X

Sampling and interpolation

Exercise 1: Create surfaces with interpolation
Estimated time: 30 minutes (excludes challenge step)

Exercise 1: Create surfaces with interpolation

Estimated time: 30 minutes (excludes challenge step)

This exercise introduces you to some surface interpolation techniques available in ArcGIS. First, you will create surfaces using different interpolation methods; then you will validate a surface by removing some of its sample points and running the interpolation method again.

In this exercise, you will:

- Use Natural Neighbor and Spline to interpolate surfaces.
- Validate surfaces created by interpolation.

Step 1: Set up the analysis environment

Before creating any surfaces, you will set some geoprocessing environment parameters to make your work more efficient.

☐ Start ArcMap and open **..\SAMI\CreateSurfaces.mxd.**.

First, you will set the default geodatabase to Surfaces.gdb.

☐ Click the Catalog window and navigate to the SAMI folder connection.

☐ Right-click the Surfaces.gdb, then click Make Default Geodatabase.

There are two layers in the Interpolation data frame that you'll be working with in this exercise:

- **Snow points**: Point features of snow depth samples in inches (liquid amount)
- **Landmask**: Raster landmask/watermask for the study area

☐ From the Geoprocessing menu, choose Environments.

☐ Click the Processing Extent heading.

☐ For Extent, choose Same as layer landmask from the drop-down list.

☐ Under Raster Analysis, set Cell Size to As specified below, then verify or type **30**.

☐ Set Mask to Landmask and click OK.

> ***Note:*** Remember that the environment settings you set here apply to the entire ArcMap application. You can always override these settings by making tool-level or model-level settings.

Now you are ready to create surfaces.

Step 2: Interpolate using Natural Neighbor

☐ In the Catalog window, expand Toolboxes, then System Toolboxes.

☐ Expand the Spatial Analyst Tools toolbox and its Interpolation toolset.

You can right-click any tool and choose Help to learn more. In this step, you will run the Natural Neighbor tool.

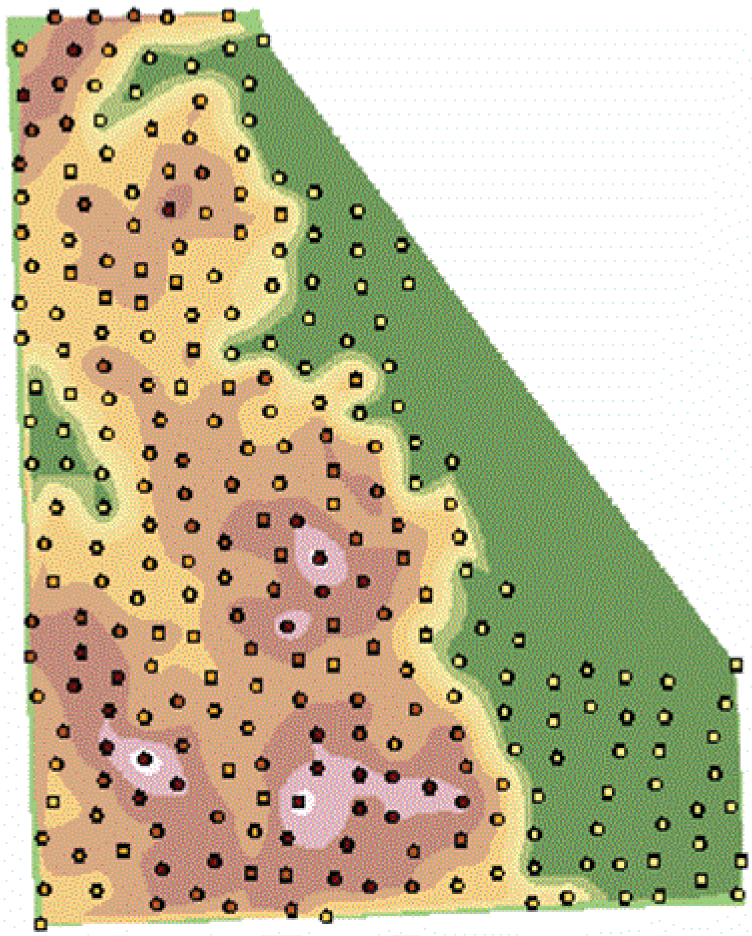
The Natural Neighbor interpolation method uses Thiessen polygons to find the "natural neighbors," or closest subset of input samples for a cell, and uses an area-weighted (not distance-weighted) method to weight the samples. It is also known as the "area-stealing" interpolation. Its basic properties are that it's local, using only a subset of samples that surround a cell, and that its interpolated heights are guaranteed to be within the range of the samples used.

This interpolator has a significant advantage over others because it can handle an extremely large number of samples that might cause other interpolators to fail.

☐ Double-click the Natural Neighbor interpolation tool.

☐ Run the tool with the following parameters:

- Input point features: **Snow points**
- Z value field: **SNOWDEPTH**
- Output raster: **natural1**



The Natural Neighbor interpolator only interpolates values for the cells that fall within the convex hull of the sample points (i.e., the polygon made up by connecting all the outermost points). For this reason, it doesn't honor the analysis mask. You'll use a geoprocessing tool to extract cells where natural1 and Landmask overlap, then manually assign symbology.

☐ In the Catalog window, navigate to System Toolboxes > Spatial Analyst Tools > Extraction > Extract By Mask.

☐ Double-click the Extract By Mask tool to open it.

☐ Run the Extract By Mask tool with the following parameters:

- Input raster: **natural1**
- Input raster or feature mask data: **Landmask**
- Output raster: **natural2**

☐ Turn off natural1 and Snow Points and collapse their legends.

The output from the Natural Neighbor interpolator is now clipped to the mask you set earlier in the environments. The natural2 surface draws with stretched renderer; you will change its symbology to match natural1.

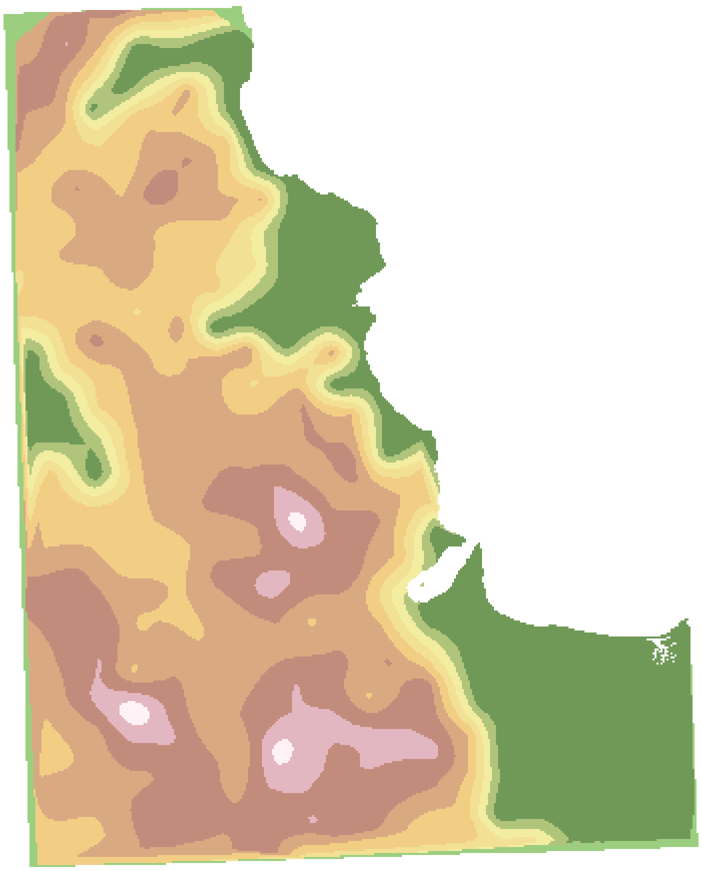
☐ Open the symbology properties for natural2 and choose Classified from the upper-left.

☐ Click the Import button.

☐ On the Import Symbology dialog box, choose natural1 from the drop-down list.

☐ Click OK.

☐ Click OK to close the Layer Properties dialog box.



This new surface is a result of the interpolation method estimating unknown values by using the values of the sample points in the Snow Points layer. Later, you will take a closer look at this surface and validate it to see how well the interpolator estimated the values.

Step 3: Interpolate using Spline

Next, you'll use the Spline interpolation method. This method uses a minimum-curvature spline interpolation. It can predict hills and valleys, and the surface it creates passes through the samples.

First you will run the Spline tool with the Regularized method. In this method, higher values used for the Weight parameter produce smoother surfaces. The Weight values are typically set between 0 and 0.5. The default is 0.1.

☐ In the Interpolation toolset, open the Spline tool.

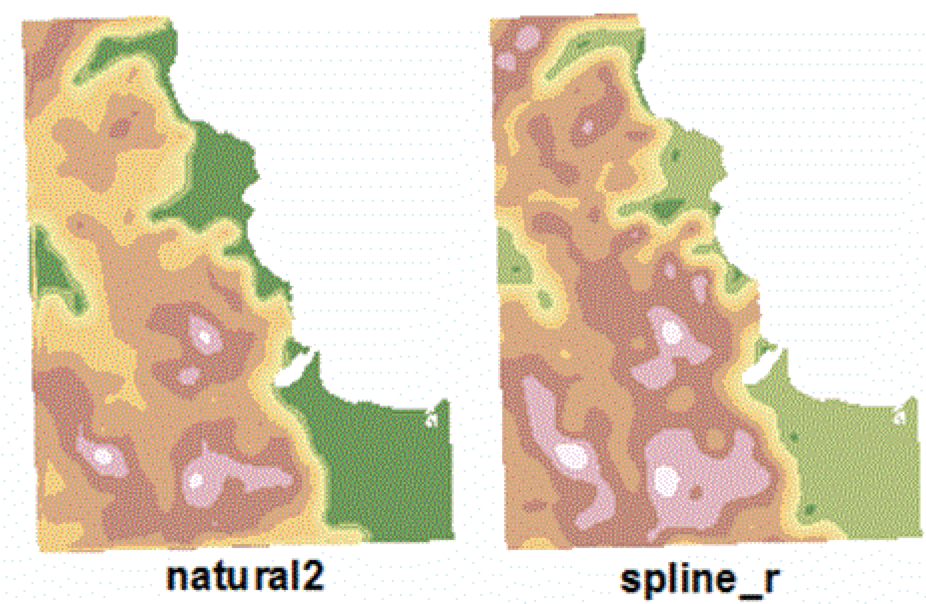
☐ Run the Spline interpolation tool with the following parameters:

- Input point features: **Snow points**
- Z value field: **SNOWDEPTH**
- Output raster: **spline_r**
- Spline type: **REGULARIZED**

☐ If necessary, turn on the Effects toolbar and dock it.

☐ On the Effects toolbar, select the spline_r layer.

☐ Click the Flicker Layer button  to compare spline_r with natural2.



☐ Click the Flicker Layer button again to turn it off.

In the legend for spline_r, notice that the surface values exceed the sample range of 0 to 28 inches. Spline has predicted hills and valleys. This interpolation method also honors the analysis mask.

☐ Turn off natural2 and collapse the legend for spline_r.

Next you will run the Spline tool using the Tension method. The Tension method tunes the stiffness of the surface according to the character of the modeled phenomenon. It creates a coarser surface with values more closely constrained by the sample data range.

A value of 0 for the Weight results in the smoothest surface (like fitting rubber plates), while higher Weight values result in a coarser surface (like fitting thin metal plates to the samples). Typical Weight values are 0, 1, 5, and 10.

You will run the Spline tool again, but this time you will run it from the geoprocessing Results window. This allows you to modify particular parameters while leaving others intact. It is faster than opening the tool and resetting all the parameters when you really only need to change one or two.

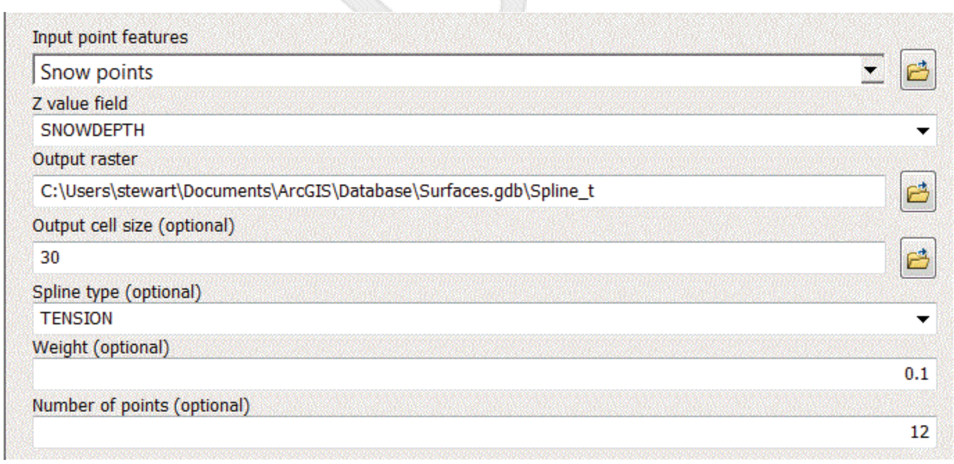
☐ From the Geoprocessing menu, choose Results.

☐ Expand Current Session, right-click Spline, and choose Open.

This dialog box contains the parameters of the tool's last run. You will modify the output raster name and the Spline type.

☐ Change the Output raster name to spline_t and the Spline type to TENSION.

☐ Your dialog box should match the following example:



☐ Click OK, then close the Results window.

☐ If necessary, open the Effects toolbar, then select the spline_t layer.

☐ Click the Flicker button to compare spline_t with spline_r.

Note in the legend that spline_t has a narrower range of values than spline_r. The Tension method created a tighter surface with shorter hills and shallower valleys.

As you can see, the two interpolation methods yield quite different results. There are many more methods to choose from in the Spatial Analyst toolbox, as well as within the Geostatistical Analyst extension. For example, the Geostatistical Analyst extension has a wizard that allows you to perform one of the most powerful interpolation methods: kriging. It also allows you to do variogram mapping and true validation of your surface. The validation you do in the next step is a much simpler way to see how well the values were predicted by the interpolator. Consult the Help documentation for more information about kriging and the Geostatistical Analyst extension.

Step 4: Evaluate interpolation results

It can be difficult to evaluate the validity of an interpolated surface because there is no correct surface with which to compare your results. Often, the only information available to you is the input sample set. This step allows you to explore some techniques to test the accuracy of your interpolation.

One way you can validate an interpolated surface is to remove some of the sample points, interpolate the surface using the remaining sample points, and then check to see how accurately the interpolator predicted the value of each of the known sample points you removed. Organizations that create critically important surfaces, like those performing global warming studies, follow this methodology exhaustively.

In this step, you will apply a simplified version of this methodology. You'll remove several samples, run each of the interpolators against the remaining samples, then check to see which interpolator came closest to predicting the value of the missing sample.

☐ From the Insert menu, choose Data Frame and then name it Evaluating Interpolation.

☐ From the Interpolation data frame, copy and paste the following layers into the new data frame:

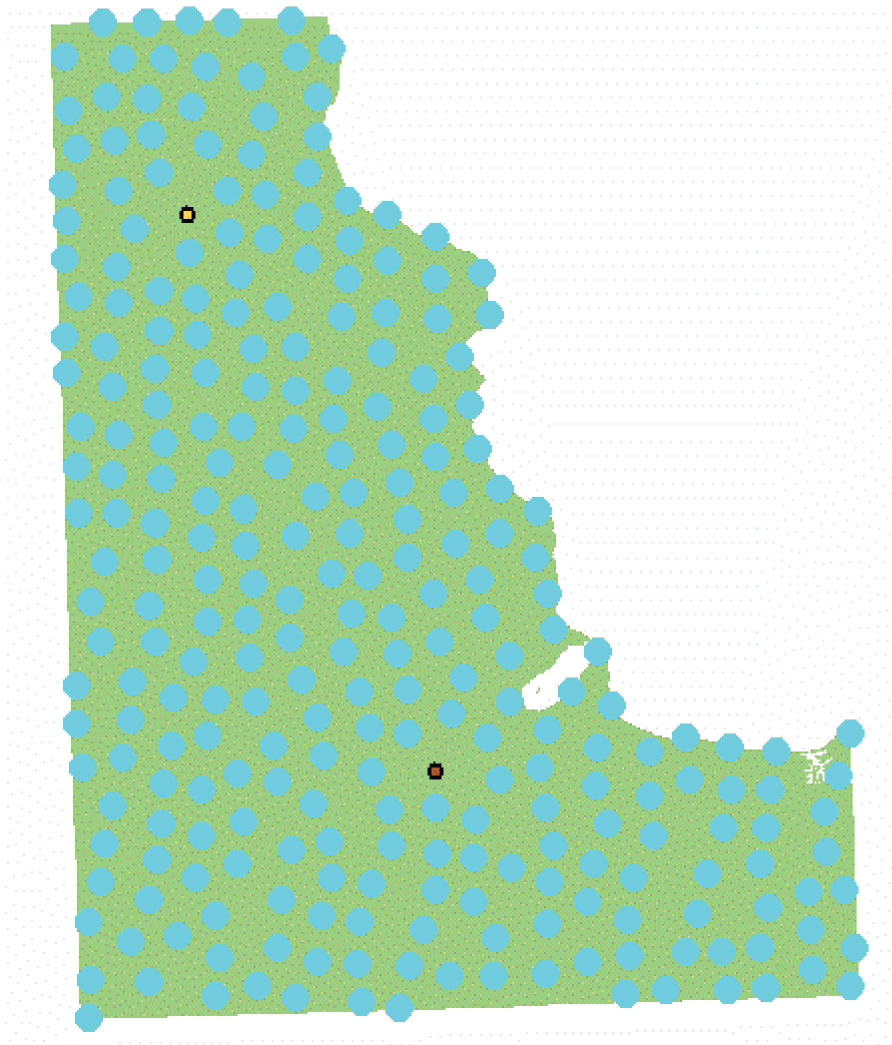
- Landmask
- Snow points

☐ Turn on Snow points and select all of the points in the layer.

☐ Click the Select Features tool .

☐ Hold down the Shift key on your keyboard and click two points to unselect them. (Note: Your result may vary from the below because you may choose to unselect any two points as long as they are fairly distant from each other.)

☐ For later use, be sure to remember which two points you unselect.



These two points will be excluded from the analysis because only the selected features will be used by the geoprocessing tools.

☐ Open the Geoprocessing Results window. (*Hint*: From the Geoprocessing menu, choose Results.)

☐ If necessary, expand Current Session.

☐ Right-click Natural Neighbor and choose Open.

☐ Change the Output raster name to **NatEval** and leave all other parameters the same.

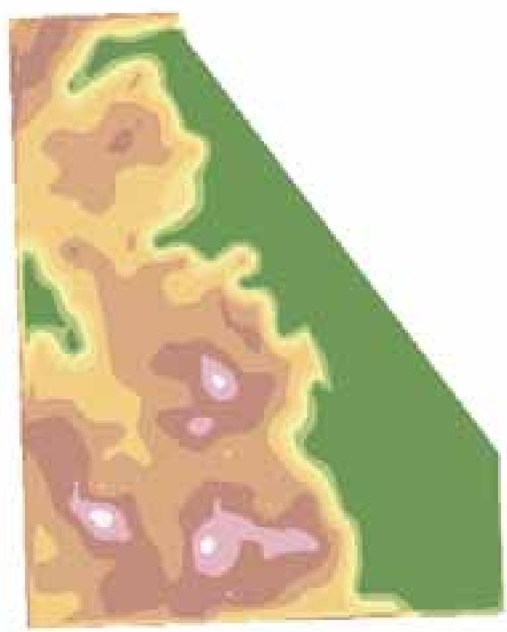
☐ Click OK to run the tool.

☐ Using the skills you've learned, open the tool dialog box for the first Spline tool you ran (i.e., the one right above Extract by Mask in Geoprocessing Results) and change its Output raster name to **SplineEval**, then run the tool.

☐ Close the Results window.

Notice that once again, the Natural Neighbor tool didn't honor the processing mask.

☐ Turn off SplineEval so you can just see NatEval.



Earlier, you fixed this with the Extract by Mask tool. Since the objective of this step is simply to evaluate the interpolation, and you won't be using this surface in any type of map or display, you can leave it as is.

Now you'll compare the results.

☐ Make sure that SnowPoints, NatEvalMask, and SplineEval are the only visible layers.

☐ Make sure Snow points is drawing on top.

☐ Click the Identify tool and click anywhere in the map to open the Identify window.

☐ From its drop-down list, set Identify to <Visible layers>.

☐ Click the first snow point you unselected to get its SNOWDEPTH value, then record all three SNOWDEPTH values (from Snow points, NatEvalMask, and SplineEval) in the following table.

Point	Original value	Natural Neighbor value	Spline value
First			
Second			

☐ Repeat this process for the second point you unselected.

Which interpolation method most accurately predicted the snow depth values along the surface?

It would be good practice to identify the interpolator that best estimates the sample, and also to compare that to which interpolator works best with the given data. In this case, removing only two points may not give you enough information to make a decision, so you would need to continue removing points from the interpolation. After going through this step, you have the skills to continue this process.

☐ Close the Identify window.

☐ Once you have explored the interpolated values and surfaces, save your map but leave it open if you want to continue to the challenge step that's next.

Step 5: Challenge ~ Explore California ozone levels with Geostatistical Analyst

Estimated time: 25 minutes

In the first part of this exercise, you used the tools in ArcGIS Spatial Analyst to interpolate snow depth surfaces from sample points. Then you evaluated the accuracy of the interpolation method by selectively removing sample points and observing how well the interpolation predicted the value at the missing sample point location.

While the Spatial Analyst tools can be very effective, users who perform a great deal of interpolation sometimes desire an even larger degree of control over the process and a wider range of tools. This is where Geostatistical Analyst comes in. Geostatistical Analyst is an ArcGIS Desktop extension specifically for interpolating surfaces from sample data. In this challenge, you'll take a quick tour through Geostatistical Analyst to see some of its functionality.

☐ In ArcMap, right-click the Geostatistical Analyst data frame and click Activate.

This data frame has two layers—an outline of the state of California and the locations of ozone concentration samples. The ozone concentrations are the points you will explore with Geostatistical Analyst.

☐ In ArcMap, click Customize > Toolbars > Geostatistical Analyst.

The Geostatistical Analyst toolbar is relatively small—only one button and one drop-down menu—but don't let its size fool you. It's very powerful. You'll begin by using the extension's data exploration tools.

☐ On the Geostatistical Analyst toolbar, click the Geostatistical Analyst menu, point to Explore Data, and choose Histogram.

This opens a histogram, or frequency chart, of the data. Understanding the range of values in your sample data can be useful in applying the best interpolation. This histogram not only displays the range of values for an attribute, it also shows their location.

☐ At the bottom of the histogram, make sure the Layer drop-down list is set to Ozone Levels.

☐ From the Attribute drop-down list, choose OZONE.

☐ On the histogram, click the leftmost bar (the one with ozone values ranging from 0.46 to 0.59).

☐ Move the histogram away from the map display.

The sample points that lie within this range (the smallest values) are highlighted on the map.

☐ Click the rightmost bar.

The sample points with the highest values are in the area east of Los Angeles (a region known locally as the Inland Empire).

☐ Click an empty part of the graph to clear the selected points and close the histogram.

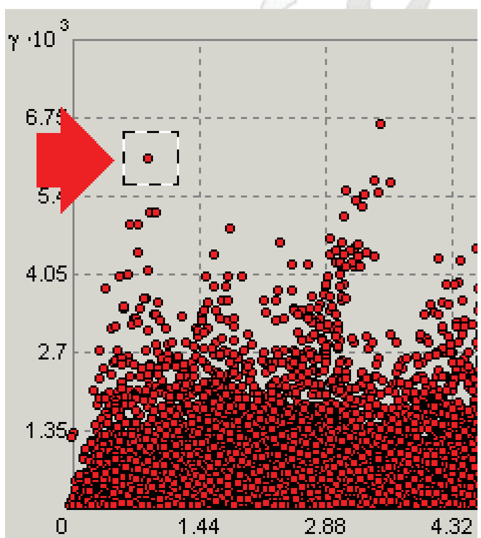
You will continue to explore the data by looking at a semivariogram.

☐ On the Geostatistical Analyst toolbar, click the Geostatistical Analyst menu, then Explore Data > Semivariogram/Covariance Cloud.

A semivariogram of the data opens. This graph can be used to examine the local characteristics of spatial autocorrelation within a dataset and to look for local outliers. Each red dot in this graph is a comparison of two ozone level sample points.

☐ At the bottom of the dialog box, make sure the Layer drop-down list is set to Ozone Levels. Then, from the Attribute drop-down list, choose OZONE.

☐ Use your mouse pointer to drag a box around the red dot shown in the following graphic.



On the graph, the dot is highlighted in blue. On the map, the two locations used to compute that red dot are highlighted and linked.

On the map, you can see that the two locations are fairly close together (which is why the highlighted dot is near zero on the semivariogram x-axis). The ozone values, however, are far apart.

☐ Leave the semivariogram open, but move it away from the map display.

☐ Zoom to and then identify the two linked points.

The Identify Results window displays the values. One has a value of 0.06409, the other a value of 0.1736.

☐ Close the Identify Results window and zoom to full extent.

☐ On the semivariogram, select some other red dots and observe the effect on the ozone level points in the map.

☐ When you are finished exploring the semivariogram, close it.

The semivariogram is a good tool for finding local outliers—locations with values that are very different from their neighbors that may or may not be outside the dataset's range of values.

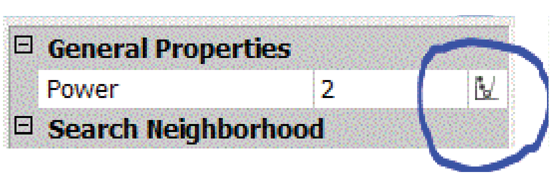
Now that you have seen some of the data exploration tools available in Geostatistical Analyst, you will examine how to interpolate a surface from sample points.

☐ From the Geostatistical Analyst menu, choose Geostatistical Wizard.

☐ On the first panel of the wizard:

- Make sure the input data is **Ozone Levels**, then set the attribute to **OZONE**.
- In the Methods box on the left, choose **Inverse Distance Weighting**.
- Click Next.

☐ On the second panel, in the upper left corner, click Optimize Power Value.



Notice that the value for the Power property changes from 2 to 3.8516.

☐ Use the defaults for the other settings.

☐ Click Finish, then click OK in the Output Layer Information dialog box.

The Inverse Distance Weighting layer is added to the map. Notice that it is clipped to the boundary of the the California layer. This is due to the extent properties of the data frame, not to Geostatistical Analyst. The actual output is really a rectangle defined by the bounding coordinates of the ca_ozone_pts layer.

> ***Note:*** One important difference between the output from Geostatistical Analyst and Spatial Analyst is that the Geostatistical Analyst output is just a layer derived on-the-fly from the sample points. If you close this map without saving, the layer will be gone. If you want to preserve the results of your interpolation for use in other maps, you can save the layer to a layer file, or export it to either vector or raster data.

Now that you have produced an IDW prediction map, you will produce another one using the Ordinary Kriging method.

☐ From the Geostatistical Analyst menu, choose Geostatistical Wizard.

☐ On the first panel:

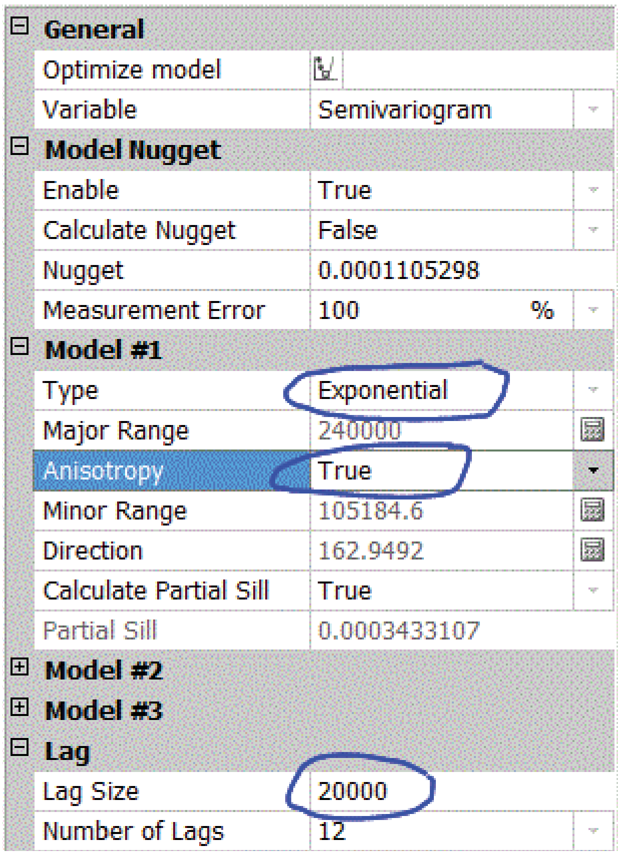
- Make sure the input data is **Ozone Levels** and set the attribute to **OZONE**.
- In the Methods box on the left, choose **Kriging/CoKriging**.
- Click Next.

☐ On the second panel, verify that the Kriging Type is set to Ordinary, then click Next.

Notice that the Kriging interpolation provides its own semivariogram to explore the data. Also notice the number and variety of parameters that can be defined for the interpolation. Kriging is a very robust interpolation method that requires some study of statistics and knowledge of the data to use it properly.

☐ On the third panel:

- Change the lag size to **20000.**
- Under Model #1:
  - From the Type drop-down list, choose **Exponential**.
  - For Anisotropy, choose **True** from the drop-down list.



The semivariogram and the semivariogram surface update in response to the changes.

☐ Click Next.

The fourth panel shows you the sample data points that are used to calculate the value of an unknown location (marked by the intersecting lines).

☐ On the Preview pane, click in a few different places.

Notice that the neighborhood moves and new sample points are selected.

☐ Click Next.

The final panel in the wizard allows you to perform diagnostics on the interpolation model.

The chart shows measured values for the data points on the x-axis and the values that the model predicts for these same locations on the y-axis. This technique, called cross-validation, tests the quality of the model. The closer the points are to the dashed gray line, the better. (The gray line represents identical values for predicted and actual values.) This is basically a quicker and more sophisticated version of the same method you used to evaluate the interpolation during a prior step in this exercise.

A summary measurement of the difference between the predicted and actual values is the root mean square error. The smaller the root mean square (RMS), the better.

☐ Click Finish, then click OK in the Output Layer Information dialog box.

☐ In the table of contents, rename the top layer **KrigingExponential**.

☐ Turn the KrigingExponential layer off and on to compare it with the Inverse Distance Weighted layer underneath.

It is obvious that the two interpolation models (Kriging and Inverse Distance Weighting) produced different results. Now you are ready to compare the models—the last stage of the analysis.

☐ In the table of contents, right-click the KrigingExponential layer, then click Compare.

The Cross Validation Comparison dialog box opens.

A cross validation chart for KrigingExponential appears on the left of the dialog box. A chart for the Inverse Distance Weighting layer is displayed on the right.

A good way to compare the models statistically is with the Root Mean Square number at the bottom of the dialog box. The smaller the root mean square, the closer the model comes, on average, to predicting measured values that were removed from the analysis. The model with the lowest RMS is probably the one you want to keep.

☐ When you're finished, close the Cross Validation Comparison dialog box.

☐ Save the map document and exit ArcMap.

You now have some idea of what Geostatistical Analyst can do. You can explore a dataset with histograms and semivariograms, fit the data to different models, adjust the model parameters, check the model with cross-validation, and compare models to each other using the root mean square measure.

In this challenge step, you were introduced to a small subset of the functionality of the extension. For more information about Geostatistical Analyst, see the ArcGIS Desktop Help and Esri training options at www.training.esri.com.

Conclusion

Interpolating a surface from point data samples is one of the most common methods for creating a surface. To properly interpolate a surface from such data, however, you need to start with a good set of sample data that accurately represents the entire study area. You also need to choose an appropriate interpolation method.

As you have learned, there are many interpolation methods to choose from. The idea is to use a method that best suits your data and your needs, and that creates an accurate surface from your sample points. To test the accuracy of your surface, you should always validate the surface after running an interpolation method.

Once you create surface rasters, you can use them with other raster analysis tools within ArcGIS Spatial Analyst. If you are not satisfied with the results when validating your surface, you should try another interpolation method. As the analyst, you are responsible for choosing a method that is suitable for both the data and the application. Researching the interpolation methods in the ArcGIS Desktop Help system will help you make an informed decision about which one to use.