Mathematical Geology, Vol. 26, No. 5, 1994

Kriging with Strings of Data[1]

Clayton V. Deutsch[2]

The concept of a random function and, consequently, the application of kriging cells for the implicit assumption that the data locations are embedded within an infinite domain. An implication of this assumption is that, all else being equal, outlying data locations will receive greater weight because they are seen as less redundant, hence, more informative of the infinite domain. A two-step kriging procedure is proposed for correcting this string effect. The first step is to establish the total kriging weight attributable to each string. The distribution of that total weight to the samples in the string is accomplished by a second stage of kriging. In the second stage, a spatial redundancy measure $r_{(n)}$ is used in place of the covariance measure in the data–data kriging matrix. This measure is constructed such that each datum has the same redundancy with the (n) data of the string to which it belongs. This paper documents the problem of kriging with strings of data, develops the redundancy measure $r_{(n)}$, and presents a number of examples.

INTRODUCTION

The short scale variability in the spatial distribution of petrophysical variables is being increasingly modeled with geostatistical conditional simulation techniques (Haldorsen and Damsleth, 1989). That is, alternative high resolution 3-D realizations are generated which reflect the level of heterogeneity observed in the data. The level of heterogeneity or spatial variability is quantified by variogram/covariance models. Geostatistical simulation techniques make use of kriging to model local conditional distributions of the variable being simulated; simulation proceeds by sequential drawing from such conditional distributions (Deutsch and Journel, 1992, p. 123; Journel 1989).

Adopting a random function (RF) model $\{Z(\mathbf{u}), \mathbf{u} \in \text{study area } A\}$ and using kriging amounts to assume that the study area A is embedded within an infinite domain (Deutsch, 1993). At first glance, this assumption has no inherent limitations since all locations outside A are never explicitly considered. A practically important consequence, however, is reflected in the kriging weights as-

[1]Received 10 January 1994; accepted 18 May 1994.
[2]Exxon Production Research Co., P.O. Box 2189, Houston, Texas 77252-2189.

0882-8121/94/0700-0623$07.00/1

signed to data contiguously aligned along finite strings. As shown in Fig. 1, the outlying samples can receive a significantly greater weight than central samples. The variogram range is as long as the string; the location being estimated is beyond that range. This overweighting becomes less important as the point being estimated gets closer to the string and as the relative nugget effect increases (this effect is fully documented in Deutsch, 1993).

The reason for outlying samples to be overweighted is that the data–data (left hand side) kriging matrix sees such samples as less redundant than the more centrally located samples. Consider the data location set (n): $\{\mathbf{u}_\alpha, \alpha = 1, \ldots, n\}$, where $\mathbf{u}_1$ is a central location and $\mathbf{u}_n$ is an outlying location (see Fig. 2), then:

$$\bar{\rho}(\mathbf{u}_n, (n)) \leq \bar{\rho}(\mathbf{u}_1, (n)) \quad (1)$$

where $\rho(\mathbf{h})$ is the stationary correlogram of $RF\ Z(\mathbf{u})$, and

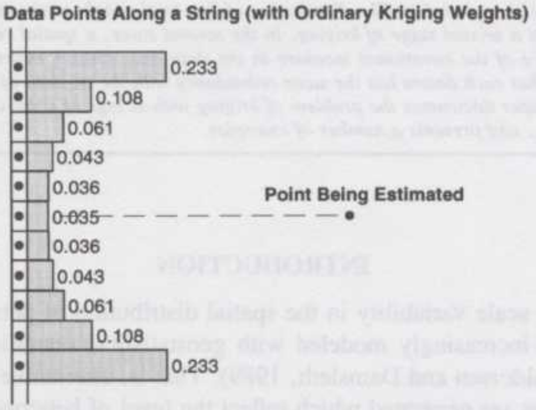


Fig. 1. Data points along a string with the ordinary kriging weights. A spherical variogram model was used with a 20% relative nugget effect and a range equal to the length of the string. The point being estimated is beyond that range (from Deutsch, 1993).

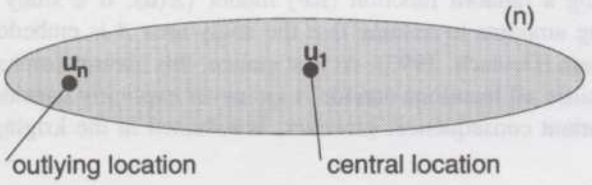


Fig. 2. Outlying and central locations within a set of data locations.

$$\bar{\rho}(\mathbf{u}_\alpha, (n)) = \frac{1}{n} \sum_{\beta=1}^{n} \rho(\mathbf{u}_\alpha - \mathbf{u}_\beta)$$

is the average correlation (redundancy of sample $z(\mathbf{u}_\alpha)$ with the data set (n)). The sample at the outlying location $\mathbf{u}_n$ is seen as less redundant than the sample at the central location $\mathbf{u}_1$. The kriging weight given to sample $z(\mathbf{u}_n)$ would then be increased correspondingly and artificially whereas, in actuality, that sample is no more or less redundant *in average over* the string (n) than any other sample. We are assuming that the string (n) has no internal clustering which would justify considering some samples as more redundant than others.

The artifact weighting is not as pronounced with simple kriging (SK). For example, SK applied in the case shown in Fig. 1 would give a weight of zero to all data points since the point being kriged is beyond the range of correlation. The artifact weighting appears, however, as the point gets within the range of correlation of the data.

This edge effect of kriging is practically important. Strings of data are frequently encountered in mining and petroleum applications where data are collected along drillholes or wells. Short strings are created when a ''long'' length of drillhole data is truncated by stratigraphic limits or the boundaries of a local search ellipsoid.

Although this odd weighting behavior is valid in a theoretical RF framework with no boundaries, it is not what a practitioner anticipates. In the case of a bounded stratigraphic layer, there is no volume of interest outside the limits justifying overweighting the end values. In the case of a limited search neighborhood, there are data beyond the limits of the data strings but these data need not be reflected by the end points of the search neighborhood. Moreover, one goal of ordinary kriging is to re-estimate locally the mean (Journel and Rossi, 1989); assigning more weight to the end points of strings would be contrary to that goal. The artifacts caused by this overweighting are exacerbated when the data values present a trend. In particular, a bias may ensue when the variable has relatively high or low values near the top and bottom of the string (often the case in contact controlled mineralization, or in fining upwards sequences).

EMPIRICAL SOLUTIONS

There are a number of empirical solutions to this problem (Deutsch, 1993). The first obvious solution is to use only two samples from any one string of data. In this case, both samples are equally redundant and equally informative. A disadvantage of this approach is the arbitrary elimination of valuable data.

Another idea is to wrap each finite string of data when building the data–data (left-hand side) kriging matrix. Consider n contiguous data $i = 1, \ldots, n$, aligned in a string, each separated from its neighbors by a vector $\mathbf{h}_s$. After wrapping the covariance between any two data points i and $j (i \geq j)$ is

$$C_{i,j} = C(h \cdot \mathbf{h}_s) \tag{2}$$

$$\text{with: } k = \min\{(j - i + n), (i - j)\}$$

It is important to note that none of the data locations are moved; the above calculation simply modifies the data–data covariance values

$$C(\mathbf{u}_\alpha - \mathbf{u}_\beta),\ \alpha = 1, \ldots, n,\ \beta = 1, \ldots, n$$

in the left-hand side kriging matrix so that all points in the string are equally redundant within that string.

The *ad hoc* partial solution of ''wrapping'' continuous strings of data has been adopted in other image processing and geophysical applications where the implicit assumption of an infinite domain has the same effect (Aki and Richards, 1980). Wrapping the data is sometimes referred to as the decision of ''circular stationarity.''

The assumption of an infinite domain is a fundamental aspect of any RF model and, consequently, of kriging. The kriging weights *always* reflect this assumption; the effects are more pronounced, however, when dealing with finite strings of contiguous data. The objective of this work is to present a modification to the kriging equations which eliminates these non-intuitive weights.

THE PROPOSED SOLUTION

The circular stationarity (or ''wrap the string'') solution consists of increasing the average redundancy term $\bar{\rho}(\mathbf{u}_n, (n))$ of the outlying sample at location $\mathbf{u}_n$ by reducing its Euclidean distance $\mathbf{u}_n - \mathbf{u}_\alpha$ to the other samples.

Capitalizing on the circular stationarity solution, the correlation function $\rho(\mathbf{h})$ could be corrected to ensure that all samples within the same string (n) have exactly the same average redundancy. Journel (private communication) has suggested that the correlation function be replaced by the measure:

$$r_{(n)}(\mathbf{u}, \mathbf{u}') = \rho(\mathbf{u}' - \mathbf{u}) + [\bar{\rho}(\mathbf{u}', (n)) - \bar{\rho}(\mathbf{u}, (n))] \qquad \forall \mathbf{u}, \mathbf{u}' \in (n) \tag{3}$$

where (n) is the set of locations of the data actually used in the kriging. Note that the set of data (n) may change from one location being estimated to another.

When kriging the unknown value $z(\mathbf{u}_0)$ from the dataset (n), the data–data left-hand side of the kriging system is constructed with $r_{(n)}$ instead of $\rho(\mathbf{h})$. The right-hand side correlation values are unchanged; they are the original data-to-unknown correlation values $\rho(\mathbf{u}_0 - \mathbf{u}_\alpha)$.

Some Properties of $r_{(n)}$

1. Using the redundancy measure (3), the average redundancy of any sample $z(\mathbf{u}_\alpha)$, $\mathbf{u}_\alpha \in (n)$, is seen to be constant:

$$\bar{r}_{(n)}(\mathbf{u}_\alpha, (n)) = \frac{1}{n} \sum_{\beta=1}^{n} r_{(n)}(\mathbf{u}_\alpha, \mathbf{u}_\beta)$$

$$= \bar{\rho}((n), \mathbf{u}_\alpha) + \bar{\rho}((n), (n)) - \bar{\rho}(\mathbf{u}_\alpha, (n)) = \bar{\rho}((n), (n))$$

$$\forall \mathbf{u}_\alpha \in (n) \tag{4}$$

2. The correction (3) amounts to increasing the redundancy of outlying samples, and decreasing that of central samples. Indeed:

$$r_{(n)}(\mathbf{u}, \mathbf{u}') - r_{(n)}(\mathbf{u}', \mathbf{u}) = 2[\bar{\rho}(\mathbf{u}', (n)) - \bar{\rho}(\mathbf{u}, (n))]$$

which is greater than 0 if $\bar{\rho}(\mathbf{u}', (n)) \geq \bar{\rho}(\mathbf{u}, (n))$, i.e., if $\mathbf{u}$ is more outlying within the string (n) than $\mathbf{u}'$. Note that the measure $r_{(n)}$ is asymmetric. This implies that $[1 - r_{(n)}(\mathbf{u} - \mathbf{u}')]$ is not a distance measure as opposed to the semivariogram $\gamma(\mathbf{h}) = 1 - \rho(\mathbf{h})$.

3. The redundancy measure $r_{(n)}(\mathbf{u}, \mathbf{u}')$ is exactly 1.0 for $\mathbf{u} = \mathbf{u}'$:

$$r_{(n)}(\mathbf{u}, \mathbf{u}') = \rho(0) = 1$$

In addition,

$$r_{(n)}(\mathbf{u}, \mathbf{u}') \leq 1 \qquad \forall \mathbf{u}, \mathbf{u}' \text{ and } (n)$$

Indeed, since $\gamma(\mathbf{h}) = 1 - \rho(\mathbf{h})$ is a distance measure:

$$\gamma(\mathbf{u} - \mathbf{u}_\alpha) \leq \gamma(\mathbf{u} - \mathbf{u}') + \gamma(\mathbf{u}' - \mathbf{u}_\alpha)$$

Thus:

$$1 - \rho(\mathbf{u} - \mathbf{u}_\alpha) \leq 2 - \rho(\mathbf{u} - \mathbf{u}') - \rho(\mathbf{u}' - \mathbf{u}_\alpha)$$

$$\rho(\mathbf{u} - \mathbf{u}_\alpha) \geq \rho(\mathbf{u} - \mathbf{u}') + \rho(\mathbf{u}' - \mathbf{u}_\alpha) - 1$$

Averaging over all $\mathbf{u}_\alpha \in (n)$ yields:

$$\bar{\rho}(\mathbf{u}, (n)) \geq \rho(\mathbf{u} - \mathbf{u}') + \bar{\rho}(\mathbf{u}', (n)) - 1$$

That is,

$$\rho(\mathbf{u} - \mathbf{u}') + \bar{\rho}(\mathbf{u}', (n)) - \bar{\rho}(\mathbf{u}, (n)) = r_{(n)}(\mathbf{u}, \mathbf{u}') \leq 1, \text{ qed}$$

Note, however, that $r_{(n)}(\mathbf{u}, \mathbf{u}')$ can be negative.

Positive Definiteness

The data–data matrix, constructed with the redundancy measure $r_{(n)}$ instead of the correlation measure ρ, must be positive semi-definite to ensure uniqueness and existence of a solution. That is,

$$\sum_{\alpha\in(n)}\sum_{\beta\in(n)}\lambda_\alpha\lambda_\beta r_{(n)}(\mathbf{u}_\alpha, \mathbf{u}_\beta) \geq 0 \quad \forall\lambda_\alpha, \lambda_\beta, (n)$$

Expanding in terms of ρ:

$$\begin{aligned}
&\sum_{\alpha\in(n)}\sum_{\beta\in(n)}\lambda_\alpha\lambda_\beta r_{(n)}(\mathbf{u}_\alpha, \mathbf{u}\beta)\\
&\quad = \sum_{\alpha\in(n)}\sum_{\beta\in(n)}\lambda_\alpha\lambda_\beta\rho(\mathbf{u}_\alpha, \mathbf{u}_\beta) + \sum_{\alpha\in(n)}\sum_{\beta\in(n)}\lambda_\alpha\lambda_\beta\bar{\rho}(\mathbf{u}_\beta, (n))\\
&\qquad - \sum_{\alpha\in(n)}\sum_{\beta\in(n)}\lambda_\alpha\lambda_\beta\bar{\rho}(\mathbf{u}_\alpha, (n))\\
&\quad = \sum_{\alpha\in(n)}\sum_{\beta\in(n)}\lambda_\alpha\lambda_\beta\rho(\mathbf{u}_\alpha, \mathbf{u}_\beta) + \left(\sum_{\alpha\in(n)}\lambda_\alpha\right)\cdot\sum_{\beta\in(n)}\lambda_\beta\bar{\rho}(\mathbf{u}_\beta, \mathbf{u}_\alpha)\\
&\qquad - \left(\sum_{\beta\in(n)}\lambda_\beta\right) + \sum_{\alpha\in(n)}\lambda_\alpha\bar{\rho}(\mathbf{u}_\alpha, \mathbf{u}_\beta)\\
&\quad = \sum_{\alpha\in(n)}\sum_{\beta\in(n)}\lambda_\alpha\lambda_\beta\rho(\mathbf{u}_\alpha, \mathbf{u}\beta) \geq 0, \quad \forall\lambda_\alpha, \lambda\beta, \forall(n) \qquad (5)
\end{aligned}$$

Therefore, $r_{(n)}$ is positive semi-definite if ρ is a positive semi-definite correlation function.

Finite Domain Kriging

Correction for the string, or finite domain, effect is done by replacing the data–data covariance values $\rho(\mathbf{u}_\alpha - \mathbf{u}_\beta)$ by the redundancy measure $r_{(n)}(\mathbf{u}_\alpha - \mathbf{u}_\beta)$, leaving the data-unknown right-hand side vector unchanged. The corrected estimator is:

$$Z^*(\mathbf{u}_0) = \sum_{\alpha=1}^{n}\nu_\alpha Z(\mathbf{u}_\alpha) \qquad (6)$$

with the corrected kriging system:

$$\begin{cases}\sum_{\beta=1}^{n}\nu_\beta r_{(n)}(\mathbf{u}_\beta, \mathbf{u}_\alpha) + \mu_1 = \rho(\mathbf{u}_0 - \mathbf{u}_\alpha), \ \alpha = 1, \ldots, n\\ \sum_{\beta=1}^{n}\nu_\beta = 1\end{cases} \qquad (7)$$

where μ_1 is a Lagrange parameter accounting for the constraint on the weights ν_β.

The measure $r_{(n)}$ is positive semi-definite, thus, system (7) yields a unique solution as long as no two data locations coincide.

The price for the correction of ρ into $r_{(n)}$ is the loss of the kriging property of exactitude. Data exactitude would call for $\nu_{\alpha 0} = 1$, $\nu_\beta = 0$, $\forall \beta \neq \alpha_0$, whenever the location $\mathbf{u}_0$ identifies a datum location $\mathbf{u}_{\alpha 0} \in (n)$. System (7) would then be written:

$$\begin{cases} \mu_1 = 0, \text{ when } \alpha = \alpha_0 \\ r_{(n)}(\mathbf{u}_{\alpha 0}, \mathbf{u}_\alpha) = \rho(\mathbf{u}_\alpha - \mathbf{u}_{\alpha 0}), \ \forall \alpha \neq \alpha_0 \end{cases} \tag{8}$$

which is true only for $r_{(n)} = \rho$, i.e., if no correction is applied to the original correlation function $\rho(\mathbf{h})$.

Even if estimation is not performed at data locations, the lack of exactitude may create artifact discontinuities next to data locations (see later section). One may consider restoring data exactitude by extending the correction to the right-hand side of the kriging system (7); however, it can be shown that this amounts to reverting to the traditional ordinary kriging system, and hence forfeiting the string effect correction.

Estimation Variance

The estimation variance of the linear estimator $Z^*(\mathbf{u}_0)$ is:

$$\begin{aligned} \sigma_{E_1}^2 &= E\{[Z(\mathbf{u}_0) - Z^*(\mathbf{u}_0)]^2\} \\ &= 1 - 2\sum_\alpha \nu_\alpha \rho(\mathbf{u}_0 - \mathbf{u}_\alpha) + \sum_\alpha \sum_\beta \nu_\alpha \nu_\beta \rho(\mathbf{u}_\beta - \mathbf{u}_\alpha) \end{aligned}$$

Accounting for relation (5)

$$\sum_\alpha \sum_\beta \nu_\alpha \nu_\beta \rho(\mathbf{u}_\beta - \mathbf{u}_\alpha) = \sum_\alpha \sum_\beta \nu_\alpha \nu_\beta r_{(n)}(\mathbf{u}_\beta, \mathbf{u}_\alpha) = \sum_\alpha \nu_\alpha \rho(\mathbf{u}_o - \mathbf{u}_\alpha) - \mu_1$$

Thus:

$$\sigma_{E_1}^2 = 1 - \sum_\alpha \nu_\alpha \rho(\mathbf{u}_0 - \mathbf{u}_\alpha) - \mu_1 \geq 0 \tag{9}$$

This estimation variance calculated from the correlation $\rho(\mathbf{h})$ is necessarily non-negative, even with the weights ν_α and Lagrange parameter μ_1 determined from the modified kriging system (7) instead of the traditional OK system. At a datum location $\mathbf{u}_0 = \mathbf{u}_{\alpha 0}$, however, the estimation variance (9) may not be zero since, in general, $z^*(\mathbf{u}_0) \neq z(\mathbf{u}_0)$.

THE CASE OF MULTIPLE STRINGS

If the data configuration includes several strings, Journel's redundancy measure (3) should be applied to each string separately and, consequently, kriging of type (7) should be performed on each.

More precisely, consider a dataset consisting of L strings of n_l data each; hence: $n = \Sigma_l n_l$. A modified kriging system of type (7) is applied to each string of data, leading to the estimator for string l:

$$Z^{*(l)}(\mathbf{u}_o) = \sum_{\alpha=1}^{n_l} \lambda_\alpha^{(l)} Z(\mathbf{u}_\alpha^{(l)}) \tag{10}$$

using the (n_l) locations $\mathbf{u}_\alpha^{(l)}$ in string l. The corresponding corrected kriging system is:

$$\begin{cases} \sum_{\beta=1}^{n_l} \nu_\beta^{(l)} r_{(n)}(\mathbf{u}_\beta^{(l)}, \mathbf{u}_\alpha^{(l)}) + \mu^{(l)} = \rho(\mathbf{u}_0 - \mathbf{u}_\alpha^{(l)}), \ \alpha = 1, \ldots, n_l \\ \sum_{\beta=1}^{n} \nu_\beta^{(l)} = 1 \end{cases} \tag{11}$$

Next, consider the L string average values:

$$\overline{Z}_{(l)} = \frac{1}{n_l} \sum_{\alpha=1}^{n_l} Z(\mathbf{u}_\alpha^{(l)}), \quad l = 1, \ldots, L$$

and kriging using these average values:

$$\hat{Z}(\mathbf{u}_o) = \sum_{\alpha=1}^{L} \omega_l \overline{Z}_{(l)} \tag{12}$$

with the (uncorrected) kriging system:

$$\begin{cases} \sum_{l'=1}^{L} \omega_{l'} \overline{\rho}(l', l) + \mu = \overline{\rho}(\mathbf{u}_0, l), \quad l = 1, \ldots, L \\ \sum_{l'=1}^{L} \omega_{l'} = 1 \end{cases} \tag{13}$$

$$\text{and: } \overline{\rho}(\mathbf{u}_0, l) = \frac{1}{n_l} \sum_{\alpha=1}^{n_l} \rho(\mathbf{u}_o - \mathbf{u}_\alpha^{(l)})$$

$$\overline{\rho}(l', l) = \frac{1}{n_l' n_l} \sum_{\beta=1}^{n_l'} \sum_{\alpha=1}^{n_l} \rho(\mathbf{u}_\beta^{(l')} - \mathbf{u}_\alpha^{(l)})$$

Last, use the weights ω_l to recombine the L estimators (10) which were corrected for their respective string effect:

$$Z^{**}(\mathbf{u}_o) = \sum_{l=1}^{L} \omega_l Z^{*(l)}(\mathbf{u}_o) \tag{14}$$

Remarks

- The estimator (14) can be written as:

$$Z^{**}(\mathbf{u}_o) = \sum_{l=1}^{L} \omega_l \sum_{\alpha=1}^{n_l} \lambda_\alpha^{(l)} Z(\mathbf{u}_\alpha^{(l)}) = \sum_{\alpha=1}^{n} k_\alpha Z(\mathbf{u}_\alpha) \tag{15}$$

with $k_\alpha = \omega_l \lambda_\alpha^{(l)}$ if $\mathbf{u}_\alpha \in$ string l, and $\Sigma_{\alpha=1}^{n} k_\alpha = \Sigma_l \omega_l \Sigma_\alpha \lambda_\alpha^{(l)} = 1$. This estimate differs from both the traditional (uncorrected) OK using all n data, and a globally corrected kriging that would use a unique redundancy measure $r_{(n)}$ as in system (7).

- Since none of the L estimators (10) are exact, the recombined estimator (14) is not exact. Again, the string effect correction comes at the price of loss of the exactitude property.
- Using the weights k_α, $\alpha = 1, \ldots, n$, *and* the correlogram $\rho(\mathbf{h})$, one can calculate the estimation variance corresponding to estimator (14).

SOME EXAMPLES

Single String Case

Figure 3 compares the kriging weights obtained with ordinary kriging and with "finite domain kriging" (FDK) using the $r_{(n)}$ redundancy measure. In both cases, the variogram model considered is spherical with a 20% nugget effect and a range equal to the length of the string. Note how "finite domain kriging" reduces the artificially higher ordinary kriging weight given to the end sample of the string; as expected, when the point to be estimated is beyond the range of correlation, all corrected weights are equal (0.091 = 1/11).

To evaluate the loss of exactitude, the evolution of the central datum weight (λ_1 for OK, ν_1 for FDK) and the extreme datum weight (λ_6 for OK, ν_6 for FDK) as the point being estimated is farther away from the string is graphed in Fig. 4. When the point being estimated, $\mathbf{u}_0$, coincides with the central datum location $\mathbf{u}_1$, the OK weight is one and all other OK weights are zero ensuring exactitude of the OK estimator; in contrast, the FDK weight ν_1 is 1.056 and ν_6 is -0.142. The OK weight given to the sample at $\mathbf{u}_1$ drops suddenly from 1.0 to less than 0.5 for $|\mathbf{u}_0 - \mathbf{u}_1| > 0$; the OK exactitude is obtained by a discontinuity at the datum location. When the distance $|\mathbf{u}_0 - \mathbf{u}_1|$ reaches the correlation range, all FDK weights become $\nu_1 = \nu_6 = 0.091$, while the OK weights show the string effect: $\lambda_6 = 0.233 >> \lambda_1 = 0.035$. The question is: Of the loss of exactitude or the string effect, which is most harmful in practical applications?

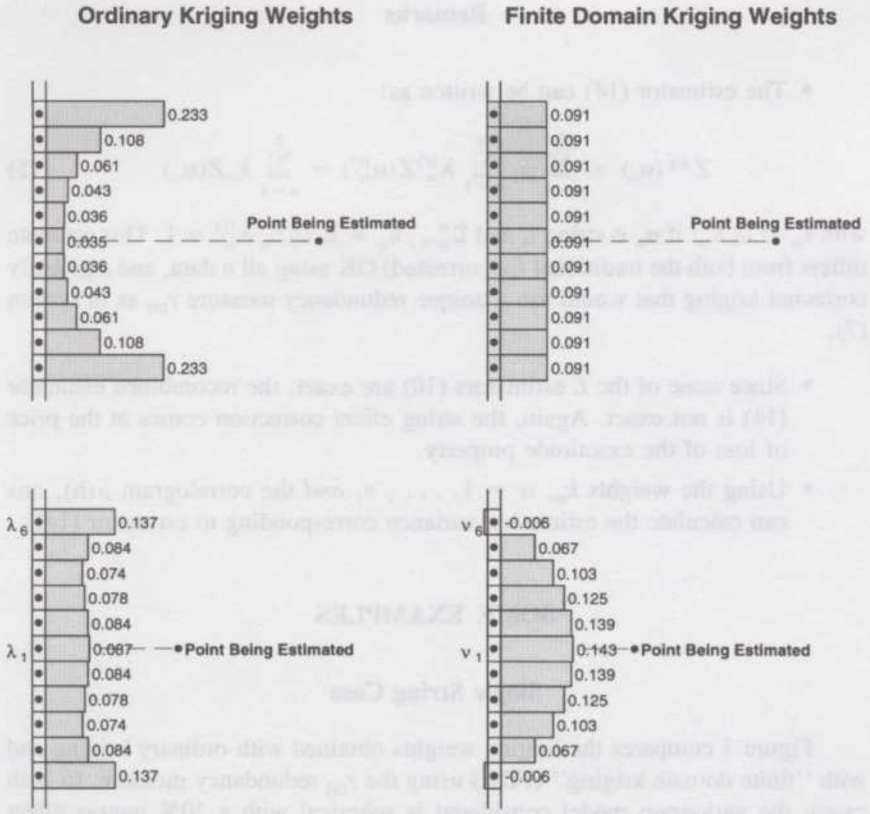


Fig. 3. The ordinary kriging weights and the kriging weights using $r_{(\alpha)}$ in the data–data covariance matrix are shown. In the top case the point being estimated is beyond the range, while in the bottom case the point being estimated is within the range of correlation of the data.

Multiple Strings Case

Figure 5 relates to a data configuration containing two strings of equal size (five data each). The point being estimated is 3 units away from the left string, 6 units away from the right string. The string length is 5 units. The variogram model used is spherical with zero nugget and range equal to 18 units, so any location of the configuration of Fig. 5 is within the range of any other location. The string effect is observed on the OK weights of both strings; it is corrected by FDK to the point of generating slightly negative weights (−0.00820) for the end points in the left string. The sum of the OK weights for the left string (0.692) is marginally different from the corresponding sum of the FDK weights (0.695), that is FDK simply redistributes the total OK weight for the string to correct for the string effect.

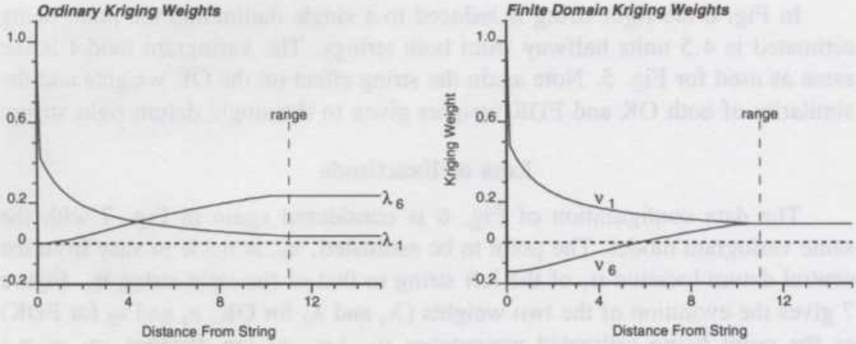


Fig. 4. Variation of the central weights (λ_1, ν_1) and the extreme weights (λ_6, ν_6) as the distance $|\mathbf{u}_0 - \mathbf{u}_6|$ between the point being estimated and the central datum location increases. OK results are on the left, FDK results are on the right.

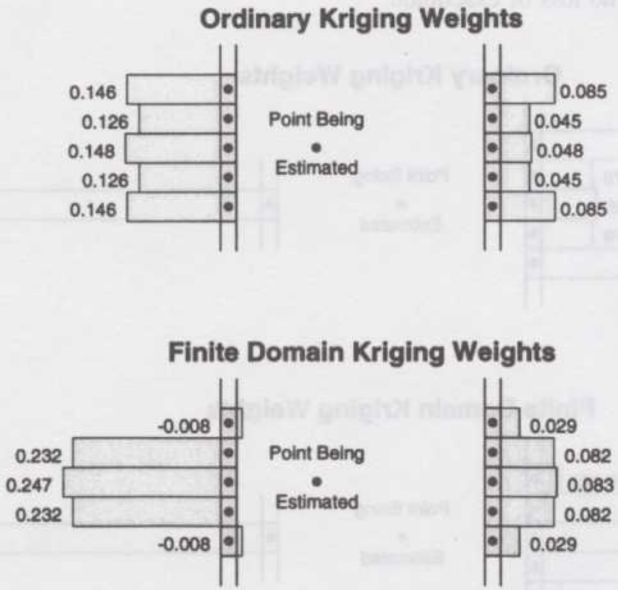


Fig. 5. The ordinary kriging weights and the finite domain kriging weights for two strings of equal size. The point being estimated is 3 units away from the left string and 6 units away from the right string. The variogram model is spherical with zero nugget effect and 18 units range.

In Fig. 6 the right string is reduced to a single datum and the point being estimated is 4.5 units halfway from both strings. The variogram model is the same as used for Fig. 5. Note again the string effect on the OK weights and the similarity of both OK and FDK weights given to the single datum right string.

Loss of Exactitude

The data configuration of Fig. 6 is considered again in Fig. 7 with the same variogram model. The point to be estimated, $\mathbf{u}_0$, is made to vary from the central datum location $\mathbf{u}_1$ of the left string to that of the right string $\mathbf{u}_2$. Figure 7 gives the evolution of the two weights (λ_1 and λ_2 for OK, ν_1 and ν_2 for FDK) as the point being estimated approaches $\mathbf{u}_2$, i.e., as the distance $|\mathbf{u}_0 - \mathbf{u}_1|$ increases from zero to 9 units. The variogram model is the same as used for Fig. 5 and 6, i.e., spherical with zero nugget and 18 units range. Ordinary kriging is an exact interpolar in that for $\mathbf{u}_0 = \mathbf{u}_1$, $\lambda_1 = 1$ and $\lambda_i = 0$, $\forall i \neq 1$. In contrast for $\mathbf{u}_0 = \mathbf{u}_1$, the FDK weights are $\nu_1 = 1.22$ and $\nu_2 = -0.020$. At the other extreme when $\mathbf{u}_0 = \mathbf{u}_2$, both OK and FDK estimators are exact ($\lambda_2 = \nu_2 = 1$, $\lambda_1 = \nu_1 = 0$). Indeed, since the right string is reduced to a single datum, there is no redistribution of the weight over that string done by FDK, and hence no loss of exactitude.

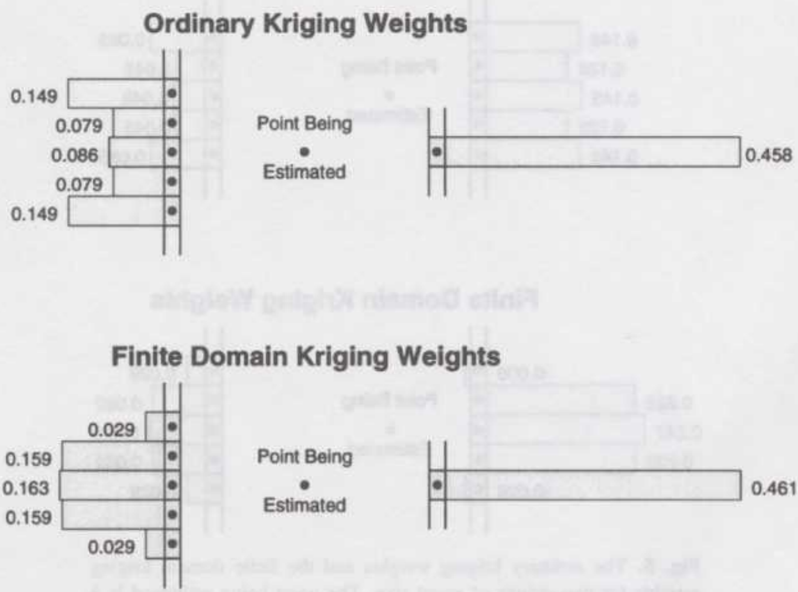


Fig. 6. The ordinary kriging weights and the finite domain kriging weights for two strings of unequal size. The point being estimated is 4.5 units halfway from both strings. The variogram model is spherical with zero nugget effect and 18 units range.

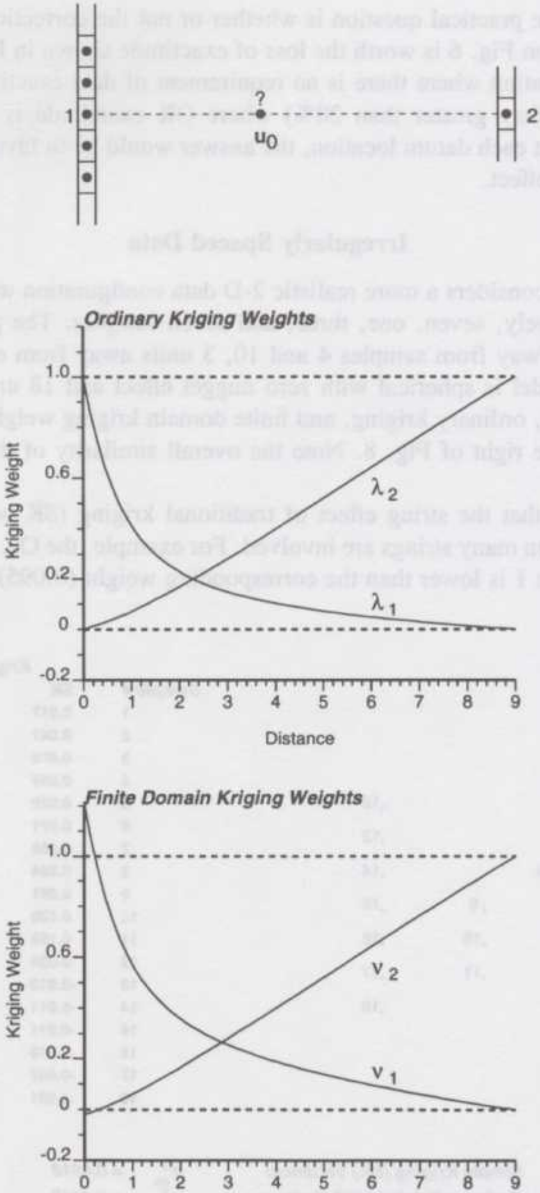


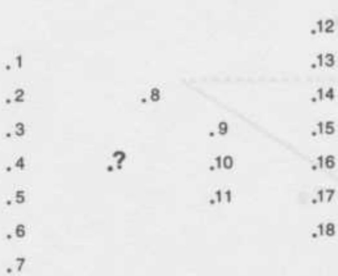
Fig. 7. Evolution of the weights for the left (λ_1, ν_1) and the right (λ_2, ν_2) central data points as the distance $|\mathbf{u}_0 - \mathbf{u}_1|$ between the point being estimated and the central datum location in the left string increases. OK results are on the top, FDK results are below.

Again, the practical question is whether or not the correction of the string effect as seen on Fig. 6 is worth the loss of exactitude shown in Fig. 7. In cases of block estimation where there is no requirement of data exactitude, or a high nugget effect (say greater than 20%) where OK exactitude is obtained by a discontinuity at each datum location, the answer would be in favor of correcting for the string effect.

Irregularly Spaced Data

Figure 8 considers a more realistic 2-D data configuration with four strings with, respectively, seven, one, three, and seven samples. The point being estimated is halfway from samples 4 and 10, 3 units away from either one. The variogram model is spherical with zero nugget effect and 18 units range. The simple kriging, ordinary kriging, and finite domain kriging weights are given in the table at the right of Fig. 8. Note the overall similarity of the three sets of weights.

It seems that the string effect of traditional kriging (SK or OK) is much attenuated when many strings are involved. For example, the OK weight (0.015) for end sample 1 is lower than the corresponding weight (0.095) for the central



Sample #	Kriging Weights SK	OK	FDK
1	0.017	0.015	0.003
2	0.047	0.046	0.070
3	0.078	0.077	0.102
4	0.096	0.095	0.114
5	0.090	0.090	0.103
6	0.071	0.071	0.071
7	0.058	0.056	0.003
8	0.264	0.264	0.242
9	0.091	0.091	0.120
10	0.120	0.121	0.144
11	0.155	0.155	0.120
12	-0.024	-0.025	-0.005
13	-0.013	-0.013	-0.012
14	-0.011	-0.011	-0.015
15	-0.011	-0.011	-0.017
16	-0.010	-0.011	-0.017
17	-0.007	-0.007	-0.015
18	-0.001	-0.003	-0.011

Simple Kriging (SK) Variance: $\sigma^2_{SK} = 0.4918$
Ordinary Kriging (OK) Variance: $\sigma^2_{OK} = 0.4919$
Finite Domain Kriging (FDK) Variance: $\sigma^2_{FDK} = 0.4985$

Fig. 8. Irregularly spaced data configuration: the 18 simple kriging, ordinary kriging, and finite domain kriging weights are given in the table. The variogram model is spherical with zero nugget effect and 18 units range.

sample 4, yet the FDK weight for that end sample 1 is even lower (0.003), that is FDK is doing its job. Samples 10 and 11 show a pronounced string effect (0.155 > 0.121 for OK) corrected by FDK (0.120 < 0.144).

Note that because each string is considered separately in FDK, see Eq. (10), there is a symmetry of FDK weights within each string, e.g., FDK weight 9 is the same as FDK weight 11; this is not the case for SK or OK.

The estimation variances corresponding to the three sets of weights are given at the bottom of Fig. 8. As expected, the smallest estimation variance is that of SK followed by OK then FDK; however, the FDK variance is only marginally greater than the OK variance.

CONCLUSIONS

The implicit assumption of an infinite domain as called for by the probabilistic random function model causes kriging to give proportionally higher weights to end samples of data strings. This effect is called the "string effect." The string effect arises because kriging sees these end samples as less redundant than central samples, everything else being equal. The string effect appears more pronounced when there are a few isolated strings than if there is a cluster of strings.

If there is no reason to consider that such end samples are more representative of the actual area of investigation, the string effect must be corrected. A redundancy measure which considers all data within a string as equally redundant is proposed to replace the traditional correlogram function. This redundancy measure is to be used only in the left-hand side matrix of the kriging systems which characterizes the redundancy between data, and the traditional correlogram function should be kept for building the right-hand side correlation values between data and the unknown.

Introduction of this redundancy measure allows correction of the string effect at the cost of loss of the exactitude property of traditional simple and ordinary kriging. The loss of exactitude is generally small and, more importantly, irrelevant if block kriging or kriging in the presence of an appreciable nugget effect. In contrast, the string effect of traditional kriging may create severe artifacts or even biases if the end samples of strings tend to be richer or poorer than central samples, which is a case often encountered when the strings represent drillholes intersecting a complete mineralization body with end samples being generally poorer in grade.

ACKNOWLEDGMENTS

The author would like to thank Professor A. G. Journel of Stanford University for his original suggestion of the spatial redundancy measure $r_{(n)}$ and helpful discussions on this subject.

REFERENCES

Aki, K., and Richards, P., 1980, *Quantitative Seismology, Theory and Methods* (Vol. I and II): W.H. Freeman and Co., San Francisco.

Deutsch, C., 1993, Kriging in a Finite Domain: Math Geol., v. 25, n. 1, p. 41–52.

Deutsch C., and Journel, A., 1992, *GSLIB: Geostatistical Software Library and User's Guide*: Oxford University Press, New York.

Haldorsen, H., and Damsleth, E., 1990, Stochastic Modeling: J. Pet. Technol., p. 404–412.

Journel, A., 1989, *Fundamentals of Geostatistics in Five Lessons. Volume 8 Short Course in Geology*: American Geophysical Union, Washington, D.C.

Journel, A., and Huijbregts, C. J., 1978, *Mining Geostatistics*: Academic Press, New York.

Journel, A., and Rossi, M., 1989, When Do We Need a Trend Model in Kriging? Math Geol., v. 21, n. 7, p. 715–739.