SADA Technical Support: Using Secondary Sampling Strategies

April 6th, 2000

Robert Stewart
Tom Purucker

1.0 Introduction

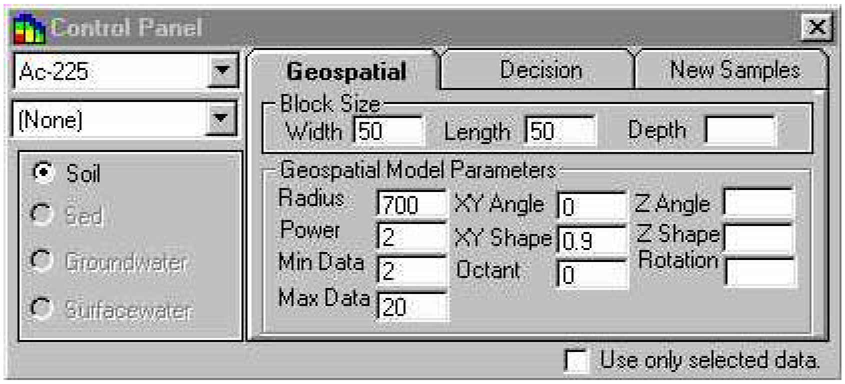
The ability to map concentration and uncertainty across a site provides an excellent framework for determining the number and location of a second round of sampling. SADA provides some basic secondary sampling frameworks that optimize sampling locations relative to a particular sampling goal. These strategies are called in SADA Adaptive Fill, Estimate Rank, Variance Rank, Percentile Rank, and Uncertainty Rank. This paper presents the methodologies behind these frameworks, shows users how to implement these strategies in SADA, and extends the topic to a cost benefit approach to determining the number of samples to choose.

2.0 Methodologies

We present now five sampling methodologies available in SADA each having a specific end goal in sampling. Before we present the particulars of each method we begin with an explanation of how to set up a grid in SADA and how to use the minimum separation distance criteria.

2.1 Creating Grids in SADA

Each of the following five methods depends on a grid definition. Defining a grid is done on the Control Panel's Geospatial tab under Block Size. The width is the size in the horizontal direction and the length is the size in the vertical direction. Depth corresponds to the size of the block in the vertical plane.



Control Panel

Ac-225

(None)

● Soil
○ Sed
○ Groundwater
○ Surfacewater

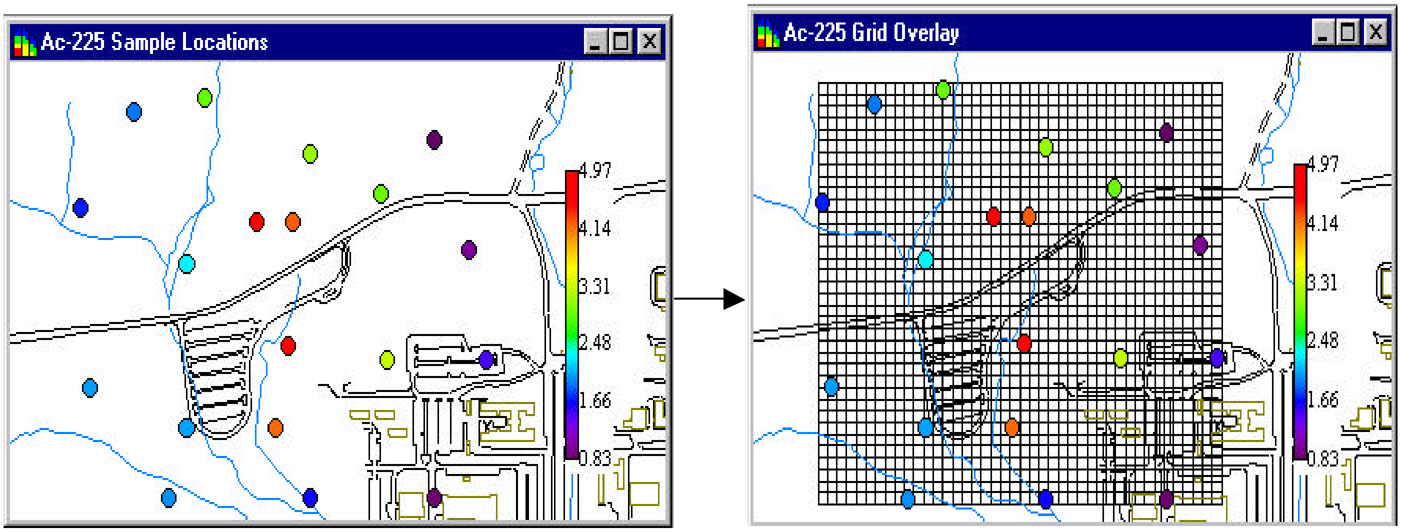
Geospatial | Decision | New Samples

Block Size
Width 50   Length 50   Depth

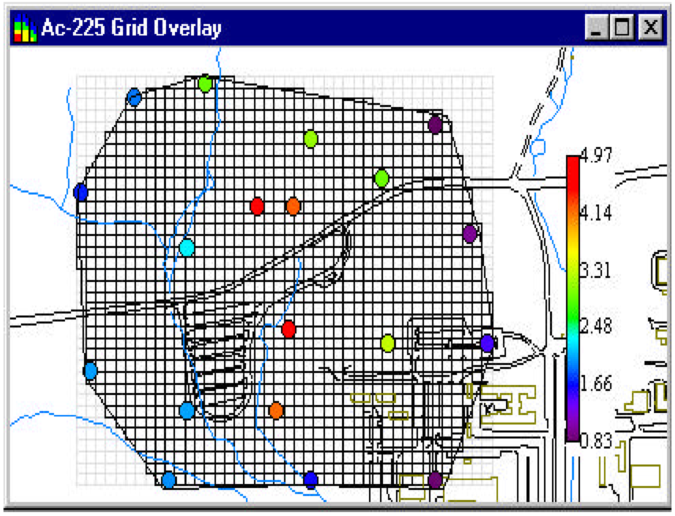
Geospatial Model Parameters
Radius 700   XY Angle 0   Z Angle
Power 2   XY Shape 0.9   Z Shape
Min Data 2   Octant 0   Rotation
Max Data 20

□ Use only selected data.

Fill in these values and press the grid button . ▦



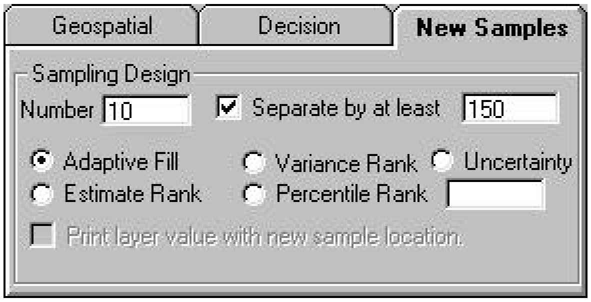
You should see the following effect. If you wish to cut away part of the grid. Then use the select tool or space definition tool to achieve this. To use the select tool press the ↖ button. Move your cursor over the picture and it should change into cross hairs. When you left click on the picture, a node is dropped and construction of a polygon begins. Each time you left mouse click a new vertex is dropped. Continue this until you have the polygon you want and the double click the left mouse button to finish.



Only those blocks within the polygon will be considered as candidates for new sample locations.

2.2 Minimum Separation Criteria and Clustering

Under certain circumstances, the strategies presented here will present clustered groups of new sample locations. Particularly for dense grids. As an example, the estimate rank places new samples where modeled block values are the highest. Higher block values usually occur around the highest sampled point. As a results, the true optimization routine will place new samples in those blocks closest to the highest sampled point. From a practical standpoint, this is not often a wise decision. What is often intended is to spread samples throughout the hot spot area. This is achieved by specifiying a separation distance between each new sample and previously sampled locations (including previously identified new sample locations). This provides the effect of spreading out the new sample locations while adhering to the goals of the sample strategy. To specify a minimum separation distance choose the New Samples tab on the control panel.



Geospatial | Decision | **New Samples**

Sampling Design

Number 10 ☑ Separate by at least 150

● Adaptive Fill ○ Variance Rank ○ Uncertainty

○ Estimate Rank ○ Percentile Rank

□ Print layer value with new sample location.

Select the option *Separate by at least* and enter the minimum distance value in the box on the right. The units for this separation should be the same as whatever units your coordinate system is in (ft, meters, etc).
The minimum separtion distance affects the results of every strategy except for adaptive fill. Typically clustering is not a problem for variance rank but one may choose to use the option regardless.

2.3 Adaptive Fill

Goal

The goal of this approach is to fill in sparsely sampled areas.

Method

Adaptive fill is the simplest of all the sampling strategies and is the only one independent of an spatial interpolant. Adaptive fill places new samples in the largest spatial data gaps. In other words, this method will place the next sample at the point that is farthest away from any previously sampled point within the bounds of the sampled region. For three dimensional data, the same approach is true as a sample will be placed in 3d space at the location farthest from any previously sampled location. SADA applys adaptive fill by first laying a grid down over the site. The center of each block within the grid system, becomes a candidate for a new sample location. SADA then cycles through each of the blocks and calculates the distance between the center of the block and the closest sample point. The block with the largest calculated distance becomes the location of the new sample. The sample is placed at the center of the block. If more than one sample is requested, this procedure is repeated again. After each round, the new sample location is considered to be a previous sample point in the next round.

Pros

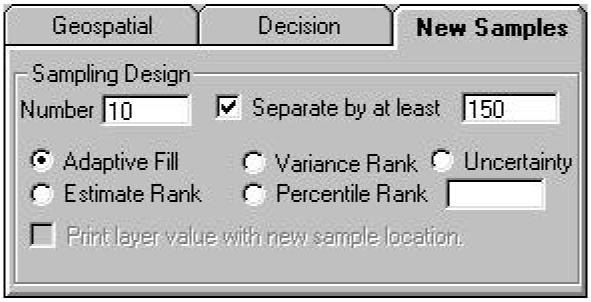
- Because adaptive fill is independent of any spatial interpolant, it is quite easy to implement.
- Quickly fills in data gaps for exploratory sampling.

Cons

- Because adaptive fill only considers distance between points and not the results of any samples, the method may place new samples in areas of little concern. In SADA, the user can overcome this through the use of polygonal definition (see implementation)

Implementation

The first step is to lay a grid of appropriate resolution over the site. The next is to select the New Samples tab on the Control Panel.



Geospatial | Decision | **New Samples**

Sampling Design

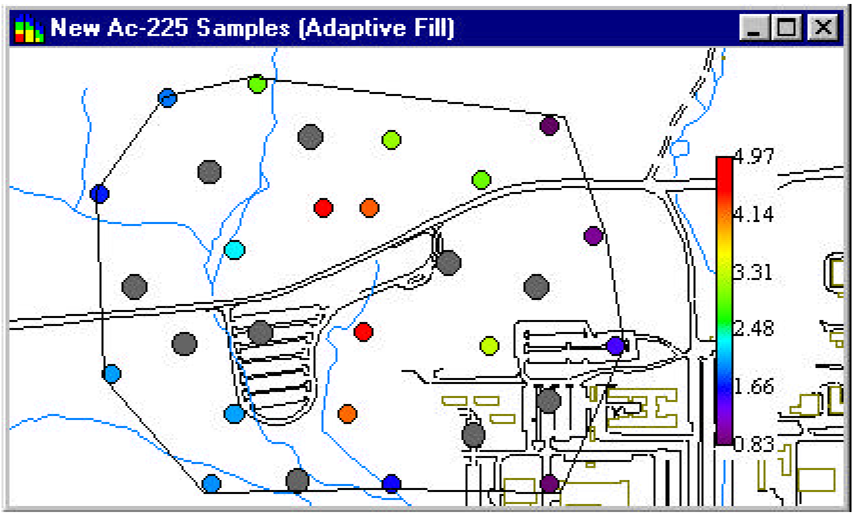
Number 10 ☑ Separate by at least 150

● Adaptive Fill ○ Variance Rank ○ Uncertainty

○ Estimate Rank ○ Percentile Rank

□ Print layer value with new sample location.

Here you will enter the number of new samples in the box beside *Number* and select the *Adaptive Fill* option. Selecting or deselecting the *Separate by at least* option will have no effect on the results as this option is ignored for Adaptive fill. For three dimensional data, you can select the *Print layer value with new sample location* option and SADA will print the depth value next to 2d view of the sample location. When these parameters are set press the . Note that if this button is already pressed the press the apply button . to reapply the analysis. The following result is an example application of the adaptive fill method. The new sample locations are identified as larger gray circles.



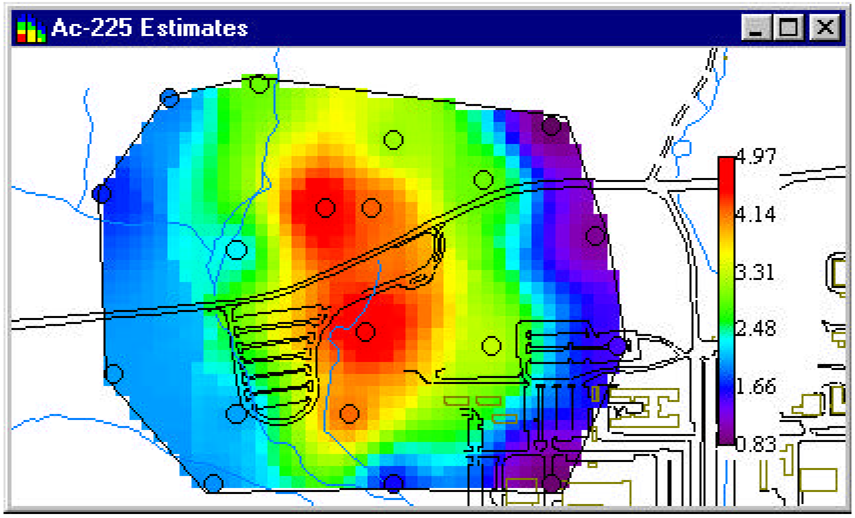
2.4 Estimate Rank

Goal

The purpose of this method is to place new samples where modeled or interpolated concentrations are the highest. This method is useful for confirming the existence or extent of hot spots.

Methodology

This approach requires an spatial interpolation be implemented. Four methods exist in SADA: Ordinary Kriging, Indicator Kriging, Inverse Distance, and Nearest Neighbor. Explanation of these methods is beyond the scope of this paper but all four are found in the SADA help system. As a review the purpose of geospatial interpolation is to contour an attributes values across the site. In many environmental applications, this is often the concentration values. The following shows the result of applying the inverse distance method to spatially distributed data. As always, the spatial interpolation as well as the estimate rank method that follows requires a grid be defined over the site.



The estimate rank approach, cycles through each of the interpolated values (one for each block) and identifies the location of the block with the highest value. The center of this block is then chosen as the next location for a new sample. If more than one sample location is requested, SADA re-interpolates the site treating this new sample as a real data value whose concentration is taken to be the predicted value. After the re-interpolation, this process is repeated for the next sample. After a block has been selected as the location of a new sample, it is no longer considered in future sample locations. Otherwise, all the new sample locations would pile up at the same location.

The exception occurs when implementing the minimum separation distance criteria. In this case, SADA cycles through all block values whose center is separated from any previously sampled value by at least the minimum separation distance.

Pros

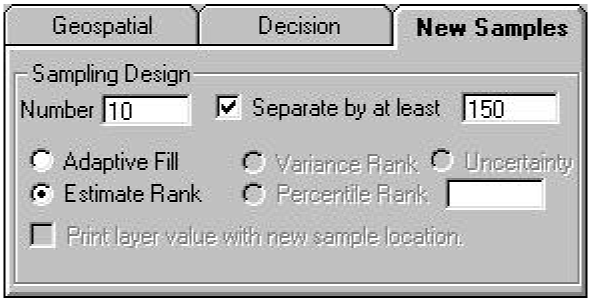
- Good for confirming the presence or extent of hot spots.

Cons

- This method does not consider model variance and may place new sample locations in well characterized locations.
- The secondary minimum separation distance constraint is often needed to prevent clustering.

Implementation

The first step requires defining a grid across the site. The next step is to set up a geospatial model that will interpolate data across the site. The last step is to select the New Samples tab on the control panel.



Sampling Design

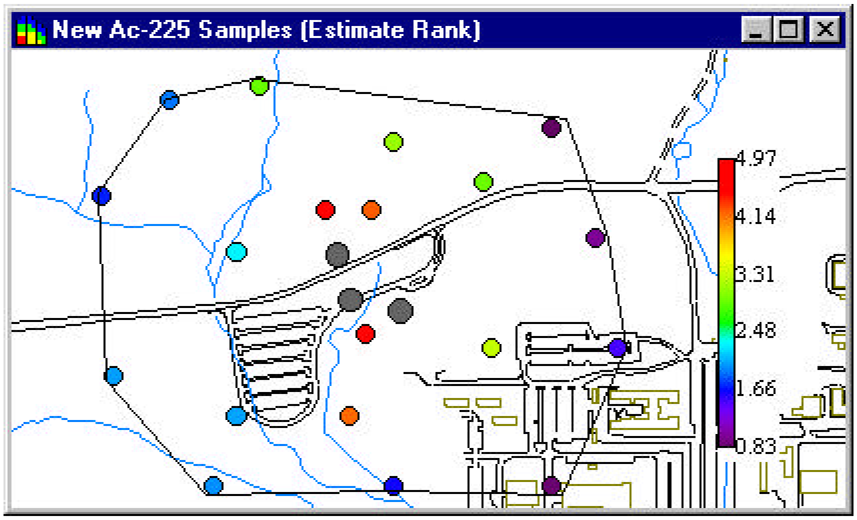
Number 10 ☑ Separate by at least 150

○ Adaptive Fill ○ Variance Rank ○ Uncertainty

● Estimate Rank ○ Percentile Rank

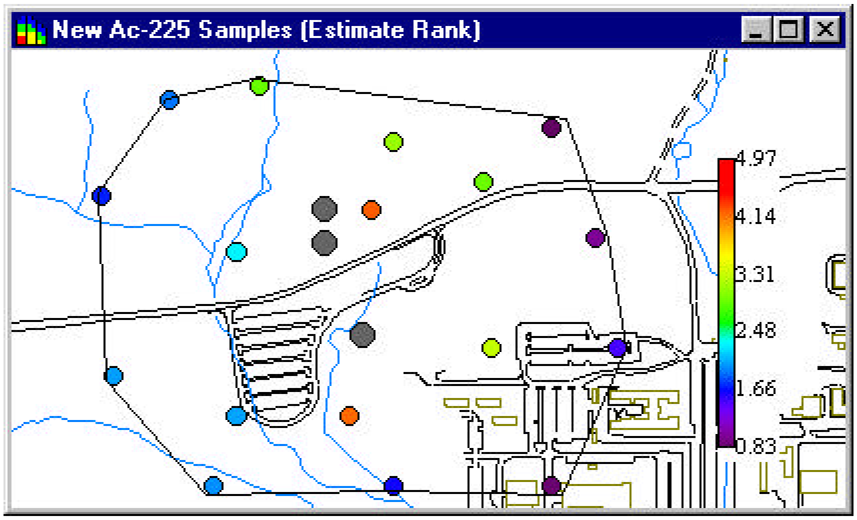
□ Print layer value with new sample location.

On this tab, enter the number of new samples into the *Number* box, select the *Estimate Rank* option and set the minimum separation distance criteria if needed. For three dimensional data, you can select the *Print layer value with new sample location* option and SADA will print the depth value next to 2d view of the sample location. When these parameters are set press the . Note that if this button is already pressed the press the apply button 📷. to reapply the analysis. The following result is an example application



of the estimate rank method with three new samples and a minimum separation of 150 feet. The new sample locations are identified as larger gray circles.

The following result show how the estimate rank will cluster new samples around the highest sample point when the minimum separation distance option is not used.



Notice how the first new sample location is located virtually on top of the highest sampled value (top new sample). The next one found lower is only one block away. The third and final sample location is found virtually on top of the second highest sampled location (lowest of the three).

2.5 Variance Rank

Goal
The goal of variance rank is to place new samples where the models local estimation variances are the highest. From a practical standpoint, this places new samples where the model is having the greatest "difficulty" in contouring the attribute. New samples in these locations will reduce the local variance and spread the effect throughout the region.

Methodology
This approach requires a grid definition and use of the ordinary kriging model. This model provides a set of model variances for each block. SADA cycles through this set of blocks and identifies the block with the highest model variance as the next sample location. If the secondary constraint is in effect, only blocks whose centers are separated from nearby data points by the minimum distance are considered. If more than one new sample location is requested, this new sample location is treated as an actual sample point whose value is equal its modeled value and the site is re-interpolated From this re-interpolation the next new sample is chosen. This iterative process is repeated for each new sample.

Pros

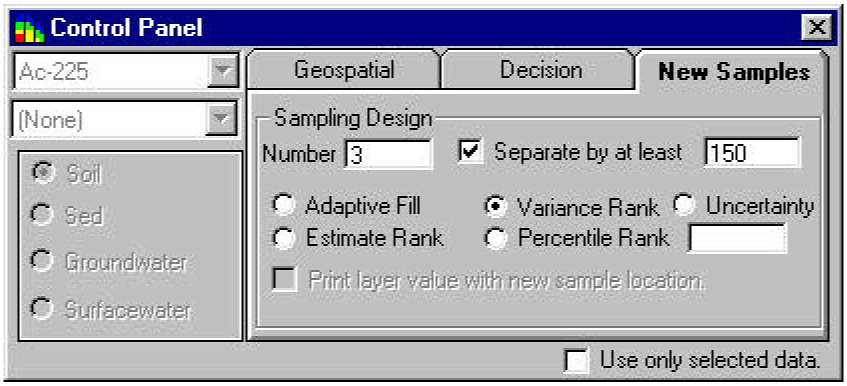
- Good for reducing model variances across the site.

Cons

- Does not consider concentration magnitudes and may place new sample values in areas where the concentration values are relatively low or are of no interest. Users may combat this by using the polygon tool to eliminate these areas.
- Only available with ordinary kriging.

Implementation

The first step is to define a grid and set up the ordinary kriging model. This is explained in the SADA help system. The next step is to select the New Samples tab on the control panel.



Control Panel

Ac-225 ▾

[None] ▾

● Soil
○ Sed
○ Groundwater
○ Surfacewater

Geospatial | Decision | New Samples

Sampling Design

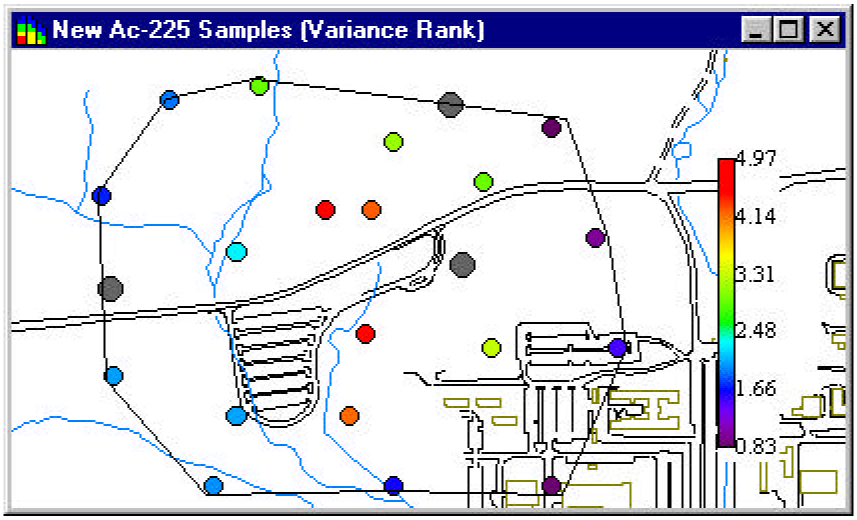
Number 3 ☑ Separate by at least 150

○ Adaptive Fill ● Variance Rank ○ Uncertainty
○ Estimate Rank ○ Percentile Rank

□ Print layer value with new sample location.

□ Use only selected data.

On this control panel, enter the number of new samples into the *Number* box, select the *Variance Rank* option and if needed setup the minimum separation criteria option at the top. Typically, variance rank won't need this option. . For three dimensional data, you can select the *Print layer value with new sample location* option and SADA will print the depth value next to 2d view of the sample location. When these parameters are set press the . Note that if this button is already pressed the press the apply button .to reapply the analysis. The following result is an example application .

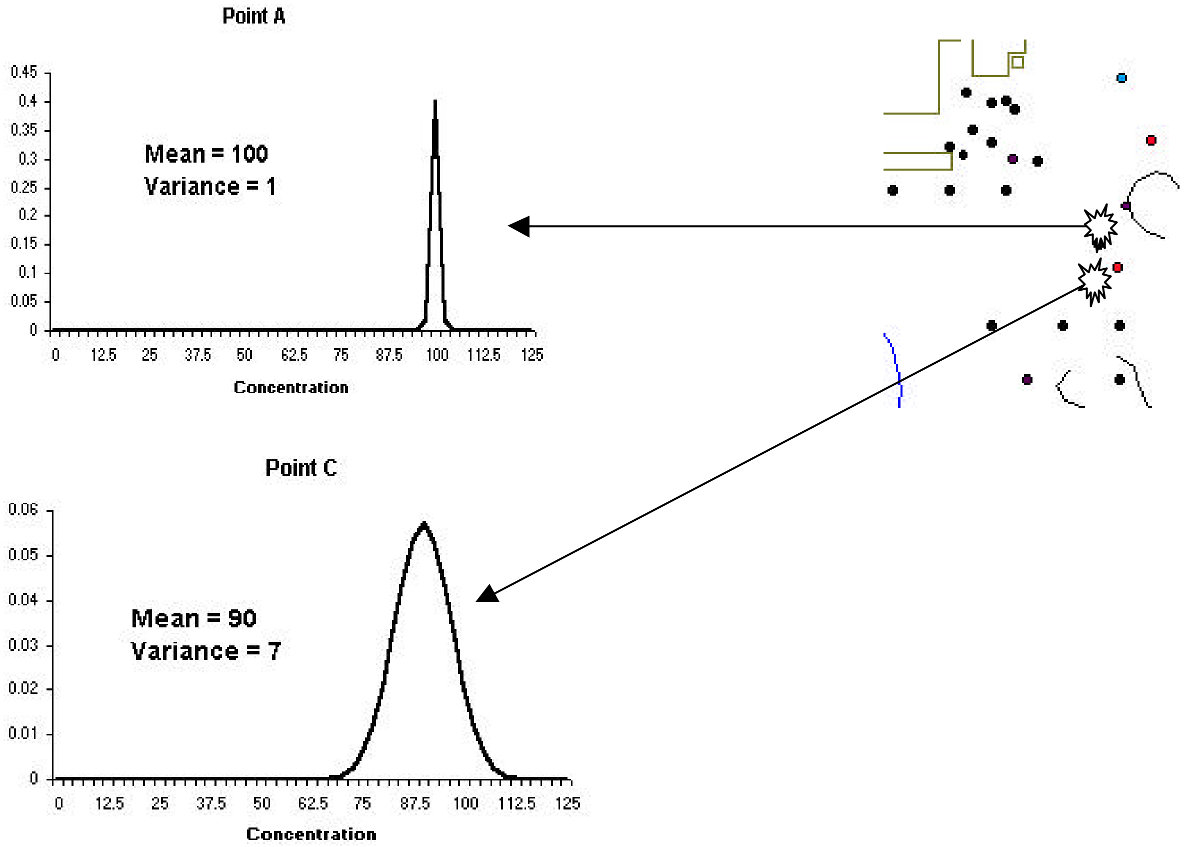


New sample locations are identifed as larger gray circles.

2.6 Percentile Rank

Goal

The percentile rank method considers both magnitude and variability when selecting new sample locations. The percentile rank is in some sense a merger of estimate rank and variance rank. This method places new samples in locations with the potential to be the highest concentrations on the site. The difference between percentile and estimate rank here is subtle but important. Estimate rank will locate new samples in locations in which the best estimated values are highest. Percentile rank places new samples in locations that might not have the highest best estimated values but because of the uncertainty (modeled here with local ccdfs) in estimation they have the potential to be extremely high. As an example consider the following two local probability distribution functions generated by ordinary kriging .



In the two probability distribution functions above, estimate rank would choose Point A because the mean (the best estimate for ordinary kriging) is greater than the mean of 90 for Point C. However, percentile rank would choose Point C because the a larger proportion of the distribution, while centered about 90 extends further up the scale. In the percentile rank, one picks a particular percentile of the distribution such as 90th to compare between distributions. The goal of this method is similar to estimate rank in terms of hot spot confirmation and for certain sites users may find little difference between the two.

Methodology

In this method, the user must specify a percentile value. Typically you will want to use a high percentile value (greater than 50th) such as 90th or 95th. Choosing low percentile values such as 5th or 10th typically are not interpretable with respect to the goal of identifying potentially high concentration values. SADA then cycles through each block in the grid, uses the conditional cumulative distribution function derived at the center of each block to calculate the 90th percentile concentration value. The block with the highest 90th percentile concentration value becomes the location of new sample point. If more samples are required, the best estimated value for this sample location is used a true sample location and the site is re-estimated (ie each ccdf in the center of every block is recalculated). The process is repeated for each new sample.

Pros

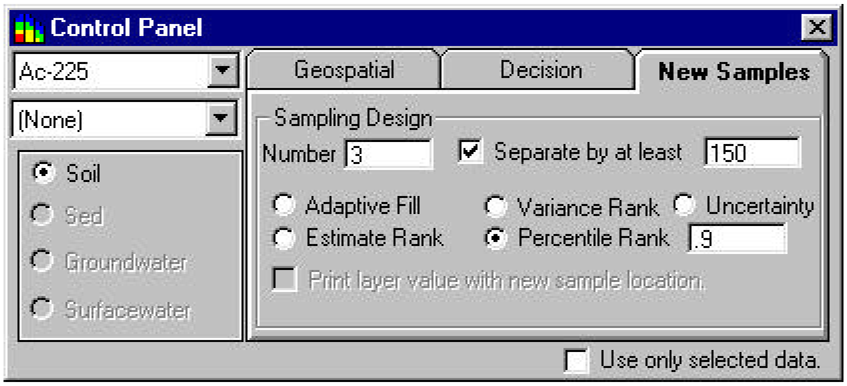
- Considers both magnitude and variability providing a tool for placing new samples in areas of potentially high concentrations.

Cons

- May often require a secondary minimum separation distance constraint to prevent clustering.
- Only available in SADA with ordinary kriging and indicator kriging.

Implementation

As with every sampling design and interpolation scheme, the first step is to define a grid. After defining the grid, choose either ordinary kriging or indicator kriging and parameterize the selected model. Next choose the New Samples tab of the control panel.



Control Panel

Ac-225

(None)

● Soil
○ Sed
○ Groundwater
○ Surfacewater

Geospatial | Decision | New Samples

Sampling Design

Number 3 ☑ Separate by at least 150

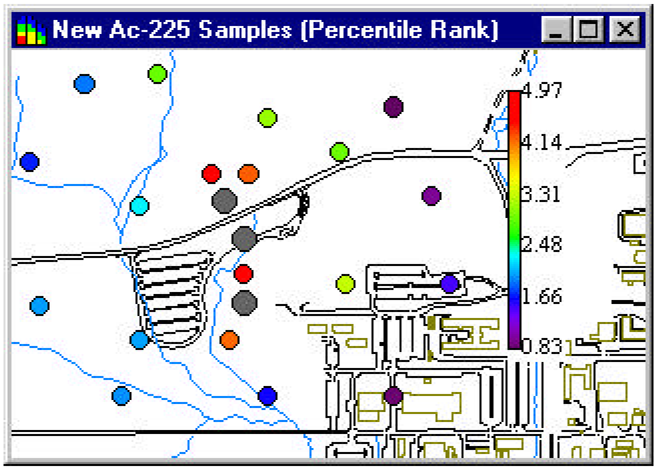
○ Adaptive Fill ○ Variance Rank ○ Uncertainty

○ Estimate Rank ● Percentile Rank .9

☐ Print layer value with new sample location.

☐ Use only selected data.

Enter the number of new samples in the *Number* box and select the *Percentile Rank* option. In the box to the right of the percentile option, enter the percentile value (e.g. .9, .95) to use in the calculations. If needed, select the minimum distance constraint and enter a minimum separation distance value. For three dimensional data, you can select the *Print layer value with new sample location* option and SADA will print the depth value next to 2d view of the sample location. When these parameters are set press the . Note that if this button is already pressed the press the apply button . to reapply the analysis. The following result is an example application .

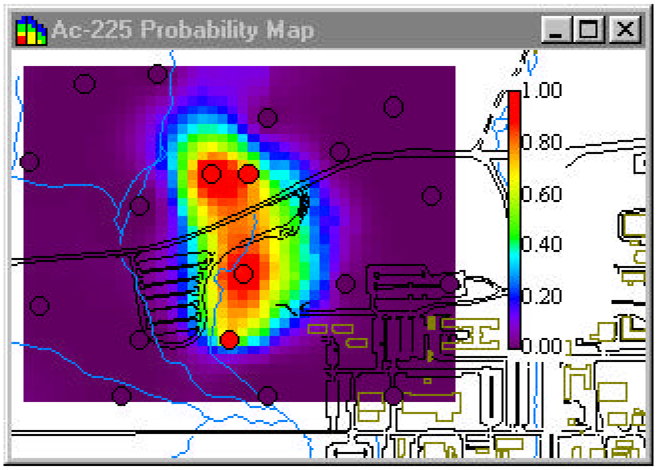


New sample locations are identified with larger gray circles.

2.7 Uncertainty Rank

Goal
The goal of uncertainty rank is to place new samples in areas where the likelihood of exceeding a specific threshold value is most uncertain. This method is useful for boundary delineation. Specifically the boundary of the area of concern.

Methodology
This approach uses the conditional cumulative distribution function determined at each point by either the ordinary kriging or indicator kriging method. Given a threshold value, the probability of exceeding this value is calculated at each point. SADA then cycles through each block and identifies the one whose probability of exceeding the threshold at its center is closest to .5. In other words the block who is as about as likely to exceed as not to exceed. The sample location is selected here and if more samples are requested, the best estimated value at this point is treated as the true sample value and the site is re-interpolated (ie each ccdf is recalculated). The process is then repeated for the next sample. The practical effect of this method is to place new samples in the green shaded areas of the probability map (see help file). The following picture shows the probability of exceeding 4 pCi/g. The percentile rank method will place new samples in the green shaded areas – those areas corresponding to probabilities near or equal to .5 (see legend on right).



In some situations, one may need to use the minimum separation distance constraint. The reason for this is that those points closest to .5 in probability of exceedance are also those points whose best estimate is likely closest to the threshold value. This is especially true for ordinary kriging who uses the 50th percentile or mean of the distribution as the best estimate. In this situation a concentration value very close or equal to the threshold value will be used as the simulated sample value. Because of this proximity to the theshold value, it may remain quite uncertain as to whether nearby blocks will be exceeding the threshold value. As a result, these nearby blocks may be selected in subsequent interations. This does not necessarily occur in every situation, but if it does this is likely the reason. To get a good spread of points around the area of concern (throughout the green shaded areas of the probability map) use the minimum separation distance constraint.

Pros

- Can be explicitly connected to cleanup goal for a site.
- Useful for delineating boundaries of areas of concern

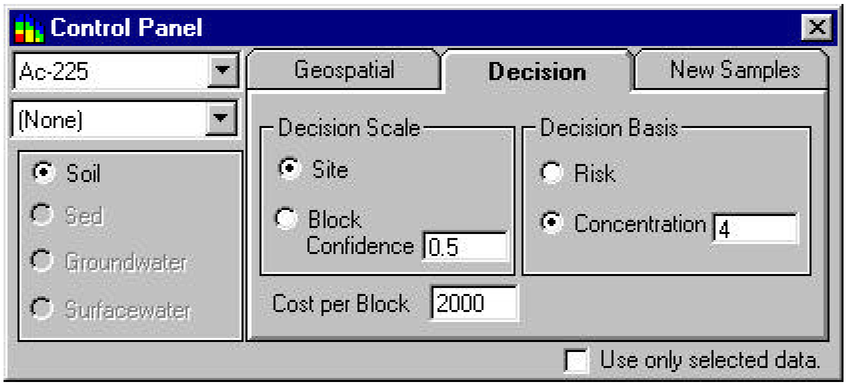
.

Cons

- In some situations, the minimum separation distance constraint may need to be implored.
- Not available for nearest neighbor or inverse distance.

Implementation

As with the percentile rank method, one must define a grid and select either ordinary kriging or indicator kriging. After parameterizing these models, click on the Decision tab of the control panel.



Ac-225

(None)

● Soil
○ Sed
○ Groundwater
○ Surfacewater

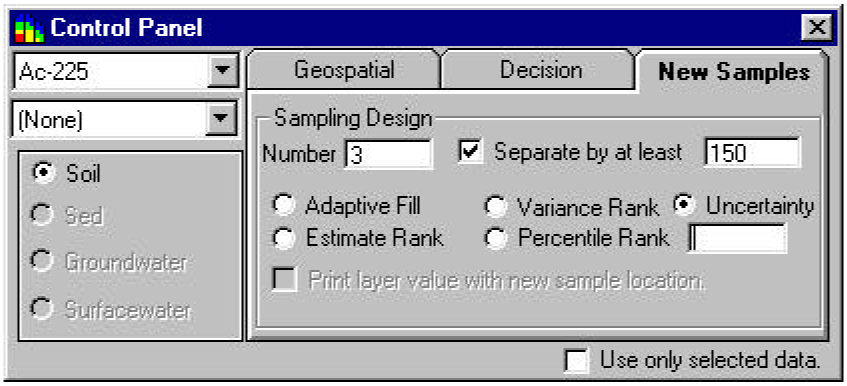
Decision Scale
● Site
○ Block
Confidence 0.5

Decision Basis
○ Risk
● Concentration 4

Cost per Block 2000

□ Use only selected data.

Under the *Decision Basis*, if the human health risk module has been setup, the *Risk* option will be available. If not only the *Concentration* option will be enabled. Choose the basis for new samples. If the concentration basis is selected, enter the concentration threshold into the box to the right of the *Concentration* option. The uncertainty rank method will be implemented relative to this decision goal. Next select the select the New Samples tab on the control panel.



Control Panel

Ac-225

(None)

● Soil

○ Sed

○ Groundwater

○ Surfacewater

Geospatial | Decision | New Samples

Sampling Design

Number 3 ☑ Separate by at least 150

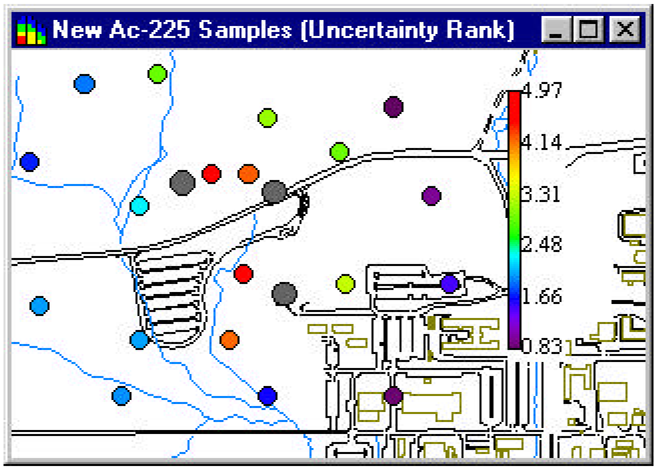
○ Adaptive Fill ○ Variance Rank ● Uncertainty

○ Estimate Rank ○ Percentile Rank

□ Print layer value with new sample location.

□ Use only selected data.

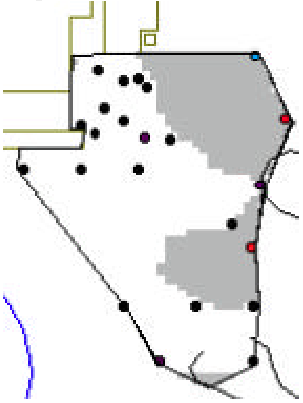
Select *Uncertainty Rank* and enter the number of samples into the *Number* box. If needed, select the minimum distance constraint and enter a minimum separation distance value. For three dimensional data, you can select the *Print layer value with new sample location* option and SADA will print the depth value next to 2d view of the sample location. When these parameters are set press the . Note that if this button is already pressed the press the apply button . to reapply the analysis. The following result is an example application .



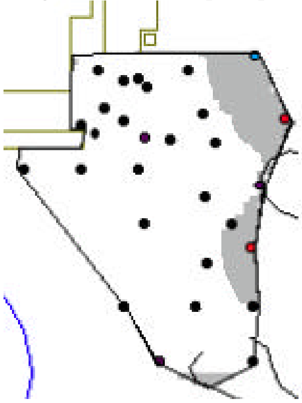
New Ac-225 Samples (Uncertainty Rank)

3.0 How many samples?

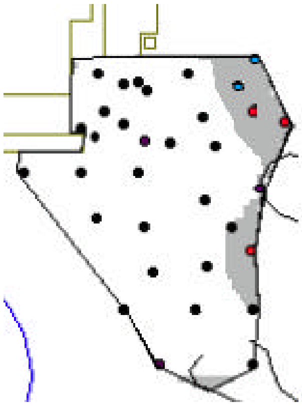
The question of how many samples can be addressed in SADA from a data worth or cost benefit perspective. Given an iterative sampling scheme, one could observe the effect of adding additional samples to the decision outcome or simply the modeling results. Consider the following cleanup maps. The first image shows the decision map after the first round of samples has occurred. The second shows the cleanup map after 6 additional samples and the final after 10 additional samples. Between the second and third maps, it is clear that with each additional sample the change to the cleanup map is reduced.



After intial sampling

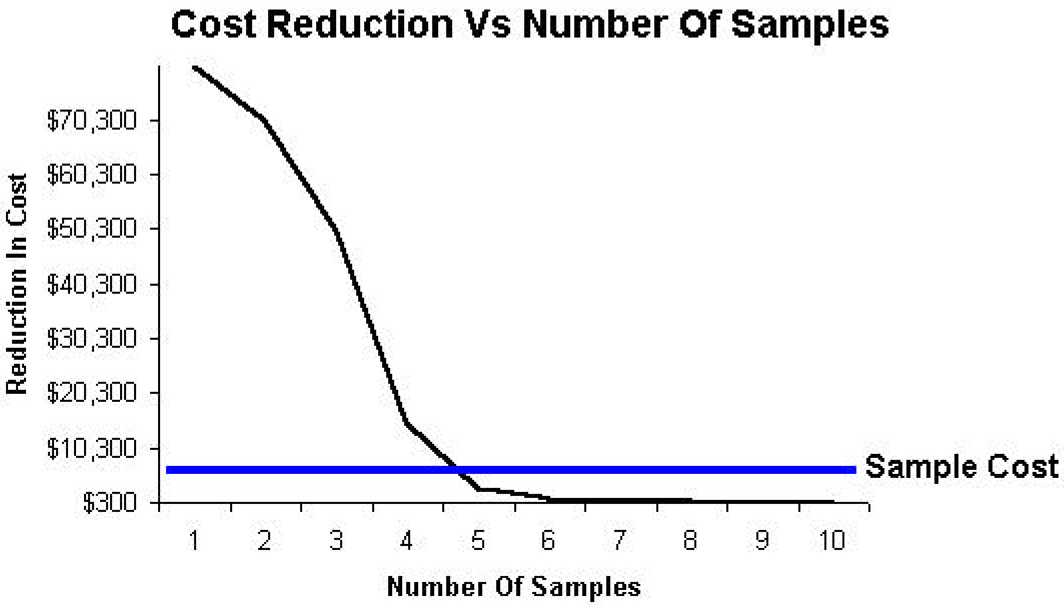


After 6 additional samples



After 10 additional samples

This can be quantified in this situation with a cost benefit curve which shows the reduction in cleanup costs per additional sample. In the graph below, the blue sample cost line represents the cost of taking a single additional sample. When the reduction in remedial costs is less than the amount spent to sample, this provides a criteria to stop sampling.



The items covered in section 3.0 are not currently available in SADA Version 1. However, the results of subsequent sampling can be output from SADA and graphs such as the one above created outside of SADA. Plans exist to formally implement these capabilities in upcoming SADA versions.