

Chapter 33: User Models

In some situations, you may want to manually create a 2 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 prior to 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 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 to you. 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 (later chapter). 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.

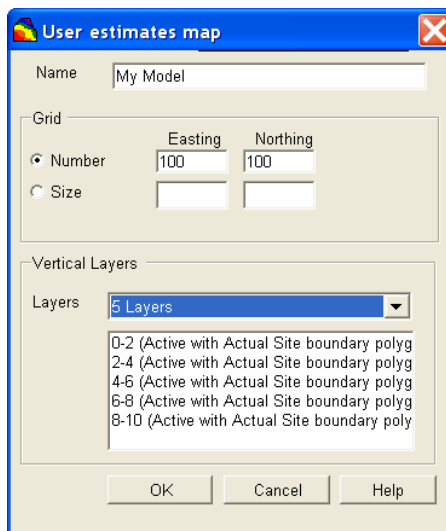
There are two kinds of user models: standard and probabilistic. A standard user model allows you to distribute any kind of values in model space. You can "paint" in 2d and 3d space any type of values you wish. They may represent concentration values or perhaps categorical values such as soil type. The possibilities are really endless. A probabilistic user model spatially expresses the probability of something being true (or false). This could be the probability that a decision criteria would be exceeded. It could also be the probability that a contaminant is present. Again the possibilities are really endless.

We will show you how 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 will take up this discussion again and demonstrate how to distribute samples according to objectives less statistical or classical in their objective. Both user model and standard models are created exactly in the same way. The majority of the discussion will be found in the standard user model section.

Standard User Model

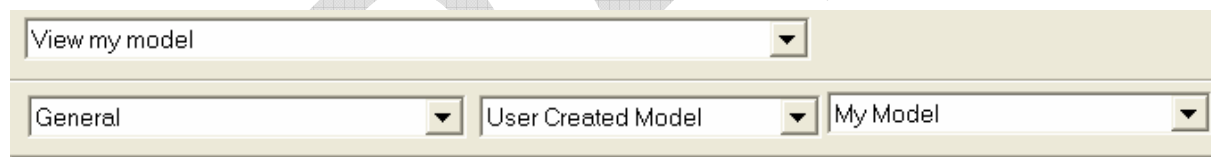
Open the file UserModels.sda. This SADA file has no data in it but the GIS layers have imported and the site boundaries have been drawn (under setup the site). We will be drawing a 3d model so we've created a vertical layering scheme called 5 Layers. Each layer in this layering design is 2 ft thick.

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 will ask you two things at this point. First, provide a name for you model ("My Model").

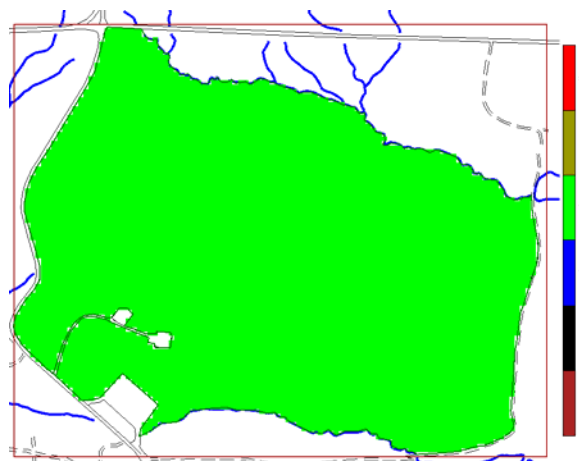


Next we'll set the grid resolution. Let's enter 100 x 100 grid cells by selecting the Number option and entering 100 and 100 for both easting and northing. This will divide the site boundary box into 100 x 100 cells. Cell width and height may be different depending on whether the site boundary box is a square or rectangle. Finally, choose the layering design you want to use. In this case 5 Layers and press the Ok button. SADA will take some time to think it over and then 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*

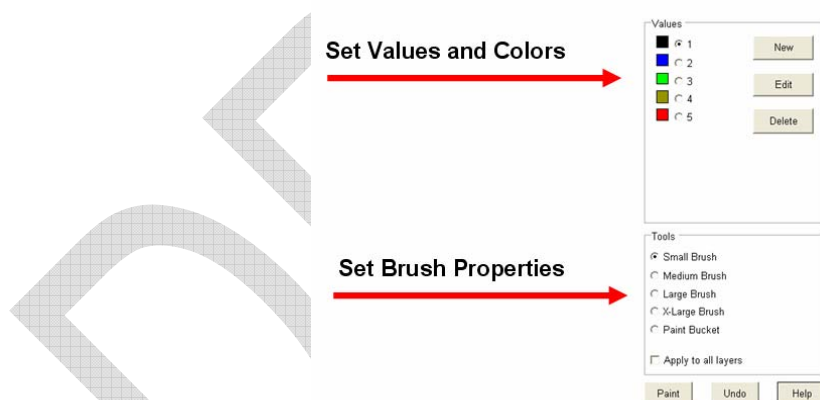


Note if you don't see the same legend as the image above, select Graphics→Legend Manager select "My Model" and press Apply. Save your file.

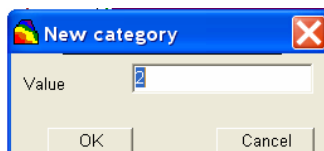
The structure of this model is permanently set in the sense that you can no longer change the grid resolution or vertical layering design. Click on the set up the site and note that the layering design box is now uneditable. We will now modify this model.

Customizing model values

All model editing features are found under the step "Edit my model". Click on this step now.



At the top of the parameter window is where you will set the values you will "paint" and the colors they will be "painted" with. 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 (blue) and click the Edit button to the right.

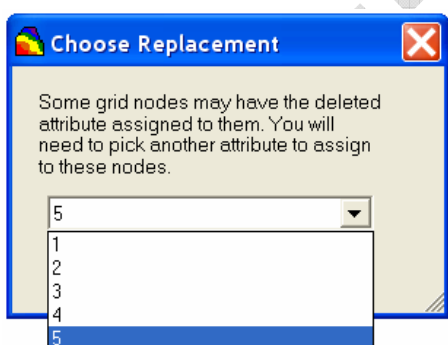


Let's enter a value of 6 here and press OK. IMPORTANT! If you have previously "painted" any 2's into your model, they all just got updated to 6's. SADA requires that all values in the model

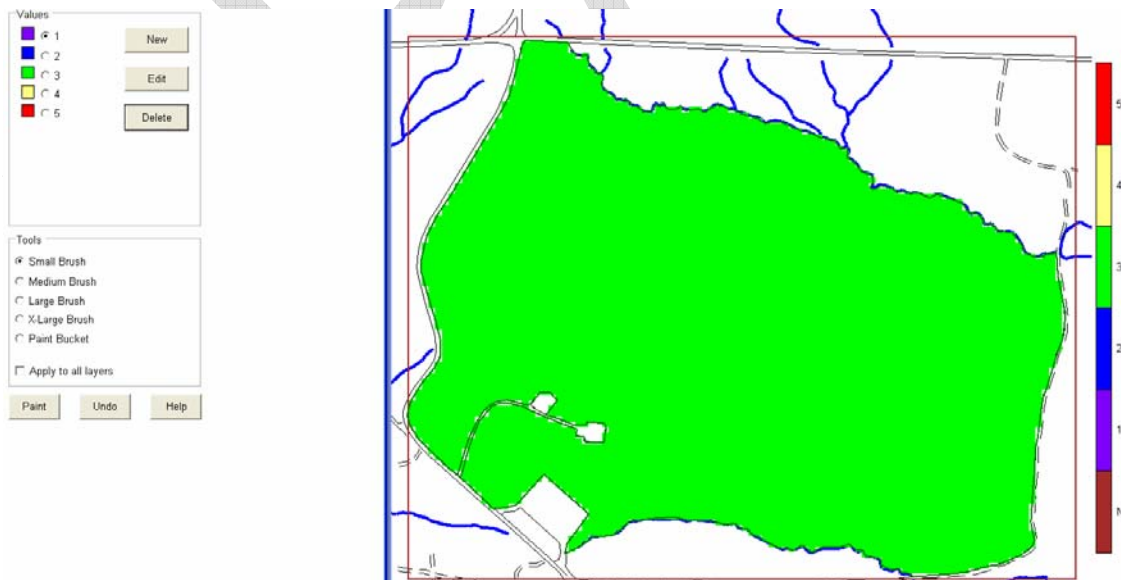
have a corresponding "paint" value in the parameter window. Notice that the values rearrange themselves in the paint selection ordering. Also recall that SADA searched the grid for all 2's and replaced them with 6's. It wouldn't have found any because we did add any yet.

Now lets 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 lets add a new value to the list. Press the New button and enter a value of 2 once again. Now we have an even 1-6. Let's now 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 replace it with. Every model cell just have a value. Select the value 5 from the drop-list and press OK. SADA will search for any cell with a value of 6 and replace it 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 $-1E+20$ 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.



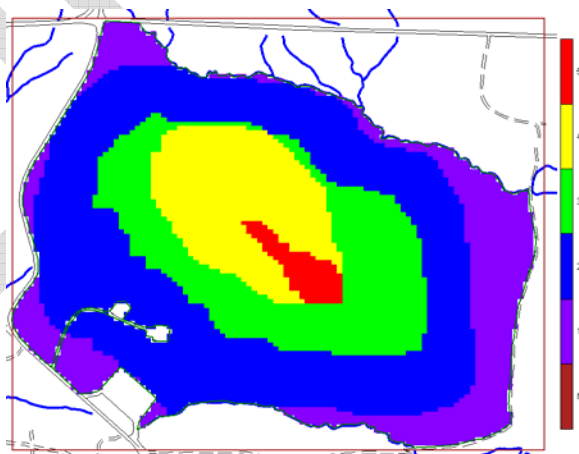
Let's move on to setting the brush size. In the lower parameter window you'll see the Tools parameter block. This includes the paint brush sizes as well as the paint bucket. These operate like most window paint tools will. We'll try each of them out. At the bottom is an Apply to all layers option. If 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, then while its deselected 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 the purple 1 value in the top parameter block and press the Paint button. 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 your 1 value, switch to the "2" value and 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.

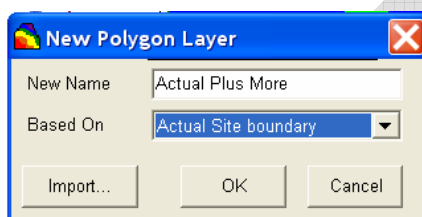


Save your SADA file and take some time to scroll through the layers. Notice that on deeper layers, the default values of 3 remain in tact. This time select the Apply to All Layers option

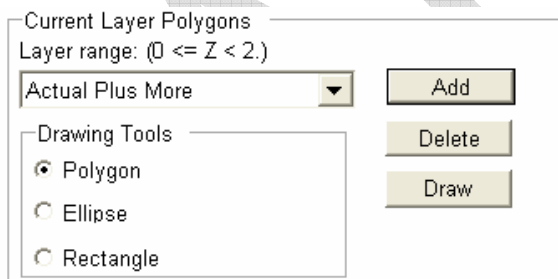
and paint a small portion of your site in any value you wish. When you are finished, press the Done button and notice the changes are found through every layer now.

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.

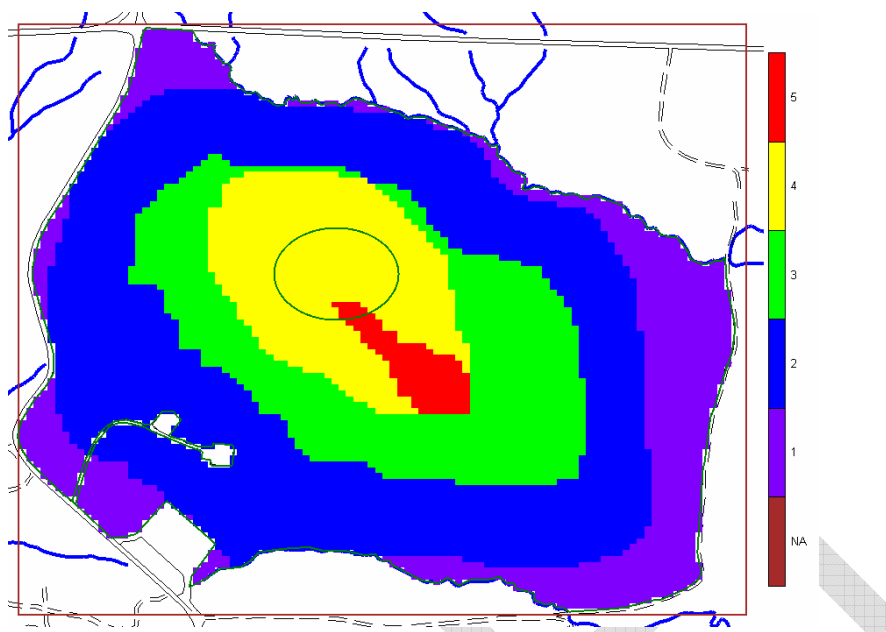
Select Set up the site step. In the Current Layer Polygon parameter block you'll see we've selected Actual Site Boundary. Click the Add button next it. Let's create a new polygon design called "Actual Plus More" that is an exact copy of the Actual Site Boundary design. In the window that appears, enter Actual Plus More into the name and select Actual Site Boundary from the "based on" drop-list.



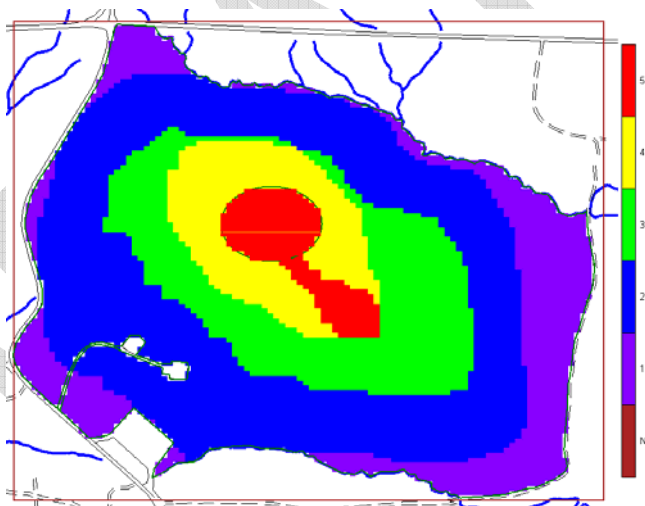
Press the Ok button. When SADA asks if you want to apply it to all layers, say YES. In the Current Layer Polygn box you should see Actual Plus More selected.



Suppose now we want to do a paint bucket dump into an elliptical area near the center of the site. Let's first draw that area by clicking the Draw button 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 "value palette" on the step Edit My Model, select the 5 value, select the paint bucket option under tools, and press the Paint button. 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 the undo button. 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 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 would include the probability that the decision criteria has been exceeded or 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 find the following step 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 by defining the site boundary box and apply any polygons, ellipses, or rectangles to more accurately set your boundaries. These have already been completed here.

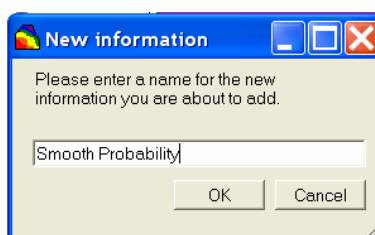
Post Processing User Models

Recall from the chapter 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.

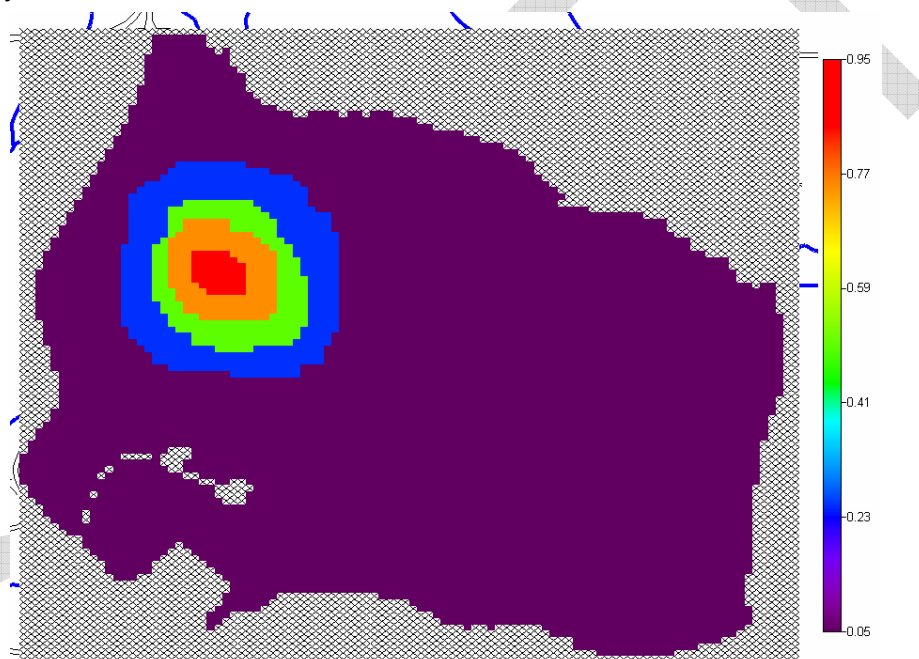
Open the file PostUseModels.sda (you can save your previous work if you like). This file has the same site boundary as the last file and a user created model called Prob > Decision. We'll use this three dimensional probabilistic model (could have just as easily been 3d) of a hypothetical source term (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 how to smooth your custom values to create a less step-wise, more continuous model.

We will first need to store a static copy of the model much like we did with the geospatial models we generated in previous chapters. Click on the step "Manage Model Results" and press the button "Save model result" in the parameter window. In the new name window enter "Smooth Probability" and press OK.



SADA will take a moment to think it over and then return control back to you. When it does, select Imported Model from the drop-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.



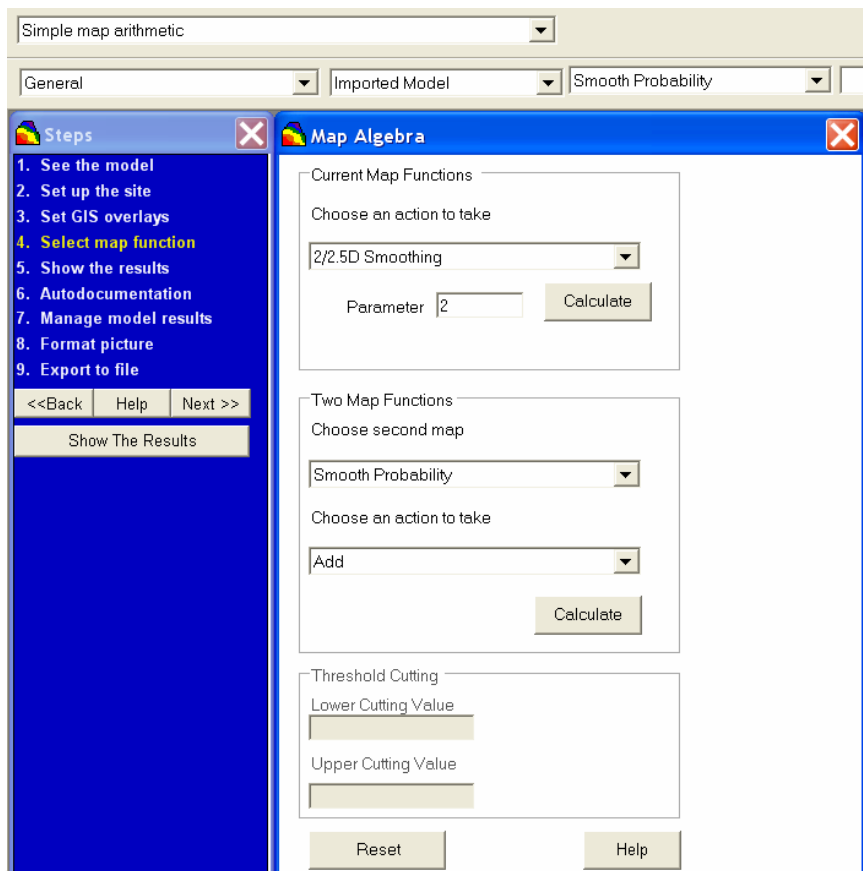
First let's turn the polygon layer back on. Click on the Set up the site step and in the current polygon layer parameter block switch to Actual Site Boundary (answer yes to apply to all layers question).

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 visit the Legend Manager chapter).

However we will not use this approach here. 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

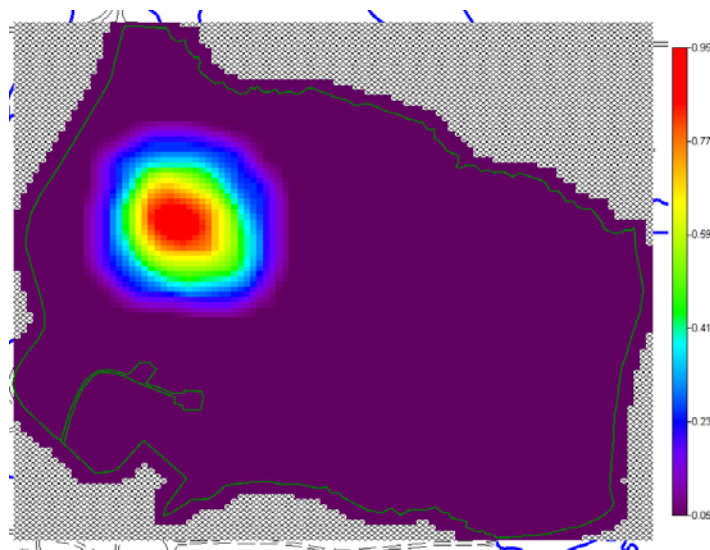
legend. Moreover we want to adopt a fixed continuous legend than extends from 0 to 1. We'll do that shortly. First let's smooth our model. You can revisit the chapter on working with models to review the rich set of post processing features there.

In the interview box select Simple Map Arithmetic and click on the step Select Map Functions.



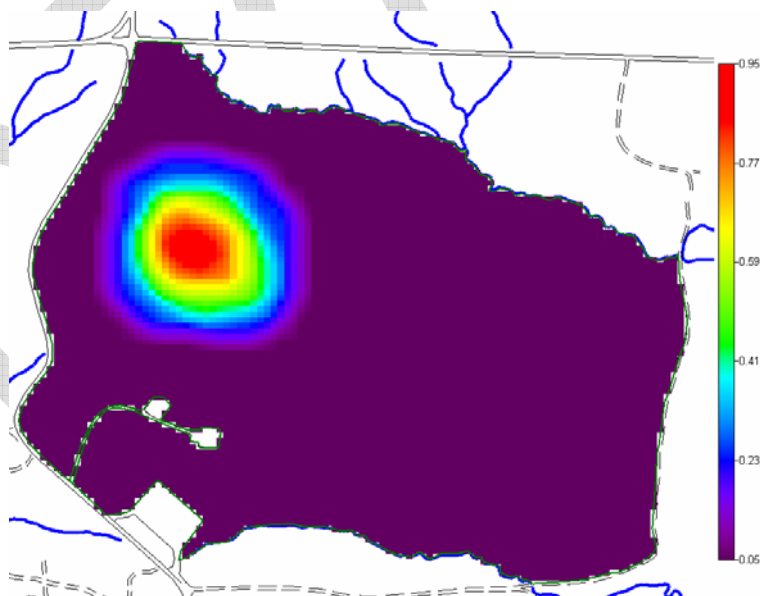
In the top parameter block choose 2/2.5D smoothing. For the parameter choose a value of 2. Recall from the Working With Models chapter that 2/2.5D smoothing only smooths by the layer (not between layers) and a parameter value of 2 means each cell value is smoothed by using the next two neighbors on all 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 the Calculate button at the top of the parameter window.

SADA responds with the following.



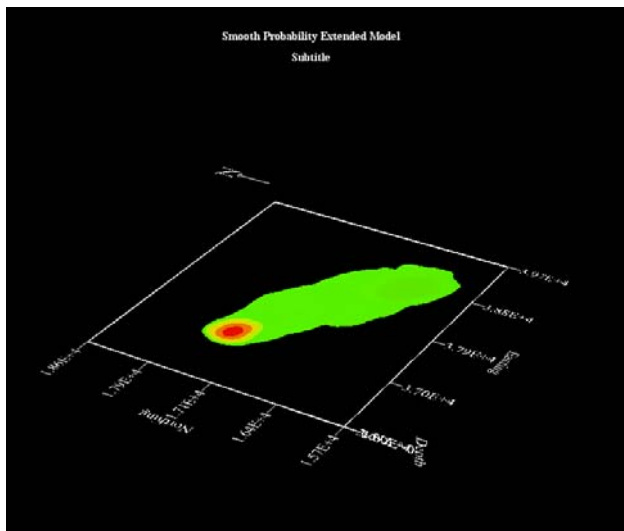
Let's now scan through the different layers and you can see that we have captured the source term as it dives down and to the southeast. Let's save this model as well. At the moment it's only a temporary result and will be lost otherwise. Click on the Manage Model Results and choose Save Model result button. Enter the name "Smoothed Model" and press OK. SADA will think it over and return control to you shortly.

When control has been returned, select Imported Model from the drop-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.

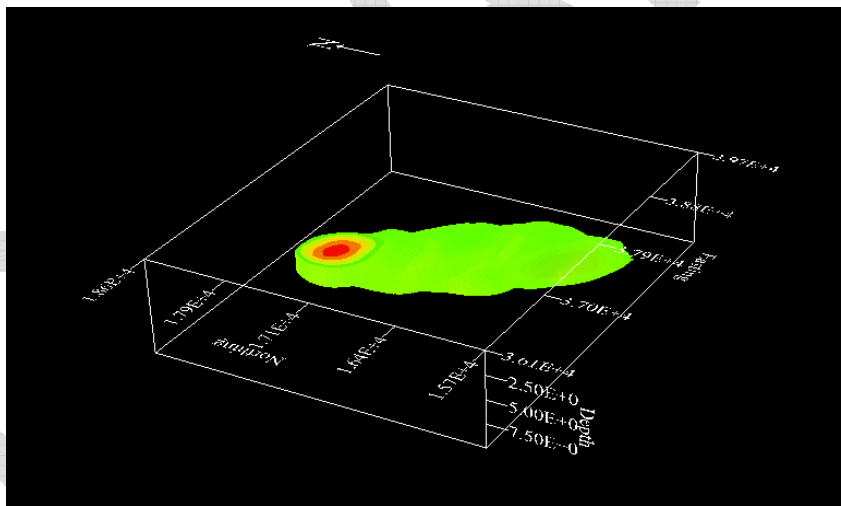


Save your SADA file (File→Save). Now let's look at it in 3d. The 3d viewer is discussed in its own chapter later, but we can use some of its features right here. Select Graphics→Show 3d View. 3d models take some time to generate so be patient. Also, some video cards have difficulty with OpenGL (in our experience) you may need to "wake up" the view by left mouse

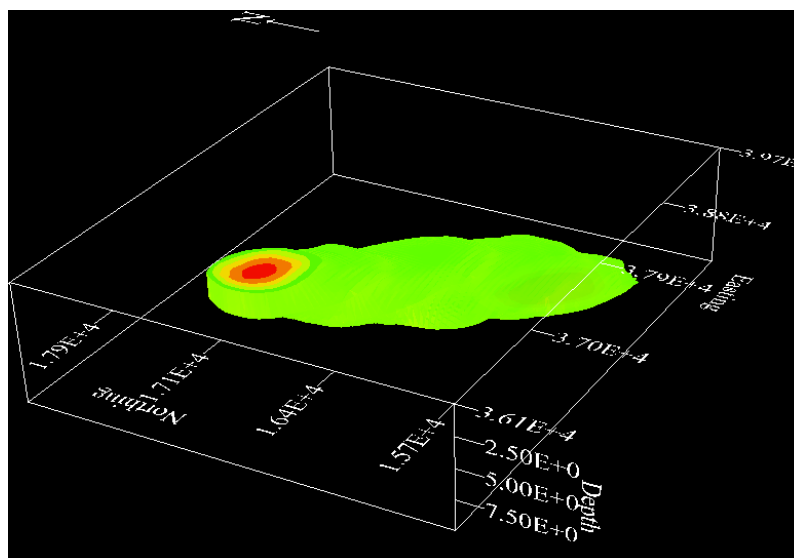
clicking in the graphics window, holding, and slightly moving your mouse. You will be presented with the following.



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 step "3D Viewer Controls" step that was added when you turned the 3d viewer. In the parameter window, click on the Scaling Tab and enter a value of 100 into the Z parameter and press the enter on your keyboard.



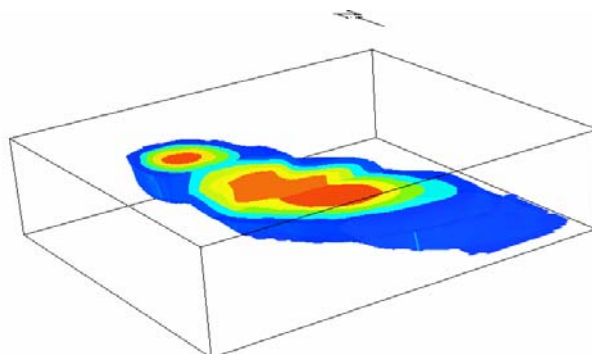
We can zoom in some by clicking on the Basic tab and pressing the Zoom In button repeatedly.



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 3d viewer chapter) 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 the Update Plot button 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 can allow 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 and the two-map function list to merge the two together. Please visit the chapter on working with models 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. A portion of that discussion is repeated here.

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 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 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 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:

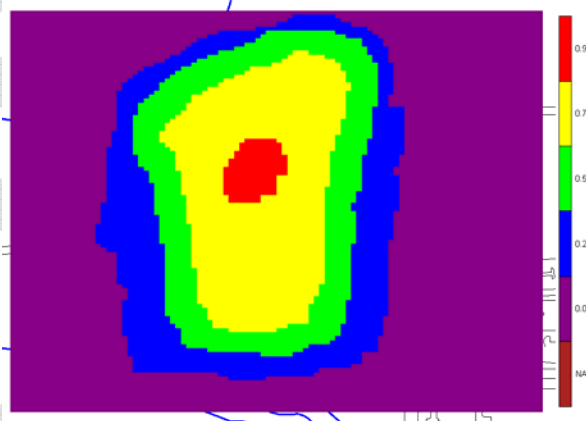
$$m^1 = E[Y(u) | I(u) = 1] \in [0,1]$$

$$m^0 = E[Y(u) | I(u) = 0] \in [0,1]$$

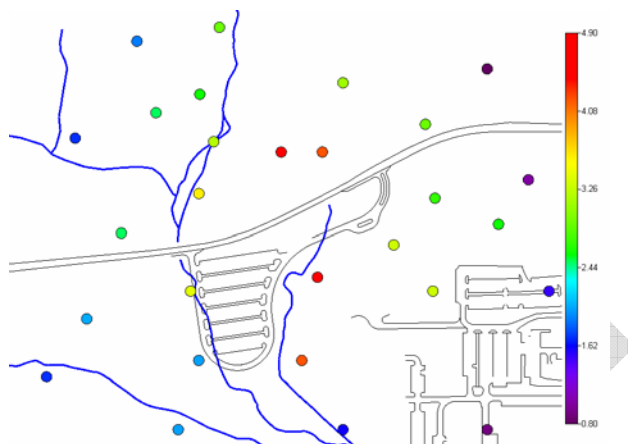
$$B = m^1 - m^0$$

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 the 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.

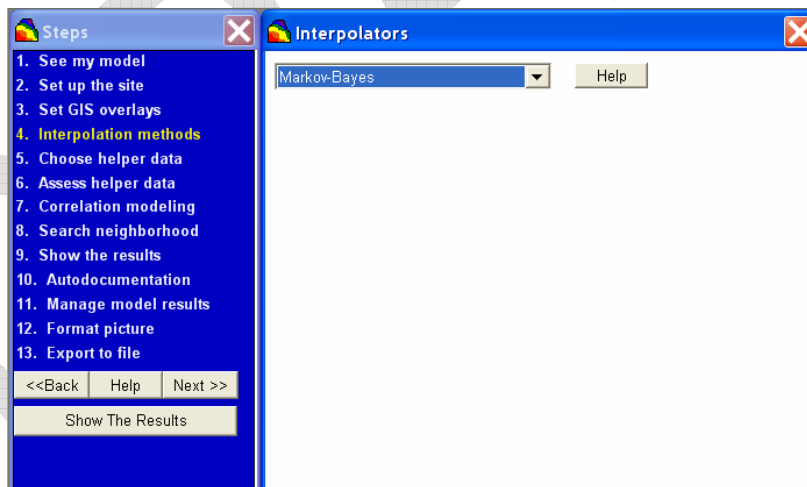


Based on this map (or not) samples of Ac-225 were collected. These can be seen by changing the data-type to soil and the dataset to Ac-225. If you do this now you should see the following.



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 then 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 10feet of slack as it is unlikely that nodes from the Prior Map and data from the Ac-225 will exactly coincide.

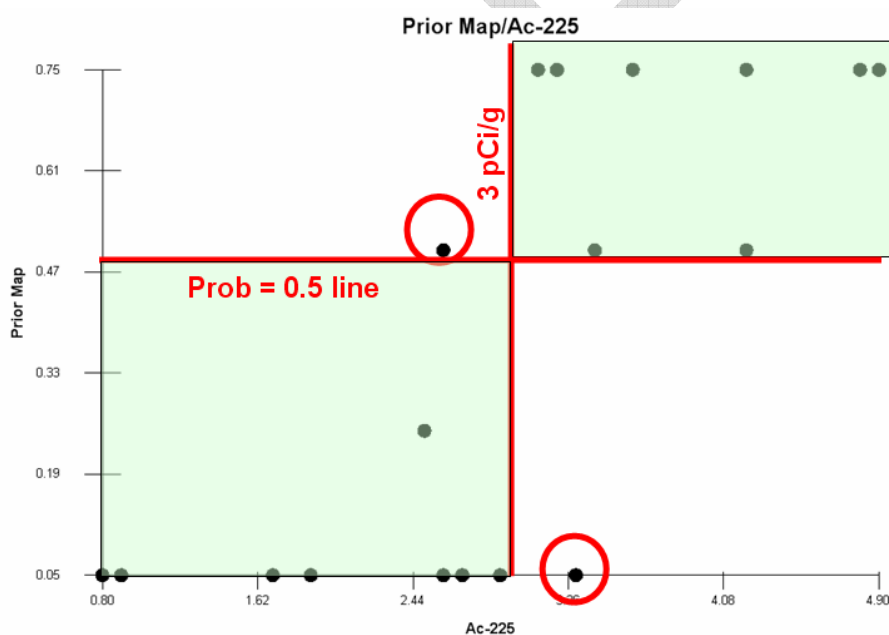
Helper Data

Soil

Ac-225

Spatial Tolerance

Next click on Assess Helper data. Under usual cokriging these 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 scatterplot which has been further annotated in the following graph.



In this graph we have Ac-225 on the horizontal axis and a red line representing our decision criteria of 3pCi/g plotted vertically from $X=3$. On the Y axis we have 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 that decision to abandon the prior 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 criteria 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 setup an indicator cutoff of 3 for you (see how to do this in the advanced geospatial modeling chapter) and developed the correlation model for you as well.

Choose Data

First Variable

Second Variable

Data Transforms

Explore Experimental Semi-variography

Previous Results

Use Direction

Name	Major	Minor
Caption	Criteria = 3	Def...
Lag Number	20	20
Lag Distance	58	58
Lag Tol	58	58
Angle	0	0
Tol	90	90
Band	2302.75	230...
Dip	0	0
ZTol	90	90
ZBand	1	1

Model Semi-variography Values

Model

Major Range	977.35067383...	
Minor Range	977.35067383...	
Angle	0	
Contribution	0.3219814241...	
Z Angle	0	
Z Range	1	
Rotation	0	

Nugget

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.

Search Radius	
Search Radius in the Major Direction	<input type="text" value="1000"/>
Search Radius in the Minor Direction	<input type="text" value="1000"/>
Vertical Search Radius	<input type="text" value="1"/>
Search Angles	
Horizontal Search Angle	<input type="text" value="0"/>
Vertical Search Angle	<input type="text" value="0"/>
Rotation About Vertical Axis (advanced)	<input type="text" value="0"/>
Geospatial Model Parameters	
Minimum Number of Sampled Data	<input type="text" value="2"/>
Maximum Number of Sampled Data	<input type="text" value="20"/>
Minimum Number of Helper Data	<input type="text" value="1"/>
Maximum Number of Helper Data	<input type="text" value="4"/>
Number of Simulated Nodes	<input type="text"/>

Help Default

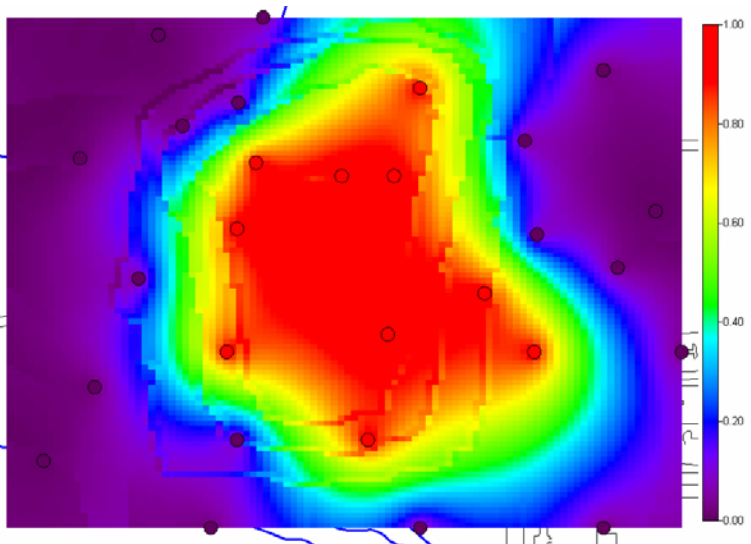
Defining the search neighborhood follows the same principles outlined in the advanced geospatial modeling chapters so we won't repeat that here.

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.

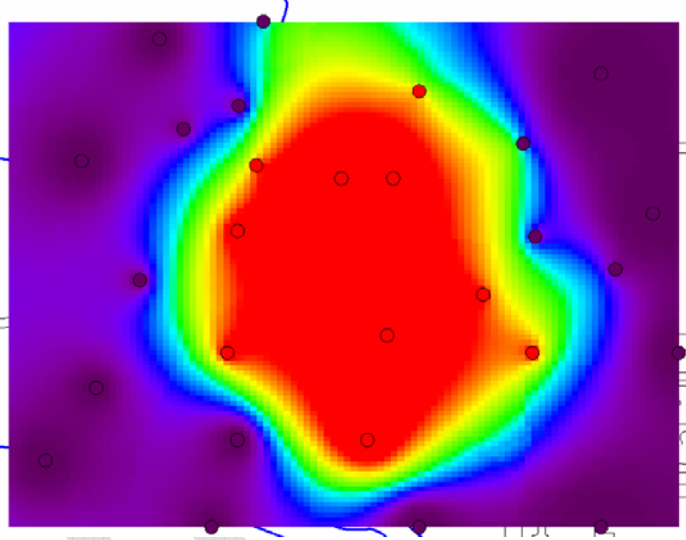
User Defined Decision Criteria
✖

Choose a IK threshold values.

Depending on what indicator threshold modeling activities you've done, you may see more than just one option here. But 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 on decision analysis).



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 to find yourself in. 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.

DRAFT