It's a Model

Quantifying uncertainty in elevation models using kriging

By Konstantin Krivoruchko and Kevin Butler, Esri

Raster based digital elevation models (DEM) are the basis of some of the most important GIS workflows: hydrologic modeling, site suitability, and cost path analysis. While there are several techniques for generating digital elevation models (DEMs), none of them can produce a true elevation surface. Locally varying measurement error and the inexactness of the interpolation methods contribute to the uncertainty of the model's estimate of the true elevation value. Kriging models and geostatistical simulations available in the Geostatistical Analyst extension for ArcGIS 10.1 for Desktop to quantify the spatially varying uncertainty of a DEM derived from lidar data.

Creating a DEM from Lidar Data

Lidar data is increasingly used as the base data for DEM creation. Topographic mapping lidar instruments direct laser pulses from an airborne platform toward the ground and record the time required for the pulse to return. Combined with accurate information about the position, roll, pitch and heading of the lidar instrument, accurate three-dimensional points are generated for any reflective surface the laser pulses hits. This includes above ground features such as buildings and trees. Through post-processing, lidar points are classified by the type of object that reflected the laser pulse such as bare earth, tree canopy, or water. For DEM creation, all above ground features are removed. This causes the sampling density of the bare earth lidar points to vary widely based on the frequency of above ground features.

The next step in the DEM creation process involves generating a surface from the bare earth points. ArcGIS for Desktop has a powerful set of tools for creating raster DEMs from lidar points. (See the ArcGIS 10.1 help topic Creating raster DEMs and DSMs from large lidar point collections.)

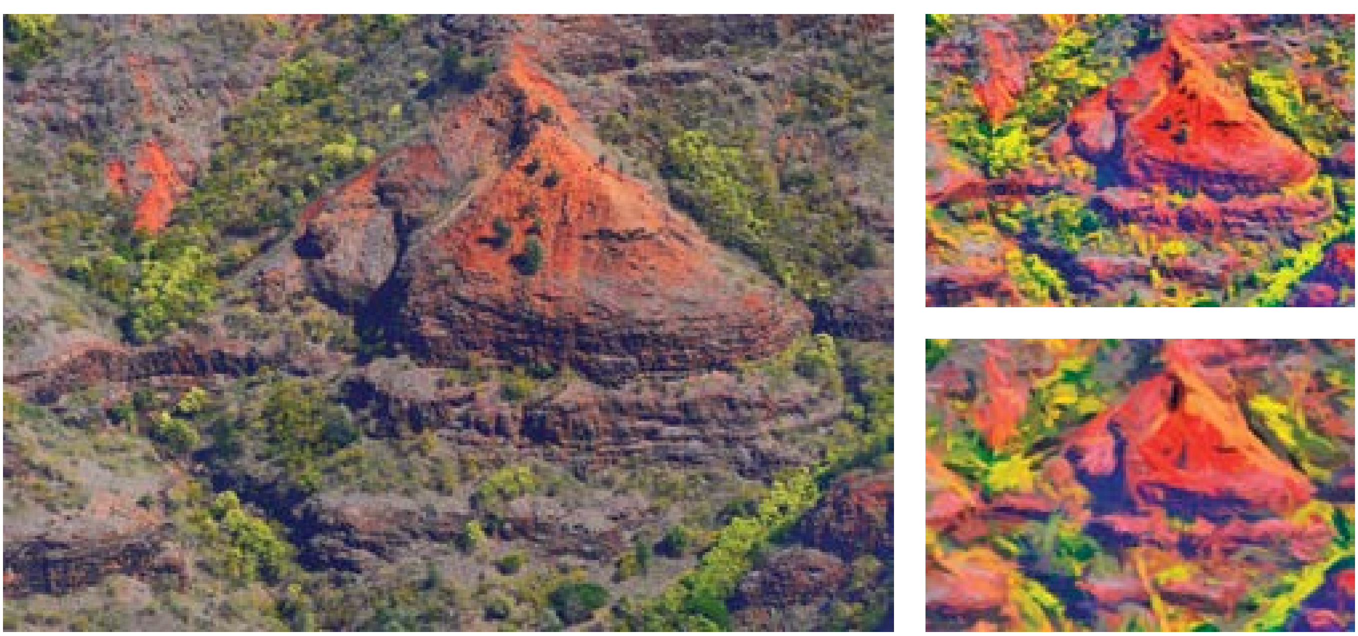
These tools use simple rules for assigning values to the raster cells such as the maximum or average elevation or use deterministic interpolation techniques to create an elevation surface from the lidar points. They produce elevation surfaces suitable for many GIS applications that are not sensitive to DEM uncertainty. GIS analysts often choose

↓ Figure 1a: Landscape represented using a small cell size.

↓ Figure 1b: Somewhat generalized landscape resulting from a moderate cell size.

↓ Figure 1c: Highly abstract landscape resulting from a large cell size.

a



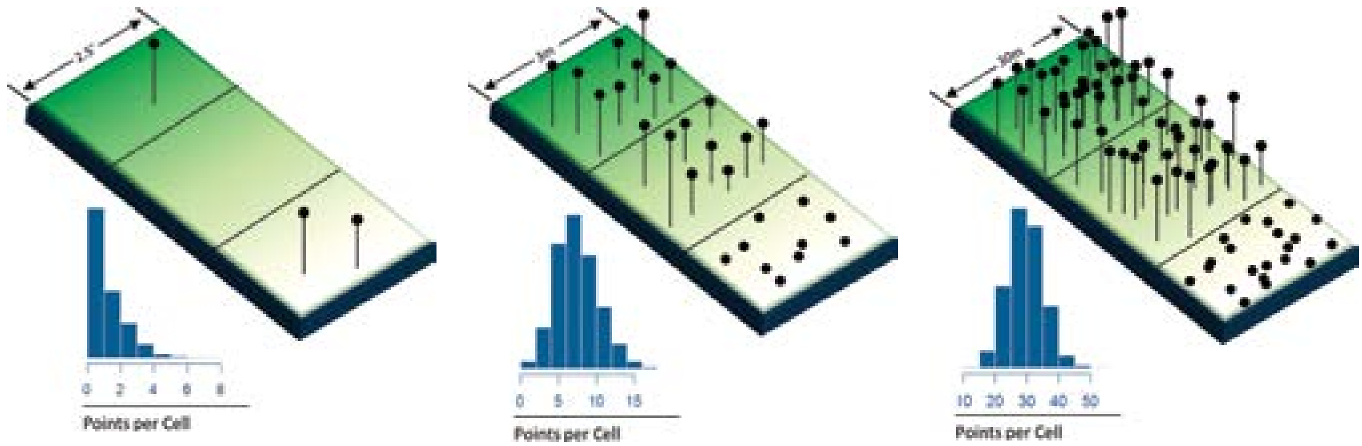
b

c

Case 1: Low number of points per raster cell

Case 2: Moderate number of points per raster cell

Case 3: Large number of points per raster cell



↑ Figure 2:Differences in the number of points per raster cell require different methods for DEM generation. Kriging is sufficient for cells with a small number of points. Block kriging can be used for moderate numbers of points. Geostatistical conditional simulations are required when cells have a large number of points.

deterministic interpolation techniques because they are based on simple formulas that don't tax computer resources and can handle a very large number of points.

An alternative to deterministic algorithms, probablistic statistical interpolation methods such as kriging, have several advantages over deterministic methods. "Empirical Bayesian Kriging: Implemented in ArcGIS Geostatistical Analyst," an article in the Fall 2012 issue of *ArcUser* magazine discusses these advantages in detail.

Historically, it was difficult to implement kriging on the massive number of points generated using lidar. Most implementations of kriging methods required manually adjusting parameters in order to receive accurate results. An important part of any interpolation workflow is to select an appropriate cell size for the output DEM raster. (Figure 1 illustrates the effects of cell size using a landscape painting displayed using different cells sizes.) Cell size determines the level of detail the raster can represent. A small cell size can represent more detail but will consume additional computer resources. A larger cell size consumes less computer resources but can overgeneralize the elevation surface. ArcGIS 10.1 introduced a new method, empirical Bayesian kriging (EBK) that automates some of the difficult aspects of building an optimal kriging model.

Generating DEMs Using Geostatistical Interpolation and Simulations

The sampling density of bare earth lidar points will vary greatly based on the frequency of above ground features such as trees or buildings. Lidar density and the desired output cell size will result in raster cells with different numbers of lidar points.

Figure 2 illustrates the effects of these differences in point density. When the number of lidar points in each raster cell is low (on average one to two points per raster cell), it is sufficient to use one kriging prediction per the cell because additional predictions and their errors should be very similar.

When the lidar point density relative to the target raster resolution is moderate (5 to 25 points per cell), it is better to use block kriging, which makes several predictions in the grid cell using the same searching neighborhood without additional model fitting.

When the lidar point density relative to the target raster resolution is high (on average greater than 30 points per raster cell), the elevation inside the cell may vary significantly. Accurately averaging many predictions, that use a variety of search neighborhoods, is required. This can be done using conditional geostatistical simulations.

Importance of Estimating DEM Uncertainty

It is common practice in GIS workflows to treat a DEM not as a model but as the true elevation surface. The following five use cases demonstrate how elevation uncertainty can dramatically impact the results of an analysis.

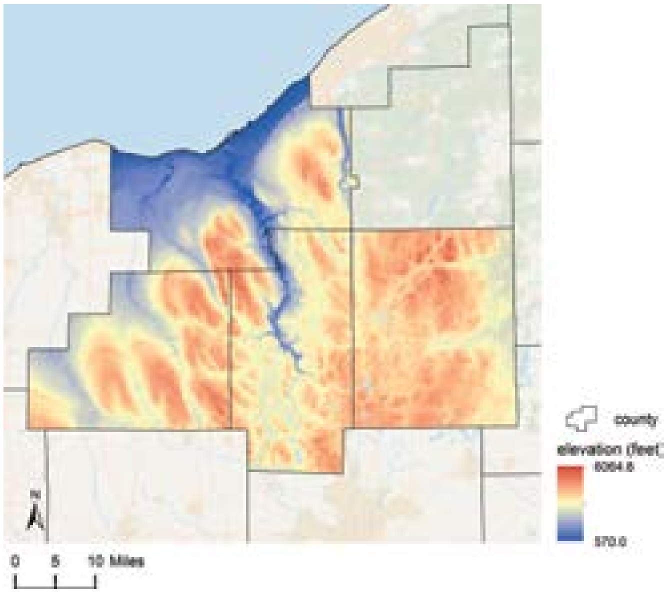
Siting a new fire lookout tower

A park manager used a GIS-based suitability model to choose ten possible locations for a new fire lookout tower. Even a small elevation error at a candidate sight could greatly change the effective viewing area for the tower. To evaluate each location, the manager needs to know the exact elevation at each candidate site.

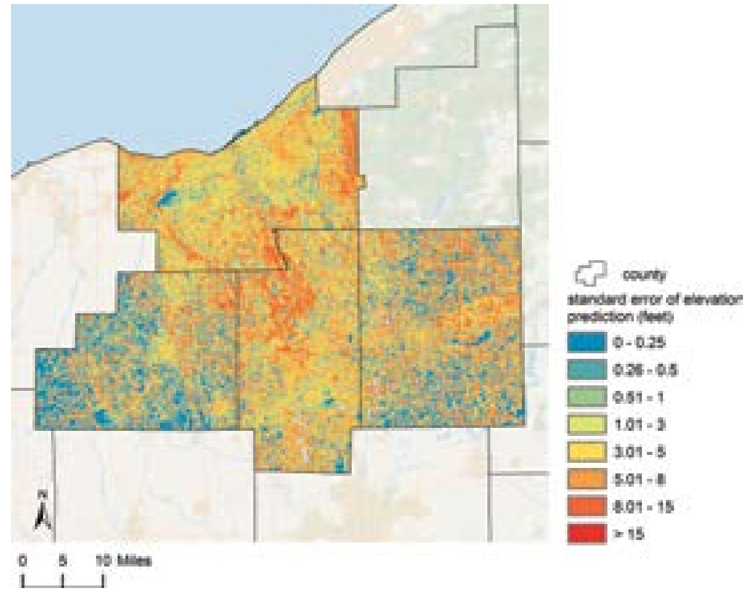
A DEM can't answer this question exactly. The value of the cell containing a candidate site is an averaged, locally smoothed or is a chosen value (i.e., minimum, maximum, or random) from the many different possible elevations that fall within the cell's extent. While kriging methods couldn't answer the elevation question either, these methods could tell that park manager that the candidate site would fall within a specific elevation range within a degree of certainty, for example, of 90 percent.

Delineating a watershed

A hydrologist is delineating a watershed and needs to identify and fill sinks to create a depressionless DEM. To help determine reasonable elevation differences between sinks and their pour points, the hydrologist needs to know the lowest elevation in the landscape. Kriging can tell the hydrologist, with a specific level of certainty, that the



↑ Figure 3a: Predicted elevation surface for northeastern Ohio created using kriging.



↑ Figure 3b: Standard error of prediction surface.

lowest elevation is smaller than some value. For example, by using kriging, the hydrologist might learn in this case that the lowest elevation is smaller than 800 feet with 99 percent certainty.

Siting a cell phone tower

An engineer, who is siting a new cell phone tower near an airport, has chosen a candidate site. The cell phone tower is 50 feet tall and the site has an elevation of 950 feet. Strict zoning regulations prohibit any structure near the airport that is more than 1,000 feet in height. The engineer needs to know the probability that the elevation of the candidate site is greater than 950 feet. Kriging methods will provide a specific probability that a cell's elevation exceeds a certain height at the proposed site or any other site in the raster.

Optimizing irrigation system design

To minimize water usage, a golf course manager used a GIS model to determine the optimal layout for an irrigation system. The model is dependent on an accurate flow direction raster. Examining the standard error of prediction surface created through kriging, the manager could identify areas where the uncertainty of the elevation was high, which would influence the accuracy of the flow direction. The manager could use these areas to assess the ground truth of the GIS model.

Assessing maximum elevation for a region

Based on the assumption that the highest elevation in the region is 1,000 feet, a military analyst has located an asset in a relatively hidden location. The analyst wants to know the probability that the elevation exceeds 1,000 feet for at least at one point in the region. Kriging methods will provide a specific probability that the elevation of 1,000 feet is exceeded.

Creating an Elevation Prediction Error Map using EBK

An elevation prediction map and prediction error map were generated for a four county area in northeast Ohio using the empirical Bayesian kriging (EBK) geoprocessing tool in ArcGIS 10.1. A model, which filters out local trends, intrinsic random function kriging, was used. Using the LAS to Multipoint tool from the 3D Analyst toolbox, approximately 842 million bare earth lidar points were converted to a multipoint feature class. This feature class was used as the input for kriging.

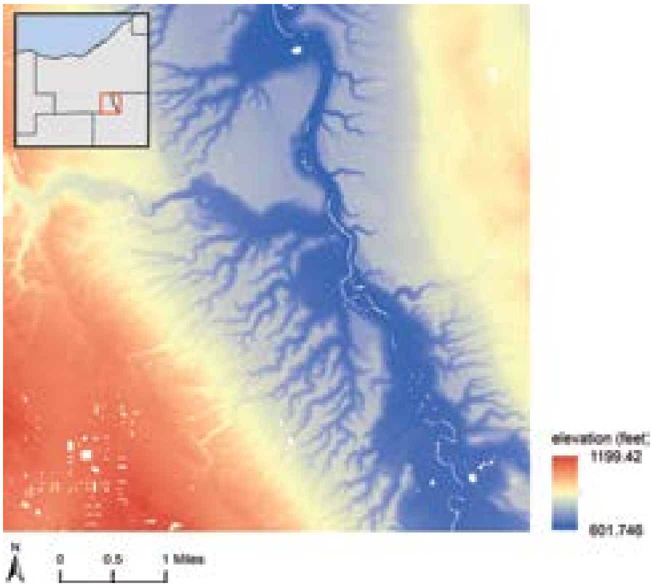
Figure 3a shows the predicted elevation surface for the glaciated topography of northeastern Ohio. Figure 3b shows a standard error of prediction surface, also produced using kriging. It is clear that the prediction errors are spatially non-homogeneous. Large portions of the Cuyahoga River valley have significant elevation errors (as estimated by kriging) in the range of 8 to 15 feet. Here, elevation is changing rapidly and abruptly thus promoting uncertainty in the elevation prediction.

Exploring the prediction error at a smaller scale demonstrates more clearly how the errors are spatially non-homogeneous. Figures 4a and 4b show the predicted elevation and prediction error for a portion of the highly eroded Cuyahoga River valley and its surrounding area. Interestingly, the topography of the area can be seen as clearly in the error surface as in the predicted surface. This example demonstrates the importance of spatially varying DEM error for hydrological applications.

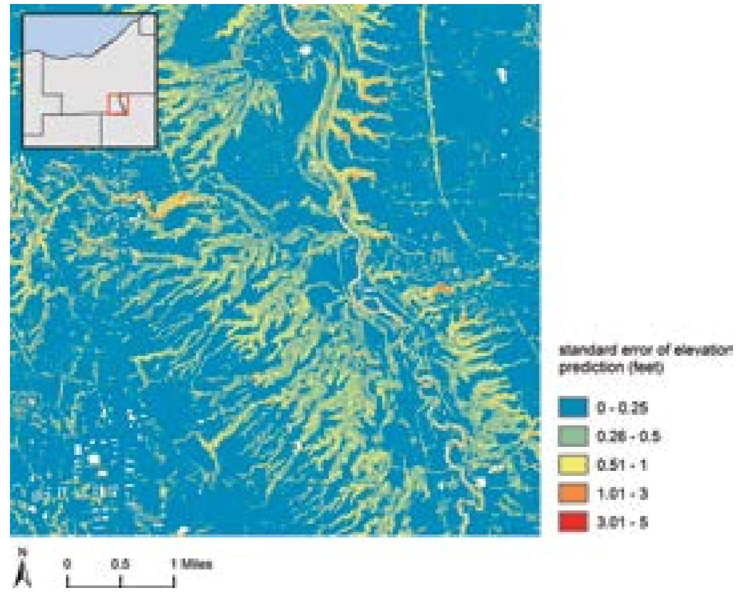
Incorporating Uncertainty in GIS Workflows

The prediction standard errors quantify the uncertainty of the elevation at each location. Assuming that predicted elevation values follow a normal distribution, 95 percent of the time the true elevation value will fall within the range of (predicted elevation value - 1.96 * prediction standard error) to (predicted elevation value + 1.96 * predicted standard error).

The DEM prediction normality assumption is usually very reasonable. However, for very important applications, the input data can be transformed to a normal distribution to guarantee that the normality assumption holds. For a particular raster cell, suppose the estimated elevation value is 1,000 feet and the standard error of prediction is 10 feet. With 95 percent confidence, the true elevation value would be somewhere in the



↑ Figure 4a: Predicted elevation surface for a portion of the Cuyahoga River valley.



↑ Figure 4b: Standard error of prediction surface for a portion of the Cuyahoga River valley.

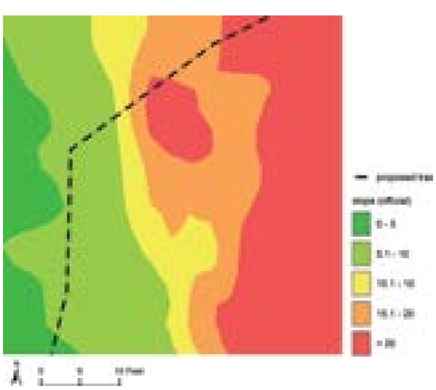
range of 980.4 feet to 1,019.6 feet.

One way to incorporate uncertainty into a GIS workflow is to use the Raster Calculator tool to adjust the elevation estimates by their prediction errors. This workflow involves running an analysis using the predicted elevation DEM, adjusting the elevation raster by the prediction standard errors, and running the analysis again. The decision to add or subtract the prediction standard error values depends on the type of problem. For a viewshed application, it might be best to increase the elevation values by the prediction error to investigate the worst case scenario. A hydrological application for identifying sinks might require a decrease the elevation values.

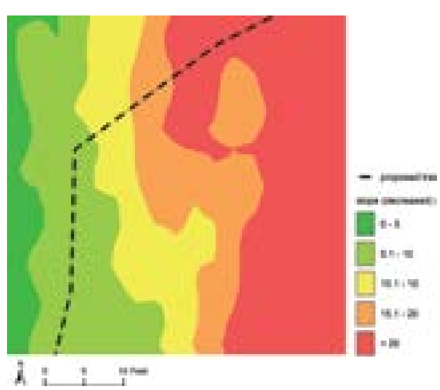
Note that random generation from the Gaussian distribution with mean value equals kriging prediction and standard deviation equals prediction standard error *should not* be used because this will destroy the spatial correlation between neighboring locations. Any potential error in elevation will be propagated to products derived from elevation such as slope and aspect. These errors can influence spatial decision making.

A modeling case study demonstrates the consequencs of treating a model as a true elevation surface. Suppose a regional park system has proposed a new recreation trail. To maintain the integrity of the trail and the safety of park patrons, park managers concluded that erosion control measures should be constructed anywhere along the trail where the slope of the adjacent embankment is greater than 20 degrees.

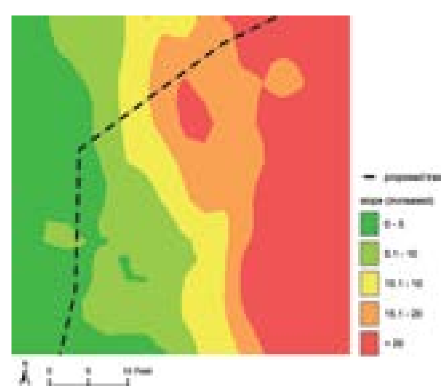
Figure 5a shows the slope derived from the publically available DEM that was derived from lidar data. Based on this map, a portion of the trail would require erosion control measures. However, when the slope is derived from elevation values decreased by the standard error (shown in Figure 5b), a much larger portion of the trail requires remediation. When the slope is derived from the elevation values increased by the standard error (shown in Figure 5c), a smaller portion of the trail would require erosion control measures. Assuming that the DEM represents the true elevation surface could potentially cause park managers to incorrectly estimate the cost of the erosion control measures and, in a worst case scenario, decide that the trail is too expensive to construct.



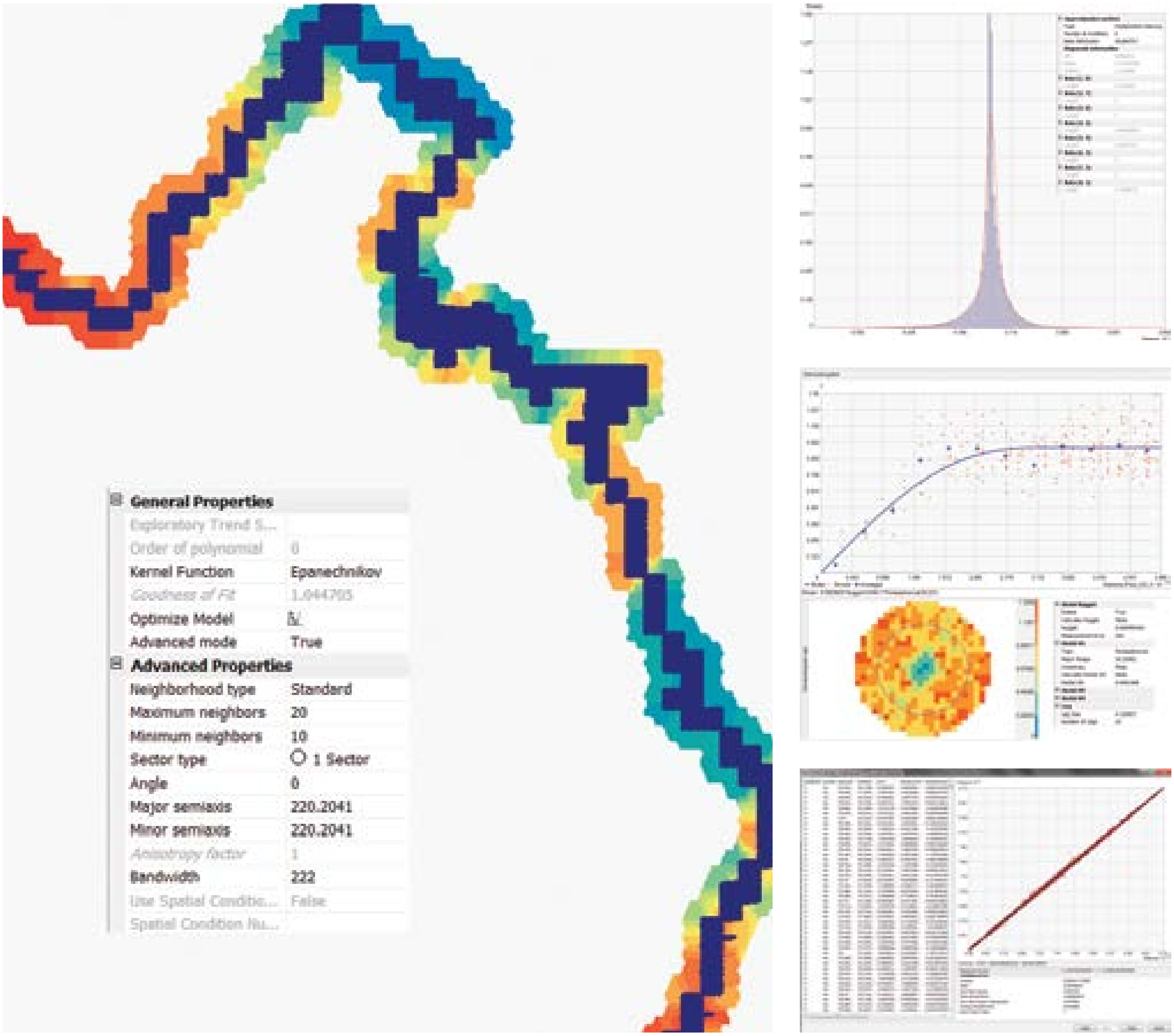
↑ Figure 5a: Slope along proposed trail based on public available DEM



↑ Figure 5b: Slope along proposed trail when elevation values were decreased by elevation error



↑ Figure 5c: Slope along proposed trail when elevation values were increased by elevation error



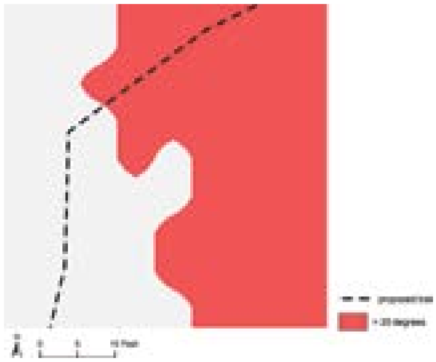
↑ Modeling fitting using the Geostatistical Wizard
↑ Figure 6a: The interpolated surface for the entire area around the proposed trail
↑ Figure 6b: Data residuals transformed to Gaussian distribution.
↑ Figure 6c: The fitted semivariogram model describing the correlation of the residuals.
↑ Figure 6d: The model diagnostics show that the kriging model was well fitted.

Another method of incorporating uncertainty into GIS workflows is geostatistical simulation. Rather than just producing a single surface with an estimate of prediction error, hundreds or even thousands of surfaces are produced. These simulated surfaces all have the same values at sampled locations (e.g., lidar points) but deviate from one another at unsampled locations. All the urfaces have the same statistical characteristics.

The average of all simulations tends toward the original surface produced by kriging. To use geostatistical simulation to quantify DEM uncertainty, create a DEM using simple kriging (only simple kriging layers are supported at this time). Next, use the Gaussian Geostatistical Simulations (GGS) tool in the Geostatistical Analyst toolbox to create 100 or more simulated surfaces. The GGS tool has an option to post-process the simulated rasters on a cell-by-cell basis and produce a new raster that summarizes the simulations. Based on the user's preference, the summary raster can contain the minimum, maximum, mean, median (and a number of other statistics) of the simulated rasters.

Use the summarized raster as the elevation surface for the remainder of the GIS workflow. How the simulations are summarized depends on the type of problem being solved. To use several of the previously cited case studies, the analyst deleanating the watershed may want to use the minimum cell values from the simulations while the engineer siting the cell phone tower may want to use the maximum cell values from the simulations.

Geostatistical simulations better reflect the data variability between measurements. Their production requires a valid kriging model. Figure 6 shows four key dialog boxes



↑ Figure 7: Areas along proposed trail with slope greater than 20 degrees in at least one of the 100 simulations

from the Geostatistical Wizard that illustrate steps in the model fitting process. Since the elevation values are, on average, slowly changing along the trail, the data detrending option was used. The data residuals (i.e., the difference between measured values and estimated trend at the data locations) were transformed to Gaussian distribution as shown in Figure 6b. The fitted semivariogram model describing the correlation of the residuals is shown in Figure 6c. The model diagnostics in Figure 6d show that the kriging model was well fitted. (Detailed instructions on fitting a kriging model can be found *Spatial Statistical Data Analysis for GIS Users*, published by Esri Press.)

The simulation workflow was applied to the problem of estimating erosion control measures needed for the park trail previously mentioned. A simple kriging model was fitted to the data and 100 simulated elevation rasters generated. Slope was calculated for all 100 rasters. Figure 7 shows locations where a slope of greater than 20 degrees was found in at least one of the 100 simulations. This area is much larger than that delineated by using the publically available raster or adjusting by the prediction error. While this is a conservative estimate of the areas that could require erosion control, it highlights the high level of uncertainty in elevation at this location. Alternatively, the number of times a slope greater than 20 degrees was found in each cell could be calculated and visualized to depict the approximate probability that the slope exceeds the threshold value.

Conclusion

Digital elevations models are just that—models. They represent elevation values but assuming they represent the true elevation values can lead to incorrect or unrealistic results from a GIS analysis. Kriging methods can help quantify the level of uncertainty associated with elevation. Once the level of uncertainty is known, it can be used to adjust the elevation value to create worst and best case scenarios. Alternatively, geostatistical simulations can be used to create many possible realizations of an elevation surface that can be further used in geoprocessing.

Further Reading

Krivoruchko, Konstantin. “Empirical Bayesian Kriging Implemented in ArcGIS Geostatistical Analyst,” *ArcUser* Fall 2012, pp.

“Modeling Contamination Using Empirical Bayesian Kriging,” *ArcUser Online*, esri.com/news/arcuser/1012/modeling-contamination-using-empirical-bayesian-kriging.html

Spatial Statistical Data Analysis for GIS Users. 2011, Esri Press