Principles of Interpolation Analysis

Prepared by: Asst. Prof. Dr. Huner Khayyat

Salahaddin University - Erbil

GIS & Spatial Analysis Course

College of Science – Department of Earth Science and Petroleum

MSc. Second Semester

Academic Year: 2023 - 2024

Table of Contents

Understanding interpolation analysis

Interpolation predicts values for cells in a raster from a limited number of sample data points. It can be used to predict unknown values for any geographic point data, such as elevation, rainfall, chemical concentrations, and noise levels.

Why interpolate to raster?

The assumption that makes interpolation a viable option is that spatially distributed objects are spatially correlated; in other words, things that are close together tend to have similar characteristics. For instance, if it is raining on one side of the street, you can predict with a high level of confidence that it is raining on the other side of the street. You would be less certain if it was raining across town and less confident still about the state of the weather in the next county.

Using the above analogy, it is easy to see that the values of points close to sampled points are more likely to be similar than those that are farther apart. This is the basis of interpolation. A typical use for point interpolation is to create an elevation surface from a set of sample measurements. Geostatistical Analyst also provides and extensive collection of interpolation methods.

Examples of interpolation applications

Some typical examples of applications for the interpolation tools follow. The accompanying illustrations will show the distribution and values of sample points and the raster generated from them.

Interpolating a rainfall surface

The input here is a point dataset of known rainfall-level values, shown by the illustration on the left. The illustration on the right shows a raster interpolated from these points. The unknown values are predicted with a mathematical formula that uses the values of nearby known points.
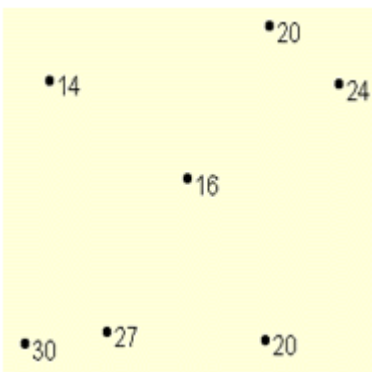
Input rainfall point data.

Interpolated rainfall surface

Interpolating an elevation surface

A typical use for point interpolation is to create an elevation surface from a set of sample measurements.

In the following graphic, each symbol in the point layer represents a location where the elevation has been measured. By interpolating, the values for each cell between these input points will be predicted.

Input elevation point data.
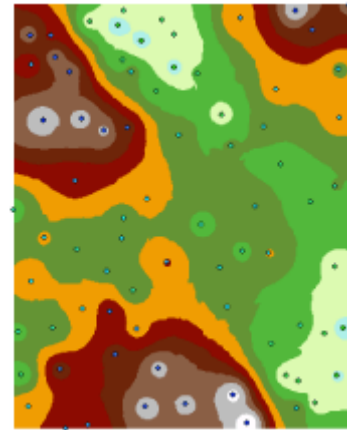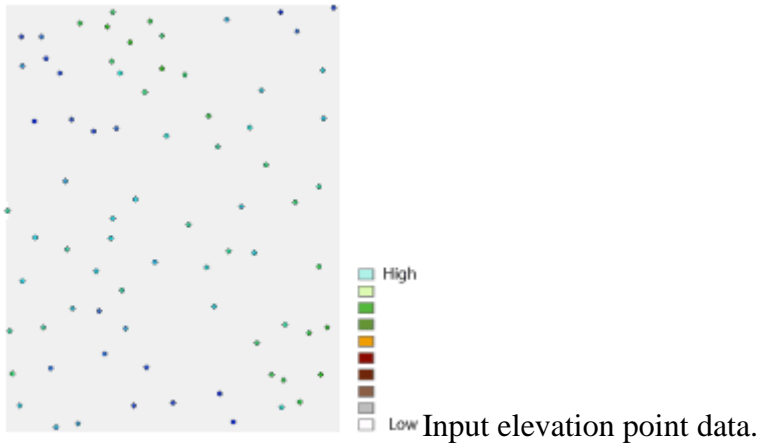
Interpolated elevation surface

Interpolating a concentration surface

In the example below, the interpolation tools were used to study the correlation of the ozone concentration on lung disease in California. The image on the left shows the locations of the ozone monitoring stations. The image on the right displays the interpolated surface, providing predictions for each location in California. The surface was derived using kri

Point locations of ozone monitoring stations
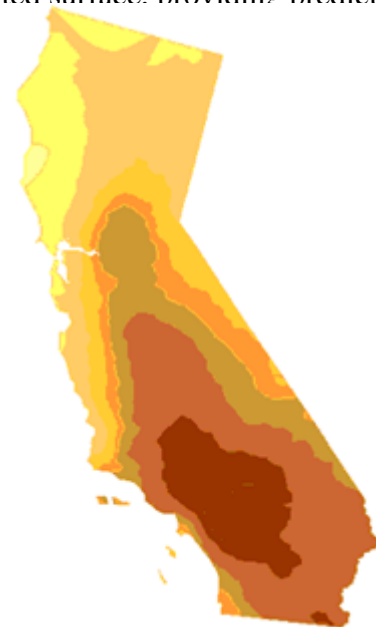
Interpolated prediction surface

Comparing interpolation methods

Interpolation predicts values for cells in a raster from a limited number of sample data points. It can be used to predict unknown values for any geographic point data, such as elevation, rainfall, chemical concentrations, noise levels, and so on.

The available interpolation methods are listed below.

IDW

The [IDW](#) (Inverse Distance Weighted) tool uses a method of interpolation that estimates cell values by averaging the values of sample data points in the neighborhood of each processing cell. The closer a point is to the center of the cell being estimated, the more influence, or weight, it has in the averaging process.

Kriging

[Kriging](#) is an advanced geostatistical procedure that generates an estimated surface from a scattered set of points with z-values. More so than other interpolation methods, a thorough investigation of the spatial behavior of the phenomenon represented by the z-values should be done before you select the best estimation method for generating the output surface.

Natural neighbour

[Natural Neighbor](#) interpolation finds the closest subset of input samples to a query point and applies weights to them based on proportionate areas to interpolate a value (Sibson, 1981). It is also known as Sibson or "area-stealing" interpolation.

Spline

The [Spline](#) tool uses an interpolation method that estimates values using a mathematical function that minimizes overall surface curvature, resulting in a smooth surface that passes exactly through the input points.

Spline with Barriers

The [Spline with Barriers](#) tool uses a method similar to the technique used in the Spline tool, with the major difference being that this tool honors discontinuities encoded in both the input barriers and the input point data.

Topo to Raster

The Topo to Raster and Topo to Raster by File tools use an interpolation technique specifically designed to create a surface that more closely represents a natural drainage surface and better preserves both ridgelines and stream networks from input contour data.

The algorithm used is based on that of ANUDEM, developed by Hutchinson et al at the Australian National University.

Trend

Trend is a global polynomial interpolation that fits a smooth surface defined by a mathematical function (a polynomial) to the input sample points. The trend surface changes gradually and captures coarse-scale patterns in the data.

IDW (3D Analyst)

Summary

Interpolates a raster surface from points using an inverse distance weighted (IDW) technique.

[Learn more about how IDW works](#)

Usage

- The output value for a cell using inverse distance weighting (IDW) is limited to the range of the values used to interpolate. Because IDW is a weighted distance average, the average cannot be greater than the highest or less than the lowest input. Therefore, it cannot create ridges or valleys if these extremes have not already been sampled (Watson and Philip 1985).
- The best results from IDW are obtained when sampling is sufficiently dense with regard to the local variation you are attempting to simulate. If the sampling of input points is sparse or uneven, the results may not sufficiently represent the desired surface (Watson and Philip 1985).
- The influence of an input point on an interpolated value is isotropic. Since the influence of an input point on an interpolated value is distance related, IDW is not ridge preserving (Philip and Watson 1982).
- The Output cell size parameter can be defined by a numeric value or obtained from an existing raster dataset. If the cell size hasn't been explicitly specified as the parameter value, it is derived from the Cell Size environment if it has been specified. If the parameter cell size or the environment cell size have not been specified, but the Snap Raster environment has been set, the cell size of the snap raster is used. If nothing is specified, the cell size is calculated from the shorter of the width or height of the extent divided by 250 in which the extent is in the output coordinate system specified in the environment.
- If the cell size is specified using a numeric value, the tool will use it directly for the output raster.

  If the cell size is specified using a raster dataset, the parameter will show the path of the raster dataset instead of the cell size value. The cell size of that raster dataset will be used directly in the analysis, provided the spatial reference of the dataset is the same as the output spatial reference. If the spatial reference of the dataset is different than the output spatial reference, it will be projected based on the selected Cell Size Projection Method value.

- Some input datasets may have several points with the same x,y coordinates. If the values of the points at the common location are the same, they are considered duplicates and have no effect on the output. If the values are different, they are considered coincident points.

  The various interpolation tools may handle this data condition differently. For example, in some cases, the first coincident point encountered is used for the calculation; in other cases, the last point encountered is used. This may cause some locations in the output raster to have different values than what you might expect. The solution is to prepare your data by removing these coincident points. The [Collect Events](#) tool in the Spatial Statistics toolbox is useful for identifying any coincident points in your data.

- The barriers option is used to specify the location of linear features known to interrupt the surface continuity. These features do not have z-values. Cliffs, faults, and embankments are

typical examples of barriers. Barriers limit the selected set of the input sample points used to interpolate output z-values to those samples on the same side of the barrier as the current processing cell. Separation by a barrier is determined by line-of-sight analysis between each pair of points. This means that topological separation is not required for two points to be excluded from each other's region of influence. Input sample points that lie exactly on the barrier line will be included in the selected sample set for both sides of the barrier.

- Barrier features are input as polyline features. IDW only uses the x,y coordinates for the linear feature; therefore, it is not necessary to provide z-values for the left and right sides of the barrier. Any z-values provided will be ignored.
- Using barriers will significantly extend the processing time.
- This tool has a limit of approximately 45 million input points. If your input feature class contains more than 45 million points, the tool may fail to create a result. You can avoid this limit by interpolating your study area in several pieces, making sure there is some overlap in the edges, then mosaicking the results to create a single large raster dataset. Alternatively, you can use a terrain dataset to store and visualize points and surfaces comprised of billions of measurement points.

  If you have the Geostatistical Analyst extension, you may be able to process larger datasets with the version of the IDW tool available there.

- The input feature data must contain at least one valid field.
- For data formats that support Null values, such as file geodatabase feature classes, a Null value will be ignored when used as input.
- References:

  Philip, G. M., and D. F. Watson. "A Precise Method for Determining Contoured Surfaces." Australian Petroleum Exploration Association Journal 22: 205–212. 1982.

  Watson, D. F., and G. M. Philip. "A Refinement of Inverse Distance Weighted Interpolation." Geoprocessing 2:315–327. 1985.

Parameters

DialogPython

Label	Explanation	Data Type
Input point features	The input point features containing the z-values to be interpolated into a surface raster.	Feature Layer
Z value field	The field that holds a height or magnitude value for each point. This can be a numeric field or the Shape field if the input point features contain z-values.	Field
Output raster	The output interpolated surface raster. It is always a floating-point raster.	Raster Dataset
Output cell size (Optional)	The cell size of the output raster that will be created. This parameter can be defined by a numeric value or obtained from an existing raster dataset. If the cell size hasn't been explicitly specified as the parameter value, the environment cell size value will be used if	Analysis Cell Size

Label	Explanation	Data Type
	specified; otherwise, additional rules will be used to calculate it from the other inputs. See the usage section for more detail.	
Power (Optional)	The exponent of distance. Controls the significance of surrounding points on the interpolated value. A higher power results in less influence from distant points. It can be any real number greater than 0, but the most reasonable results will be obtained using values from 0.5 to 3. The default is 2.	Double
Search radius (Optional)	Defines which of the input points will be used to interpolate the value for each cell in the output raster. There are two options: Variable and Fixed. Variable is the default. • Variable Uses a variable search radius in order to find a specified number of input sample points for the interpolation. ◦ Number of points—An integer value specifying the number of nearest input sample points to be used to perform interpolation. The default is 12 points. ◦ Maximum distance—Specifies the distance, in map units, by which to limit the search for the nearest input sample points. The default value is the length of the extent's diagonal. • Fixed Uses a specified fixed distance within which all input points will be used for the interpolation. ◦ Distance—Specifies the distance as a radius within which input sample points will be used to perform the interpolation. The value of the radius is expressed in map units. The default radius is five times the cell size of the output raster. ◦ Minimum number of points—An integer defining the minimum number of points to be used for interpolation. The default value is 0. If the required number of points is not found within the specified distance, the search distance will be increased until the specified minimum number of points is found.	Radius

Label	Explanation	Data Type
	When the search radius needs to be increased it is done so until the Minimum number of points fall within that radius, or the extent of the radius crosses the lower (southern) and/or upper (northern) extent of the output raster. NoData is assigned to all locations that do not satisfy the above condition.	
Input barrier polyline features (Optional)	Polyline features to be used as a break or limit in searching for the input sample points.	Feature Layer

How IDW works

Inverse distance weighted (IDW) interpolation determines cell values using a linearly weighted combination of a set of sample points. The weight is a function of inverse distance. The surface being interpolated should be that of a locationally dependent variable.
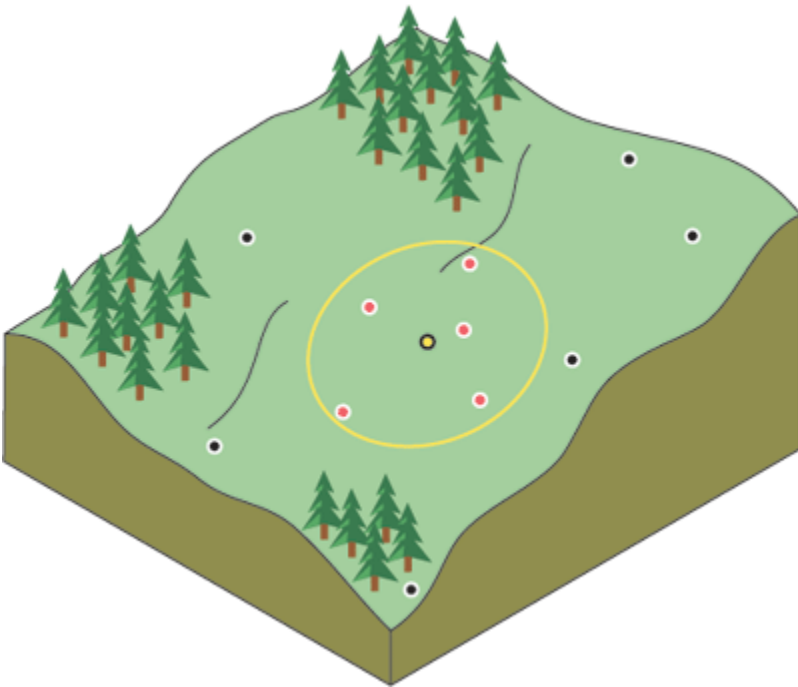
IDW neighborhood for selected point

This method assumes that the variable being mapped decreases in influence with distance from its sampled location. For example, when interpolating a surface of consumer purchasing power for a retail site analysis, the purchasing power of a more distant location will have less influence because people are more likely to shop closer to home.

Controlling the influence with the Power parameter

IDW relies mainly on the inverse of the distance raised to a mathematical power. The Power parameter lets you control the significance of known points on the interpolated values based on their distance from the output point. It is a positive, real number, and its default value is 2.

By defining a higher power value, more emphasis can be put on the nearest points. Thus, nearby data will have the most influence, and the surface will have more detail (be less smooth). As the power increases, the interpolated values begin to approach the value of the nearest sample point. Specifying a lower value for power will give more influence to surrounding points that are farther away, resulting in a smoother surface.

Since the IDW formula is not linked to any real physical process, there is no way to determine that a particular power value is too large. As a general guideline, a power of 30 would be considered extremely large and thus of questionable use. Also keep in mind that if the distances or the power value are large, the results may be incorrect.

An optimal value for the power can be considered to be where the minimum mean absolute error is at its lowest. The ArcGIS Geostatistical Analyst extension provides a way to investigate this.

Limiting the points used for interpolation

The characteristics of the interpolated surface can also be controlled by limiting the input points used in the calculation of each output cell value. Limiting the number of input points considered can improve processing speeds. Also consider that input points far away from the cell location where the prediction is being made may have poor or no spatial correlation, so there may be reason to eliminate them from the calculation.

You can specify the number of points to use directly, or specify a fixed radius within which points will be included in the interpolation.

Variable search radius

With a variable search radius, the number of points used in calculating the value of the interpolated cell is specified, which makes the radius distance vary for each interpolated cell, depending on how far it has to search around each interpolated cell to reach the specified number of input points. Thus, some neighborhoods will be small and others will be large, depending on the density of the measured points near the interpolated cell. You can also specify a maximum distance (in map units) that the search radius cannot exceed. If the radius for a particular neighborhood reaches the maximum distance before obtaining the specified number of points, the prediction for that location will be performed on the number of measured points within the maximum distance. Generally, you will use smaller neighborhoods or a minimum number of points when the phenomenon has a great amount of variation.

Fixed search radius

A fixed search radius requires a neighborhood distance and a minimum number of points. The distance dictates the radius of the circle of the neighborhood (in map units). The distance of the radius is constant, so for each interpolated cell, the radius of the circle used to find input points is the same. The minimum number of points indicates the minimum number of measured points to use

within the neighborhood. All the measured points that fall within the radius will be used in the calculation of each interpolated cell. When there are fewer measured points in the neighborhood than the specified minimum, the search radius will increase until it can encompass the minimum number of points. The specified fixed search radius will be used for each interpolated cell (cell center) in the study area; thus, if your measured points are not spread out equally (which they rarely are), there are likely to be different numbers of measured points used in the different neighborhoods for the various predictions.

Using barriers

A barrier is a polyline dataset used as a breakline that limits the search for input sample points. A polyline can represent a cliff, ridge, or some other interruption in a landscape. Only those input sample points on the same side of the barrier as the current processing cell will be considered.

References

Philip, G. M., and D. F. Watson. "A Precise Method for Determining Contoured Surfaces." Australian Petroleum Exploration Association Journal 22: 205–212. 1982.

Watson, D. F., and G. M. Philip. "A Refinement of Inverse Distance Weighted Interpolation." Geoprocessing 2:315–327. 1985.

Kriging (3D Analyst)

Summary

Interpolates a raster surface from points using kriging.

The [Empirical Bayesian Kriging](#) tool provides enhanced functionality or performance.

[Learn more about how Kriging works](#)

Usage

- Kriging is a processor-intensive process. The speed of execution is dependent on the number of points in the input dataset and the size of the search window.
- Low values within the optional output variance of prediction raster indicate a high degree of confidence in the predicted value. High values may indicate a need for more data points.
- The Universal kriging types assume that there is a structural component present and that the local trend varies from one location to another.
- The Semivariogram properties allow control of the semivariogram used for kriging. A default value for Lag size is initially set to the default output cell size. For Major range, Partial sill, and Nugget, a default value will be calculated internally if nothing is specified.
- The optional output variance of prediction raster contains the kriging variance at each output raster cell. Assuming the kriging errors are normally distributed, there is a 95.5 percent probability that the actual z-value at the cell is the predicted raster value, plus or minus two times the square root of the value in the variance raster.
- The Output cell size parameter can be defined by a numeric value or obtained from an existing raster dataset. If the cell size hasn't been explicitly specified as the parameter value, it is derived from the Cell Size environment if it has been specified. If the parameter cell size or the environment cell size have not been specified, but the Snap Raster environment has been set, the cell size of the snap raster is used. If nothing is specified, the cell size is calculated from the shorter of the width or height of the extent divided by 250 in which the extent is in the output coordinate system specified in the environment.
- If the cell size is specified using a numeric value, the tool will use it directly for the output raster.

  If the cell size is specified using a raster dataset, the parameter will show the path of the raster dataset instead of the cell size value. The cell size of that raster dataset will be used directly in the analysis, provided the spatial reference of the dataset is the same as the output spatial reference. If the spatial reference of the dataset is different than the output spatial reference, it will be projected based on the selected Cell Size Projection Method value.

- Some input datasets may have several points with the same x,y coordinates. If the values of the points at the common location are the same, they are considered duplicates and have no effect on the output. If the values are different, they are considered coincident points.

  The various interpolation tools may handle this data condition differently. For example, in some cases, the first coincident point encountered is used for the calculation; in other cases, the last point encountered is used. This may cause some locations in the output raster to have different values than what you might expect. The solution is to prepare your data by

removing these coincident points. The [Collect Events](#) tool in the Spatial Statistics toolbox is useful for identifying any coincident points in your data.

- For data formats that support Null values, such as file geodatabase feature classes, a Null value will be ignored when used as input.

Parameters

Label	Explanation	Data Type
Input point features	The input point features containing the z-values to be interpolated into a surface raster.	Feature Layer
Z value field	The field that holds a height or magnitude value for each point. This can be a numeric field or the Shape field if the input point features contain z-values.	Field
Output surface raster	The output interpolated surface raster. It is always a floating-point raster.	Raster Dataset
Semivariogram properties	The Semivariogram model to be used. There are two methods for kriging: Ordinary and Universal. Ordinary kriging can use the following semivariogram models: • Spherical—Spherical semivariogram model. This is the default. • Circular—Circular semivariogram model. • Exponential—Exponential semivariogram model. • Gaussian—Gaussian or normal distribution semivariogram model. • Linear—Linear semivariogram model with a sill. Universal kriging can use the following semivariogram models: • Linear with Linear drift—Universal Kriging with linear drift. • Linear with Quadratic drift—Universal Kriging with quadratic drift. There are options available via the Advanced Parameters dialog box. These parameters are: • Lag size—The default is the output raster cell size. • Major range—Represents a distance beyond which there is little or no correlation. • Partial sill—The difference between the nugget and the sill. • Nugget—Represents the error and variation at spatial scales too fine to detect. The nugget effect is seen as a discontinuity at the origin.	KrigingModel

Label	Explanation	Data Type
Output cell size (Optional)	The cell size of the output raster that will be created. This parameter can be defined by a numeric value or obtained from an existing raster dataset. If the cell size hasn't been explicitly specified as the parameter value, the environment cell size value will be used if specified; otherwise, additional rules will be used to calculate it from the other inputs. See the usage section for more detail.	Analysis Cell Size

Search radius (Optional)	Defines which of the input points will be used to interpolate the value for each cell in the output raster.
	
	There are two options: Variable and Fixed. Variable is the default.
	
	- Variable
	
	  Uses a variable search radius in order to find a specified number of input sample points for the interpolation.
	
	    o Number of points—An integer value specifying the number of nearest input sample points to be used to perform interpolation. The default is 12 points.
	    o Maximum distance—Specifies the distance, in map units, by which to limit the search for the nearest input sample points. The default value is the length of the extent's diagonal.
	
	- Fixed
	
	  Uses a specified fixed distance within which all input points will be used for the interpolation.
	
	    o Distance—Specifies the distance as a radius within which input sample points will be used to perform the interpolation.
	
	      The value of the radius is expressed in map units. The default radius is five times the cell size of the output raster.
	
	    o Minimum number of points—An integer defining the minimum number of points to be used for interpolation. The default value is 0.
	
	      If the required number of points is not found within the specified distance, the search distance will be increased until the specified minimum number of points is found.
		Radius

Label	Explanation	Data Type
	When the search radius needs to be increased it is done so until the Minimum number of points fall within that radius, or the extent of the radius crosses the lower (southern) and/or upper (northern) extent of the output raster. NoData is assigned to all locations that do not satisfy the above condition.	
Output variance of prediction raster (Optional)	Optional output raster where each cell contains the predicted variance values for that location.	Raster Dataset

How Kriging works

Kriging is an advanced geostatistical procedure that generates an estimated surface from a scattered set of points with z-values. Unlike other interpolation methods in the Interpolation toolset, to use the [Kriging](#) tool effectively involves an interactive investigation of the spatial behavior of the phenomenon represented by the z-values before you select the best estimation method for generating the output surface.

What is kriging?

The IDW (inverse distance weighted) and Spline interpolation tools are referred to as deterministic interpolation methods because they are directly based on the surrounding measured values or on specified mathematical formulas that determine the smoothness of the resulting surface. A second family of interpolation methods consists of geostatistical methods, such as kriging, which are based on statistical models that include autocorrelation—that is, the statistical relationships among the measured points. Because of this, geostatistical techniques not only have the capability of producing a prediction surface but also provide some measure of the certainty or accuracy of the predictions.

Kriging assumes that the distance or direction between sample points reflects a spatial correlation that can be used to explain variation in the surface. The Kriging tool fits a mathematical function to a specified number of points, or all points within a specified radius, to determine the output value for each location. Kriging is a multistep process; it includes exploratory statistical analysis of the data, variogram modeling, creating the surface, and (optionally) exploring a variance surface. Kriging is most appropriate when you know there is a spatially correlated distance or directional bias in the data. It is often used in soil science and geology.

The kriging formula

Kriging is similar to IDW in that it weights the surrounding measured values to derive a prediction for an unmeasured location. The general formula for both interpolators is formed as a weighted sum of the data:

$$\hat{Z}(s_0) = \sum_{i=1}^{N} \lambda_i Z(s_i)$$

- where:

    $Z(s_i)$ = the measured value at the ith location

    λ_i = an unknown weight for the measured value at the ith location

    s_0 = the prediction location

    N = the number of measured values

In IDW, the weight, λ_i, depends solely on the distance to the prediction location. However, with the kriging method, the weights are based not only on the distance between the measured points and the prediction location but also on the overall spatial arrangement of the measured points. To use the spatial arrangement in the weights, the spatial autocorrelation must be quantified. Thus, in ordinary kriging, the weight, λ_i, depends on a fitted model to the measured points, the distance to the prediction location, and the spatial relationships among the measured values around the prediction location. The following sections discuss how the general kriging formula is used to create a map of the prediction surface and a map of the accuracy of the predictions.

Creating a prediction surface map with kriging

To make a prediction with the kriging interpolation method, two tasks are necessary:

- Uncover the dependency rules.
- Make the predictions.

To realize these two tasks, kriging goes through a two-step process:

1. It creates the variograms and covariance functions to estimate the statistical dependence (called spatial autocorrelation) values that depend on the model of autocorrelation (fitting a model).
2. It predicts the unknown values (making a prediction).

It is because of these two distinct tasks that it has been said that kriging uses the data twice: the first time to estimate the spatial autocorrelation of the data and the second to make the predictions.

Variography

Fitting a model, or spatial modeling, is also known as structural analysis, or variography. In spatial modeling of the structure of the measured points, you begin with a graph of the empirical semivariogram, computed with the following equation for all pairs of locations separated by distance h:

Semivariogram(distance$_h$) = 0.5 * average((value$_i$ – value$_j$)^2)

The formula involves calculating the difference squared between the values of the paired locations.

The image below shows the pairing of one point (the red point) with all other measured locations. This process continues for each measured point.
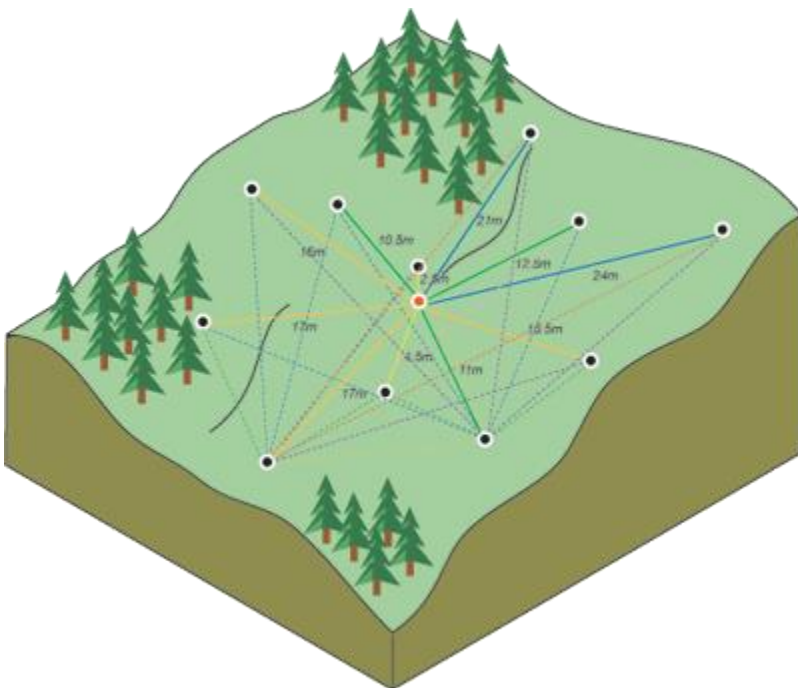
Calculating the difference squared between the paired locations

Often, each pair of locations has a unique distance, and there are often many pairs of points. To plot all pairs quickly becomes unmanageable. Instead of plotting each pair, the pairs are grouped into lag bins. For example, compute the average semivariance for all pairs of points that are greater than 40 meters apart but less than 50 meters. The empirical semivariogram is a graph of the averaged semivariogram values on the y-axis and the distance (or lag) on the x-axis (see diagram below).
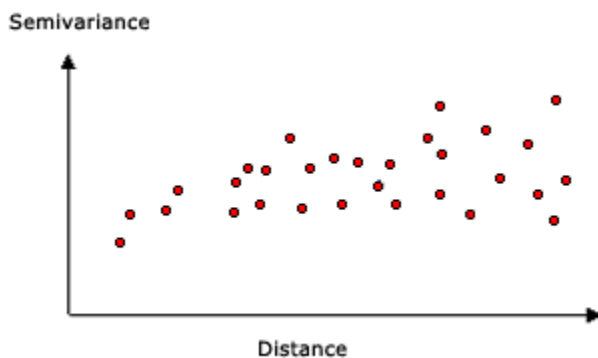
Empirical semivariogram graph example

Spatial autocorrelation quantifies a basic principle of geography: things that are closer are more alike than things farther apart. Thus, pairs of locations that are closer (far left on the x-axis of the semivariogram cloud) should have more similar values (low on the y-axis of the semivariogram cloud). As pairs of locations become farther apart (moving to the right on the x-axis of the semivariogram cloud), they should become more dissimilar and have a higher squared difference (moving up on the y-axis of the semivariogram cloud).

Fitting a model to the empirical semivariogram

The next step is to fit a model to the points forming the empirical semivariogram. Semivariogram modeling is a key step between spatial description and spatial prediction. The main application of kriging is the prediction of attribute values at unsampled locations. The empirical semivariogram provides information on the spatial autocorrelation of datasets. However, it does not provide information for all possible directions and distances. For this reason, and to ensure that kriging predictions have positive kriging variances, it is necessary to fit a model—that is, a continuous function or curve—to the empirical semivariogram. Abstractly, this is similar to regression analysis, in which a continuous line or curve is fitted to the data points.

To fit a model to the empirical semivariogram, select a function that serves as your model—for example, a spherical type that rises and levels off for larger distances beyond a certain range (see the spherical model example below). There are deviations of the points on the empirical semivariogram from the model; some points are above the model curve, and some points are below. However, if you add the distance each point is above the line and add the distance each point is below the line, the two values should be similar. There are many semivariogram models from which to choose.

Semivariogram models

The Kriging tool provides the following functions from which to choose for modeling the empirical semivariogram:

- Circular
- Spherical
- Exponential
- Gaussian
- Linear

The selected model influences the prediction of the unknown values, particularly when the shape of the curve near the origin differs significantly. The steeper the curve near the origin, the more influence the closest neighbors will have on the prediction. As a result, the output surface will be less smooth. Each model is designed to fit different types of phenomena more accurately.

The diagrams below show two common models and identify how the functions differ:

A spherical model example

This model shows a progressive decrease of spatial autocorrelation (equivalently, an increase of semivariance) until some distance, beyond which autocorrelation is zero. The spherical model is one of the most commonly used models.

Spherical model example
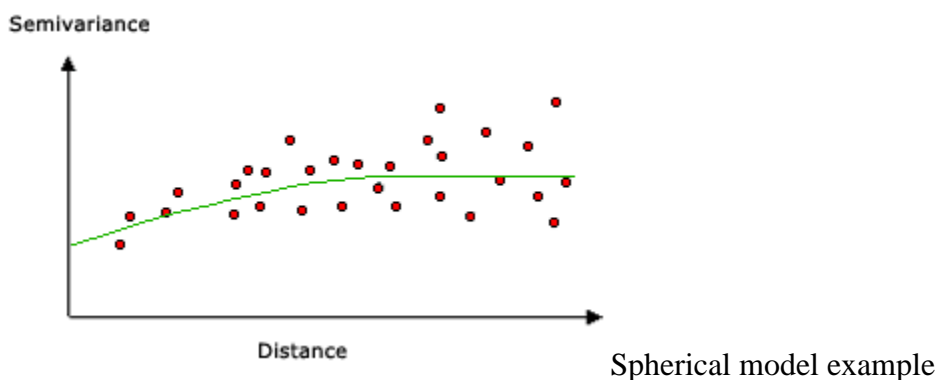
An exponential model example

This model is applied when spatial autocorrelation decreases exponentially with increasing distance. Here, the autocorrelation disappears completely only at an infinite distance. The exponential model is also a commonly used model. The choice of which model to use is based on the spatial autocorrelation of the data and on prior knowledge of the phenomenon.

Exponential model example
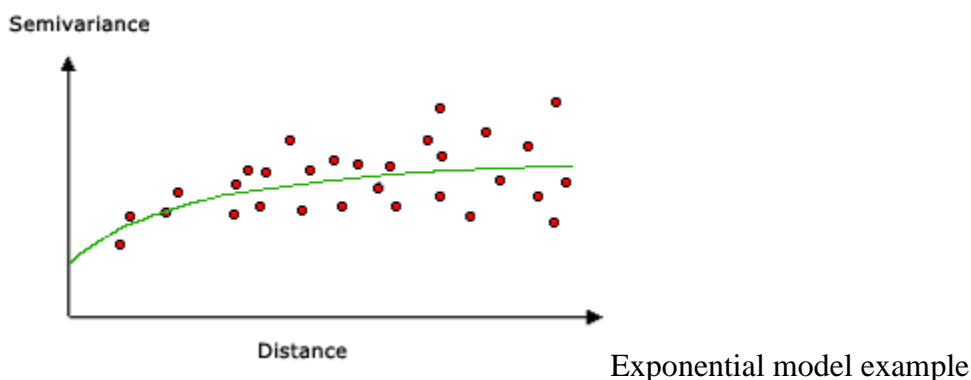
More of the mathematical models are illustrated below.

Understanding a semivariogram—Range, sill, and nugget

As previously discussed, the semivariogram depicts the spatial autocorrelation of the measured sample points. Because of a basic principle of geography (things that are closer are more alike), measured points that are close will generally have a smaller difference squared than those farther apart. Once each pair of locations is plotted after being binned, a model is fit through them. Range, sill, and nugget are commonly used to describe these models.

Range and sill

When you look at the model of a semivariogram, you will notice that at a certain distance the model levels out. The distance where the model first flattens is known as the range. Sample locations separated by distances closer than the range are spatially autocorrelated, whereas locations farther apart than the range are not.
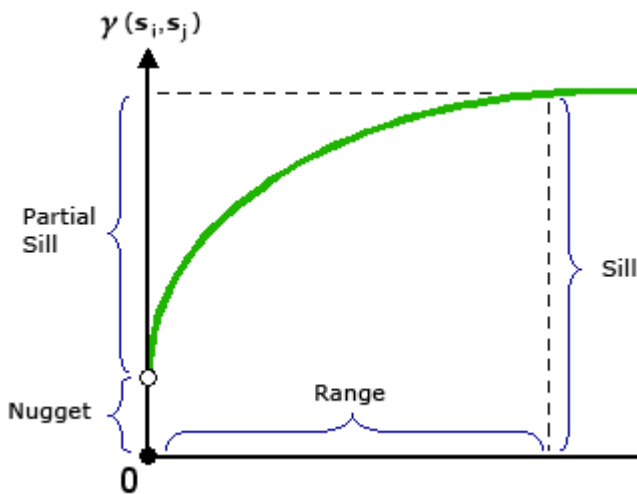
Illustration of Range, Sill, and Nugget components

The value at which the semivariogram model attains the range (the value on the y-axis) is called the sill. A partial sill is the sill minus the nugget. The nugget is described in the following section.

Nugget

Theoretically, at zero separation distance (for example, lag = 0), the semivariogram value is 0. However, at an infinitely small separation distance, the semivariogram often exhibits a nugget effect, which is a value greater than 0. If the semivariogram model intercepts the y-axis at 2, then the nugget is 2.

The nugget effect can be attributed to measurement errors or spatial sources of variation at distances smaller than the sampling interval (or both). Measurement error occurs because of the error inherent in measuring devices. Natural phenomena can vary spatially over a range of scales. Variation at microscales smaller than the sampling distances will appear as part of the nugget effect. Before collecting data, it is important to gain an understanding of the scales of spatial variation in which you are interested.

Making a prediction

After you have uncovered the dependence or autocorrelation in your data (see Variography section above) and have finished with the first use of the data—using the spatial information in the data to compute distances and model the spatial autocorrelation—you can make a prediction using the fitted model. Thereafter, the empirical semivariogram is set aside.

You can now use the data to make predictions. Like IDW interpolation, kriging forms weights from surrounding measured values to predict unmeasured locations. As with IDW interpolation, the measured values closest to the unmeasured locations have the most influence. However, the kriging weights for the surrounding measured points are more sophisticated than those of IDW. IDW uses a simple algorithm based on distance, but kriging weights come from a semivariogram that was developed by looking at the spatial nature of the data. To create a continuous surface of the phenomenon, predictions are made for each location, or cell centers, in the study area based on the semivariogram and the spatial arrangement of measured values that are nearby.

Kriging methods

There are two kriging methods: ordinary and universal.

Ordinary kriging is the most general and widely used of the kriging methods and is the default. It assumes the constant mean is unknown. This is a reasonable assumption unless there is a scientific reason to reject it.

Universal kriging assumes that there is an overriding trend in the data—for example, a prevailing wind—and it can be modeled by a deterministic function, a polynomial. This polynomial is subtracted from the original measured points, and the autocorrelation is modeled from the random errors. Once the model is fit to the random errors and before making a prediction, the polynomial is added back to the predictions to give meaningful results. Universal kriging should only be used when you know there is a trend in your data and you can give a scientific justification to describe it.

Semivariogram graphs

Kriging is a complex procedure that requires greater knowledge about spatial statistics than can be conveyed in this topic. Before using kriging, you should have a thorough understanding of its fundamentals and assess the appropriateness of your data for modeling with this technique. If you do not have a good understanding of this procedure, it is strongly recommended that you review some of the references listed at the end of this topic.

Kriging is based on the regionalized variable theory that assumes that the spatial variation in the phenomenon represented by the z-values is statistically homogeneous throughout the surface (for example, the same pattern of variation can be observed at all locations on the surface). This hypothesis of spatial homogeneity is fundamental to the regionalized variable theory.

Mathematical models

Below are the general shapes and the equations of the mathematical models used to describe the semivariance.
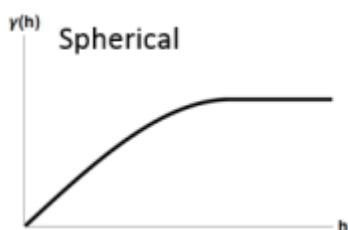
$$\gamma(h) = \begin{cases} c_0 + c\left(\frac{3h}{2a} - \frac{1}{2}\left(\frac{h^3}{a^3}\right)\right) & 0 < h \leq a \\ c_0 + c & h > a \\ 0 & h = 0 \end{cases}$$

Spherical semivariance model illustration
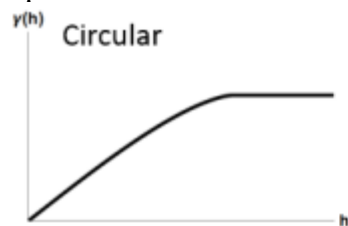
$$\gamma(h) = \begin{cases} c_0 + \frac{2c}{\pi}\left(\frac{h}{a}\sqrt{1 - \left(\frac{h^2}{a^2}\right)} + arcsin\left(\frac{h}{a}\right)\right) & 0 < h \leq a \\ c_0 + c & h > a \\ 0 & h = 0 \end{cases}$$

Circular semivariance model illustration
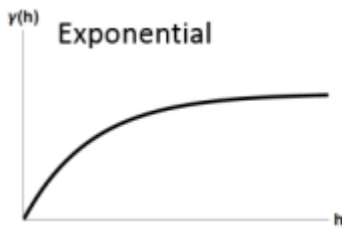
$$\gamma(h) = \begin{cases} c_0 + c\left(1 - exp\left(\frac{-h}{a}\right)\right) & h > 0 \\ 0 & h = 0 \end{cases}$$

Exponential semivariance model illustration

$$\gamma(h) = \begin{cases} c_0 + c\left(1 - exp\left(-\frac{h^2}{a^2}\right)\right) & h > 0 \\ 0 & h = 0 \end{cases}$$

Gaussian semivariance model illustration
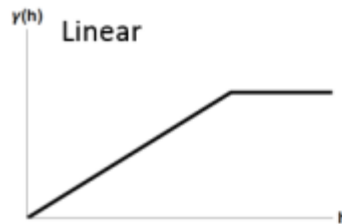
$$\gamma(h) = \begin{cases} c_0 + c\left(\frac{h}{a}\right) & 0 < h \leq a \\ c_0 + c & h > a \\ 0 & h = 0 \end{cases}$$

Linear semivariance model illustration

References

Burrough, P. A. Principles of Geographical Information Systems for Land Resources Assessment. New York: Oxford University Press. 1986.

Heine, G. W. "A Controlled Study of Some Two-Dimensional Interpolation Methods." COGS Computer Contributions 3 (no. 2): 60–72. 1986.

McBratney, A. B., and R. Webster. "Choosing Functions for Semi-variograms of Soil Properties and Fitting Them to Sampling Estimates." Journal of Soil Science 37: 617–639. 1986.

Oliver, M. A. "Kriging: A Method of Interpolation for Geographical Information Systems." International Journal of Geographic Information Systems 4: 313–332. 1990.

Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. Numerical Recipes in C: The Art of Scientific Computing. New York: Cambridge University Press. 1988.

Royle, A. G., F. L. Clausen, and P. Frederiksen. "Practical Universal Kriging and Automatic Contouring." Geoprocessing 1: 377–394. 1981.

Natural Neighbor (3D Analyst)

Summary

Interpolates a raster surface from points using a natural neighbor technique.

[Learn more about how Natural Neighbor works](#)

Usage

- If the cell center of the perimeter cells of the output raster fall outside the convex hull (defined by the input points), then those cells will be assigned NoData values. If an input point falls within one of these perimeter cells and the cell center falls outside the convex hull, the cell will still be assigned a value of NoData.
- The Output cell size parameter can be defined by a numeric value or obtained from an existing raster dataset. If the cell size hasn't been explicitly specified as the parameter value, it is derived from the Cell Size environment if it has been specified. If the parameter cell size or the environment cell size have not been specified, but the Snap Raster environment has been set, the cell size of the snap raster is used. If nothing is specified, the cell size is calculated from the shorter of the width or height of the extent divided by 250 in which the extent is in the output coordinate system specified in the environment.
- If the cell size is specified using a numeric value, the tool will use it directly for the output raster.

  If the cell size is specified using a raster dataset, the parameter will show the path of the raster dataset instead of the cell size value. The cell size of that raster dataset will be used directly in the analysis, provided the spatial reference of the dataset is the same as the output spatial reference. If the spatial reference of the dataset is different than the output spatial reference, it will be projected based on the selected Cell Size Projection Method value.

- Some input datasets may have several points with the same x,y coordinates. If the values of the points at the common location are the same, they are considered duplicates and have no effect on the output. If the values are different, they are considered coincident points.

  The various interpolation tools may handle this data condition differently. For example, in some cases, the first coincident point encountered is used for the calculation; in other cases, the last point encountered is used. This may cause some locations in the output raster to have different values than what you might expect. The solution is to prepare your data by removing these coincident points. The [Collect Events](#) tool in the Spatial Statistics toolbox is useful for identifying any coincident points in your data.

- For data formats that support Null values, such as file geodatabase feature classes, a Null value will be ignored when used as input.
- It is recommended that the input data be in a projected coordinate system rather than in a geographic coordinate system.
- An alternative approach is to use a [TIN](#) dataset. First, [create a TIN surface](#) from your source data. Then, convert the resulting TIN to a raster with the [TIN To Raster](#) tool, using the Natural Neighbors option. This is particularly useful if you have breaklines or an irregularly shaped data area.

Parameters

DialogPython

Label	Explanation	Data Type
Input point features	The input point features containing the z-values to be interpolated into a surface raster.	Feature Layer
Z value field	The field that holds a height or magnitude value for each point. This can be a numeric field or the Shape field if the input point features contain z-values.	Field
Output raster	The output interpolated surface raster. It is always a floating-point raster.	Raster Dataset
Output cell size (Optional)	The cell size of the output raster that will be created. This parameter can be defined by a numeric value or obtained from an existing raster dataset. If the cell size hasn't been explicitly specified as the parameter value, the environment cell size value will be used if specified; otherwise, additional rules will be used to calculate it from the other inputs. See the usage section for more detail.	Analysis Cell Size

How Natural Neighbor works

The algorithm used by the Natural Neighbor interpolation tool finds the closest subset of input samples to a query point and applies weights to them based on proportionate areas to interpolate a value (Sibson 1981). It is also known as Sibson or "area-stealing" interpolation. Its basic properties are that it's local, using only a subset of samples that surround a query point, and interpolated heights are guaranteed to be within the range of the samples used. It does not infer trends and will not produce peaks, pits, ridges, or valleys that are not already represented by the input samples. The surface passes through the input samples and is smooth everywhere except at locations of the input samples.

Tip:

You can also use the TIN To Raster tool to incorporate breaklines into the interpolation. Breaklines can augment results of the interpolation, creating linear discontinuities where appropriate such as along roadsides and water bodies. It adapts locally to the structure of the input data, requiring no input from the user pertaining to search radius, sample count, or shape. It works equally well with regularly and irregularly distributed data (Watson 1992).

The natural neighbors of any point are those associated with neighboring Voronoi (Thiessen) polygons. Initially, a Voronoi diagram is constructed of all the given points, represented by the olive-colored polygons. A new Voronoi polygon, beige color, is then created around the interpolation point (red star). The proportion of overlap between this new polygon and the initial polygons is then used as the weights.
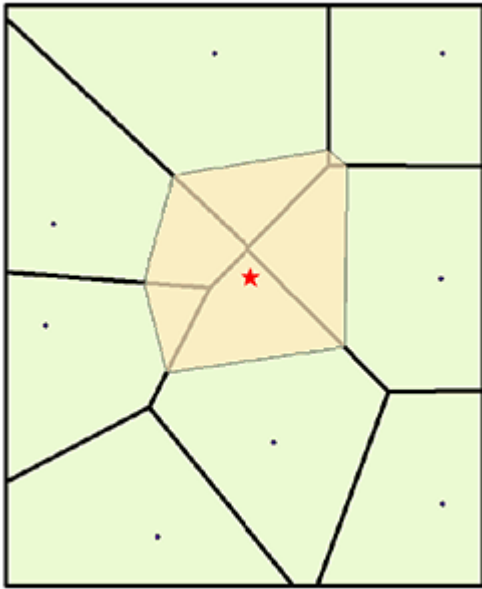
Example of Voronoi polygon created around interpolation point

By comparison, a distance-based interpolator tool such as IDW (inverse distance weighted) would assign similar weights to the northernmost point and to the northeastern point based on their similar distance from the interpolation point. Natural neighbor interpolation, however, assigns weights of 19.12 percent and 0.38 percent, respectively, which is based on the percentage of overlap.

References

Sibson, R. "A Brief Description of Natural Neighbor Interpolation," chapter 2 in *Interpolating Multivariate Data*. New York: John Wiley & Sons, 1981. 21–36.

Watson, D. *Contouring: A Guide to the Analysis and Display of Spatial Data*. London: Pergamon Press, 1992.

Spline (3D Analyst)

Summary

Interpolates a raster surface from points using a two-dimensional minimum curvature spline technique.

The resulting smooth surface passes exactly through the input points.

[Learn more about how Spline works](#)

Usage

- The greater the value of Number of Points, the smoother the surface of the output raster.
- The Output cell size parameter can be defined by a numeric value or obtained from an existing raster dataset. If the cell size hasn't been explicitly specified as the parameter value, it is derived from the Cell Size environment if it has been specified. If the parameter cell size or the environment cell size have not been specified, but the Snap Raster environment has been set, the cell size of the snap raster is used. If nothing is specified, the cell size is calculated from the shorter of the width or height of the extent divided by 250 in which the extent is in the output coordinate system specified in the environment.
- If the cell size is specified using a numeric value, the tool will use it directly for the output raster.

  If the cell size is specified using a raster dataset, the parameter will show the path of the raster dataset instead of the cell size value. The cell size of that raster dataset will be used directly in the analysis, provided the spatial reference of the dataset is the same as the output spatial reference. If the spatial reference of the dataset is different than the output spatial reference, it will be projected based on the selected Cell Size Projection Method value.

- The Regularized option of Spline type usually produces smoother surfaces than those created with the Tension option.
  - With the Regularized option, higher values used for the weight parameter produce smoother surfaces. The values entered for this parameter must be equal to or greater than zero. Typical values used are 0, 0.001, 0.01, 0.1, and 0.5. The Weight is the square of the parameter referred to in the literature as tau (t).
  - With the Tension option, higher values entered for the weight parameter result in somewhat coarser surfaces, but surfaces that closely conform to the control points. The values entered must be equal to or greater than zero. Typical values are 0, 1, 5, and 10. The Weight is the square of the parameter referred to in the literature as phi (Φ).
- Some input datasets may have several points with the same x,y coordinates. If the values of the points at the common location are the same, they are considered duplicates and have no effect on the output. If the values are different, they are considered coincident points.

  The various interpolation tools may handle this data condition differently. For example, in some cases, the first coincident point encountered is used for the calculation; in other cases, the last point encountered is used. This may cause some locations in the output raster to have different values than what you might expect. The solution is to prepare your data by

removing these coincident points. The [Collect Events](#) tool in the Spatial Statistics toolbox is useful for identifying any coincident points in your data.

- For data formats that support Null values, such as file geodatabase feature classes, a Null value will be ignored when used as input.

Parameters

DialogPython

Label	Explanation	Data Type
Input point features	The input point features containing the z-values to be interpolated into a surface raster.	Feature Layer
Z value field	The field that holds a height or magnitude value for each point. This can be a numeric field or the Shape field if the input point features contain z-values.	Field
Output raster	The output interpolated surface raster. It is always a floating-point raster.	Raster Dataset
Output cell size (Optional)	The cell size of the output raster that will be created. This parameter can be defined by a numeric value or obtained from an existing raster dataset. If the cell size hasn't been explicitly specified as the parameter value, the environment cell size value will be used if specified; otherwise, additional rules will be used to calculate it from the other inputs. See the usage section for more detail.	Analysis Cell Size
Spline type (Optional)	The type of spline to be used. • Regularized—Yields a smooth surface and smooth first derivatives. • Tension—Tunes the stiffness of the interpolant according to the character of the modeled phenomenon.	String
Weight (Optional)	Parameter influencing the character of the surface interpolation. When the Regularized option is used, it defines the weight of the third derivatives of the surface in the curvature minimization expression. If the Tension option is used, it defines the weight of tension. The default weight is 0.1.	Double
Number of points (Optional)	The number of points per region used for local approximation. The default is 12.	Long

How Spline works

The [Spline](#) tool uses an interpolation method that estimates values using a mathematical function that minimizes overall surface curvature, resulting in a smooth surface that passes exactly through the input points.

Conceptual background

Conceptually, the sample points extrude to the height of their magnitude. The spline bends a sheet of rubber that passes through the input points while minimizing the total curvature of the surface. It fits a mathematical function to a specified number of nearest input points while passing through the sample points. This method is best for generating gently varying surfaces such as elevation, water table heights, or pollution concentrations.

The basic form of the minimum curvature Spline interpolation imposes the following two conditions on the interpolant:

- The surface must pass exactly through the data points.
- The surface must have minimum curvature. The cumulative sum of the squares of the second derivative terms of the surface taken over each point on the surface must be a minimum.

The basic minimum curvature technique is also referred to as thin plate interpolation. It ensures a smooth (continuous and differentiable) surface, together with continuous first-derivative surfaces. Rapid changes in gradient or slope (the first derivative) can occur in the vicinity of the data points; hence, this model is not suitable for estimating second derivative (curvature).

The basic interpolation technique can be applied by using a value of zero for the Weight argument to the Spline tool.

Spline types

There are two Spline types: Regularized and Tension. The Regularized type creates a smooth, gradually changing surface with values that may lie outside the sample data range. The Tension type controls the stiffness of the surface according to the character of the modeled phenomenon. It creates a less smooth surface with values more closely constrained by the sample data range.

Regularized spline type

The Regularized option modifies the minimization criteria so third-derivative terms are incorporated into the minimization criteria. The Weight parameter specifies the weight attached to the third-derivative terms during minimization, referred to as τ (tau) in the literature. Higher values of this term lead to smoother surfaces. Values between 0 and 0.5 are suitable. Using this option ensures a smooth surface together with smooth first-derivative surfaces. This technique is useful if the second derivative of the interpolated surface needs to be computed.

Tension spline type

The Tension option modifies the minimization criteria so first-derivative terms are incorporated into the minimization criteria. The Weight parameter specifies the weight attached to the first-derivative terms during minimization, referred to as Φ (phi) in the literature. A weight of zero results in the basic thin plate Spline interpolation. Using a larger value of weight reduces the stiffness of the plate, and in the limit as phi approaches infinity, the surface approximates the shape of a membrane or rubber sheets, passing through the points. The interpolated surface is smooth. First derivatives are continuous but not smooth.

Additional spline parameters

Further control of the output surface is accomplished through two additional parameters: weight and number of points.

Weight parameter

For the Regularized Spline method, the Weight parameter defines the weight of the third derivatives of the surface in the curvature minimization expression. The higher the weight, the smoother the output surface. The values entered for this parameter must be equal to or greater than zero. The typical values that may be used are 0, 0.001, 0.01, 0.1, and 0.5.

For the Tension Spline method, the Weight parameter defines the weight of tension. The higher the weight, the coarser the output surface. The values entered must be equal to or greater than zero. The typical values are 0, 1, 5, and 10.

Number of points parameter

The Number of points identifies the number of points used in the calculation of each interpolated cell. The more input points you specify, the more each cell is influenced by distant points and the smoother the output surface is. The larger the number of points, the longer it takes to process the output raster.

Spline equation

The algorithm used for the Spline tool uses the following formula for the surface interpolation:

$$S(x,y) = T(x,y) + \sum_{j=1}^{N} \lambda_j R(r_j)$$

- where:

  j = 1, 2, ..., N.

  N is the number of points.

  λ_j are coefficients found by the solution of a system of linear equations.

  r_j is the distance from the point (x,y) to the jth point.

T(x,y) and R(r) are defined differently, depending on the selected option.

For computational purposes, the entire space of the output raster is divided into blocks or regions equal in size. The number of regions in x- and y-directions are the same, and are rectangular in shape. The number of regions is determined by dividing the total amount of points in the input point dataset by the value specified for the number of points. For data less uniformly distributed, the regions may contain a significantly different number of points, with the value for the number of points being only the rough average. If in any region, the number of points is smaller than eight, the region is expanded until it contains a minimum of eight points.

For the Regularized option

$T(x,y) = a_1 + a_2x + a_3y$

- where:

  a_i are coefficients found by the solution of a system of linear equations.

and,

$$R(r) = \frac{1}{2\pi} \left\{ \frac{r^2}{4}\left[\ln\left(\frac{r}{2\tau}\right) + c - 1\right] + \tau^2 \left[K_o\left(\frac{r}{\tau}\right) + c + \ln\left(\frac{r}{2\pi}\right)\right]\right\}$$

- where:

  r is the distance between the point and the sample.

  τ^2 is the Weight parameter.

  K_o is the modified Bessel function.

  c is a constant equal to 0.577215.

For the Tension option

$T(x,y) = a_1$

- where:

  a_1 is a coefficient found by the solution of a system of linear equations.

and,

$$R(r) = -\frac{1}{2\pi\varphi^2} \left[\ln\left(\frac{r\varphi}{2}\right) + c + K_o(r\varphi)\right]$$

- where:

  r is the distance between the point and the sample.

  φ^2 is the Weight parameter.

  K_o is the modified Bessel function.

  c is a constant equal to 0.577215.

Regional processing of output

For computational purposes, the entire space of the output raster is divided into blocks or regions equal in size. The number of regions in x- and y-directions are the same, and are rectangular in shape. The number of regions is determined by dividing the total amount of points in the input point dataset by the value specified for the number of points. For data less uniformly distributed, the regions may contain a significantly different number of points, with the value for the number of points being only the rough average. If, in any region, the number of points is smaller than eight, the region expands until it contains a minimum of eight points.

References

Franke, R. 1982. Smooth Interpolation of Scattered Data by Local Thin Plate Splines. *Computer and Mathematics with Applications*. Vol. 8. No. 4. pp. 273–281. Great Britain.

Mitas, L., and H. Mitasova. 1988. General Variational Approach to the Interpolation Problem. *Computer and Mathematics with Applications*. Vol. 16. No. 12. pp. 983–992. Great Britain.

Spline with Barriers (3D Analyst)

Summary

Interpolates a raster surface, using barriers, from points using a minimum curvature spline technique. The barriers are entered as either polygon or polyline features.

[Learn more about how Spline with Barriers works](#)

Usage

- The output raster cannot have more than 65,536 columns or rows.
- The Output cell size parameter can be defined by a numeric value or obtained from an existing raster dataset. If the cell size hasn't been explicitly specified as the parameter value, it is derived from the Cell Size environment if it has been specified. If the parameter cell size or the environment cell size have not been specified, but the Snap Raster environment has been set, the cell size of the snap raster is used. If nothing is specified, the cell size is calculated from the shorter of the width or height of the extent divided by 250 in which the extent is in the output coordinate system specified in the environment.
- If the cell size is specified using a numeric value, the tool will use it directly for the output raster.

  If the cell size is specified using a raster dataset, the parameter will show the path of the raster dataset instead of the cell size value. The cell size of that raster dataset will be used directly in the analysis, provided the spatial reference of the dataset is the same as the output spatial reference. If the spatial reference of the dataset is different than the output spatial reference, it will be projected based on the selected Cell Size Projection Method value.

- Some input datasets may have several points with the same x,y coordinates. If the values of the points at the common location are the same, they are considered duplicates and have no effect on the output. If the values are different, they are considered coincident points.

The various interpolation tools may handle this data condition differently. For example, in some cases, the first coincident point encountered is used for the calculation; in other cases, the last point encountered is used. This may cause some locations in the output raster to have different values than what you might expect. The solution is to prepare your data by removing these coincident points. The [Collect Events](#) tool in the Spatial Statistics toolbox is useful for identifying any coincident points in your data.

For the Spline with Barriers tool, by default the values for each set of coincident points will be averaged.

- For data formats that support Null values, such as file geodatabase feature classes, a Null value will be ignored when used as input.

Parameters

DialogPython

Label	Explanation	Data Type
Input point features	The input point features containing the z-values to be interpolated into a surface raster.	Feature Layer
Z value field	The field that holds a height or magnitude value for each point. This can be a numeric field or the Shape field if the input point features contain z-values.	Field
Input barrier features (Optional)	The optional input barrier features to constrain the interpolation.	Feature Layer
Output cell size (Optional)	The cell size of the output raster that will be created. This parameter can be defined by a numeric value or obtained from an existing raster dataset. If the cell size hasn't been explicitly specified as the parameter value, the environment cell size value will be used if specified; otherwise, additional rules will be used to calculate it from the other inputs. See the usage section for more detail.	Analysis Cell Size
Output raster	The output interpolated surface raster. It is always a floating-point raster.	Raster Dataset
Smoothing Factor (Optional)	The parameter that influences the smoothing of the output surface. No smoothing is applied when the value is zero and the maximum amount of smoothing is applied when the factor equals 1. The default is 0.0.	Double

How Spline with Barriers works

The [Spline with Barriers](#) tool applies a minimum curvature method, as implemented through a one-directional multigrid technique that moves from an initial coarse grid, initialized in this case to the average of the input data, through a series of finer grids until an approximation of a minimum curvature surface is produced at the desired row and column spacing.

At each grid refinement level, the current grid-based surface model is treated as an elastic membrane, and a convergent linear iterative deformation operator is applied repeatedly at each node to achieve an approximation to a minimum curvature surface that honors both the input point data and discontinuities encoded in the barriers. The deformation that is applied to each cell is calculated on the basis of a molecular summation (Terzopoulos, 1988) that compares the weighted summation of 12 neighboring cells with the current value of a central target cell to calculate a new value for the target cell.

References and further reading

Briggs, I. C., 1974. Machine contouring using minimum curvature, *Geophysics*, Vol. 39, pages 39–48.

Terzopoulos, D., 1988. The computation of visible-surface representations, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 10, No. 4, (July), pages 417–438.

Smith, W. H. F., and P. Wessel, 1990. Gridding with continuous curvature splines in tension, *Geophysics*, Vol. 55, No. 3 (March 1990), pages 293–305.

Zoraster, S., A surface modeling algorithm designed for speed and ease of use with all petroleum industry data, *Computers & Geosciences*, 2003, Vol. 29, No. 9, pages 175–182.

Topo to Raster (3D Analyst)

Summary

Interpolates a hydrologically correct raster surface from point, line, and polygon data.

[Learn more about how Topo to Raster works](#)

Usage

- The best results will be obtained if all input data is stored in the same planar coordinate system and has the same z-units. Unprojected data (latitude-longitude) can be used; however, the results may not be as accurate, particularly at high latitudes.
- If more than one input point falls within an output cell,Topo to Raster will use the average value for the interpolation (only the first 100 points that fall within a cell will be considered and the rest will be ignored). If too many points are encountered by the algorithm, an error may occur, indicating the point dataset has too many points. The maximum number of points that can be used is NRows * NCols, where NRows is the number of rows in the output raster and NCols is the number of columns.
- When the input feature data type is Contour, the algorithm first generates a generalized morphology of the surface based on the curvature of the contours. The algorithm then implements the contours as a source of elevation information. Contours are best suited for large-scale data where the contours and corners are reliable indicators of streams and ridges. At smaller scales it can be just as effective, and less expensive, to digitize corner points of contours and use them as an input point feature class.
- The Output cell size parameter can be defined by a numeric value or obtained from an existing raster dataset. If the cell size hasn't been explicitly specified as the parameter value, it is derived from the Cell Size environment if it has been specified. If the parameter cell size or the environment cell size have not been specified, but the Snap Raster environment has been set, the cell size of the snap raster is used. If nothing is specified, the cell size is calculated from the shorter of the width or height of the extent divided by 250 in which the extent is in the output coordinate system specified in the environment.
- If the cell size is specified using a numeric value, the tool will use it directly for the output raster.

  If the cell size is specified using a raster dataset, the parameter will show the path of the raster dataset instead of the cell size value. The cell size of that raster dataset will be used directly in the analysis, provided the spatial reference of the dataset is the same as the output spatial reference. If the spatial reference of the dataset is different than the output spatial reference, it will be projected based on the selected Cell Size Projection Method value.

- Representing braided streams or using arcs to represent two sides of a stream may not produce reliable results. Stream data always takes priority over point or contour data; therefore, elevation data points that conflict with descent down each stream are ignored. Stream data is a powerful way of adding topographic information to the interpolation, further ensuring the quality of the output DEM.
- For data formats that support Null values, such as file geodatabase feature classes, a Null value will be ignored when used as input.
- Typical values for the Tolerance 1 and Tolerance 2 settings are:
  - For point data at 1:100,000 scale, use 5.0 and 200.0.

- o For less dense point data up to 1:500,000 scale, use 10.0 and 400.0.
- o For contour data with contour spacing of 10, use 5.0 and 100.0.

Tolerance 2 should be at least 6 times greater than Tolerance 1.

- To make experimentation with the inputs and parameters easier, use the Topo to Raster dialog box to create an output parameter file, which can be modified in any text editor and used as input to the Topo to Raster by File tool.
- This tool is a memory-intensive application and it is therefore not possible to create large output rasters. When large output is required, use the Margin parameter to produce smaller output rasters. For more details on how to do this, the creating and mosaicking rasters section of the How Topo to Raster works help.

Parameters

Label	Explanation	Data Type
Input feature data	The input features containing the z-values to be interpolated into a surface raster. Each feature input can have a field specified that contains the z-values and one of six types specified. - Feature layer—The input feature dataset. - Field—The name of the field that stores the attributes, where appropriate. - Type—The type of input feature dataset. There are nine types of accepted inputs: - Point elevation—A point feature class representing surface elevations. The Field stores the elevations of the points. - Contour—A line feature class that represents elevation contours. The Field stores the elevations of the contour lines. - Stream—A line feature class of stream locations. All arcs must be oriented to point downstream. The feature class should only contain single arc streams. There is no Field option for this input type. - Sink—A point feature class that represents known topographic depressions. The tool will not attempt to remove from the analysis any points explicitly identified as sinks. The Field used should be one that stores the elevation of the legitimate sink. If NONE is selected, only the location of the sink is used. - Boundary—A feature class containing a single polygon that represents the outer boundary of the output raster. Cells in the output raster outside this boundary will be NoData. This option can be used for clipping out water areas along coastlines before making the final output raster. There is no Field option for this input type.	TopoInput

Label	Explanation	Data Type
	• Lake—A polygon feature class that specifies the location of lakes. All output raster cells within a lake will be assigned to the minimum elevation value of all cells along the shoreline. There is no Field option for this input type. • Cliff—A line feature class of the cliffs. The cliff line features must be oriented so that the left-hand side of the line is on the low side of the cliff and the right-hand side is the high side of the cliff. There is no Field option for this input type. • Exclusion—A polygon feature class of the areas in which the input data should be ignored. These polygons permit removal of elevation data from the interpolation process. This is typically used to remove elevation data associated with dam walls and bridges. This enables interpolation of the underlying valley with connected drainage structure. There is no Field option for this input type. • Coast—A polygon feature class containing the outline of a coastal area. Cells in the final output raster that lie outside these polygons are set to a value that is less than the user-specified minimum height limit. There is no Field option for this input type.	
Output surface raster	The output interpolated surface raster. It is always a floating-point raster.	Raster Dataset
Output cell size (Optional)	The cell size of the output raster that will be created. This parameter can be defined by a numeric value or obtained from an existing raster dataset. If the cell size hasn't been explicitly specified as the parameter value, the environment cell size value will be used if specified; otherwise, additional rules will be used to calculate it from the other inputs. See the usage section for more detail.	Analysis Cell Size
Output extent (Optional)	Extent for the output raster dataset. Interpolation will occur out to the x and y limits, and cells outside that extent will be NoData. For best interpolation results along the edges of the output raster, the x and y limits should be smaller than the extent of the input data by at least 10 cells on each side. • Left—The default is the smallest x coordinate of all inputs. • Bottom—The default is the smallest y coordinate of all inputs. • Right—The default is the largest x coordinate of all inputs. • Top—The default is the largest y coordinate of all inputs. The default extent is the largest of all extents of the input feature data.	Extent

Label	Explanation	Data Type
Margin in cells (Optional)	Distance in cells to interpolate beyond the specified output extent and boundary. The value must be greater than or equal to 0 (zero). The default value is 20. If the Output extent and Boundary feature datasets are the same as the limit of the input data (the default), values interpolated along the edge of the DEM will not match well with adjacent DEM data. This is because they have been interpolated using one-half as much data as the points inside the raster, which are surrounded on all sides by input data. The Margin In Cells option allows input data beyond these limits to be used in the interpolation.	Long
Smallest z value to be used in interpolation (Optional)	The minimum z-value to be used in the interpolation. The default is 20 percent below the smallest of all the input values.	Double
Largest z value to be used in interpolation (Optional)	The maximum z-value to be used in the interpolation. The default is 20 percent above the largest of all input values.	Double
Drainage enforcement (Optional)	The type of drainage enforcement to apply. The drainage enforcement option can be set to attempt to remove all sinks or depressions so a hydrologically correct DEM can be created. If sink points have been explicitly identified in the input feature data, these depressions will not be filled. • Enforce—The algorithm will attempt to remove all sinks it encounters, whether they are real or spurious. This is the default. • Do not enforce—No sinks will be filled. • Enforce with sink—Points identified as sinks in Input feature data represent known topographic depressions and will not be altered. Any sink not identified in input feature data is considered spurious, and the algorithm will attempt to fill it.Having more than 8,000 spurious sinks causes the tool to fail.	String
Primary type of input data (Optional)	The dominant elevation data type of the input feature data. Specifying the relevant selection optimizes the search method used during the generation of streams and ridges. • Contour—The dominant type of input data will be elevation contours. This is the default. • Spot—The dominant type of input will be point.	String

Label	Explanation	Data Type
Maximum number of iterations (Optional)	The maximum number of interpolation iterations. The number of iterations must be greater than zero. A default of 20 is normally adequate for both contour and line data. A value of 30 will clear fewer sinks. Rarely, higher values (45–50) may be useful to clear more sinks or to set more ridges and streams. Iteration ceases for each grid resolution when the maximum number of iterations has been reached.	Long
Roughness penalty (Optional)	The integrated squared second derivative as a measure of roughness. The roughness penalty must be zero or greater. If the primary input data type is Contour, the default is zero. If the primary data type is Spot, the default is 0.5. Larger values are not normally recommended.	Double
Discretisation error factor (Optional)	The discrete error factor is used to adjust the amount of smoothing when converting the input data to a raster. The value must be greater than zero. The normal range of adjustment is 0.25 to 4, and the default is 1. A smaller value results in less data smoothing; a larger value causes greater smoothing.	Double
Vertical standard error (Optional)	The amount of random error in the z-values of the input data. The value must be zero or greater. The default is zero. The vertical standard error may be set to a small positive value if the data has significant random (non-systematic) vertical errors with uniform variance. In this case, set the vertical standard error to the standard deviation of these errors. For most elevation datasets, the vertical error should be set to zero, but it may be set to a small positive value to stabilize convergence when rasterizing point data with stream line data.	Double
Tolerance 1 (Optional)	This tolerance reflects the accuracy and density of the elevation points in relation to surface drainage. For point datasets, set the tolerance to the standard error of the data heights. For contour datasets, use one-half the average contour interval. The value must be zero or greater. The default is 2.5 if the data type is Contour and zero if the data type is Spot.	Double
Tolerance 2 (Optional)	This tolerance prevents drainage clearance through unrealistically high barriers. The value must be greater than zero. The default is 100 if the data type is Contour and 200 if the data type is Spot.	Double

Label	Explanation	Data Type
Output stream polyline features (Optional)	The output line feature class of stream polyline features and ridge line features. The line features are created at the beginning of the interpolation process. It provides the general morphology of the surface for interpolation. It can be used to verify correct drainage and morphology by comparing known stream and ridge data. The polyline features are coded as follows: 1. Input stream line not over cliff. 2. Input stream line over cliff (waterfall). 3. Drainage enforcement clearing a spurious sink. 4. Stream line determined from contour corner. 5. Ridge line determined from contour corner. 6. Code not used. 7. Data stream line side conditions. 8. Code not used. 9. Line indicating large elevation data clearance.	Feature Class
Output remaining sink point features (Optional)	The output point feature class of the remaining sink point features. These are the sinks that were not specified in the sink input feature data and were not cleared during drainage enforcement. Adjusting the values of the tolerances, Tolerance 1 and Tolerance 2, can reduce the number of remaining sinks. Remaining sinks often indicate errors in the input data that the drainage enforcement algorithm could not resolve. This can be an efficient way of detecting subtle elevation errors.	Feature Class
Output diagnostic file (Optional)	The output diagnostic file listing all inputs and parameters used and the number of sinks cleared at each resolution and iteration.	File
Output parameter file (Optional)	The output parameter file listing all inputs and parameters used, which can be used with Topo to Raster by File to run the interpolation again.	File
Profile curvature roughness penalty (Optional)	The profile curvature roughness penalty is a locally adaptive penalty that can be used to partly replace total curvature. It can yield good results with high-quality contour data but can lead to instability in convergence with poor data. Set to 0.0 for no profile curvature (the default), set to 0.5 for moderate profile	Double

Label	Explanation	Data Type
	curvature, and set to 0.8 for maximum profile curvature. Values larger than 0.8 are not recommended and should not be used.	
Output residual point features (Optional)	The output point feature class of all the large elevation residuals as scaled by the local discretisation error. All the scaled residuals larger than 10 should be inspected for possible errors in input elevation and stream data. Large-scaled residuals indicate conflicts between input elevation data and streamline data. These may also be associated with poor automatic drainage enforcements. These conflicts can be remedied by providing additional streamline and/or point elevation data after first checking and correcting errors in existing input data. Large unscaled residuals usually indicate input elevation errors.	Feature Class
Output stream and cliff error point features (Optional)	The output point feature class of locations where possible stream and cliff errors occur. The locations where the streams have closed loops, distributaries, and streams over cliffs can be identified from the point feature class. Cliffs with neighboring cells that are inconsistent with the high and low sides of the cliff are also indicated. This can be a good indicator of cliffs with incorrect direction. Points are coded as follows: 1. True circuit in data streamline network. 2. Circuit in stream network as encoded on the out raster. 3. Circuit in stream network via connecting lakes. 4. Distributaries point. 5. Stream over a cliff (waterfall). 6. Points indicating multiple stream outflows from lakes. 7. Code not used. 8. Points beside cliffs with heights inconsistent with cliff direction. 9. Code not used. 10. Circular distributary removed. 11. Distributary with no inflowing stream. 12. Rasterized distributary in output cell different to where the data stream line distributary occurs.	Feature Class

Label	Explanation	Data Type
	13. Error processing side conditions—an indicator of very complex streamline data.	
Output contour error point features (Optional)	The output point feature class of possible errors pertaining to the input contour data. Contours with bias in height exceeding five times the standard deviation of the contour values as represented on the output raster are reported to this feature class. Contours that join other contours with a different elevation are flagged in this feature class by the code 1; this is a sure sign of a contour label error.	Feature Class

How Topo to Raster works

The Topo to Raster tool is an interpolation method specifically designed for the creation of hydrologically correct digital elevation models (DEMs). It is based on the ANUDEM program developed by Michael Hutchinson (1988, 1989, 1996, 2000, 2011). See Hutchinson and Dowling (1991) and ANU Fenner School of Environment and Society and Geoscience Australia (2008) for applications of ANUDEM to continent-wide DEM production. Applications of DEMs to environmental modeling are discussed in Hutchinson and Gallant (2000) and Hutchinson (2008). Further developments of ANUDEM are discussed in Hutchinson et al. (2009, 2011). The current version of ANUDEM used in ArcGIS is 5.3.

Topo to Raster interpolates elevation values for a raster while imposing constraints that ensure:

- A connected drainage structure
- Correct representation of ridges and streams from input contour data

As such, it is the only ArcGIS interpolator specifically designed to work intelligently with contour inputs.

The Topo to Raster by File tool is useful for executing the Topo to Raster tool multiple times, since it is often easier to change a single entry in the parameter file and rerun the tool than to repopulate the tool dialog box each time.

The interpolation process

The interpolation procedure has been designed to take advantage of the types of input data commonly available and the known characteristics of elevation surfaces. This method uses an iterative finite difference interpolation technique. It is optimized to have the computational efficiency of local interpolation methods, such as inverse distance weighted (IDW) interpolation, without losing the surface continuity of global interpolation methods, such as Kriging and Spline. It is essentially a discretized thin plate spline technique (Wahba, 1990) for which the roughness penalty has been modified to allow the fitted DEM to follow abrupt changes in terrain, such as streams, ridges and cliffs.

Water is the primary erosive force determining the general shape of most landscapes. For this reason, most landscapes have many hilltops (local maximums) and few sinks (local minimums),

resulting in a connected drainage pattern. Topo to Raster uses this knowledge of surfaces and imposes constraints on the interpolation process that results in a connected drainage structure and correct representation of ridges and streams. This imposed drainage condition produces higher accuracy surfaces with less input data. The quantity of input data can be up to an order of magnitude less than that normally required to adequately describe a surface with digitized contours, further minimizing the expense of obtaining reliable DEMs. The global drainage condition also virtually eliminates any need for editing or postprocessing to remove spurious sinks in the generated surface.

The program acts conservatively in removing sinks and will not impose the drainage conditions in locations that would contradict the input elevation data. Such locations normally appear in the diagnostic file as sinks. Use this information to correct data errors, particularly when processing large datasets.

The drainage enforcement process

The purpose of the drainage enforcement process is to remove all sink points in the output DEM that have not been identified as sinks in the input sink feature dataset. The program assumes that all unidentified sinks are errors, since sinks are generally rare in natural landscapes (Goodchild and Mark, 1987).

The drainage enforcement algorithm attempts to clear spurious sinks by modifying the DEM, inferring drainage lines via the lowest saddle point in the drainage area surrounding each spurious sink. It does not attempt to clear real sinks as supplied by the Sink function. Since sink clearance is subject to the elevation tolerance, the program is conservative when attempting to clear spurious sinks. In other words, it does not clear spurious sinks that would contradict input elevation data by more than the value of Tolerance 1.

Drainage enforcement can also be supplemented with the incorporation of stream line data. This is useful when more accurate placement of streams is required. Stream distributaries are modeled by allowing each cell to have up to two downstream directions.

The drainage enforcement can be turned off, in which case the sink clearing process is ignored. This can be useful if you have contour data of something other than elevation (for example, temperature) for which you want to create a surface.

Use of contour data

Contours were originally the most common method for storage and presentation of elevation information. Unfortunately, this method is also the most difficult to properly utilize with general interpolation techniques. The disadvantage lies in the undersampling of information between contours, especially in areas of low relief.

At the beginning of the interpolation process, Topo to Raster uses information inherent in the contours to build an initial generalized drainage model. This is done by identifying the points of local maximum curvature in each contour. A network of curvilinear streams and ridges intersecting these points is then derived using the initial elevation grid (Hutchinson, 1988). The locations of these lines are iteratively updated as the DEM elevations are iteratively updated. This information is used to ensure proper hydrogeomorphic properties of the output DEM and may also be used to verify accuracy of the output DEM.

The contour data points are also used in the interpolation of elevation values at each cell. All contour data are read and generalized. A maximum of 100 data points are read from the contours within each cell, with the average elevation value used as the unique elevation data point for each cell intersecting the contour line data. At each DEM resolution, only one critical point is used for each cell. For this reason, having a contour density with several contours crossing output cells is redundant.

After the general morphology of the surface has been determined, contour data is also used in the interpolation of elevation values at each cell.

When the contour data is used to interpolate elevation information, all contour data is read and generalized. A maximum of 50 data points are read from these contours within each cell. At the final resolution, only one critical point is used for each cell. For this reason, having a contour density with several contours crossing output cells is redundant.

Use of lake data

Lake polygons in earlier versions of Topo to Raster were simple masks that set the elevation of each lake surface to the minimum elevation of all DEM values immediately neighboring the lake. The lake boundary algorithm has been upgraded to enable automatic determination of lake heights that are fully compatible with connecting stream lines and neighboring elevation values.

The revised lake boundary method also treats each lake boundary as a contour with unknown elevation and iteratively estimates the elevation of this contour from the cell values on the lake boundary. At the same time, the elevation of each lake boundary is made to conform to the elevations of any upstream and downstream lakes. The elevation of each lake boundary is also made to be consistent with the neighboring DEM values. Cell values immediately outside the lake are made to lie above the elevation of the lake boundary and cell values on the interior of the lake made to lie below the elevation of the lake boundary.

Lake boundaries are permitted to include islands within lakes, and lakes within islands. All DEM values that lie inside lakes, as determined by the lake boundary polygons, are set to the estimated height of the DEM on the boundary of the lake.

Use of cliff data

Cliff lines permit a complete break in continuity between neighboring cell values on each side of the data cliff lines, as they are encoded into the output raster. Cliff lines must be supplied as directed lines, with the low side of each cliff line on the left and the high side of the cliff line on the right. This permits removal of elevation data points that lie on the wrong side of the cliffs, as they are encoded onto the raster, and better placement of cliffs in relation to streamlines.

It has also been found that the minor shifts in position that are imposed on streams and cliffs as they are incorporated into the raster can lead to spurious interactions between these data. An automated method has therefore been developed to make small adjustments in the placement of both streams and cliff lines in the raster to minimize these spurious interactions.

Use of coastline data

Cells in the final output DEM that lie outside the polygons specified by this polygon feature class are set to an internally determined special value that is less than the user-specified minimum height limit. The result of this is that a complete coastal polygon can be used as input and it will automatically be clipped to the processing extent.

Multi-resolution interpolation

The program uses a multi resolution interpolation method, starting with a coarse raster and working toward the finer, user-specified resolution. At each resolution, drainage conditions are enforced, interpolation is performed, and the number of remaining sinks is recorded in the output diagnostic file.

Processing stream data

The Topo to Raster tool requires that stream network data has all arcs pointing downslope and that there are no polygons (lakes) in the network.

The stream data should be composed of single arcs in a dendritic pattern, with any parallel stream banks, lake polygons, and so on, cleaned up by interactive editing. When editing lake polygons out of the network, a single arc should be placed from the beginning to the end of the impounded area. The arc should follow the path of an historic streambed if one is known or exists. If the elevation of the lake is known, the lake polygon and its elevation can be used as a Contour input.

Creating and mosaicking adjacent rasters

Sometimes it's necessary to create DEMs from adjacent tiles of input data. Normally this happens when input features are derived from a map sheet series or when, due to memory limitations, the input data must be processed in several pieces.

The interpolation process uses input data from surrounding areas to define the morphology and drainage of the surface and interpolate output values. However, the cell values at the edges of any output DEM are not as reliable as in the central area because they are interpolated with one-half as much information.

To make the most accurate predictions at the edges of the area of interest, the extent of the input datasets should be greater than the area of interest. The Margin in cells parameter provides a method for trimming the edges of output DEMs based on a user-specified distance. The edges of overlapping areas should be at least 20 cells wide.

There should be some overlap of input data into the adjacent areas when multiple output DEMs are to be combined into a single raster. Without this overlap, the edges of merged DEMs may not be smooth. The extents of the input datasets from each of the interpolations should have an even larger area than if only an interpolation for a single interpolation were to be done, so as to ensure that the edges can be predicted as accurately as possible.

When the DEMs have been created, they can best be combined using the geoprocessing Mosaic tool with the Blend or Mean options. This function provides options for handling overlapping areas to smooth the transition between datasets.

Evaluating output

Every created surface should be evaluated to ensure that the data and parameters supplied to the program resulted in a realistic representation of the surface. There are many ways to evaluate the quality of an output surface, depending on the type of input available to create the surface.

The most common evaluation is to create contours from the new surface with the [Contour](#) tool and compare them to the input contour data. It is best to create these new contours at one-half the original contour interval to examine the results between contours. Drawing the original contours and the newly created contours on top of one another can help identify interpolation errors.

Another method of visual comparison is to compare the optional output drainage cover with known streams and ridges. The drainage feature class contains the streams and ridges that were generated by the program during the drainage enforcement process. These streams and ridges should coincide with known streams and ridges in the area. If a stream feature class was used as input, the output streams should almost perfectly overlay the input streams, although they may be slightly more generalized.

A common method for evaluating the quality of a generated surface is to withhold a percentage of the input data from the interpolation process. After generating the surface, the height of these known points can be subtracted from the generated surface to examine how closely the new surface represents the true surface. These differences can be used to calculate a measure of error for the surface, such as the root mean squared (RMS) error.

Topo to Raster has a comprehensive set of procedures for assessing the quality of the fitted DEM, for optimizing DEM resolution, and for detecting errors in the input data.

- The optional Output diagnostic file can be used to evaluate how effectively the tolerance settings are clearing sinks in the input data. Decreasing the values of the tolerances can make the program behave more conservatively at clearing sinks.
- The Output remaining sink point feature class contains the locations of all remaining spurious sinks. It should be inspected together with the output stream polyline features to check for errors in all input topographic data.
- The Output residual point feature class contains the locations of all large elevation data residuals as scaled by the local discretisation error. Large scaled residuals indicate conflicts between input elevation data and stream line data. These may also be associated with poor automatic drainage enforcements. These conflicts can be remedied by providing additional streamline and/or point elevation data after first checking and correcting errors in existing input data. Large unscaled residuals usually indicate input elevation errors.
- The Output contour error point feature class contains the locations of points on input contours with significantly biased residuals from the fitted DEM. An ErrorValue of 1 most often indicates the location of points where contours with different elevations are connected, a sure indicator of a contour label error.
- The Output stream and cliff error point feature class is an important indicator of stream line and cliff line data quality, particularly of stream direction errors and cliff direction errors, and it should always be inspected.

The feature class has the following codes:

1. True circuit in data streamline network.
2. Circuit in stream network as encoded on the out raster.
3. Circuit in stream network via connecting lakes.
4. Distributaries point.
5. Stream over a cliff (waterfall).
6. Points indicating multiple stream outflows from lakes.
7. Code not used.
8. Points beside cliffs with heights inconsistent with cliff direction.
9. Code not used.
10. Circular distributary removed.
11. Distributary with no inflowing stream.
12. Rasterized distributary in output cell different to where the data stream line distributary occurs.
13. Error processing side conditions—an indicator of very complex streamline data.

- The Output stream polyline feature class contains all drainage constraints imposed by Topo to Raster as determined from input stream line data, stream lines and ridge lines inferred from contour data and stream lines obtained by automatic drainage enforcement. These can be inspected to check for location errors in input stream lines and to verify appropriate concordance with constraints associated with input stream lines and automated drainage enforcements. Each type of derived stream line is given a different code. Stream lines that cross cliff lines are indicated by short stream lines of length, one cell with a separate code. The feature class also includes lines flagging large source elevation data clearances via connecting stream lines and lakes that exceed the second elevation tolerance. These can be a useful indicator of source elevation data errors.

The polyline features are coded as follows:

1. Input stream line not over cliff.

2. Input stream line over cliff (waterfall).

3. Drainage enforcement clearing a spurious sink.

4. Stream line determined from contour corner.

5. Ridge line determined from contour corner.

6. Code not used.

7. Data stream line side conditions.

8. Code not used.

9. Line indicating large elevation data clearance.

Contour biasing

There is a minor bias in the interpolation algorithm that causes input contours to have a stronger effect on the output surface at the contour. This bias can result in a slight flattening of the output surface as it crosses the contour. This may result in misleading results when calculating the profile curvature of the output surface but is otherwise not noticeable.

Likely causes of problems with Topo to Raster

If you encounter any problems when running Topo to Raster, check the following points for explanations and solutions for the most commonly encountered issues.

- There are insufficient system resources available. The algorithms used in Topo to Raster hold as much information as possible in memory during processing. This allows point, contour, sink, stream, and lake data to be accessed simultaneously. To facilitate processing of large datasets, it is recommended that unnecessary applications be closed before running the tool to free up physical RAM. It is also important to have sufficient amounts of system swap space on disk.
- The contour or point input may be too dense for the output cell size specified. If one output cell covers several input contours or points, the algorithm may not be able to ascertain a value for that cell. To resolve this, try any of the following:
    - Decrease the cell size, then resample back to the larger cell size after Topo to Raster.
    - Rasterize smaller sections of the input data using the Output extent and Margin in cells. Assemble the resulting component rasters with the Mosaic tool.
    - Clip the input data into overlapping sections, and run Topo to Raster separately on each section. Assemble the resulting component rasters with the Mosaic tool.
- The application of a surface interpolator may not be consistent with the input dataset. For example, if there is a sinks input with more points than there would be cells in the output raster, the tool will fail. Densely sampled data sources, such as lidar data, may have similar problems. Using the Do not enforce option may help in this case, but a proper understanding of how the interpolator works is important to prevent misapplication.

References

ANU Fenner School of Environment and Society and Geoscience Australia, 2008. GEODATA 9 Second DEM and D8 Digital Elevation Model and Flow Direction Grid, User Guide. Geoscience Australia, 43 pp. See: http://www.ga.gov.au/image_cache/GA11644.pdf.

Goodchild, M. F., and D. M. Mark. 1987. The fractal nature of geographic phenomena. Annals of Association of American Geographers. 77 (2): 265–278.

Hutchinson, M. F. 1988. Calculation of hydrologically sound digital elevation models. Paper presented at Third International Symposium on Spatial Data Handling at Sydney, Australia.

Hutchinson, M. F. 1989. A new procedure for gridding elevation and stream line data with automatic removal of spurious pits. Journal of Hydrology, 106: 211–232.

Hutchinson, M. F., and T. I. Dowling. 1991. A continental hydrological assessment of a new grid-based digital elevation model of Australia. Hydrological Processes 5: 45–58.

Hutchinson, M. F. 1993. Development of a continent-wide DEM with applications to terrain and climate analysis. In Environmental Modeling with GIS, ed. M. F. Goodchild et al., 392–399. New York: Oxford University Press.

Hutchinson, M. F. 1996. A locally adaptive approach to the interpolation of digital elevation models. In Proceedings, Third International Conference/Workshop on Integrating GIS and Environmental Modeling. Santa Barbara, CA: National Center for Geographic Information and Analysis. See: http://www.ncgia.ucsb.edu/conf/SANTA_FE_CD-ROM/sf_papers/hutchinson_michael_dem/local.html.

Hutchinson, M.F. 2000. Optimising the degree of data smoothing for locally adaptive finite element bivariate smoothing splines. ANZIAM Journal 42(E): C774–C796.

Hutchinson, M.F. and Gallant, J.C. 2000. Digital elevation models and representation of terrain shape. In: J.P. Wilson and J.C. Gallant (eds) Terrain Analysis. Wiley, New York, pp. 29–50.

Hutchinson, M.F. 2008. Adding the Z-dimension. In: J.P. Wilson and A.S. Fotheringham (eds), Handbook of Geographic Information Science, Blackwell, pp 144–168.

Hutchinson, M.F., Stein, J.A., Stein, J.L. and Xu, T. 2009. Locally adaptive gridding of noisy high resolution topographic data. In Anderssen, R.S., R.D. Braddock and L.T.H. Newham (eds) 18th World IMACS Congress. Modelling and Simulation Society of Australia and New Zealand and International Association for Mathematics and Computers in Simulation, July 2009, pp. 2493–2499. See: http://www.mssanz.org.au/modsim09/F13/hutchinson.pdf.

Hutchinson, M.F., Xu, T. and Stein, J.A. 2011. Recent Progress in the ANUDEM Elevation Gridding Procedure. In: Geomorphometry 2011, edited by T. Hengel, I.S. Evans, J.P. Wilson and M. Gould, pp. 19–22. Redlands, California, USA. See: http://geomorphometry.org/HutchinsonXu2011.

Wahba, G. 1990. Spline models for Observational data. Paper presented at CBMS-NSF Regional Conference Series in Applied Mathematics. Philadelphia: Soc. Ind. Appl. Maths.

Trend (3D Analyst)

Summary

Interpolates a raster surface from points using a trend technique.

[Learn more about how Trend works](#)

Usage

- As the order of the polynomial is increased, the surface being fitted becomes progressively more complex. A higher-order polynomial will not always generate the most accurate surface; it is dependent on the data.
- For the Logistic option of Type of regression, the z-value field of input point features should have codes of zero (0) and one (1).
- The Output cell size parameter can be defined by a numeric value or obtained from an existing raster dataset. If the cell size hasn't been explicitly specified as the parameter value, it is derived from the Cell Size environment if it has been specified. If the parameter cell size or the environment cell size have not been specified, but the Snap Raster environment has been set, the cell size of the snap raster is used. If nothing is specified, the cell size is calculated from the shorter of the width or height of the extent divided by 250 in which the extent is in the output coordinate system specified in the environment.
- If the cell size is specified using a numeric value, the tool will use it directly for the output raster.

  If the cell size is specified using a raster dataset, the parameter will show the path of the raster dataset instead of the cell size value. The cell size of that raster dataset will be used directly in the analysis, provided the spatial reference of the dataset is the same as the output spatial reference. If the spatial reference of the dataset is different than the output spatial reference, it will be projected based on the selected Cell Size Projection Method value.

- The optional RMS file output contains information on the RMS (root mean square) error of the interpolation. This information can be used to determine the best value to use for the polynomial order, by changing the order value until you get the lowest RMS error. See [How Trend works](#) for information on the RMS file.
- Some input datasets may have several points with the same x,y coordinates. If the values of the points at the common location are the same, they are considered duplicates and have no effect on the output. If the values are different, they are considered coincident points.

  The various interpolation tools may handle this data condition differently. For example, in some cases, the first coincident point encountered is used for the calculation; in other cases, the last point encountered is used. This may cause some locations in the output raster to have different values than what you might expect. The solution is to prepare your data by removing these coincident points. The [Collect Events](#) tool in the Spatial Statistics toolbox is useful for identifying any coincident points in your data.

- For data formats that support Null values, such as file geodatabase feature classes, a Null value will be ignored when used as input.

Parameters

DialogPython

Label	Explanation	Data Type
Input point features	The input point features containing the z-values to be interpolated into a surface raster.	Feature Layer
Z value field	The field that holds a height or magnitude value for each point. This can be a numeric field or the Shape field if the input point features contain z-values. If the regression type is Logistic, the values in the field can only be 0 or 1.	Field
Output raster	The output interpolated surface raster. It is always a floating-point raster.	Raster Dataset
Output cell size (Optional)	The cell size of the output raster that will be created. This parameter can be defined by a numeric value or obtained from an existing raster dataset. If the cell size hasn't been explicitly specified as the parameter value, the environment cell size value will be used if specified; otherwise, additional rules will be used to calculate it from the other inputs. See the usage section for more detail.	Analysis Cell Size
Polynomial order (Optional)	The order of the polynomial. This must be an integer between 1 and 12. A value of 1 will fit a flat plane to the points, and a higher value will fit a more complex surface. The default is 1.	Long
Type of regression (Optional)	The type of regression to be performed. • Linear—Polynomial regression is performed to fit a least-squares surface to the set of input points. This is applicable for continuous types of data. • Logistic—Logistic trend surface analysis is performed. It generates a continuous probability surface for binary, or dichotomous, types of data.	String
Output RMS file (Optional)	File name for the output text file that contains information about the RMS error and the Chi-Square of the interpolation. The extension must be .txt.	File

How Trend works

The [Trend](#) tool uses a global polynomial interpolation that fits a smooth surface defined by a mathematical function (a polynomial) to the input sample points. The trend surface changes gradually and captures coarse-scale patterns in the data.

Conceptual background

Conceptually, trend interpolation is like taking a piece of paper and fitting it between raised points (raised to the height of value). This is demonstrated in the diagram below for a set of sample points of elevation taken on a gently sloping hill. The piece of paper is magenta.
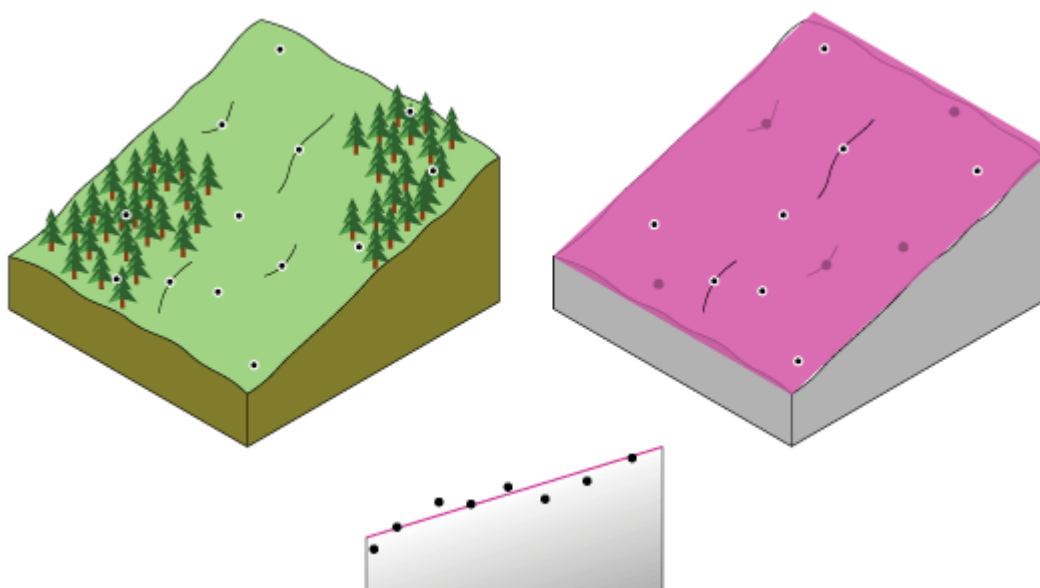
Illustration of a flat trend surface

A flat piece of paper will not accurately capture a landscape containing a valley. However, if you bend the piece of paper once, you get a better fit. Adding a term to the mathematical formula produces a similar result, a bend in the plane. A flat plane (no bend in the piece of paper) is a first-order polynomial (linear). Allowing one bend is a second-order polynomial (quadratic), two bends a third-order (cubic), and so forth. A maximum of 12 bends (twelfth order) are allowed with this tool. The following image conceptually demonstrates a second-order polynomial fitted to a valley.
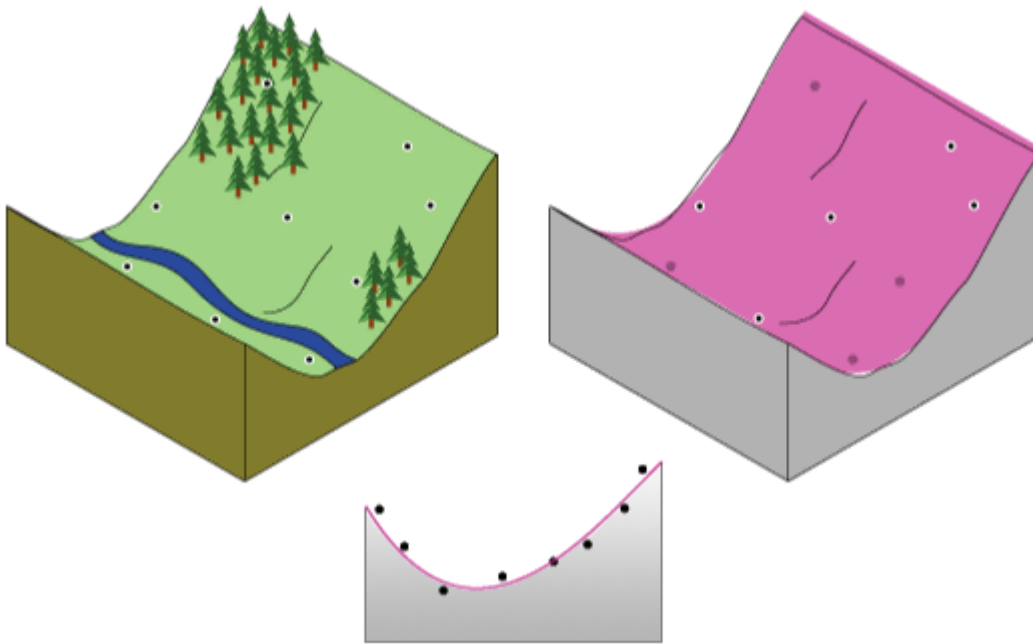
Illustration of a curved trend surface

Rarely will the piece of paper pass through the actual measured points, thus making trend interpolation an inexact interpolator. Some points will be above the piece of paper, and others will be below. However, if you add up how much higher each point is above the piece of paper and add up how much lower each point is below the piece of paper, the two sums should be similar. The surface, given in magenta, is obtained by using a least-squares regression fit. The resulting surface minimizes the squared differences among the raised values and the sheet of paper.

The lower the root mean square (RMS) error, the more closely the interpolated surface represents the input points. The most common orders of polynomials are one through three. Trend surface interpolation creates smooth surfaces.

When to use trend interpolation

Trend interpolation results in a smooth surface that represents gradual trends in the surface over the area of interest. This type of interpolation can be used for:

- Fitting a surface to the sample points when the surface varies gradually from region to region over the area of interest—for example, pollution over an industrial area.
- Examining or removing the effects of long-range or global trends. In such circumstances, the technique is often referred to as trend surface analysis.

Trend interpolation creates a gradually varying surface using low-order polynomials that describe a physical process—for example, pollution and wind direction. However, the more complex the polynomial, the more difficult it is to ascribe physical meaning to it. Furthermore, the calculated surfaces are highly susceptible to outliers (extremely high and low values), especially at the edges.

Types of trend interpolation

There are two basic types of Trend interpolation: Linear and Logistic.

Linear trend

The Linear trend surface interpolator creates a floating-point raster. It uses a polynomial regression to fit a least-squares surface to the input points. This method allows you to control the order of the polynomial used to fit the surface. To understand this method, consider a first-order polynomial. A first-order linear trend surface interpolation performs a least-squares fit of a plane to the set of input points.

Trend surface interpolation creates smooth surfaces. The surface generated seldom passes through the original data points, since it performs a best fit for the entire surface. When a polynomial order higher than one is used, the interpolator may generate a raster whose minimum and maximum exceed the minimum and maximum of the input file of the input feature data.

Logistic trend

The Logistic option for generating a trend surface is appropriate for prediction of the presence or absence of certain phenomena (in the form of probability) for a given set of locations (x,y) in space. The z-value is a categorized random variable with only two possible outcomes—for example, the existence of an endangered species or the lack of existence of that species. These two z-values can be coded as one and zero, respectively. This option creates a continuous probability grid with cell values between one and zero.

A maximum likelihood estimation is used to calculate the nonlinear probability surface model without first converting the model into linear form.

Output RMS file

The RMS error file contains the root mean square error of the interpolation by comparing the value of the locations in the input dataset against the value of those same locations in the interpolated raster surface.

The RMS error value can be used to determine the best value to use for the {order} parameter of the interpolation by changing the order value until you get the lowest RMS error. The Chi-square value is also reported.

Example

An example of the output RMS file from a run of the Trend tool with the Polynomial order set to 3 is:

```
coef #      coef
------ ----------------
   0 -1192066.7888371
   1 -1.78479492586755
   2 -0.195982103615487
   3 -8.87072249743903e-1
   4 -2.0538267625596e-1
   5 -3.85610088343239e-1
   6 -1.46420255709888e-2
   7 -5.31539027745154e-2
   8 -2.59261094879031e-3
   9 9.71651459136166e-4
------ ----------------
RMS Error  = 296.957857221845
Chi-Square  = 17019506.0103975
```

The predicted value at any location in the output raster for this third-order trend can be determined by summing up the values that result from a series of equations. For these equations, x = the longitude of the location, y = the latitude, and the c_n term is the coefficient value from the preceding table. The equations for this third-order interpolation are:

$$
\begin{aligned}
\text{Prediction}(x,y) = c_0 + \\
x \cdot c_1 + y \cdot c_2 + \\
x^2 \cdot c_3 + x \cdot y \cdot c_4 + y^2 \cdot c_5 + \\
x^3 \cdot c_6 + x^2 \cdot y \cdot c_7 + x \cdot y^2 \cdot c_8 + y^3 \cdot c_9
\end{aligned}
$$

Essentially, coefficient 0 is always the intercept. From there, iterate through the first-order terms starting with the highest x and ending with no xs. Move to the second-order terms and iterate down the xs again, and similarly for the third-order terms.

Note:

These equations follow a similar pattern as the order increases. For example, if the fourth order was specified, there would be five more values specified in the RMS table (coefficients 10, 11, 12, 13 and 14) and commensurately more equations that utilize them.