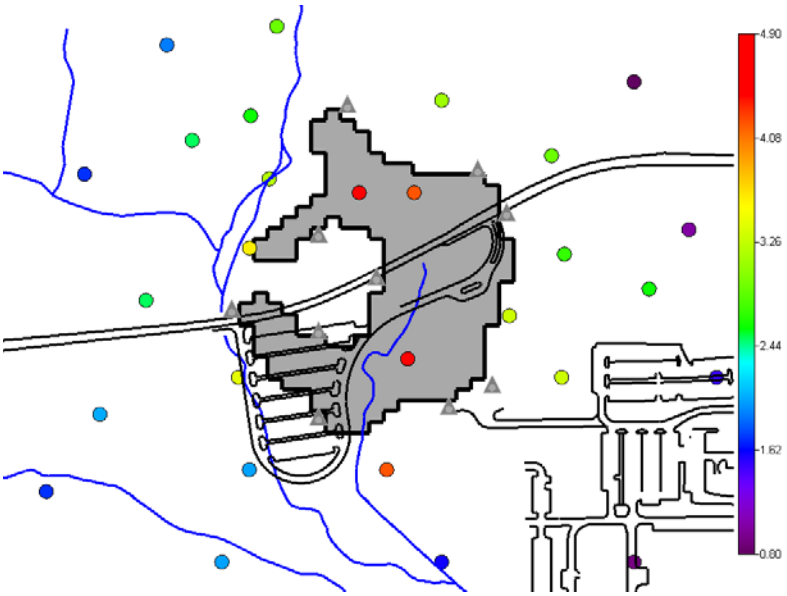
This sampling design places samples along the boundary line in the AOC result. In particular, those nodes that have a value closest to the decision criterion are the targets of the design. They are selected in order to more readily distinguish between contaminated and uncontaminated zones. Area of Concern maps are discussed in Chapter 35. If you have not read this chapter yet, you should do so before proceeding.

In the file SampleDesigns.sda, select *Soil* and *Ac-225* and the interview *Develop a sample design*. Click on the *Set sampling parameters* step and choose the *Area of Concern Boundary* from the drop-down list of available designs. We will place 10 new samples and will constrain them to be 150 feet apart. Without this constraint, there can be some clustering as observed with the high value design. If there are ties, we'll choose the one that is found in the largest data gap (*Tie Break Options* set to *Maximize spatial coverage*). Your parameter window should look like the following image:



This approach is based on an underlying geospatial model, and you will notice some additional steps have appeared that address this. In order to simplify the example, we'll choose a stored model so that we can dispense with calibrating a geospatial model as part of this exercise. While you certainly could calibrate a geospatial model under this interview, it is not advised. Largely, because it will simply slow you down. If you try to calibrate here, you will also have to wait while SADA attempts to place the new samples each time. Rather, select the interview *Interpolate my data* and calibrate the model there first. Those same parameters will then appear here as well. An alternative is to store your model or import a model from outside SADA. This is the approach we will use now.

Click on the *Interpolation methods* step. From the drop-down list, select *User Stored* result. *Field Detection* is a geospatial model that we've previously imported for you. Notice that when this imported model is used, the steps list adjusts and no longer asks for calibration parameters from you. Press *Show The Results*. Enter a value of 12 into the decision criterion (see the decision analysis chapters for all possible decision criteria approaches under *Human Health*, *Ecological*, and *Custom* analysis). The following result is produced:



Notice that new samples closely follow the AOC boundary line, and they are spaced out rather evenly when the minimum distance constraint has been imposed.

Now practice some with exporting these results to an Excel friendly CSV file and copy this image into PowerPoint. Finally, store this sampling design as Field Detection AOC (hint: look at the threshold radial example above for a quick review on each of these).

Stored Results

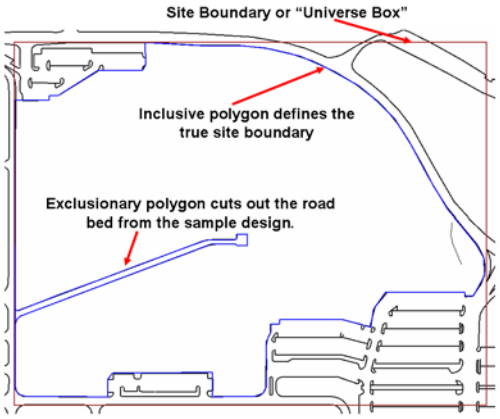
As you have seen in the previous exercises, it is possible to store any sample design by pressing the *Store design as….* button in the parameter window of the *Set sampling parameters* step. Recall that even though this stores your result, your SADA file is not stored until you save it. To use a previously stored result, simply choose *Store Results* from the drop-down list of available sample designs found under the *Set sampling parameters* step. It is important to note that when the sample designs were created they will have been created, with a particular vertical layering scheme and/or polygon scheme in place. If you now attempt to recall the design with a different vertical layering scheme, you may not see all the points in the design at different depths. So be careful about that. Once you have selected your stored result, simply press *Show The Results* and the result is restored.

Chapter 39: Standard Initial Sampling Designs

SADA provides a set of classic initial sampling designs with various endpoint objectives. These designs in one way or another assume little or no knowledge about site conditions. Some are exploratory in nature, some are searching for contamination, and some meet a statistical objective. SADA allows you to administer targeted initial designs that use some spatially explicit prior knowledge about the site. These methods are very similar to secondary designs and are discussed in the next chapter. If you have not already done so, please read Chapters 1-4. These are important in understanding how SADA files are created, setup, and managed. You will also need to have read Chapter 37, "Overview of Sample Designs." This covers important concepts in establishing designs in 2d and 3d space.

Note: Initial sample designs are only available under the Develop a sample design interview, when (None) has been selected as the current data set. Otherwise, secondary designs are presented.

We now present initial sampling designs found in SADA. Each of these designs will use the file InitialDesigns.sda. This is a SADA file that has been created and set up already for you. If you are not familiar with how this is done, you should revisit Chapters 1-4. Please open this file now and we'll point out a couple of things about it.

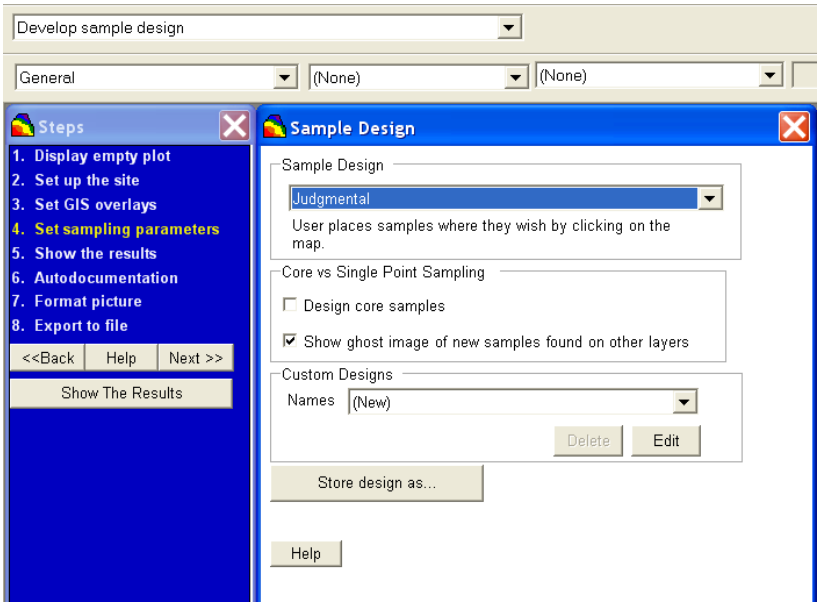


We have imported a layer showing roads in the vicinity and have overlaid this with the *Site Boundary* box (the brown rectangle). If you don't see the *Site Boundary* box, you can unhide it in the *Setup the site* step. This is really the so-called "universe box" that defines the maximum extents of the site. Within this box, we carved out the "true" site boundaries using polygon tools that can accurately trace the curvature of the site.

Judgmental Designs

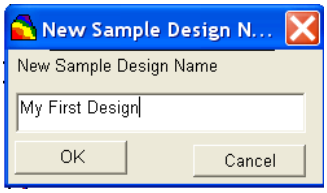
The judgmental design feature allows you to place samples exactly where you think they should be. Often, there may be sampling objectives that are not met by any of the initial or targeted designs in SADA. You can add your samples in a graphic, interactive manner, and then export them out to files for others to use in the field. Judgmental designs are also included in the list of secondary sample designs. The best way to learn is to simply start using the design. If you have not already done so, open the file InitialDesigns.sda.

You can see that the site has already been set up for us. This is an "empty" SADA file in the sense that no data base has been imported into it yet. In the interview drop-down list, select *Develop sample design*. Click on *Set sampling parameters*.

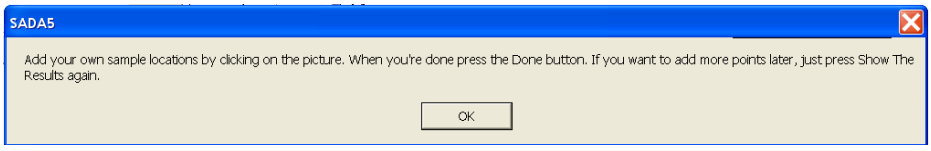


At the top of the parameter window is a drop-down list of available initial sample designs. Since *Judgmental* is the first in the list, it is the default selection. Below this is the *Core vs. Single Point Sampling* option. We are working in 2d, so this is not relevant to us. Below this is the *Custom Designs* box. The *Custom Designs* drop-down list contains any previously created *judgmental* designs. At the moment, there are none and our only option is *(New). Note: this is a different feature than the Stored Results feature you'll see later on. Custom Designs stores only judgmental designs that can be reopened and edited.*

With *(New)* selected you can either press the *Show The Results* button or click on the *Edit* button below the *Custom Designs* drop-down list. Press *Edit* now. We'll need to provide a name. Let's enter "My First Design" and hit *Ok*.



You are now in edit mode and SADA will provide a friendly reminder about how to add your samples.



Press *OK*. Notice when you are in judgmental edit mode, most of SADA becomes disabled. This is normal because it wouldn't be good for you to start some other action in the middle of editing. Notice that the *Edit* button has become a *Done* button. When you are finished with your design, press the *Done* button and SADA will become enabled and your design will be stored (keep in mind nothing is really saved until you save your SADA file).

Adding Samples

Adding a new sample is easy. Simply left-mouse-click on the location where you wish the sample to go. Each time you click, a new sample is added. Add a few samples now.

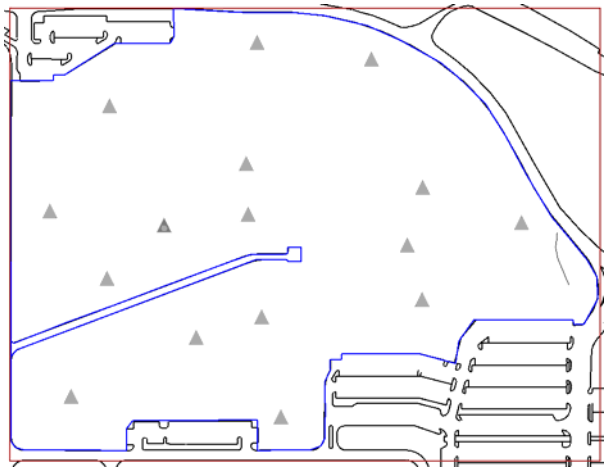
Deleting Samples

To delete a sample, left-mouse-click in the center of the sample. It will turn red. Press the Delete key and the sample is gone. Try this now with one of your samples.

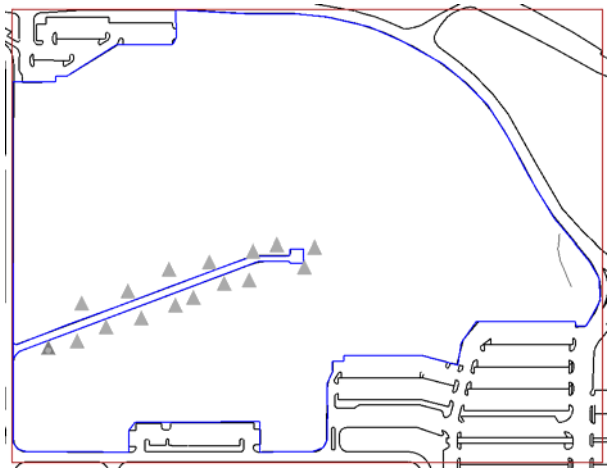
Moving Samples

To move a sample, left-mouse-click and hold on the center of the sample. It will turn red. While still holding the mouse down, simply move the sample to where you want it to go.

If you click outside the site boundary or in an exclusionary polygon, SADA will warn you but allow the sample to be added. When you are done, press the *Done* button. Do this now (your design will look different than the image below).



Now your result is stored as *My First Design*. Let's create another design called *My Second Design*. In the drop-down list of custom designs, select *(New)* and press the *Edit* button. Create another design. When you are done press *Done* (again, your design will look different than the image below).



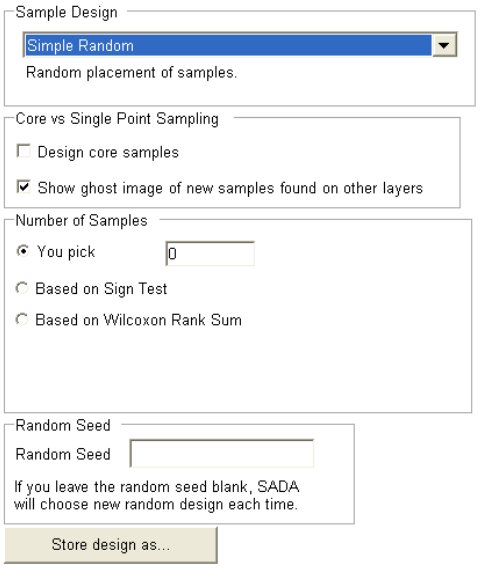
Now in the *Custom Designs* drop-down list, select *My First Design*. The design is immediately recalled. To continue editing it, simply press the Edit button. Editing works just as it did before.

It is possible to statically store your result as well. With *My First Design* selected, press the *Store design as ….* button. Enter "My First Design Finished" as the name. Now in the drop-down list of available initial designs at the top of the page, select *Store results* (at bottom of list). You will see the stored static copy of your design there. You can see it at any time by selecting

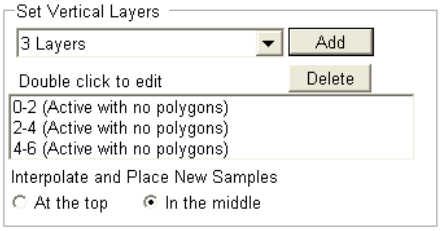
here and pressing *Show The Results* or *Show Me*. Of course you could continue editing this design under the *Judgmental* design selection, but “My First Design” and “My First Design Finished” are no longer connected. The latter is a permanent copy of what you finished to that point. Save your SADA file now to make all these changes permanent.

Simple Random

The *Simple Random* design will distribute a set of new samples in a random pattern across the site. The number of samples can be determined by three methods: *You pick*, *Based on Sign Test,* and *Based on Wilcoxon Rank Sum*. Each of these methods is discussed in Chapter 37, “Overview of Sampling Designs,” and will not be repeated here. To try the simple random design, select it from the drop-down list of designs under the *Set sampling parameters* step. A number of parameters become available.

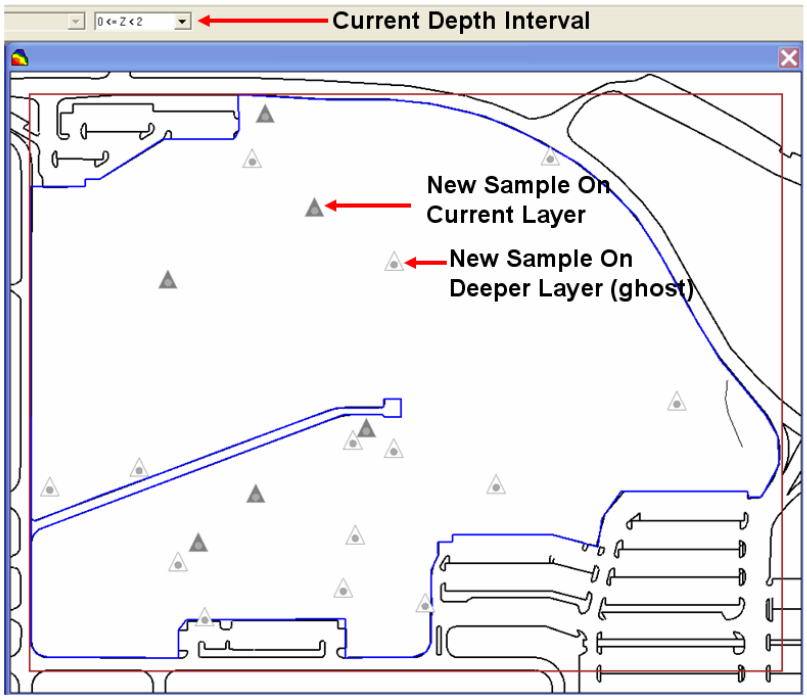


With this example, we will demonstrate the effect of core versus true 3d sampling. Click on the *Setup the site* step and select the *3 Layer* design under the *Set Vertical Layers* drop-down list.



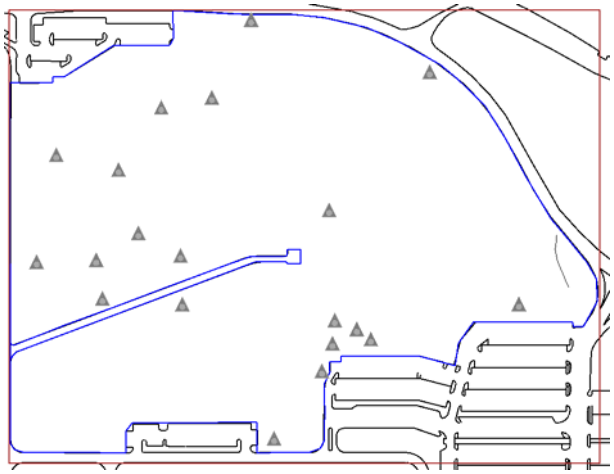
This is a vertical layering scheme of three layers, each only 2 feet thick. Let’s begin by applying the sample design without using the core option. This will randomly distribute the samples both horizontally and by layer. This later point warrants additional emphasis. In the vertical direction, layers are randomly selected, not depths. If a layer is selected, the sample is placed either in the middle of the layer or on the top of the layer (depending on what you specified in the *Set Vertical Layers* parameter block under the *Setup the site* step).

Deselect the *Design core samples* option. Select the *You pick* option and enter 20 into the adjacent parameter box. Without the *Design core samples* option on, this will generate 20 new samples randomly across three layers. Press the *Show The Results* button in the steps window.



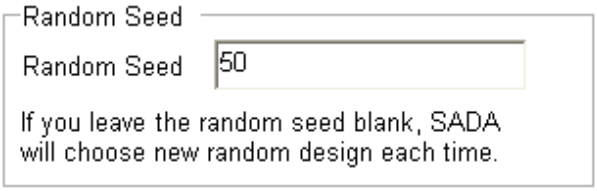
In the results window (may look different than yours), you will notice solid and empty triangles that represent new sample locations. The solid locations are samples that were distributed to the current layer (0<=z<2). The empty triangles are “ghost” images of samples located further down. If you switch to the (2<=z<4) layer, some of these will become solid while the current solid ones will become ghosts.

Without the *Design core samples* option selected, 20 samples results in 20 distinct sample locations at all different layers. If we now choose the *Design core samples* option, we will instead generate 20 cores at 20 random locations. The total number of samples produced will be the number of cores times the number of active layers. So in our case, with 3 layers, a total of 60 samples will be produced. Select this option now and 20 random cores are located.



Let’s switch now to the surface layer. Click on the *Set up the site* step and select *Surface only* as the current vertical layer design (under *Set Vertical Layers* parameter block). We’ll now discuss the role of the random seed. When a random number is drawn, it actually comes from a sequence of randomly generated numbers. Each number from the sequence is then mapped

into an (x,y,layer) coordinate. If you repeatedly push the *Show The Results* button, you will notice that the locations continually jump around. This is because each time the button is pressed, the next number in the sequence is selected, resulting in a new location. A random seed is just any positive integer you like that tells the random number generator where to start in the sequence. In reality, you will have no idea where your seed lands you in the sequence. What it does though is create a repeatable random draw. For example, choose some positive integer such as 50 and enter it into the random seed parameter box (back on the *Set sampling parameters* step).



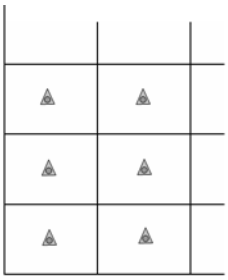
Press *Show The Results*. You will see a new design. Press *Show The Results* again and the same design is generated. There is no way to predict what the random design will be, given a seed value; it's just a way to repeat the same random draw. Another way is to simply press the *Store Design as…* button and save a static copy of the design to be used later. The design will appear when you select *Stored Results* as the current sample design.

Simple Grid and Simple Grid (unaligned)

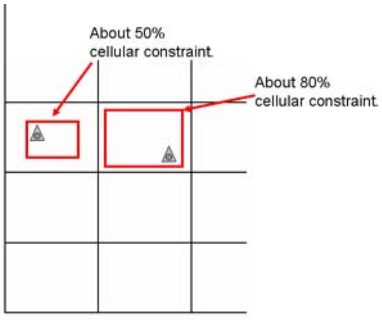
Many of the initial sample designs found in SADA are essentially grids. The difference between them is in how they are computed and what their objectives are. In a simple grid, you specify the number of samples you want to distribute in a grid-like fashion across the site. SADA first identifies a random start location somewhere on your site to locate the first grid node.

Second, the grid spacing must be determined. This is no trivial task when you consider irregular boundaries and the presence of exclusionary polygons. If the estimated grid spacing fails to adequately cover the site, the grid spacing is adjusted. In an inadequate placement, open spaces exist where additional nodes could be located or the spacing was too wide and not all samples could be included. This process of refining the grid spacing continues until SADA either arrives at a successful distribution of the grid, or the grid is impossible. The idea that a certain number of nodes could not be arranged in a grid may itself seem impossible, but consider this simple example. Suppose you had a perfectly square site and you asked for a square grid of 17 samples. There is no way that a square grid system can be created in a square area with 17 samples (16 would have been ideal: 4 x 4). If SADA encounters this problem, it will increase the number of samples until a grid can be accomplished. You will be informed if this happens.

There are two types of simple grids: simple grid and simple unaligned grid. In simple grid, nodes are identified with regular spacing, as we just discussed. The geometry of the resulting grid can be either square or triangular. If you think of each sample grid node as the center of a grid cell, the samples would look like the following image:

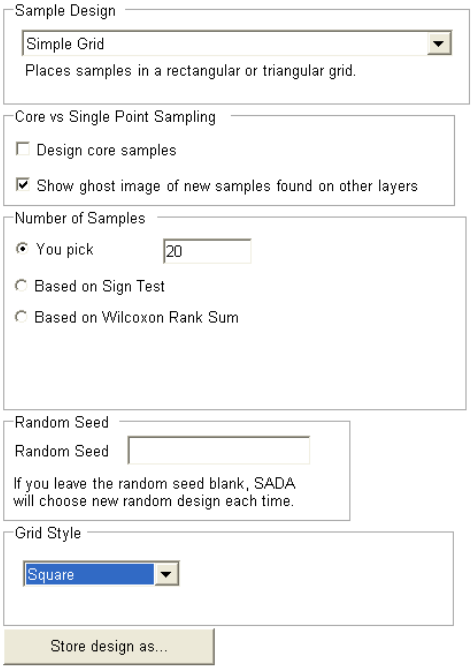


Simple unaligned grid randomizes the new sample location within each cell. You can control how close the randomization is permitted to come to the cell wall by specifying a “sub-cell” size. A sub-cell is simply a smaller cell located within the grid cell in which randomization can occur. The sub-cell parameter is specified as a percentage of the grid cell. If we plotted the sub-cells, they would appear as follows.

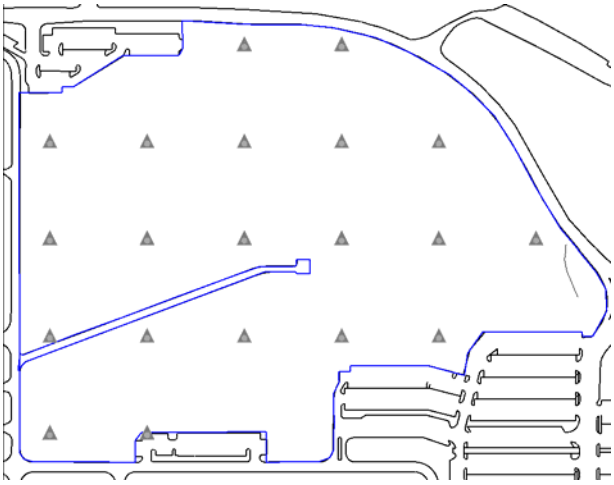


Any time a random location is selected, the random seed parameter becomes relevant. As previously discussed with the *Simple Random* design, when a random number is drawn, it actually comes from a sequence of randomly generated numbers. Each number from the sequence is then mapped into an (x,y,layer) coordinate. If you repeatedly push the *Show The Results* button, you will notice that the locations continually jump around. This is because each time the button is pressed, the next number in the sequence is selected, resulting in a new location. A random seed is just any positive integer you like that tells the random number generator where to start in the sequence. In reality, you will have no idea where your seed lands you in the sequence. What it does though is create a repeatable random draw.

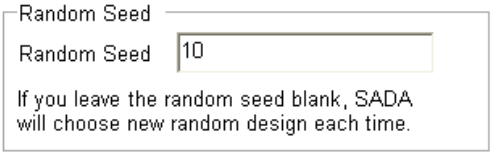
If you have not already done so, open the file InitialDesigns.sda, select the *Develop a sample design* interview, and click on *Set up the site*. On this step, make sure that you have the *Surface Only* vertical layer design selected. Click on the *Set sampling parameters* step and select *Simple Grid* from the drop-down list of available sample designs. In this example, we’ll place 20 samples in a square grid across the site. We won’t use a random seed at first. Make sure your parameter window looks like this and press *Show The Results*.



Your result may differ greatly from the following image. This is due to the randomization of the grid starting point each time you press the *Show The Results* button.

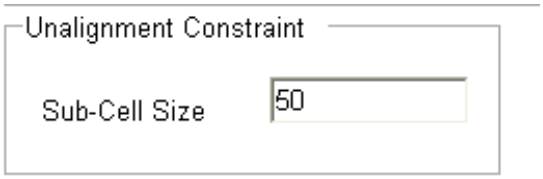


Now choose a seed value of 10 and press *Show The Results* a couple of times.



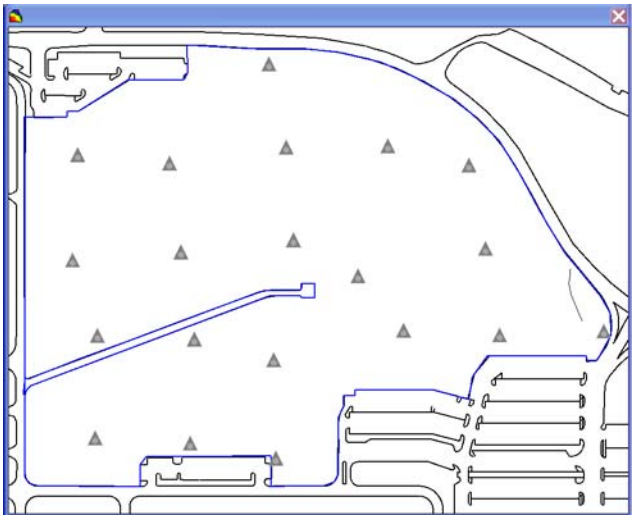
You will see that the random draw is repeatable. Next, try a Triangular grid (in the drop-down list of the *Grid-Style* parameter block). For extra practice, try storing a result by using the *Store Design as…* button. Can you recall where the stored design is kept? (hint: if you can't recall, revisit the discussion on Simple Random design).

Now, select *Simple Unaligned Grid* from the drop-down list of available sample designs. You will see only one additional parameter requirement: the *Unalignment Constraint* parameter.

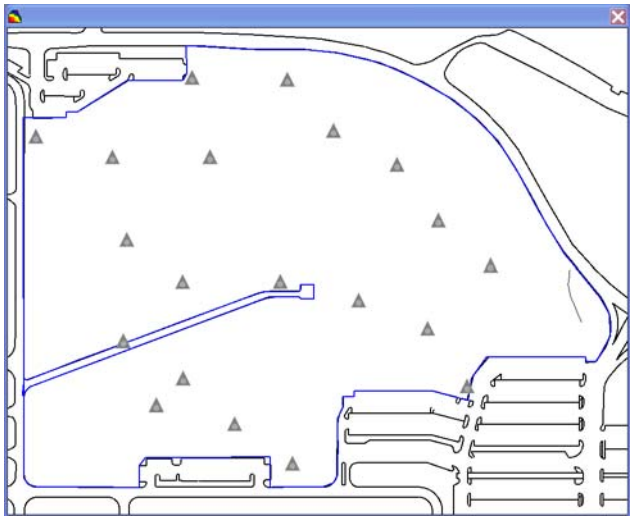


Keep in mind that this parameter is expressed as a percentage of the regular grid cell size. The default value is 50, which means that a sub-cell half the size (height and width) of the regular cell is centered at about the center of the grid cell. Randomization of the new sample point will occur only within this sub-cell.

You may also notice that you no longer have an option between square and triangular. For simple unaligned grid, the cell is strictly a square geometry. *Press Show The Results* and you'll see a slight perturbation of a regular grid.



Now try assigning a *Sub-Cell Size* parameter of 90. You may see clustering of points together as points are able to appear closer to the original cell border.



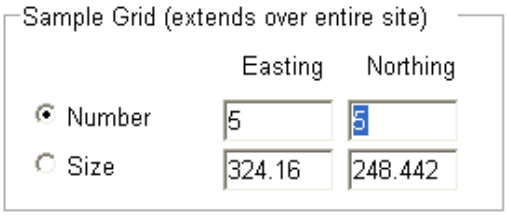
The higher the *Sub-Cell Size*, the more random the sample design becomes. The smaller the *Sub-Cell Size*, the more regular the design becomes.

Standard Grids and Standard Unaligned Grids

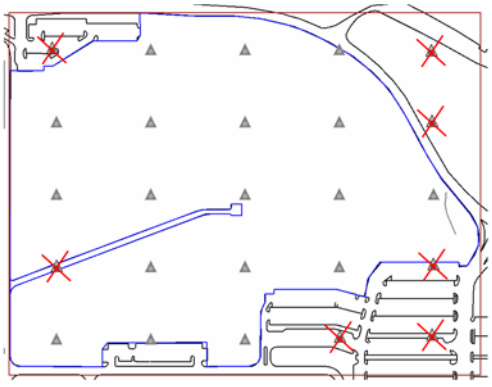
If you have not already done so, please read the previous section on simple grids. Standard grids and standard unaligned grids behave very similarly. There are two primary differences.

1) You specify the spacing of the sample nodes, either by stating node separation distances or by number of nodes in each direction.
2) The grid is not randomly positioned. Rather, the grid is centered as well as possible over the site such that the distance between the west side of the site and the first node in a row is about the same as the distance between the east side of the site and the last node of a row. The same is true for the N-S direction; however, the balance may not be exactly symmetrical, as irregular boundaries may require a reasonable balance.

Select *Standard Grid* from the drop-down list of available sample designs. You will see a grid definition parameter block. If you have read the geospatial chapters or secondary design chapters, this parameter set should be familiar to you. There are two ways to specify a grid: by number of nodes in each direction or by separation between nodes in each direction.

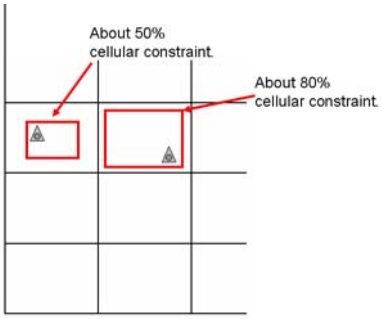


Either way, the grid is stretched over the entire site (the universe box or site boundary box) and then polygons are applied in order to either include or exclude points. Let's define a grid by number first. All grids in SADA have a default number of nodes equal to 50 x 50. As sample designs go, however, this is likely unreasonable (i.e., if your site has 2500 samples). Let's suppose we can afford to take 25 samples. In the *Sample Grid* parameter block, select *Number* and enter 5 into both the *Easting* and *Northing* direction. Press the *Show The Results* button. SADA reports that 18 samples have been identified. So, if we specified 25 (5x5) what happened to the other 7?

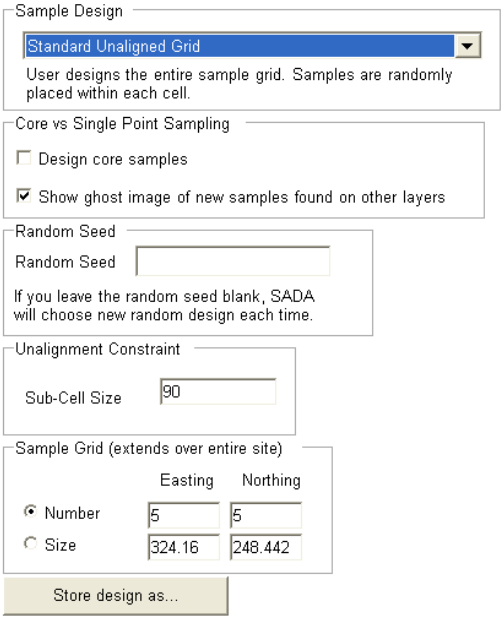


Recall that SADA's grid systems work by first stretching the specified grid over the entire site as defined by the rectangular site boundary box (usually brown). The sample design is then applied first and then any polygon or layering protocols are applied to eliminate samples from consideration. In the image above, those samples that were eliminated have been graphically added back in with red x marks across them. If you want 25 in your actual site, you'll need to either increase the resolution or perhaps use the simple grid method instead.

The standard unaligned grid operates in much the same way as the simple unaligned grid. A sub-cell is specified within each standard grid cell. The sample location is then randomized within the cell.

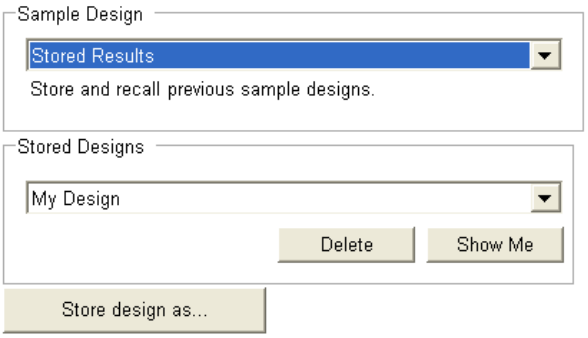


The sub-cell is defined as a size percentage of the original cell. As discussed in simple unaligned grid, the greater the percentage the more random the design becomes. The smaller the constraint, the more grid-like it becomes. As with all previous randomization methods, a random seed will be available to provide the option of a repeatable random draw (see more detailed discussion under Simple Random). Select *Standard Unaligned Grid* now.



You will see the *Random Seed* parameter and *Sub-Cell Size* constraint boxes appear again. Leave the random seed blank and set the *Sub-Cell Size* to 90. Press *Show The Results* and observe the strong perturbation in the grid structure. Now constrain the *Sub-Cell Size* to 50 and notice it returns to a more normal grid pattern.

Play around with the *Random Seed*. Choose some positive integers and notice that for any given integer, pressing *Show The Results* produces the same design again and again. If you leave this parameter blank, a new design is created with each execution. Try storing one of your designs by using the *Store design as…* button. Provide a name for the design and say *OK*. Your design will appear when you select *Stored Results* as the sample design type and select it from the drop-down list of available stored designs.



MARSSIM Design

The Multi-Agency Radiological Site Survey Investigation Manual (MARSSIM) is a comprehensive approach agreed upon by a number of federal agencies, including DOE, EPA, and NRC, for determining whether radiologically contaminated sites can be released to the public (www.marssim.com). Among the many methods and procedures described therein, you will find a set of sample design strategies that are followed up by a formal assessment framework. The importance of MARSSIM is such that the entire next chapter is devoted to this subject. If you are strictly interested in doing a MARSSIM sampling design, you should visit Chapter 40 now. Make sure you have also read the introductory Chapters (1-4) and Chapter 37, "Overview of Sample Designs."

Hot Spot Searches

The hot spot search designs choose sample node placement so as to optimally search for hidden objects (e.g., buried waste). These methods have a long history in the environmental field, much of which is captured in the list of references at the end of this chapter. The algorithms and technical details will not be repeated here. You are invited to take a look at those papers/books if you are interested at that level of detail. In this discussion, we will present a practical overview of the methods. You should have a good understanding of what each does and how to use them by the end of this section.

Note: this feature was originally encoded as Ellipgrid PC by the Oak Ridge National Laboratory and is available in the public domain in that form.

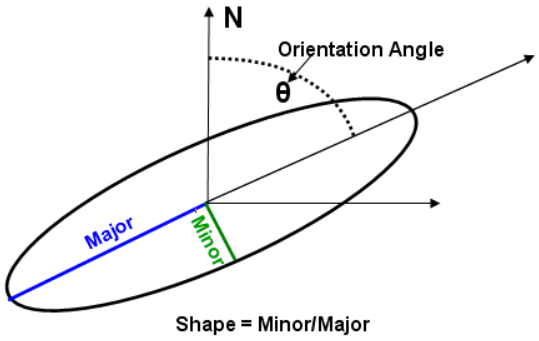
There are three factors considered in the search for an object:

Object Geometry

The object is assumed to be an elliptical shape with geometry specified by size, shape, and orientation. The shape parameter sets the ratio of the major elliptical axis to the minor elliptical axis. A ratio of 1 creates a circle. A ratio of 0.5 means that the ellipse is half as wide as it is long. In addition to the ratio, we need to specify the area of the ellipse. This can be done by specifying either a) the major radius length or b) specifying the area inside the ellipse. Converting between radius and area is a trivial calculation

$$Area = \pi \bullet \text{major} \times \text{minor}$$

where major and minor are the major and minor radii, respectively. If the object is not a circle, then its position can affect the outcome. You can specify a certain degree of rotation (from due north) or use random degrees. Under random degrees, any elliptical orientation is considered.



Grid Geometry

Grid density and arrangement (triangular or rectangular) are central features. The grid is actually established by specifying the separation distances between grid nodes. For a square or triangular grid, you only specify a single separation distance, as an equilateral triangle is assumed. For a rectangular grid, you can specify both an E-W and N-S separation distance.

Probability of Discovery

This is the chance of finding the object given grid and object geometry.

If the user can specify any two of these quantities, the code will calculate the third. These three states form the three named search strategies.

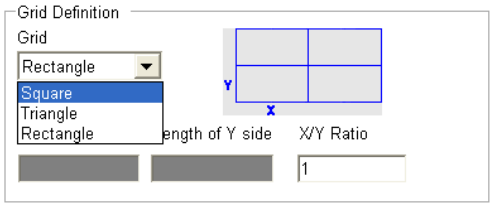
Minimize Sample by Hot Spot Definition	**Probability**	**Size & Shape**	**??**
Unknown Hot Spot	**Probability**	**??**	**Grid Specs**
Calculate Probability	**??**	**Size & Shape**	**Grid Specs**

Hot spot Search Examples

If you have not already done so, please open InitialDesigns.sda and select the *Develop sample design* interview. Click on the step *Set up the site* and make sure you have *surface only* selected as the current vertical layering design. Hot spot searches are two-dimensional only. If you have a multiple vertical layering design in place, SADA will apply the sample design to the currently selected layer only. Click on *Set sampling parameters*.

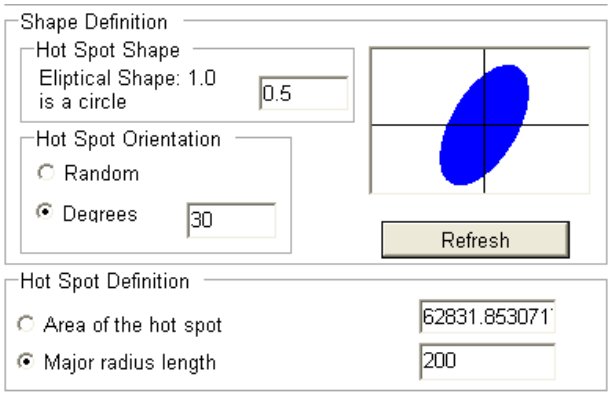
Minimize Sample By Hot Spot Definition Example

In the drop-down list of available sample designs, select *Hot Spot: Minimize Sample By Hot Spot Definition*. In this design, you will specify probability and object geometry and SADA will determine the grid specs. Now, you do have some say in the grid before the algorithm is run. You can specify a square, rectangular, or triangular grid.

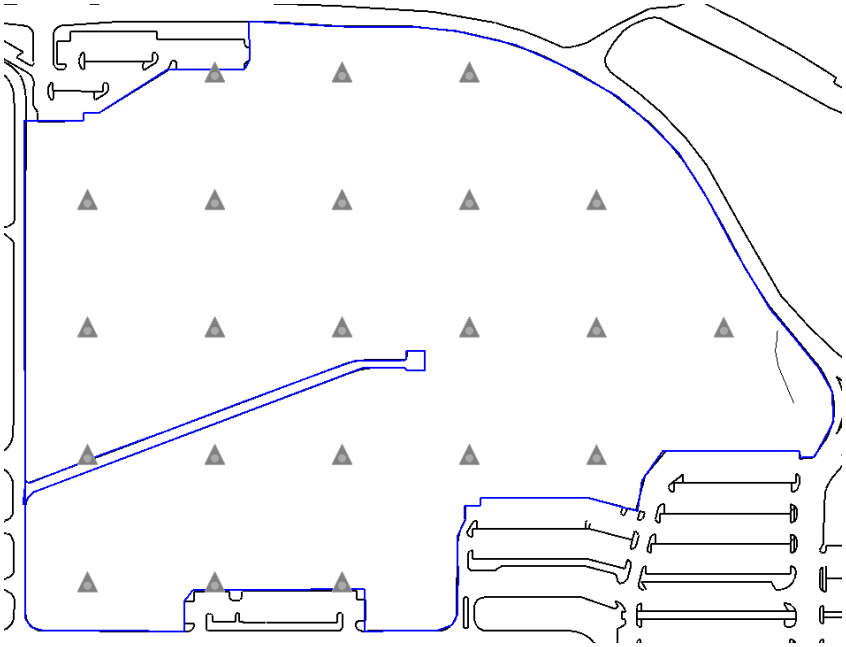


Notice that the Length of X side and Length of Y side boxes are grayed out for any geometry choice. This is because SADA will calculate these for you. If you select rectangle, you can apply some control by specifying the width to height ratio in whatever resulting grid is derived. Choose a *Square* geometry.

Under the *Shape Definition*, let's choose an *Eliptical Shape* with a ratio of 0.5. In the *Hot Spot Orientation* block, choose *Degrees* and enter a value of 30. Finally, under the *Hot Spot Definition* block, choose to specify the size by *Major radius length* and enter a value of 200. Notice that the corresponding area is entered into the *Area of the hot spot* line. Press the *Refresh* button to see how the ellipse will appear.



The last thing we will need to do is specify the probability of finding such an object on the site. Let's enter a value of 90 (%). Press the *Show The Results* button and 22 samples are distributed.



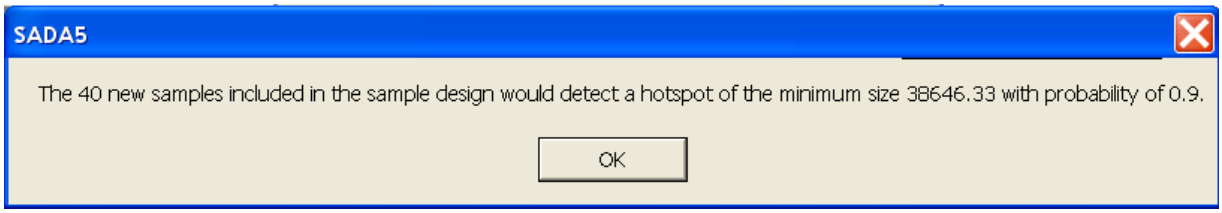
Now change the *Hot Spot Orientation* from only 30 degrees to *Random* and reapply. Notice that more samples are required to account for any elliptical rotation.

Unknown Hot Spot Example

Now select *Hot spot: Unknown Hot Spot* from the drop-down list of available designs. In this design, you can specify the grid and the probability of discovery. SADA will then determine the

smallest elliptical object you can find with stated probability given the grid you are using. Now, you do still have some control over the object in the sense of specifying its shape ratio and orientation. SADA will determine the size of the object.

In the *Grid Definition* block, choose a *Triangle* grid with a side length of 200. The units will be in feet because our shape file and corresponding site is in feet. We'll leave all the other parameters the same for both the limited object geometry and probability blocks. Press *Show The Results* and SADA will distribute 40 new samples (determined by the side length of 200) and report the smallest hot spot size (38646ft$^{2)}$ that could be found with 90% probability.

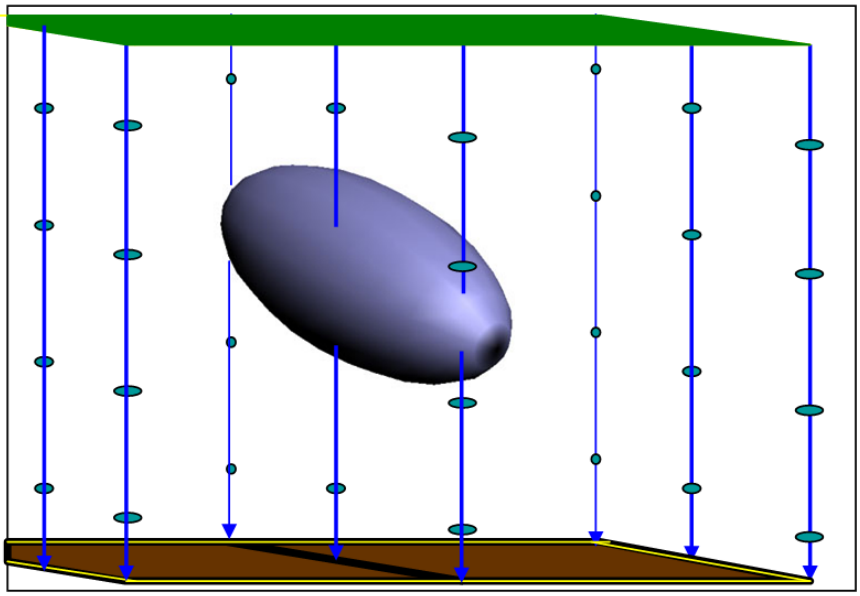


Calculate Probability Example

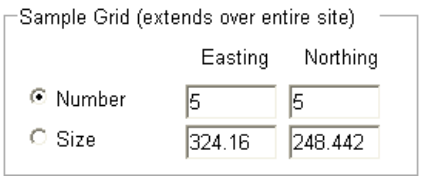
Now switch to *Hot spot: Calculate probability sample design*. In this design, you will specify the object geometry and the grid you wish to use. SADA will determine the probability that you could find such an object. Notice that the probability parameter block disappears completely. Let's use a rectangular grid with a height (Y) of 150. We'll leave the shape geometry largely unchanged but switch the size defined by the major of axis parameter value of 200 to 150. Press *Show The Results* and SADA reports the object would be found by such a grid about 90% of the time.

3d Hot Spot Search

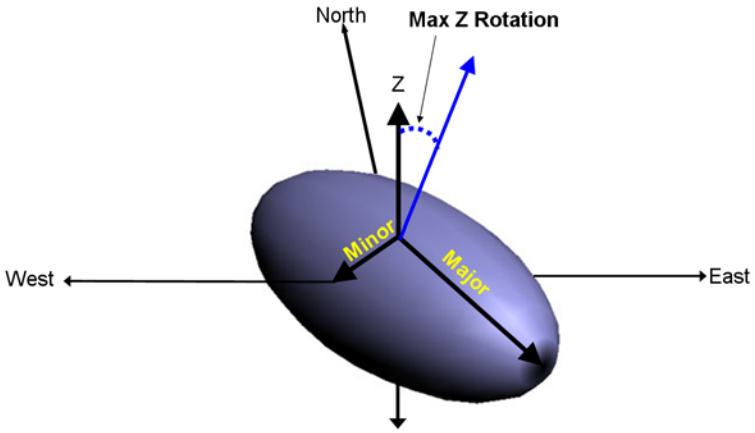
The 2d search algorithm has been extended into 3d for one of the three scenarios: calculate probability. In the 3d hot spot search, you will specify the 3d grid and the 3d object and SADA will determine the probability of discovery.



The grid definition is specified as a standard grid. You can choose to set grid spacing either by number of nodes or by actual spacing distance itself. The vertical frequency of the search grid is specified by the number of vertical layers.



The object geometry is specified by you as well but in a slightly different way. In this formulation, you will specify the x radius, y radius, and vertical radius. You can also specify a maximum z rotation. This will limit the ellipsoid's vertical rotation range to be more in line with environmental geometries (e.g., plumes).

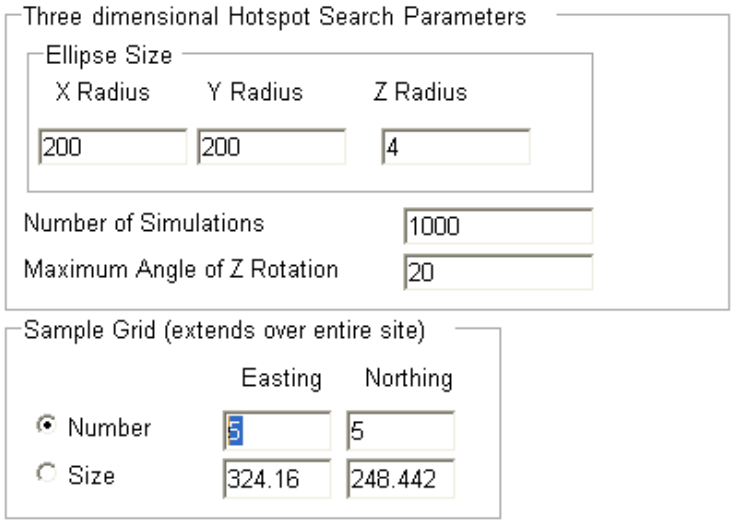


For 2d searches, the solution is a closed form. For 3d, simulation of ellipsoids is used as a numerical approach in calculating the probability of discovery. Simulations of 500-1000 should not be a problem.

3d Search Example

If you have not already done so, please open InitialDesigns.sda and select the *Develop sample design* interview. Click on the step *Set up the site* and make sure you have *3D layers* selected as the current vertical layering design. The vertical layering design is important because it determines the vertical density of your 3d grid.

Click on the *Set sampling parameters* step and select *3d Hot Spot Search* from the drop down list of available initial designs. We will calculate the probability of finding a circular hot spot in the shallow subsurface. The geometry will be a major and minor radius both equal to 200 and a thickness of 4. We will run 1000 simulations and restrict the vertical dip of the ellipse to be no more than 20 degrees. We will search for the ellipse with a standard grid defined by number of nodes. We will lay down a 5 x 5 grid for a total of 18 samples (after accounting for boundaries). Make sure your parameter window looks like the following image:

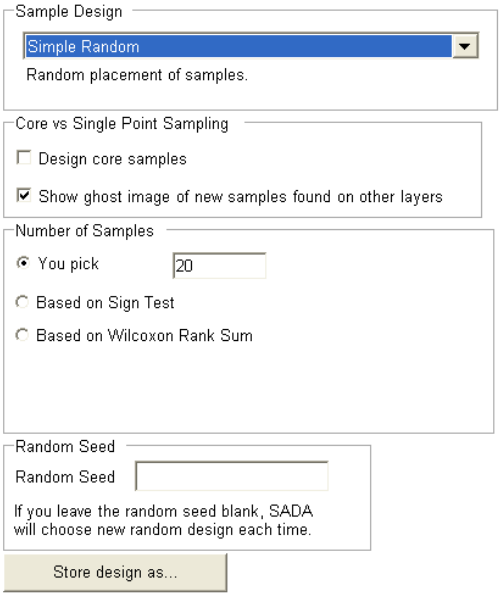


Pres the *Show The Results* button and SADA will report about a 70% chance of finding the object. Notice that your answers will vary slightly each time you reapply the algorithm. This is because the solution is numerical. It should converge as you increase the number of simulations.

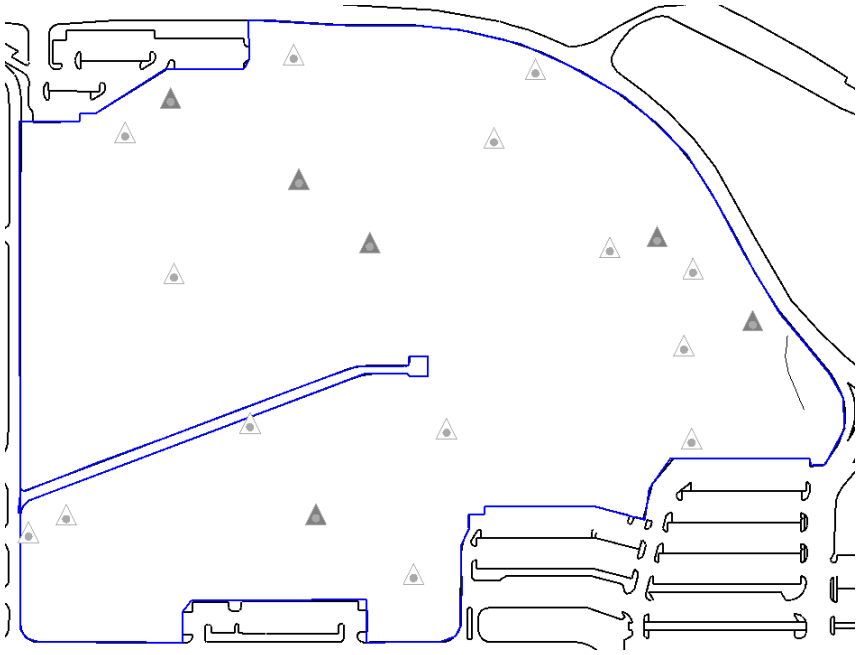
Stored Sample Designs

If you have tried any of the initial or secondary designs, you have likely already encountered this feature. You can store a static copy of any sampling design within your SADA file. To try this, make sure you have InitialDesigns.sda open and click on *Set up the site* step. Make sure the *Vertical Layer* design is set to 3 layers. Click on *Set sampling parameters* and select *Simple Random* from the drop-down list of available designs.

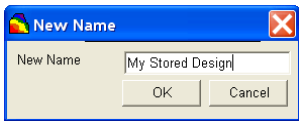
Deselect *Design core samples* and place 20 new samples randomly (*You pick*) throughout this 3 layer design. Your parameter window should look like the following image:



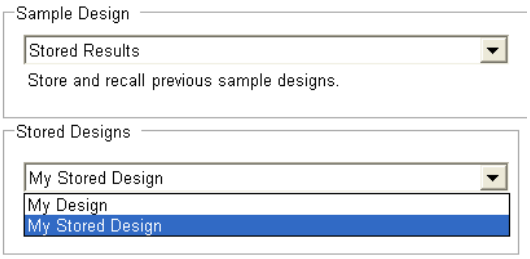
Press *Show The Results* and SADA produces a design like the following image (yours will not necessarily be the same).



The solid gray triangles are those samples found on the current layer. The light, empty triangles are samples found at lower (or higher) levels. To save this result, simply press the *Store Design As…* button.



Enter “My Stored Design” as the name of the result and press *OK*. It’s very important to emphasize that nothing is really stored or saved until you save your SADA file. In the drop-down list of available sample designs, select *Stored Results* and a list of previously stored results will appear.



Select *My Stored Design* and press either the *Show Me* button or the *Show The Results* button. The sample design is automatically restored. To test this fully, click on the step *Display empty plot*. Then return to the *Set sampling parameters* step and restore the design by pressing the *Show Me* button. To permanently store the result, you’ll need to save your SADA File. Do this now.

One important point remains. We created and stored this sample design under the 3 layer vertical layer design. If you restore this result under another vertical design, you may not be able to see all your sample points. This will be especially true if you have the *Show ghost image of*

new samples found on other layers deselected. This is not necessarily an error but+ can create misunderstandings. It's best to use the vertical layer design with which you created the result.

Hotspot Search References

Conover, WJ. 1980. Practical Nonparametric Statistics, Second Edition. John Wiley & Sons, New York.

Davidson, JR. 1994. ELIPGRID-PC: A PC Program for Calculating Hot Spot Probabilities. ORNL/TM-12774. Oak Ridge National Laboratory, Oak Ridge, Tennessee.

Davidson, JR. 1995a. Monte Carlo Tests of the ELIPGRID-PC Algorithm. ORNL/TM-12899. Oak Ridge National Laboratory, Oak Ridge, Tennessee.

Davidson, JR. 1995b. ELIPGRID-PC: Upgraded Version. ORNL/TM-13103. Oak Ridge National Laboratory, Oak Ridge, Tennessee.

Gerald, CF and PO Wheatley. 1989. Applied Numerical Analysis. Addison-Wesley Publishing Company, Reading, Massachusetts.

Gilbert, RO. 1987. Statistical Methods for Environmental Pollution Monitoring. Van Nostrand Reinhold, New York.

Gogolak, CV, GE Powers, and AM Huffert. 1998. A Nonparametric Statistical Methodology for the Design and Analysis of Final Status Decommissioning Surveys. NUREG-1505. U.S. Nuclear Regulatory Commission, Washington DC.

Law, AM and WD Kelton. 2000. Simulation Modeling and Analysis, Third Edition. McGraw Hill, Boston.

Nuclear Regulatory Commission (NRC). 2000. Multi-Agency Radiation Survey and Site Investigation Manual. (MARSSIM). NUREG-1575, Rev. 1. U.S. Nuclear Regulatory Commission, Washington DC. As of Jan. 2003 a PDF copy of MARSSIM may be obtained at: http://www.epa.gov/radiation/marssim/index.html

Singer, DA. 1972. ELIPGRID, a FORTRAN IV Program for Calculating the Probability of Success in Locating Elliptical Targets with Square, Rectangular, and Hexagonal Grids. Geocom Programs, 4:1-16.

Singer, DA, and FE Wickman. 1969. Probability Tables for Locating Elliptical Targets with Square, Rectangular and Hexagonal Point Nets. Pennsylvania State University, University Park, Pennsylvania.

Chapter 40: Multi-Agency Radiation Survey and Site Investigation Manual (Scenario A)

The Multi-Agency Radiation Survey and Site Investigation Manual (MARSSIM) is the product of interagency cooperation among five key US federal agencies responsible for managing radioactive materials: the Department of Energy, the Department of Defense, the Environmental Protection Agency, and the Nuclear Regulatory Commission. This document provides guidance for determining whether a site is in compliance with a radiation dose or risk-based regulation. You can obtain a copy of MARSSIM at this website http://www.epa.gov/rpdweb00/marssim/. The chapters and discussions most relevant to this user guide are found in Chapter 5 section 5 (Final Status Surveys) and Chapter 8 sections 2-5. Also you may find the document NUREG 1505 particularly helpful as well (http://www.orau.org/PTP/PTP%20Library/library/NRC/NUREG/NUREGS.htm). US Nuclear Regulatory Commission Report NUREG 1505 provides the technical background for the survey design and statistical tests used in MARSSIM. It focuses more closely on the methods we present here.

This chapter is neither a substitute for any guidance document or for any official MARSSIM training. It assumes a reasonable level of knowledge about MARSSIM and shows you how to implement those features in SADA. This chapter covers those features in SADA associated with Scenario A. Scenario B is not yet implemented in SADA. We begin by reviewing two important principles: survey unit and release criterion.

Survey Unit

A survey unit is a geographical area with a specific size and shape for which a release decision will be made. A survey unit may be the entire site or a smaller portion of the site. The important thing to emphasize here is that the survey unit will be the geographic target for decision making. You may have more than one survey unit on your site. Each one then becomes a separate decision. Specifically, each one must pass a release criterion in order to be released for public use. If a survey unit fails the release criterion additional steps must be taken prior to release.

Release Criterion

The criterion for release is that the amount of residual radioactivity must be present at levels that produce less than the total effective dose equivalent (TEDE) limit or risk limit set by the cognizant regulatory agency. The dose equivalent or excess risk is not a physically measureable quantity. Rather a pathway model is used to model the dose or risk over time. More importantly, this model is used to determine the maximum allowable concentration associated with the TEDE. This concentration limit is referred to as the Derived Concentration Guideline (DCGL) and is expressed in units that can be measured now (e.g. Bq/g or Bq/m^2).

There are two concerns in assessing radiological contamination in MARSSIM. First, the survey unit-wide average concentration should not exceed the DCGL. For this reason, it is usually written as the $DCGL_W$ (the w indicates survey unit area-wide limit). In addition, localized areas of elevated activity may also exist that may cause the dose or risk criteria to be exceeded. A separate DCGL known as the $DCGL_{EMC}$ is used to screen against local elevated areas. It is obtained using the same dose model used to derive the $DCGL_W$, but with the smaller area of contamination used. The two values can be related to each other through an *area factor*. More specifically we have,

$$DCGL_{EMC} = F_A \times DCGL_W$$

The derivation of the area factor (F_A) is beyond the scope of this user guide. The important point here is that we will be dealing with two decision criteria ($DCGL_W$, $DCGL_{EMC}$) at two different scales (survey unit-wide, local). Any DCGL is obtained by finding the concentration of activity within a *specific* area that would cause the dose or risk limit to be exceeded.

For the survey unit-wide comparison, non-parametric statistical tests will be used. Either the Sign Test or Wilcoxon Rank Sum (WRS) test will be used to test the hypothesis that the survey unit-wide average is less than the $DCGL_W$. Comparisons for the local activity levels will be conducted by comparing scanning results and/or sample measurements directly against the $DCGL_{EMC}$. Any result that exceeds the $DCGL_{EMC}$ requires additional investigation and/or remedial action before the survey unit can be released. The number of samples you will need to take must be sufficient to achieve agreed upon Type I and Type II error rates for the statistical tests, and also have a sufficient density in the survey unit to ensure that any potential elevated areas will be discovered with high probability. We will now discuss laying out samples in order to apply these two decision approaches simultaneously.

Classifying the Survey Unit

Survey units are classified according the likelihood that they are contaminated. More sampling resources will be spent on the survey units most likely to be contaminated above the release criterion.

Class I Areas

These are areas containing locations where, prior to remediation, the concentrations of residual radioactivity may have exceeded the $DCGL_W$. Class I areas have the highest potential for containing small areas of elevated activity exceeding the release criterion. Therefore, both the number of sampling locations and the extent of scanning effort is greatest. The sampling effort is driven by the goal of finding areas with concentrations in exceedance of the $DCGL_{EMC}$. Sampling is done on a systematic grid and the distance between sampling locations is made small enough so that any elevated area that might be missed by sampling would be found by scanning.

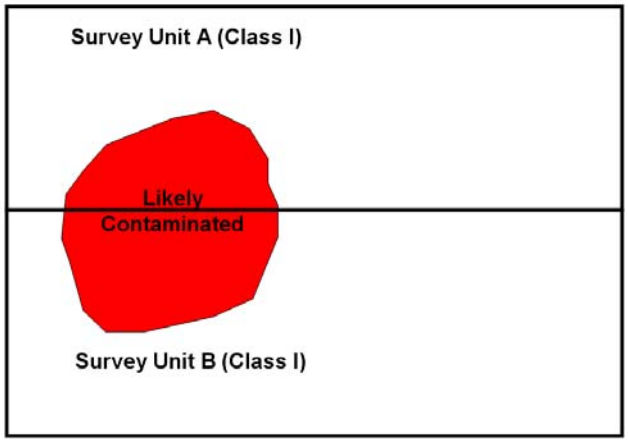
Class II Areas

These are areas containing no locations where, prior to remediation, the concentrations of residual radioactivity may have exceeded the $DCGL_W$. Class II areas may contain residual radioactivity, but the potential for elevated areas is very small. Sampling is done on a systematic grid and the distance between samples is limited by limiting the maximum size of the survey unit. Scanning coverage ranges from 10-100% depending on the potential for elevated areas.

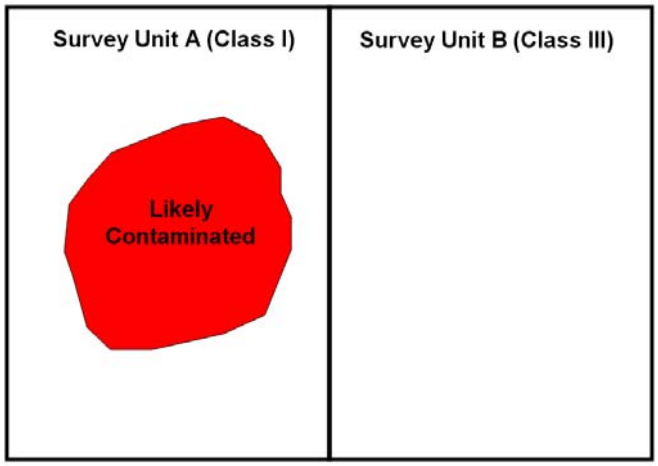
Class III Areas

These are areas with a low probability of containing any locations with residual radioactivity. Class III areas should contain little, if any, residual radioactivity. There should be virtually no potential for elevated areas. Sampling is random across the unit and the sampling density can be very low. Scanning is performed on a judgmental basis.

You should choose your survey units with these classifications in mind. You don't want to divide a site into survey units in such a way that virtually every unit is a Class I. If you have areas that are likely uncontaminated you may want to spatially define a separate survey unit for this area. Consider the following figure where a contaminated area has been divided into two survey units. This now forces the classification for both to be Class I, and requires both to have a higher level of sampling and analysis.



Rather you may want to divide the site into survey units that are more manageable. The following image shows the same site divided into two survey units. However this arrangement will require fewer samples as Survey Unit B is now a Class III.



Part I: Sample Design

The number of samples and their distribution is dependent on survey classification, the size of the survey unit, and the sensitivity of the scanning equipment used to check activity between sample locations. Sample designs are either systematic grids or random designs depending on the classification of the survey area. In addition to sampling, scanning is done on part or all of the survey unit for Class I and II units. Very limited scanning is done in Class III.

Class	Design	Scan
I	Grid	100%
II	Grid	10-100%
III	Random	Judgmental

Determine Number of Samples/Choose a Statistical Test

When sampling has been completed, you will use either a Sign Test or Wilcoxon Rank Sum (WRS) test to determine if the mean concentration exceeds the $DCGL_W$. Technically, these two methods test hypothesis about median and not the mean. If the distribution of values is roughly symmetrical then the mean and median area are about the same. If not, then MARSSIM argues that the results are still more robust than if a parametric test is used (MARSSIM, sec 8.2.3). If the contaminant is present in background, then data for the contaminant is collected from a reference area as well as the survey unit and the Wilcoxon Rank Sum test is used. If the contaminant is not present in background and radionuclide-specific measurements are made, then the Sign Test is used.

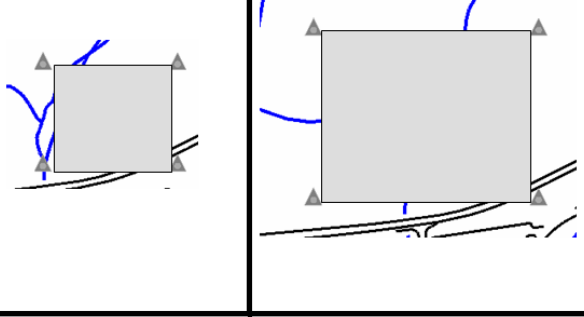
As you recall from both Chapter 7, "Statistics," and Chapter 37, "Overview of Sample Designs," it is possible to estimate *a priori* the number of samples to take if you are willing to make some assumptions about the data. Power curves for both the Sign and WRS test require an estimate for the data variance, the $DCGL_W$, and tolerances for Type I and Type II errors. The result is an estimate for the number of samples to collect both in the survey unit and if necessary in the reference area. Please visit the the section on determining the number of samples in Chapter 37 for a more detailed discussion of Sign and WRS power curves.

Adjust the Number of Samples for scanning (Class I only)

For Class I areas, scanning the surface plays an important role in determining if contamination exists in the unit. Recall that the $DCGL_{EMC}$ is calculated as

$$DCGL_{EMC} = F_A \times DCGL_W$$

The $DCGL_W$ is determined by an exposure pathway model. The F_A can be calculated by RESRAD (http://web.ead.anl.gov/resrad/home2/) or other model approved by the appropriate regulatory agency. The F_A depends on the area in which you are interested. For us, the *grid* area is related to the distance between the sample points. The bigger the spaces between samples, the smaller the F_A (and the $DCGL_{EMC}$), the smaller the spaces between the samples, the higher the F_A (and the $DCGL_{EMC}$).



Area = 100m²	Area = 200m²
Area Factor = $F_A(100)$	Area Factor = $F_A(200)$
$F_A(100) > F_A(200) \rightarrow DCGL_{emc}(100) > DCGL_{emc}(200)$	

The greater the value of the $DCGL_{EMC}$ the easier it becomes to determine if it's been exceeded. Very small $DCGL_{EMC}$ values may be lower than the sensitivity levels of some scanning devices. So given the spacing we do have between our samples, is our scanning device good enough?

The more sensitive the scanning instrument the more likely it can detect if the $DCGL_{EMC}$ has been exceeded. The scanning ability is entered into the decision process through the minimum detectable concentration (MDC). If the MDC is less than the $DCGL_{EMC}$ then we can proceed. If the MDC is greater than the $DCGL_{EMC}$, this means we won't be able to tell whether we've exceeded the $DCGL_{EMC}$ or not. There are three options at this point:

Use a better scanning device.

This option is obvious. However, there may be physical limitations that prevent the availability of a more sensitive device.

Increase the number of samples

Increase the number of samples so that the scan area is reduced. Since the area factor and survey area are inversely related, this will increase the $DCGL_{EMC}$. However this relationship is generally non-linear. At a minimum the sample density should be increased until the $DCGL_{EMC}$ and MDC are equivalent. How many more samples are needed is determined by back calculating from the MDC through the area factor to the actual area.

$$DCGL_{EMC\ (new)} = F_{A(new)} \times DCGL_W = MDC$$

giving

$$F_{A(new)} = MDC / DCGL_W$$

The area associated with $F_{A(new)}$ must be determined outside of SADA in software such as RESRAD. Once the new, smaller, between-sample area is produced by RESRAD, this will correspond to an increase in the number of samples.

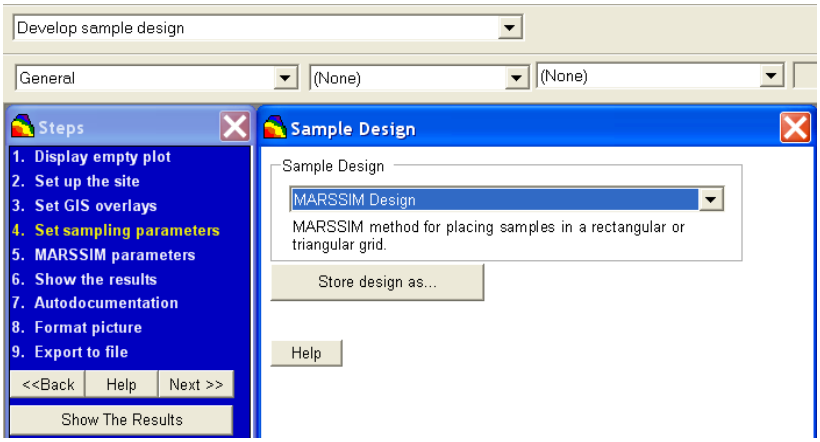
Keep things as they are

Use the scanning device and current sample design and accept some risk that you may miss a violation of the original $DCGL_{EMC}$. The scanning device could determine only exceedances above the MDC. For values below the MDC, some areas that exceed the $DCGL_{EMC}$ but not the MDC may be missed.

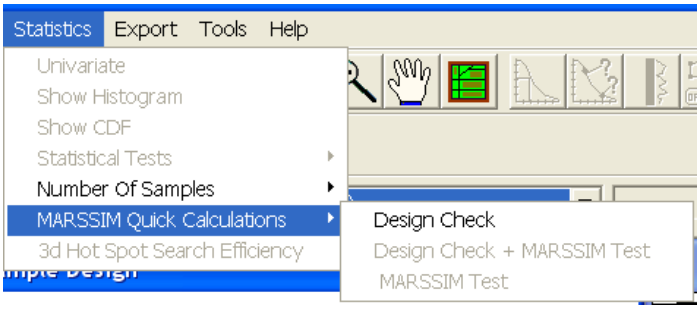
These probabilities can be calculated using the hot spot search algorithm described in Chapter 39, "Standard Initial Sampling Designs."

MARSSIM Sample Designs in SADA

MARSSIM features can be accessed in two ways. First you can select MARSSIM as a sample design under the *Develop sample design* interview. This is the formal method where a sample design is created, stored, and later retrieved after data has been collected.

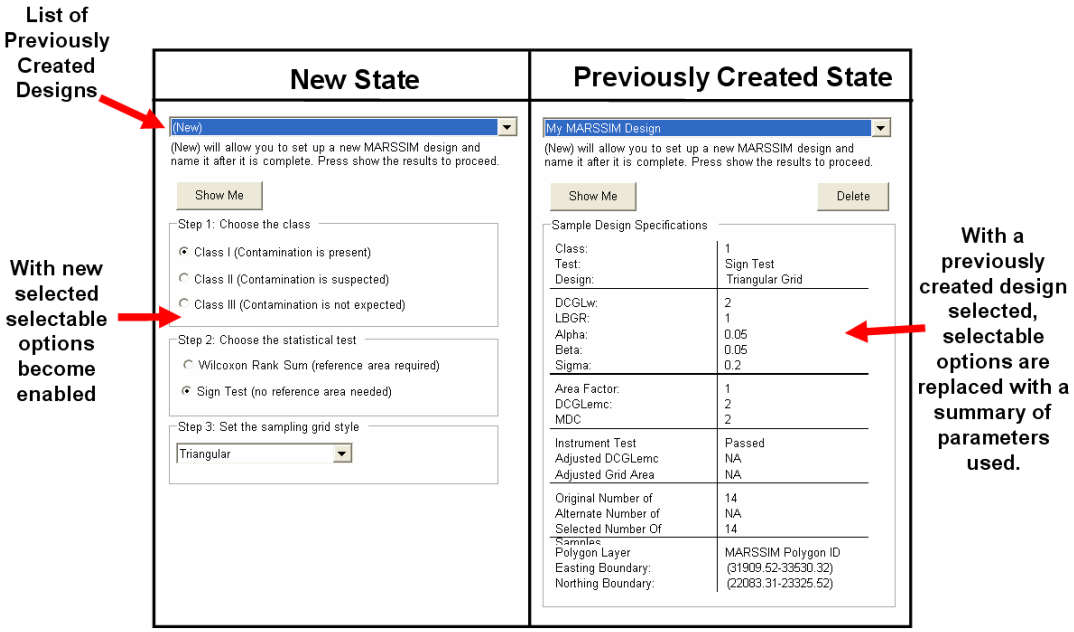


This process is available in smaller pieces using the MARSSIM quick calculation tools under the *Statistics* menu item. The quick calculation features are designed more with regulatory users in mind. With the quick calculation feature you can easily check some numbers in a previously conducted MARSSIM analysis.



The formal method is better suited for individuals carrying out the MARSSIM analysis. We'll begin with this approach first. At the end of the chapter, we'll cover all the quick tools separately.

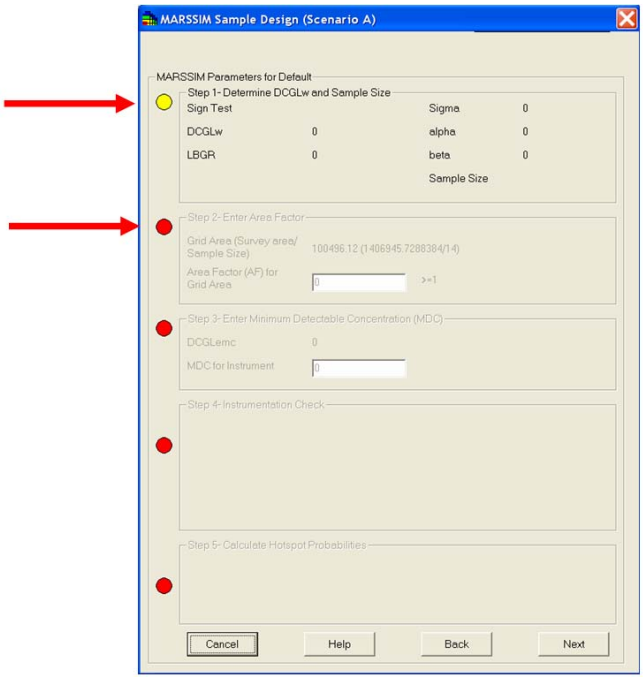
Please open the file InitialDesigns.sda. Select the *Develop sample design* interview and in the *Sample Design* parameter box, select MARSSIM Design. Click on the step *MARSSIM parameters*. This parameter window can exist in two states. At the top of the parameter window there is a drop-down list of previously created designs and the *(New)* option. When you have *(New)* selected, SADA understands that you want to create a new sample design and presents you with options you can customize. If you select a previously stored design, the customizable options are replaced with a static report of parameter values that were used to create the design. You cannot edit these.



At this point, we don't have any previously stored results. Let's divide the example into two sub-examples: Class I/II and a Class III survey unit. The radionuclide(s) could be anything. The only place the radionuclide matters is in the calculation of the area factor and possibly the MDC. We'll use hypothetical values for each of these. The survey unit has been delineated in the graphics window by the *Actual Site Boundary* polygon (see *Setup the site* step).

Class I/II Example

In *Step 1: Choose the class*, select *Class I (Contamination is present)*. For our first example, we'll assume the radionuclide is not present in background. In *Step 2: Choose the statisticsl test*, we'll choose *Sign Test (no reference area needed)*. Since the survey unit is a Class I (or II), we are also presented with *Step 3: Set the sampling grid style*. Select *Triangular*. Press the *Show The Results* button to proceed to more sample design parameters.

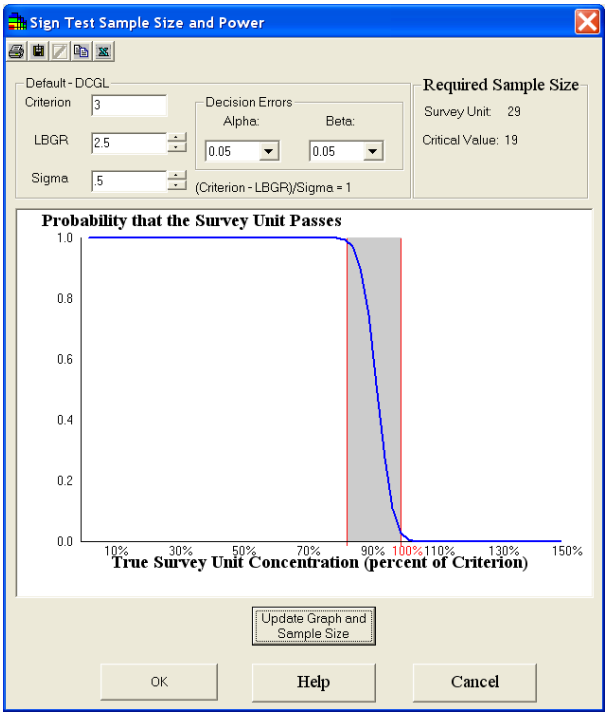


In this window we are presented with some additional steps that we will follow top to bottom. Each step has a "light" just to its left. A red light means that the step has not been completed. A yellow light means the step is the current step. A green light means that the step is complete. Each step is dynamic and may change depending on the answers provided to previous steps. For this reason, all steps except for the current step are disabled.

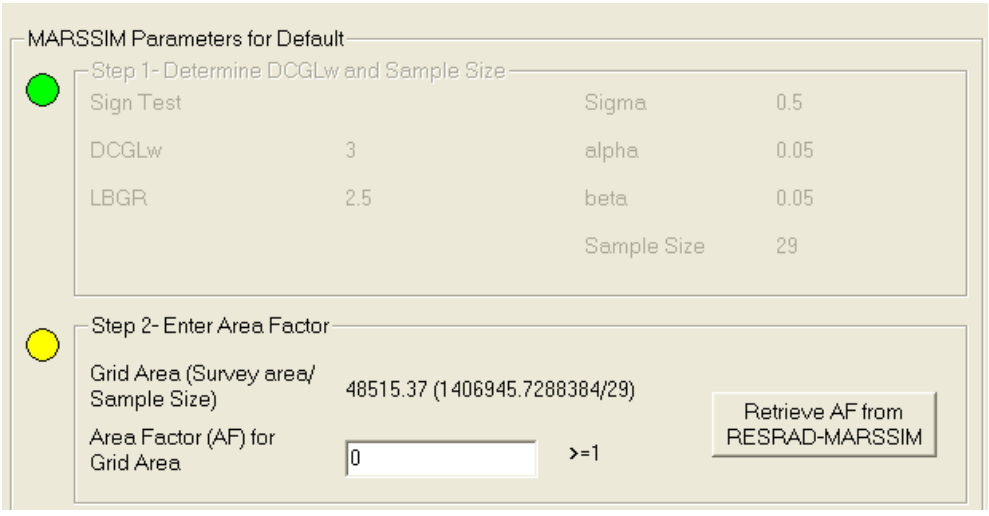
Step 1 is to determine the sample size for testing the following hypothesis under the Sign Test.

$$H_0: \mu_s > DCGL_w (\text{fails criterion})$$
$$H_1: \mu_s \leq DCGL_w (\text{passes criterion})$$

Here, μ_s is the mean concentration for the survey unit. Press the *Next* button and we are presented with the statistical power curve window for the Sign Test. This window was covered in some detail in Chapter 39. We will not cover all aspects of it here. You should return to that chapter now if you are not comfortable with the power curve. The important point here is that when we finish this window we will know an estimate for the number of samples we must take to test whether the mean concentration exceeds the $DCGL_W$ or not. Make sure your parameters match those in the following image. You will want to press the *Update Graph* and *Sample Size* button when you've entered the correct values.



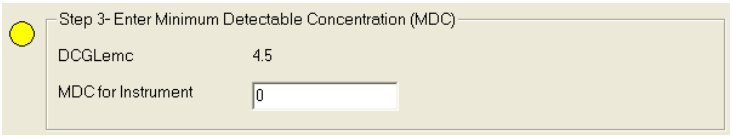
We see in the upper-right-hand corner that we will need to collect 29 samples in order to test the hypothesis that the survey unit does not meet the release criterion. We have a critical value of 19. This means that if more than 19 of the 29 samples is less than the criterion then we must reject the null hypothesis. Press *OK* and we are returned to the previous parameter window. Notice that the first step regarding sample size now has a green light.



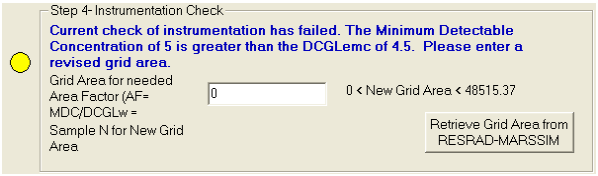
Step 2 is to determine our $DCGL_{EMC}$. The $DCGL_{EMC}$ is our local allowable limit given our site-wide limit $DCGL_W$. Recall that

$$DCGL_{EMC} = F_A \times DCGL_W$$

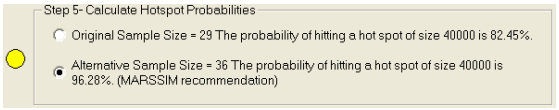
So we need to get the F_A from somewhere. The F_A is radionuclide specific and can be calculated using RESRAD or other approved dose or risk models if we know the area of interest. The area of interest lies between our samples, and in this case is about 48,515 square feet. We would normally use this value in RESRAD along with the radionuclide of concern to arrive at the area factor. In fact, SADA can import an entire RESRAD area-factor-curve and pick the area off the curve for your specific area size. In our example, we'll just use a hypothetical value of 1.5. Enter this value and press *Next*.



The local limit is 4.5pCi/g. The next step is to determine if our scan instrument can detect a value as low as 4.5pCi/g. Let's suppose for a moment that the MDC is higher than 4.5. This means our scanning device can't detect concentrations that low over such a large sample gap. Enter a value of 5 for the *MDC for Instrument* and press *Next*.



Step 4 tells us that the instrument is inadequate. We must now adjust our $DCGL_{EMC}$ to be equal to that of the MDC. This will result in a smaller area factor. From our equation we know that the new area factor must be equal to the $MDC/DCGL_W$ or in this case 5/3= 1.7. We will need to determine what area corresponds to this area factor of 1.7. This will again come from outside of SADA (RESRAD). For this hypothetical example, we will enter a new grid area of 40000 . This will mean smaller spacing between samples and therefore more samples are required to fill the survey unit. Press *Next*.



In this next step, SADA asks us to make a decision. We can go with the original 29 samples determined with the $DCGL_W$ in mind, or we can increase the sample size to 36 (34 for me) and produce a better analysis.

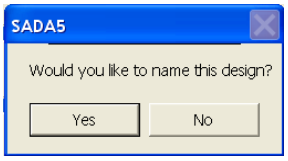
If we go with the first choice (29 samples), our scan device would not have a high probability of finding elevated areas with concentrations between 4.5 and 5.0. The chance of finding an area with a concentration > 4.5 using the grid samples alone using SADA's ELIPGRID calculation, however, is about 82% (87%for me) .

If we go with the revised alternate sample size of 36 (34 for me), then the Sign test will have increased power and the probability of finding an elevated area of size (30000 for me) increases to about 96%. The scanning instrument will find any elevated area smaller than this.

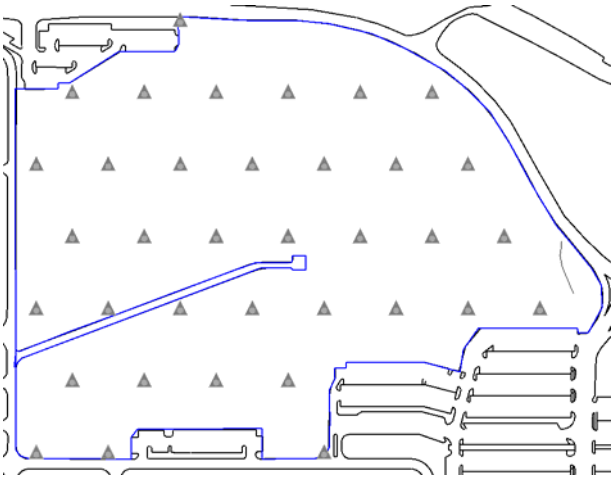
Since in this instance the increase in sample size is modest, we'll choose this option and press *OK*. SADA then uses the simple grid approach (behind the scenes) to place the 36 samples. As mentioned in the simple grid design in the previous chapter, sometimes it is not geometrically possible to place a requested number of samples (recall that 17 samples cannot be distributed on a square survey unit as a square grid). If this occurs, SADA will add one or two more

samples to make the design geographically complete. In this case, it wasn't necessary – 36 samples are placed (my 34 went up to 36).

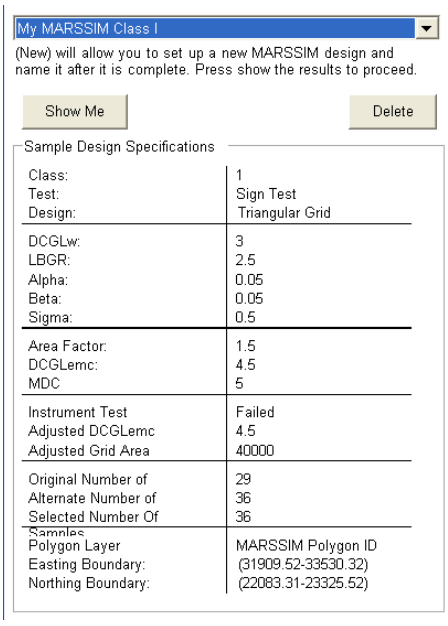
The placement of your 36 samples may differ from those presented here. Recall that a simple grid is a random start grid. You do have an option now to save your design as a MARSSIM design.



Answer *Yes* and provide the name "*My MARSSIM Class I*." You MUST save the design here if you intend to continue later with a full MARSSIM analysis once the data becomes available.



In the MARSSIM parameters window, select *My MARSSIM Class I* from the drop-down list of MARSSIM designs. You will now see a static report of the parameters used in the design.

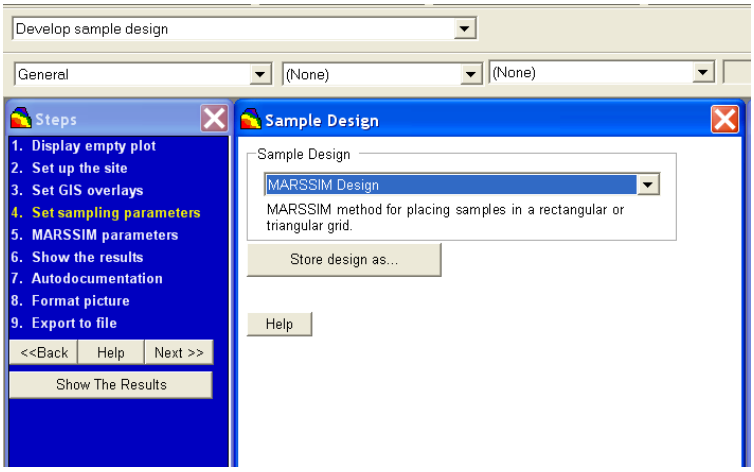


Save your SADA file now, and we'll use this result later. Click on the step *Set sampling parameters* and note that there is also a *Store Design as…* option. You can certainly store the

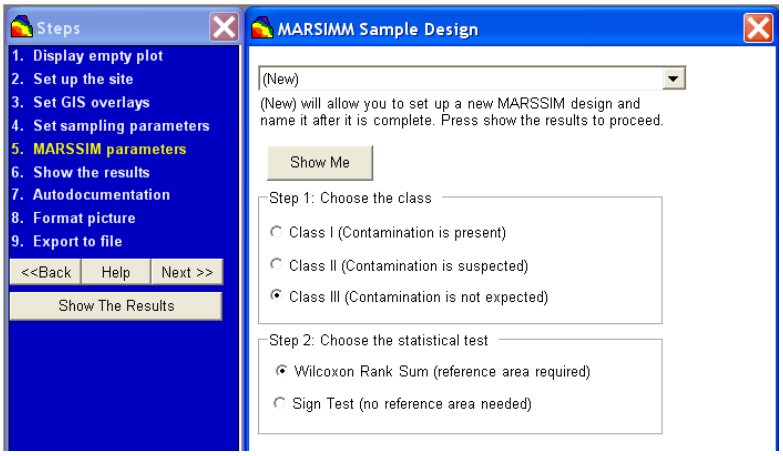
design, but SADA does a “dumb” save in this case. It will neither recall the parameters used to create the design nor that it is a MARSSIM design. You should not use *Store Design as…* if you are going to continue with a MARSSIM analysis later.

Class III Example

A Class III survey unit is not expected to have any contamination, and so the sampling effort here is less. In this case, a random design is used rather than a systematic grid, and scanning is done only if deemed necessary by site investigators. In this example, we will assume our radionuclide is present in the background and will demonstrate the Wilcoxon Rank Sum approach. If you have not already done so, open InitialDesign.sda, switch to *Develop sample design*, and select *MARSSIM Design* from drop-down list of available designs under the *Set sampling parameters* step.



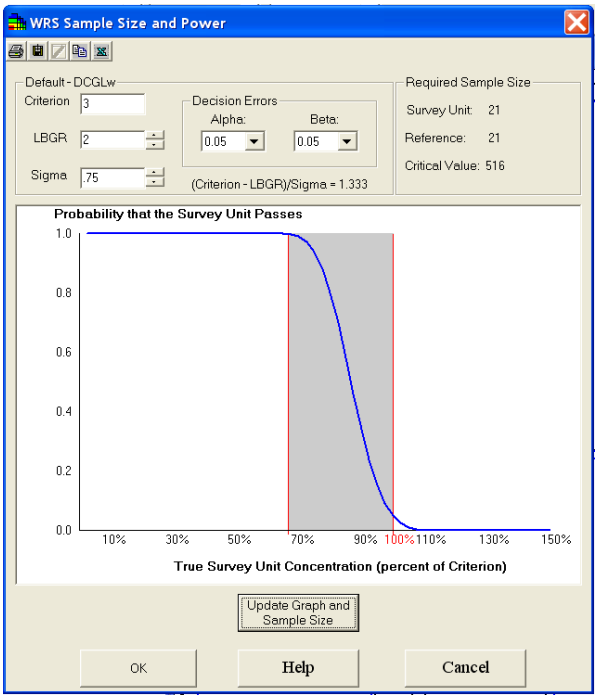
Click on the *MARSSIM parameters* step, and select *(New)* from the drop-down list at the top. Then select *Class III (Contamination is not expected)* and *Wilcoxon Rank Sum*.



Press *Show The Results* to bring up the next set of parameters. As with Class I or Class II survey units, we need to estimate the number of samples it will take to test the hypothesis that passes the release criterion. In this case, the background is involved. So we are testing to find out whether or not the difference between the survey-unit mean and the reference-area-background mean is less than the $DCGL_W$. In other words, is onsite contamination adding an excessively large amount to the naturally-occurring background already there?

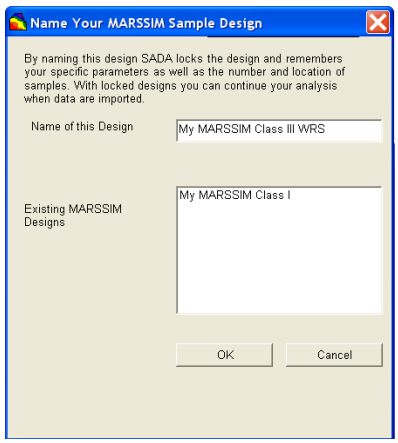
$$H_0 : \mu_s - \mu_B > DCGL_w \text{ (fails criterion)}$$
$$H_1 : \mu_s - \mu_B \leq DCGL_w \text{ (passes criterion)}$$

Where μ_s and μ_B are the mean survey-unit concentration and mean background concentration respectively. For this exercise, we will use a $DCGL_W$ of 3pCi/g (calculated outside of SADA) and LBGR of 2pCi/g. Recall that the LBGR or lower bound of the gray region from the overview of initial designs relaxes the number of samples you must take. In this case, we can take fewer samples because we are betting that the mean concentration will not fall between 2 and 3pCi/g. If in the end it does, then we might possibly have to take more samples. We'll enter a sigma of 0.75 based on similar assessments in the past, and we'll use the customary 0.05 for both Type I (alpha) and Type II (beta) error rates. If you not comfortable with these concepts, please review Chapter 37, "Overview of Sample designs" or Chapter 39, "Standard Initial Sampling Designs.".

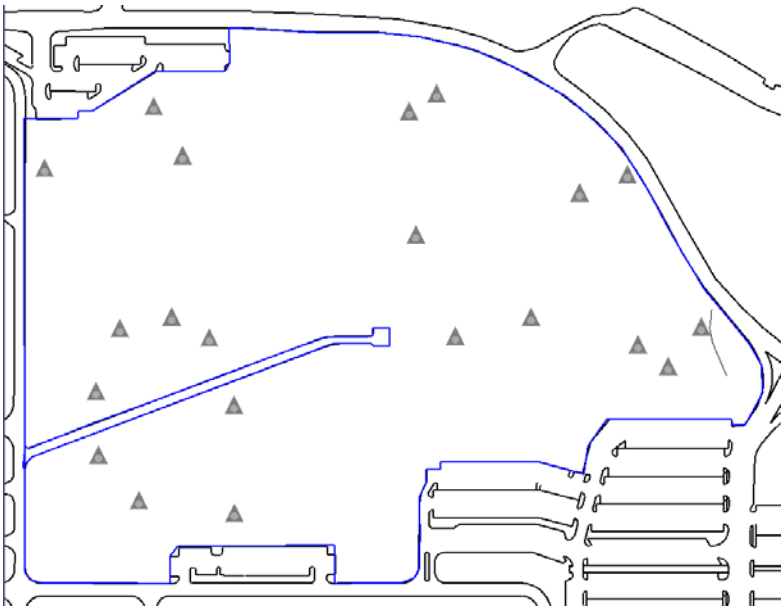


Pres the *Update Graph and Sample Size* button. There are now two samples sizes: one for the survey unit, and one for the reference area. Both require 21 samples for a total of 42 samples. Press the *OK* button and SADA distributes the 42 samples. Note that we are not required to answer any questions regarding scanning sensitivity because this is a Class III survey unit (no scanning required).

You will be immediately asked if you want to save the design. In order to carry a MARSSIM design through to final analysis you will need to store it at this point. Let's say *Yes* and enter the name "My MARSSIM Class III WRS." Press *OK* and SADA will store both the designs and the parameters, and it will recognize later when you have data that you may want to use to complete the MARSSIM analysis.



Note that your random design may look differently than the one presented here because of the randomness of the process.



Click on the step *Set sampling parameters* and note that there is also a *Store Design as…* option. You can certainly store the design, but SADA does a "dumb" save in this case. It will neither recall the parameters used to create the design nor that it is a MARSSIM design. You should not use *Store Design as…* if you are going to continue with a MARSSIM analysis later. More importantly, save your SADA file now or none of this will be truly saved.

We've seen an overview of the sample design process and how to implement it in SADA. Now we'll turn to the analysis of actual data. The sample designs you've just created can be exported to a CSV file by clicking on the *Export To file* step or pressing the Export to File button . A sampling team can now take these coordinates and collect the data. The results can then be imported back into this same file where the MARSSIM analysis will continue.

Part II: Analysis of Data

In the previous section we discussed how to create a MARSSIM survey sample design for Class I, II, and III with and without background data. In this section we will use the sample designs generated there to collect and analyze the data as we continue through the MARSSIM process.

Testing for Compliance with $DCGL_W$

We know from the previous discussion that we must first test to find if the survey unit average is above the $DCGL_W$. For radionuclides that do not occur in background we have the Scenario A hypothesis:

$$H_0 : \mu_s > DCGL_w (\text{fails criterion})$$
$$H_1 : \mu_s \leq DCGL_w (\text{passes criterion})$$

Here, μ_s is the mean concentration for the site (technically a Sign Test is for the median and not the mean, however MARSSIM uses median as proxy to mean under assumption of symmetry). This hypothesis is tested using the Sign test. For radionuclides that are present in the background we have the following Scenario A hypothesis:

$$H_0 : \mu_s - \mu_B > DCGL_w (\text{fails criterion})$$
$$H_1 : \mu_s - \mu_B \leq DCGL_w (\text{passes criterion})$$

Where μ_s and μ_B are the mean survey unit concentration and mean background concentration respectively. This hypothesis is tested using the Wilcoxon Rank Sum (WRS) test.

Testing for Compliance with $DCGL_{EMC}$

Assuming compliance with the $DCGL_W$ is met, then for Class I and Class II survey units, individual measurements, including scan measurements must be less than the $DCGL_{EMC}$. Recall that the $DCGL_{EMC}$ is a limit for localized activity. This test is accomplished by simply comparing the results to the $DCGL_{EMC}$. If the $DCGL_{EMC}$ is exceeded by one or more points, more investigation is needed, such as additional sampling or scanning. Please refer to MARSSIM for more details.

Testing for MARSSIM Compliance in SADA

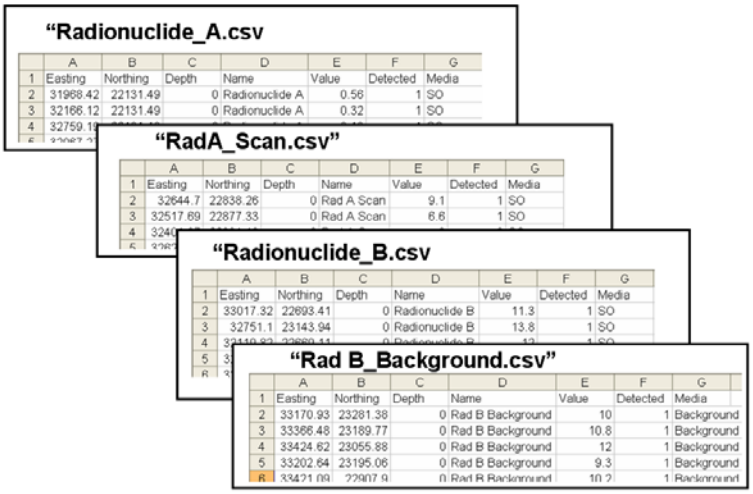
As with MARSSIM sample design, there are two ways to access the MARSSIM compliance testing features. The first way is using the formal MARSSIM analysis feature where you first create a sample design, store it, import your data, and perform the analysis against your recorded design parameters. This is the most thorough, but also the most restrictive. If you are interested in doing a MARSSIM analysis but the original design was not constructed or saved in SADA (but in some outside means), you will want to use the MARSSIM quick tools feature. This feature is found under the *Statistics* menu. If this is the case, it will be a good idea to keep reading through this section to understand how SADA conducts the analysis. At the end of this chapter we discuss the quick calculation features separately. We continue now discussing the more formal, thorough approach.

Once you've created a MARSSIM sample design and stored it in your SADA file, you can export the design, and then import the results when they're collected. You are then ready to continue with your analysis. We will begin at the point where the data have been imported. Open the file MARSSIM.sda. If SADA asks you if you want to set up the MARSSIM analysis say *No* for now.

This file has the same MARSSIM sample designs created in the previous discussion along with the corresponding sampling measurements. In the previous section we provided two examples. In the first example, we demonstrated a Class I/II type design for a radionuclide not present in the background. This design was based on the Sign Test and stored as My MARSSIM Class I. This sample design was targeted at Radionulcide A which also required some scan results "Rad A Scan." The sample locations for scanning were not produced but were left to the judgment of the site investigator. We imported the measurements associated with both of these designs. You can find them in the data set drop-down list.

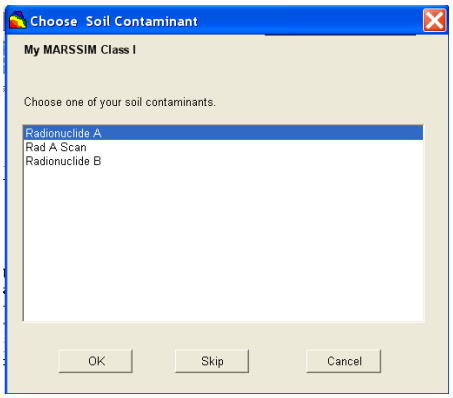
We also demonstrated a Class III design for a radionuclide that was present in background. This design was based on the Wilcoxon Rank Sum test and stored as My MARSSIM Class III WRS. This sample design was targeted at Radionuclide B for which data also must be collected in a suitable reference area. This background/reference area data was called Rad B Background. Both Radionuclide B and Rad B Background data sets were imported into SADA. If you are only reading the MARSSIM chapter in the user guide, you will likely also need to read the first six chapters on setting up a SADA file and importing data.

Spend a moment or two and look at each of the data sets we imported. The important thing to note is that our background data set Rad B Background has a media type of *Background* and not *Soil* even though it was sampled in the soil. This is important to remember because when doing the WRS test, SADA will ask you to select from among the list of data sets with media type Background. If you import your Rad B Background as a *Soil* media type, SADA would not recognize that you wanted to use it as such. The following images show screenshots from the comma separated value (csv) files that we imported. The first few lines of each of these csv files as seen in Microsoft Excel are shown below:

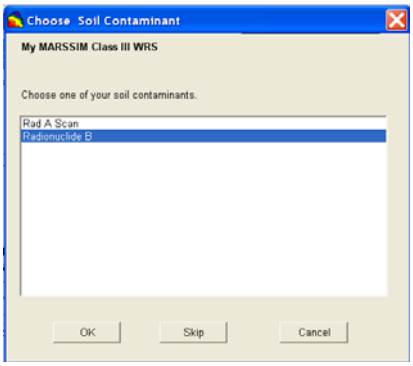


Setting up the MARSSIM Analysis

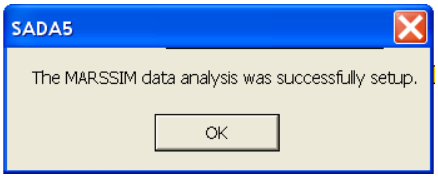
In order to access the formal MARSSIM analysis you will need to setup the analysis. Select from the main menu *Setup→MARSSIM Data Analysis*. SADA needs to know which data set is associated with each stored MARSSIM design. The first one is *My MARSSIM Class I*. You will want to associate this with the dataset *Radionuclide A*.



Select *Radionuclide A* in the list of data sets and press *OK*. Now SADA wants to know which data set is associated with *My MARSSIM Class III* WRS. Select *Radionuclide B* and press *OK*.



If everything goes well you should see the following message box:

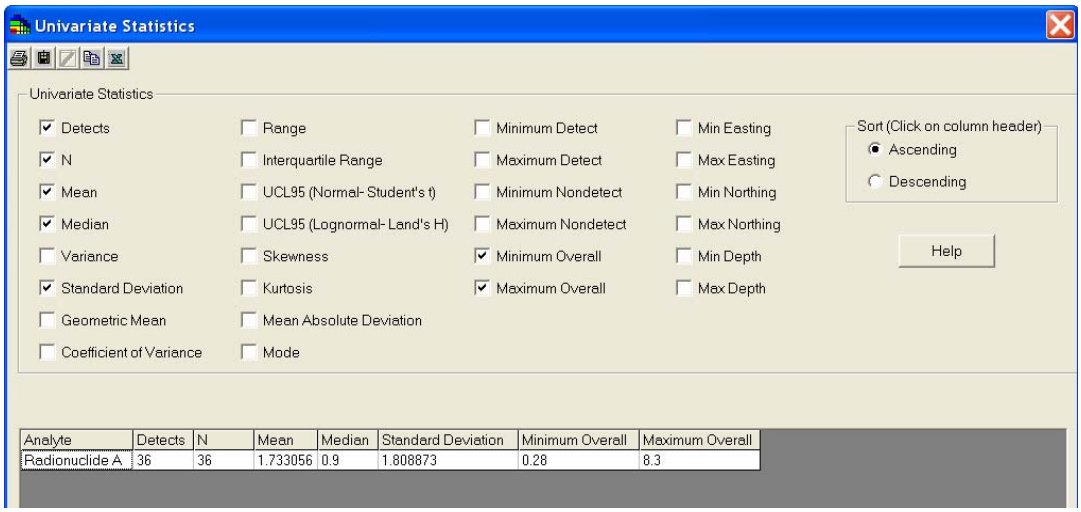


But what about the Rad A Scan data set and the Rad B Background? The MARSSIM analysis will ask you later which data set is the background for Radionuclide B. The Rad B Background doesn't show here because SADA recognizes that its role will enter the process later. The data screen interview will be used to confirm whether any scan results exceed the $DCGL_{EMC}$. We'll get to both of these momentarily. Press *OK* on the message box. Save your SADA file.

Evaluating Data Quality Objectives

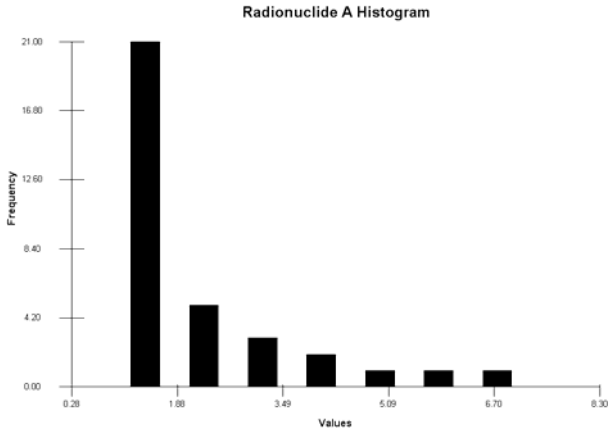
MARSSIM Chapter 8 suggests a review of some of the DQO objectives planned for in the survey. This includes computing some exploratory statistics: mean, standard deviation, and median. We can view these even before using the MARSSIM analysis tools. To demonstrate how to do this in SADA, in the analysis drop-down list select *General*. In the data type drop-

down list select *Soil*, and in the data set drop-down list select *Radionuclide A*. We could do the same thing for Radionuclide B. Select *Statistics→Univariate*.

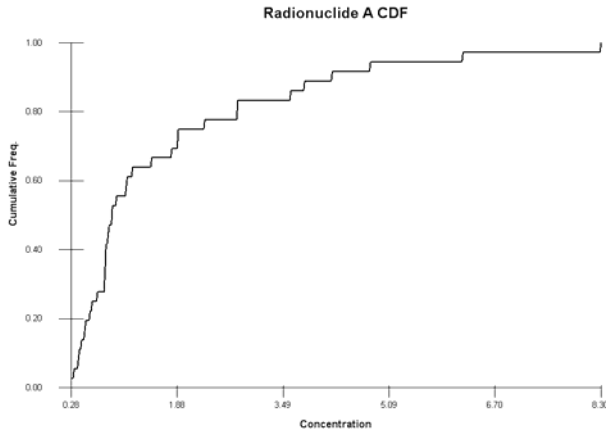


For *Radionuclide A*, we have mean of 1.73, a median of 0.90, and a standard deviation of 1.81. We can see right away that since the mean of 1.73 is less than the $DCGL_W$ of 3 pCi/g, then we will pass the survey unit average criteria under the Sign Test. The standard deviation of 1.81 is higher than the 0.5 we anticipated. This may reduce our statistical power some and may lead us to accept the null hypothesis under Scenario A (null survey unit is contaminated) when it really should be rejected. We'll have to look at the retrospective power curves for the number of data we collected and see how it compares. The mean and median are somewhat different indicating a skewed distribution--which is not uncommon in environmental data. A quick look at the histogram will also reveal this. Click the *X* button in the upper-right-hand corner of the statistics window to close it.

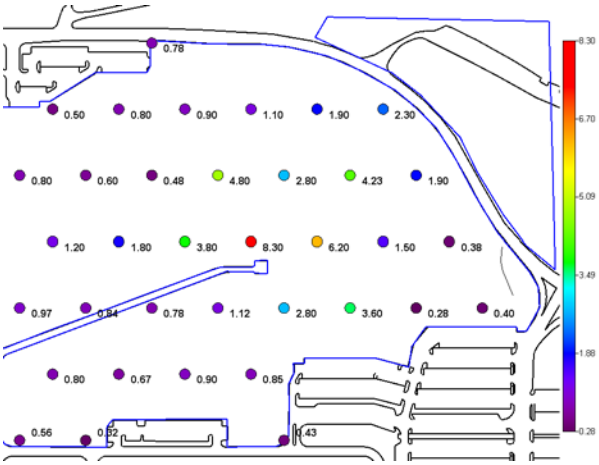
Select *Statistics→Show Histogram* to reveal the following histogram for Radionuclide A data:



This is a clearly skewed distribution. Fortunately, the Sign Test requires no assumption about the underlying distribution. We can also view the cumulative distribution function. Select *Statistics→Show CDF*.

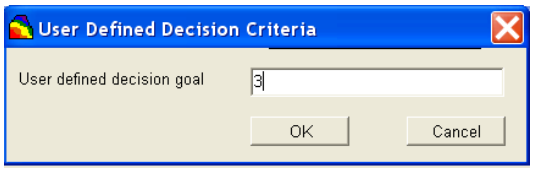


MARSSIM also suggests a posting plot of your data (Chapter 8). This is done by default in SADA. To see the plot of samples within a GIS environment just click the step *See the data*. You can position any field data you have along with the sample locations. To add the measured value to the plot, select *Value* from the meta-data drop-down list.

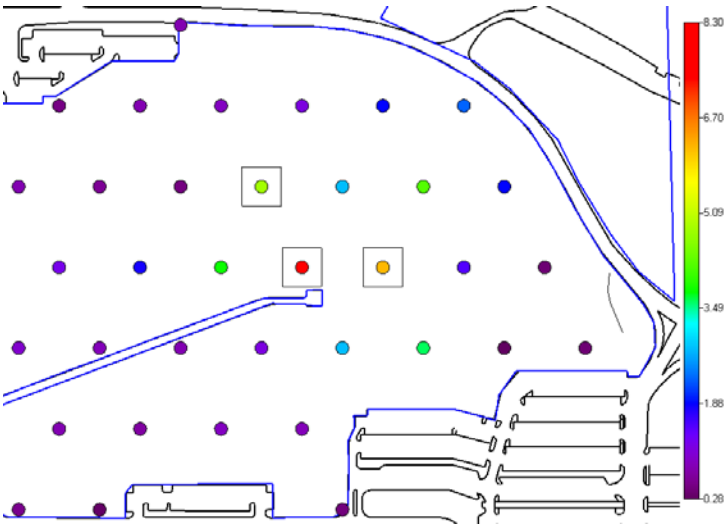


In this plot we are using the default continuous legend. The legend can play a useful role in visualizing your data. Please see Chapter 43, “Graphical Tools,” for helpful tips on how to customize the legend so that important features are brought to visual prominence.

At this point we could do a couple of quick data screens to see where we exceed the $DCGL_W$ and the $DCGL_{EMC}$. This is actually conducted as part of a larger process in MARSSIM, but we can do a sneak peak. In the interview drop-down list, select *Draw a data screen map*. Press *Show The Results* and enter the $DCGL_W$ value of 3pCi/g into the *User defined decision goal* box.



Press *OK* and SADA reports that 6 points exceed the criteria of 3. You can see where they are by the boxes that encircle them in the GIS plot. Press *Show The Results* to repeat this for $DCGL_{EMC}$ of 4.5 pCi/g. In this case 3 points exceed the $DCGL_{EMC}$.



Click on the step *MARSSIM parameters*. In the drop-down list of available MARSSIM designs there is only one (*My MARSSIM Class I*). You can see there the parameters used to create the design.

Sample Design Specifications

Class:	1
Test:	Sign Test
Design:	Triangular Grid
DCGLw:	3
LBGR:	2.5
Alpha:	0.05
Beta:	0.05
Sigma:	0.5
Area Factor:	1.5
DCGLemc:	4.5
MDC	5
Instrument Test	Failed
Adjusted DCGLemc	4.5
Adjusted Grid Area	40000
Original Number of	29
Alternate Number of	36
Selected Number Of Samples	36
Polygon Layer	MARSSIM Polygon ID
Easting Boundary:	(31955.5656-33544.41406)
Northing Boundary:	(22131.48893-23350.0918)

Click on the *Setup the site* step. Notice that with the *MARSSIM Data Analysis* selected and the *My MARSSIM Class I* selected, you cannot edit the boundaries, vertical layers, or the polygons that were used to create the design. Notice in the graphics window that the original sample design is overlaid on top of the measured value coordinates. This allows you to quickly see if any deviations from the original design were committed.

Checking Test Assumptions and Diagnostics

The advantage of using non-parametric methods such as Sign and WRS tests is that they have *fewer* assumptions than parametric methods. This does not imply they have no assumptions.

Spatial independence

This simply means that knowing the value at one location does not tell you anything about the value in another location. Independence can be visually checked by noticing any spatial trends in the post-plots. The values should be more or less randomly distributed. More sophisticated methods such as semi-variography analysis can be used to measure spatial dependence (see Chapter 30, “Advanced Geospatial Methods Part I”). In the case of Radionuclide A, there does appear to be a trend in the data. The center of the site near the end of the road is consistently high, trailing off to lower values at the edge of the site.

Symmetry

Symmetry is not really a requirement of the tests but rather a requirement due to the way we are using the tests. MARSSIM uses tools that test whether the median exceeds a value in order to determine whether the mean exceeds the same value. This is fine as long as there is symmetry in the analysis where the mean and median are about the same (exactly the same for perfect symmetry) as in normal distribution. You can check for symmetry by looking at the histogram or by observing the skewness value for the data (calculated in the univariate statistics feature). If skewness is present to a significant degree so that the mean is substantially higher than the median, there must be some high data values causing this. The elevated measurement is conducted to prevent a survey unit from being erroneously released when this occurs.

Data Variance

Sample sizes were generated initially based on an estimated data standard deviation. One should compare the estimated and actual standard deviation to determine how the test may be affected. If the sample standard deviation is less than the estimated deviation, then you should have enough statistical power in the test. If the sample standard deviation is greater than the deviation you estimated during the design, then depending on the median value you may or

may not have enough statistical power. This is only important if the null hypothesis is not rejected because it will be unclear whether it was not rejected because the null hypothesis is true or simply because the data was not sufficient to detect that the null hypothesis was not true. A finding of insufficient power for a survey design can impact how other designs are conducted at the site, even when the null hypothesis has been rejected.

Retrospective Power Curve

When you created your sample design, the number of samples was computed such that your power curve met an alpha and beta level for a DCGlw and associated LBGR. This was based on the standard deviation as well. Now that we have collected the data, we know more about the standard deviation and can compare the power curve with the more recent value. However, instead of using the number of samples required for the test, we will use the actual number of samples taken. This might be more for Class I units where scanning requirements demanded more samples.

Determining Compliance

Compliance is determined by the following tables (based on MARSSIM Chapter 8).

Sign Test (no background)

Survey Result	Conclusion
All values less than $DCGL_W$	Survey unit meets release criterion
Average greater than $DCGL_W$	Survey unit does not meet release criterion
Any measurement greater than $DCGL_W$ and average is less than $DCGL_W$	▪ Conduct sign test and elevated measurement comparison.

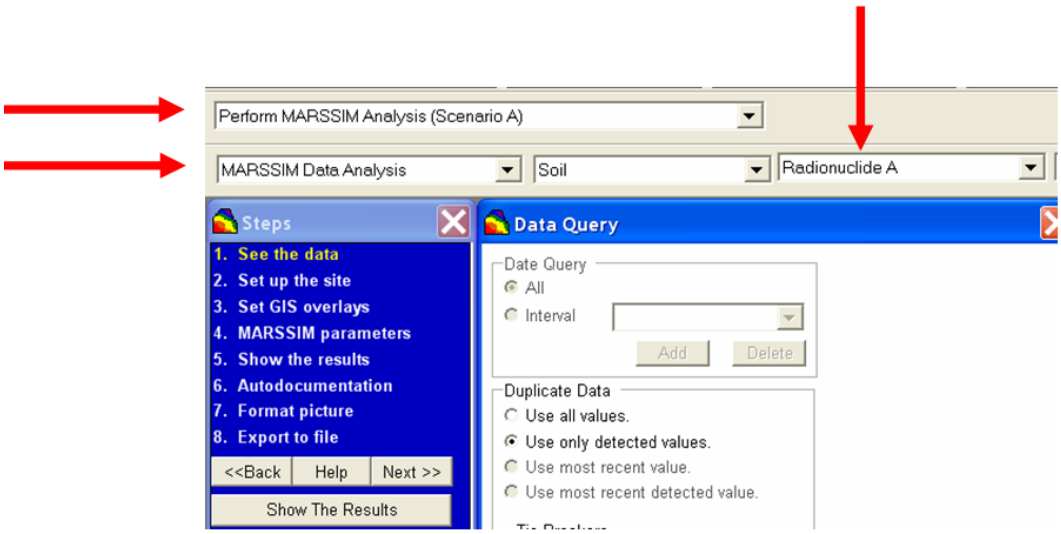
Wilcoxon Rank Sum Test (with background)

Survey Result	Conclusion
Difference between largest survey unit measurement and smallest reference area measurement is less than $DCGL_W$	Survey unit meets release criterion
Difference of survey unit average and reference area is greater than the $DCGL_W$	Survey unit does not meet release criterion
Difference between any survey unit measurement and any reference area measurement greater than $DCGL_W$ and the difference of survey unit average and reference area average is less than $DCGL_W$	Conduct WRS test and elevated measurement comparison.

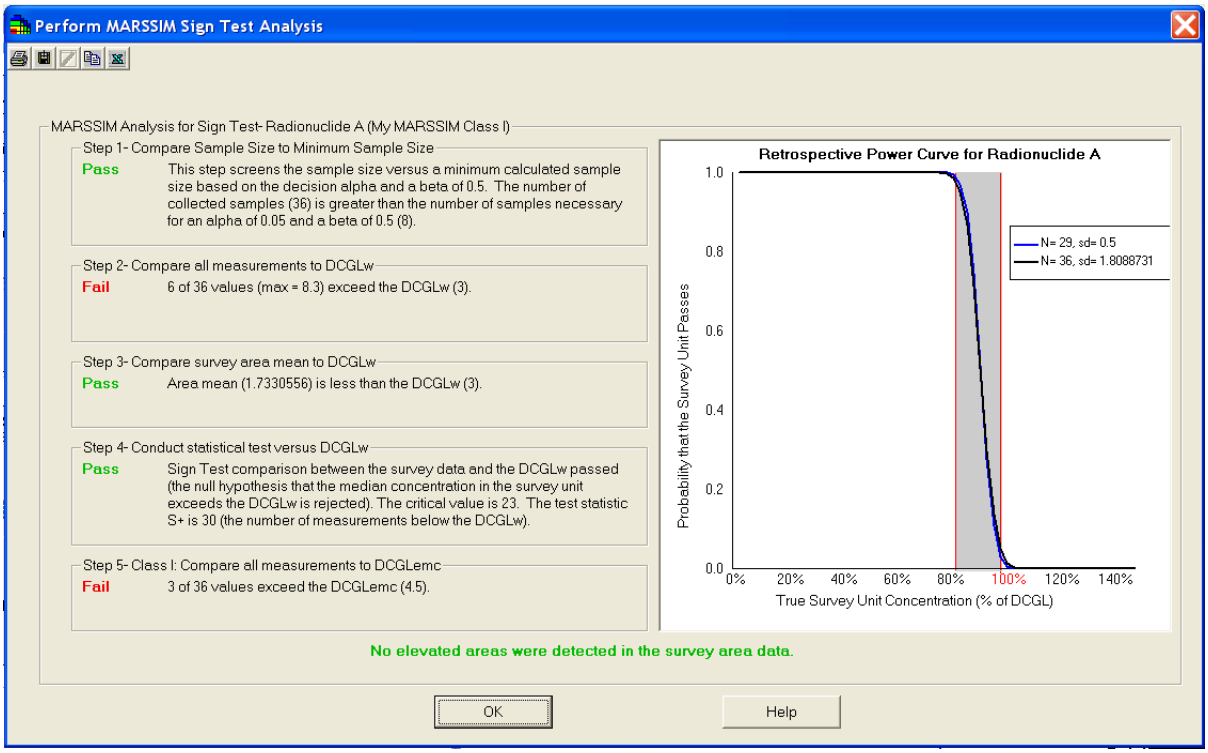
We will now show how to test for compliance in SADA for the previously created Class I/II and Class III example designs.

A Class I/II Example

Switch from the *General* analysis to the *MARSSIM Data* analysis. The only analysis permitted is *Perform MARSSIM Analysis (Scenario A).*



Press *Show The Results* and the following window is presented:

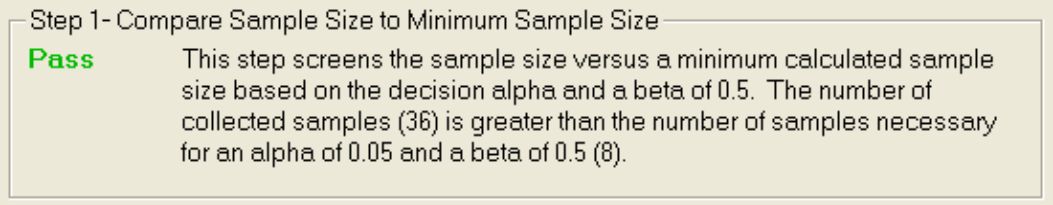


SADA conducts the test for the $DCGL_W$ and $DCLG_{EMC}$ (measured values only) in six instantaneous steps. Before we discuss the steps, notice the *Retrospective Power Curve for*

Radionuclide A graph on the right side. This graph shows the original power curve for 29 samples in blue. Recall, however, that we needed to increase the number of samples to 36 in order to meet the $DCGL_{EMC}$ scanning requirements. We did collect 36 samples, but the standard deviation was not 0.5 as we estimated. In fact it was much higher (1.81). In this case, it does not appear that the power was reduced significantly.

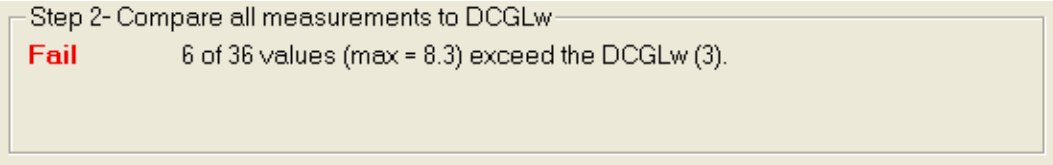
Step 1: Enough Samples?

In this step we re-compute the number of samples required using the new standard deviation and a relaxed beta value of 0.5. Beta refers to the likelihood that you may accept the null hypothesis (survey unit is contaminated) when it isn't true. This step will generate a much smaller number of samples than the original sample design. In a sense it provides a lower bound on the number of samples through relaxing the constraint against incorrectly accepting the null (survey unit is contaminated). The spirit of this step is to prevent you from having to return to the survey unit and sample further, if the standard deviation in fact was much higher. This step is only relevant if you fail it. This means that even with a relaxed beta, you didn't take enough samples. If you pass this step, you know you at least made it over a very low-bar sampling requirement. In this case, we pass it because we took 36 samples and the low-bar requirement is 8 samples. Of course if you did only take 8 samples, you are in considerable danger of saying the survey unit is contaminated when it is not. You can actually revisit the power curve with your known standard deviation by choosing *Statistics→Number of Samples→Sign Test* (or WRS Test). This will tell you how many samples you really needed. You can also observe the retrospective power curve to see if enough were taken.



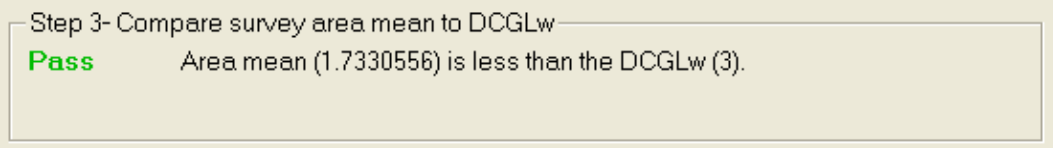
Step 2: How many samples exceeded the $DCGL_W$?

Step 2 informs us how many times the $DCGL_W$ was exceeded. If no measured values exceeds the $DCGL_W$, then the site meets the release criterion. You don't need to examine any of the remaining steps. They are presented only for completeness. In this case we had six samples in excess. This does not mean the survey unit fails (as suggested by the red *Fail* text). It only means that we will need to continue through the steps by looking at the average concentration next.



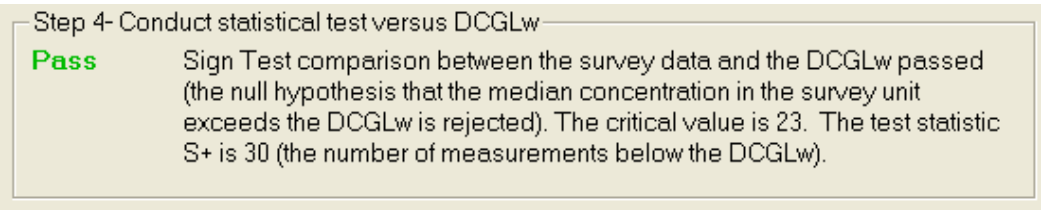
Step 3: What is the survey mean concentration?

Step 3 is really the first opportunity where the survey unit could fail the release criterion. If the mean concentration exceeds the $DCGL_W$, there is no point in continuing. The survey unit fails. The remaining steps will be presented, but they are only provided in the interest of completeness. If the mean is less than the $DCGL_W$, then we don't necessarily pass either; we just need to perform the Sign Test. In this case, we did have some values exceeding the $DCGL_W$ (Step 2), so even though our mean was less than the $DCGL_W$, we are required to do the Sign Test.



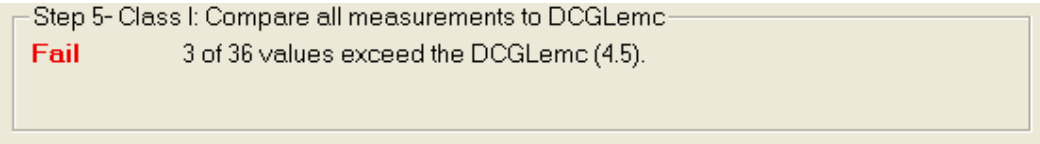
Step 4: Sign Test for $DCGL_W$

Step 4 is where we get a final answer on the pass/fail of the $DCGL_W$. In this case if we pass, the hypothesis is rejected, and the survey unit is classified as clean. If we fail, then we have problems. You might recall from the sample design phase that a critical value of 19 was estimated. So if we had 19 samples fall below the $DCGL_W$, then the survey unit would pass. Because of the larger standard deviation, this critical value was bumped to 23 in order to retain an alpha and beta of 0.05. Thirty measurements actually fell below the $DCGL_W$.



Step 5: Compare All Measurements to $DCGL_{EMC}$

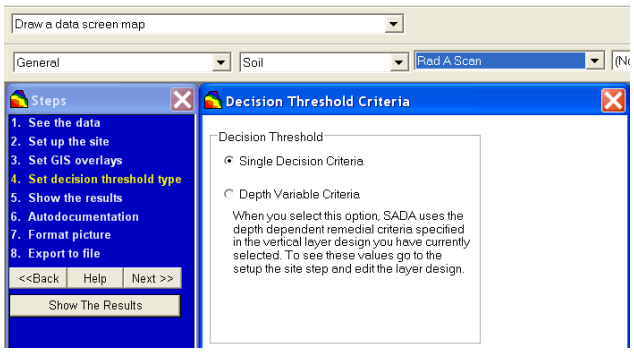
Step 5 is only necessary if Step 4 was necessary (it was in this example). If you are put into the position of using the Sign Test, then you also must screen against the $DCLG_{EMC}$. In this step, we compare all measured values against the $DCGL_{EMC}$. But this is only half the story for Class I/II survey units. Those survey units must also compare the scan results against the $DCGL_{EMC}$. We'll do that shortly as an example. It wouldn't necessarily be required, however, because we already fail here with 3 of the 36 samples exceeding the $DCGL_{EMC}$. Step 5 failure does not mean failure to meet the release criterion. It is essentially a warning that caveats the results of the Sign Test (MARSSIM, section 8.5.1).



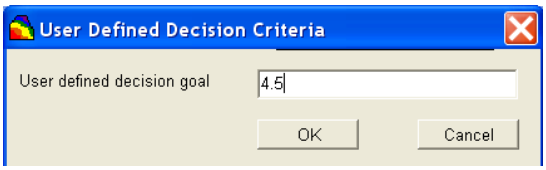
Step 6: Did the Survey unit Pass?

Compare Scan Data Against $DCGL_{EMC}$

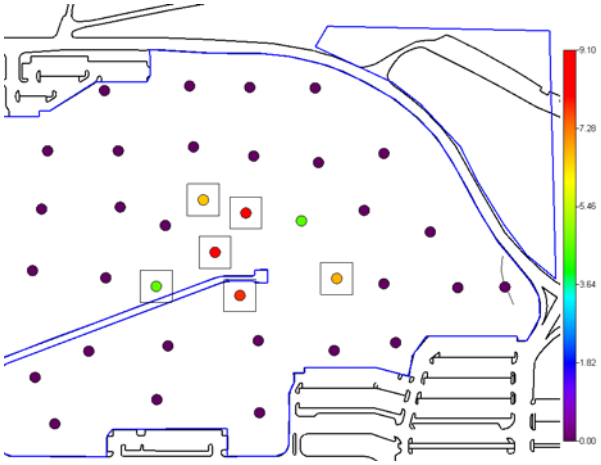
We already know that some of the discrete measurements on the grid sample locations have exceeded the $DCGL_{EMC}$, so more actions may need to be taken. However, we demonstrate here how to easily compare the scan results against the $DCGL_{EMC}$ as well. A spatially informative way is to use the interview *Draw a data screen map*. Unfortunately at this time, we can't access this interview under a *MARSSIM* analysis. We need to switch quickly back to the *General* analysis then select *Rad A Scan* and *Draw a data screen map* interview.



Recall that under the *General* analysis you can have a single-decision criteria or a depth-variable criteria. Click on the *Set decision threshold type* step and just confirm that we have single-decision criteria selected. *Press Show The Results* and enter or $DCGL_{EMC}$ value of 4.5 into the *User defined decision goal*.



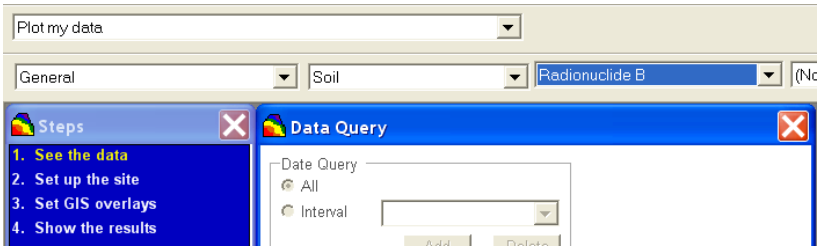
Press *OK* and SADA reports that 6 scan values exceed the $DCGL_{EMC}$. They are identified by boxes around each point.



It is clear from both the $DCGL_W$ and $DCLG_{EMC}$ evaluations that this survey unit will require some additional investigation before release to the public. At this point you might bring some of the geospatial models to bear in order to better refine the local elevated area and perhaps inform a remedial design. After the WRS example, we'll demonstrate how you can do this.

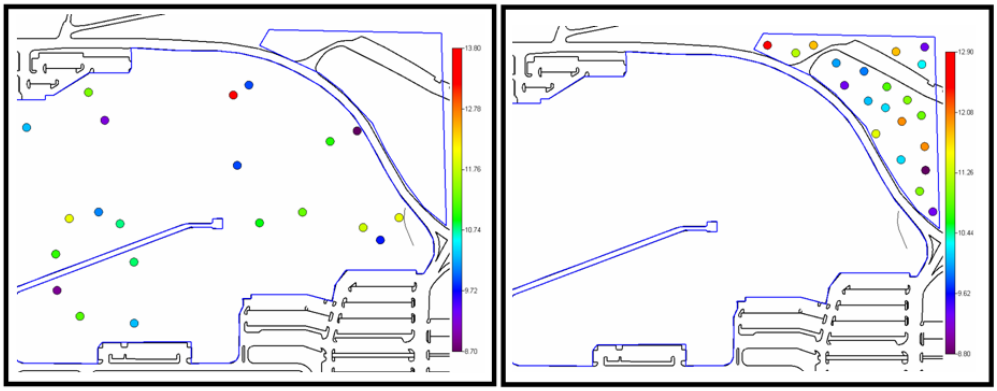
Class III Example (with WRS)

Now we'll take a look at Radionuclide B under a Class III scenario. For this example we assumed B was present in the background as well and so a WRS test was necessary. Let's switch back to the *General* analysis and select *Radionuclide B*.



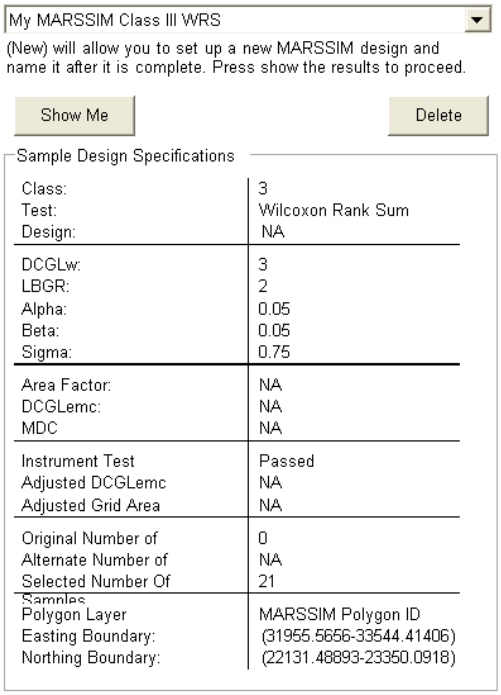
At this point you would want to do some of the same DQO checks and verification of assumptions we previously talked about and demonstrated with the Radionuclide A. We won't repeat those here. Take a look at both the Radionuclide B plots and the Rad B Background data (select *Background* instead of *Soil*). They are presented here side-by-side.

Radionuclide "B" **"B" Background**

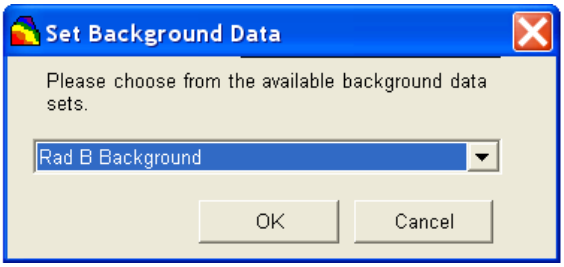


If your background data does not have geographic coordinates, you can still use it. Simply create fake coordinates to get it past SADA's data requirements. The coordinates won't enter into the MARSSIM analysis. However, the reference area and the survey unit should have roughly comparable sizes.

Select *Soil* again and choose *Radionuclide B*. Switch the analysis to *MARSSIM Data Analysis*. Click on the *MARSSIM parameters* step to see the details of our WRS Class III sample design.



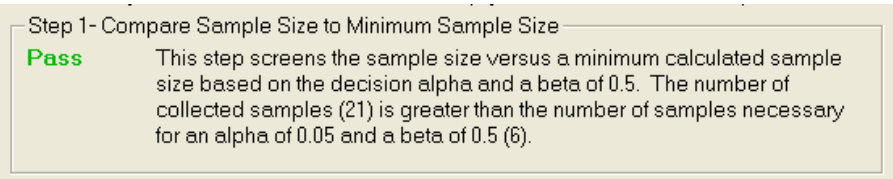
The next step is to specify the background data. Click on this step now and select *Rad B Background* from the drop-down list of available background data sets.



Press *OK*. Press *Show The Results* and the *MARSSIM* analysis window will reappear. We'll investigate these results step-by-step.

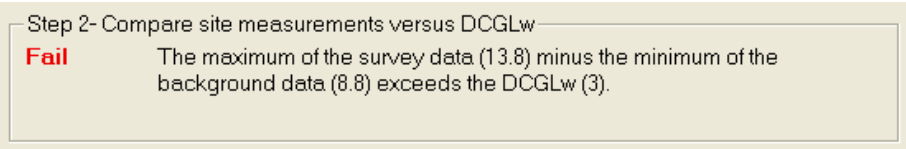
Step 1: Compare Sample Size to Minimum Sample Size.

In this step we re-compute the number of samples required using the new standard deviation and a relaxed beta value of 0.5. Beta refers to the likelihood that you may accept the null hypothesis (survey unit is contaminated) when it isn't true. This step will generate a much smaller number of samples than the original sample design. In a sense it provides a lower bound on the number of samples through relaxing the constraint against incorrectly accepting the null (survey unit is contaminated). The spirit of this step is to prevent you from having to return to the survey unit and sample further, if the standard deviation in fact was much higher. This step is only relevant if you fail it. This means that even with a relaxed beta you didn't take enough samples. If you pass this step, you know you at least made it over a very low-bar sampling requirement. In this case, we pass it because we took 21 samples and the low-bar requirement is 6 samples. Of course if you did only take 6 samples, you are in considerable danger of saying the survey unit is contaminated when it is not. You can actually revisit the power curve with your known standard deviation by choosing *Statistics→Number of Samples→ WRS Test*. This will tell you how many samples you really needed. You can also observe the retrospective power curve to see if enough were taken.



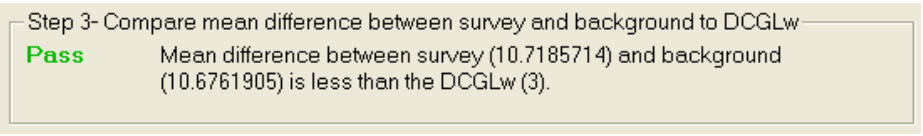
Step 2: Compare Survey unit Measurements Versus $DCGL_W$

When background is involved, we look at the differences between survey unit and reference area rather than measured values themselves. In this case we want to examine the worst case scenario in some sense. If we take the highest survey value and the smallest background value, is their difference more than the $DCGL_W$? If it isn't, then we know the survey unit passes the criterion because all other possible differences will certainly be below the $DCGL_W$ meaning that the average will also be less. If we pass, the survey unit does meet the release criterion and the remaining steps are provided only in the interest of completeness. If we fail, then we may or may not fail to meet the release criterion. We'll need to continue through the steps to find out. In this case, we did fail. We will need to continue to Step 3.



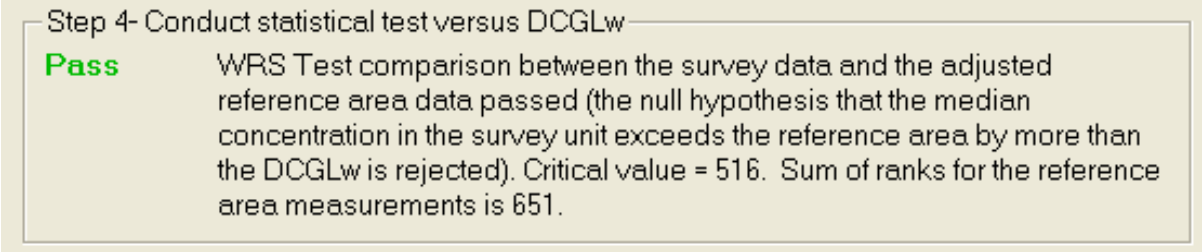
Step 3: Compare mean difference between survey and background

This is the first step in which we can unequivocally fail. If the mean of the survey data, minus the mean of the background data, is greater than the $DCGL_W$, then we fail to meet the release criterion and no further steps are necessary. If we pass, then we need to continue with the remaining steps.



Step 4: Conduct statistical test versus $DCGL_W$

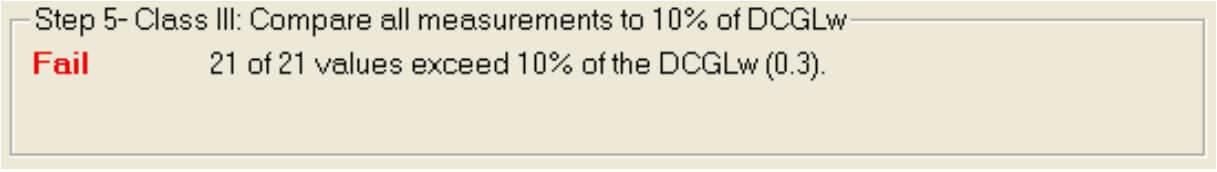
If the survey unit passes the WRS test (reject null hypothesis that survey unit is contaminated) then the release criterion is met. Otherwise, the release criterion has failed. In this example, we passed the test. The critical value of 516 was exceeded by 651 combinations of survey and background data differences falling below the $DCGL_W$.



Step 5: Class III compare all measurements to 10% of $DCGL_W$

In Step 5 we look to see if any survey unit measurement exceeds the reference area average by more than the $DCGL_{EMC}$. In this case, no survey unit measurement exceeds 4.5 + 10.6 =14.6.

The possibility of survey unit misclassification also exists, when measurements are found that clearly show activity above background. To do this, in this particular situation (Class III, background), the survey unit measurements will be compared to the UCL95 on the mean of the reference area plus 10% of the DCGLw. At the time of this writing this was not implemented yet. Rather the values are compared directly to 10% of the DCGLw. This will be corrected in the next update.



Connecting Geospatial Decision Analysis and MARSSIM

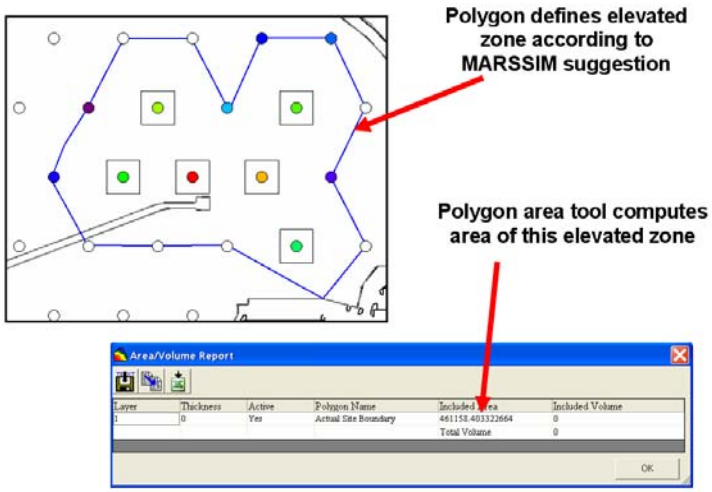
SADA provides a number of tools that can further support or even extend a MARSSIM analysis particularly when there is a failure when comparing against $DCGL_{EMC}$. Chapters 28-37 discuss geospatial decision support, including determining the area of concern (in this case a local elevated area), placement of additional samples, and rudimentary cost-benefit analysis. In this section we will point out how SADA can be used to bound local areas of elevated activity. If you are interested in exploring these tools further, please visit the previous chapters.

Determining the Area of the Elevated Zone for determining Area Factor (simple way).

SADA can help you determine the area of the elevated activity in direct accordance with MARSSIM. In section 8.5.1 we have the following statement regarding the *actual area of elevated concentration*.

"The area of elevated activity is generally bordered by concentration measurements below the $DCGL_W$. An individual elevated measurement on a systematic grid could conceivably represent an area four times as large as the systematic grid area used to define the DCGLEMc. This is the area bounded by the nearest neighbors of the elevated measurement location."

To address this situation, SADA provides a simple tool for determining the actual area of elevated concentration. In the *Set up the site* step, simply use the polygon drawing tools to trace around the elevated points using those points less than the $DCGL_W$ to guide you. When you are done, select *Tools→Area of polygons*. SADA will generate a report that gives the area of every polygon in place.



Determining the Area of the Elevated Zone for determining Area Factor (geospatial way).

It is possible to use the full geospatial decision models in SADA to determine the area of the elevated zone. One can even develop uncertainty bands on the elevated zone. In the following image, we have the same elevated zone identified by a dark boundary line. At the center is a gray region that is most likely elevated. The green boundary indicates uncertainty about where the boundary line may fall.

SADA can also produce area estimates that can include or not include the uncertain boundary regions.

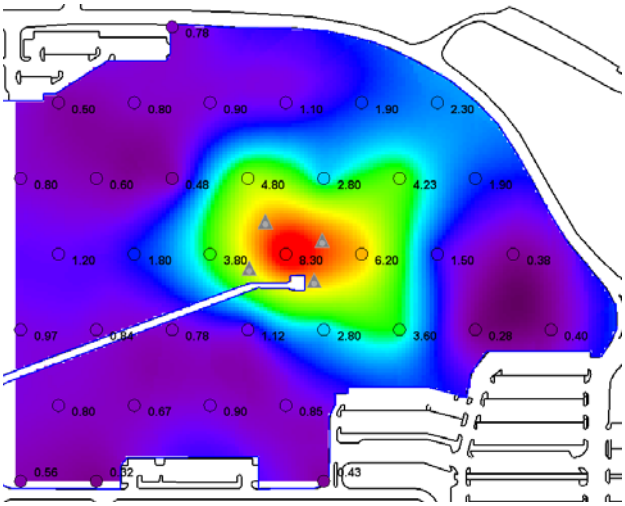
Placing Additional Samples

SADA provides a couple of methods for determining where to place new samples that are relevant to a MARSSIM investigation.

1. Threshold radial. If you have only one or two exceedances, use the threshold radial method to bound the offending samples to determine if an elevated zone really exists.

2. High value design. Use this method to place new samples in the area identified by geospatial modeling as a high-activity area.
3. Boundary design. Use this method to place new samples so that the exact location of the boundary line between elevated and un-elevated zones can be better determined. This is useful particularly if you are considering dividing the unit into different survey units.

In the example below, we've used a high value design to optimally locate four new samples in the high-activity area.



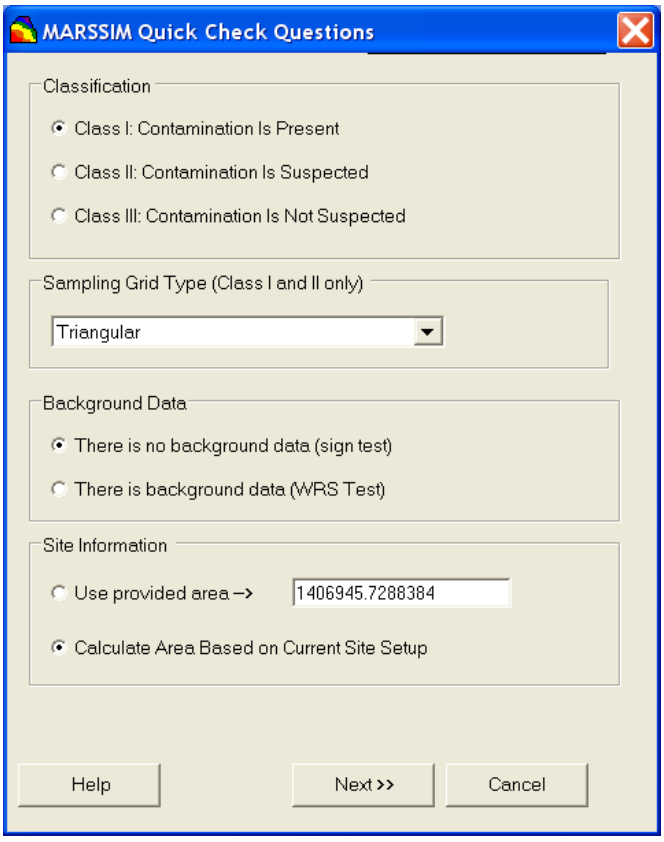
MARSSIM Quick Tools

This formal process can be divided into smaller pieces more readily accessible. To use these tools you don't have to have the *MARSSIM* analysis set up. You may not even need any data imported into SADA if you are just doing a design check. These tools can be found under the *Statistics→MARSSIM Quick Calculations*.

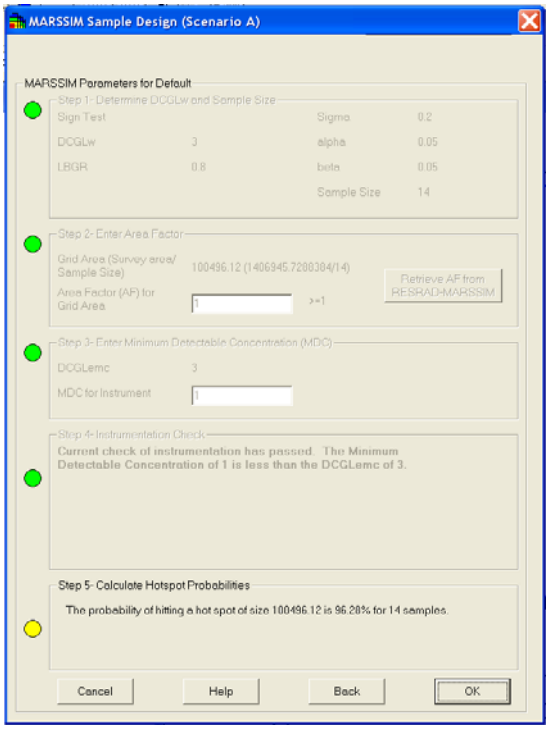
Quick Design Check

This tool simply calculates the number of samples a MARSSIM test would require. It does not place them. This is useful to quickly check that an investigator has used an appropriate number of samples.

Select *Statistics→MARSSIM Quick Calculations→Design Check*. This brings up the first of two parameter sets with which you may already be familiar. You'll pick the class, the grid type (if needed), and set the statistical test just as before. The last parameter block, *Site Information*, asks about the area of your survey unit. You can simply enter the survey unit area if you have it calculated elsewhere. If you already have the survey unit set up in SADA, you can use that as well. Click on *Use provided area*.



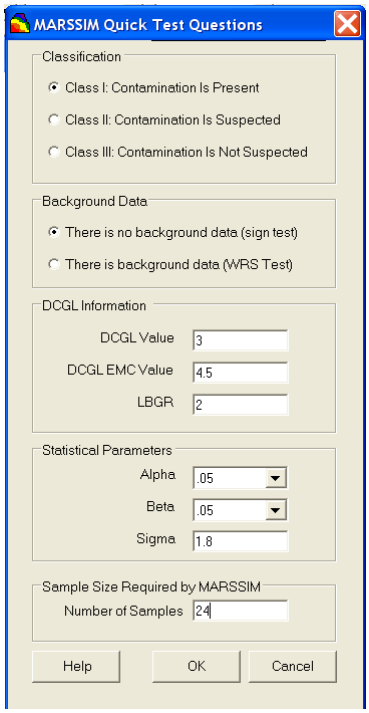
Press the *Next>>* button and you will be presented with exactly the same parameters you've seen before. We won't repeat them here. Enter values of your choosing into each of the steps. Step 1 will give you the number of samples for the test. The remaining steps will consider any scanning needs as before.



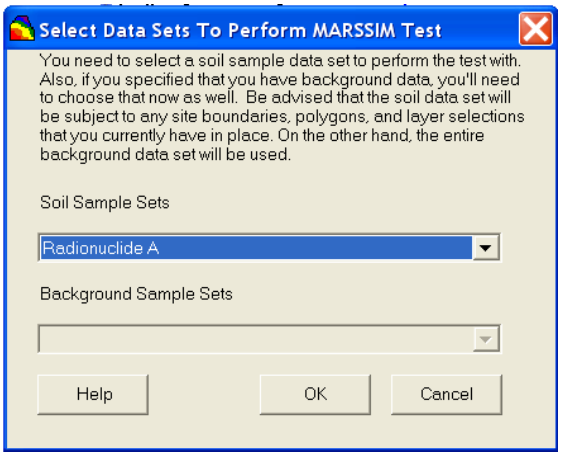
Press the OK button and the tool is dismissed.

Quick MARSSIM Test

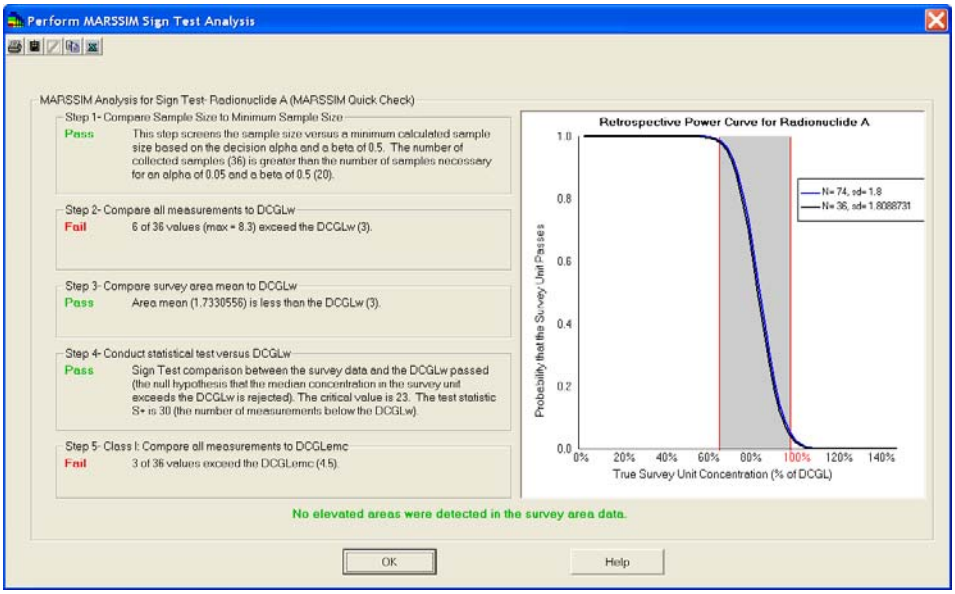
This feature allows you to perform a MARSSIM test without having previously stored a MARSSIM design in SADA and/or set up the MARSSIM analysis. Select *Statistics→MARSSIM Quick Calculations→MARSSIM Test*.



Enter all the information about the design here including the number of required samples. You can enter anything you wish. Press the *OK* button.



You will need to select a data set to run the MARSSIM test on. Therefore you will need to have imported the data before hand. Select Radionuclide A and press OK. You will be presented with the usual MARSSIM analysis window.



Quick Design Check + MARSSIM Test

This tool just combines the last two features into one continuous stream. The only difference will be that the first screen encountered with the MARSSIM Test is unnecessary.

Chapter 41: Informed and Targeted Initial Designs

Previously, we discussed the standard initial sample designs available in SADA. These designs are classics in the sense that they have been around for many years, their objective is to meet some customary statistical endpoint. Such endpoints include hypothesis testing under assumptions of independent and identically distributed data, searching for objects, or some spatial coverage objective. Most individuals work under the premise (stated or implied) that absolutely nothing is known about the site. This is often not true but adopted in the interest of being conservative. In reality, many environmental data sets are not independent, the costs of evaluation are too high to ignore prior knowledge about contamination events, and the objectives may be highly geographical and entirely outside the realm of traditional statistical objectives.

This chapter discusses initial designs that use prior, spatially-definable information to distribute sample locations across the site to meet some target objective. These methods (with the exception of the judgmental design) borrow heavily from the secondary design strategies. In addition, a new method for searching for objects when there is some information about where it most likely exists is introduced.

Sources of Prior Knowledge

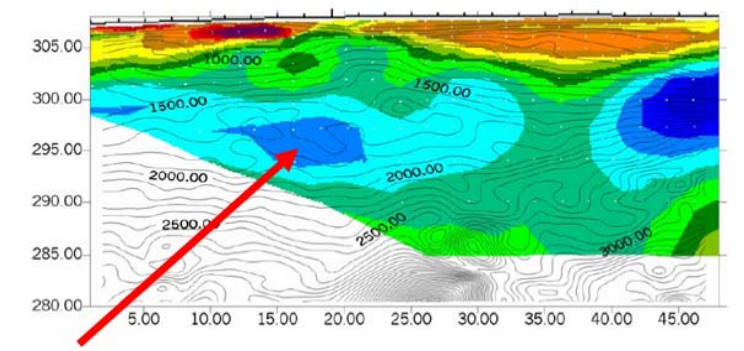
Spatially-definable prior information essentially equates to a 2d or 3d model of what site conditions are thought to be prior to any current evaluation. Of course, the issue is how to come up with this model. There are many ways, but one must be able to accept a degree of uncertainty about the quality of the model. Here are four example sources:

1. Field detection methods

   For surface contamination events, it may be possible to use cheap field detection technologies as a precursor to sampling. Such technologies may include X-Ray Fluorescent (XRF) detectors, various “sniffers,” and radiological scans. Field detection methods that provide less accurate but more abundant data are gaining traction within the regulatory community. For example, the TRIAD model developed and supported by EPA (www.triadcentral.org, last accessed 6/10/2009) encourages this type of increase in data supply over accuracy. The emphasis is shifted to adequate accuracy to support the decision. *Note: SADA is often recognized as a TRIAD support tool (see for example http://www.triadcentral.org/tech/dsp_sub.cfm?id=13, last accessed 6/10/2009).* These detection results can be used to form a model that may indicate the location of contamination in a reasonably rigorous way.

2. Geophysical measurements

   In some cases, geophysical measurements may give some insight into the location of subsurface contamination. The following image, taken from Watson et al., (2002), shows a nitrate plume identified with the use of electrical resistivity (paper can be found at http://public.ornl.gov/orifc/other/DollSpectrum.pdf, last accessed 6/9/2009).



Nitrate Plume

(Taken from Watson et al. 2002. Nitrate plume annotation added here.)

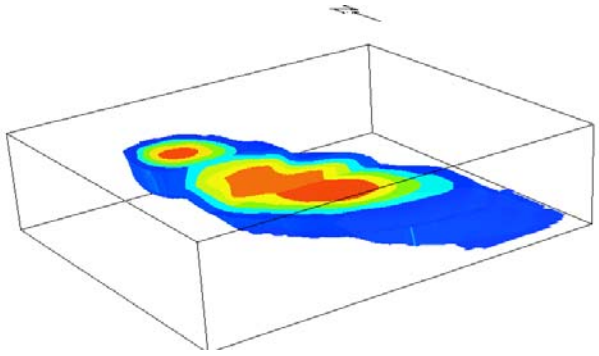
With this kind of geophysical information, it is possible to distribute samples so that they have a greater chance of intersecting and even bounding such a subsurface volume.

3. Previous sampling efforts

   Some sites may have been under investigation for a longer period of time. Historical data may be available that may remain relevant even after accounting for an extended time lapse. This kind of prior data can be imported into SADA, modeled, and used in a secondary or initial targeted design.

4. Professional expertise (user models)

   It may be possible to spatially delineate a "site conceptual" model that indicates where contamination is located based on an assimilation of several lines of evidence. This can include historical documents, previous sampling efforts (that may have lost some relevance), or even some geophysical measurements. In this case, it may be possible to create a user model (see Chapter 33) that captures these different lines of information and creates a conceptual endpoint map/volume. The following image is a user-created delineation of the probability of contamination in the subsurface assimilated from a variety of sources.



If these forms of information can be brought or created in SADA as models, the following initial sample designs can be brought to bear.

Targeted High Value Design (simple)

This design places new samples in those locations most likely to be contaminated based on the prior map. This design is exactly the same as the secondary high-value design discussed in the secondary sample design chapter (Chapter 38) and will not be repeated here. If you are

interested, please visit that chapter now. There is no difference between the approaches. If you import a prior model (see Chapter 7, “Importing Modeled or Gridded Data”), then you are in good shape. Simply select the model under the data type *Imported Models*, select *Develop sample design* as the interview and select *High Value Design* from the list of available designs. If you create the model in SADA using a geospatial model, the process is exactly the same. To be more efficient, you might consider first storing the geospatial model as a static design. Otherwise, SADA will rebuild it each time you build a sample design.

If you create the model in SADA as a user model, you have a couple of new options: *High Value Design (simple)* and *High Value Design*. In the case of *High Value Design (simple)*, SADA treats the user model as if it were a static stored model. If you select *High Value Design*, SADA provides you an opportunity to first update the user model with some real data and then distribute samples.

Area of Concern Design (simple)

This design places new samples in those locations found along the boundary between likely-contaminated and likely-uncontaminated areas. This of course implies a decision criterion and a decision framework (see Chapter 35). The design is exactly the same as the area of concern boundary design discussed in the secondary sample design chapter and will not be repeated here. If you are interested, please visit chapter 38 now. There is no difference between the approaches. If you import a prior model (see Chapter 7, “Importing Modeled or Gridded Data”), then you are in good shape. Simply select the model under the data type *Imported Models*, select *Develop sample design* as the interview, and select *Area if Concern Design* from the list of available designs. If you create the model in SADA using a geospatial model, the process is exactly the same. To be more efficient, you might consider first storing the geospatial model as a static design. Otherwise, SADA will rebuild it each time you build a sample design.

Target High Value and Area of Concern Design (non-simple)

User-Created probability models allow you to update your model using real measured values just before distributing the new samples. Updating user-defined models is discussed in Chapter 33, “User Models,” and won’t be repeated here. It is advisable to update your probability map and then store it as a static copy (see previous section). With an imported or stored model, you can more efficiently apply the simple versions just discussed. To use this non-simple implementation, you would follow exactly the same parameterization methods discussed in the section on updating the user model in Chapter 33 along with exactly the same parameterization techniques discussed in the high-value and area-of-concern designs in Chapter 38.

Bayesian Ellipgrid

In standard ellipgrid applications, such as the hotspot search routines, the underlying assumption is that the elevated zone does exist, and the grid will discover it with some prescribed probability. In this sample design, we do not assume that the elevated zone does exist. Instead, the user provides probability coverage for the entire site, indicating spatially where the zone is likely to be. This is done by creating a user-defined probability map. SADA then uses this probability map as the basis for a revised ellipgrid approach.

In the ellipgrid world, since P(E) = 1 for every ellipsoid, the problem is reduced to a geometric calculation. This same geometric calculation can be used to infer the number and location of samples when prior knowledge is introduced.

The informed ellipgrid approach first accepts the probability map as previously described, along with the standard parameters that describe the size (and possibly shape) of the elevated zone of concern. When the routine is executed, SADA first creates polygons encompassing each zone individually. Within each polygon, the original ellipgrid code already available in SADA may be utilized to consider the probability within each polygon in the following manner:

When including probabilities that an elevated zone exists at all, the discussion becomes more intuitive to consider the probability that an elevated zone exists and is missed rather than the probability that an elevated zone exists and is found. This is particularly true for the search for contaminated zones, and it transitions well into the other SADA discussions, which consider the probability that an elevated zone exists given none was discovered by sampling. Therefore, the algorithm is described from this later vantage point.

The original ellipgrid code needs the value for probability of finding the elevated zone. This must be adjusted to account for the probability that it might not be there at all. In particular, we must determine the equivalent probability of discovery assuming that the elevated zone is definitely there. To accomplish this we begin first with the following definitions:

F = The object is found

DF = The object is not found

E = The object exists

DE = The object does not exist

In reality, an object will exist or not. Our probability of finding it can be completely enumerated as:

$$P(F) = P(F| E) \times P(E) + P(F|DE) \times P(DE)$$

This indicates that the probability of finding an object is really a function of our chances of finding it when it's definitely there P(F|E) and our chances of finding it when it is definitely not there, P(F|DE). The latter is obviously zero, but we include it in the mathematics to provide a complete enumeration. In order to properly qualify P(F), the probability of finding it when it is there, P(F|E), must be adjusted by the probability that it's there at all. This is done by multiplying P(F|E) by P(E). The same is true for P(F|DE).

As mentioned previously, in the latter term we have P(F|DE) = 0. So the equation can be reduced to simply:

$$P(F) = P(F| E) \times P(E)$$

Now, P(F) is really the probability parameter the code is expecting. That is, the probability of discovery. If P(E) = 1 then we simply have P(F) = P(F|E), and we have a normal ellipgrid situation.

It is certainly true that to extend the model to account for the probability that it might not be there, we would just need the user to provide P(E). This is certainly possible, but the meaning of P(F|E) is not quite intuitive. This is different from asking the user to specify the probability of finding an elevated zone, P(F). So, for Bayesian ellipgrid, we ask "What is the probability that we miss it when it's really there?" In other words, what is P(DF|E)? This question really turns the question into one of an acceptable risk level for the user. All that is required is to move from P(DF|E), P(E) to P(F).

We have that:

$$P(F|E) + P(DF|E) = 1$$

So, applying a bit of trivial algebra we have:

$$P(F|E) = 1 – P(DF|E)$$

Plugging into the equation above we have:

$$P(\text{you find it}) = [1-P(DF|E)] \times P(\text{exists})$$

So, in the Bayesian ellipgrid approach, P(exists) is provided by the user-defined prior probability map. The user must also provide the level of risk they are willing to take, P(DF|E). The equation above then provides the equivalent P(F) needed for the ellipgrid model.

The standard probability map provides the P(A). Here, P(A) is a constant within each polygonal region. The value P(A,B) must be provided by the user. From a technical standpoint, the algorithm calculates the P(C|A), where C is the probability that we detect the elevated zone. This value is automatically calculated for the user by SADA behind the scenes. At this point, there is a constraint imposed by the current version of SADA that says P(C|A) must be greater than 10%. That is, the probability that an elevated zone exists and we missed it cannot be less than 10%.

The remaining parameters specify the size and geometry of the elevated zone and are the same as the standard hot spot search model.

Suppose we have a region that has a probability of containing an elevated zone of only 50/50. Suppose further that we wish the chance of missing such an elevated zone to be no more than 10%. What is the grid size that would accomplish this for a circular zone of radius 100ft?

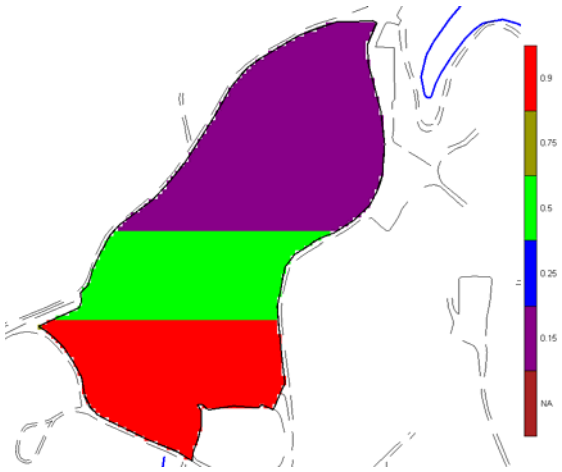
In SADA, we would first create a user defined map that includes the 50/50 region of interest. Then, we would choose the Bayesian ellipgrid model and indicate a probability value of missing the hot spot equal to 10. SADA then performs any necessary transformations behind the scenes and produces a map of the necessary grid spacing.

Then, for each polygonal area, SADA will calculate the grid required to meet the user need for P(A,B) given P(A). This number will usually be smaller than the standard ellipgrid model, which assumes that the elevated zone does exist. Similarly, areas with smaller probabilities of containing an elevated zone will be searched with less frequency than those with higher probabilities. This makes sense, as intuitively one will spend greater resources in areas of greater potential payoff.

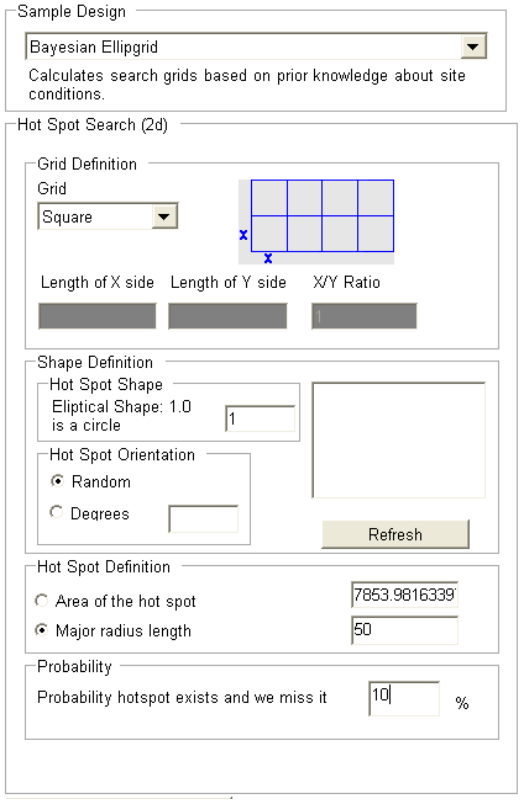
Let’s do a quick example.

We recommend that you first read the section on hot spot search designs in Chapter 39. The parameters and approach there are almost exactly the same as here. Therefore, we will cover them here very briefly.

Open the file BayesianEllipgrid.sda. When the file opens, you are presented with a probabilistic user model called Probability Exists. This surface-only site has been populated with three value regions (0.15, 0.5, and 0.95), corresponding roughly to “I don’t think it’s here,” “I don’t know if it’s here,” and “I think it’s here,” respectively.



Select the interview *Develop sample design*, click on the *Set sampling parameters* step, and choose *Bayesian Ellipgrid* from the drop-down list of available sample designs.



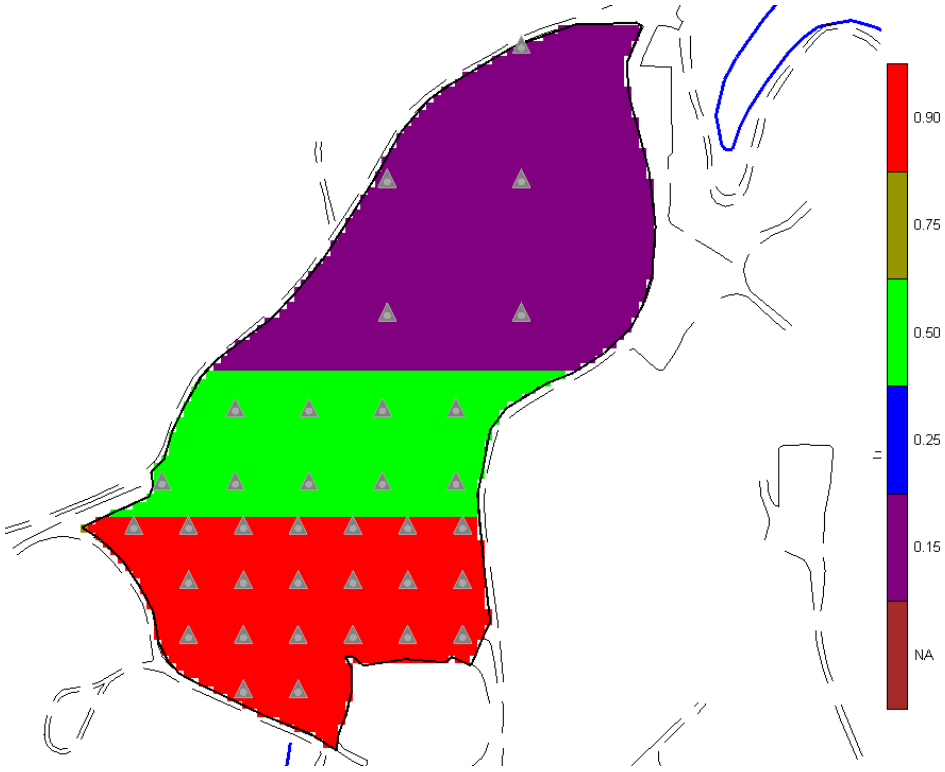
We wish to search this site for a circular object with at least a 50ft radius using a square grid. We don't think it exists with equal probability everywhere on the site, but when we're done we want the chance that it exists and we simply missed it with sampling to be only 10%. Notice that this is fundamentally different than the probability parameter in standard ellipgrid. There, the probability parameter was the probability that we discover it (assuming it does exist).

We begin with the *Hot Spot Search (2d)* parameter block. Bayesian ellipgrid technically does apply to a 3d model, but all layers must be the same model. In a 3d context, you are drawing the probability a little differently. In 3d when all layer models must be the same, you are really saying the chance is X% that it will be found somewhere below this area of the site. Dealing with

a sampling model that can also use information about how far down it may be is the subject of a future research effort.

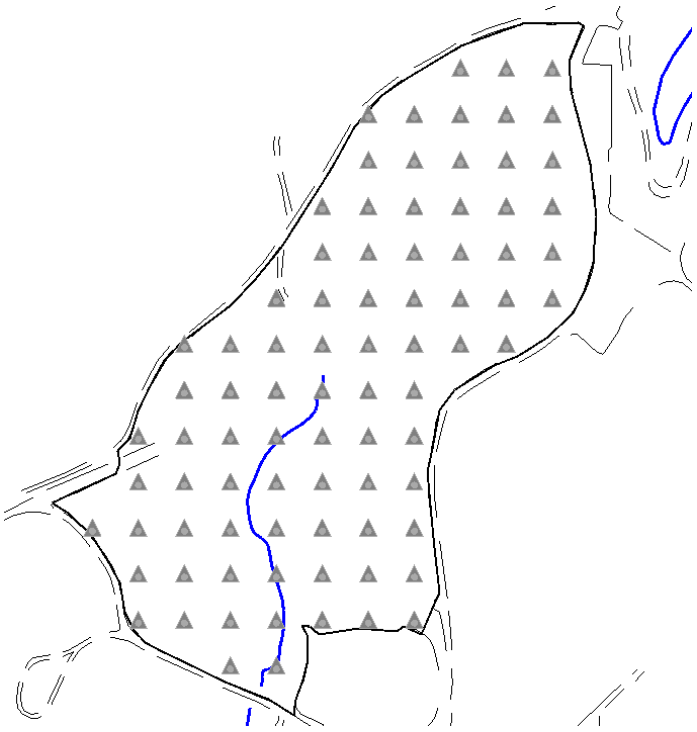
Select a *Square* grid. The lengths are irrelevant, since they will all be the same for a square. In the *Shape Definition* box, we will choose a value of 1.0. This is the ratio of the major to minor axis. For a selection of 1, major/minor = 1, which means that we are talking about a circle. Press the *Refresh* button to see a preview. Since the object is circular, the orientation parameter is irrelevant. Leave it as *Random*. Under the *Hot Spot Definition*, select *Major radius length* and choose 50 (ft). Finally, under the probability that a hot spot exists and we miss it, select 10%. This means that there is a 90% chance that either it doesn't exist at all, or it does exist and we find it. Press the *Show The Results* button.

SADA places 37 new samples in accordance with the spatially delineated probability that the elevated area exists at all.



This set of samples taken together produces a 10% chance that the hot spot does exist and we missed it by sampling. Notice in the northern region, there is only a 15% chance the object exists at all. Yet with a 10% limit, we'll need to take at least a few samples. In the southern area, we obviously spent more effort sampling. In fact, the closer the probability of existence is to 1, the closer we are to just a traditional ellipgrid model.

By comparison, the traditional uninformed ellipgrid requires 87 samples to search for a hotspot of the same size with a 90% probability of discovery.



This is a 57% reduction in sampling requirements gained strictly by acknowledging some prior knowledge about the site.

Check and Cover Sample Design

Check and cover sample design is a new sample design that arose during the development of SADA. A manuscript for this design is in preparation for *Journal of Environmental Modeling and software* (Stewart, 201_). The essential details are also presented here. Check and cover seeks to provide an exploratory balance between sampling areas that are of the greatest concern while providing some coverage in those areas thought unaffected.

Suppose that an investigator wishes to place samples where contamination is known or suspected to exist. This appears to be a reasonable course of action, but validity depends on how close expert opinion or available lines of evidence coincide with reality. For this reason, investigators may wish to also place some samples in those locations believed to be uncontaminated. Sampling at both locations does two things. First, it mitigates the risk that current lines of evidence are wrong. Secondly, if contamination is encountered in anticipated locations, samples in uncontaminated regions can provide some geographic limit to how widespread the contamination may be. This approach to sampling is biased by design, since the intent is to locate and characterize the extent of contamination. Any attempt to impart some statistical test or moment on this data may first require de-clustering. In summary, two simultaneous objectives are at play:

- Check: Sample where contamination is known or suspected to exist.
- Cover: Provide some sample coverage across the rest of the site.

This problem is extraordinarily close to a defining problem found in an entirely different application domain. Location theorists have long studied the *facility location* problem. The facility

location problem is spatial by nature and concerns finding the optimal or near-optimal locations for certain activities to occur (Miller and Shaw, 2001). In its most general sense, one has a spatial domain in which importance, interest, or value varies across space. One can then respond to this spatially-varying importance by locating an event or activity that maximizes the return on investment. Most often, this problem is formulated within an economic context where the spatially-varying interest is *demand* (e.g., consumers) and the activity is *supply* (e.g., retail) placement. There, the objective is to optimally locate supply facilities within the demand field to minimize some accessibility constraint, whether economic, distance, or a combination.

Within an environmental context, the spatially-varying interest is the likely location of contamination, and the activity is sampling. In particular, the CCM provides both the spatial domain (i.e., the site) and the spatially-varying interest pattern (i.e., likelihood of contamination). Sampling is then the activity that responds to this spatially-varying demand for information.

Facility location theory provides a number of mathematical approaches to this problem. One of the more viable approaches is the *Median problem* formulation. Median problem algorithms seek to minimize the demand-weighted distance between locations of importance and the associated activity. In this approach, a balance is struck across the varying levels of importance (i.e., likelihood of contamination) and the activity that serves them (i.e., sampling). Normally, discussions of facility location algorithms are described using the economic language of supply and demand for retailers and consumers. Those have been used here as well to introduce the concepts. For the remainder of this discussion, the language of environmental sampling will be used instead.

One of the most commonly applied approaches is the *p-median* problem. Drawing on the formulation presented in Neema et al. (2008), the continuous case p-median (within an environmental sampling context) is formally written as:

$$\text{Minimize} \quad \sum_{\text{i-1}}^{\text{p}} \sum_{j=1}^{n} a_{ij} d_{ij} w_{ij}$$

$$\text{subject to} \sum_{\text{i=1}}^{\text{p}} a_{ij} = 1,\ j = 1, \ldots, n,$$

where:

p = number of samples,

n = number of grid nodes in contamination concern model,

i = sample location point,

j = grid node point,

w_j = likelihood of exceedance at node j,

$$a_{ij} = \begin{cases} 1 \text{ if node j is associated with sample i} \\ 0 \text{ otherwise} \end{cases},$$

$$d_{ij} = \left[(x_i - x_j)^2 + (y_i - y_j)^2 \right]^{1/2}$$

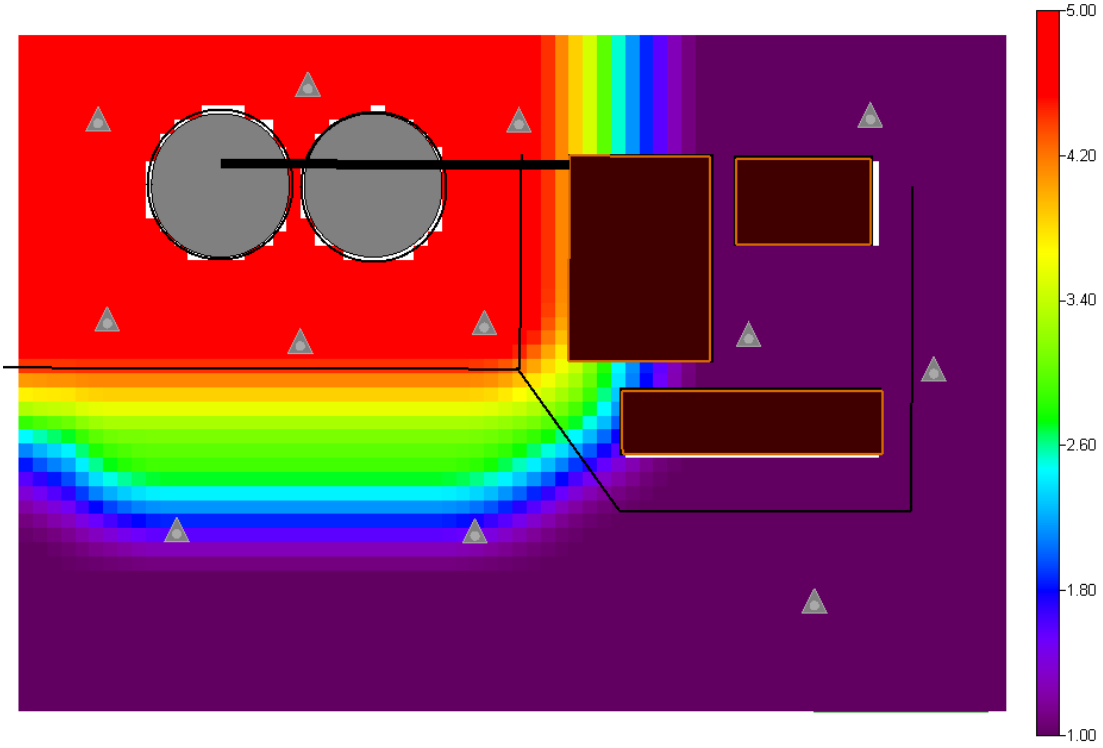
where:

(x_i, y_i) is location of sample point i

(x_j, y_j) is location of grid node j

The solution to the p-median is classified as *nondeterministic polynomial complete* (*NP-complete).* This means that an exact or optimal solution technique is not feasible for the solution due to the potentially enormous computing time. As an alternative, heuristic approaches have been developed that can produce near-optimal solutions in a feasible amount of time. One common requirement in these heuristic approaches is that an initial guess be made. For example, if one wishes to place P samples, one must begin by making a guess at where those P locations might be. The guess can be rough, perhaps even naïve, in some cases. The algorithms will then move the locations toward an optimal solution.

The following shows a p-median placement of 12 samples for a hypothetical site using the existing contour map and an initial grid guess. P-median assumes that the underlying demand map is accurate and pursues those areas of interest.



Applications of p-median to environmental contamination are already available in the literature. For example, Meyer and Brill (1988) apply a variation of p-median, called the maximum covering location problem to the optimal placement of groundwater wells. More recently, p-median is applied to ecological diversity sampling (Hortal et al. 2009). In both of these scenarios, there is a spatial model with some continually varying attribute of importance (e.g.,

modeled groundwater concentrations or ecological diversity). P-median is then applied to optimally locate a set number of samples.

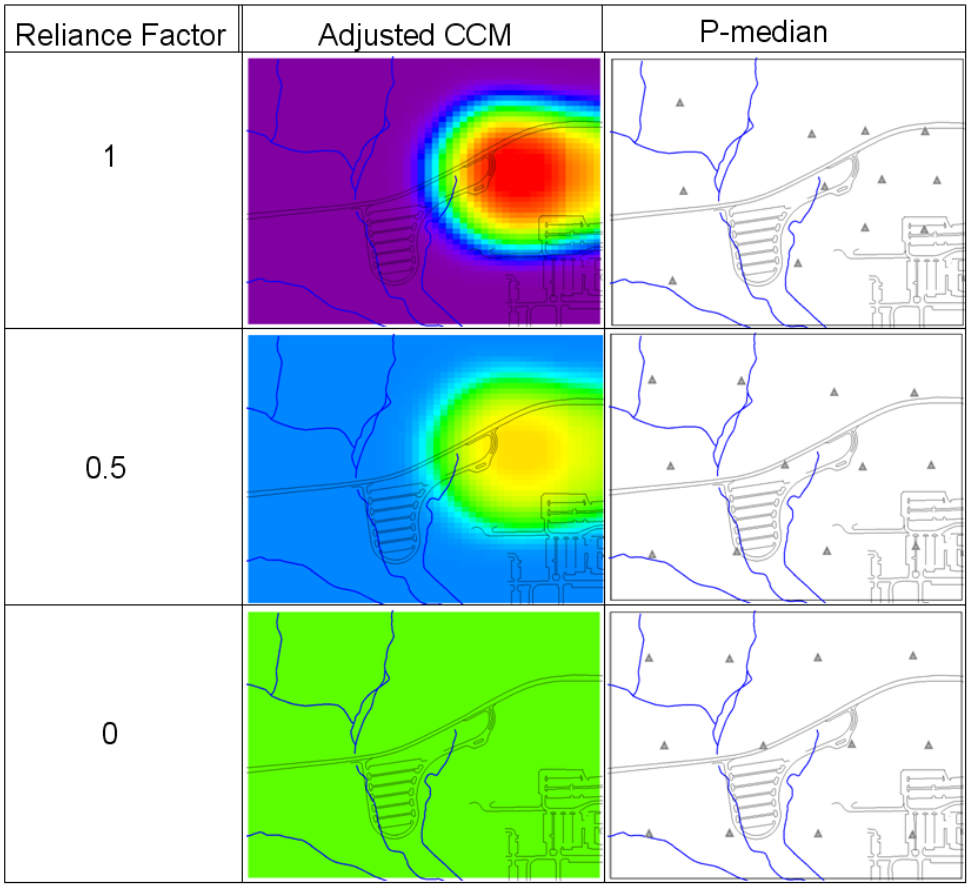
There may be some doubt about the validity of the underlying model, however, particularly in the scoping phase. What is needed is a way to allow the map to influence the decision without completely relying on it. Check and cover extends the p-median algorithm by accounting for the investigator's desired reliance on the contamination concern map. In practice, the lines of evidence that inform the map may have varying levels of credibility. For maps with low levels of credibility, one may wish to rely on the map less and migrate away from a map informed p-median approach towards an uninformed grid design. On the other hand, strong lines of evidence may justify a strong reliance on the map and p-median prevails. Both of these are possible under check and cover. Credibility is a qualitative component to the process. While it is difficult to measure or quantify such a fuzzy parameter as credibility within a practical application, it can provide great flexibility in the investigation.

If the focus is on whether or not selected points may exceed a decision criteria, then from a probabilistic viewpoint, the case of "no known information" corresponds to 0.5. Therefore, if one wishes to disregard all information about the location of contamination, a 0.5 value would be placed at every node.

Check and cover uses this principle to balance the location of p-median samples. Given a map, a reliance factor (0-1), and a know-nothing value (e.g., 0.5), check and cover will adjust the map values toward a "know nothing state," as the reliance factor decreases from one to zero. More specifically, the equation below is used to transform the CCM map according to the reliance factor.

$$Map_{adjusted}(i) = Map_i + (S_{nothing} - Map(i))(1 - R_{factor}) \;\; for\; i = 1,\ldots,N_{nodes}$$
$$R_{factor} \in [0,1]$$

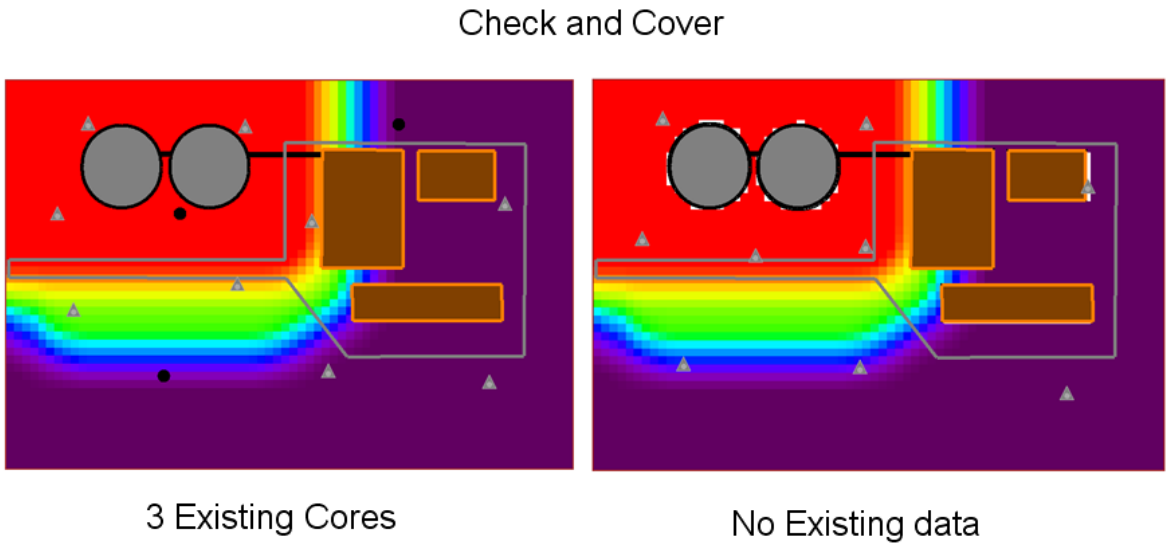
In this equation, Map(i) is the model value at node i, and $Map_{adjusted}(i)$ is the adjusted value for the reliance factor, R_{factor}. $S_{nothing}$ is the state value for "no information," and N_{nodes} is the number of nodes in the map. Notice that, when the reliance factor is 1, the $Map_{adjusted}$ and the Map are the same. When the reliance factor is 0, the $Map_{adjusted}$ becomes the $S_{nothing}$ state ($Map_{nothing}$) everywhere. For intermediary values between 0 and 1, adjusted values between Map and $Map_{nothing}$ are used. This is demonstrated in the following figure where the map is adjusted for select values of R_{factor} and a "know nothing" value of 0.5.



Notice that as the adjusted map moves from a “knowledge” state to a “no knowledge state,” the samples move from a clustered design to a triangular grid just as one should expect.

Accounting for Existing Samples

Check and cover can account for existing samples in SADA 5.1. The algorithm has been equipped to recognize the location of previous samples in the optimization routine. In the following figure, check and cover is applied to a site under two scenarios: 1) 3 cores have already been taken, and 2) no previous cores are available.

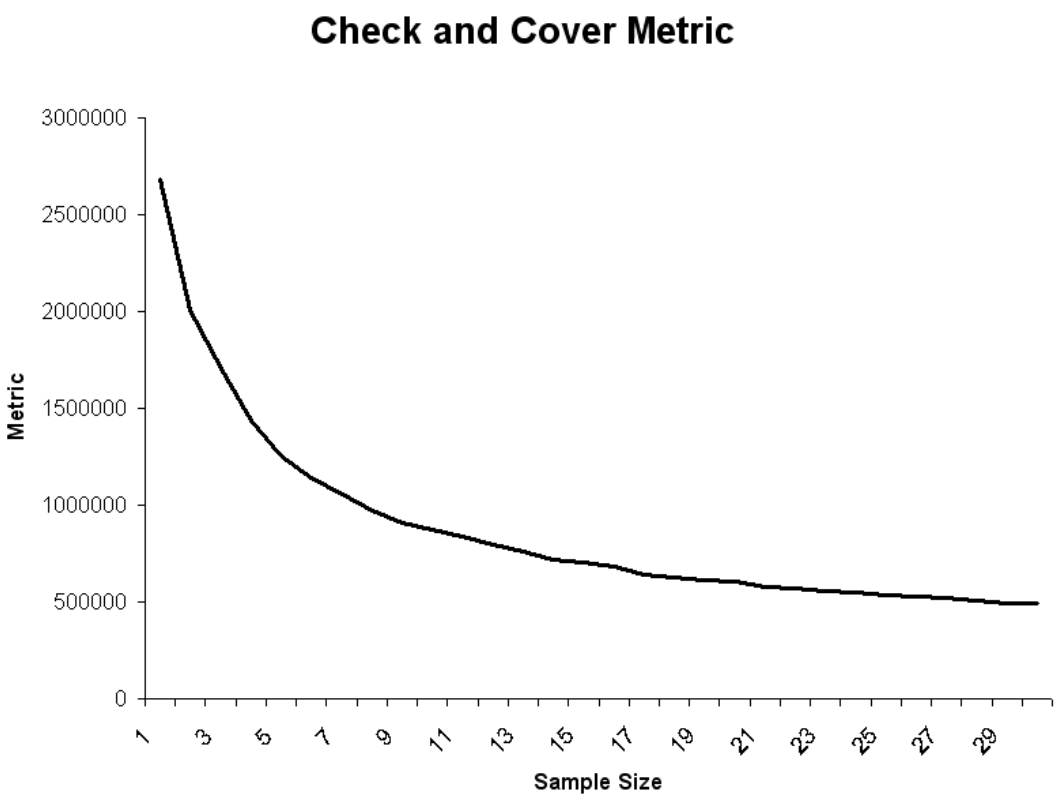


Determining Number of Samples

With a continuous framework such as the maps above, where emphasis is placed on understanding processes at a very granular level, determining the number of samples can be problematic. One way is to estimate the number of samples based on their "worth" or based on some metric of interest (e.g., Back, et al. 2007). There are several ways in which the number of samples might be determined in this phase.

Metric Based

The size of samples during a scoping phase may be a qualitative decision influenced primarily by cost considerations. Some sample designs, such as check and cover, can provide a *metric based* determination of the number of samples. Designs that place new samples based on minimizing or maximizing some value can report the progress as each new sample is produced. For example, check and cover seeks to minimize the maximum sum of value-weighted distances. In the following figure, the addition of each new sample produces a smaller sum. Eventually, the curve will flatten out, indicating a smaller return on investment for each additional sample.



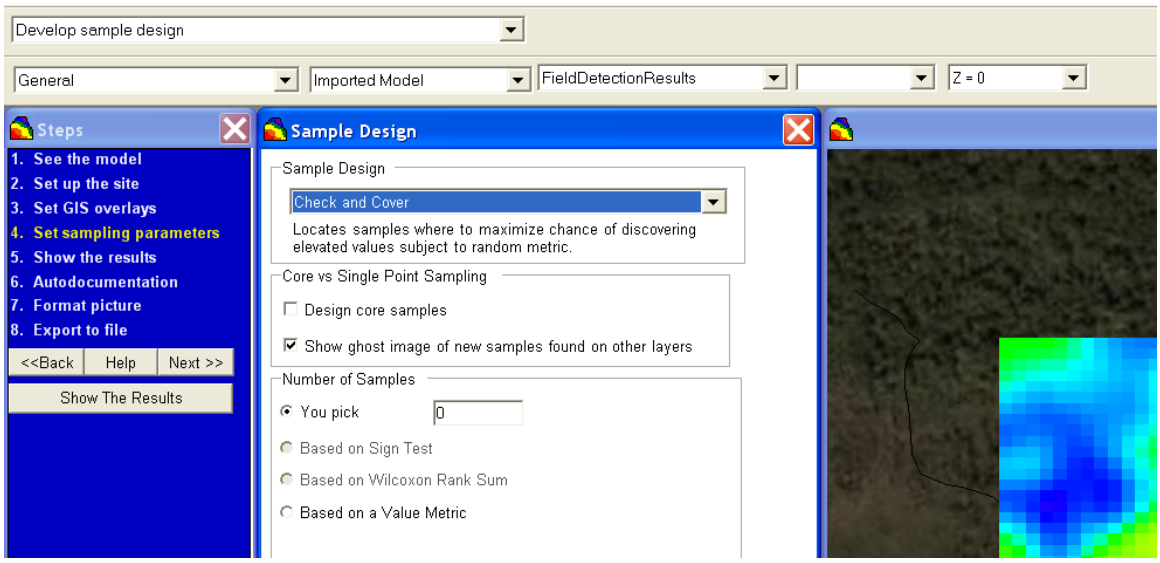
Performing Check and Cover in SADA

Check and cover requires that a spatial model be present. The user can either select a stored model or a point data set. In the case of the latter, a spatial model must be selected to first interpolate the data as part of the sample design process.

In this demonstration, a stored model approach will be used. Open TwoDimensional.sda and select *Data→Import Gridded Data*. In the open dialogue box, navigate to FieldDetectionResults.csv and select open. Give the imported model the name "FieldDetectionResults" and say *OK*.

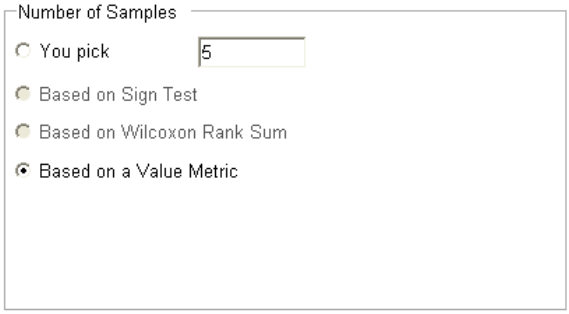
When SADA returns control to the user, select *Imported Model* from the data types box and *FieldDetectionResults* should be the first data set under this type. *Select File→Save as…* and name the SADA file "CheckAndCover.sda."

From the interview drop-down list select *Develop sample design* and in the sample design parameters step, choose *Check and Cover* from the drop-down list of available designs.



The first block instructs SADA whether to design core samples or not. If the user selects *Design core samples*, then samples will be placed above and below any optimal sample location to form a continuous core. Don't select this option now as this is a two dimensional example. If the option is selected under a 2d scenario, the algorithm will work normally since SADA recognizes that cores don't apply in 2d scenarios. The *Show ghost image of new samples found on other layers* option allows one to see new samples placed above or below the currently selected layer. This option can be selected or not for a 2d scenario. It doesn't matter since there is only one layer.

The *Number of Samples* parameter block tells SADA how to choose the number of samples. The *You pick* option allows the user to enter the value directly.

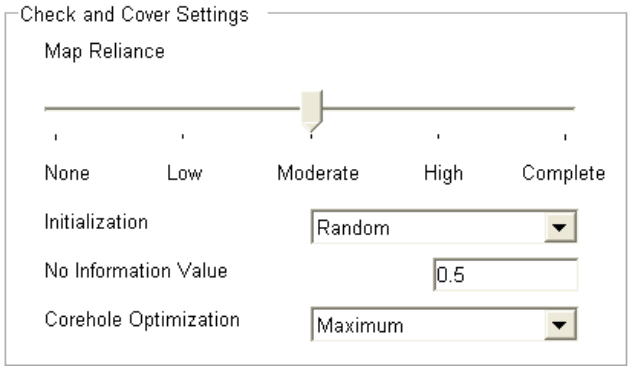


Neither Sign Test nor Wilcoxon Rank Sum tests are possible options here as they are specific to hypothesis tests regarding the median of the data. This is not the objective for *Check and Cover.*

The other option is *Based on a Value Metric* where the worth of each additional sample is valued by how much it reduces the value distance as described above. Select this option now.

The next parameter block is where *Check and Cover Settings* are found. The large slider bar allows the user to select a *Map Reliance* factor. There are 5 levels from 1 (no reliance) to 5 (complete reliance).

The next option is how to provide an initial guess. If Random, SADA places the specified number of samples randomly and then adjusts them to optimize the design. If Gridded, SADA places the samples in a triangular grid pattern first and then adjusts them.



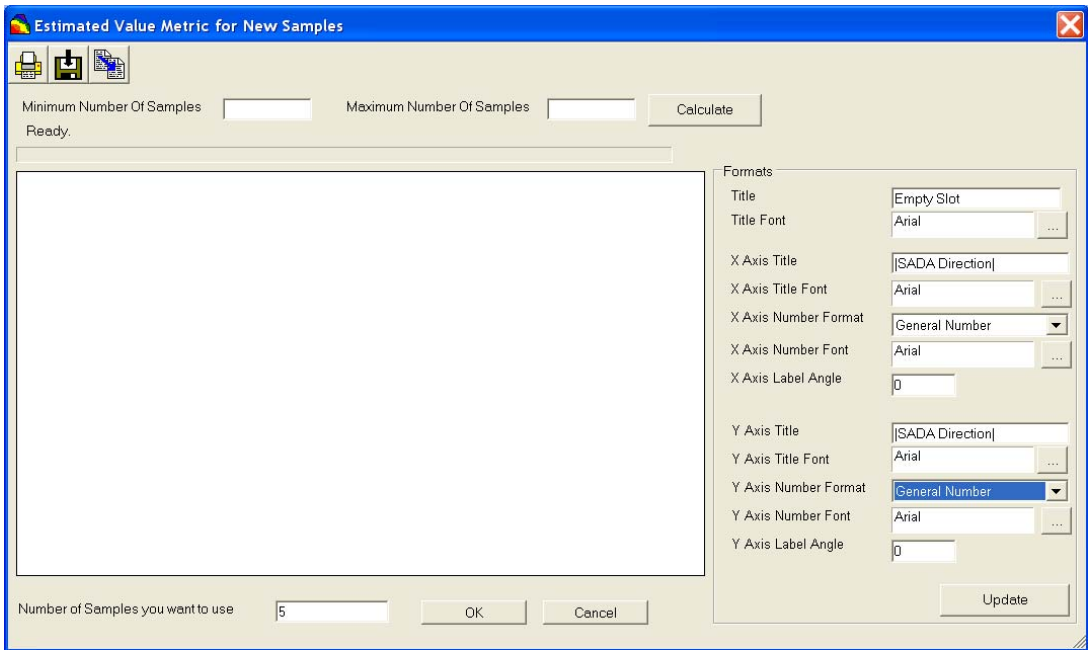
The next option is the *No Information Value* parameter. As one moves from complete to none in the map reliance setting, the map must be adjusted so that a single uniform value is obtained. This uniform value is the “know nothing” value and indicates no spatial preference in the sample design. If the underlying map is a probability map, a value of 0.5 might make sense for this value. If the underlying map is a concentration contour map, the selection is not as obvious but should be connected to some type of decision endpoint.

P-median is a two dimensional algorithm. If the method is applied to a 3d (multiple layers) scenario, the *Corehole Optimization* becomes important. If the core option is not selected, then the method is applied to the currently selected layer. If the core option is selected, then the volume content must be projected to a 2d plane at the beginning. Later when the samples are physically located, the 3d extent is recognized by the code.

There are three ways to project the 3d content to the 2d place: maximum value, average value, and sum. For any given node, a complete set of included nodes from the surface to the bottom are collected and either averaged, summed, or the maximum is selected according to the user’s choice. This is done for every node to create a 2d map. *Check and Cover* operates on this map and then extends the result back into 3d at the end.

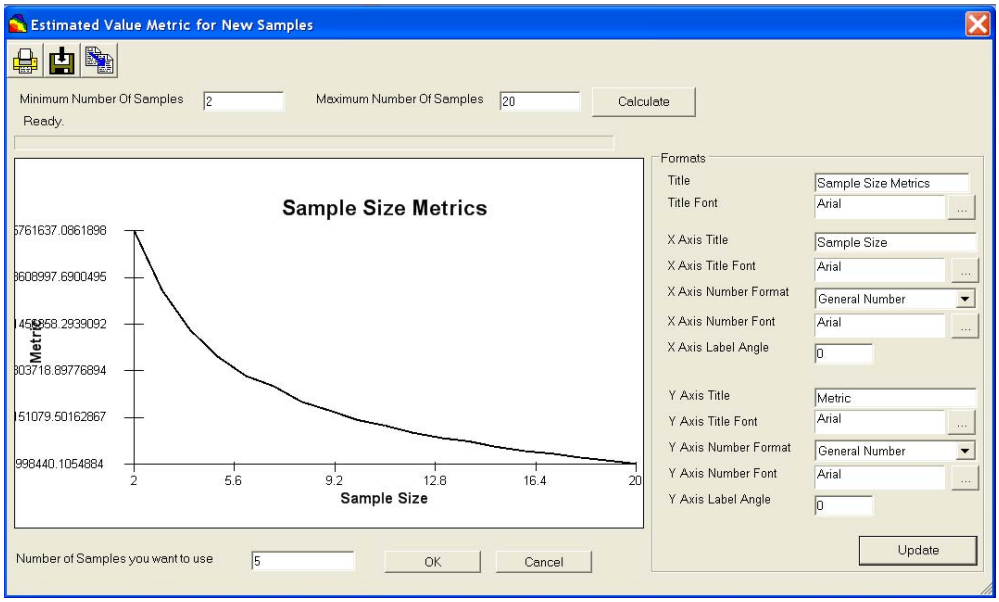
For this demonstration, select *Complete* reliance, *Grid* initialization, *No Information Value* of 0.5 and a *Corehole Optimization of Maximum*. Press *Show The Results*.

Since the *Based on Value Metric* was selected, the user is presented with the metric determination window first.



In this window, specify the *Minimum* and *Maximum Number of Samples* in the top two parameter boxes. For now choose a value *Minimum* of 2 and a *Maximum of* 20. On the right side are formatting options. If an older SADA file is opened, there may be default and unfortunately nonsensical values in these parameter boxes. The sample metric graph in this version of SADA is a new picture type and older SADA files don't know about it yet. Therefore some default values are selected.

To correct these enter "Sample Size Metrics" in the *Title* parameter box, "Sample Size" in the *X Axis Title* box, "Metric" in the *Y Axis Title* box. Press the *Calculate* button.



To print this graph, press the Print toolbar button. To save the sample size/metric values as a comma delimited file, press the save to disk button (disk with arrow) and provide a file name. To copy this image to the clipboard, press the copy button (last toolbar button).

To make other formatting changes, enter them into the parameter boxes on the right and simply press the *Update* button.

As the number of samples increase, the decrease in the value-weighted distance sum decreases. At some point, the additional decrease is not worth the additional cost in adding another sample. In this example, 16 samples will be the decision point. Enter this value into the *Number of Samples you want to use* box and press *OK*. SADA responds by placing the samples according to the adjusted map/p-median algorithm described above.

References

Back, P.-E., L. Rosen, et al. (2007). "Value of Information Analysis in Remedial Investigations." AMBIO: A Journal of the Human Environment 36(6): 486-493.

Hortal, J., M. B. Araújo, et al. (2009). "Testing the effectiveness of discrete and continuous environmental diversity as a surrogate for species diversity." Ecological Indicators 9(1): 138-149.

Meyer, P. D. and E. D. Brill, Jr. (1988). "A Method for Locating Wells in a Groundwater Monitoring Network Under Conditions of Uncertainty." Water Resour. Res. 24.

Miller, H. and S.-L. Shaw (2001). Geographic Information Systems for Transportation. London, Oxford University Press.

Neema, M., K. Maniruzzaman, et al. (2008). "New Genetic Algorithms Based Approaches to Continuous p -Median Problem." Networks and Spatial Economics.

Part VIII: Miscellaneous Topics

Chapter 42: Local Index of Spatial Association Tools (LISA)

Local Index of Spatial Association (LISA) methods are a set of functions that give some sense of spatial sampling density and spatial variability. Three types are available in SADA: Moran's I, Ripley's K, and Geary's C.

Ripley's K

A formal Ripley's K implementation is used to assess the spatial pattern of point data. This method is used often in epidemiological studies to determine if there is any clustering in disease events. We use a similar method here to determine if there is any clustering in the sample data. It is more accurate to say that we use it to determine where areas of low sampling density are found. These areas may be candidates for new sample locations (see Chapter 38). This provides some measure of improvement over adaptive fill, which considers only gaps defined by nearest neighbors.

Let's begin with a description of Ripley's K. This method basically measures the average number of points within a certain distance of each other. In particular, a window of size *h* is centered about each point and the number of points found within the window is computed. This window is then moved to every point and the number is recomputed. These values are then averaged, producing an average value for the distance window *h*. The estimator for K(h) is given as

$$\hat{K}(h) = \frac{\sum_{i=1}^{N} \sum_{j=1(j \neq i)}^{N} w_{ij} I(h_{ij} \leq h)}{\lambda N}$$

where λ= N/|A| where N is the number of samples, A is the area of the site and w_{ij} is a spatial weight used to account for edge effects near the boundary.

In SADA, we produce a moving window of sample counts over an extent of grid nodes.

For a specific distance of *h*, SADA will create a continuous map of count data using a defined base grid. For each grid node, SADA will compute the number of points within a distance *h*. This can provide a sense of the spatial distribution of clusters for a given distance *h*.

Moran's I

Moran's I is another measure of spatial autocorrelation, usually used in exploratory data analysis (geostatistical measures are discussed Chapter 31). The statistic is calculated in much the same spirit as Ripley's K, in the sense that a moving window of radius *d* is positioned at data points around the site and the weighted variance of data points within the window are computed. The weighting methods can vary. If this is repeated for overall measurements for a distance *d*, then we have Moran's I estimate at distance *d*. The statistic is expressed as follows.

$$I(d) = \frac{\frac{\sum_i^N \sum_j^N w_{ij}(d)(x_i - \bar{x})(x_j - \bar{x})}{\sum_i^N \sum_j^N w_{ij}(d)}}{\frac{\sum_i^N (x_i - \bar{x})^2}{N}}$$

In SADA for specific value d_0, we show the local variance within a window centered at each node.

Geary's C

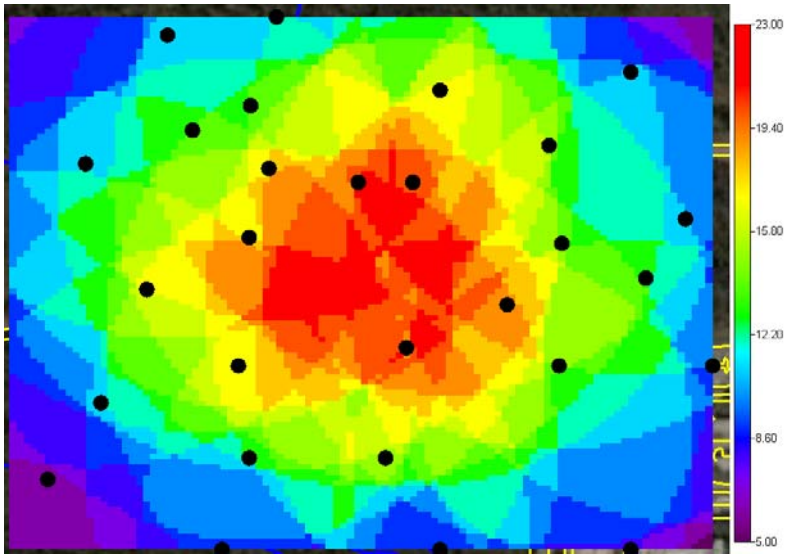
This is another method in assessing spatial variance. In Geary's C, we compute semi variance as follows.

$$c(d) = \frac{\frac{\sum_i^N \sum_j^N w_{ij}(d)(x_i - x_j)^2}{2\sum_i^N \sum_j^N w_{ij}(d)}}{\frac{\sum_i^N (x_i - \bar{x})^2}{(N-1)}}$$

In SADA for specific value d_0, we show the local variance within a window centered at each node

Using LISA Tools in SADA

Open the file LISA.sda. Make sure you've selected *General*, *Soil*, and *Ac-225*. Select *Draw a LISA map* from the drop-down list of available interviews. First, we'll set up the grid. The moving windows in each of these three methods will be centered over each node of the grid. Click on the *Setup the Grid* step. Note that the grid is 100x100. Now click on the step *Set LISA parameters*. Set the search radius to 1000. First, select *Ripley's K* and then press *Show The Results.*



The result is a map of the number of data encountered within a window size of 1000 when centered over each grid node. Now try both *Moran's I* and *Geary's C* yourself.

Chapter 43: Graphical Tools

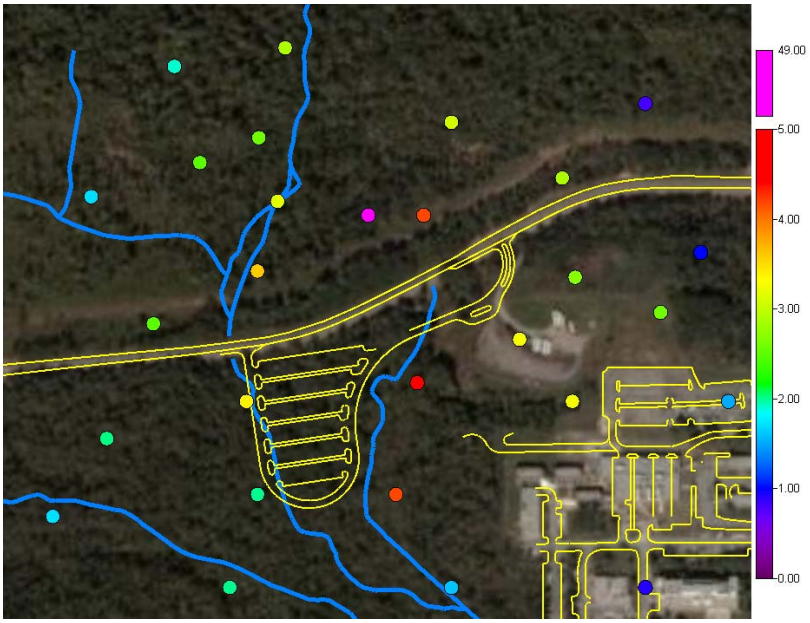
The display of geographical data presents an excellent opportunity to maximize the use of color. This really cannot be overstated, as color selection can not only present the end consumer with a clear and concise picture but can assist and even improve the analysis itself. Color selection, symbology, and the like have been long standing concerns in the field of cartography well before GIS. Cartography is largely concerned with the appropriate construction of maps and is rooted in a combination of both aesthetics and technique (see, for example, Robinson et al. 1995 and Harvey 2008). This chapter deals with some issues that often arise in the modeling and visualization of environmental data and in particular with the use of SADA. We will begin by showing you how to use the Legend Manager (a central figure in the graphics package) and continue with a host of other smaller graphical tools. Then, we will present some typical situations you may encounter.

Graphical Choices

The color choices and graphical decisions used in the representation of spatial data are very important in communicating to the decision maker the “story” that you, as an analyst, wish to convey. This really goes beyond map aesthetics (which are also important). We are more concerned here that the decision process is properly supported by the analysis, which necessarily includes map interpretations. Within SADA, we provide graphical elements that have a central focus on map interpretation. The cartographic quality of the maps may or may not be adequate for your needs. Major GIS packages, such as ESRI’s family of ArcInfo products, are dedicated to publication of quality maps. If a high degree of cartographic quality is required by your analysis, we recommend that you export your SADA results and use them in one of these of packages.

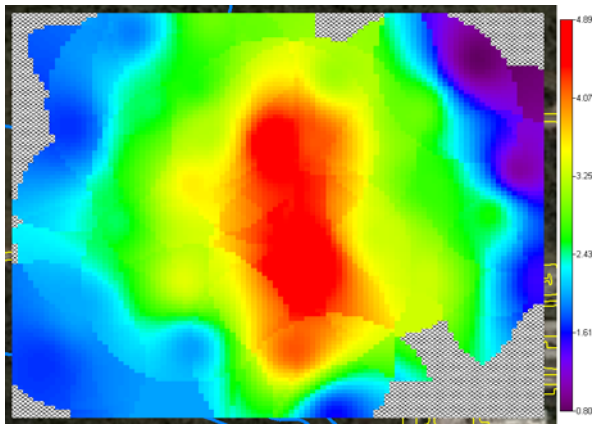
Let’s begin by showing some examples of bad and good graphical choices. What defines good and bad depends on the objective of the map. Certain choices may be excellent for some objectives and poor for others. Let’s consider the following map of measured Ac-225 samples.

Here the objective is to present the decision maker with a good sense of the distribution of sample values across the site and where they are in proximity to roads, streets, and buildings. With this in mind, the map needs some work. First, both the roads and waterways are colored black and virtually invisible with the photograph in the background. Secondly, we see only a single bright-red sample with a value of 49pCi/g. The other samples have values below 5pC/g and show up in the purple part of the spectrum. This really does not give a good sense of how the values vary in space. If we make a few simple changes, notice the difference in the result in the following image:



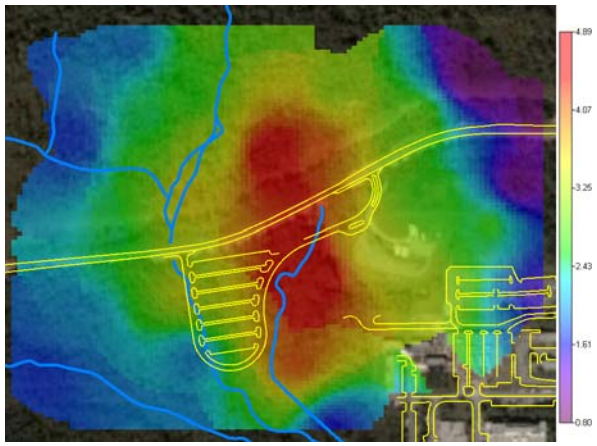
Now the roads and water are much easier to see. With the fixed legend, we can now see variation in sample values and note the highest value as bright purple. In this view, it is easier to see that a N-S trend in contamination may be present near the center of the site. In the previous image, it appeared that maybe only a single isolated hot spot was found. The numerical results in both of these images are exactly the same.

Now, let's take a look at the following model result. The objective here is to overlay the GIS elements with the estimated values for Ac-225.



The image has problems because the opaque property of both unestimated values (cross hatches) and modeled values is blocking the photograph. Geographical context is all but lost. Furthermore, the unestimated values are drawing too much visual attention. In SADA, we can

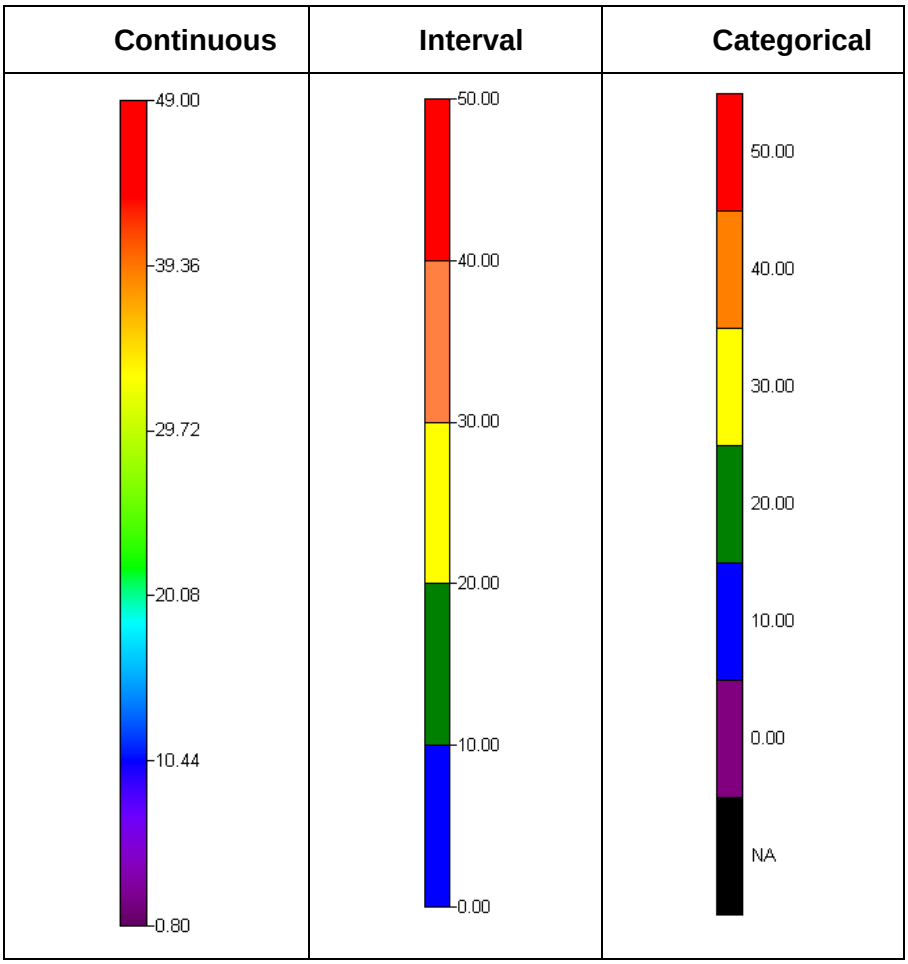
control how unestimated values are displayed, the order that results are drawn, and modify the transparency of modeled values. The following result shows the update.



We'll now show you how each of these transformations were accomplished and introduce other tips as well. Some individual principles may have been covered in other chapters, but we will bring all of it together here. To begin, please open the file GraphicalTools.sda.

Legends

The best place to begin the discussion is with map legends. Legends specify how numerical values are presented by using either a color spectrum or individually assigning colors to specific values. There are three basic legend types in SADA: continuous, interval, and categorical. The three basic forms are illustrated below.



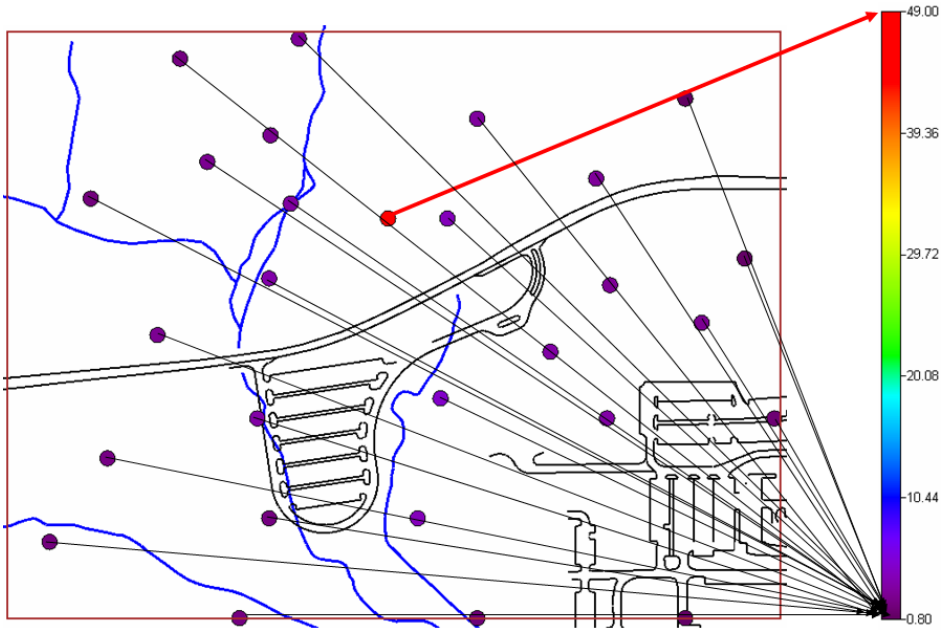
Legends are managed and customized by using the Legend Manager feature. We'll approach this topic by focusing on each type of legend (pros and cons) and show how the Legend Manager works as we go.

Continuous Legends

In its simplest form, a continuous legend stretches itself from the lowest to highest numerical value. It is therefore dynamic, always adjusting itself to the range of the currently viewed result. This has the benefit that no matter what data set you look at, the legend will always "fit."

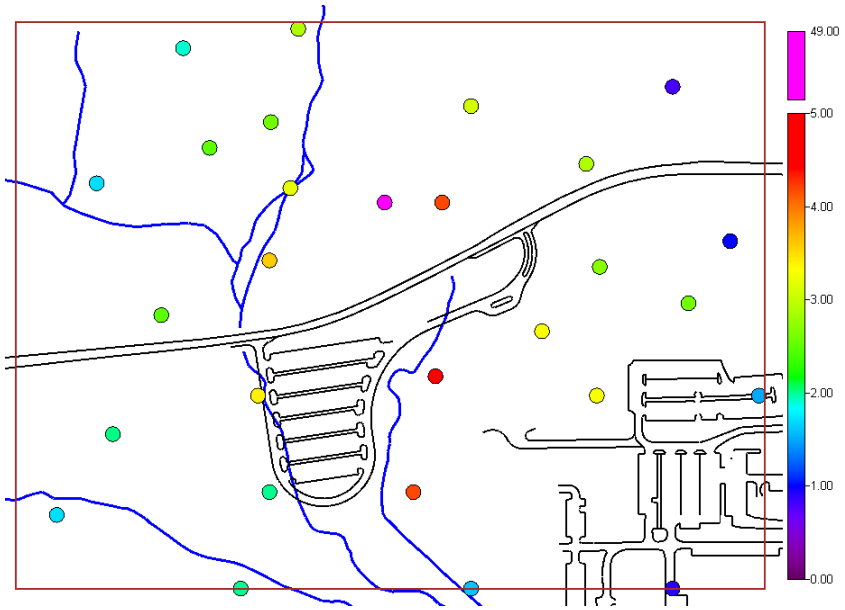
All values, however, are represented in a relative way in the sense that the high value is always shown with an alarming red color while the low value is always shown with a cooler purple color. What is high and what is low is really dictated by the decision process and not the legend. Consider the following result for some contaminant X. The concentration of contaminant X does not become an issue until it exceeds 200 mg/kg. Yet the representation of the data, particularly the red values, may suggest that the values are dangerously high even though they are half the criterion limit.

Another problem with simple continuous legends is that they give greatest priority to the highest and lowest values. As we've seen before, if a single value is significantly higher than the rest of the data, it can stretch the spectrum well beyond the majority of the values. In the following figure, each sample value is connected to its location on the color spectrum.

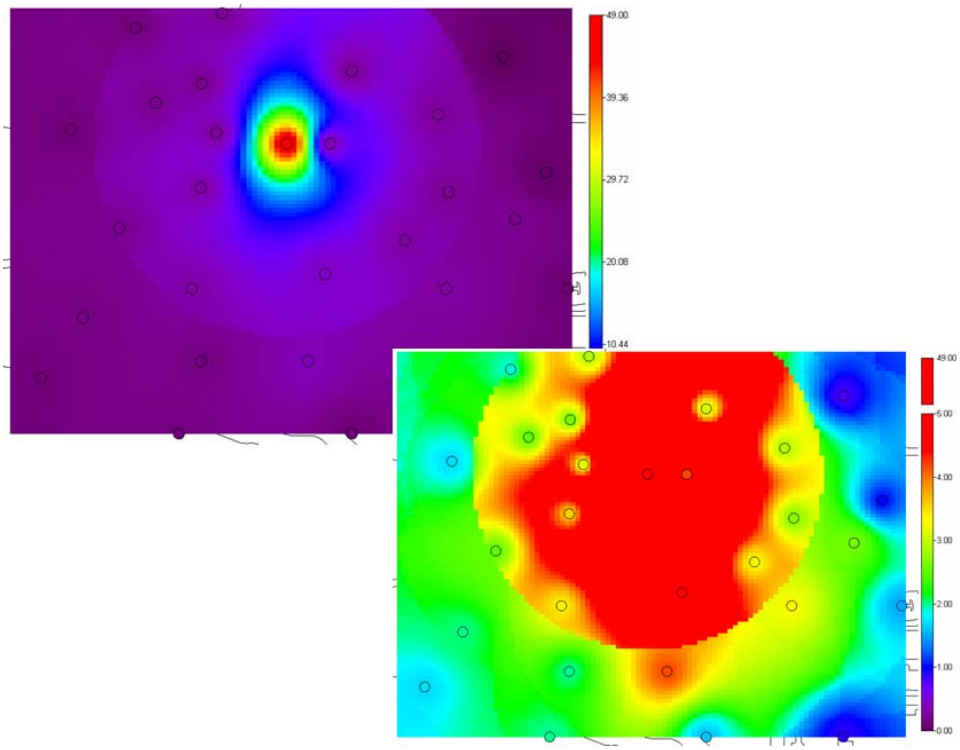


Notice that the majority of the color spectrum is not used, and any variations in the data found at the lower end of values are lost in the purple haze.

This problem is handled by *fixed continuous legends*. In a fixed continuous legend, you fix the lower and upper bound of the color spectrum. For values that exceed the upper bound, a single color is selected. The same is true for the lower bound. When applied to the same data, the variation in values is much easier to discern.



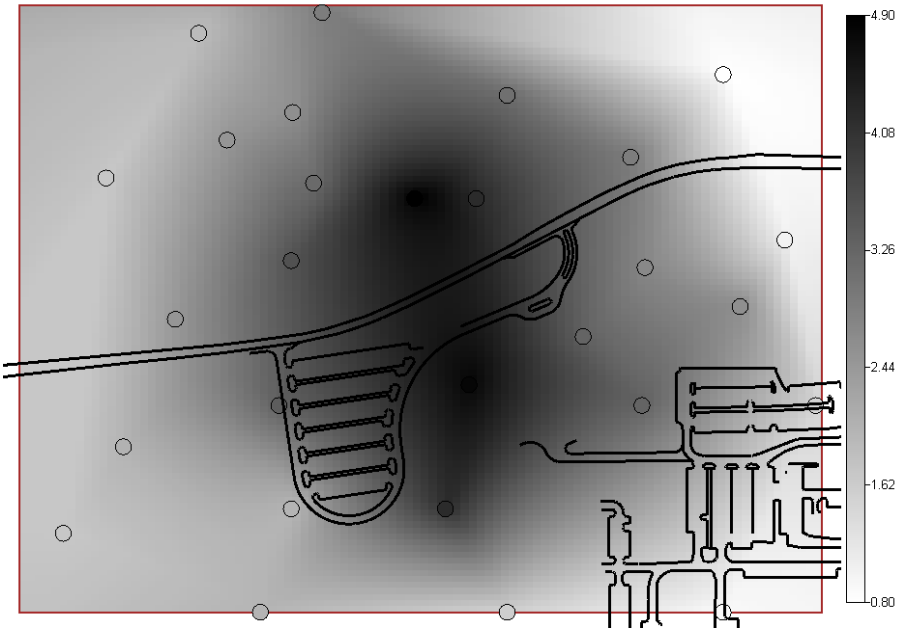
This principle can be very important in assessing how well a model is fitting the data. In the following image, we have a simple and a fixed-continuous legend applied to the same model.



In the upper-left example, the model appears to adequately match the sample values in the sense that we see purple estimates covering purple measurements. The model, however, is actually performing very poorly, especially regarding a decision criteria of 5pCi/g. In the lower-right example, the effect of the extreme value is creating a large nonsensical artifact (large circle) in the center of the image. Furthermore, regarding the criterion of 5pCi/g, this extreme value is smearing high values across areas with values well below the decision criterion. This second image should lead to an alternate parameterization of the model. These extreme value

effects are taken up in greater detail in the chapters on geospatial modeling. The important thing here is that a proper choice of legend revealed the problem.

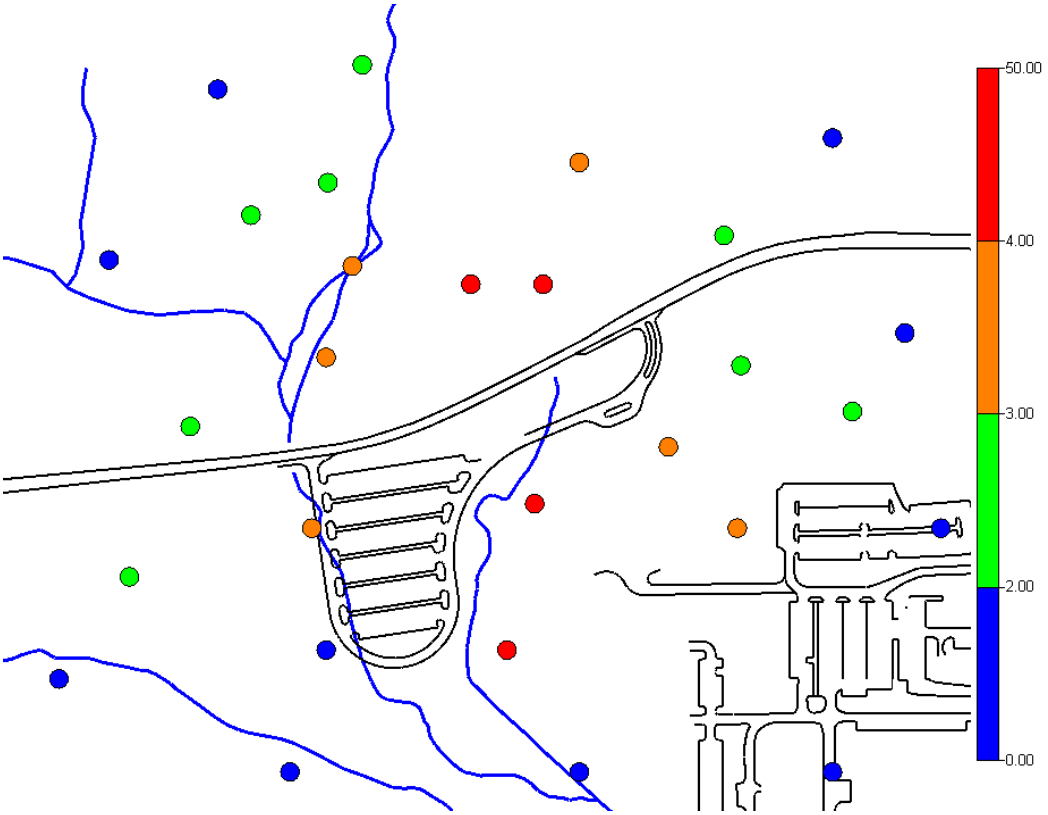
Continuous legends are offered as both color and grayscale. The ability to create simple or fixed scales is the same for both. In some cases, gray scale is preferred, particularly when submitting images to journal publications.



Interval Legends

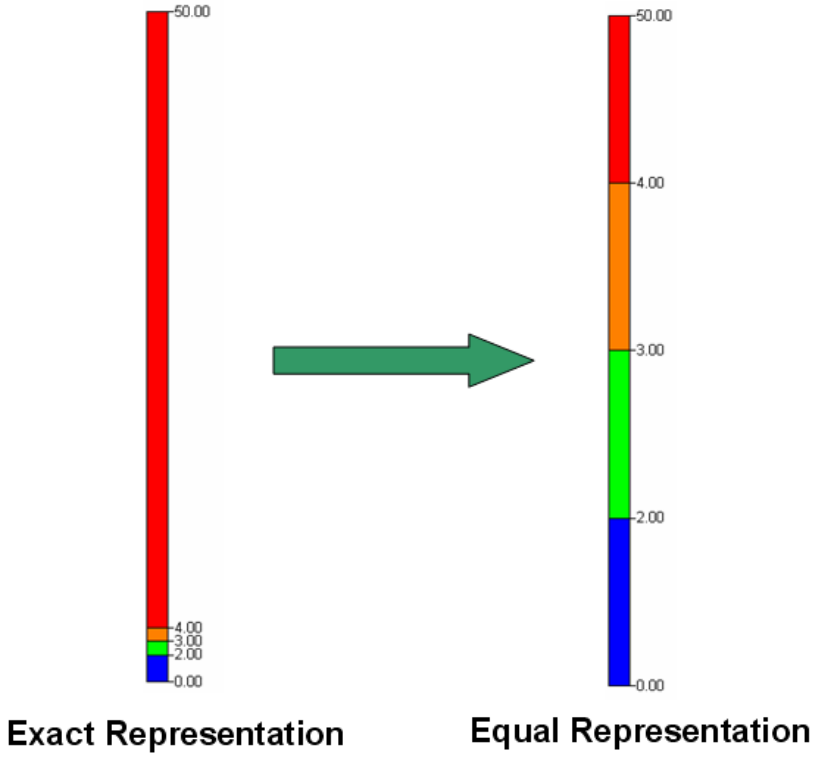
An interval legend essentially divides a continuous range of numbers into a set of smaller ranges. Each sub-range or interval receives a color. When applied to a specific numerical value, the color for the interval containing the value is used. Like fixed-continuous legends, these must also have colors for values that fall above the highest range and below the lowest range.

Interval legends are particularly useful for showing compliance with multiple decision criteria. For example, suppose that three decision criteria are under consideration for contaminant X: 2, 3, and 4 pCi/g. You could certainly apply a spatial data screen, but it would have to be applied as three separate screens. Alternatively, you could create an interval legend with ranges specified by each of these three criteria. It is then easy to see which values exceed one or more of the criteria.



In this image, any sample value greater than 4 pCi/g is shown in red. Any value greater than 3pCi/g is shown as either orange or red, and so forth.

Interval legends can contain numerically unbalanced ranges. The legend shown in the image above is an example. We have a set of ranges from 0-2, 2-3, 3-4, and 4-50. As a default, interval legends are created with exact range representation. That is, each sub-range occupies the legend space in proportion to its size. You can elect to assign each interval equal spacing, regardless of any numerical imbalance, to create a more useable view.



Categorical Legends

A categorical legend is appropriate when you have a small set of values. Usually, these values are coded values that actually represent something (1 = sand, 2 = clay, 3 = sandstone). They normally are not applied to continuous data. In the following image, we see a Landcover model just imported into a SADA file. The coded values are as follows (1 = industrial, 2 = grass, 3 = forest). When they are first imported, a default color legend is assigned.

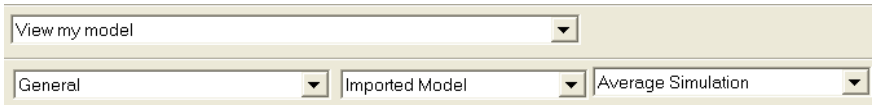
This is a good start, but applying a continuous color spectrum to coded values doesn't make much sense. A custom categorical legend was then created that colored the industrial (1) values

gray, the grass values (2) brown, and the forest values (3) green. Apply a bit of transparency to the values and we get a much better result.

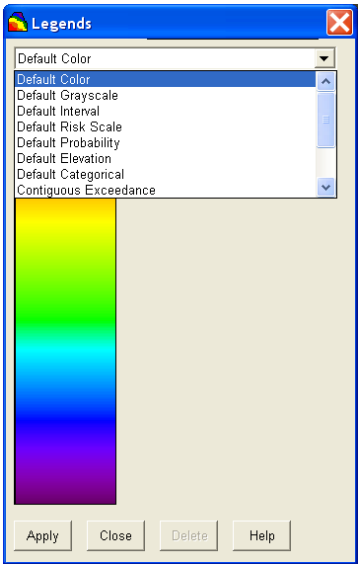
Notice that at the bottom of the legend a category called NA is added. This will be applied to any values that are not 1, 2, or 3.

Legend Manager

Legends are created, edited and deleted in the Legend Manager. Before we begin, make sure you've selected *Imported Model* and *Average Simulation*.



This is the average of a set of geostatistical simulations (covered Chapter 31, "Advanced Geospatial Methods"). *Select Graphics→Legend Manager* and the manager window appears.



At the top of the window, a drop-down list contains a set of template legends as well as choices that lead to new custom legends. Each template legend begins with the word *Default* (at the time this guide was written, *Contiguous Exceedance* is the exception) and is not editable. These

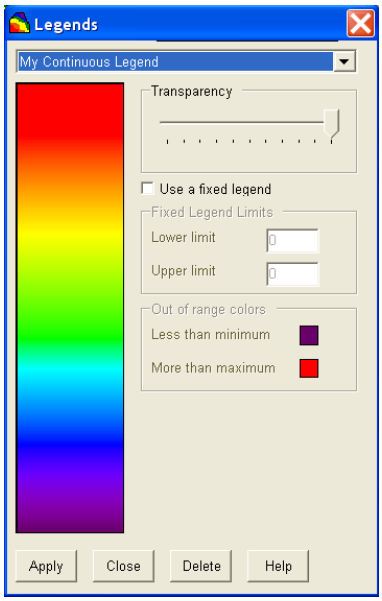
legends are primarily there to serve as a basis for your custom legends; although you can apply templates as if they were your own.

Further down at the bottom of the list, you'll find some legends called *(New…)*. These allow you to create new legends. We'll do that momentarily. First, let's go through some of the templates to see what they look like. With *Soil* and *Ac-225* selected, switch to *Default Grayscale* and observe the change in the graphics window. Switch to *Default Interval* and continue through each of the templates. Notice the dramatic differences in how each presents the same data set.

New Continuous Legend

Let's create a new continuous legend. Select (*New color palette*) in the drop-down list and the following window appears.

At the top, enter "My Continuous Legend" as the name of your new legend. In the lower drop down list are previously created continuous legends that you can base your new legend on. At the moment, the only options are the template legends. If you had created and customized others beforehand, you could select them here. SADA will make the new legend as an exact copy of this selection that you can continue to customize. Press the *OK* button. In the Legend Manager, the new legend appears. On the left is a preview of the legend and on the right are all the features that permit customization.



Transparency

Let's begin with Transparency. This feature allows you to control how well you "see through" the measured or modeled values. This is useful when you have a photographic background (for shape or DXF layers, you can simply push the SADA result to the bottom of the drawing order

and avoid using the transparency, see Chapter 4). The default for this feature is no transparency (bar all the way to the right). Let's set the transparency to about 50% by moving the glider bar to the middle. Press the *Apply* button and note that you can now see through the result to the photograph behind.

Fixed Legend

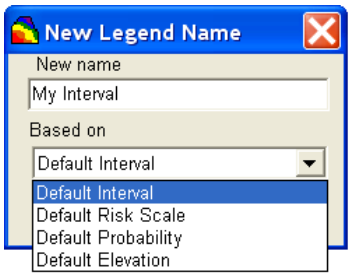
The next option is whether or not to *Use a fixed legend*. Currently, this is unchecked and so the spectrum is just stretched between the lowest and highest value. Let's check this and establish a *Lower limit* of 0 and an *Upper limit* of 4. Press the *Apply* button. The legend is now "fractured" to show that values were found above 4, as high as 4.89; and all these values will receive a red value. This color choice for the upper values makes it difficult to distinguish from those values near the top of the color spectrum. In the Legend Manager, click on the *More than maximum* color block and choose a light, baby blue. Press *Apply* and those values greater than 4 are much easier to see.

New Grayscale

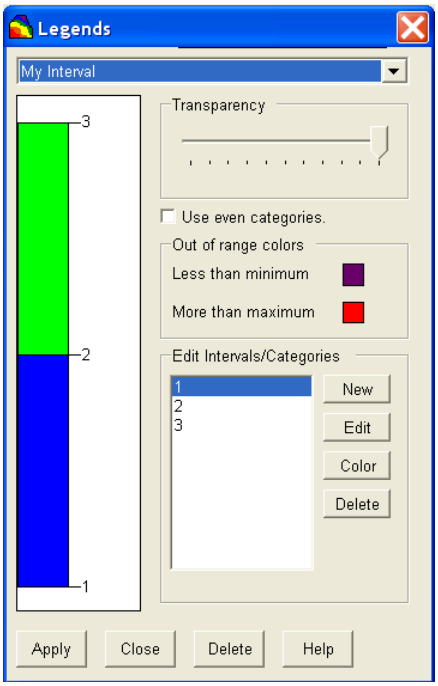
The process is exactly the same for the grayscale. Repeat this exercise by creating a new grayscale legend. To get started, select *(New Grayscale)* from the drop-down list of available legends.

New Interval

From the drop-down list, select *(New Interval)* and enter the new name "My Interval" in the *New Legend Name* window. Notice that in the *Based On* drop-down list there are 4 options to choose from.



Your choice will depend on how well it suits your current needs. For example, the *Default Risk Scale* might be a good place to start if you are modeling risk. We'll choose *Default Interval* and press *OK*.



As with continuous legends, you get a preview to the left and the customization tools on the right.

Transparency

We've seen transparency already in the continuous example and won't repeat that here.

Use even categories

At the moment, we have two default ranges: 1-2 and 2-3. These are equal in size, and so presenting them in a legend is easy; but what if the ranges were 1-2 and 2-200? If you represented them proportionally, they would appear as follows.

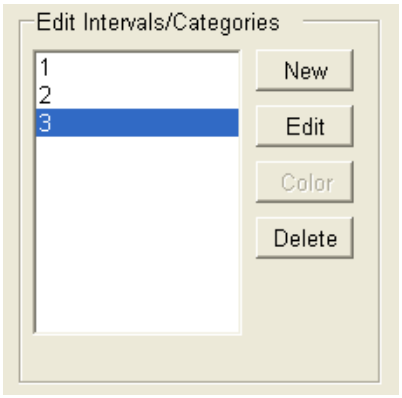
This is unacceptable. The *Use even categories* option allows you to equally represent each range in the drawing of the legend. Let's select this option now.

Out of Range Colors

As with fixed continuous legends, we'll need to specify what color to use for those values encountered above or below our largest and smallest range values, respectively. Click on the *More than maximum* color block and change this to baby blue. We'll keep the *Less than minimum* color as it is for now.

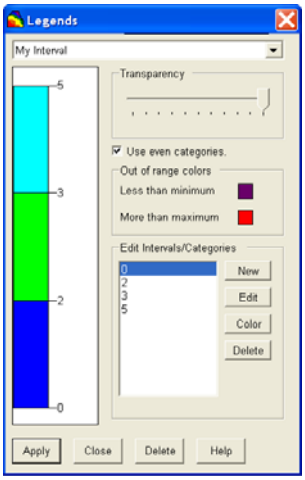
Edit Intervals/Categories

This is where you will specify your break points. They will always be continuous. You cannot, for example, have 1-2 and 3-5. The buttons to the right of the interval break points list is where you will add, edit, and delete them.



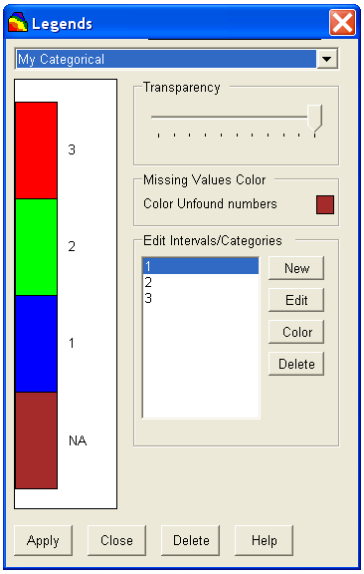
Let's add a new interval from 3-5. To do this press the *New* button, enter a value of 5, and press *OK*. SADA assigns a default color to the new interval. Let's change the color. To change the color, you'll want to select the number at the bottom of the interval. In this case, choose 3. Press the *Color* button and choose baby blue again and press *OK*.

Select 1 and press the *Edit* button. Change this value to zero and press *OK*. Make sure you're parameters look like the following and press the *Apply* button.



New Categorical

Finally, we'll create a new categorical legend. This legend type is very similar to the interval type. Select *(New Categorical)* from the drop-down list, enter "My Categorical" as the new name, and select *Default Categorical* as the basis. Press *OK*.



Transparency

This operates the same as with both the continuous and interval legend types. Set the transparency to about 50%.

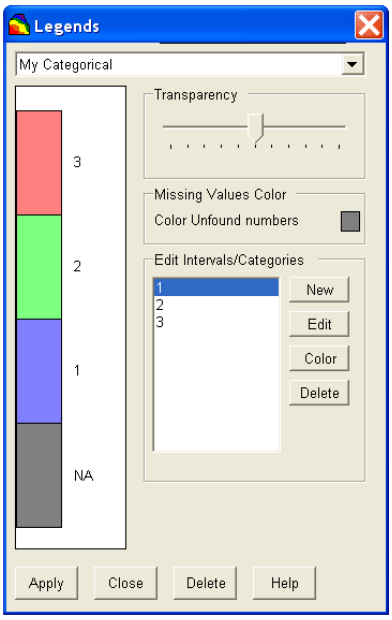
Missing Values Color

This operates the same as the *Out of range colors* option seen earlier. The difference here is that categorical legends don't need to be continuous, and so there is no notion of more-than or less-than the highest and lowest values; there is only "not found." For example, we have default categories of 1, 2, and 3. If a value of 2.3 is encountered, it will be colored brown. To change this, click on the brown color box and change it to gray.

Edit Intervals/Categories

This specifies the exact categories you want to use. Functionally, it is exactly the same as with the interval legend, and we won't repeat the discussion here.

Make sure your parameter window looks like the following image and press the *Apply* button.



Most of the map will turn gray because this result is continuous, the average of simulation values. Press the Close button for a moment and choose the *Landcover dataset* instead. Select *Graphics→Legend Manager*. Choose *My Categorical* and press *Apply*.

This is a more suitable use of categorical data, as the Landcover model contains only three values (1=industrial, 2=grass, 3=forest).

Deleting Legends

This is fairly straightforward. Select the legend you want to delete from the drop down list, and press the *Delete* button. SADA will prompt for confirmation. For now, we'll keep each of the legends we've created.

Remember that nothing is really saved until you save your SADA file. Let's save it now as MyGraphicalTools.sda.

Setting Other Colors

You can control the color of a number of other entities as well. These include Polygons, New Samples, Screening Box, and Remedial Map Colors. We'll briefly enumerate these.

Access to these color controls is found in the menu paths, as follows, with each one presenting the standard Windows color palette options.

- *Graphics→Set Various Colors→Axis Color*

  This sets the color for the axis for non-GIS plots or GIS/XY blended plots.

- *Graphics→Set Various Colors→Label Colors*

  This sets the color for the metadata selected via the Labels box.

- *Graphics→Set Various Colors→New Samples*

  This sets the color for all new sample design locations.

- *Graphics→Set Various Colors→Polygons*

Sets the color of all polygons drawn by the user.

- *Graphics→Set Various Colors→Pooled Data Point Color*

  This sets the color for pooled data points.

- *Graphics→Set Various Colors→Remedial Map Colors→Block Is Contaminated*

  Under the remedial design, this controls the color for those blocks that will require remediation under all uncertainty scenarios.

- *Graphics→Set Various Colors→Remedial Map Colors→Block Might Not Be Contaminated*

  Under the remedial design, this controls the color for those blocks currently identified in the area of concern but may in fact not be contaminated.

- *Graphics→Set Various Colors→Remedial Map Colors→Block Might Not Be Clean*

  Under the remedial design, this controls the color for those blocks currently identified outside the area of concern but may in fact be contaminated.

- *Graphics→Set Various Colors→Remedial Map Colors→Block Is Part of Overburden*

  Under the remedial design, this controls the color for those blocks currently outside the area of concern but part of the overburden that must be removed.

- *Graphics→Set Various Colors→Search Label Color*

  This sets the color for the highlight circle used in the search labels feature.

- *Graphics→Set Various Colors→Screening Box*

  This sets the color for the graphical exceedance box in data screen plots.

- *Graphics→Set Various Colors→Site Boundary Box*

  This sets the color for the site box boundary.

- *Graphics→Set Various Colors→Unestimated Model Values→Show With Hash Mark*

  When SADA is unable to estimate a grid cell value, it still needs to fill the cell. SADA will highlight these unestimated cells with a hash mark if this option is selected (by selecting *Yes*). If *Yes* is selected, this hash mark will be drawn along with any unestimated color you have chosen (see *Graphics→Set Various Colors→Unestimated Model Values→Show Color*).

- *Graphics→Set Various Colors→Unestimated Model Values→Show Color*

  SADA can unilaterally color any unestimated grid cell with a single color of your choice (see *Graphics→Set Various Colors→Unestimated Model Values→Set Color*). This option allows the user to turn the color feature on or off. If *No* is selected, the unestimated grid cell will be transparent.

- *Graphics→Set Various Colors→Unestimated Model Values→Set Color*

  This sets the color for the unestimated grid cells.

- *Graphics→Set Various Colors→Vertical Profile ID Color*

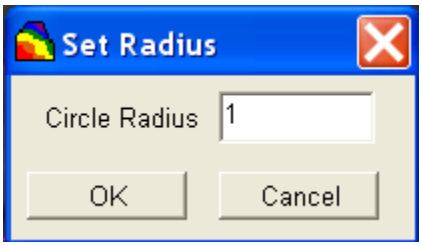
  This sets the box color for the vertical profiles that are spatially identified in the vertical profile feature.

Setting Radius for Plot Circles

Menu item *Set radius for various plots circles*, found under the *Graphics* menu, allows users to determine how large graphical dots or circles appear in the output. There are three options currently (shown by menu path).

- *Graphics→set radius for various plot circles→ Data Plot Circles*
- *Graphics→set radius for various plot circles→ Helper Data Analysis Circles*
- *Graphics→set radius for various plot circles→ Semivariogram Circles*

Each of these three options presents the user with a window where the size can be entered.



The default is 1. The value is a relative value and is therefore unitless. For example, a value of two would produce a circle twice as large as a value of one.

As one might guess, the *Data Plot Circles* sets the size of point sample locations whenever they are displayed. *Helper Data Analysis Circles* sets the size of the points found in the helper data scatter plot. Finally, the *Semi-variogram circles* menu item controls the size of the experimental variogram points when doing correlation modeling.

Draw Vertical Samples in Order

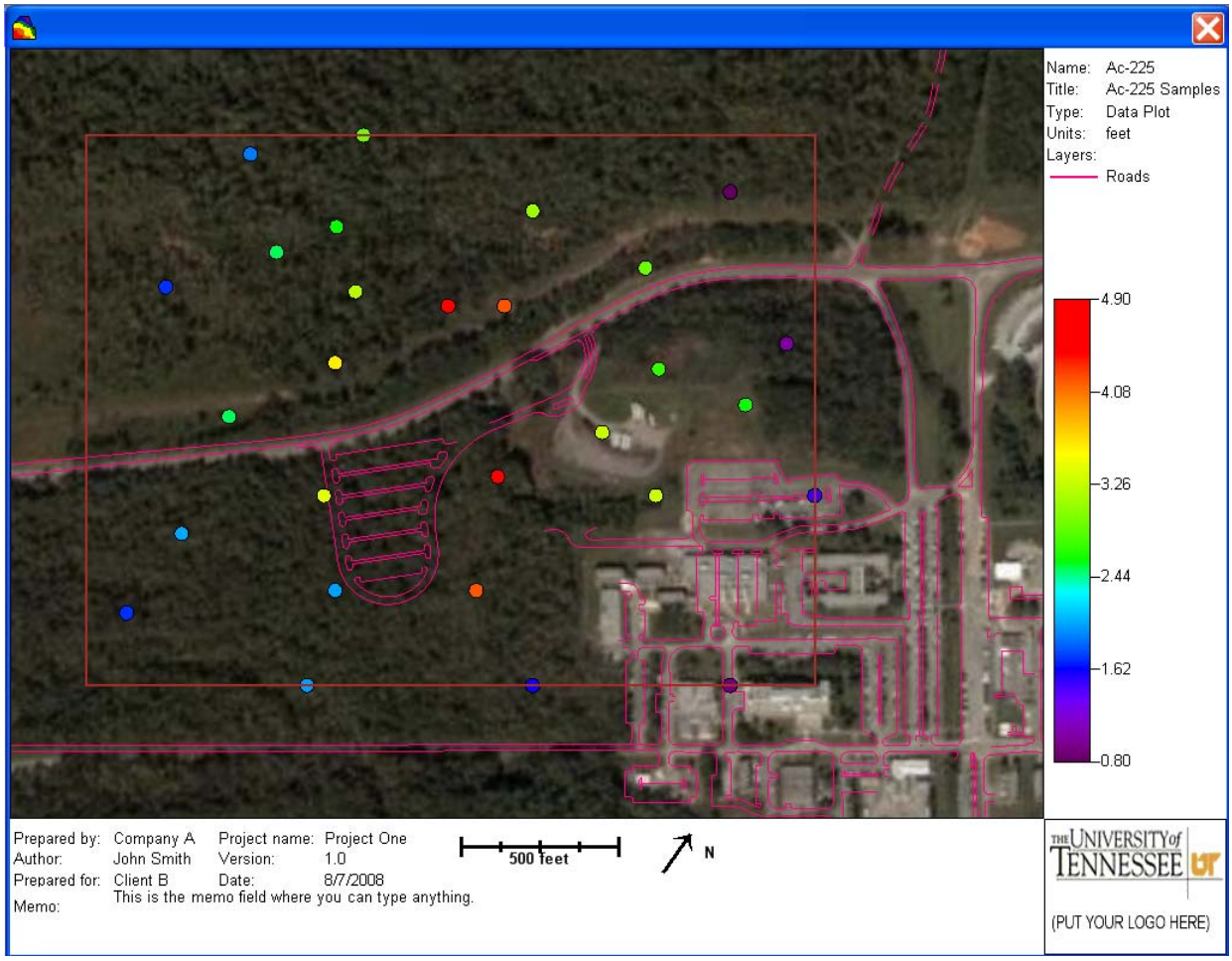
If this menu is checked, then SADA will order the samples queried from your database by depth before drawing them in the 2d viewer. This will guarantee that within any given layer, the top or most shallow data point will appear drawn over a lower point. This feature may slow down rendering for very large data sets. If you think this might be true in your case, try unchecking this menu item. If the menu is deselected, SADA plots the data points within any given layer simply in the order it encounters them in the data set. This may mean that deeper values may appear over the top of more shallow points when both are found in the same layer. Note that SADA does not draw any points in the 2d viewer that are not located in the currently selected layer. This feature is irrelevant in the 3d viewer as data points are drawn in the correct vertical order at all times.

Reset SADA Windows

This menu item restores the size and position of the steps window, the parameters window, and the graphics window to their default location and size.

Setting the Viewer Layout

You can enhance your SADA results by adding various annotations to the lower and right margins of mapping results. These annotations include a variety of items such as spatial scale, authorship, legends, and so forth.



Default View

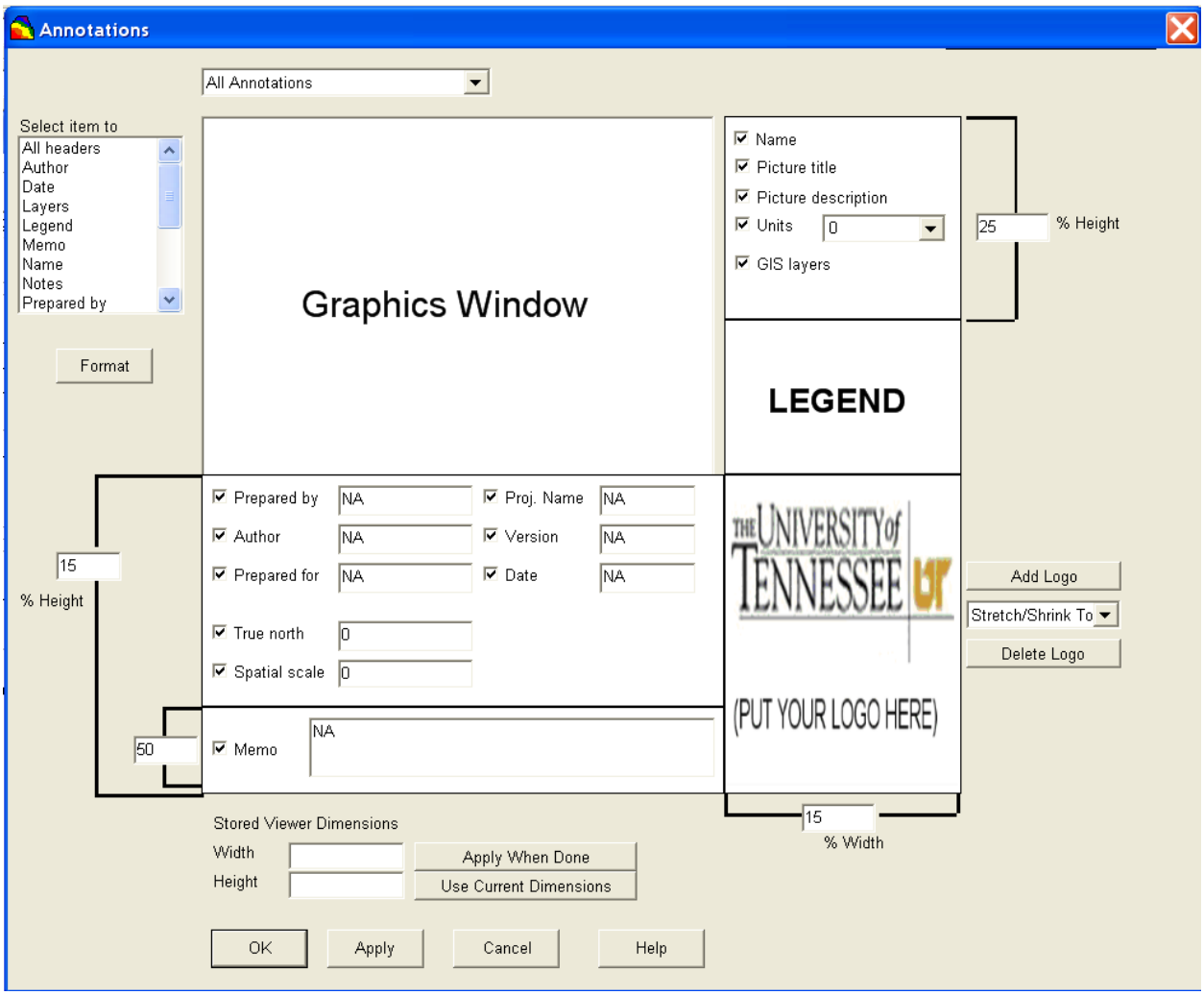
To see how these work, open the file TwoDimensional.sda. Unless specified otherwise, SADA opens with the 2d Default view. This is the most familiar view in SADA with the image to the left and the legend to the right. You may or may not have GIS layers in your viewer. If you do not, don't worry about it and just continue.

This can be accessed by the menu item *Graphics→Viewer Settings→Default View*.

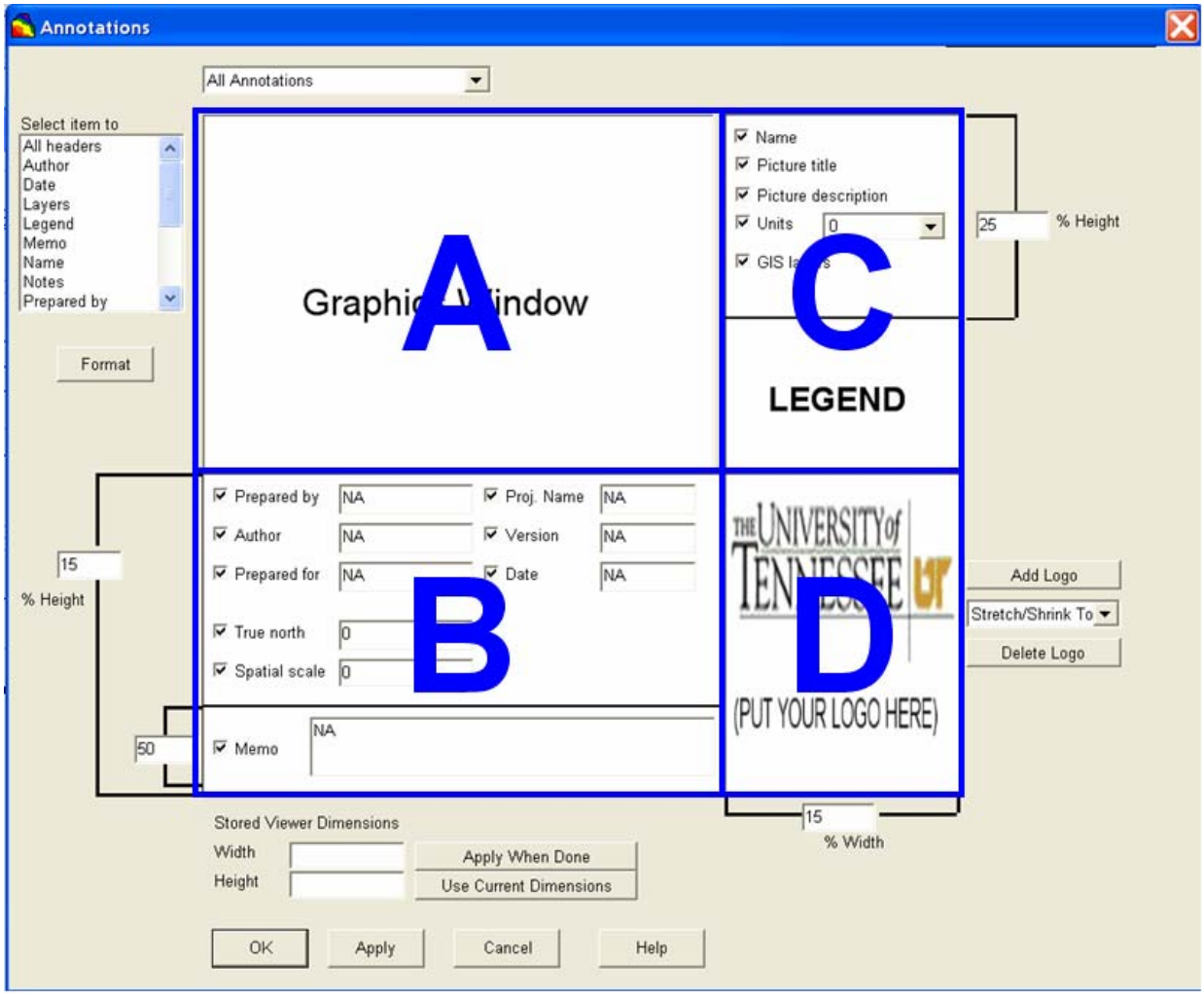
Annotated View

Let's enhance this result with some annotations. First, select from the main menu *Graphics → Viewer Settings → Engineering → Annotations*.

This presents a window where you can specify the types of annotations you want to see and determine how much space of the total result they should occupy.



Drawing attention to the center of the window, four major groups are identified with major letters A, B, C, and D:



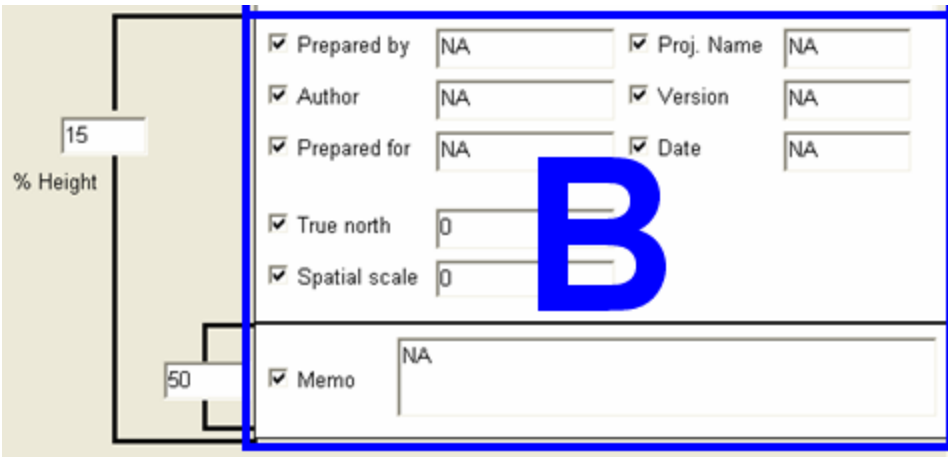
Section A

This simply represents the relative position of the actual graphical result within the annotations.

Section B

This is the lower annotation and all its options. Here, you can specify *Prepared by*, *Prepared for*, *Author*, *Project Name*, *Version*, *Date*, *True north*, *Spatial scale*, and *Memo*. Each of these can be displayed in the lower annotation simply by checking the appropriate box and entering the needed information.

True north and *Spatial scale* are particularly important to understand. SADA does not manage any particular coordinate system or projection. This is left up to the user to decide. Therefore, the user must also specify in what direction the true north arrow should be pointing (e.g., 15 degrees, 45 degrees, etc.). SADA DOES NOT DETERMINE THIS AUTOMATICALLY. The valid range is 0-360 degrees.



For *Spatial scale*, you should enter the distance you wish to represent in the spatial scale legend. For example, if you enter 500, then SADA will automatically draw a legend line equal to 500 units in length. SADA DOES NOT DETERMINE IF IT SHOULD BE METERS OR FEET, ETC. The user must specify this in Section C (see below).

The relative height of Section B is specified by the % Height parameter found to its far left. This is the percentage of the total result height the annotation should occupy. You can adjust this and make the annotation smaller or larger in order to accommodate the text you might choose.

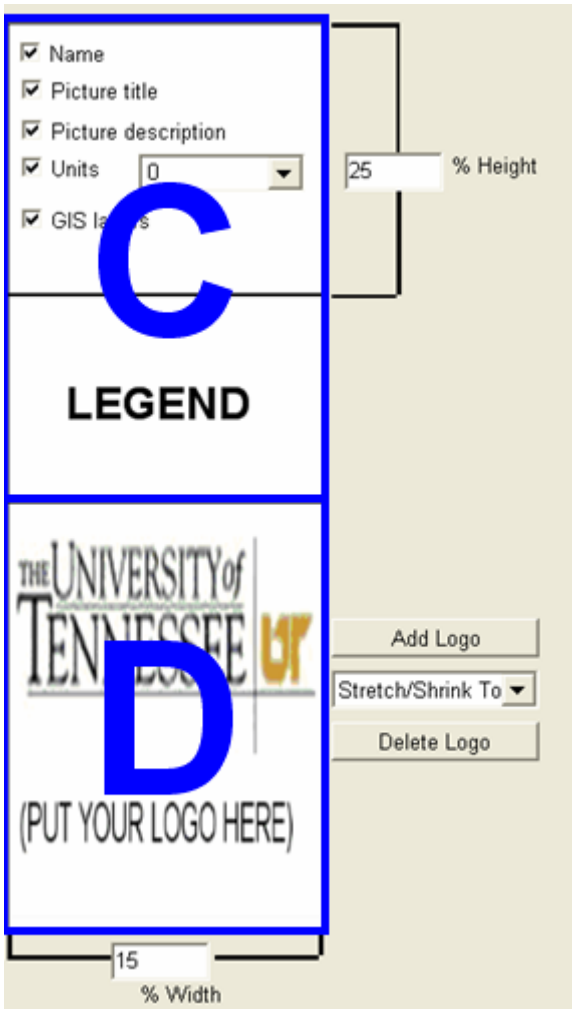
The height of the *Memo* section is specified separately in the parameter box to its immediate left. This is the percentage of the total height B occupies. So, if you specify that B should be 20% of the total height and specify that the memo should be 50% of B, then the actual memo will be 10% of the total image.

These height parameters are intuitive with a little practice.

Section C

This is the upper annotation and the legend. Here, the data set (contaminant) *Name* will be added as well as the *Picture title*, *Picture description*, and *GIS layers* if there are any. Note: If you wish to change the *Name* of the *Picture title*, it is actually done in the *Format* step of the control panel and not here, because this value is dynamically generated and dependent on the image shown. Finally, if used, the *Units* MUST BE SPECIFIED BY THE USER.

The width and height of Section C is specified in a similar manner as in Section B; however, the % height parameter applies only to the upper annotation portion. The legend portion simply occupies the space that remains.



Section D

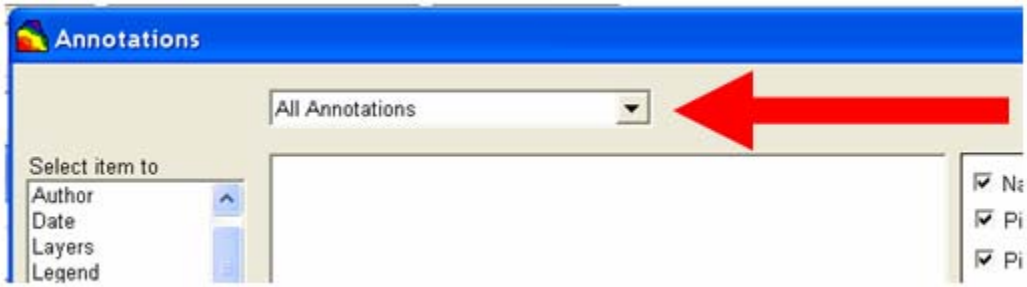
This is the logo box. You can enter any graphic here you wish. To add a picture, press the *Add Logo* button to the right of the logo box.

Choose a graphic file, such as a bitmap (.bmp) or jpeg (.jpg file), and say *OK*. You can then control how the image is shown in the logo box by using the drop-down list found just below the *Add Logo* button. Images can be presented *Normally* (shown in actual bitmap size), *Centered* (centered and shown in normal bitmap size), or sized-to-fit the logo box: *Stretch/Shrink To Fit*. The later is likely the most useful choice, as it facilitates pictures of any size.

To delete a graphic image, press tthe *Delete Logo* button.

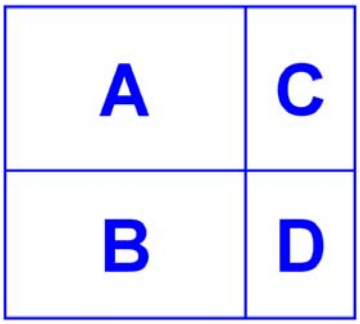
Section D's height will be the same as Section B. The width of the logo box (and Section C) is determined by specifying the *% Width* parameter just below the logo box. This is the percentage of the total image width occupied by the logo box (D).

You can control which combination of annotation blocks are shown by choosing from the options in the drop-down list at the top of the screen.

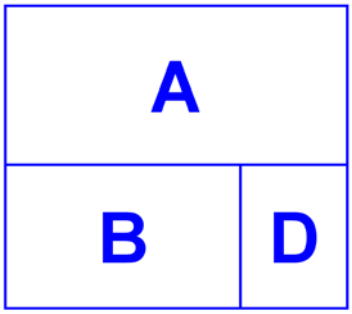


Using our previous block notation, these options will do the following.

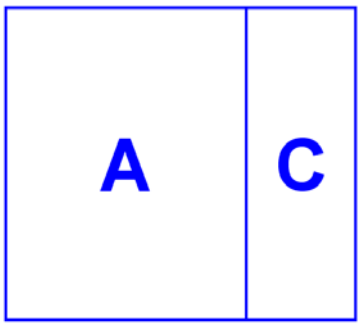
- All Annotations



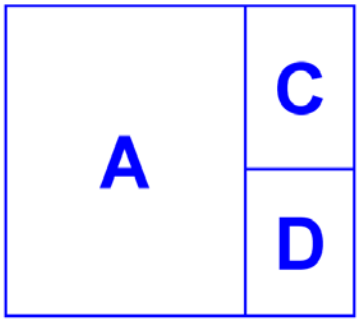
- Only Lower Annotations



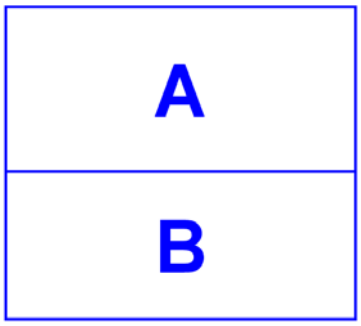
- Only Right Annotations



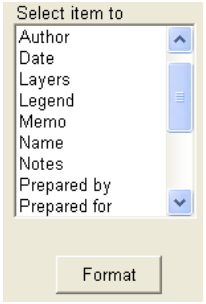
- Only Right + Logo Box



- Only Lower – Logo Box

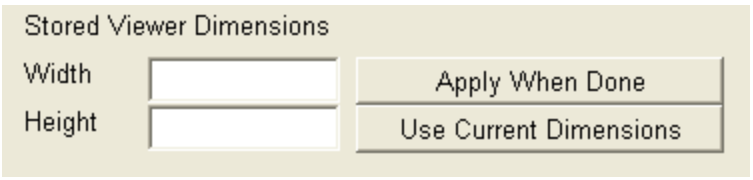


The annotations for each section can be formatted by using the format box found in the upper-left-hand corner.



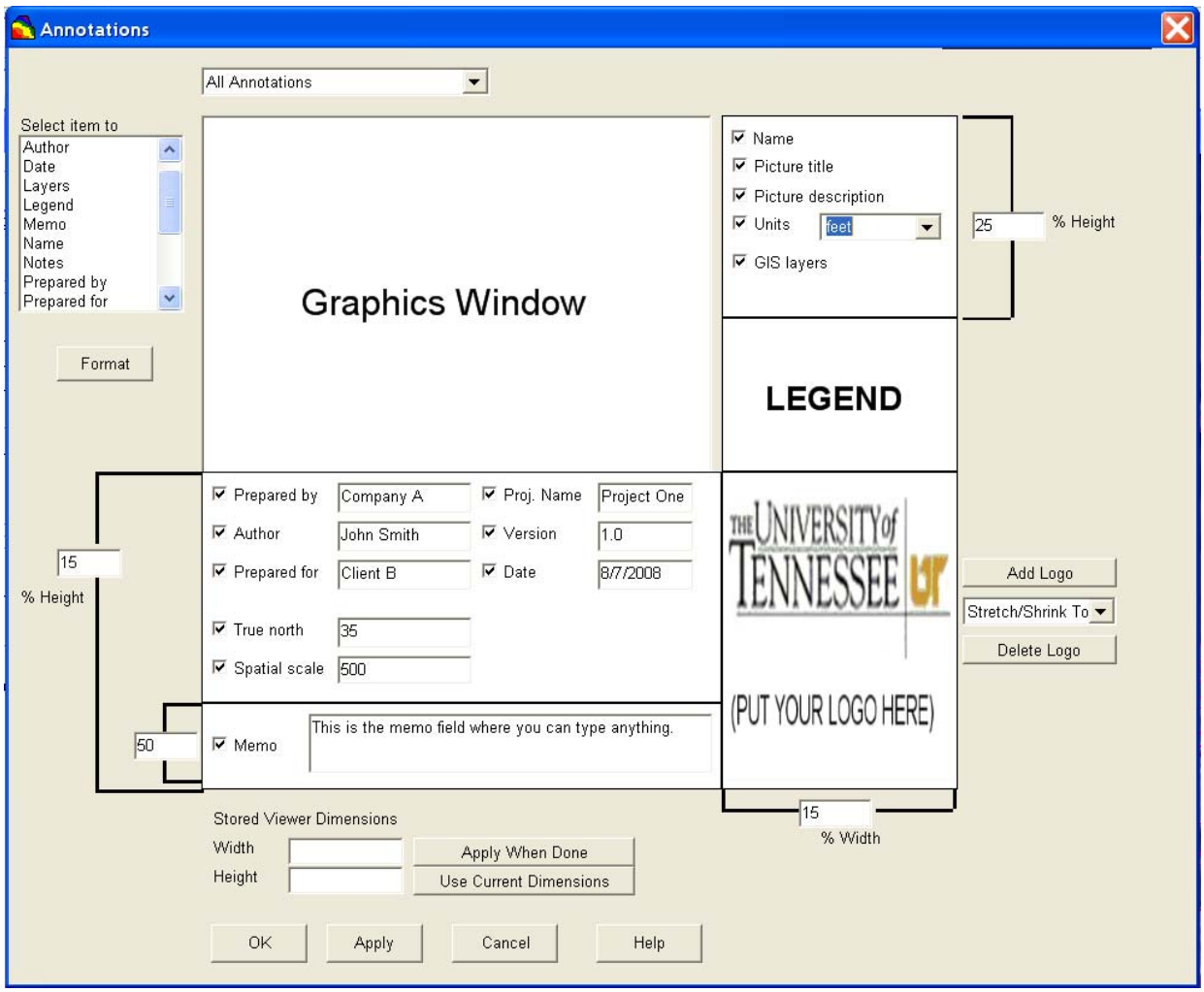
Simply select the item you wish to format and press the *Format* button. You are presented with the usual Windows font window.

In some cases, the exact image you see in your result is best shown with a particular height and width in your graphics viewer. If you have the perfect image for a given result, you can preserve the exact height and width of the graphics viewer and recall it again. Management of this feature is found at the bottom of the annotations window.



To store the current extent of the graphics viewer, press *Use Current Dimensions*. SADA will enter the values into the Width and Height boxes automatically. You may also enter them manually. Later, you can reapply these extents to recreate the exact image you previously had by pressing *Apply When Done*.

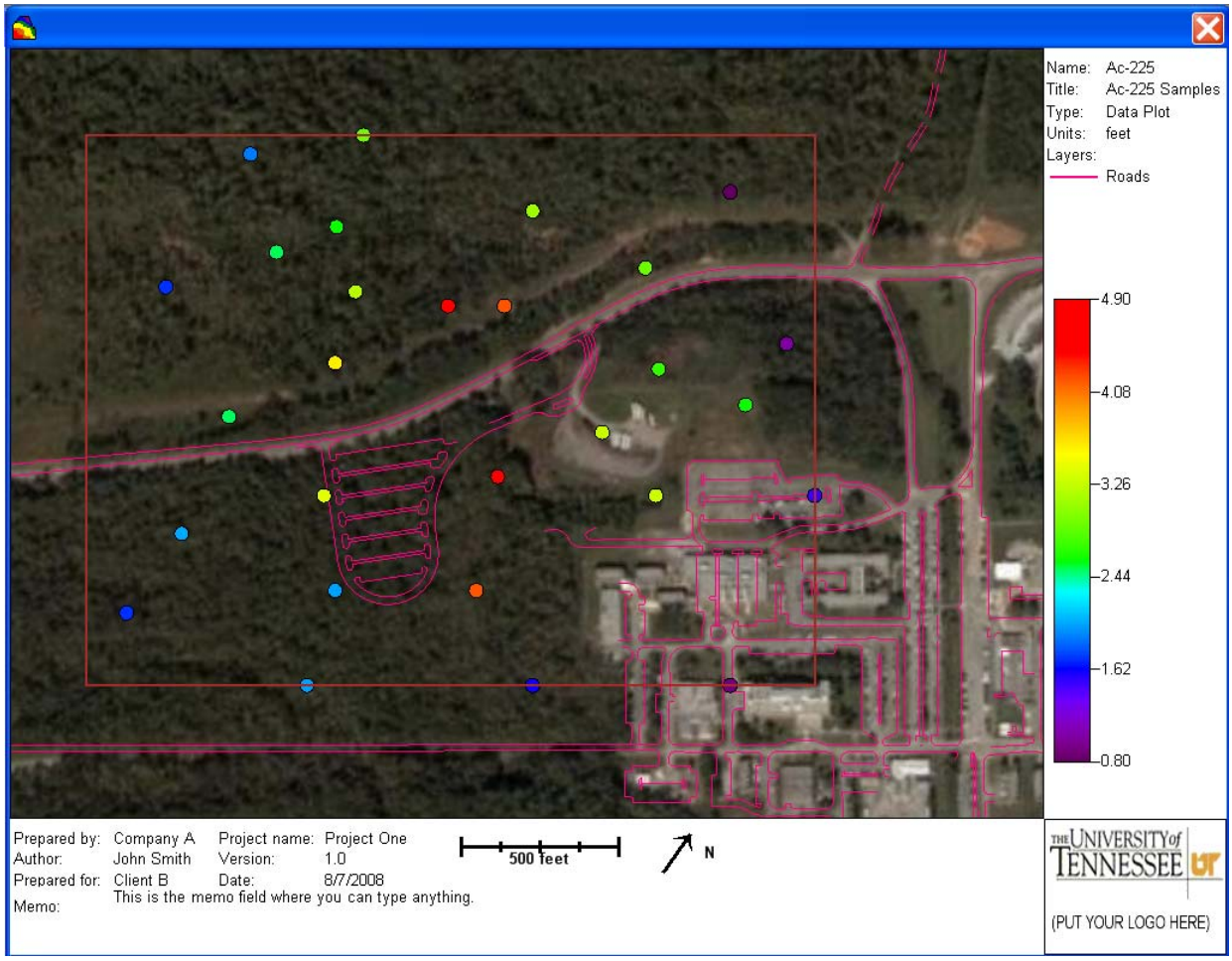
Let's try filling in these sections with the values on the following image.



You can use other values if you like. Your image may be different as well. When you have all the options selected and valid values for each of them, press the *OK* button to return to the main part of SADA.

To see these changes in your result, you will need to select the menu option *Graphics → Viewer Settings → Engineering → Standard*.

SADA will respond with the following image (or something similar depending on your choices).



Now, you can try returning to the *Annotations* window (hint: a shortcut is to click anywhere in the annotations and it will come up automatically) and trying different combinations from the drop-down list (e.g., *All Annotations*, *Only Bottom Annotations*, etc.).

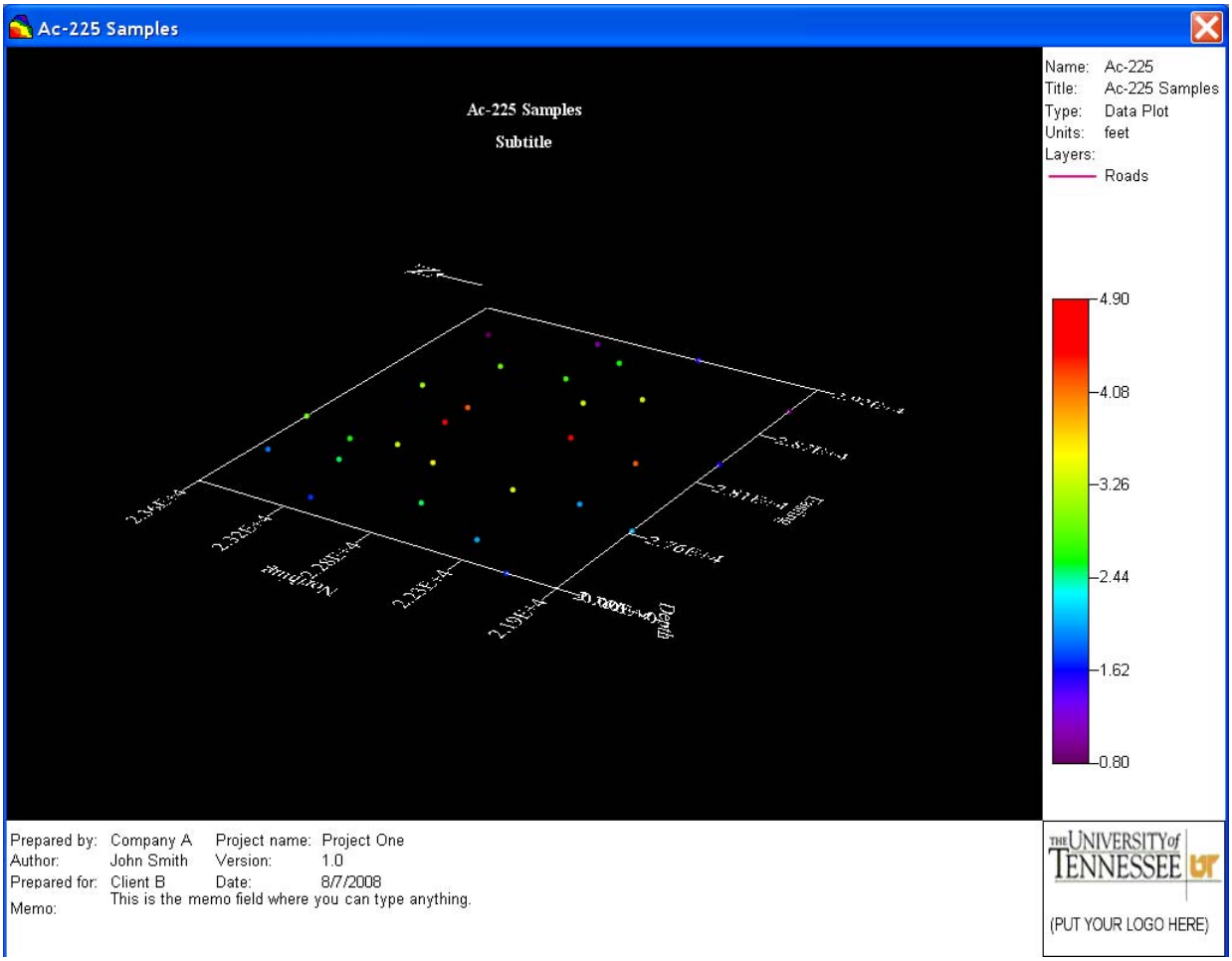
SADA can also put your engineering view into the proper aspect ratio for a landscape or portrait image. This is done by simply choosing either *Graphics → Viewer Settings → Engineering → Portrait* or *Graphics → Viewer Settings → Engineering → Landscape.*

With your viewer in this shape, it should be in good position to be copied to a landscape or portrait oriented paper style in a program such as Word or PowerPoint.

The 3D Viewer

The 3d viewer will reside within the annotations just as the two dimensional views do.

Select Graphics → Show 3d View.



Notice, however, that the *True north* and *Spatial scale* disappear. This is because in a true 3d environment, it is not easy to determine true north or scale with rotations and perspectives in place.

To return to the usual SADA view, select *Graphics → Viewer Settings →Default View*.

References

J., Muehrcke , P., Kimerling, A. Guptill, S. (1995) Elements of Cartography (6th Edition). John Wiley & Sons.

Harvey, F. (2008) A Primer of GIS: Fundamental Geographic and Cartographic Concepts, Guilford Press, NY.

Chapter 44: The 3D Viewer

SADA comes with a three dimensional viewer that allows you to produce some reasonably good 3d images. The objective of this viewer was not to produce publication quality views but rather to support the modeling and analysis activities. Commercial packages are available, should you require higher cartographic fidelity; SADA can export modeling results to many of these packages. If you read no further in this chapter, you should know the following things about 3d rendering and SADA's 3d viewer in particular.

1. Rendering in 3d can be very slow.

   Drawing in 3d is very complicated relative to 2d. In SADA, you should do as much of your modeling or data analysis as you can with the 2d viewer. When you think you are close, then turn the 3d view on. The first time you view a model in 3d, try using a very low grid resolution. If things look good, then you can increase the resolution.

2. Creating great 3d views is an art.

   The 3d viewer comes with many, many options for customizing your 3d view. You will need to take some time and learn how they operate together. This is accomplished primarily by playing around in the 3d viewer and seeing what works for your current application.

3. SADA's 3d viewer and some video graphics cards don't get along.

   In the worst case, it simply won't work. This is actually very rare. More often a 3d render suddenly disappears. Don't worry; it's still there. Take your mouse and touch the 3d viewer screen to rotate the view some. It will wake up. We continually work on this and hope to iron it out at some point.

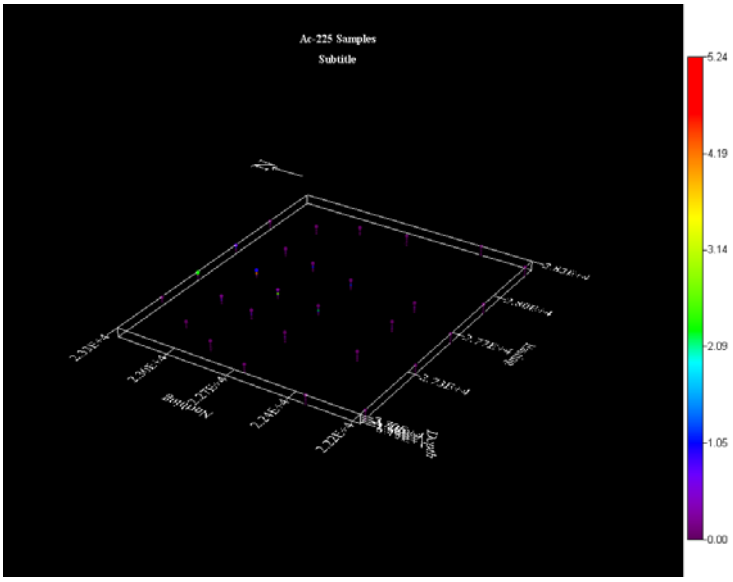
4. 3d Views/Schemas

   The 3d viewer rendering settings must be saved within the SADA file as a schema. You can have several schemas, but if you don't save your view settings in this way, they will be lost. This is taken up later but should be pointed out now to avoid confusion.

This chapter does not address every single 3d viewer feature. Rather, we discuss the most commonly used options and why they are important. If you can master these common functions, you will be well on your way. You can continue with the remaining features by referring to the help file. We are going to present the 3d viewer by demonstration. We'll take a data set and a 3d model and show what kinds of views are possible. Please open the file ThreeDViewer.sda.

The Basics

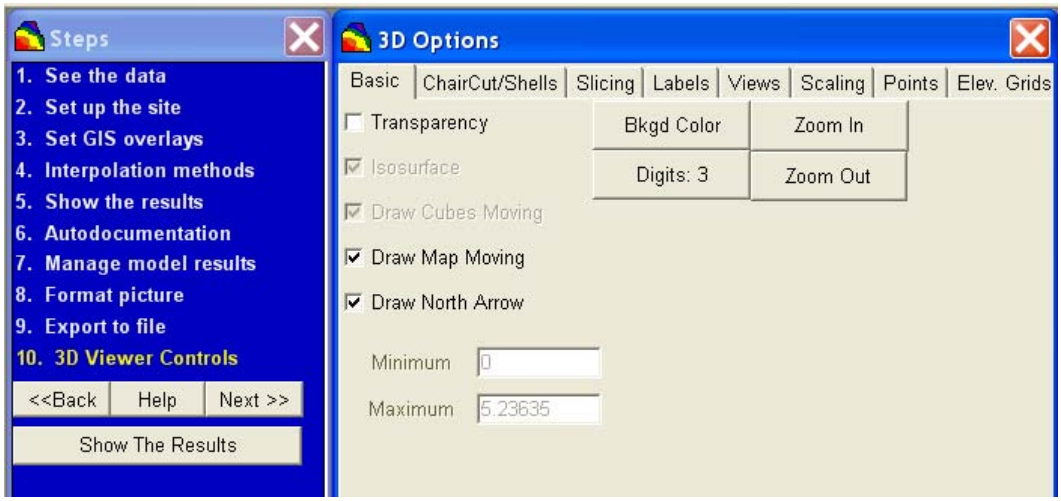
We'll start by showing some of the basics that apply to any render (point, model, or both). As a matter of speed, we'll use point data to demonstrate, but these functions all work for modeled data as well. Make sure you have *Soil* and *Ac-225* selected. To turn on the 3d viewer, select *Graphics→Show 3D View*. The 2d view in the graphics window is replaced by a 3d view. At this point, you will get some indication of how well the 3d viewer and your graphics device are getting along. If you see your data in a squished box like the following, then you're OK (for the moment).



If you see a completely blank screen or a box frame with nothing in it, then left-mouse-click in the center of the 3d view and while holding the button down, move your mouse very slightly. This should make the render appear in most cases. If nothing is happening or you are getting an error, the viewer may not work for your circumstances.

Your site will be contained within a box frame. This frame represents the site boundary box (see Chapter 44 on setting up your SADA file). At this time, polygons do not appear in the 3d viewer, although they are still present. This will be added at some point in the future.

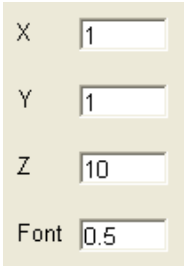
Whenever the 3d viewer is active, *a 3d Viewer Controls* step will be added at the end of your steps list. Click on this step now.



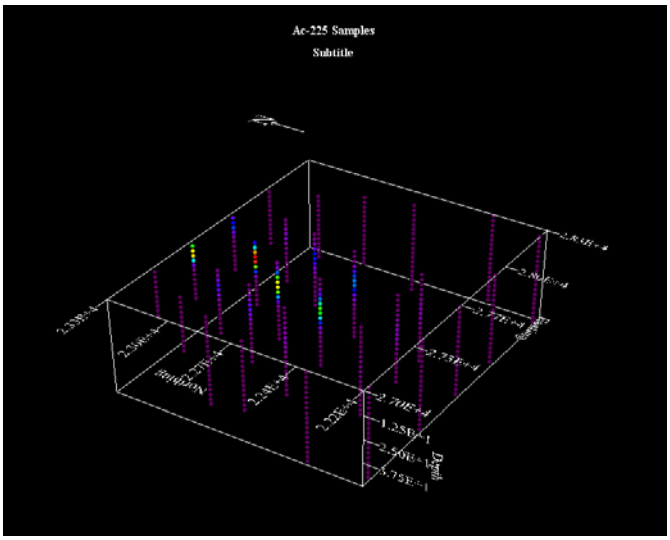
The parameter window for this step is comprised of a series of tabs that organize the many viewer features. Rather than explaining the features of the 3d viewer, we will explain applications of the materials: we will follow a path of parameter selection that you may find yourself often doing. We will be bouncing around these tabs as we discuss the most common features.

Exaggerating the Axes

The first thing you may notice about your site is how flat it is. This is normal and occurs because, in most cases, the horizontal extents are much greater than the vertical extents. Here, we have a site that is thousands of feet across and 50 feet deep. Still, this is not a very helpful view. So, the first thing you will want to do is exaggerate the vertical direction. Click on the *Scaling* tab (you may have to scroll to see it) and enter a value of 10 for the Z parameter.



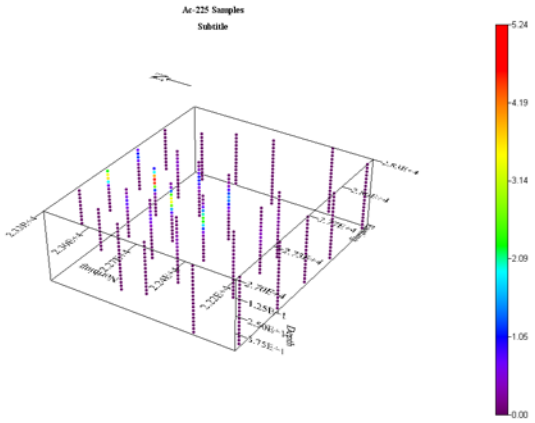
Press Enter on your keyboard and the picture is updated by exaggerating the z axis by a factor of 10.



You can also exaggerate the X and Y directions, although this will rarely be needed in most environmental applications.

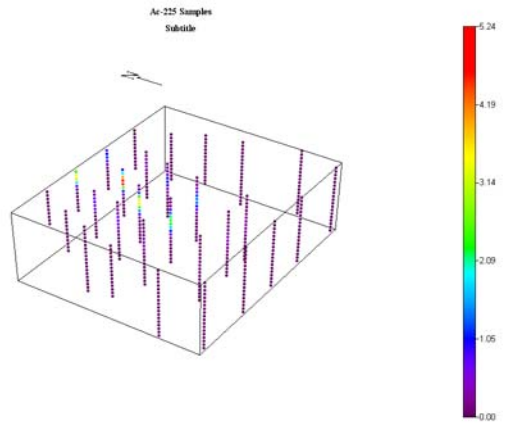
Background

Depending on your preferences, you may wish a different background color. In this situation, a white may be better, as it will then match the legend background. Click on the *Basic* tab and press *Bkgd Color*. Choose *White* and press *OK*.



Dealing with the Axis Labels

One of the less attractive features of the 3d viewer is the tic mark labels. In most cases, they tend to be low resolution. We are working to improve this. In the meantime, if the labels are more of a distraction than a benefit, you can turn them on or off by right-mouse-clicking the axis itself. Try now right-mouse-clicking on each of the axes with labels to turn those labels off. You can change the relative font size by entering a new value for the *Font* parameter under the *Labels* tab.



You can also turn the labels off by choosing the *Labels* tab and unchecking the checkbox next to each axis of interest.

Changing Labels

There is a loose connection between the labeling feature of the 3d viewer and the formatting options step. 3d viewer labels are actually stored in the *Labels* tab. The default values are informed by the formatting step; however, changes in this label tab are actually stored as part of the schema. To insure you maintain the label you wish, you'll need to store it in the schema. Click on this *Labels* tab now. Let's delete the subtitle. Delete the word *Subtitle* from the box and press the Enter key on your keyboard.

Setting the Redraw Option

Whenever you rotate or zoom a 3d view (which we will do shortly), SADA will attempt to render the image constantly as it moves. This is a nice, because you can see the volume as you spin or zoom on it. Unfortunately, for very dense renders, this can be impractical, as the movements become dramatically slower and have a "jerky" quality to them. If this is happening to you, you'll

want to deselect the *Draw Maps While Moving* option under the *Basic* tab. For now, we'll leave it because with only 400 data points, the image will move nicely in our next steps.

Rotating

Rotating is relatively easy. Left-mouse-click anywhere in the 3d viewer and move your mouse around. The best advice is to just play with starting your mouse in different locations. The response is fairly intuitive. If the rotation is not smooth, deselect the *Draw Map While Moving* option under the *Basic* tab.

Zooming

There are two ways to zoom in or zoom out. Under the *Basic* tab, simply click on the *Zoom In* or *Zoom Out* button. An easier way is to select the *3d view* in the graphics window (click on it), press the Shift key, and while holding also click on the arrow up or arrow down buttons on your keyboard. If the zoom effect is too great, you can control the "jump" distance by clicking on the *Translation* tab and setting a new delta for the z.

Shifting

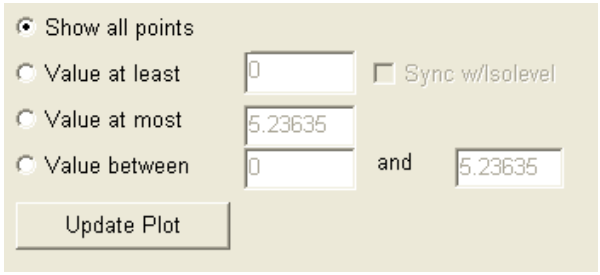
To shift the picture left, right, up, or down, hold the control key and press the left, right, up, or down arrows, respectively. Again, the amount of the jump can be controlled by the delta parameters for x and y, found in the *Translation* tab.

Saving the View/Schema

It is very easy to quickly lose your 3d settings. Once you have them where you like, be sure to save the view as a schema in the 3d viewer parameter set. Click on the *View* tab. Press *Save Current View* and provide a name and even a comment if you like. Press *OK*. Later, select the newly-saved view and press *Load Selected View*. You can also save the view to a file and recall it later in the same or different SADA file.

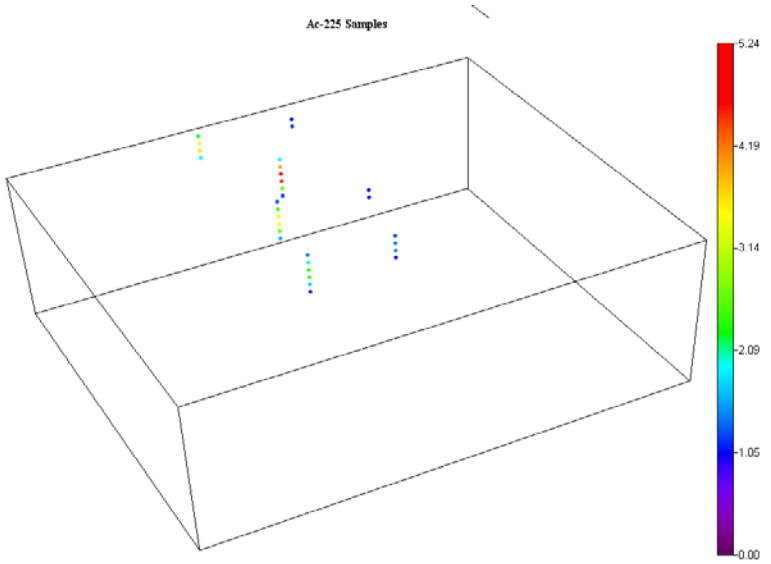
Rendering Points

When you are showing points separately or as part of a model result, there are several features you can use to control the rendering. One of the most useful features is to restrain the rendering to certain ranges of values. Click on the *Points* tab, and we'll discuss the parameters near the bottom of the window.



The default is *Show all points*. You can also choose to show only those points greater than some minimum value, less than some value, or between certain values. Click on the *Value at least* option and enter a value of 3. Recall from earlier chapters that the decision criterion for Ac-225 ingestion was 3pCi/g. We can quickly emulate the data screen feature right here by only drawing those points greater than 3. There is also a Sync w/Isolevel associated with this feature. The isolevel is used in model rendering. If you specify some value, an isolevel is the surface in the 3d volume where the model is exactly 3. We'll see that shortly. Enter 3 and press

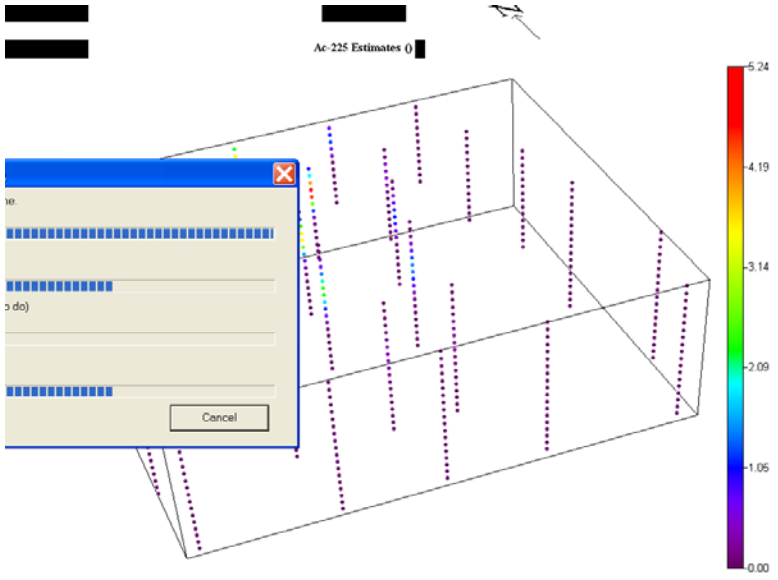
Update Plot. Notice that only a few samples now appear. Try a value of 1 now and press *Update Plot*.



Now select *Value at most* and enter a value of 1. Press *Update Plot*; this shows all those values below 1pCi/g. Try now the *Value between* option on your own. Now return to *Show all points*.

Drawing Models

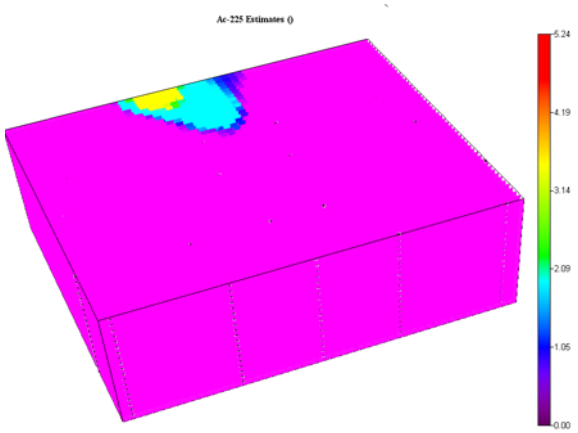
When you draw a model, you have two options. You can either draw the blocks or the isosurfaces. The blocks option will draw each cube and color it according to the legend. The isosurface (isolevel) option will determine the inner surface for all those points in the model exactly equal to a given value. In the interview drop-down list, select *Interpolate my data*. We've previously created a model for Ac-225 to save time and have elected to use it for the interpolation method here. Press *Show The Results*. We'll take this opportunity to show a common problem with interfacing and certain video cards. Sometimes, SADA appears to just crash when you show a 3d result. For example consider the following image:



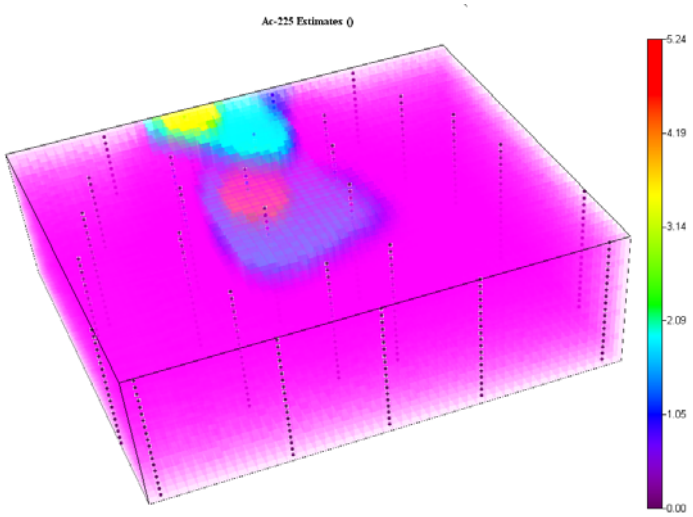
Actually, SADA has not failed. Just click your mouse anywhere and move it slightly to refresh the view.

Block View

Click on the *3d Viewer Controls* step. Click on the *Basic* tab and deselect the *isosurface* option. This switches SADA to block view.

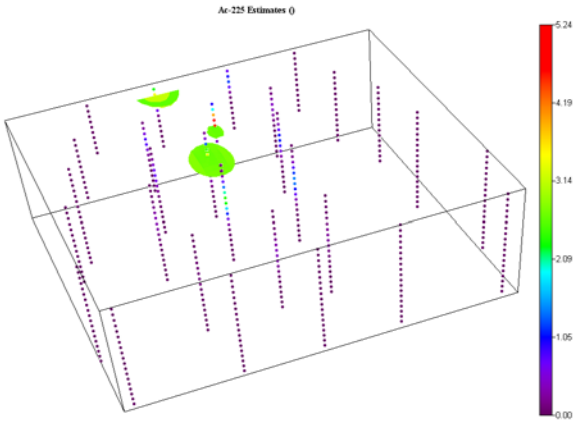


The problem with the blocks view is that it's difficult to explore the volume. To ease this, you can turn the transparency on. The transparency makes lower values more transparent and higher values more opaque. In most situations, this will allow you to see through the volume better. This feature is found on the *Basic* tab directly above the *isosurface* option. Select this now.

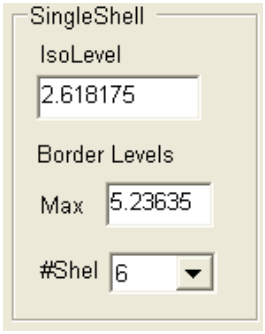


Isosurfaces View

The more common view of 3d space is with the use of isosurfaces or isolevel. An isosurface is that surface where the volume values are exactly equal to a given value. Let's try this now. In the *Basic* tab, deselect *transparency* and select *Isosurface*.



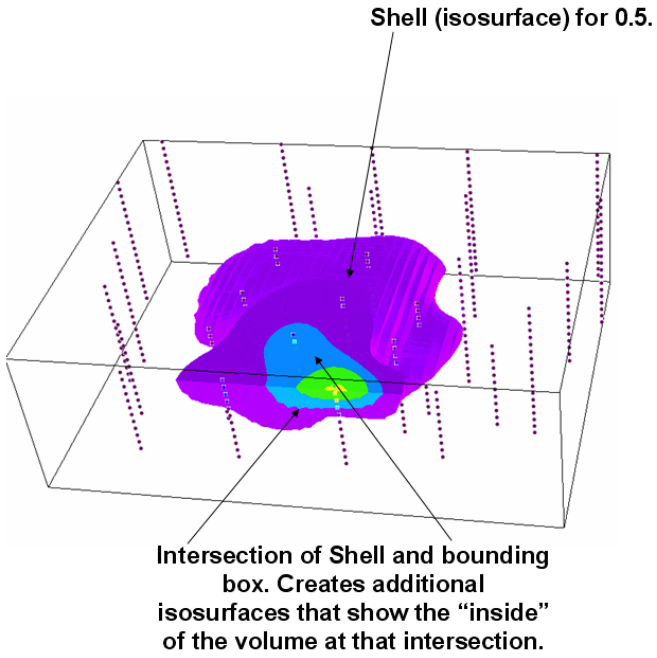
At the moment, SADA has chosen an arbitrary isolevel for us. Click on the *ChairCuts/Shells* tab. At the top of this tab, you'll see a parameter block called *SingleShell*.



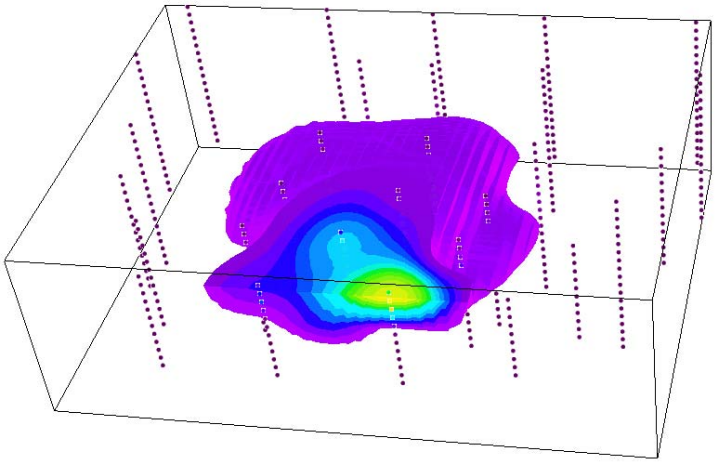
The *IsoLevel* is found here (currently 2.618175). You will also see a feature called *Border Levels*. This handles the situation where the isosurface intersects the side walls, top or bottom. When this happens, the wall is expressed as a range of isosurfaces. You can specify the

number of shells (another word for isosurface). Change the *IsoLevel* to 0.5 and press *Update Plot*.

In the following image, we've rotated the view so you are looking due south. You can see from here where the isosurface (0.5) has intersected or "splashed" up against the walls. These intersection points are filled with those isosurfaces on the interior of the volume. At present, we are using up to 6 shells to delineate this inside intersection.



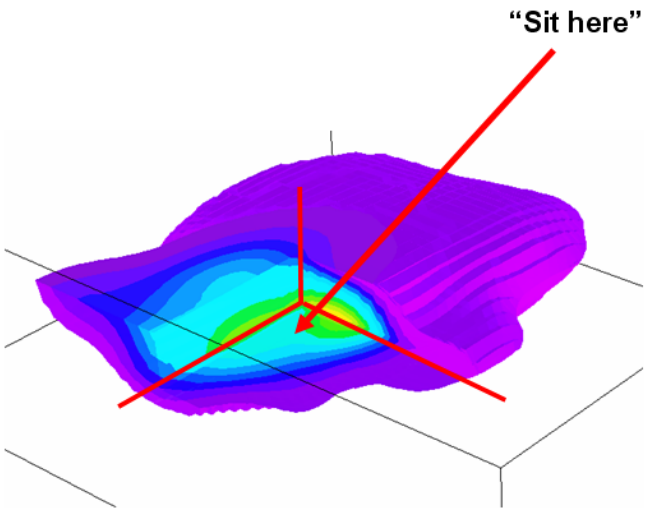
In the *Single Shell* block, change the number of shells from 6 to 20 and press *Update the Plot*. Now you have a finer delineation.



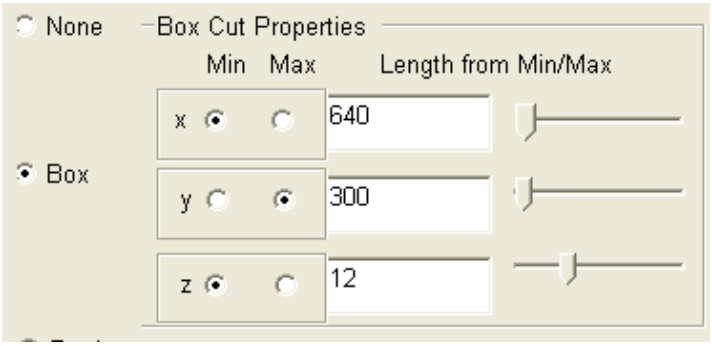
Chair Cuts

The last feature we'll discuss is chair cuts. Chair cuts and isosurfaces work hand in hand. Recall how when an isosurface intersects a wall, we show the interior of the volume at that intersection. With chair cuts, we can create a cubical cut that we can move around, allowing us

to see the inside of the volume at any point of interest. The following picture shows why this cut away is referred to as a chair (looks like you could sit down on it).



Click on the *ChairCut/Isosurface* tab. On that tab you'll find the options for using a chair cut.



Click on the *Box* option and the parameters become enabled. A chair cut is created by sliding a "glass chair" into the volume from one of the axis corners. In this case, we will slide the chair in from the corner (xmin, ymax, zmin). That's the surface in the northwest corner. You chose the corner by selecting either *Min* or *Max*.

Next, you'll choose how far to slide the glass chair in from that corner. This is the distance parameter that is specified for each direction. Enter 640 for X, 300 for Y, and 12 for Z. Press *Update Plot*. You should see the same image as we presented above. Now, play around with different distances and different corners on your own.

Exporting the Image

Recall that there are several ways to export the image.

- To copy this image into Word or PowerPoint, use the Copy button on the main toolbar ().
- To save this to a bitmap file, use the export to BMP button on the main toolbar ().

As always, to save the numerical results for use in another package use the *Export to file* step in the steps window.

Summary

Spend some time working with these basic tools and then venture out into some of the lesser used features. They are all documented in the help file.

Chapter 45 Quick References

Chapter Summaries and Prerequisites

This section summarizes each chapter, assigns a level of difficulty, and identifies prerequisite chapters required for each chapter. Levels of difficulty are as follows: 1 = little or no prior experience, 2 = comfortable with SADA basics and interested in building on that, 3 = advanced user.

Table 45.1 Chapter Summaries (1 of 3 tables)

Name	#	Level	Prerequisite Chapters
Part I Overview			
Introduction	1	1	None
A Quick Tour	2	1	None
Understanding SADA Files	3	1	A Quick Tour
Setting Up A SADA File	4	1	Understanding SADA Files
Part II: Importing Data			
Importing Sampled Data	5	1	Setting Up A SADA File
Import New Data Over Existing Data	6	1	Importing Sampled Data
Importing Modeled or Gridded Data	7	1	Understanding SADA Files
Importing Elevation Data	8	1	Understanding SADA Files
Managing your Data Sets	9	1	One of Previous Import Chapters
Automatically Importing Information into SADA from Your Database Management System	10	3	Importing Sampled Data
Part III: Exploring Data			
Visualizing and Exploring Your Data	11	1	A Quick Tour An Import Chapter Graphical Tools
Statistics	12	1	A Quick Tour An Import Chapter Visualizing and Exploring Your Data
How to Perform a Quick Spatial Data Screen	13	1	Importing Sampled Data Visualizing and Exploring Your Data Graphical Tools

Table 45.2 Chapter Summaries (2 of 3 tables)

Chapter		Level	Prerequisite Chapters
Part IV: Risk			
Importing Custom Screening and Remediation Criteria	14	1	Importing Sampled Data Visualizing and Exploring Your Data
Screening Data Against Custom Criteria	15	1	Importing Sampled Data Visualizing and Exploring Your Data Importing Custom Screening and Remediation Criteria
Creating Ratio/Sum of Fraction Maps Using Custom Criteria	16	1	Screening Data Against Custom Criteria (Custom Analysis) Visualizing and Exploring Your Data Graphical Tools
Overview of Human Health Risk Assessment	17	2	A Quick Tour Importing Sampled Data
Preparing for Human Health Risk	18	2	Overview of Human Health Risk Assessment
Calibrating The Risk Model	19	2	Preparing for Human Health Risk
Tabular PRG, Screen, and Risk Calculations	20	2	Calibrating The Risk Model
Spatial Risk	21	2	Visualizing and Exploring Your Data Graphical Tools Tabular PRG, Screen, and Risk Calculations
Overview of Ecological Risk Assessment	22	2	A Quick Tour Importing Sampled Data
Preparing for Ecological Risk	23	2	Preparing for Ecological Risk
Calibrating the Ecological Risk Model	24	2	Calibrating the Ecological Risk Model
Conducting Tabular Ecological Risk Assessments	25	2	Conducting Tabular Ecological Risk Assessments in SADA
Ecological Risk: Terrestrial Dose Modeling	26	2	Ecological Risk: Terrestrial Dose Modeling
Spatial Ecological Risk	27	2	Visualizing and Exploring Your Data Graphical Tools Conducting Tabular Ecological Risk Assessments in SADA
Part V: Spatial Modeling			
Geospatial Methods	28	2	A Quick Tour
Basic Geospatial Methods	39	2	Geospatial Methods
Advanced Geospatial Methods Part I: Overview and Correlation Modeling	30	3	Basic Geospatial Methods
Advanced Geospatial Methods Part II: Geospatial Modeling, Uncertainty Analysis, and Simulation	31	3	Advanced Geospatial Methods Part I: Overview and Correlation Modeling
Working With Models	32	2	Importing Modeled or Gridded Data and/or Chapters on Geospatial Methods
User Models	33	2	Setting Up A SADA File
Spatial Risk	34	2	Choose from your focus area: custom, human health, or ecological risk chapters

Table 45.3 Chapter Summaries (2 of 3 tables)

Chapter		Level	Prerequisite Chapters
Part VI: Decision Analysis			
Decision Analysis	35	2	If doing risk then choose from your focus area: custom, human healh, or ecological risk chapters If building your own spatial models, then chapters on geospatial methods. If importing models then the chapter on Importing Modeled or Gridded Data
Cost Benefit Analysis	36	2	Decision Analysis If doing risk then choose from your focus area: custom, human healh, or ecological risk chapters If building your own spatial models, then chapters on geospatial methods. If importing models then the chapter on Importing Modeled or Gridded Data
Part VII: Sample Design			
Overview of Sample Designs	37	2	A Quick Tour
Secondary Sampling Designs	38	3	Overview of Sampling Designs Decision Analysis If doing risk then choose from your focus area: custom, human healh, or ecological risk chapters If building your own spatial models, then chapters on geospatial methods. If importing models then the chapter on Importing Modeled or Gridded Data
Standard Initial Sampling Designs	39	2	Setting Up A SADA File Overview of Sampling Designs
Multi-Agency Radiation Survey and Site Investigation Manual (Scenario A)	40	3	Setting Up A SADA File Overview of Sampling Designs
Informed and Targeted Initial Designs	41	3	Overview of Sampling Designs Decision Analysis If doing risk then choose from your focus area: custom, human healh, or ecological risk chapters If building your own spatial models, then chapters on geospatial methods. If importing models then the chapter on Importing Modeled or Gridded Data
Part VIII: Miscellaneous Topics			
Local Index of Spatial Association Tools (LISA)	42	2	A Quick Tour Setting Up A SADA File Importing Sampled Data
Graphical Tools	43	1	A Quick Tour Visualizing and Exploring Your Data Graphical Tools
The 3D Viewer	44	2	A Quick Tour Visualizing and Exploring Your Data Graphical Tools
Quick Reference	45	1	NA

Summary of File Types

This section presents the types of data that you can import to and export from SADA. A brief description and the chapter where more information can be found is provided. Each result in SADA can be exported and the format varies slightly. Not all of these are enumerated in the export table. The emphasis is on major export formats.

Table 45.4 Importing Data

DATA TYPE	FILE FORMAT	CHAPTER	COMMENT
Data points	Comma Separated Value (.csv) Microsoft Access Database (.mdb)	5 & 6	These are (x,y,z) measured point values.
Gridded Data	SADA Grid Format (.csv) Float Grid (*.hdr/*.flt) ASCII Grid (*.txt)	7	Use these types to import spatial models built in other software packages.
Elevation Data	SADA Grid Format (.csv) Float Grid (*.hdr/*.flt) ASCII Grid (*.txt) Digital Elevation Model (.dem)	8	This is gridded elevation data used to adjust the height of the measurements in 3d viewer.
GIS Layers	Data eXchange Format (.dxf) Shape File (.shp) Image (.jpg, .bmp, .tiff, gif)	4	Use these as base maps for your analysis.
Custom Criteria	Comma Separated Value (.csv) Microsoft Access Database (.mdb)	14	These are user specific decision criteria imported to compare against data point values.
Toxicological Data	Microsoft Access Database (.mdb)	18	These data contain toxicity and radiological data for a large set of analytes. This is needed for human health risk.
Scenario Data	Microsoft Access Database (.mdb)	18	These data contain behavioral parameters to calibrate the

			exposure models in human health.
Benchmark Data	Microsoft Access Database (.mdb)	23	These contain benchmark values for ecological risk assessment.
Area Factors	Microsoft Excel (.xls)	40	Contains area factors from Resrad for the MARSSIM module.

Table 45.5 Exporting out of SADA

DATA TYPE	FILE FORMAT	CHAPTER	COMMENT
Data Points	Microsoft Excel (.xls) Comma Separated Values (.csv)	5, Various	Exporting data points is available whenever a data point map is available and you press the export to file button (). You can also export from most tabular views of the data using the button with the Excel symbol.
Gridded Data	SADA Grid Format (.csv) Float Grid (*.hdr/*.flt) ASCII Grid (*.txt)	7, various	Use these types to export spatial models to other software packages. To do this, press the export to file button () when a spatial model is in the graphics window. Or use the Export Menu
Line Graphs	Comma Separated Values (.csv)	Various	At various points, xy line graphs are presented (e.g. cost benefit analysis). You can export these to be used in another graphing package (e.g. Excel) where you can more closely control the formatting. To do this, press the export to file button () when a spatial model is in the

			graphics window.
Images	Bitmap, jpg	Various	Whenever you wish to export an image to a file, simply press the export to image button () when the image you are interested in is visible.
Copy to Clipboard	NA	Various	When you want to copy an image result to the clipboard to be pasted in another program, press the copy image button () when the image you are interested in is visible.
Parameter Output	Hypertext Markup Language (.html)	Various	This topic is covered in the help file. There is an Autodocumentation Step with every interview. Click on this to access reporting features.

Toolbar Button Summary

The following table describes the major toolbar buttons found in the interface.

Table 45.6

ICON	DESCRIPTION
	Opens a SADA file.
	Saves a SADA file.
	Prints the information in the graphics window.
	Copy to File – Copies Ascii results to a file.
	Copy Image to File – Copies Ascii results to a file
	Copy to Clipboard – Copies current image to the clipboard. It can then be pasted into most Window Packages.
	Displays relevant information on the current graphic.
	Displays statistical information on the current graphic.
	Screen Table– Displays results of screening data against risk-based screening criteria. Only enabled for human health, ecological, or custom analyses
	Risk Table – Displays results of calculating risk. Only available for human health or ecological analyses.
	Measures Button – Measures the distance between two points on the graphics window.
	Zoom In – Zooms in an area on the graphics window.
	Zoom Out – Zooms out an area on the graphics window.

	Shift Picture – shifts the picture on the graphics window.
	Restore Picture – Restores the graphics window to the default view.
	Line Pointer - The line pointer button is used to define the coordinates of a line on the graphics window.
	Line Query – The line query button is used to type in the value of one coordinate for a line in the graphics window and find the value for the other coordinate.
	Vertical Profile Tool – The vertical profile button is used to access single vertical profile features in SADA.
	Copies selected polygon (within SADA only)
	Pastes copied polygon (only polygons copied within SADA)
	Export to excel

Index