Some Implementation Aspects of Kriging with a Trend

Jason A. McLennan and Clayton V. Deutsch

Centre for Computational Geostatistics (CCG)
Department of Civil and Environmental Engineering
University of Alberta

Kriging is still a valuable estimation technique. A map of kriged estimates is based on the available data and provides a smoother representation of variability than the truth. Kriged estimates are appropriate for detecting and quantifying large-scale geological trends. Universal kriging (UK) or kriging with a trend (KT) provides the least-squares estimate of the attribute of interest concurrently with the least-squares fit of a previously established deterministic trend function. There are, however, several significant challenges facing the implementation of universal kriging algorithm. Four issues are identified and described in this work: (1) the spatial law and stationarity, (2) simulation, (3) solution existence or singularity, and (4) data conditioning. Particular attention is given to the issue of choosing the correct variogram model.

Kriging Background

Early methods of estimation included hand contouring, polygonal estimation, triangulation, inverse distance and moving window averages. A description and comparative study of these conventional estimation techniques is given in [1]. Estimation algorithms have evolved to more sophisticated approaches based on optimization. One particular problem with the earliest subjective mapping techniques was the sacrifice of local or conditional bias for global unbiasedness. Conditional bias is the systematic overestimation of high grades and underestimation of low grades [2]. Correcting this bias was of great concern to the pioneers of geostatistics. In ore reserve valuation, the existence of truly lower grade stopes where higher grades were predicted was obviously undesirable. The pioneering work of Danie Krige during the 1950s to correct conditional bias was the seed for the currently most popular group of estimation techniques collectively referred to as kriging [3]. The theory of kriging was formalized by Georges Matheron in 1961.

The kriging estimator is often referred to with the **BLUE** acronym: **B**est since the estimate is optimum in that the error variance is a minimum; **L**inear since the estimate is a weighted linear combination of surrounding data; **U**nbiased since the expected value of the estimate and true value are the same; and **E**stimate.

From the 1960s, kriging was used both for large scale trend visualization and calculating recoverable reserves for mine planning and economic forecasting. The classic application of kriging to mine planning is discussed in [4]. From the early 1980s until present, however, the use of kriging or any type of estimation for production planning is no longer as strongly recommended due to the inherent smoothing effect and persistence of conditional bias [5] in these methods. Simulation is becoming increasingly popular for production planning due to the ability to reproduce the correct amount of spatial variability through multiple equally probable geological realizations.

The Place of Kriging

Kriging provides an estimate and an estimation variance at unsampled locations. Under the assumption of a multivariate Gaussian distribution, the local kriging estimate and estimation variance identify the mean and variance and full parameterization of the local Gaussian ccdf. This is sometimes referred to as Multi-Gaussian (MG) kriging. If all that is required is local uncertainty, MG kriging is appropriate. The local Gaussian distributions of uncertainty are back transformed to original units.

The assessment of joint uncertainty within some arbitrary volume requires sampling the multivariate distribution of multiple locations with simulation. Kriging is often applied in a multivariate Gaussian context; however, values are sampled at each location rather than retaining a single best estimate. The simulated values can then be combined to obtain a realization at the arbitrary support volume. The addition of random residuals is repeated for a prescribed number of equally probable realizations from which a distribution of uncertainty can be constructed.

Kriging is also used for detecting and quantifying large-scale geological trends. The random function theory underlying the derivation of kriging equations is amenable to a wide variety of estimation schemes using conditioning data of different types, volume support, and quality [8].

The discussion in this paper is aimed at universal kriging (UK). UK is sometimes referred to as kriging with a trend (KT). UK amounts to specify the functional form of the locally varying mean model, then optimal estimation proceeds with an implicit fitting of this functional form with minimum squared error. The UK algorithm allows the incorporation of varying types of deterministic trend information in the estimation of a single geological attribute of interest. There are a number of implementation challenges.

Kriging with a Trend

Universal Kriging (UK) was developed by Matheron in 1969 [9]. The trend is a predetermined deterministic function of the coordinates. The description below highlights the UK equations. Consider a random function (RF) $Z(\mathbf{u})$ in a domain $\mathbf{D}$ composed of the set of random variables (RV), $\{Z(\mathbf{u}_s)$, for all $\mathbf{u}_s \in \mathbf{D}\}$. UK implies the spatial distribution of the continuous random function $Z(\mathbf{u})$ is of dual character: partly structured and partly stochastic. This notion of dual character can be represented analytically within the RF through the following additive decomposition:

$$Z(\mathbf{u}) = m(\mathbf{u}) + R(\mathbf{u}) \tag{1}$$

where $m(\mathbf{u})$ is the structured component or trend and $R(\mathbf{u})$ is the random component or residual. This decision to split the spatial variability observed into a smoothly varying trend component and a random component is arbitrary [10]. Moreover, the particular additive decomposition in (1) is a necessary implication of the kriging algorithm, that is, all kriging imply an additive decomposition.

The kriging estimator at an unsampled location $\mathbf{u}_0$ takes on the form:

$$z^*_{\mathrm{K}}(\mathbf{u}_0) = \mathrm{A} + \sum_{s=1}^{n} \lambda_{\mathrm{K}}(\mathbf{u}_s) \cdot z(\mathbf{u}_s) \tag{2}$$

A is a constant shift parameter and the $\lambda_K(\mathbf{u}_s)$'s are the kriging weights assigned to the n surrounding $z(\mathbf{u}_s)$ sample data. The estimate $z^*_K(\mathbf{u}_0)$ and data $z(\mathbf{u}_s)$ can also be represented in probabilistic notation:

$$Z^*_K(\mathbf{u}_0) = \mathrm{A} + \sum_{s=1}^{n} \lambda_K(\mathbf{u}_s) \cdot Z(\mathbf{u}_s) \tag{3}$$

The actual error of estimation $e(\mathbf{u}_0)$ is:

$$e(\mathbf{u}_0) = z(\mathbf{u}_0) - z^*_K(\mathbf{u}_0) \tag{4}$$

Little can be done about this error unless its probabilistic version $E(\mathbf{u}_0)$ is considered:

$$E(\mathbf{u}_0) = Z(\mathbf{u}_0) - Z^*_K(\mathbf{u}_0) \tag{5}$$

In this case, the expected value and variance of $E(\mathbf{u}_0)$ can be calculated and thus acted upon. In particular, we require the expected value of $E(\mathbf{u}_0)$ to be zero for unbiasedness and the variance of $E(\mathbf{u}_0)$ to be a minimum for optimality. The expected value of $E(\mathbf{u}_0)$ is:

$$\begin{aligned} \mathrm{E}\{E(\mathbf{u}_0)\} &= \mathrm{E}\{Z(\mathbf{u}_0)\} - \mathrm{E}\{\mathrm{A}\} - \mathrm{E}\left\{\sum_{s=1}^{n} \lambda_K(\mathbf{u}_s) z_K(\mathbf{u}_s)\right\} \\ &= m(\mathbf{u}_0) - \mathrm{A} - \sum_{s=1}^{n} \lambda_K(\mathbf{u}_s) m(\mathbf{u}_s) \end{aligned} \tag{6}$$

In order for the kriging estimator $Z^*_K(\mathbf{u}_0)$ to be unbiased, this expected error must be zero. For this, the shift parameter A is set to:

$$\mathrm{A} = m(\mathbf{u}_0) - \sum_{s=1}^{n} \lambda_K(\mathbf{u}_s) m(\mathbf{u}_s) \tag{7}$$

$\mathrm{E}\{E(\mathbf{u}_0)\}$ is now zero. The kriging estimator in (3) is then:

$$\begin{aligned} Z^*_K(\mathbf{u}_0) &= m(\mathbf{u}_0) - \sum_{s=1}^{n} \lambda_K(\mathbf{u}_s) m(\mathbf{u}_s) + \sum_{s=1}^{n} \lambda_K(\mathbf{u}_s) Z(\mathbf{u}_s) \\ &= m(\mathbf{u}_0) + \left(\sum_{s=1}^{n} \lambda_K(\mathbf{u}_s) \left[Z(\mathbf{u}_s) - m(\mathbf{u}_s) \right] \right) \\ R^*(\mathbf{u}_0) &= \sum_{s=1}^{n} \lambda_K(\mathbf{u}_s) R(\mathbf{u}_s) \end{aligned} \tag{8}$$

The kriging estimator then appears as the result of linear estimation of the residual value $r(\mathbf{u}_0)$ from the $r(\mathbf{u}_s)$ residual data. This is true for any type of kriging. This estimator requires the decomposition (1). Although this decomposition is arbitrary in the sense that there is no physical evidence for it or actual $r(\mathbf{u}_s)$ sample data, it is necessary for an estimator to be formulated.

The other moment of $E(\mathbf{u}_0)$ required to develop the kriging equations is the variance:

$$\begin{aligned} \mathrm{VAR}\{E(\mathbf{u}_0)\} &= \sum_{s=0}^{n} \sum_{s'=0}^{n} \lambda_K(\mathbf{u}_s) \lambda_K(\mathbf{u}_{s'}) C_R(\mathbf{u}_s - \mathbf{u}_{s'}) \\ &= C_R(\mathbf{0}) - 2\sum_{s=1}^{n} \lambda_K(\mathbf{u}_s) C_R(\mathbf{u}_0 - \mathbf{u}_s) + \sum_{s=1}^{n} \sum_{s'=1}^{n} \lambda_K(\mathbf{u}_s) \lambda_K(\mathbf{u}_{s'}) C_R(\mathbf{u}_s - \mathbf{u}_{s'}) \end{aligned} \tag{9}$$

The error variance is interpreted as the sum of the overall variance (first term), closeness (second term), and redundancy (last term) of the nearby $r(\mathbf{u}_s)$ residual data. The error variance in (9) is general for all types of kriging. Different flavors of kriging correspond to different models for the mean $m(\mathbf{u})$, which require different constraints to attain unbiasedness in (6) and different procedures for minimizing the error variance in (9).

The current approach to UK or KT is to assume the $m(\mathbf{u})$ component is a smoothly varying deterministic function of the coordinates vector $\mathbf{u}$ whose unknown parameters are fit from the data within local search windows:

$$m(\mathbf{u}) = \sum_{v=0}^{V} a_v(\mathbf{u})\, f_v(\mathbf{u}) \tag{10}$$

The $f_v(\mathbf{u})$'s are known and constant functions of the coordinate vectors over the domain $\mathbf{D}$. The $a_v(\mathbf{u})$'s are estimated and constant within local search windows centered on the unsampled locations. The actual $m(\mathbf{u})$ trend values are unknown since the $a_v(\mathbf{u})$'s are also unknown.

The shift parameter A then becomes:

$$\mathrm{A} = \sum_{v=0}^{V} a_v(\mathbf{u}_0)\, f_v(\mathbf{u}_0) - \sum_{s=1}^{n} \lambda_{\mathrm{UK}}(\mathbf{u}_s) \sum_{v=0}^{V} a_v(\mathbf{u}_s)\, f_v(\mathbf{u}_s) \tag{11}$$

And the universal kriging estimator $Z^*_{\mathrm{UK}}(\mathbf{u}_0)$ is then:

$$\begin{aligned} Z^*_{\mathrm{UK}}(\mathbf{u}_0) &= \sum_{v=0}^{V} a_v(\mathbf{u}_0)\, f_v(\mathbf{u}_0) - \sum_{s=1}^{n} \lambda_{\mathrm{UK}}(\mathbf{u}_s) \sum_{v=0}^{V} a_v(\mathbf{u}_s)\, f_v(\mathbf{u}_s) + \sum_{s=1}^{n} \lambda_{\mathrm{UK}}(\mathbf{u}_s) Z(\mathbf{u}_s) \\ &= \sum_{v=0}^{V} a_v(\mathbf{u}_0)\, f_v(\mathbf{u}_0) + \left(\sum_{s=1}^{n} \lambda_{\mathrm{UK}}(\mathbf{u}_s) \left[Z(\mathbf{u}_s) - \sum_{v=0}^{V} a_v(\mathbf{u}_s)\, f_v(\mathbf{u}_s) \right] \right) \end{aligned} \tag{12}$$

There are many ways to ensure the universal kriging estimator $Z^*_{\mathrm{UK}}(\mathbf{u}_0)$ is unbiased. The classic approach is to impose the following $V + 1$ constraint equations:

$$\sum_{s=1}^{n} \lambda_{\mathrm{UK}}(\mathbf{u}_s)\, f_v(\mathbf{u}_s) = f_v(\mathbf{u}_0) \quad v = 0, \ldots, V \tag{13}$$

where $f_v(\mathbf{u}_0)$ are the monomial trend functions evaluated at the unsampled locations within $\mathbf{D}$ and $f_v(\mathbf{u}_s)$ are the monomial trend functions evaluated at the sampled locations. By considering these constraints (13), the resulting universal kriging estimator $Z^*_{\mathrm{UK}}(\mathbf{u}_0)$ is then unbiased:

$$
\begin{aligned}
\mathrm{E}\{E(\mathbf{u}_0)\} &= \mathrm{E}\{Z(\mathbf{u}_0)\} - \mathrm{E}\{Z^*_{\mathrm{UK}}(\mathbf{u}_0)\} \\
&= \sum_{v=0}^{V} a_v(\mathbf{u}_0)\, f_v(\mathbf{u}_0) - \sum_{v=0}^{V} a_v(\mathbf{u}_0)\, f_v(\mathbf{u}_0) - \left(\sum_{s=1}^{n} \lambda_{\mathrm{UK}}(\mathbf{u}_s) \left[Z(\mathbf{u}_s) - \sum_{v=0}^{V} a_v(\mathbf{u}_s)\, f_v(\mathbf{u}_s) \right] \right) \\
&= -\left(\sum_{s=1}^{n} \lambda_{\mathrm{UK}}(\mathbf{u}_s) \left[Z(\mathbf{u}_s) - \sum_{v=0}^{V} a_v(\mathbf{u}_s)\, f_v(\mathbf{u}_s) \right] \right) \\
&= \sum_{s=1}^{n} \lambda_{\mathrm{UK}}(\mathbf{u}_s) \sum_{v=0}^{V} a_v(\mathbf{u}_s)\, f_v(\mathbf{u}_s) - \sum_{s=1}^{n} \lambda_{\mathrm{UK}}(\mathbf{u}_s) \sum_{v=0}^{V} a_v(\mathbf{u}_s)\, f_v(\mathbf{u}_s) \\
&= \sum_{v=0}^{V} a_v(\mathbf{u}_0) \left[f_v(\mathbf{u}_0) - f_v(\mathbf{u}_0) \right] \\
&= 0
\end{aligned}
\tag{14}
$$

UK can be referred to as constrained kriging since there are $V + 1$ additional constraint equations in (13) that need to be imposed on the system in order to achieve unbiasedness in (14). There are, however, many other possible constraints that can be imposed in order to achieve unbiasedness. Notice the universal kriging estimator in (12) is significantly reduced to a linear combination of the n universal kriging weights $\lambda_{\mathrm{UK}}(\mathbf{u}_s)$ and RVs $Z(\mathbf{u}_s)$, s = 1,…, n, due to the unbiasedness constraints in (13).

There remains to determine the universal kriging weights $\lambda_{\mathrm{UK}}(\mathbf{u}_s)$. These n weights are determined so that the error variance in relation (9) is a minimum. However, the minimization must be performed subject to the $V + 1$ constraint equations in (13). These constraints call for the definition of a Lagrangian function $G(\mathbf{u}_0)$ that depends on the n universal kriging weights $\lambda_{\mathrm{UK}}(\mathbf{u}_s)$ in addition to the Lagrange parameters $2\mu_v(\mathbf{u}_0)$:

$$
\mathrm{G}(\mathbf{u}_0) = \mathrm{VAR}\{E(\mathbf{u}_0)\} + 2\mu_v(\mathbf{u}_0) \left[\sum_{s=1}^{n} \lambda_{\mathrm{UK}}(\mathbf{u}_s)\, f_v(\mathbf{u}_s) - f_v(\mathbf{u}_0) \right] \quad v = 0,...,V \tag{15}
$$

The optimal weights $\lambda_{\mathrm{UK}}(\mathbf{u}_s)$ are obtained by setting the n partial derivatives of (15) with respect to $\lambda_{\mathrm{UK}}(\mathbf{u}_s)$ to zero as well as the $(V + 1)$ partial derivatives of (15) with respect to $\mu_v(\mathbf{u}_0)$ to zero:

$$
\begin{aligned}
\frac{\partial\left(\mathrm{VAR}\{E(\mathbf{u}_0)\}\right)}{2\partial\left(\lambda_{\mathrm{UK}}(\mathbf{u}_s)\right)} &= \mathrm{COV}\{R(\mathbf{u}_s)R(\mathbf{u}_0)\} - \sum_{s'=1}^{n} \lambda_{\mathrm{UK}}(\mathbf{u}_{s'})\mathrm{COV}\{R(\mathbf{u}_{s'})R(\mathbf{u}_s)\} - \sum_{v=0}^{V} \mu_v(\mathbf{u}_0) f_v(\mathbf{u}_s) = 0 \\
\frac{\partial\left(\mathrm{G}(\mathbf{u}_0)\right)}{2\partial\left(\mu_v(\mathbf{u}_0)\right)} &= \sum_{s'=1}^{n} \lambda_{\mathrm{UK}}(\mathbf{u}_{s'})\, f_v(\mathbf{u}_{s'}) - f_v(\mathbf{u}_0) = 0
\end{aligned}
\tag{16}
$$

This results in the following system of universal kriging equations:

$$
\begin{aligned}
&\sum_{s'=1}^{n} \lambda_{\mathrm{UK}}(\mathbf{u}_{s'})\mathrm{COV}\{R(\mathbf{u}_{s'})R(\mathbf{u}_s)\} + \sum_{v=0}^{V} \mu_v(\mathbf{u}_0) f_v(\mathbf{u}_s) = \mathrm{COV}\{R(\mathbf{u}_s)R(\mathbf{u}_0)\} \\
&\sum_{s'=1}^{n} \lambda_{\mathrm{UK}}(\mathbf{u}_{s'})\, f_v(\mathbf{u}_i) = f_v(\mathbf{u}_0)
\end{aligned}
\tag{17}
$$

There are $(n + V + 1)$ equations with n universal kriging weights $\lambda_{\mathrm{UK}}(\mathbf{u}_s)$ and $(V + 1)$ Lagrange parameters $\mu_v(\mathbf{u}_0)$ to be determined. From (9) and (17), the minimized universal kriging variance $\sigma^2_{\mathrm{UK}}(\mathbf{u}_0)$ is then:

$$\sigma^2_{\mathrm{UK}}(\mathbf{u}_0) = C_R(\mathbf{0}) - \sum_{\mathrm{s}=1}^{n} \lambda_{\mathrm{UK}}(\mathbf{u}_\mathrm{s}) C_R(\mathbf{u}_0 - \mathbf{u}_\mathrm{s}) - \sum_{v=0}^{V} \mu_v(\mathbf{u}_0) \tag{18}$$

Ordinary kriging is a particular type of universal kriging whereby convention $f_0(\mathbf{u}) = 1$ and $m(\mathbf{u}) = a_0(\mathbf{u})$. This corresponds to the case where $m(\mathbf{u})$ is re-estimated to a constant $a_0(\mathbf{u})$ value within local often overlapping search windows.

Implementation Issues

Four implementation issues are identified and described in this work:

1. It is not the spatial law of the original $Z(\mathbf{u})$ RF that is required to interpolate $z(\mathbf{u}_\mathrm{s})$ data using the unconstrained UK/KT equations. It is actually the spatial law of $R(\mathbf{u})$ that is required. The usual assumption of stationarity for $Z(\mathbf{u})$ must then be transferred to the residual RF $R(\mathbf{u})$. Inference of this spatial law is difficult since $R(\mathbf{u})$ is not sampled in reality and its realizations are only a product of the artificial construct in (1). The emphasis of this paper is this first issue. The logic and reasoning of this group of challenges are developed and explained with examples.
2. Constrained UK/KT is not theoretically correct for implementation of simulation.
3. There are several situations in which the estimation setting creates a singular matrix within the universal kriging system of equations. Two typical circumstances for this situation are described with examples.
4. The set of n conditioning data retained for the least squares estimation of the mean $m(\mathbf{u})$ is not typically the same as for ordinary and simple kriging in practice. The reason for this is explained with an example.

Stationarity and the Spatial Law

Just like a single RV $Z(\mathbf{u})$ at just one $\mathbf{u}_0$ location is fully characterized by its ccdf model $F(\mathbf{u}; z|(n))$, the entire RF $Z(\mathbf{u})$ is fully characterized by the set of all L-variate ccdfs:

$$F\left(\mathbf{u}_1,...,\mathbf{u}_L; z_1\middle|(n),...,z_L\middle|(n)\right) = \mathrm{Prob}\left\{Z(\mathbf{u}_1) \le z_1\middle|(n),...,Z(\mathbf{u}_L) \le z_L\middle|(n)\right\} \tag{19}$$

for any number of RV locations L and any choice of the L locations $\mathbf{u}_l$, $l = 1,\ldots, L$. The multivariate ccdf in (19) describes the joint uncertainty about the L unknown $z(\mathbf{u}_1),\ldots, z(\mathbf{u}_L)$ values and is referred to as the spatial law of the RF $Z(\mathbf{u})$. For most practical geostatistical applications, inference of the spatial law is limited to the first two orders of the $Z(\mathbf{u})$ RF corresponding to $L = 1$ and $L = 2$ in (19). When $L = 2$, the two locations are typically denoted $\mathbf{u}$ and $\mathbf{u}' = \mathbf{u} + \mathbf{h}$ where $\mathbf{h}$ is a separation or lag vector. The univariate ccdfs $F(\mathbf{u}; z|(n))$ and first moment expected values are written:

$$\begin{aligned} F\left(\mathbf{u}; z\middle|(n)\right) &= \mathrm{Prob}\left\{Z(\mathbf{u}) \le z\middle|(n)\right\} \\ \mathrm{E}\left\{Z(\mathbf{u})\right\} &= m(\mathbf{u}) \end{aligned} \tag{20}$$

And all bivariate ccdfs $F(\mathbf{u},\mathbf{u}'; z|(n), z'|(n'))$ and second order covariance, correlation, and variogram measures are written as:

$$
\begin{aligned}
F\left(\mathbf{u},\mathbf{u}';z\left|(n),z'\right|(n')\right) &= \text{Prob}\left\{Z(\mathbf{u})\le z\left|(n),Z(\mathbf{u}')\le z'\right|(n')\right\} \\
C(\mathbf{u},\mathbf{u}') &= \text{E}\left\{Z(\mathbf{u})\cdot Z(\mathbf{u}')\right\}-\text{E}\left\{Z(\mathbf{u})\right\}\cdot\text{E}\left\{Z(\mathbf{u}')\right\} \\
\rho(\mathbf{u},\mathbf{u}') &= C(\mathbf{u},\mathbf{u}')\left[\left(C(\mathbf{u},\mathbf{u})\right)\left(C(\mathbf{u}',\mathbf{u}')\right)\right]^{-0.5} \\
2\gamma(\mathbf{u},\mathbf{u}') &= \text{E}\left\{\left(Z(\mathbf{u})-Z(\mathbf{u}')\right)^2\right\}
\end{aligned}
\tag{21}
$$

The problem of evaluating the spatial law of $Z(\mathbf{u})$ reduces to that of inferring the ccdfs and moments in (20) and (21). Geostatisticians typically use the experimental variogram $\gamma_Z(\mathbf{h})$ instead of the covariance $C_Z(\mathbf{h})$ to quantify spatial correlation in practice. The variogram is considered more robust in the presence of trends and departures from stationarity since it filters away any non-stationary or locally varying mean. The covariance on the other hand is sensitive to the mean.

Stationarity

The spatial law of $Z(\mathbf{u})$ is estimated with the $z(\mathbf{u}_s)$ data vector. A decision of stationarity must be made in order to substitute the need for repetitive realizations at all $\mathbf{u}_l$ locations for scattered (single realizations) sampling at $\mathbf{u}_s$ locations. A strong assumption of stationarity entails invariance of the full L-variate joint distribution function under any translation $\mathbf{h}$ within the domain $\mathbf{D}$:

$$
F\left(\mathbf{u}_1,...,\mathbf{u}_L;z(\mathbf{u}_1)\,|\,(n),...,z(\mathbf{u}_L)\,|\,(n)\right)=F\left(\mathbf{u}_1+\mathbf{h},...,\mathbf{u}_L+\mathbf{h};z(\mathbf{u}_1+\mathbf{h})\,|\,(n),...,z(\mathbf{u}_L+\mathbf{h})\,|\,(n)\right) \tag{22}
$$

Geostatistical techniques are normally applied with a less stringent second-order assumption of stationarity. This entails relation (22) with $L = 2$. All single-variate ccdfs $F(\mathbf{u}_l; z(\mathbf{u}_l)|(n))$ are then equivalent to the ccdf formed by all available $z(\mathbf{u}_s)$ sample values within $\mathbf{D}$. And all bivariate ccdfs $F(\mathbf{u}_l, \mathbf{u}_l + \mathbf{h}; z(\mathbf{u}_l)|(n), z(\mathbf{u}_l + \mathbf{h})|(n))$ are equivalent to the joint distribution of all possible pairs of sample data approximately separated by $\mathbf{h}$. This assumption or decision of second-order stationarity then implies the following first-order mean and second-order covariance properties:

1. The mean is independent of location,

$$
\text{E}\left\{Z(\mathbf{u}_l)\right\}=\text{m}\qquad\forall\mathbf{u}_l\in\mathbf{D} \tag{23}
$$

2. The covariance is independent of location depending only on the lag vector $\mathbf{h}$,

$$
C_Z(\mathbf{h})=\text{E}\left\{Z(\mathbf{u}_l+\mathbf{h})\cdot Z(\mathbf{u}_l)\right\}-\text{m}^2\qquad\forall\mathbf{u}_l,\mathbf{u}_l+\mathbf{h}\in\mathbf{D} \tag{24}
$$

The assumption of stationarity allows inference. In particular, the spatial law of $Z(\mathbf{u})$ can be assessed by evaluating $C\{Z(\mathbf{u}_l), Z(\mathbf{u}_l) + \mathbf{h}\}$ with $C_Z(\mathbf{h})$ which is simply the experimental covariance of all pairs of sample data approximately separated by $\mathbf{h}$ ($z(\mathbf{u}_s)$, $z(\mathbf{u}_s + \mathbf{h})$).

Decisions of stationarity are a necessary consequence of the random function approach. These decisions amount to assume the geology is homogeneous within certain spatial domains, (23) and (24). Decisions of stationarity are necessarily subjective and cannot be tested a-priori, but they can be argued inappropriate a-posteriori.

The Spatial Law for Non-Stationary Kriging

Making an assumption of first order stationarity (SK), the spatial law of $R(\mathbf{u})$ and $Z(\mathbf{u})$ are the same; however, this is not the case when a locally varying mean such as in (10) is used in UK/KT. Interestingly, it is actually the spatial law of the unknown and unsampled $R(\mathbf{u})$ RF that is required and not the spatial law of $Z(\mathbf{u})$. The residuals are also assumed second-order stationary, replace Z with R in (19) to (24). The spatial law of $R(\mathbf{u})$ is calculated with all residual data approximately separated by $\mathbf{h}$, $(r(\mathbf{u}_s), r(\mathbf{u}_s + \mathbf{h}))$.

Significant problems arise in calculating either the variogram $\gamma_Z(\mathbf{h})$ or covariance $C_Z(\mathbf{h})$. These can all be rooted to the fact that the residual data, $r(\mathbf{u}_s) = z(\mathbf{u}_s) - m(\mathbf{u}_s)$, are not actually *data* since they exist only as an artifact of the construct assumed in (1), which was necessary to perform non-stationary kriging. This key section of the paper is devoted to investigating this notion further and revealing specific reasons why the spatial law of residuals is notoriously difficult to capture in practice.

To start, the residual covariance and variogram decomposition following (1) is presented. Next an analytical calculation of the decomposed second order moments for an entirely parameterized 1D example is performed. This exercise shows the residual and original variable variogram are roughly equal up to reasonable separation vectors $\mathbf{h}$. A simple modification of the example then shows that this is seldom the case in practice. And finally, two reasons for the separation between theory and practice are then presented.

Residual Covariance and Variogram Decomposition

Decomposition of the $Z(\mathbf{u})$ RF according to (1) leads to the following covariance relation:

$$\begin{aligned} C\{Z(\mathbf{u}), Z(\mathbf{u}+\mathbf{h})\} &= C\{m(\mathbf{u}), m(\mathbf{u}+\mathbf{h})\} + C\{R(\mathbf{u}), R(\mathbf{u}+\mathbf{h})\} + C\{m(\mathbf{u}), R(\mathbf{u}+\mathbf{h})\} + C\{m(\mathbf{u}+\mathbf{h}), R(\mathbf{u})\} \\ &= C_m(\mathbf{h}) + C_R(\mathbf{h}) + 2C_{mR}(\mathbf{h}) \end{aligned} \quad (25)$$

Similarly, the variogram is decomposed as:

$$\begin{aligned} \gamma\{Z(\mathbf{u}), Z(\mathbf{u}+\mathbf{h})\} &= \gamma\{m(\mathbf{u}), m(\mathbf{u}+\mathbf{h})\} + \gamma\{R(\mathbf{u}), R(\mathbf{u}+\mathbf{h})\} + \gamma\{m(\mathbf{u}), R(\mathbf{u}+\mathbf{h})\} + \gamma\{m(\mathbf{u}+\mathbf{h}), R(\mathbf{u})\} \\ &= \gamma_m(\mathbf{h}) + \gamma_R(\mathbf{h}) + 2\gamma_{mR}(\mathbf{h}) \end{aligned} \quad (26)$$

The amount of variability modeled by the trend $m(\mathbf{u})$ is a subjective balance between determinism and stochasticity. In all cases, the trend should model no more variability than what our deterministic understanding of the geological processes suggests.

Analytically Determined Residual Spatial Law

Consider the porosity profile shown in the Figure 1. There is a clear fining downward trend in the porosity variable. For this part of the example, the mean function is fully defined. The locally varying mean/trend function $m(\mathbf{u})$ is represented by red in the figure. Its analytical form exhibits a linear increase from 10% to 25% porosity over the 20m elevation interval, that is: $m(\mathbf{u}) = 10 + 15(Z/20)$. In order to generate the porosity variable RF $Z(\mathbf{u})$, the residual RF $R(\mathbf{u})$ is needed, see (1). A 20m long string of residuals is generated according to a 10cm elevation increment and a spherical variogram with 5m range. The distribution of $r(\mathbf{u})$ values is normal with mean and variance of zero and 0.25. *SGSIM* from GSLIB [12] was used to impart spatial correlation to the residuals. The $Z(\mathbf{u})$ random function (jagged blue line) is then obtained by adding $m(\mathbf{u})$ and $r(\mathbf{u})$. The 5m residual range and variance are apparent.

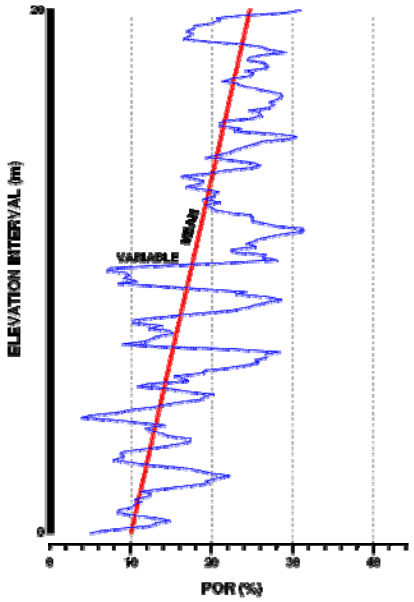

Figure 1: A simple 1D porosity profile with fully parameterized linear mean function (red). The example shows trend modeling practice is often dangerous.

The decomposed variograms in (26) are calculated and plotted in Figure 2 for the porosity profile shown in Figure 1. The most important observation is that the original variable and residual variable variograms are the same for reasonably small lag vectors. The point where their separation begins to be significant occurs at $\mathbf{h} \sim 3$m and continues for larger $\mathbf{h}$. The reason for the separation, the point where it begins, and its severity all depend on the mean variogram model.

Consider another example where the mean model is a constant arithmetic average, $m(\mathbf{u}) = 17.5\%$. This would be the case for an SK implementation. Here, the variogram model would be zero and the residual and original variable variograms would coincide over all $\mathbf{h}$. As we consider non-stationarities and model more variability in the mean and later variogram, the separation between residual and original variable variograms increases in magnitude at smaller $\mathbf{h}$. We investigate this relationship further with a numerical example in the next section.

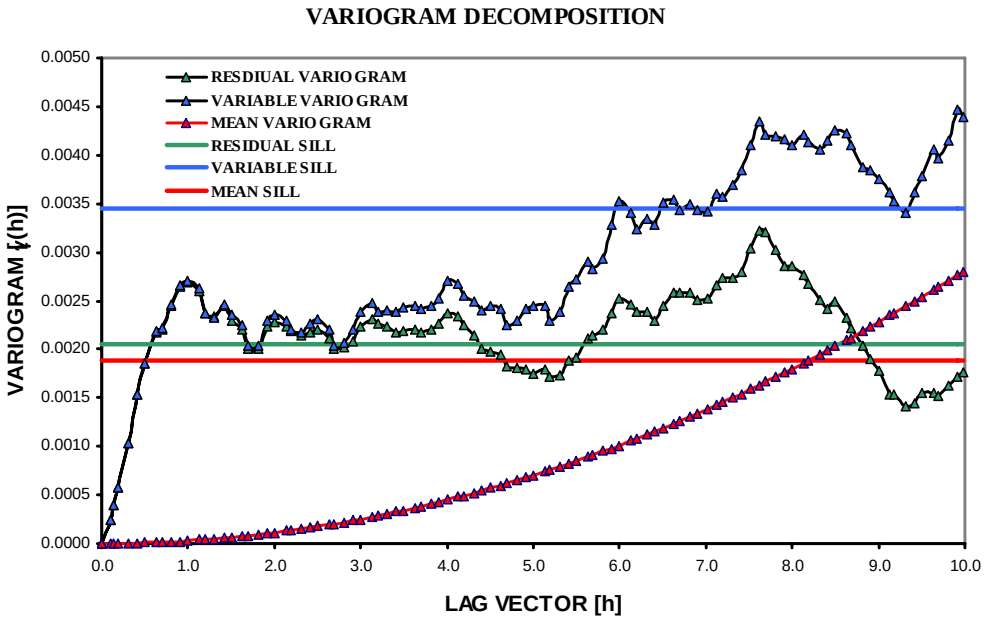

Figure 2: The original variable (blue), residual variable (green), and mean variable (red) variograms for the porosity profile previously shown in Figure 1.

Experimentally Determined Residual Spatial Law

The variograms for the original variable and residual are the same for reasonably small lag vectors. As more variability is modeled by $m(\mathbf{u})$ and $\gamma_m(\mathbf{h})$, the separation between $\gamma_Z(\mathbf{h})$ and $\gamma_R(\mathbf{h})$ is more severe and occurs at shorter $\mathbf{h}$. This is part of the problem in practice. Too much variability is modeled by $m(\mathbf{u})$ and $\gamma_m(\mathbf{h})$. Most techniques for trend modeling rely heavily on the available $z(\mathbf{u})$ data and not enough on deterministic geological knowledge for the trend model. In the most extreme yet not uncommon cases, the residual variogram $\gamma_R(\mathbf{h})$ even appears as almost a pure nugget effect. Clearly this is incorrect for predicting within the UK system of equations that requires the spatial law of residuals $R(\mathbf{u})$. Modeling the trend must be revisited.

Consider again the setting in Figure 1. This time we assume the $z(\mathbf{u})$ data (jagged blue line) is a given profile from a logging tool or core analysis. A typical trend modeling technique is applied to obtain a locally varying mean model. The method chosen here is a block ordinary kriging estimating along the 20m string at 10cm elevation increments with 10 discretization locations for each of the 200 estimation sites. A larger 30% nugget effect, large search, and longer 10m range are used for the model variogram to ensure the trend model is smooth. *KT3D* from GSLIB [12] is used to perform the kriging. Other smooth trend modeling techniques are commonly implemented in practical cases; similar methods create the same problems. The resulting trend model is represented by the red line in Figure 3. The $R(\mathbf{u})$ random function is then obtained by simply subtracting the locally varying mean $m(\mathbf{u})$ from the original variable $z(\mathbf{u})$.

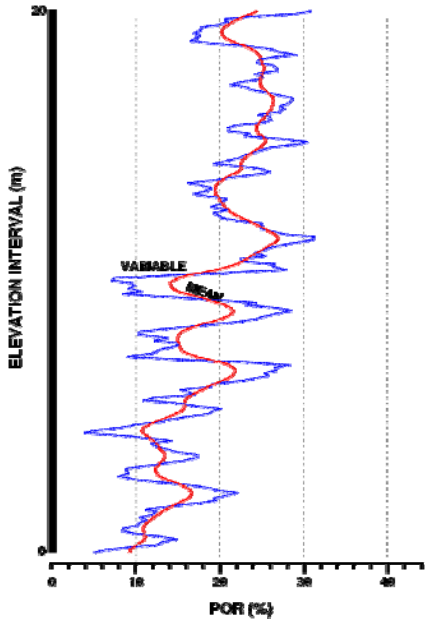

Figure 3: The porosity profile from Figure 1 with an experimentally determined trend model created with block ordinary kriging.

The decomposed variograms in (26) are calculated and plotted in Figure 4 for the porosity profile shown in Figure 3. Notice first the additional variability in the mean variogram. This can be seen by comparing the mean profiles in Figures 1 and 3 or the mean variograms in Figures 2 and 4. This additional variability does not change the variability in the original variable; however, it does affect the variability of the residual. That is, the residual variable has lost spatial correlation to the mean variable. The residual variogram for instance has a range of only about 3 **h** increments or 30cm and departs significantly from the original variable variogram immediately at about the same distance.

The porosity variable in this 1D example is structured. Certainly there is a fining downward trend in the porosity variable. Kriging with a trend (KT) or universal kriging (UK) can be used to automatically account for this trend. For this, the residual spatial law captured by the residual variogram is required. However, it would be incorrect to use the almost pure nugget effect residual variogram in Figure 4 caused by modeling too much variability in the mean. It is possible to use a combination of the residual and variable variogram for modeling the spatial law, for example, the shorter-scale residual variogram and longer-range variable variogram. Perhaps a better approach, however, is to model only the variability afforded through deterministic observations (Figure 1 and 2) rather than a combination of determinism and hard data conditioning (Figure 3 and 4). This subjective choice is certainly a key implementation challenge facing non-stationary kriging.

Reasons for Difference between Theory and Practice

The difference between the residual variogram from the original variable variogram is attributable to (1) an overfitting of the trend to the original data, and (2) significant correlation between the mean and residual. The first reason is made clear in the previously presented examples. The second reason is more subtle. Notice the covariance terms in (25) and (26). Of course if these

terms are significant, the original variable variability is difficult to match. Artifacts due to the residual-mean remnant correlation potentially exist [6] when these covariance terms are significant.

Simulation and Non-Stationary Kriging

Consider the following criteria for a map of any particular petrophysical property z predicting at unsampled $\mathbf{u}_0$ locations with $Z^*(\mathbf{u}_0)$ from a set of $z(\mathbf{u}_s)$ samples:

1. The data-to-data covariance $C(\mathbf{u}_s - \mathbf{u}_{s'})$ involving the $z(\mathbf{u}_s)$ samples is reproduced;
2. The estimate-to-estimate covariance $C(\mathbf{u}_0 - \mathbf{u}_0)$ or equivalently the variance of the estimates VAR$\{Z^*(\mathbf{u}_0)\}$ reproduces the stationary variance σ^2;

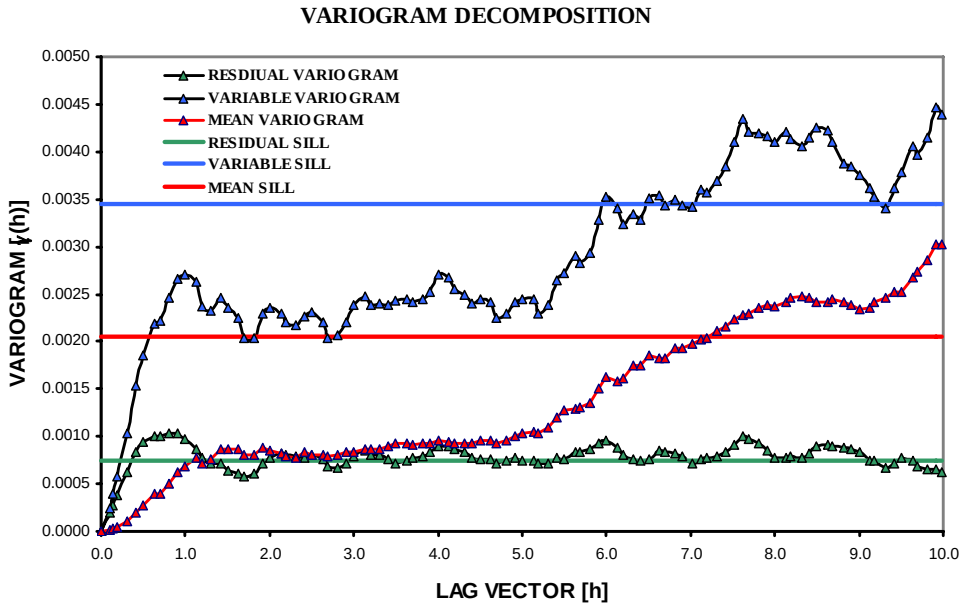

Figure 4: The original variable (blue), residual variable (green), and mean variable (red) variograms for the porosity profile previously shown in Figure 3.

3. The estimate-to-data covariance $C(\mathbf{u}_0 - \mathbf{u}_s)$ involving the $z^*(\mathbf{u}_0)$ values and $z(\mathbf{u}_s)$ samples is reproduced before and after estimation.

The motivation for simulation can be derived by evaluating these criteria for SK. The first of these three conditions is easily met. The $z(\mathbf{u}_s)$ sample data do not change from before to after estimation; therefore, the data-to-data covariance $C(\mathbf{u}_s - \mathbf{u}_{s'})$ remains the same. The second condition involves the new $z^*_{SK}(\mathbf{u}_0)$ estimates and should be checked:

$$
\begin{aligned}
\mathrm{VAR}\left\{Z^*(\mathbf{u})\right\} &= \mathrm{E}\left\{R^*(\mathbf{u}_0)R^*(\mathbf{u}_0)\right\} \\
&= \sum_{s=1}^{n}\sum_{s'=1}^{n}\lambda_{\mathrm{SK}}(\mathbf{u}_s)\lambda_{\mathrm{SK}}(\mathbf{u}_{s'})C_R\left(\mathbf{u}_s-\mathbf{u}_{s'}\right) \\
&= \sigma^2_{\mathrm{SK}}(\mathbf{u}_0) - C_R\left(\mathbf{0}\right) + 2\sum_{s=1}^{n}\lambda_{\mathrm{SK}}(\mathbf{u}_s)C_R\left(\mathbf{u}_0-\mathbf{u}_s\right) \\
&= \sigma^2_{\mathrm{SK}}(\mathbf{u}_0) - C_R\left(\mathbf{0}\right) + 2\left(C_R\left(\mathbf{0}\right) - \sigma^2_{\mathrm{SK}}(\mathbf{u}_0)\right) \\
&= C_R\left(\mathbf{0}\right) - \sigma^2_{\mathrm{SK}}(\mathbf{u}_0) \\
&= \sigma^2 - \sigma^2_{\mathrm{SK}}(\mathbf{u}_0)
\end{aligned}
\tag{27}
$$

The variance of the estimates VAR{$Z^*(\mathbf{u}_0)$} is underestimated by an amount equal to the SK estimation variance. This is known as the smoothing effect of kriging, which is prevalent in all flavors of kriging. The degree of smoothing depends on location. In particular, far away from data, the covariance decreases, the kriging variance increases, and there is more smoothing. Close to data, the covariance increases, the kriging variance decreases, and there is less smoothing. Ideally, the variance of the kriging estimator $Z^*(\mathbf{u}_0)$ would be the global variance σ^2 in expected value. This motivates simulation.

Simulation acts to correct the variance via the addition of a random residual $Q(\mathbf{u})$:

$$
R_{\mathrm{SGS}}(\mathbf{u}_0) = R^*(\mathbf{u}_0) + Q(\mathbf{u}_0) \tag{28}
$$

The residual is drawn randomly from a Gaussian distribution with an expected value of zero and variance equal to the kriging variance $\sigma^2{}^*_{\mathrm{SK}}(\mathbf{u}_0)$. This does not change the optimal kriging estimate $z^*_{\mathrm{SK}}(\mathbf{u}_0)$, but does act to increase the variance by the amount missing in (27), that is, the variance of the newly simulated values $z_{\mathrm{SGS}}(\mathbf{u}_0)$ is the correct global variance σ^2. The third condition also involves new $z^*_{\mathrm{SK}}(\mathbf{u}_0)$ estimates and should also be checked:

$$
\begin{aligned}
C_R\left(\mathbf{u}_0-\mathbf{u}_s\right) &= \mathrm{E}\left\{R^*_{\mathrm{SK}}(\mathbf{u}_0)R(\mathbf{u}_s)\right\} - \mathrm{E}\left\{R^*_{\mathrm{SK}}(\mathbf{u}_0)\right\}\mathrm{E}\left\{R(\mathbf{u}_s)\right\} \\
&= \mathrm{E}\left\{\sum_{s'=1}^{n}\lambda_{\mathrm{SK}}(\mathbf{u}_{s'})R(\mathbf{u}_{s'})R(\mathbf{u}_s)\right\} \\
&= \sum_{s'=1}^{n}\lambda_{\mathrm{SK}}(\mathbf{u}_{s'})C_R\left(\mathbf{u}_{s'}-\mathbf{u}_s\right) \\
&= C_R\left(\mathbf{u}_0-\mathbf{u}_s\right)
\end{aligned}
\tag{29}
$$

Therefore, in probabilistic terms, the third condition is also satisfied for SK. Moreover, it can be shown that the addition of the random residual $Q(\mathbf{u})$ does not change the estimate-to-data covariance reproduction. Therefore, while the simulated values $z_{\mathrm{SGS}}(\mathbf{u}_0)$ satisfy all three criteria, the simple kriging values $z^*_{\mathrm{SK}}(\mathbf{u}_0)$ satisfies only the first two.

Note the use of the SK system of equations with no V constraints to make the last substitution. This third condition of estimate-to-data covariance reproduction is only met with unconstrained simple kriging. For UK/KT:

$$
C_R\left(\mathbf{u}_0-\mathbf{u}_s\right) \neq C_R\left(\mathbf{u}_0-\mathbf{u}_s\right) - \sum_{v=0}^{V}\mu_v(\mathbf{u}_0)f_v(\mathbf{u}_s) \tag{30}
$$

The constrained kriging system of equations does not allow the third condition to be satisfied because of the inclusion of Lagrange parameters in the systems of equations. This means that constrained kriging should not theoretically be used in simulation-based algorithms. The popular alternative of integrating an estimated trend model by performing geostatistical simulation of the residuals and adding the simulated model back to the trend model is a result of this limitation.

Singularity and Solution Existence

There are some constrained estimation settings that produce singular matrices within the subsequent UK/KT systems. The left-hand-side redundancy covariance and constraint matrix in (17) must be a non-singular matrix in order for its inverse to exist and to calculate the $\lambda_{UK}(\mathbf{u}_s)$ weights, $\mu_v(\mathbf{u}_0)$ Lagrange parameters, $z^*_{UK}(\mathbf{u}_0)$ universal kriging estimate, and $\sigma^2_{UK}(\mathbf{u}_0)$ universal kriging variance. With SK, the possibilities for singular redundancy matrices are small barring computer precision issues. However, the constraint equations involved in UK/KT (13) can potentially create singular matrices.

Two situations in which singular UK redundancy covariance-constraint matrices occur are addressed. Figure 5 shows a schematic 2D illustration of the setting for each situation. Although computer precision can also cause singular matrices, numerical instabilities are not discussed here in detail. This type of situation is too sensitive to computer resources to discuss in detail.

The first typical cause of a singular matrix is when the number of data n is greater than the number of constraint equations V. Figure 5 (left) shows an estimation location with $n = 2$ sample data and $V = 2$ constraint equations (3 in total). The constraint equations correspond to a linear mean in both the X and Y coordinates. The missing covariances $C(R(\mathbf{u}_1),R(\mathbf{u}_s)) = C(R(\mathbf{u}_s),R(\mathbf{u}_1))$, $s = 1, 2, 3$, create a rectangular redundancy matrix that theoretically has no inverse. In general, the number of sample data n and number of constraint equations V must follow $n \geq (V + 1)$ in order for a solution to exist.

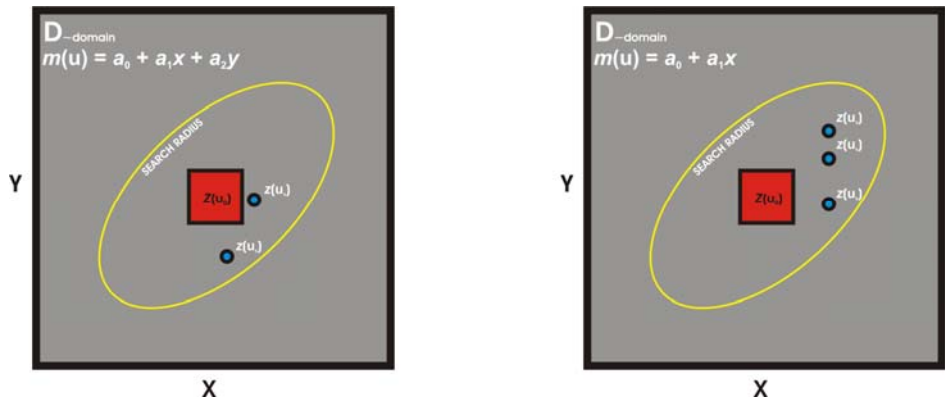

Figure 5: An illustration of two common causes of singular non-invertible redundancy covariance-constraint matrices within the constrained UK or KT systems of equations.

The other typical cause of a singular matrix is when the sample data is perfectly aligned in a straight line perpendicular to the direction in which the mean is a function. For example, consider fitting a linear mean model as a function of the X coordinate in Figure 5 (right) where $n = 3$ data are shown perfectly aligned in the Y direction. It is impossible to fit a mean that is a function of the X coordinate using the three data with equal X coordinate vector components. The resulting redundancy matrix would be singular and not permit an inverse to be calculated. In general, the

sample data must have various location vector components in the direction in which the mean is a function in order for a solution to exist.

Sample Data Conditioning

A well understood but poorly implemented principle in the practice of modeling non-stationarity is that the trend should model no more variability than what our deterministic understanding of the geological processes suggests [6]. Often the trend represents too much variability in practice. Indeed, modeling the trend with an automatic least squares optimum fit of a deterministically established polynomial form for the locally varying mean $m(\mathbf{u})$ is then illusive. Two major reasons why the KT system can potentially yield misleading $m(\mathbf{u})$ fits are given here.

The first reason arises from a subjective choice of the explicit polynomial form for the mean in (10). There are infrequently cases where the balance between the degree of homogeneity and available data is such that a polynomial form with $V > 0$ is necessary. That is, the data may intrinsically reproduce an anticipated linear geological trend in the X coordinate without explicitly specifying the $m(\mathbf{u})$ polynomial up to $V = 1$. Or smaller estimation domains can be considered so that the trend is less significant.

The second reason is the automatic least squares fitting of the $m(\mathbf{u})$ polynomial function. The $a_v(\mathbf{u})$ polynomial trend coefficients are fit using a least squares weighted linear combination of the n surrounding data. The BLUE criteria for the consequent $m^*(\mathbf{u})$ estimates in this case are dangerously objective in that past deciding the functional form of $m(\mathbf{u})$ there is no direct or deterministic control on the final mean $m^*(\mathbf{u})$. Inevitably $m^*(\mathbf{u})$ is more objective and variable than what our deterministic understanding of the geological processes suggest. Discontinuities in $m^*(\mathbf{u})$ are particularly frequent and problematic. These discontinuities typically arise from search windows retaining less data than the total available within the overall estimation domain. At estimation locations where new/different data is retained, $m^*(\mathbf{u})$ can show a discontinuity. Therefore, although it is not theoretically required, best practice is to use a global search so that n coincides with the number of available data within the domain.

Figure 6 shows a small 1D example. There are 5 data along a 100m 1D line in the X direction (green bullets). It is previously known that there is a linear geological trend in the variable. This knowledge is particularly important where there are few data from about $X = 40$m to $X = 70$m (between vertical red lines). UK is used to account for the linear trend automatically where a $V = 1$ polynomial trend function, $m(\mathbf{u}) = a_0(\mathbf{u}) + a_1(\mathbf{u})x$, is to be fit. The top two figures show the estimated trend (left) and variable (right) using a limited search retaining at least two ($V + 1$) and at most 3 sample data. The bottom two figures show the estimated trend and variable using a global search retaining all 5 data in each case. Note the discontinuity in the trend and variable estimated profiles where the trend is important; these profiles are not consistent with previous deterministic trend knowledge. Discontinuities such as these are often not realistic in practice. In contrast, the global search produces a trend and variable estimated profile consistent with prior deterministic knowledge. In general, a global search is recommended to fit the trend functionals.

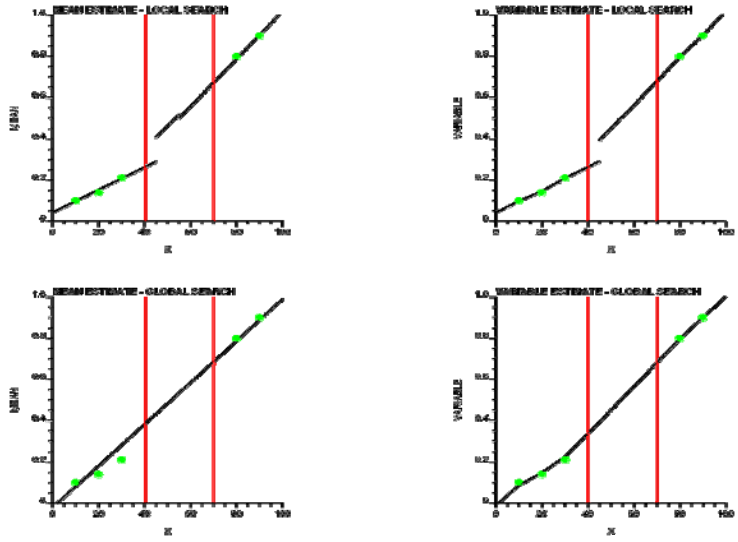

Figure 6: A simple 1D illustration of the importance of a global search in kriging with a trend (KT).

Conclusion

Kriging is a *best* estimator in the sense that the error variance between true values and estimates is a minimum. Although inappropriate for making production-based decisions, kriging is still remarkably flexible for detecting and quantifying large-scale geological trends. This work dealt with universal kriging (UK). Four implementation challenges associated with UK system of equations were addressed in this paper. These include: (1) The subjective choice for the mean model and its implication on the residual spatial law for the KT or UK systems of equations; (2) The use of non-stationary kriging in the implementation of simulation; (3) Situations when the setting creates a singular matrix and no solution exists; and (4) The group of conditioning data retained for the least squares estimation of the mean.

Due to these and other implementation details and especially the limitation for simulation applications, the authors still do not recommend the use of KT/UK for any other purpose other than to grasp large-scale geological trends.

References

[1] Isaaks, E.H. and Srivastava, R. M. *An Introduction to Applied Geostatistics*. Oxford University Press, New York, 1989.

[2] Journel, A.G. and Huijbregts, Ch.J. *Mining Geostatistics*. The Blackburn Press, New Jersey, 1978.

[3] Krige, D.G. *A Statistical Approach to some Mine Valuations and Allied Problems at the Witwatersrand*. Master's Thesis, University of Witwatersrand, South Africa, 1951.

[4] David, M. *Geostatistical Ore Reserve Estimation*, Elsevier, Amsterdam, 1977.

[5] McLennan, J.A. and Deutsch, C.V., *Conditional Bias of Geostatistical Simulation for Estimation of Recoverable Reserves*. CIM Bulletin, April 2004, vol. 97, no. 1080, pp. 68-72.

[6] McLennan, J.A., Leuangthong, O., and Deutsch, C.V., *Trend Modeling Techniques and Guidelines*. CCG Report 7, September 2005.

[7] Leuangthong, O. and Deutsch, C. V., *Transformation of Residuals to Avoid Artifacts in Geostatistical Modelling with a Trend*, Mathematical Geology, 36 (3), April 2004, pp. 287-305.

[8] Matheron, G. *The Theory of Regionalized Variables and Its Applications*. Ecole de Mines, Fontainbleau, France, 1971.

[9] Matheron, G. *Le Krigeage Universel: Fascicule 1*. Cahiers du CMM, 1969, 82 p.

[10] Journel, A.G. and Rossi, M.E. *When do we Need a Trend Model in Kriging?*, Mathematical Geology, vol. 21, no. 7, 1989.

[11] Journel, A.G. *Fundamentals of Geostatistics in Five Lessons*. American Geophysical Union 1989.

[12] Deutsch, C.V. and Journel, A.G. *GSLIB: Geostatistical Software and User's Guide (Second Edition)*, Oxford University Press, New York, 1998.

[13] Armstrong, M. *Basic Linear Geostatistics*, Springer-Verlag Berlin Heidelberg, New York, 1998.