Chapter 37: Secondary Sampling Designs

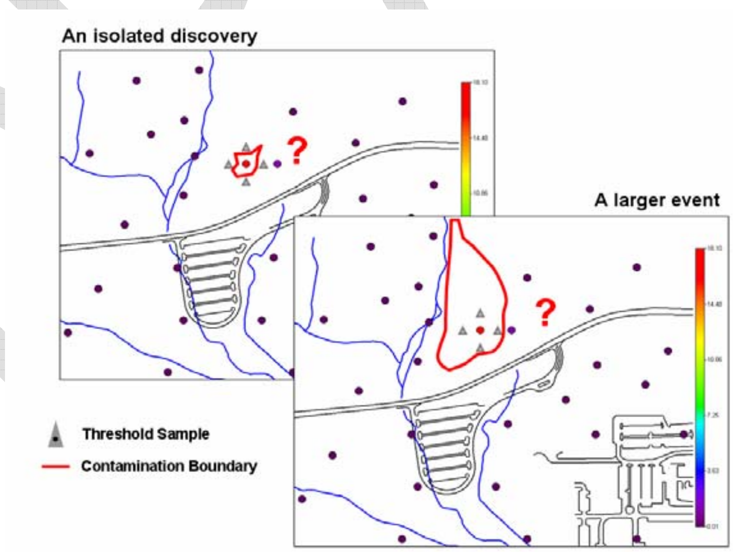
Secondary sampling designs are applied after some data or other information has already been obtained. Generally, the objective of any secondary sampling design is to further refine the model or the decision in some very specific way. Secondary designs can be either point (sample) or model (geospatial model) based. Before reading this chapter please make sure you've read the previous chapter Overview of Sampling Designs. 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 occassions warrant a strictly 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

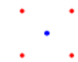
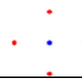
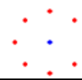
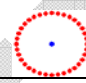
Threshold Radial (also known as Adaptive Cluster Sampling) is a straightforward secondary sample design that places samples in a radial pattern around existing data points that exceed a decision threshold. The user has control over the pattern of the surrounding new sample points. They can be circular 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 that can help determine whether the result is an isolated event or the edge of a larger contamination event.



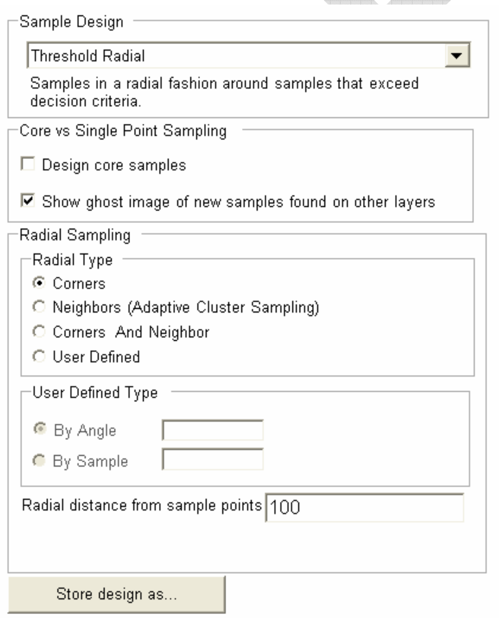
When you execute the threshold design, SADA will ask you for a decision criteria. Those measured values that exceed this criteria will be the target of the sampling design. Normally this decision criteria is a single value typically the screening value (see chapters on performing data screens). However recall that under the General analysis it is possible to specify depth specific decision criteria. In the example that follows we will use a two dimensional (surface)

Copyright © 2008 University of Tennessee. All Rights Reserved

example so depth specific criteria isn't necessary. Please see the chapter on Performing a quick data screen for where and how to set up and use depth specific criteria. The following configurations are possible:

Name	Appearance
Corners	
Neighbors	
Corners + Neighbors	
User Defined	

In 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 list of available sample designs 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 type 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 criteria. If you wish to use depth specific criteria (we don't) you can select this option under set decision threshold type found in the next step. Press the Show the Results button.

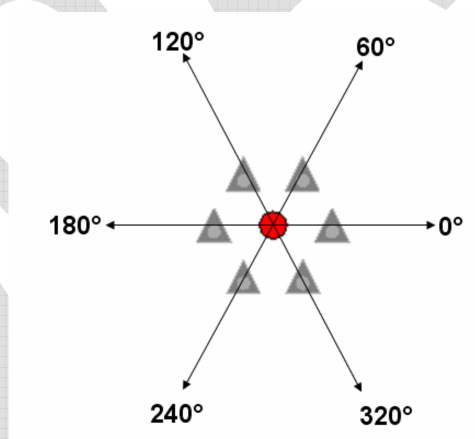
Since we are under the General analysis, SADA will now ask for the decision criteria to manually entered. Had we been under the human health risk, ecological risk, or custom analysis settings we would receive a window with options for selecting appropriate scenarios for each of these. If you have read 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

Copyright © 2008 University of Tennessee. All Rights Reserved

reports that only one value exceeded our decision criteria. New samples appear as triangles in the image below.

Try now neighbors and corners-and-neighbors now on your own.

When you use User Defined arrangements, you can specify 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 angle increment you specify. 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.

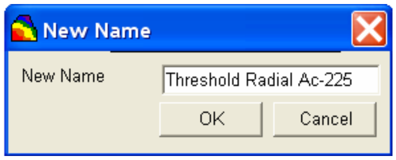


Select User Defined in the radial sampling block, by angle in the User Defined type, and 60 for the angle increment. Press show the results and enter 4.8 as the decision criteria. SADA reports one sample exceeded the criteria and 6 new samples are generated.

You can now store this result by pressing the Store Design As… Enter Threshold Radial Ac-225 as the name.

DRAFT

Copyright © 2008 University of Tennessee. All Rights Reserved



To see this result, select Stored Results from the drop-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.

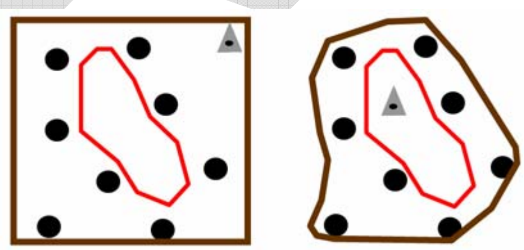
Adaptive Fill Design

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 previous chapter). 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 satisifies 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.

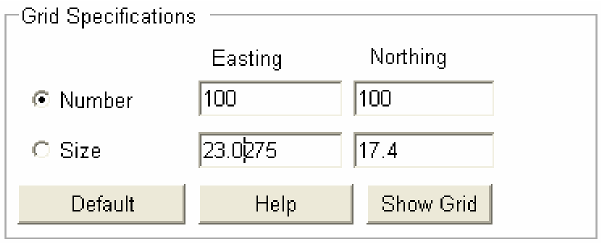
This design requires special care be taken in defining the study area. If the study area is arbitrarily chosen then spatial data gaps map appear artificially along the boundaries. Consider the following example.



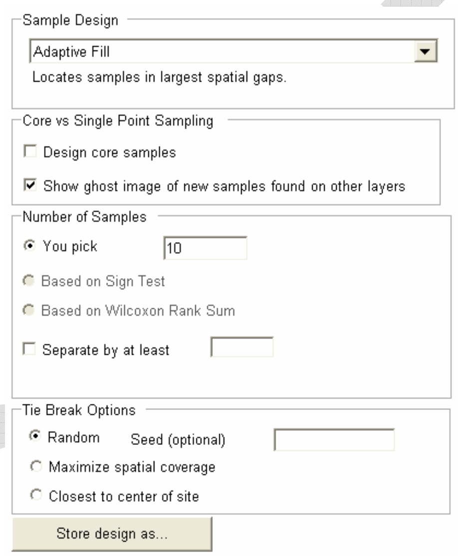
On the left we have a set of data points collected such that an area in the center has been undersampled (red boundary). We want to use the adaptive fill method to place one more sample in this area to create optimal spatial coverage. However, because of our rigid rectangular definition of the site boundary (thick brown line), SADA places the new sample (gray triangle) in the upper left 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-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.

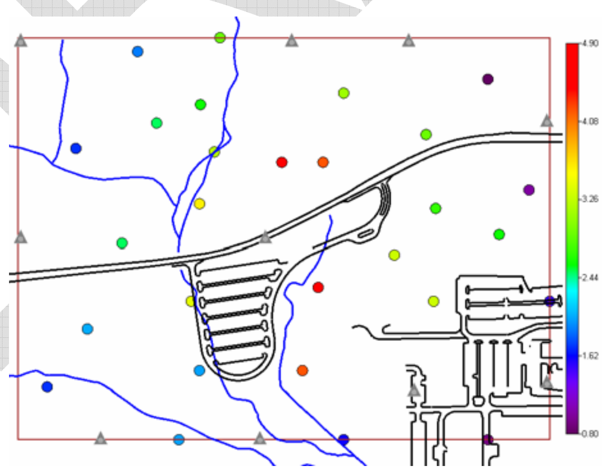
Copyright © 2008 University of Tennessee. All Rights Reserved



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 maybe decided each time we reapply the design. Note that it is difficult to determine if ties are occuring just by applying the design. Your parameter window should look like this.



Press show the results buttton.



The results seem to favor the area along the site boundary. To see the site boundary click on set up the site 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).

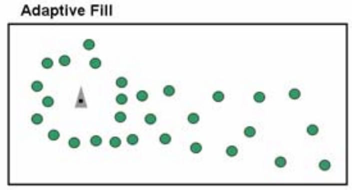
Copyright © 2008 University of Tennessee. All Rights Reserved

Let's export these results to .csv file now. Click on the export to text button (). Enter the name AdaptiveFill in to 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.

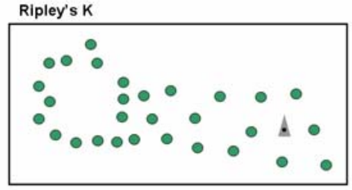
Ripley's K

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 the chapter on Local Index of Spatial Association (LISA). If you have not read that chapter yet, you are encouraged to do so now. Ripley's k value is evaluated at node in the grid by specifying a simple search neighorhood (anisotropic neighborhoods not yet available) about that node 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 left.

Adaptive Fill



Ripley's K

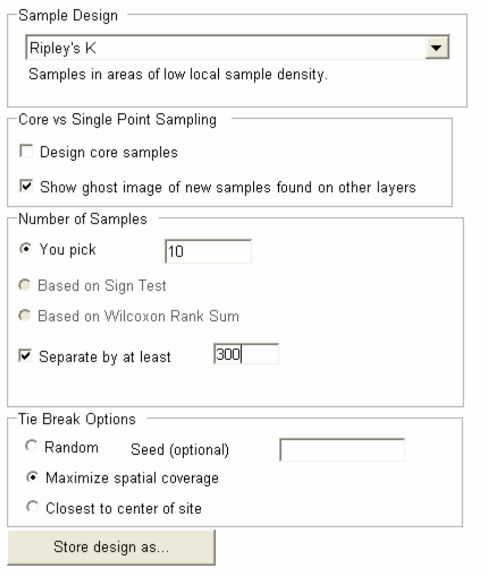


Ripley's K may require use of the minimum distance constraint as nodes of low sampling density will likely be clustered together. Previous versions of SADA simulated the placement of Ripley k values to offset this problem. This approach will be reinstalled into SADA in the near future. In the meantime, 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 this.

DRAFT

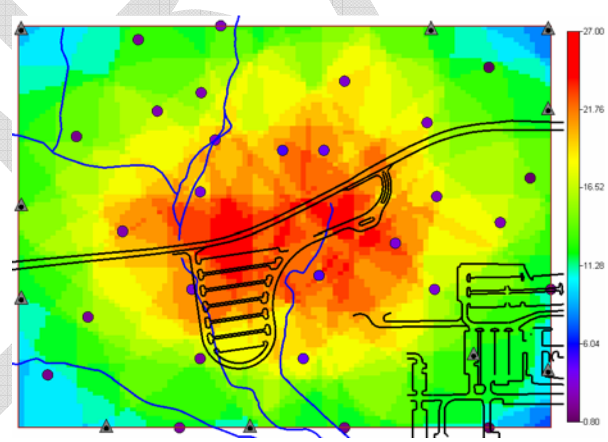
Copyright © 2008 University of Tennessee. All Rights Reserved



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 that at the time this text was written, a bug appears in the interface at this point (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 button. SADA places the 10 new samples over the top of the underlying Ripley's K map from which these locations are based.



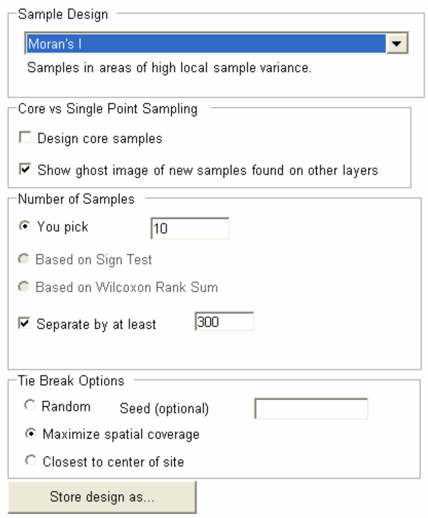
Moran's I

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 the chapter on Local Index of Spatial Association 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.

DRAFT

Copyright © 2008 University of Tennessee. All Rights Reserved

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 this.



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.

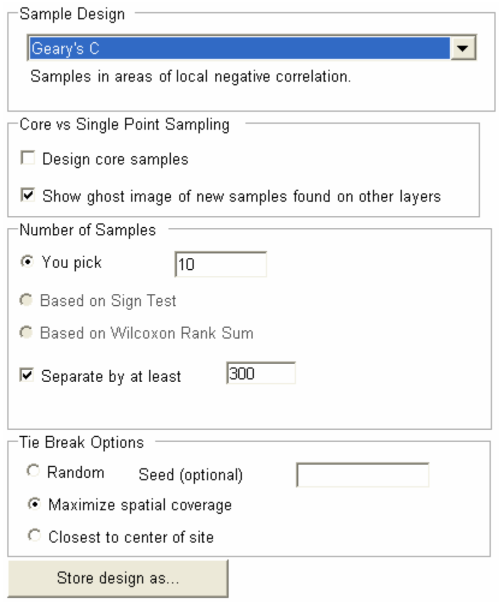
Geary's C

This sample design places samples in those areas of with greater (in magnitude) negative correlation among samples found in the search neighborhood. The Geary's C map is discussed in the chapter on Local Index of Spatial Association 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 hetereogeneity is measured not by local variance but 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 this.

DRAFT

Copyright © 2008 University of Tennessee. All Rights Reserved

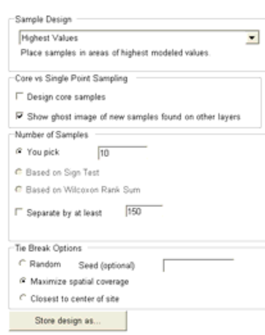


From the previous result, you should have the LISA search radius set to 1000 and grid specs at 100x100. Press Show The Results. The new sample locations are plotted along with the Geary's C map on which each location was based.

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 contraint 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.

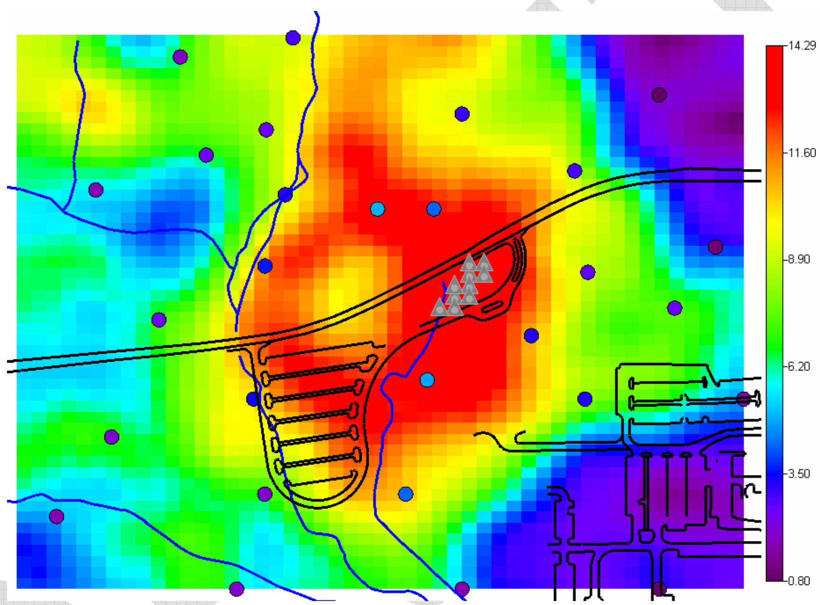
If you have not already done so, open SampleDesigns.sda. With Soil and Ac-225 selected, choose the interview Develop a Sample design. Click on the Set sampling parameters step and choose Highest Values Design from the drop-list of available methods. We will place 10 new samples and choose to separate them by a minimum of 150 feet. However, for demonstration purposes, 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 this.



Copyright © 2008 University of Tennessee. All Rights Reserved

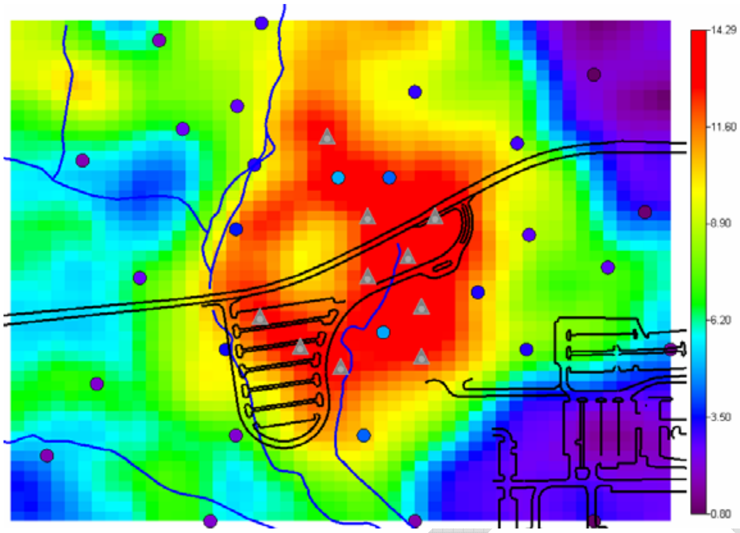
Notice that a number of steps appear that you've seen before when doing geospatial modeling. The next step is to specify grid parameters. The next one is to choose an interpolation method and so forth. In order to simplify the example, we'll choose a stored model so that we can dispense with calibrating 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 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 approch we will use now.

Click on the Interpolation Methods step. From the drop list select User Stored result. Field Detection is a geospatil 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.

Copyright © 2008 University of Tennessee. All Rights Reserved

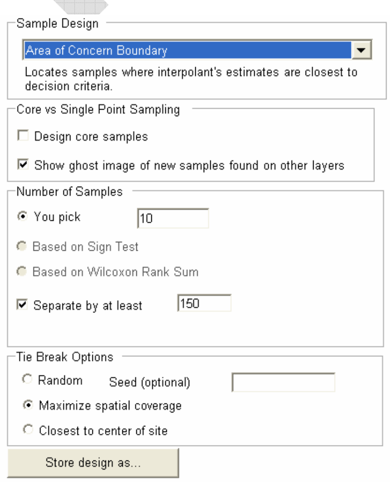


Thre result is a more reasonable distribution of new sample locations throughout the high value region providing a balance between strict mathematical optimization and practical sample design.

Area of Concern

This sampling design places samples along the boundry line in the AOC result. In particular, those nodes that have a value closest to the decision criteria 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 the chapter on Determing the Area of concern. If you have not read this chapter yet, you should do so now before proceeding.

In the file SampleDesigns.sda, select soil and Ac-225 and the inteview Develop a Sample Design. Click on the Set sampling parameters step and choose the Area of concern boundary design from the drop-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 breaker method set to maximize spatial coverage) Your parameter window should look like this.

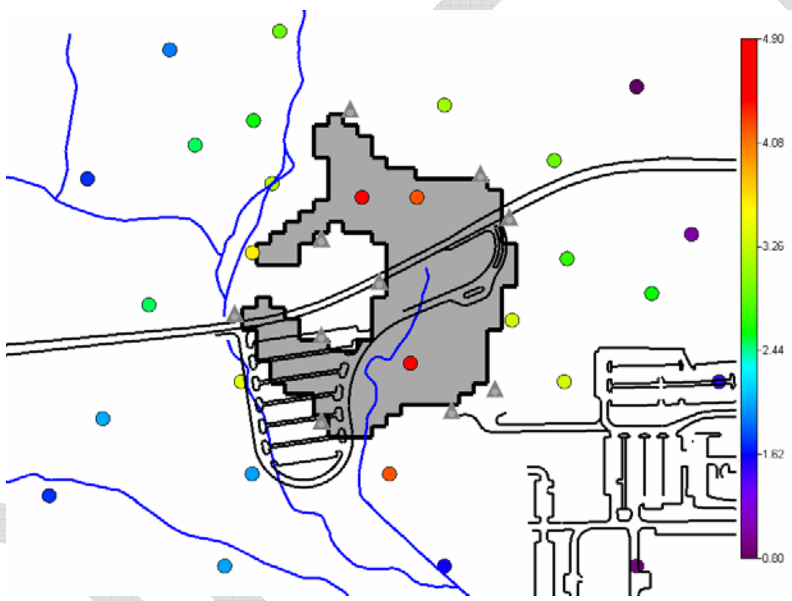


DRAFT

Copyright © 2008 University of Tennessee. All Rights Reserved

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 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 approch we will use now.

Click on the Interpolation Methods step. From the drop list select User Stored result. Field Detection is a geospatil 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 criteria (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 sample closely follow the AOC boundary line and space them out rather evenly when the minimum distance constraint has been imposed.

Practice some now on exporting these results to 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-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

Copyright © 2008 University of Tennessee. All Rights Reserved

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, simple press Show The Results and the result is restored.

DRAFT

Copyright © 2008 University of Tennessee. All Rights Reserved