The Geostatistical Framework for Spatial Prediction

ZHANG Jingxiong  YAO Na

Abstract  Geostatistics provides a coherent framework for spatial prediction and uncertainty assessment, whereby spatial dependence, as quantified by variograms, is utilized for best linear unbiased estimation of a regionalized variable at unsampled locations. Geostatistics for prediction of continuous regionalized variables is reviewed, with key methods underlying the derivation of major variants of uni-variate Kriging described in an easy-to-follow manner. This paper will contribute to demystification and, hence, popularization of geostatistics in geoinformatics communities.

Keywords  geostatistics; Kriging; regionalized variables; variograms; spatial prediction

CLC number  P228.4

Introduction

Geostatistics used to mean statistics applied to geology, or more generally, to past problems in the earth sciences. Beginning in the mid 1960’s and especially in the mid 1970’s, however, it became closely affiliated with the work of Georges Matheron. The connection is still prevalent today. Because much of his early work and that of his students appeared primarily in French, it was not as well known in other countries. However, several events began to change that. In 1975, a NATO ASI was held near Rome, Italy on Advanced Geostatistics in the Mining Industry. A more definitive theoretical article appeared in the J Applied Probability in 1973.

Matheron’s work was not very well accepted in the statistical community for a period of time although a number of prominent statisticians were visitors at Fountainebleau in the 1970’s, 1980’s, and 1990’s. Nowadays, geostatistics has established a place for itself both within statistics journals and at national meetings.

Geostatistics is definitely an applied discipline. Its development has been the handiwork of mining engineers, petroleum engineers, hydrologists, soil scientists, geologists as well as statisticians. There are applications in epidemiology and plant pathology, as well as forestry, atmospheric sciences, global change, and geography. There is some overlap with geographic information systems (GIS) and spatial statistics in general. These journals should especially be noted: Mathematical Geology, Water Resources Research and the J Soil Science Society of America. More recently, articles have begun appearing in environmetrics, remote sensing of the environment, and numerous others.

Geostatistics is concerned with spatial data. That is, each data value is associated with a location in space and there is at least an implied connection between the location and the data value. Location refers often to a point in space (in an abstract mathematical sense).
and can be associated with an area or volume in space. For example, a data value associated with an area might be the average value of an observed variable. In the latter case the area is often called the “support” of the data. This is closely related to the idea of the support of a measure. Let \( x \) be a point (not just coordinates) in 1, 2, or 3 dimensional space and \( z(x) \) denote observed values of a variable \( Z \) at location \( x \). For example, this might be the grade of copper, temperature, hydraulic conductivity, or concentration of a pollutant.

Now suppose that \( t \) is a location that is not “sampled”. The objective then is to estimate/predict the value \( Z(t) \) (and the data locations as well as the location \( t \)). If only this information is given then the problem is ill-posed, i.e. it does not have a unique solution. One way to obtain a unique solution is to introduce a model into the problem. There are two ways to do this; one is deterministic and the other is stochastic or statistical. Both approaches must somehow incorporate the idea that there is uncertainty associated with the estimation/prediction step. The value at the unsampled location is not itself random but our knowledge of it is uncertain. One approach then is to treat \( Z \) as being a random variable. If its distribution were known then the “best” estimator (“best” meaning unbiased and having minimal variance of the error of estimation) at location \( t \) would be the conditional expectation of \( Z(t) \), given the values of the other random variables. However, observational data consist of only one observation of the random variable \( Z \) at data locations and none of \( Z(t) \), hence it is not possible to estimate or model this distribution using standard ways of modeling or fitting probability distributions.

Geostatistics provides a coherent framework for spatial prediction and uncertainty assessment\(^{[1-9]}\). Spatial prediction or interpolation is concerned with how to estimate the variable under study at an un-sampled location, given sample observations at nearby locations. This question aims at finding a single number, termed the ‘estimate’, at the un-sampled location. Estimation is possible due to spatial correlation, i.e. the underlying biophysical phenomenon causes observations that are measured closely to be dependent on one and another. If the unknown values at the un-sampled location were dependent on the known sample value at another location, then those sample values carry information about the unknown.

This paper provides an overview of the most used geostatistical estimation following a description of variograms. The Kriging techniques discussed include simple, ordinary, and universal Kriging. The underlying principles of variants of Kriging will be seen as unified in terms of implementing the criteria of best linear unbiased estimation and manipulating basic linear algebra.

### 1 Spatial dependency and variogram models

The texts in this section follow Caers\(^{[10]}\). Statistics of random variables is based on the fact that when something is unknown or uncertain, one turns it into a random variable \( Z \) and describes its possible outcomes \( z \) through a probability model, often as a cumulative distribution function (cdf):

\[
\Pr(Z \leq z) = F_Z(z)
\]

One of the most popular concepts is that of stationarity. That is, in order to perform estimation, one requires having multiple observations of the same variable. For example, in order to estimate the expected value of a population, or to determine the probability of a certain event occurring, or to draw a regression line between two variables, one needs multiple outcomes of the same random variables. This requires that all outcomes be alternative representations of the same random variable, which implies a rather strong assumption that all outcomes come from the same population. The stationarity principle is particularly relevant to the study of spatial phenomena. In a spatial context, only one single outcome \( z \) exists at each sample location \( x=(X,Y,Z) \). The unknown outcome at the un-sampled locations \( x \) is modeled by a random variable \( Z(x) \), determined by a distribution function \( F_Z(z; x) \). How can one then estimate the expected value at \( Z(x) \) or the probability that exceeds a certain threshold when no alternative outcomes are observed? This calculation requires a
decision of stationarity. For example, to estimate the mean of $Z$ at $x$, one uses information at nearby locations. These locations are termed the zone of stationarity, where statistics about $Z$ are assumed to be similar. Such a zone is not only dependent on the particular phenomenon under study, but also on the amount of data. It should be stressed that stationarity is not a property of the biophysical phenomenon, rather it is a decision for modeling that allows us to perform estimation with spatial data. Such a decision is based on sound judgment and cannot be statistically tested.

Geological genesis of a reservoir causes observations to be correlated in space. The underlying geological continuity causes porosity at location $x$ to be dependent on porosity at location $x+h$, as long as distance vector $h$ is not too large. In geostatistics, one attempts at quantifying that geological continuity. The simplest way is to consider correlation or degree of association between any two points in space. The correlogram is such a measure of spatial continuity and represents the correlation between $Z(x)$ and $Z(x+h)$.

$$
\rho(Z(x), Z(x+h)) = \frac{\text{Cov}(Z(x), Z(x+h))}{\sqrt{\text{Var}(Z(x))\text{Var}(Z(x+h))}}
$$

(2)

with

$$
\text{Cov}(Z(x), Z(x+h)) = E[(Z(x) - E[Z(x)])(Z(x+h) - E[Z(x+h)])]
$$

To infer this probabilistic measure of association from the data, one requires multiple outcomes of the pair $(Z(x), Z(x+h))$; hence one requires a decision of stationarity to model the experimental or sample correlogram. Indeed, to calculate the experimental correlogram, the dataset is searched for various pairs of sample locations that are approximately a lag $h$ apart. Note that $h$ is a vector, hence pairs of data within a certain distance and along a certain direction should be distinguished. In a layered reservoir, one would expect a higher degree of association between observations in any horizontal direction compared to the vertical. The decision of stationarity allows for pooling all sample locations that are $h$ apart into one scatterplot from which the correlation coefficient is an estimate of the location-independent correlogram:

$$
\rho(Z(x), Z(x+h)) = \rho(h)
$$

(3)

An estimate of $\rho(h)$ provides a quantification of continuity along various directions. Note that $\rho(h)$ quantifies patterns between any two spatial locations, hence it is termed a two-point statistic. In many cases, a reservoir with heterogeneous and complex geology cannot be described accurately by considering two-point patterns only. Hence, two-point geostatistics should be extended to multiple-point geostatistics in order to describe better the actual geological continuity of the reservoir.

Geostatistics traditionally uses a different but related measure of spatial continuity than $\rho(h)$, which is termed the variogram:

$$
\gamma(h) = \text{var}(Z(x) - Z(x+h)) = \text{var}(Z) (1 - \rho(h))
$$

(4)

The variogram is a measure of dissimilarity between the property at locations $x$ and $x+h$, hence it tends to increase with increasing $|h|$. The variogram is small for small $|h|$ (small dissimilarity) and tends to increase with increasing $|h|$, as samples taken at larger distances become uncorrelated (perfectly dissimilar).

The variogram or correlogram are also termed moments of the bi-variate distribution between properties at any two spatial locations. The bi-variate distribution is a more complete measure of spatial dependency than the variogram:

$$
\Pr(Z(x) \leq z, Z(x+h) \leq z') = F(z, z')
$$

(5)

While the variogram considers the overall spatial continuity between any two locations, the bi-variate distribution quantifies spatial continuity between specific classes of the property at any two locations. For example, the bi-variate distribution allows quantifying the spatial correlation between the low property values (when $z$, $z'$ is low) separately from the high (when $z$, $z'$ is high) property values. The overall variogram $\gamma(h)$ measures the dissimilarity between all ranges of the data taken together. Note that in general, the continuity of high values need not be the same as that for low values. To quantify the bi-variate distribution, one employs so-called indicator random variables:

$$
I(x,z) = \begin{cases} 
1, & \text{if } Z(x) \leq z \\
0, & \text{else}
\end{cases}
$$

(6)

One can define a multiple of these indicator variables.
for each threshold \( z \). The indicator correlogram or variogram is related to the bi-variate distribution, i.e.:

\[
\Pr(Z(x) \leq z, Z(x+h) \leq z) = E[I(x,z)I(x+h,z)] = p(\rho_{x,h}(z))(1-p) + p
\]

with \( p \) the marginal proportion of \( Z \) exceeding the cutoff \( z \), hence:

\[
\Pr(Z(x) \leq z) = F_z(z)
\]

The indicator variograms or correlograms can be estimated from the sample data using the same procedure as outlined for the overall variogram. This would require transforming the \( z \)-values (e.g. porosity) measured at will into multiple indicator data, each describing different ranges of the property value.

### 2 Spatial prediction with Kriging

Spatial correlation allows making predictions about the property at un-sampled locations from sample data. If samples are not correlated in space, the best linear estimate of \( Z(x) \) would be the global mean \( m_Z = E[Z] \). Kriging, basically a form of generalized linear regression, is a name for a spatial estimation technique, which uses the variogram as a model of geospatial continuity and estimates un-sampled locations on that basis.

The simplest way of Kriging is termed simple Kriging and consists of estimating the un-sampled location as a linear combination of the neighboring data values. Prediction with known stationary mean \( m_Z \) can be described as:

\[
\hat{Z}(x) = m_Z + \sum_{i=1}^{n} \lambda_i [z(x_i) - m_Z] = m_Z + \lambda^T \mathbf{z} - m_Z
\]

where \( \lambda_i \) are termed the Kriging weights. Kriging provides a single estimate \( \hat{Z}(x) \) for the true unknown value \( Z(x) \) at location \( x \), hence the error of the estimate can be determined in terms of the mean and variance, to ensure unbiasedness and minimum dispersion of the estimator.

The average prediction error is bound to be zero on its own:

\[
E[\hat{Z}(x) - z(x)] = m_Z + \sum_{i=1}^{n} \lambda_i [E[z(x_i)] - m_Z] - E[z(x_i)] = 0
\]

and variance of the prediction error, i.e., the Kriging variance, can be written as:

\[
\sigma_Z^2(x) = \text{var}[\hat{Z}(x) - z(x)]
\]

\[
= \text{var}[m_Z + \sum_{i=1}^{n} \lambda_i [z(x_i) - m_Z] - z(x_i)]
\]

\[
= \text{var} \left[ \sum_{i=1}^{n} \lambda_i [z(x_i) - m_Z] - [z(x) - m_Z] \right]
\]

To derive a reduced form for the Kriging variance, some linear algebra is put in place. Denote an extended weight vector \( \lambda^* \) and an extended vector of regionalized variables at locations \( \{x, x_1, \ldots, x_s\} \) \( \mathbf{Z}^* \):

\[
\lambda^* = \begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_s \\ -1 \end{bmatrix}, \quad \mathbf{Z}^* = \begin{bmatrix} z(x_1) - m_Z \\ \vdots \\ z(x_s) - m_Z \\ z(x) - m_Z \end{bmatrix}
\]

Thus, the Kriging variance formulated in Eq.(11) can be rewritten as:

\[
\text{var}[\hat{Z}^T \mathbf{Z}^*] = \lambda^T \mathbf{Z}^* \begin{bmatrix} \text{cov}(x_i) & \text{cov}(x_i, x) \\ \text{cov}(x_i, x)^T & \text{var}[z(x)] \end{bmatrix} \lambda^* - 2 \lambda^T \text{cov}(x_i, x) + \text{var}[z(x)]
\]

where \( \text{cov}(x_i) \) and \( \text{cov}(x_i, x) \) stand for the covariance matrix of the sample locations and the covariance vector between sampled locations and the unsampled location \( x \), respectively.

To minimize prediction variance, it is necessary to set the \( n \) partial derivatives of prediction variance to zero:

\[
\frac{\partial \text{var}[\sigma^2_Z(x)]}{\partial \lambda} = 0
\]

which gives rise to the following simple Kriging equation:

\[
\text{cov}(x_i) \lambda - \text{cov}(x_i, x) = 0
\]

and hence the weight vector:

\[
\lambda = \text{cov}(x_i)^{-1} \text{cov}(x_i, x)
\]

Based on the weight vector, it is possible to get the simple Kriging estimation:
\[ \hat{z}(x) = m_x + \sum_{i=1}^{n} \lambda_i [z(x_i) - m_x] \]

(18)

\[ \hat{z}(x) = m_x + \lambda^T [Z(x_i) - m_x] \]

and the Kriging variance:

\[ \sigma^2_z(x) = \text{var}[z(x)] - \lambda^T \text{cov}(x_i, x) \]

(19)

When the stationary mean is unknown, it is necessary to perform estimation of \( Z(x) \) through linear combination of the sampled values \( \{z(x_i), s = 1, \ldots, n\} \). Thus, the simple Kriging estimation formulated by Eq.(9) should be reformatted as:

\[ \hat{z}(x) = \sum_{i=1}^{n} \lambda_i z(x_i) + (1 - \sum_{i=1}^{n} \lambda_i) m_x \]

(20)

where the factor for \( m_x \) should be imposed with a constraint of zero:

\[ \sum_{i=1}^{n} \lambda_i - 1 = 0 \]

(21)

so that the effect of the unknown \( m_x \) is cancelled, resulting in the so-called ordinary Kriging procedure. To streamline the algebraic treatment, it is helpful to construct a column vector of 1’s:

\[ 1 = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}_{n \times 1} \]

(22)

so that the constraint in Eq.(21) is rewritten as:

\[ \lambda^T 1 - 1 = 0 \]

(23)

The objective function for minimization of ordinary Kriging variance becomes:

\[ \phi = \lambda^T \text{cov}(x_i, x) \lambda - 2 \lambda^T \text{cov}(x_i, x) z(x) + \text{var}[z(x)] - 2 \nu (\lambda^T 1 - 1) \]

(24)

where \( \nu \) is the Lagrange multiplier. To solve for the minimum variance, the objective function is subject to partial derivation against the vector of \( n \) weights \( \{\lambda_i, s = 1, \ldots, n\} \) and is set to zero:

\[ \frac{\partial \phi}{\partial \lambda} = 0 \]

(25)

which gives rise to the following equations:

\[ \text{cov}(x_i, x) \lambda - 1 \nu = \text{cov}(x_i, x) \]

(26)

and setting the partial derivative of \( \phi \) against the Lagrange multiplier \( \nu \) leads to:

\[ \frac{\partial \phi}{\partial \nu} = 0 \]

(27)

\[ \lambda^T 1 = 1 \]

(28)

Then Eq.(26) and (28) can be combined to produce the following ordinary Kriging equation system:

\[
\begin{bmatrix}
\text{cov}(x_i) & 1 \\
1 & 0
\end{bmatrix}
\begin{bmatrix}
\lambda \\
-\nu
\end{bmatrix}
=
\begin{bmatrix}
\text{cov}(x_i, x) \\
1
\end{bmatrix}
\]

(29)

from which \( n \) weights \( \{\lambda_i, s = 1, \ldots, n\} \) can be solved, so the estimator for \( z(x) \) is

\[ \hat{z}(x) = \sum_{i=1}^{n} \lambda_i z(x_i) \]

(30)

and the Kriging variance is

\[ \sigma^2_z(x) = \text{var}[z(x)] - \lambda^T \text{cov}(x_i, x) + \nu \]

(31)

Both simple and ordinary Kriging assume stationary means for the problem domains. However, there may be evidence that the regionalized variable is better described as the sum of a non-stationary mean \( m_x(x) \) and a zero-mean, but with a spatially correlated error \( R(x) \) having covariance model \( c_r(h) = \text{cov}(R(x), R(x + h)) \).

This prompts the development of a non-stationary Kriging or a universal Kriging:

\[ \hat{z}(x) = m_x(x) + \sum_{i=1}^{n} \lambda_i [z(x_i) - m_x(x_i)] \]

(32)

where the local means are modeled as a trend surface:

\[ m_x(x) = \sum_{b=0}^{B} \beta_b f_b(x) \]

(33)

where the functions \( \{f_b(x), b=0, B\} \) of known forms constitute the following column vector:

\[ f(x) = \begin{bmatrix} f_0(x) \\ \vdots \\ f_B(x) \end{bmatrix} \]

(34)

for which a quadratic trend surface example may be written as \( f_0(x) = 1, f_1(x) = x, f_2(x) = y, f_3(x) = x^2, f_4(x) = y^2, f_5(x) = xy \).

To develop the universal Kriging equation systems, denote trend coefficients vector \( \beta \) as:

\[ \beta = \begin{bmatrix} \beta_0 \\ \vdots \\ \beta_B \end{bmatrix} \]

(35)

and a matrix of column vectors \( f(x) \):

\[ F(x_i) = \begin{bmatrix} f_0(x_i) & \cdots & f_B(x_i) \\ \vdots & \ddots & \vdots \\ f_0(x_i) & \cdots & f_B(x_i) \end{bmatrix} \]

(36)

It is possible to write the residual at an unsampled location \( x \) as a linear combination below:

\[ \hat{z}(x) - f(x)^T \beta = \lambda^T z(x) - \lambda^T F(x)^T \beta \]

(37)
Re-arranging the equation above, we arrive at:
\[
\hat{z}(x) = \lambda^T z(x) + (f(x)^T - \lambda^T F(x)^T) \beta \tag{38}
\]

To account for the unknown \( \beta \), enforce a constraint for unbiased estimation in universal Kriging:
\[
f(x)^T - \lambda^T F(x)^T = 0 \tag{39}
\]

The following universal Kriging equation is derived
\[
\begin{bmatrix}
\text{cov}_s(x_i, x_j) & F(x_i)^T \\
F(x_i) & 0
\end{bmatrix}
\begin{bmatrix}
\hat{z} \\
-\psi
\end{bmatrix} = 
\begin{bmatrix}
\text{cov}_s(x_i, x) \\
F(x)
\end{bmatrix} \tag{40}
\]

with weights derived from the above, the universal Kriging is obtained:
\[
\hat{z}(x) = \sum_{i=1}^{n} \lambda_i z(x_i) \tag{41}
\]

while the Kriging variance is
\[
\sigma_z^2(x) = \text{var}[R(x)] - \lambda^T \text{cov}_s(x_i, x) + F(x)^T \psi \tag{42}
\]

3 Conclusion

Variants of Kriging have been described neatly by applying the optimum criteria of best linear unbiased estimation and linear algebra. Multivariate geostatistics is yet to be discussed in the second paper of this sequence, which will be useful for making use of cross-variable covariation in spatial prediction. Also, indicator geostatistics will be discussed for predicting the occurrences of categorical variables, such as certain soil types or classes of intervals of originally continuous variables.

Kriging is known to be able to generate not only optimal estimation for a regionalized variable at unsampled locations but also measures (i.e., Kriging variance) of precision concerning the estimation. The information we have about a spatially varying phenomenon is usually incomplete. Most often, only few samples of the variable under study are available, next to an abundance of indirect information gathered with remote sensing devices. This implies that one cannot determine with full confidence the exact unknown true outcome of that variable at every location. Geostatistics allows quantifying of the uncertainty about the unknown spatial phenomenon in terms of a set of alternative representations, termed realizations, of the unknown truth\(^{[11-13]}\). The aim of stochastic simulations is to enforce various properties on these realizations, which will be the focus of the third paper in the sequence.

References

[1] Matheron G (1962) Traité de géostatistique appliquée[M]. Paris: Editions Technip
[2] David M (1977) Geostatistical ore reserve estimation [M]. Amsterdam: Elsevier Scientific Publishing Company
[3] Journel A G, Huijbregts C (1978) Mining Geostatistics[M]. Liverpool: World Academic Press
[4] Isaaks E, Srivastava S (1989) An introduction to applied geostatistics[M]. Oxford: Oxford University Press
[5] Journel A G (1989) Fundamentals of geostatistics in five lessons[M]. Washington D C: American Geophysical Union
[6] Olea R (1991) Geostatistical glossary and multilingual dictionary[M]. Oxford: Oxford University Press
[7] Kitanidis P K (1997) Introduction to geostatistics: applications in Hydrogeology[M]. Cambridge: Cambridge University Press
[8] Chiles J-P, Delfiner P (1999) Geostatistics: modeling spatial uncertainty[M]. New York: Wiley & Sons
[9] Goovaerts P (1997) Geostatistics for natural resources estimation[M]. Oxford: Oxford University Press
[10] Caers J (2001) Geostatistical reservoir modelling using statistical pattern recognition[J]. Journal of Petroleum Science and Engineering, 29(3-4): 177-188
[11] Deutsch C V (2002) Geostatistical reservoir modeling [M]. Oxford: Oxford University Press
[12] Lantuéjoul C (2002) Geostatistical simulation: models and algorithms[M]. Berlin: Springer-Verlag
[13] Wackernagel H (2003) Multivariate geostatistics[M]. Berlin: Springer-Verlag