Geostatistical Mapping With Continuous Moving Neighborhood[1]

Alexander Gribov[2] and Konstantin Krivoruchko[2]

An issue that often arises in such GIS applications as digital elevation modeling (DEM) is how to create a continuous surface using a limited number of point observations. In hydrological applications, such as estimating drainage areas, direction of water flow is easier to detect from a smooth DEM than from a grid created using standard interpolation programs. Another reason for continuous mapping is esthetic; like a picture, a map should be visually appealing, and for some GIS users this is more important than map accuracy. There are many methods for local smoothing. Spline algorithms are usually used to create a continuous map, because they minimize curvature of the surface. Geostatistical models are commonly used approaches to spatial prediction and mapping in many scientific disciplines, but classical kriging models produce noncontinuous surfaces when local neighborhood is used. This motivated us to develop a continuous version of kriging. We propose a modification of kriging that produces continuous prediction and prediction standard error surfaces. The idea is to modify kriging systems so that data outside a specified distance from the prediction location have zero weights. We discuss simple kriging and conditional geostatistical simulation, models that essentially use information about mean value or trend surface. We also discuss how to modify ordinary and universal kriging models to produce continuous predictions, and limitations using the proposed models.

INTRODUCTION

Surface discontinuity is an undesirable effect in some applications, including a digital elevation model (DEM), which must be continuous and have no artifacts (Wilson and Gallant, 2000) to be used for hillshading and hydrological modeling. A DEM is a grid of square cells, where each cell value represents elevation. In hillshading, the illumination values for each cell in an elevation grid are calculated by evaluating the relationship between the position of the light source and the direction and steepness of the terrain. In hydrological applications, such as estimating drainage areas, direction of water flow is detected from DEM as a drop

[2]Environmental Systems Research Institute, 380 New York Street, Redlands, California; e-mail: agribov@esri.com; kkrivoruchko@esri.com

of water is followed from where it falls, to a stream, then to a river, and finally to the sea. Water flow direction is defined perpendicular to the elevation isolines in the direction to the steepest descent. Consequently, interpolation of the elevation measurements resulting in a DEM surface must be continuous to simplify the analysis of hydrological processes.

After detecting and removing systematic and rough errors, users of geostatistical software usually expect to see a continuous surface from continuous data, such as temperature observations. However, kriging prediction and prediction standard errors in nearby locations are significantly different if their local neighborhoods are different. Consider simulated configuration and values of 10 points in a 20 by 20 units area, Figure 1. Positions of points were simulated from uniform distribution, and values are from normal distribution with zero mean and exponential covariance with the following parameters: nugget = 3, sill = 1, and effective range = 30.

Consider prediction at two nearby points, represented as centers of two circles. Prediction to each point uses data inside each corresponding circle. The only difference between local neighborhoods is inclusion or not of location with value −3.60, and this may lead to discontinuity of the prediction and prediction standard error surfaces between two points under consideration. Generally speaking, kriging cannot produce continuous surfaces with local neighborhoods, but breaks are clearly seen if data have significantly different values in nearby local neighborhoods.
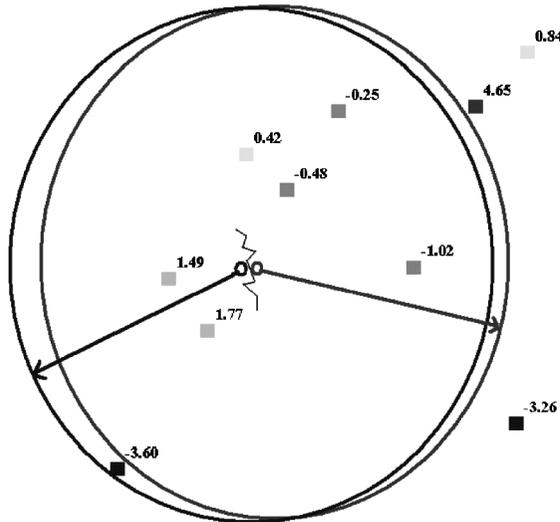
Figure 1. Simulated data for comparison of simple and continuous simple kriging.
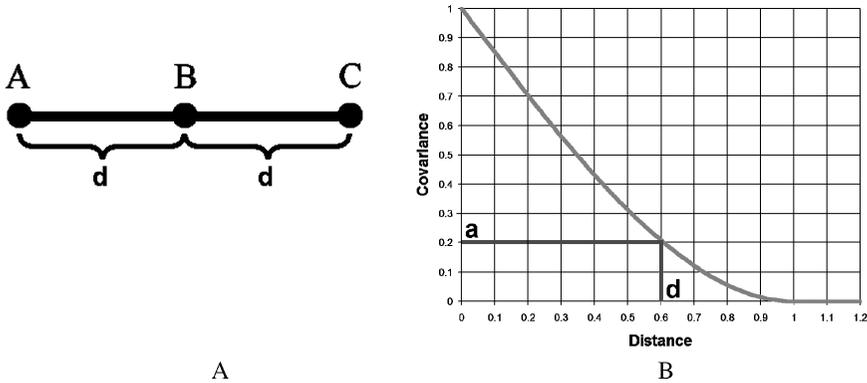
Figure 2. Data interpolation using three points on a line (A) using covariance (B).

Kriging prediction depends on data outside the range of correlation (Cressie, 1993; Gandin, 1963). Consider three points on a line, Figure 2(A), where distances from point B to points A and C equals d. Let $Z(A)$, $Z(B)$, and $Z(C)$ be random variables with means $E\{Z(A)\} = E\{Z(B)\} = E\{Z(C)\} = 0$ and variances $\text{var}\{Z(A)\} = \text{var}\{Z(B)\} = \text{var}\{Z(C)\} = 1$. Let range of the covariance model be less than $2d$ and $\text{cov}\{Z(A), Z(B)\} = \text{cov}\{Z(B), Z(C)\} = a$; $\text{cov}\{Z(A), Z(C)\} = 0$. Figure 2(B) shows an example of the spherical covariance model with such features.

Prediction and prediction variance at the point A using point B only are

$$\hat{Z}(A) = a \cdot Z(B) \quad \text{and} \quad E\{(\hat{Z}(A) - Z(A))^2\} = 1 - a^2.$$

Prediction and prediction variance at the point A using point C only are

$$\hat{Z}(A) = 0 \quad \text{and} \quad E\{(\hat{Z}(A) - Z(A))^2\} = 1.$$

Prediction and prediction variance in the point A using both points B and C are

$$\hat{Z}(A) = \frac{a}{1 - a^2} \cdot Z(B) - \frac{a^2}{1 - a^2} \cdot Z(C) \quad \text{and} \quad E\{(\hat{Z}(A) - Z(A))^2\} = \frac{1 - 2a^2}{1 - a^2}.$$

Thus, data outside range influence on prediction if at least one datum at distance less than a range of correlation exists.

Continuous surfaces can be produced using global neighborhood with all input data. However, there are two main reasons why prediction using local neighborhood should be preferred in practice. First, kriging with large number of points

(larger than 100–200 observations) leads to the computational problem of solving a large system of linear equations. Second, in practice true covariance is estimated, and measurements include uncertainty, and this uncertainty is not known exactly. Then it is possible that interpolation with a large number of neighbors produces larger mean-squared prediction error than interpolation with a relatively small number of neighbors.

In GIS, a continuous surface is usually produced by predicting values on a regular grid and then using smoothing filters. This is equivalent to adding (often uncontrollable) noise to the prediction. Then prediction and prediction standard error surfaces do not match, which could be a problem if they are used in the decision-making. Another method of continuous mapping is spline interpolation. For instance, Hutchinson (1997), implemented splines in tension in the ANUDEM program, which interpolates "the elevation data onto a regular grid by minimizing a suitably weak roughness penalty on the fitted grid values." Resulting surfaces are over smooth, and the best parametric covariance model to describe the spatial structure of the DEM is Gaussian. DEM is used as input data in many applications. However, many geostatisticians recommend that the Gaussian covariance model not to be used for practical applications; see for example Wackernagel (1998).

We propose modification of kriging to produce continuous, without breaks, prediction and prediction standard error surfaces. The idea is to modify kriging system so that data outside a specified distance from the prediction location have zero weights. We discuss simple kriging and conditional geostatistical simulation, models that essentially use information about mean value or trend surface. We also discuss how to modify ordinary and universal kriging models to produce continuous predictions, and limitations using the proposed models.

CONTINUOUS SIMPLE KRIGING

Consider spatial data Z_i, measured at the locations s_i. In most applications, data include measurement error, and our goal is to predict signal $Y(s)$ at many new locations s using the data Z_i, that is, we are using filtering to remove noise or measurement error if they exist. If there is no measurement error, prediction to the data locations gives exact measurements: $\hat{Y}(s_i) = Z_i$. We use the following simple kriging model for signal $Y(s)$ (Cressie, 1993; Gandin, 1963; Krivoruchko, Gribov, and Ver Hoef, 2000):

$$Y(s) = m(s) + S(s), \tag{1}$$

where $m(s)$ is a large scale variation (trend) and $S(s)$ is a random process with zero mean and known covariance (small scale variation).

Measurement at the locations s_i is a sum of the signal and the independent random error with zero mean and known variance.

$$Z_i = Y(s_i) + \varepsilon_i, i = \overline{1, n}, \tag{2}$$

where n is a number of measurements. Formula (2) allows for more than one measurement at the same data location; see the discussion on the difference between exact and filtered kriging models in Krivoruchko, Gribov, and Ver Hoef (2000). Exact kriging cannot be used when there are different measurements at the same location.

Simple kriging (also called simple filtered kriging) uses a linear combination of the data $\hat{Y}(s) = m(s) + \sum_{i=1}^{n} \lambda_i(s) \cdot (Z_i - m(s_i))$ and finds the weights that minimize the mean-squared prediction error $E\{(\hat{Y}(s) - Y(s))^2\}$. The optimal weights

$$\lambda(s) = \begin{bmatrix} \lambda_1(s) \\ \lambda_2(s) \\ \vdots \\ \lambda_n(s) \end{bmatrix}$$

are a solution of the following system:

$$\begin{bmatrix} \text{var}\{S(s_1)\} + \text{var}\{\varepsilon_1\} & \text{cov}\{S(s_1), S(s_2)\} & \cdots & \text{cov}\{S(s_1), S(s_n)\} \\ \text{cov}\{S(s_2), S(s_1)\} & \text{var}\{S(s_2)\} + \text{var}\{\varepsilon_2\} & \cdots & \text{cov}\{S(s_2), S(s_n)\} \\ \vdots & \vdots & \ddots & \vdots \\ \text{cov}\{S(s_n), S(s_1)\} & \text{cov}\{S(s_n), S(s_2)\} & \cdots & \text{var}\{S(s_n)\} + \text{var}\{\varepsilon_n\} \end{bmatrix}$$

$$\times \begin{bmatrix} \lambda_1(s) \\ \lambda_2(s) \\ \vdots \\ \lambda_n(s) \end{bmatrix} = \begin{bmatrix} \text{cov}\{S(s), S(s_1)\} \\ \text{cov}\{S(s), S(s_2)\} \\ \vdots \\ \text{cov}\{S(s), S(s_n)\} \end{bmatrix}$$

There are no differences between exact and filtered models when predicting at locations where data have not been observed. If data are not precise, kriging can filter out measurement error at the data location.

Because the data outside the range of correlation do influence prediction, we propose to use kernel function to smooth local fluctuations so that the influence of measurements far from the prediction location is reduced and starting from

some specified distance data weights equal to zero. However, instead of using this smoother after the kriging surface is produced, we derive a new predictor that incorporates this desired feature: continuous prediction and prediction standard error surfaces.

The following is one possible way to modify the data for simple kriging prediction at the location s:

$$Z_i^*(s) = w(s_i - s, s) \cdot (Z_i - m(s_i)) + \sqrt{1 - w^2(s_i - s, s)} \cdot \sigma_i \cdot \eta_i, \qquad (3)$$

where $|w(s_i - s, s)| = |w(r, s)| \leq 1$ is a kernel function; $w(0, s) = 1 \forall s$ and such distance $\chi(s)$ exists, that $w(r, s) = 0 \forall ||r|| \geq \chi(s)$; $\sigma_i^2 = \text{var}\{Z_i\} = \text{var}\{S(s_i)\} + \text{var}\{\varepsilon_i\}$; η_i is an independent random variable with zero mean and variance one.

We defined $w(r, s)$ as a two-parameter function, because ideally the distance where kernel becomes equal to zero should depend on location. For example, $w(r, s)$ may be anisotropic, depend on data density, or be function of the explanatory variable. However, in this paper, we consider a simple situation where $w(r, s) = w(r)$ is a function of distance between two locations only and $\chi(s)$ is a constant, $\chi(s) = \chi$. Function $w(r)$ is close to 1 near $r = 0$ such that at least several data are unchanged. One such kernel function, which will be used in the simulation study below, is the following:

$$w(r) = \begin{cases} 1, \ ||r|| \leq 7.5 \\ 156.25 - 84.375 \cdot ||r|| + 18 \cdot ||r||^2 - 1.88 \cdot ||r||^3 \\ \quad + 0.096 \cdot ||r||^4 - 0.00192 \cdot ||r||^5, \ 7.5 < ||r|| < 12.5 \\ 0, \ ||r|| \geq 12.5 \end{cases} \qquad (4)$$

Figure 3(A) shows transect of this kernel.

Then covariance between modified data is

$$\text{cov}\{Z_i^*(s), Z_j^*(s)\} = \begin{cases} w(s_i - s) \cdot w(s_j - s) \cdot \text{cov}\{S(s_i), S(s_j)\}, \ i \neq j \\ \sigma_i^2, \ i = j \end{cases} \qquad (5)$$

Variance of the modified data remains the same as variance of the original data, while covariance decreases such that modified data outside distance χ do not correlate. The same is true for covariance between datum and prediction location:

$$\text{cov}\{Y(s), Z_j^*(s)\} = w(s_j - s) \cdot \text{cov}\{S(s), S(s_j)\}.$$
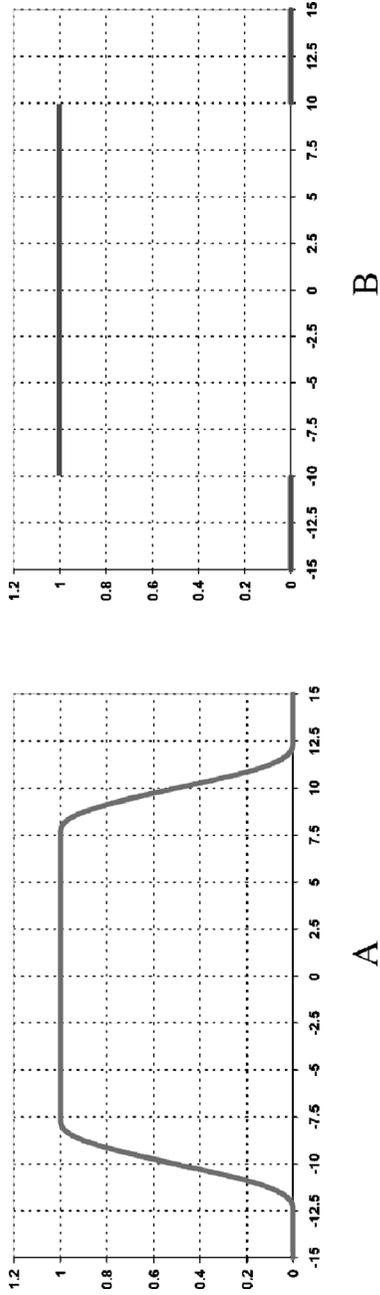
Figure 3. Example of kernel functions $w(r)$. A, Continuous kernel (4). B, Kernel that produces the same predictions as classical simple kriging.

The continuous version of a simple kriging prediction of the signal is

$$\hat{Y}(s) = m(s) + \sum_{i=1}^{n} \lambda_i(s) \cdot Z_i^*(s), \tag{6}$$

where weights

$$\lambda(s) = \begin{bmatrix} \lambda_1(s) \\ \lambda_2(s) \\ \vdots \\ \lambda_n(s) \end{bmatrix}$$

are the solution of the following system:

$$A(s) \times \lambda(s) = b(s), \tag{7}$$

where

$A(s) =$

$$\begin{bmatrix} \mathrm{var}\{S(s_1)\} + \mathrm{var}\{\varepsilon_1\} & w(s_1 - s) \cdot w(s_2 - s) \cdot \mathrm{cov}\{S(s_1), S(s_2)\} & \cdots & w(s_1 - s) \cdot w(s_n - s) \cdot \mathrm{cov}\{S(s_1), S(s_n)\} \\ w(s_2 - s) \cdot w(s_1 - s) \cdot \mathrm{cov}\{S(s_2), S(s_1)\} & \mathrm{var}\{S(s_2)\} + \mathrm{var}\{\varepsilon_2\} & \cdots & w(s_2 - s) \cdot w(s_n - s) \cdot \mathrm{cov}\{S(s_2), S(s_n)\} \\ \vdots & \vdots & \ddots & \vdots \\ w(s_n - s) \cdot w(s_1 - s) \cdot \mathrm{cov}\{S(s_n), S(s_1)\} & w(s_n - s) \cdot w(s_2 - s) \cdot \mathrm{cov}\{S(s_n), S(s_2)\} & \cdots & \mathrm{var}\{S(s_n)\} + \mathrm{var}\{\varepsilon_n\} \end{bmatrix}$$

$$b(s) = \begin{bmatrix} w(s_1 - s) \cdot \mathrm{cov}\{S(s), S(s_1)\} \\ w(s_2 - s) \cdot \mathrm{cov}\{S(s), S(s_2)\} \\ \vdots \\ w(s_n - s) \cdot \mathrm{cov}\{S(s), S(s_n)\} \end{bmatrix}$$

$Z_i^*(s)$ contains the unknown random variable η_i. The best linear prediction (6), which minimizes mean-squared error, will be reached for $\eta_i = 0 \, \forall i$:

$$\hat{Y}(s) = m(s) + \sum_{i=1}^{n} \lambda_i(s) \cdot w(s_i - s) \cdot (Z_i - m(s_i)) \tag{8}$$

Then the continuous kriging prediction variance is the following:

$$E\{(\hat{Y}(s) - Y(s))^2\} = E\left\{\left(m(s) + \sum_{i=1}^{n} \lambda_i(s) \cdot w(s_i - s) \cdot (Z_i - m(s_i)) - Y(s)\right)^2\right\}$$

$$= \mathrm{var}\{S(s)\} - \sum_{i=1}^{n} \lambda_i(s) \cdot w(s_i - s) \cdot \mathrm{cov}\{S(s), S(s_i)\}$$

$$- \sum_{i=1}^{n} \lambda_i^2(s) \cdot (1 - w^2(s_i - s)) \cdot \sigma_i^2 \tag{9}$$

If trend $m(s)$, covariance of the process $S(s)$, and kernel $w(r)$ are continuous, then the weights $\lambda(s)$ in (7) are also continuous. Thus, the prediction and prediction standard error surfaces are also continuous. In locations s, where $m(s)$, $S(s)$, and $w(r)$ have continuous first derivatives, prediction and prediction standard error have continuous first derivatives as well.

If data are separated from location s by a distance larger than χ, corresponding rows and columns of the matrix $A(s)$ are equal to zero, except of diagonal values, and corresponding rows of vector $b(s)$ are also zeroes. Solution of such a system is equivalent to solution of the system using data inside distance χ from the prediction location only. This guarantees that data outside distance χ do not influence prediction. In contrast, classical kriging weights may depend on data far beyond the range of data correlation. Using the kernel function

$$w(r) = \begin{cases} 1, & \|r\| < \chi \\ 0, & \|r\| \geq \chi \end{cases}, \tag{10}$$

(see Fig. 3(B), prediction will be equivalent to prediction using standard simple kriging with a searching neighborhood where the radius of the circle equals χ. This kernel is not continuous and consequently prediction and prediction standard error are noncontinuous in classical kriging models, see Equation (7).

Using the simulated data described in the introduction and displayed in Figure 1, and the standard simple kriging with zero mean and known covariance model (which was used for data simulation), and a radius of the searching neighborhood equaling 10 units, produced the prediction map presented in Figure 4(A). A continuous simple kriging prediction map is presented in Figure 4(B) using kernel (4).

Figure 5 shows prediction standard error maps using classical simple kriging (Fig. 5(A), and continuous simple kriging (Fig. 5(B).

Figure 6 compares continuous simple kriging variance to variance estimated using classical simple kriging with several different searching neighborhood radii for transect plane $X = 10$ in Figure 5.
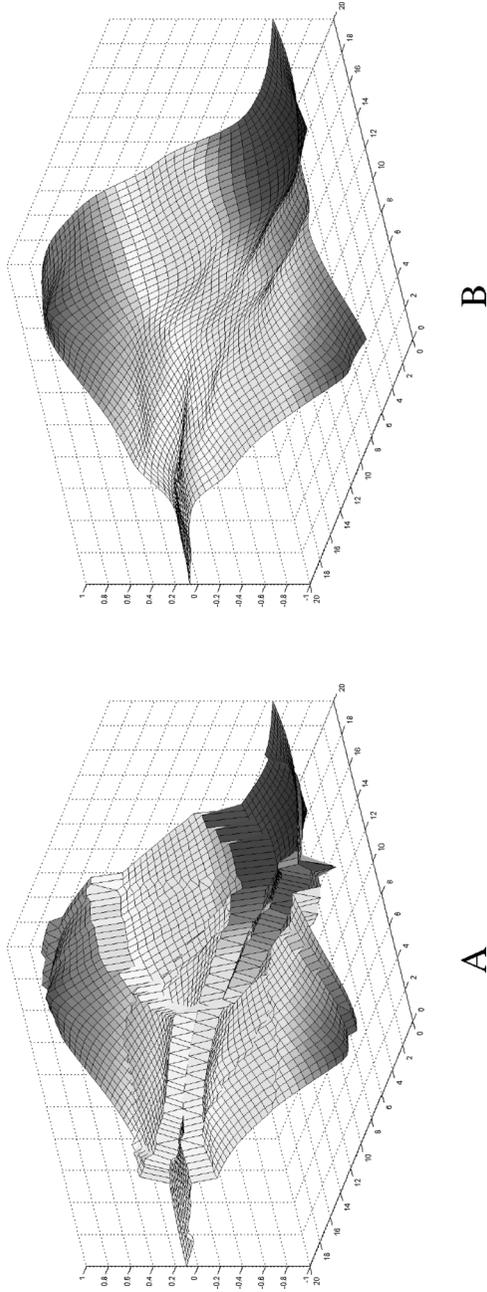
Figure 4. Simple kriging using the input data displayed in Figure 1. A, Standard simple kriging. B, Continuous simple kriging using kernel (4).
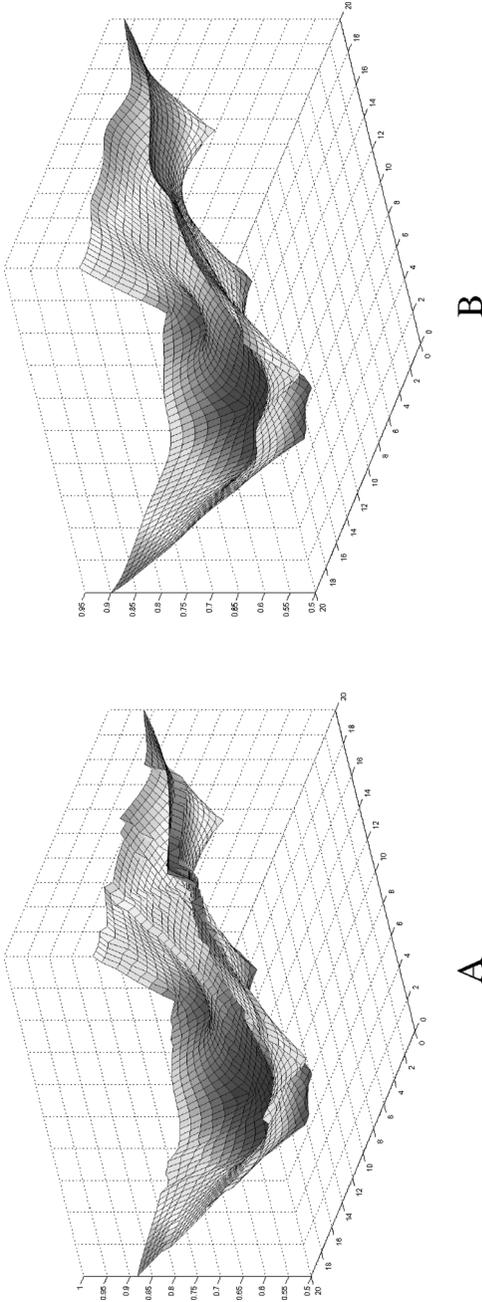
Figure 5. Prediction standard error maps using classical (A) and continuous (B) simple kriging.
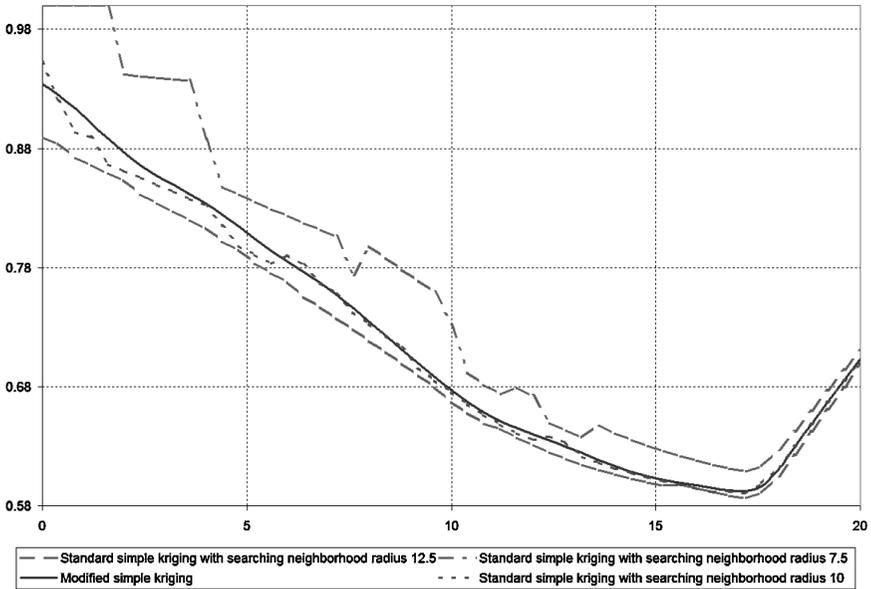
Figure 6. Comparing kriging variances using plane YZ transect at $X = 10$.

It is important that continuous simple kriging variance is between classical simple kriging variance with searching neighborhood radii 7.5 and 12.5 units, see Figure 3 and formula (4). Classical simple kriging variance using a searching neighborhood radius equal to 10 is also displayed in Figure 6. It is close to the continuous simple kriging variance.

CONDITIONAL GEOSTATISTICAL SIGNAL SIMULATION USING CONTINUOUS SIMPLE KRIGING

Kriging predicts a single value, close to the true but unknown value. Geostatistical conditional simulations describe local data variability based on many variants (realizations) of the predictions, consistent with the data and its statistical characteristics. For any of such realizations, the prediction error is larger than kriging standard error.

If data are not precise, which is common in environmental sciences, it makes no sense to reproduce inaccurate measurements. We use a modification of the conditional geostatistical simulation approach for data measured with error (Aldworth, 1998), which is a generalization of (Journel, 1974) and Journel and Huijbregts (1978), conditional simulation approach using exact simple kriging.

In conditional geostatistical simulation practice, simple kriging is usually used with local neighborhood. As a result, the simulated surfaces have breaks. Using the continuous simple kriging approach, simulation surfaces will be continuous.

Conditional simulation using a continuous simple kriging model is as follows:

$$Y_C(s) = m(s) + S_N(s) + \sum_{\|s_i - s\| < \chi, i = \overline{1,n}} \lambda_i(s) \cdot w(s_i - s)$$
$$\times [(Z_i - m(s_i)) - (S_N(s_i) + \varepsilon_{N,i})], \tag{11}$$

where $Y_C(s)$ is a signal, $\lambda_i(s)$ solution of (7), $w(r)$ kernel function, and $S_N(s)$ is a realization of the independent random Gaussian process with zero mean and known covariance of the process $S(s)$, and $\varepsilon_{N,i}$ is a realization of independent normal random variable with zero mean and variance var$\{\varepsilon_i\}$. Note that simulated values are distributed slightly differently from the distribution of simulated values using standard simple kriging model.

If the variable being studied is continuous by its nature, the result of a conditional Gaussian simulation should be also continuous. We expect that our approach will find interesting applications.

DISCUSSION AND CONCLUSIONS

Noncontinuous kriging prediction and prediction standard error surfaces can have a considerable effect on decision-making in applications where the data gradient is calculated. In practice, kriging with local neighborhood always produces noncontinuous surfaces. We proposed a modification of the simple kriging algorithm to produce continuous surfaces. A model requires information on mean value or continuous trend surface and can be used with disjunctive kriging (Rivoirard, 1994), where the mean value is assumed to be known.

Simple kriging is the most popular model in geostatistical conditional simulation, but most researchers are using ordinary kriging for interpolation. If mean value and trend surface are unknown, one possible way to modify data to make the prediction continuous is the following (this formula is also valid for the universal kriging model):

$$Z_i^*(s) = Z_i + \frac{\sqrt{1 - w^2(s_i - s)}}{w(s_i - s)} \cdot \sigma_i \cdot \eta_i \tag{12}$$

This formula is a generalization of formula (3) when the mean value is unknown. Formula (12) reduces to (3) by subtracting the mean value and multiplying by $w(s_i - s)$. Both modifications of data, (3) and (12), will produce the same result in a continuous simple kriging system (7)–(9).

A continuous version of the ordinary kriging prediction of the signal $Y(s)$ using (12) is

$$\hat{Y}_{OK}(s) = \sum_{i=1}^{n} \lambda_i(s) \cdot w(s_i - s) \cdot Z_i, \tag{13}$$

where weights

$$\lambda_{OK}(s) = \begin{bmatrix} \lambda_1(s) \\ \lambda_2(s) \\ \vdots \\ \lambda_n(s) \\ \mu \end{bmatrix}$$

are a solution of the following system

$$A_{OK}(s) \times \lambda_{OK}(s) = b_{OK}(s), \tag{14}$$

where

$A_{OK}(s) =$

$$\begin{bmatrix} \mathrm{var}\{S(s_1)\} + \mathrm{var}\{\varepsilon_1\} & w(s_1 - s) \cdot w(s_2 - s) \cdot \mathrm{cov}\{S(s_1), S(s_2)\} & \cdots & w(s_1 - s) \cdot w(s_n - s) \cdot \mathrm{cov}\{S(s_1), S(s_n)\} & w(s_1 - s) \\ w(s_2 - s) \cdot w(s_1 - s) \cdot \mathrm{cov}\{S(s_2), S(s_1)\} & \mathrm{var}\{S(s_2)\} + \mathrm{var}\{\varepsilon_2\} & \cdots & w(s_2 - s) \cdot w(s_n - s) \cdot \mathrm{cov}\{S(s_2), S(s_n)\} & w(s_2 - s) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ w(s_n - s) \cdot w(s_1 - s) \cdot \mathrm{cov}\{S(s_n), S(s_1)\} & w(s_n - s) \cdot w(s_2 - s) \cdot \mathrm{cov}\{S(s_n), S(s_2)\} & \cdots & \mathrm{var}\{S(s_n)\} + \mathrm{var}\{\varepsilon_n\} & w(s_n - s) \\ w(s_1 - s) & w(s_2 - s) & \cdots & w(s_n - s) & 0 \end{bmatrix}$$

$$b_{OK}(s) = \begin{bmatrix} w(s_1 - s) \cdot \mathrm{cov}\{S(s), S(s_1)\} \\ w(s_2 - s) \cdot \mathrm{cov}\{S(s), S(s_2)\} \\ \vdots \\ w(s_n - s) \cdot \mathrm{cov}\{S(s), S(s_n)\} \\ 1 \end{bmatrix}$$

The problem is that Eq. (14) does not have a solution when the kernel function is equal to zero, $w(s_i - s) = 0 \forall i$, and in the nearby area solution of (14) is unstable.

Thus, in the case of ordinary kriging, the kernel function should be selected carefully so that the distance χ will be large enough to prevent areas where prediction is not possible. Similar data modification can be used for universal kriging.

Research of using different kernels for simple and ordinary kriging and case study using simulated and real data are in progress, and we are working on practical recommendations for the continuous kriging.

ACKNOWLEDGMENTS

REFERENCES

Aldworth, J., 1998, Spatial prediction, spatial sampling, and measurement error: PhD Thesis, Iowa State University, 186 p.

Cressie, N., 1993, Statistics for spatial data: Wiley, New York, 900 p.

Gandin, L. S., 1963, Objective analysis of meteorological fields: Gidrometeorologicheskoe Izdatel'stvo (GIMIZ), Leningrad (translated by Israel Program for Scientific Translations, Jerusalem, 1965), 242 p.

Hutchinson, M. F., 1997, ANUDEM Version 4.6: Centre for Resource and Environmental Studies, The Australian National University, Canberra, Available from http://cres.anu.edu.au/outputs/anudem.html.

Journel, A. G., 1974, Geostatistics for conditional simulation of ore bodies: Econ. Geol., v. 69, p. 673–687.

Journel, A. G., and Huijbregts, C. J., 1978, Mining geostatistics: Academic Press, London, 600 p.

Krivoruchko, K., Gribov, A., and Ver Hoef, J., 2000, Predicting exact, filtered, and new values using kriging: Available from ESRI online at http://www.esri.com/software/arcgis/arcgisxtensions/geostatistical/research_papers.html.

Rivoirard, J., 1994, Introduction to disjunctive kriging and nonlinear geostatistics: Clarendon Press (Spatial Information Systems), Oxford, 181 p.

Wilson, J. P., and Gallant J. C., eds., 2000, Terrain analysis. Principles and applications, Wiley, New York, 479 p.

Wackernagel, H., 1998, Multivariate geostatistics: Springer, Berlin, 292 p.