Non-stationary Geostatistical Modeling Based on Distance Weighted Statistics and Distributions

David F. Machuca-Mory · Clayton V. Deutsch

Abstract A common assumption in geostatistics is that the underlying joint distribution of possible values of a geological attribute at different locations is stationary within a homogeneous domain. This joint distribution is commonly modeled as multi-Gaussian, with correlations defined by a stationary covariance function. This results in attribute maps that fail to reproduce local changes in the mean, in the variance and, particularly, in the spatial continuity. The proposed alternative is to build local distributions, variograms, and correlograms. These are inferred by weighting the samples depending on their distance to selected locations. The local distributions are locally transformed into Gaussian distributions embedding information on the local histogram. The distance weighted experimental variograms and correlograms are able to adapt to local changes in the direction and range of spatial continuity. The automatically fitted local variogram models and the local Gaussian transformation parameters are used in spatial estimation algorithms assuming local stationarity. The resulting maps are rich in nonstationary spatial features. The proposed process implies a higher computational effort than traditional stationary techniques, but if data availability allows for a reliable inference of the local distributions and statistics, a higher accuracy of estimates can be achieved.

D.F. Machuca-Mory · C.V. Deutsch
Centre for Computational Geostatistics, Department of Civil and Environmental Engineering, University of Alberta, Edmonton, Alberta, Canada

D.F. Machuca-Mory (✉)
Stochastic Mine Planning Laboratory, Department of Mining and Materials Engineering, McGill University, Frank Dawson Adams Building, Room 112 3450 University Street, Montreal, Quebec, Canada H3A 2A7
e-mail: david.machucamory@mail.mcgill.ca

1 Introduction

The assessment of the spatial uncertainty in geostatistics relies on the Random Function (RF) model (Matheron 1970). This is defined as an ensemble of multiple spatially correlated random variables (RV), $Z(\mathbf{u})$, located at multiple points $\mathbf{u}$ within a selected domain D. This domain is usually assumed homogeneous under statistical, geological, and/or practical criteria (Journel 1986). This corresponds to the decision of stationarity and allows for the inference of a spatially invariant distribution of grouped data within a domain, or of only some of its statistics (Myers 1989; Journel and Huijbregts 1978). The multi-Gaussian distribution is commonly adopted for characterizing the RF due to its convenient properties (Deutsch and Journel 1998). The mean field, and the variance and covariance matrix fully define the multi-Gaussian RF. Commonly used geostatistical techniques such as simple multi-Gaussian kriging (Verly 1983) and the various Gaussian-based simulation methods (Alabert 1987; Gómez-Hernández and Cassiraga 1993; Dimitrakopoulos and Luo 2004) rely on the assumption of a stationary multi-Gaussian RF. The basis of these techniques is the local conditioning of the global distribution of uncertainty by neighboring data. The resulting maps show the homogeneous spatial continuity resulting from the adoption of a global distribution and spatial covariance models.

Nevertheless, when modeling complex geological settings, it may be unfeasible to delimit homogeneous domains due to the lack of information about their limits or to the complexity of the geological processes affecting the geological attribute. In these cases, a realistic numerical modeling may require including locally changing means, variances, ranges, and orientations of the spatial continuity and other parameters.

Different approaches have been proposed to include the nonstationarity of particular statistics or parameters required for spatial estimation. Techniques like Universal Kriging incorporate a trend in the mean (Goovaerts 1997). Techniques such as variogram modeling within moving windows (Haas 1990), presimulation of the local anisotropy angles of the variogram model (Xu 1996) and spatial transformation (Sampson and Guttorp 1992; Boisvert et al. 2009; Monestiez and Switzer 1991) have been proposed for incorporating locally changing anisotropies in the spatial continuity. Although these approaches may deal satisfactorily with the nonstationarity of particular statistics and parameters, none of them offers an integrated treatment of all aspects of nonstationarity.

This paper proposes an integrated approach for incorporating local changes in the cumulative distribution function (cdf) and all the required statistics. This approach is based on the assumption of local stationarity. Under this assumption, the RF multivariate cdf is deemed stationary only in reference to an anchor point. The local cdf and its 1-point and 2-point statistics are inferred by weighting the samples depending on their distance to a respective anchor point. The modeling of local normal scores transformation functions by Hermite polynomials allows for translating the changes in the univariate cdf into a multi-Gaussian locally stationary framework. Additionally, the parameters of the local variogram models are intended to reflect the local changes in the spatial continuity. Locally stationary estimation and simulation techniques work with local transformation functions and local variograms at each unsampled location. This paper describes the details of the proposed approach. The performance of locally stationary multi-Gaussian kriging is compared with its stationary counterpart

in terms of the accuracy of estimates. The last part of this paper discusses several key aspects of the locally stationary approach and briefly describes possible future developments.

2 The Decision of Local Stationarity

An initial step in standard geostatistical modeling is the definition and delimitation of geologically and statistically homogeneous spatial domains (McLennan 2007). Within each of these domains, methods that account for the trend in the mean can be used effectively. Although a trend model may capture the variations of the local mean, this may not necessarily remove variations in other statistics, particularly in the spatial correlation. Moreover, when these variations are continuous, building sub-domains to capture them may be unfeasible. Modeling under the decision of local stationarity is proposed as an alternative to the standard modeling process.

Given a geologically homogeneous domain D, the proposed decision amounts to strict stationarity defined in relation to any of P reference points $\mathbf{o}_i, i = 1, \ldots, P$ within such domain. Thus, the RF multivariate distribution is invariant by translation when anchored to a point $\mathbf{o}_i$

$$
\begin{aligned}
&\text{Prob}\big\{ Z(\mathbf{u}_1) < z_1, \ldots, Z(\mathbf{u}_n) < z_n; \mathbf{o}_i \big\} \\
&= \text{Prob}\big\{ Z(\mathbf{u}_1 + \mathbf{h}) < z_1, \ldots, Z(\mathbf{u}_n + \mathbf{h}) < z_n; \mathbf{o}_j \big\}, \\
&\mathbf{u}_\alpha, \mathbf{u}_\alpha + \mathbf{h}, \mathbf{o}_i \text{ and } \mathbf{o}_j \in D, \text{ and only if } i = j, \alpha = 1, \ldots, n,
\end{aligned}
\tag{1}
$$

where $\mathbf{h}$ is a translation vector. The RF multivariate distribution is therefore redefined at each location and its statistics are pertinent only for the reference point where they are anchored. This is different to assuming stationarity within finite moving windows (Haas 1990) and inferring the corresponding RF using only the data that falls inside that window. This approach could result in difficult and unstable inference, particularly in areas where data is scarce. Instead, what is proposed is to build different RFs at each anchor point $\mathbf{o}_i \in D$ using all data within the domain, with closer samples to $\mathbf{o}_i$ having greater influence than far away ones in the inference of the local cdf and its statistics.

3 A Distance Weighting Approach for the Inference of Local Distributions and Statistics

The use of distance based weights for obtaining local statistics has been proposed in spatial statistics as the methodology of Geographically Weighted Regression (Fotheringham 1997; Fotheringham et al. 2002). This paper explores an extension of this idea within a geostatistical framework.

The weighting of samples for the inference of the locally stationary RF cdf can be achieved by any distance weighting function that fulfills a set of desirable properties: smooth distance decay, strict positivity, unbiasedness, independence of units, and global consistency for all statistics. The continuous decrease of weights as the

distance to the anchor point increases represents the idea that closer samples should have a greater contribution to the corresponding anchored cdf and its statistics. The distance based weights assigned to each sample are proportional to their contribution to the local cdf. Therefore, they should be positive and the sum of contributions from all samples should sum to 1. The weights used for the inference of the univariate cdf and its statistics should be also used for the inference of 2-point statistics within the same domain. This allows the 2-point statistics to revert to 1-point statistics when the 2-point separation distance becomes zero. Additionally, the weights should be independent of the distance units used, but dependent only on the relative distances. One distance function that fulfills all these properties is the Gaussian kernel. Given a set of n samples at locations $\mathbf{u}_\alpha$, $\alpha = 1, \ldots, n$ within a domain D, their corresponding weights in relation to an anchor point $\mathbf{o}$ are obtained from

$$\omega_{GK}(\mathbf{u}_\alpha; \mathbf{o}) = \frac{\varepsilon + \exp\left(-\frac{(d(\mathbf{u}_\alpha; \mathbf{o}))^2}{2s^2}\right)}{n\varepsilon + \sum_{\alpha=1}^{n} \exp\left(-\frac{(d(\mathbf{u}_\alpha; \mathbf{o}))^2}{2s^2}\right)}, \quad \mathbf{u}_\alpha, \mathbf{o} \in D \text{ and } \alpha = 1, \ldots, n. \quad (2)$$

This is akin to a Nadraya–Watson estimator (Wasserman 2006). The background constant ε is included here in order to avoid computational problems when the distance $d(\mathbf{u}_\alpha; \mathbf{o})$ is very large and also to permit some contribution of distant samples. The parameter s, known as the kernel variance or bandwidth, can be assumed the same for all directions if there is no evidence of a strong global anisotropy.

In geological datasets, sample locations commonly appear in clusters (Davis 2002; Borradaile 2003). In order to minimize the bias induced by the clustered sampling, the original distance weights can be corrected by declustering weights. Suitable spatial declustering techniques include cell declustering (Isaaks and Srivastava 1989; Deutsch and Journel 1998), polygonal declustering (Isaaks and Srivastava 1989; Deutsch 2002), and weights obtained from the cross-validation simple kriging variance (Bourgault 1997). If the distance weights are able to minimize the biases caused by clustered sampling, the average of weights assigned to each one of the samples in relation to P anchor points is equivalent to its corresponding declustering weight, w_α, that is,

$$\bar{\omega}(\mathbf{u}_\alpha) = \frac{1}{P} \sum_{j=1}^{P} \omega(\mathbf{u}_\alpha; \mathbf{o}_j) \approx w_\alpha \quad \forall \alpha = 1, \ldots, n. \quad (3)$$

This equality can be enforced if the original distance weights are scaled by

$$\hat{\omega}(\mathbf{u}_\alpha; \mathbf{o}) = \omega(\mathbf{u}_\alpha; \mathbf{o}) \frac{w_\alpha}{\bar{\omega}(\mathbf{u}_\alpha)} \quad \forall \alpha = 1, \ldots, n. \quad (4)$$

Whereas polygonal and cell declustering methods only take into account sample geometry, weights obtained from the cross-validation simple kriging variance consider also the redundancy and the global anisotropy of data. Thus, the latter represents a more complete approach for scaling the distance weights for local statistics.

2-point statistics, such as variograms, covariances, and correlograms measure the variability between pairs of samples separated by a given distance. Pairs of distance weights assigned to single samples can be combined in order to obtain the weights required for locally weighted 2-point statistics. A mixture rule (Korvin 1982) can

Fig. 1 2-point weight profiles for different values of the parameter t in the mixture rule. The tail sample location is fixed in the origin and has a weight of 1, while the head sample location is moving along the line
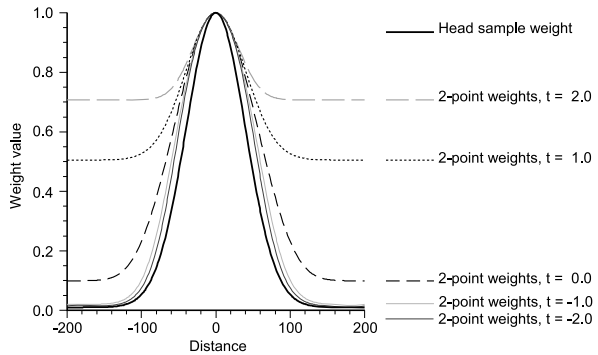
be applied for this purpose. Given two property values g_1 and g_2, and volume fractions ϕ, and $1 - \phi$, respectively, the general mixture rule is expressed as

$$M(g_1, g_2, \phi, t) = \left[\phi g_1^t + (1 - \phi)g_2^t\right]^{1/t}, \quad \text{with } t \neq 0, \ t \in R. \tag{5}$$

For obtaining 2-point weights for a pair separated by a vector $\mathbf{h}$, it can be assumed that the volume fractions of each sample are equivalent (that is, $\phi = 0.5$) and the individual sample weights are the property values

$$\omega(\mathbf{u}_\alpha, \mathbf{u}_\alpha + \mathbf{h}; \mathbf{o}) = \left[\frac{\omega^t(\mathbf{u}_\alpha; \mathbf{o}) + \omega^t(\mathbf{u}_\alpha + \mathbf{h}; \mathbf{o})}{2}\right]^{1/t}. \tag{6}$$

For $t = 1$, the mixture rule is the arithmetic average, if $t = -1$ it is the harmonic average. As seen in Fig. 1, the higher the t value, the closer the 2-point weight value approaches the highest 1-point weight. The inverse is observed when t is increasingly negative. A value of zero for t may be preferred in order to avoid these extremes. The limit of the mixture rule when t tends to zero is the geometric average, that is,

$$\lim_{t \to 0} \omega(\mathbf{u}_\alpha, \mathbf{u}_\alpha + \mathbf{h}; \mathbf{o}) = \sqrt{\omega(\mathbf{u}_\alpha; \mathbf{o}) \cdot \omega(\mathbf{u}_\alpha + \mathbf{h}; \mathbf{o})}. \tag{7}$$

A critical aspect of the local statistics inference is the choice of the distance weighting function parameters. Although this choice depends on data density and can be supported by numerical measures, it is mostly left to the practitioner's judgment. When choosing these parameters, the practitioner must be aware of the related variance/bias trade-off. A narrow bandwidth and very low background constant may capture smaller nonstationary features, but they may render the local statistics unreliable and cause overfitting. On the other hand, a very wide bandwidth could result in excessive smoothing of the local statistics that may mask nonstationary features. The distance weighting function must capture local trends informed by groups of samples rather than reflect the local influence of a few individual values. Therefore, if data is sparse, using a wider bandwidth would be preferred.

These weights modify the usual expressions of the 1-point and 2-point statistics. Thus, the local univariate cdf is obtained by

$$\text{Prob}\{Z(\mathbf{u}) \leq z_k \mid \mathbf{o}\} \approx \hat{F}(z_k; \mathbf{o}) = \sum_{\alpha=1}^{n} \omega(\mathbf{u}_\alpha; \mathbf{o}) \cdot I(\mathbf{u}_\alpha; z_k) \in [0, 1],$$
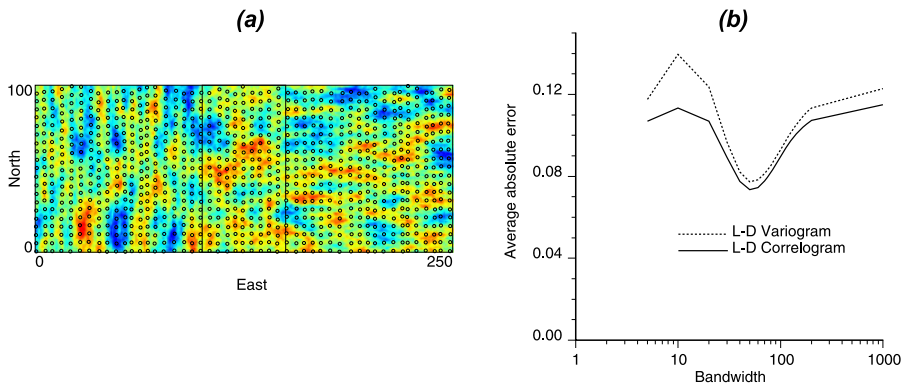$$\forall \mathbf{u}_\alpha \in D, \ k = 1, \ldots, K, \tag{8}$$

Fig. 2 (**a**) Synthetic image composed by three zones of different anisotropy (the *small circles* indicate sample locations). (**b**) Absolute errors for location-dependent 2-point statistics obtained using different kernel bandwidths

where $I(\mathbf{u}_\alpha; z_k)$ is one if $z(\mathbf{u}_\alpha) \leq z_k$, otherwise it is zero. The local mean, $\hat{m}_Z(\mathbf{o})$, variance, $\hat{\sigma}^2(\mathbf{o})$, and other univariate statistics are obtained using the same set of weights.

Omre (1984) proposed the estimation of variograms as weighted averages to account for data clustering. Here, the 2-point weights obtained using Eq. (7) are used to estimate local experimental variograms, $\hat{\gamma}(\mathbf{h}; \mathbf{o})$, covariances, $\hat{C}(\mathbf{h}; \mathbf{o})$, and correlograms, $\hat{\rho}(\mathbf{h}; \mathbf{o})$. The expressions for locally weighted 1-point and 2-point statistics are presented in Appendix A.

A synthetic image produced using sequential Gaussian simulation with different variogram model parameters at each one of its three regions allows for illustrating the impact of the kernel bandwidth on the local variograms and correlograms (Fig. 2(a)). This image was sampled by a semiregular grid of 5×5 pixels. The location-dependent experimental 2-point statistics were obtained at multiple anchor points located in a 10×10 pixels regular grid using different bandwidths. Figure 2(b) shows the average absolute errors between the experimental local 2-point statistics and the true ones calculated for each region of the exhaustive image. When narrow kernel bandwidths are used, the fluctuations in the local 2-point statistics increase (Fig. 3(a)). Instead, for very wide bandwidths, they converge toward the global experimental 2-point statistics (Fig. 3(c)). Provided that data is abundant enough, a bandwidth that is roughly the size of the regions of different anisotropy allows a reasonable reproduction of the true local spatial continuity (Fig. 3(b)). Narrow bandwidths and data scarcity result in increased uncertainty about the local distributions and statistics. Its quantification is beyond the scope of this paper, but the discussion section contains a more extended commentary on this subject.

Local experimental correlograms (Expression (A.6)) tend to be more robust than local experimental variograms (Expression (A.3)). This is due to the incorporation of the local means and variances in the calculation of the former. In this example this is translated in a smaller absolute error for the local correlograms. Moreover, their interpretation in terms of local spatial coefficients of correlation is straightforward.
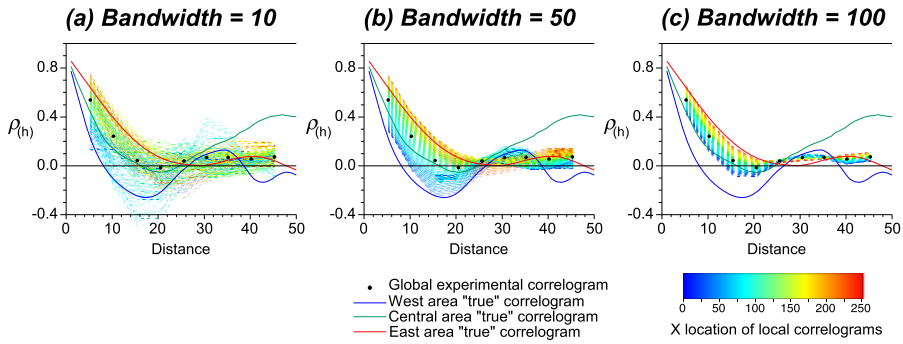
Fig. 3 Experimental global, true local and experimental local correlograms inferred using Gaussian kernel bandwidths of 10 (**a**), 50 (**b**), and 100 (**c**) distance units. All the correlograms are in the direction E–W

4 Modeling of Local Normal Scores Transformation and Local Variograms

Local normal scores transformation is required for adapting Gaussian-based geostatistical methods to the locally stationary framework. Each location dependent cdf $\hat{F}(z; \mathbf{o})$ must be locally transformed into the standard normal cdf, $G(y)$. The local normal scores transformation is expressed as

$$y = G^{-1}\big(\hat{F}(z; \mathbf{o})\big) = \varphi_Y(z; \mathbf{o}). \tag{9}$$

This allows for matching of the local quantiles $z_p(\mathbf{o})$ with the standard Gaussian quantiles, y_p, such as

$$\hat{F}\big(z_p(\mathbf{o}); \mathbf{o}\big) = G(y_p) = p \quad \forall p \in [0, 1]. \tag{10}$$

The local normal scores transformation functions must be stored for their further use in estimation and simulation. An efficient option is to use Hermite polynomial series with stored local coefficients to model the local transformation functions. Appendix B presents the equations required for this purpose.

The modeling of experimental variograms allows for an exhaustive description of the spatial continuity for different directions and distances. The variogram modeling process also allows for incorporating geological information on spatial continuity. These models provide a suitable covariance function for the kriging equations (Gringarten and Deutsch 2001). Any of the valid variogram models can be used for fitting the local experimental 2-point statistics. The same variogram model should be chosen for all anchor point locations. Only the parameters of this model change locally in response to changes in the spatial continuity informed by the experimental 2-point statistics. The stable model provides the capability of changing shapes according to its power parameter (Chilès and Delfiner 1999)

$$\gamma(\mathbf{h}; \mathbf{o}) = c(\mathbf{o}).\big[1 - \exp\big(-\big(3\big|\mathbf{h}'\big|\big)^{b(\mathbf{o})}\big)\big] \quad 0 < b(\mathbf{o}) \le 2. \tag{11}$$

In this expression, $c(\mathbf{o})$ is the local sill contribution, $\mathbf{h}'$, is the vector separation modified by the local orientations and ranges of anisotropy, and $b(\mathbf{o})$ controls the shape of the local variogram model. $b(\mathbf{o})$ yields the exponential model when it is 1 and the Gaussian model when it reaches 2.

The local variogram models must be fitted for several directions at multiple anchor point locations. This is only feasible with the help of a semiautomatic fitting algorithm. Popular semiautomatic fitting algorithms are based in the minimization of the weighted square error between the model and the experimental variogram (Cressie 1985; Pardo-Igúzquiza 1999; Emery 2010). Moreover, abrupt fluctuations in the parameters of the local variogram models must be also avoided if the experimental local 2-point statistics change smoothly between anchor points.

Strictly, the processes of location-dependent variogram inference and modeling, and local normal scores transformation and modeling should be performed at all the locations to be estimated or simulated; however, these processes can be very demanding in time and computer resources. Therefore, these are performed only at a limited number of anchor points. The locations of these anchor points are chosen in order to provide exhaustive and smoothly changing local statistics and parameters by interpolating their inferred values between those points. The local variogram parameters and local Hermite coefficients do not necessarily average linearly, but it is reasonable to reconstruct their variation between anchor points by interpolation if they change smoothly from one anchor point to another. An adequate anchor point separation minimizes the number of required anchor points while keeping the error introduced by the interpolation within tolerable limits. Since local 1-point statistics, such as the mean and the variance, are relatively straightforward to infer, these are used to assess the trade-off between the number of anchor points and the error introduced by interpolation.

5 Locally Stationary Spatial Estimation

Estimation and simulation under the locally stationary assumption use the interpolated local variogram parameters and Hermite coefficients to update the covariance matrix and the normal scores transformation function at each location. Locally stationary simple kriging (LSSK) is central for estimation and simulation under the assumption of local stationarity. Its equations are akin to the traditional simple kriging (Goovaerts 1997), but with spatial correlation models that are locally dependent. The locally stationary simple kriging system of equations is expressed as

$$\sum_{\beta=1}^{n(\mathbf{o})} \lambda_\beta^{(LSSK)}(\mathbf{o}) \rho_Z(\mathbf{u}_\beta - \mathbf{u}_\alpha; \mathbf{o}) = \rho_Z(\mathbf{o} - \mathbf{u}_\alpha; \mathbf{o}), \quad \alpha = 1, \ldots, n(\mathbf{o}), \qquad (12)$$

where $\lambda_\beta^{(LSSK)}$ are the kriging weights and $\rho_Z(\mathbf{h}; \mathbf{o})$ is the location dependent correlogram model and $n(\mathbf{o})$ is the number of samples within an ellipsoidal search neighborhood centered at $\mathbf{o}$. Thus, $\mathbf{o}$ is the estimation location, the location where the local correlations $\rho_Z(\mathbf{h}; \mathbf{o})$ are defined, and the center of the search ellipsoid. For the search ellipsoid, a simple option is to use a unique isotropic moving neighborhood with fixed dimensions large enough to accommodate the different local variogram ranges throughout the estimation domain. A more elaborated option would be to use locally changing neighborhoods with dimensions and orientations related to range and anisotropy of the local variogram models.

Equation (12) requires the data–data and data-estimated location correlation matrices change from one location to another. The correlation for the same pair of locations will differ depending on which location is being estimated. This difference will be larger as the separation between two estimate locations grows or if the location-dependent correlograms change abruptly between adjacent locations. The resulting inconsistencies that may arise can be minimized by using wide bandwidths to infer smoothly changing local correlograms and fitting them with models with continuously changing parameters. Moreover, the local correlation between nearby samples is more relevant for estimation. The expressions for the locally-stationary simple kriging estimator, $Z_{LSSK}^*(\mathbf{o})$, and variance, $\sigma_{LSSK}^2(\mathbf{o})$, are given in Appendix C.

In locally stationary multi-Gaussian kriging (LSMGK), the local normal scores transformation is performed at every estimation point, locally stationary simple kriging is then used to obtain the conditional cdf (ccdf) in Gaussian space. This is defined by the LSMGK estimate, $Y_{LSSK}^*(\mathbf{o})$, and variance, $\sigma_{LSSK}^2(\mathbf{o})$. The ccdf in original units is subsequently built by back-transforming the local Gaussian quantiles $y_p(\mathbf{o})$, that is,

$$z_p(\mathbf{o}) = \varphi_Z\big(y_p(\mathbf{o}); \mathbf{o}\big) \approx \sum_{q=0}^{Q} \phi_q(\mathbf{o}) H_q\big[y_p(\mathbf{o})\big]$$

$$= \sum_{q=0}^{Q} \phi_q(\mathbf{o}) H_q\big[Y_{LSSK}^*(\mathbf{o}) + \sigma_{LSSK}(\mathbf{o}) \cdot t_p\big]. \qquad (13)$$

In this expression, t_p is the p-quantile in standard Gaussian units, and $H_q[y_p(\mathbf{o})]$ and $\phi_q(\mathbf{o})$ are the Hermite polynomials and their local coefficients, respectively (see Appendix B). The mean and variance of the ccdf in original units can be approximated by the mean and the variance of the local quantiles $z_p(\mathbf{o})$.

6 Example

The well-known Walker Lake dataset (Isaaks and Srivastava 1989) is used for illustrating the proposed methodology and algorithms. Figure 4(a) presents the exhaustive Walker Lake dataset overlaid by the clustered samples that are used for locally stationary modeling. The scale and the elevation attribute in this map have been distorted. Thus, their corresponding units will be assumed generic.

A Gaussian kernel with a 30 units bandwidth and a background of 0.01 was used for the inference of location-dependent cdfs and its statistics for each one of the anchor points located in a 20×20 units grid. Figure 4(b) shows the location of the anchor points as well as a circle of radius equal to the kernel bandwidth centered in one of them. The original distance weights were corrected by declustering weights, as in Eq. (4). Figure 5 presents the location-dependent cdfs for different anchor points and the global cdf. Each one of these location-dependent cdfs was locally normal scores transformed and their corresponding transformation function modeled by series of 40 Hermite polynomials. The location-dependent experimental correlograms, $\hat{\rho}_Y(\mathbf{h}; \mathbf{o})$ were inferred from the locally transformed values for 6 different
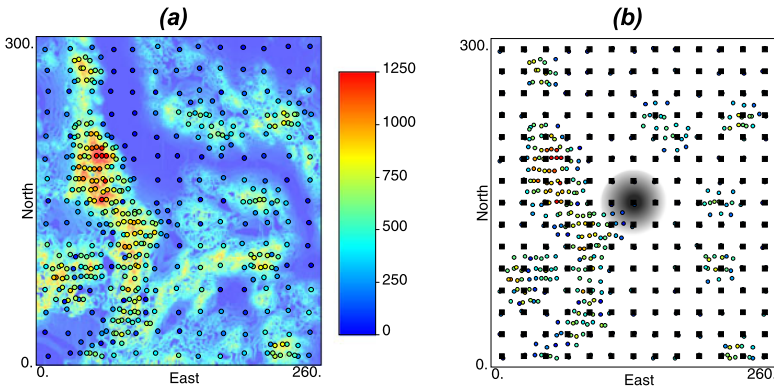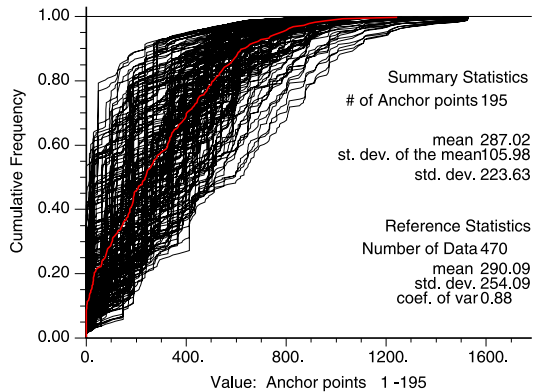
Fig. 4 (**a**) Clustered samples overlying the exhaustive Walker Lake dataset. (**b**) Anchor points locations (*small squares*) overlying the clustered dataset (*small circles*). The fading *grey circle* in the middle of the figure represents the bulge of the moving Gaussian kernel

Fig. 5 Location-dependent cdfs inferred at 195 different anchor points

directions at each anchor point. The $1 - \hat{\rho}_Y(\mathbf{h}; \mathbf{o})$ experimental curves were fitted semi-automatically using a nugget effect and an exponential model with local parameters. Figure 6 shows the interpolated maps of the resulting local variogram parameters. The orientation of maximum continuity for the global normal scores variogram model is Az. 165°. This global variogram model is defined as

$$\gamma(\mathbf{h}) = 0.25 + 0.75 \cdot \mathrm{Exp}_{\substack{\mathbf{h}\,\mathrm{max}=119 \\ \mathbf{h}\,\mathrm{min}=28}}(\mathbf{h}). \qquad (14)$$

The moving circular search window used for both, MGK and LSMGK, has a radius of 135 distance units, which is large enough to accommodate any global or local variogram range. Figure 7 shows a comparison of the cross validation results between MGK based on simple kriging (MGSK, Fig. 7(a)), ordinary kriging (MGOK, Fig. 7(b)) and LSMGK (Fig. 7(c)). For this dataset, LSMGK cross-validation results present a noticeable reduction in the mean square error. Also, despite the slight increase in the standard deviation of estimates, LSMGK shows an increment of the coefficient of correlation between true and estimated values from around 0.77, for
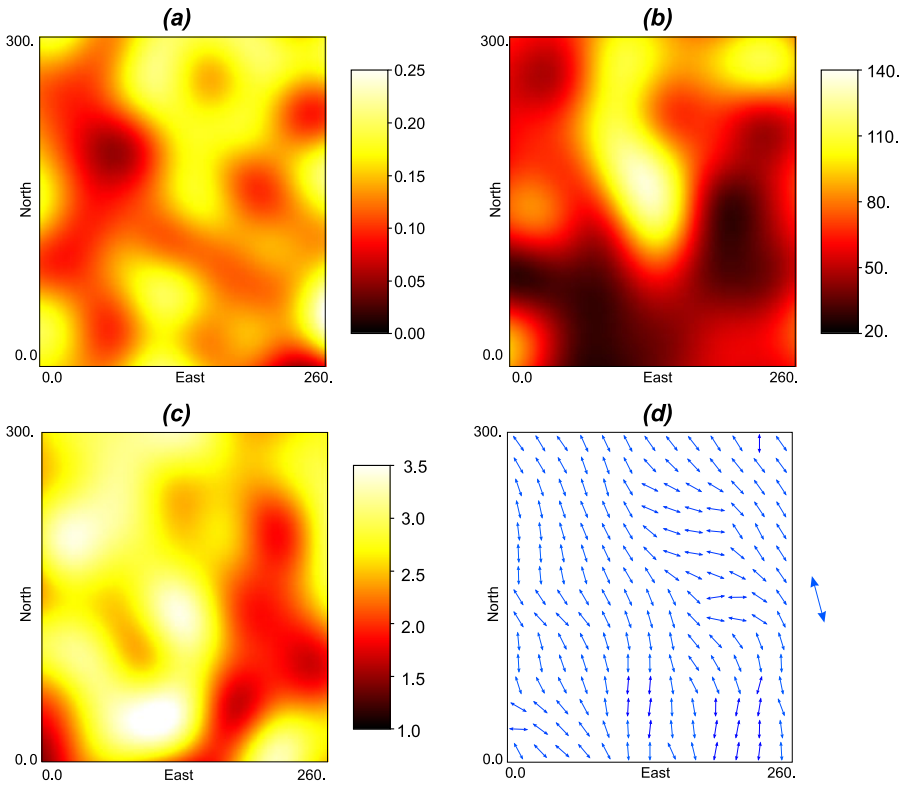
Fig. 6 Local parameters of the exponential model fitted on the experimental correlograms of locally transformed normal scores values: (**a**) nugget effect, (**b**) maximum range, (**c**) anisotropy ratio, (**d**) anisotropy orientation
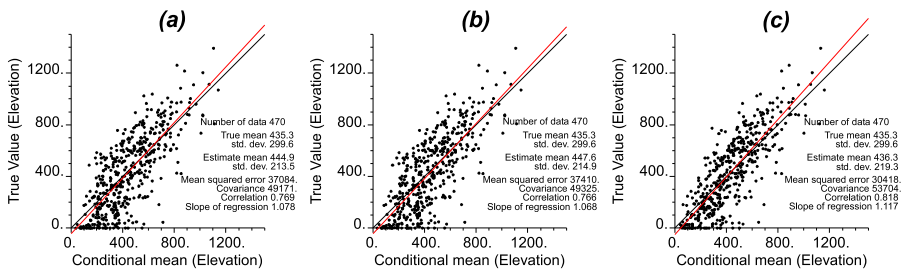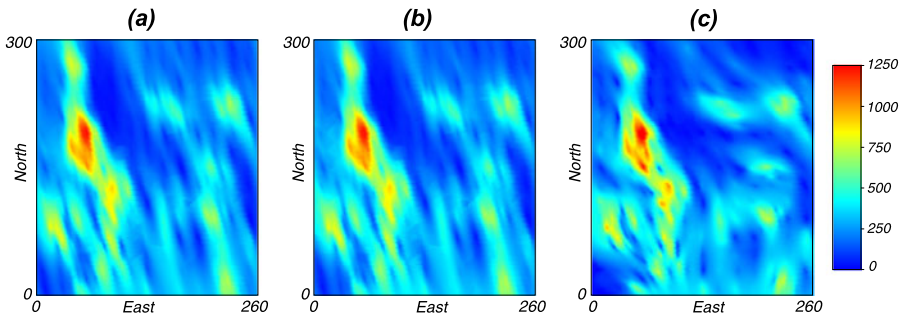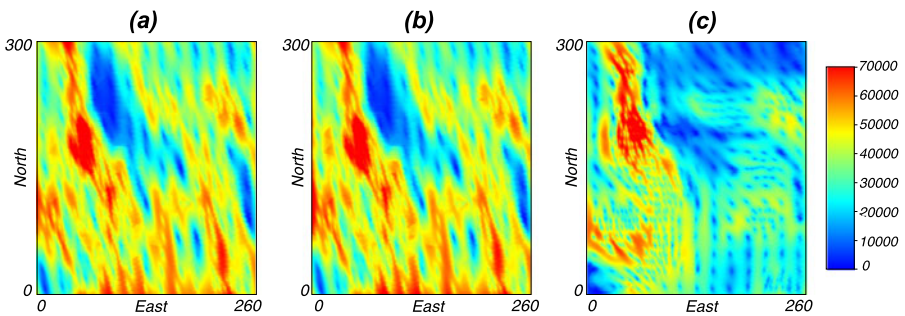

Fig. 7 Cross-validation scatterplots for (**a**) multi-Gaussian simple kriging, (**b**) multi-Gaussian ordinary kriging, and (**c**) locally stationary multi-Gaussian kriging

both MGSK and MGOK, to 0.82. This is because the covariance between true and estimated values is larger for LSMGK cross-validation results. Table 1 presents the summary statistics for the cross-validation results using the two traditional kriging methods and the proposed one.

Table 1 Cross-validation statistics for multi-Gaussian simple kriging, multi-Gaussian ordinary kriging, and locally stationary multi-Gaussian kriging

	Data	MGSK	MGOK	LSMGK
Mean (clustered)	435.4	444.9	447.6	436.3
Standard deviation (clustered)	299.6	213.5	214.9	219.3
Mean square error		37084	37410	30418
Covariance between true and estimated values		49171	49325	53704
Correlation between true and estimated values		0.769	0.766	0.818

Fig. 8 Estimates maps for (**a**) multi-Gaussian simple kriging, (**b**) multi-Gaussian ordinary kriging, and (**c**) locally stationary multi-Gaussian kriging

Fig. 9 Conditional variance maps for (**a**) multi-Gaussian simple kriging, (**b**) multi-Gaussian ordinary kriging, and (**c**) locally stationary multi-Gaussian kriging

While both MGSK and MGOK estimates maps (Figs. 8(a) and 8(b), respectively) show a dominant direction of spatial continuity imposed by an invariant model of spatial correlation, the LSMGK map (Fig. 8(c)) presents the full range of variation of the directions of spatial continuity informed by the local variogram models. The incorporation of local cdfs and variograms models in LSMGK causes the noticeable differences in the resulting variances map (Fig. 9(c)) when compared with the MGSK and MGOK variances maps (Figs. 9(a) and 9(b), respectively).

Table 2 Basic statistics for multi-Gaussian simple kriging, multi-Gaussian ordinary kriging, and locally stationary multi-Gaussian kriging estimates

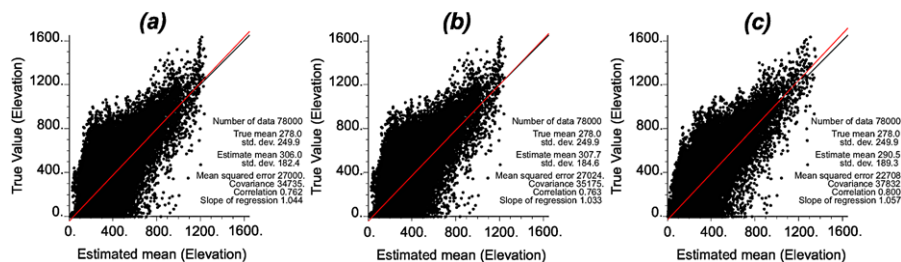|	Mean	Variance	Average of ccdf variances in Gaussian units	Average of ccdf variances in original units
Data	290.1	64,562		
MGSK	306	33,263	0.602	39,444
MGOK	307.7	34,067	0.605	39,615
LSMGK	290.5	35,794	0.630	29,592

Fig. 10 True vs. estimated values scatterplots for (**a**) multi-Gaussian simple kriging, (**b**) multi-Gaussian ordinary kriging, and (**c**) locally stationary multi-Gaussian kriging

Table 3 True and estimated values univariate and bivariate statistics for multi-Gaussian simple kriging, multi-Gaussian ordinary kriging, and locally stationary multi-Gaussian kriging

	Data	MGSK	MGOK	LSMGK
Mean	278	306	307.7	290.5
Standard deviation	249.9	182.4	184.6	189.3
Mean square error		27000	27024	22708
Covariance between true and estimated values		34735	35175	37832
Correlation between true and estimated values		0.762	0.763	0.800

Table 2 presents some summary statistics of the traditional and locally stationary multi-Gaussian kriging estimates and ccdf variances. The variance of LSMGK estimates is larger, which is an indication of reduced smoothing. Although the average of LSMGK conditional variances is slightly larger in Gaussian units, the back-transformation constrained by the a priori local distributions results in a reduction of the average of conditional variances in original units.

The comparison of scatterplots of MGSK, MGOK (Figs. 10(a) and 10(b)) and LSMGK (Fig. 10(c)) estimates versus the values of the exhaustive dataset shows that, for this particular case, LSMGK provides a more accurate estimation. This is evidenced by a reduced mean square error, and by an increased covariance and correlation coefficient between true and estimated values, as Table 3 shows.

7 Discussion

Local statistics should capture actual local trends of the attribute rather than the influence of a few nearby samples. This is better accomplished with very dense sampling. In the case of the Walker Lake dataset, the number of samples and their locations allow for the modeling of trends in the mean, variance, and spatial continuity by the location-dependent statistics. Particularly, the changes in the orientations of spatial continuity are satisfactorily reproduced by the 2-point statistics. Moreover, kriging variances reflect not only the samples configuration and availability around estimates, but also the local variability. Local normal scores transformation carry information of the changes in the local statistics for their use in Gaussian-based estimation algorithms. These features allow LSMGK to outperform simple and ordinary MGK in terms of accuracy and reduced uncertainty; this last expressed as the average of ccdf variances. The improvement obtained by locally stationary modeling is observed for both cross-validation results and in the posterior confirmation of the model by the exhaustive dataset. This may not be the case when samples are scarce, the sampling design is poor, or the kernel bandwidth too narrow in relation to the samples separation. In those cases, the inference of location-dependent statistics may be poor and the consequent increased fluctuations in them reflect a higher associated uncertainty. These fluctuations are related to the error that arises from differences between the estimated statistics and the statistics of the unique spatial distribution of a geological attribute (Marchant and Lark 2004). The effect of reducing the bandwidth on the estimation of locally stationary statistics is similar to the effect of reducing the sample size and the ratio between the domain size and the spatial correlation range for stationary statistics (Muñoz Pardo 1987; Webster and Oliver 1992). A measure of the uncertainty in the estimation of local statistics is a necessary objective for forthcoming research. This could be helpful to define an appropriate kernel bandwidth and it would allow for transferring the uncertainty about the local statistics to the uncertainty about the estimates.

Currently, the choice of the kernel bandwidth relies mostly on the practitioner's judgment. If the local statistics vary abruptly from one anchor point to another, this may be an indication of the issues discussed above. Unless a hard domain boundary or another discontinuity is present, the local statistics should vary smoothly across the domain. This also allows for minimizing inconsistencies and artifacts that may arise from assigning different local covariance values to the same pairs of locations in relation to adjacent anchor points. A related pending challenge is the development of semiautomatic algorithms for the joint fitting of variogram models with local parameters that change smoothly within limits imposed by the user.

8 Conclusions

The use of location-dependent distributions and statistics under an assumption of local stationarity with locally conditioned distributions and statistics is a viable approach for building numerical models that account for different aspects of nonstationarity in an integrated way. The resulting models show a greater richness of spatial features, clearly contrasting with the homogeneous spatial continuity imposed by global variogram models and the assumption of a stationary distribution.

The distance weighted distributions and statistics can adapt to nonstationary aspects of the attribute. This capability depends on data availability and the choice of the distance weighting function parameters. If data is abundant enough to allow for a robust inference of the location-dependent statistics, the resulting models are able to reproduce actual non-stationary patterns of the attribute. Moreover, these models can result in increased accuracy and reduced uncertainty and, therefore, they have the potential to provide an improved support for engineering decisions. Some future areas of research include the determination of the uncertainty about location-dependent statistics and its transfer to final models, and improved joint fitting algorithms for local variograms.

Appendix A: Locally Weighted Statistics

Weighted local univariate statistics have been previously presented as geographically weighted statistics (Brunsdon et al. 2002). For a set of calibrated weights $\hat{\omega}(\mathbf{u}_\alpha; \mathbf{o})$, $\alpha = 1, \ldots, n$, anchored at point $\mathbf{o}$, the local mean and variance are obtained from

$$\hat{m}_Z(\mathbf{o}) = \sum_{\alpha=1}^{n} \hat{\omega}(\mathbf{u}_\alpha; \mathbf{o}) \cdot z(\mathbf{u}_\alpha) \tag{A.1}$$

and

$$\hat{\sigma}^2(\mathbf{o}) = \sum_{\alpha=1}^{n} \hat{\omega}(\mathbf{u}_\alpha; \mathbf{o})\big[z(\mathbf{u}_\alpha) - \hat{m}(\mathbf{o})\big]^2. \tag{A.2}$$

Given a separation vector $\mathbf{h}$, the locally weighted variograms are estimated by (Machuca-Mory and Deutsch 2008; Harris et al. 2010)

$$\hat{\gamma}(\mathbf{h}; \mathbf{o}) = \frac{1}{2} \sum_{\alpha=1}^{N(\mathbf{h})} \omega'(\mathbf{u}_\alpha, \mathbf{u}_\alpha + \mathbf{h}; \mathbf{o}) \cdot \big[z(\mathbf{u}_\alpha) - z(\mathbf{u}_\alpha + \mathbf{h})\big]^2, \tag{A.3}$$

where $\omega'(\mathbf{u}_\alpha, \mathbf{u}_\alpha + \mathbf{h}; \mathbf{o})$ are the standardized 2-point distance weights and $N(\mathbf{h})$ is the number of pairs separated by $\mathbf{h}$. The local experimental covariance can be estimated as

$$\hat{C}(\mathbf{h}; \mathbf{o}) = \sum_{\alpha=1}^{N(\mathbf{h})} \omega'(\mathbf{u}_\alpha, \mathbf{u}_\alpha + \mathbf{h}; \mathbf{o})\big[z(\mathbf{u}_\alpha) - \hat{m}_{-\mathbf{h}}(\mathbf{o})\big]\big[z(\mathbf{u}_\alpha + \mathbf{h}) - \hat{m}_{+\mathbf{h}}(\mathbf{o})\big]. \tag{A.4}$$

The tail and head local means are obtained from

$$\begin{aligned}
\hat{m}_{-\mathbf{h}}(\mathbf{o}) &= \sum_{\alpha=1}^{N(\mathbf{h})} \omega'(\mathbf{u}_\alpha, \mathbf{u}_\alpha + \mathbf{h}; \mathbf{o}) \cdot z(\mathbf{u}_\alpha), \\
\hat{m}_{+\mathbf{h}}(\mathbf{o}) &= \sum_{\alpha=1}^{N(\mathbf{h})} \omega'(\mathbf{u}_\alpha, \mathbf{u}_\alpha + \mathbf{h}; \mathbf{o}) \cdot z(\mathbf{u}_\alpha + \mathbf{h}).
\end{aligned} \tag{A.5}$$

And the local experimental correlogram is estimated as

$$\hat{\rho}(\mathbf{h}; \mathbf{o}) = \frac{\hat{C}(\mathbf{h}; \mathbf{o})}{\sqrt{\hat{\sigma}^2_{-\mathbf{h}}(\mathbf{o}) \cdot \hat{\sigma}^2_{+\mathbf{h}}(\mathbf{o})}} \in [-1, +1]. \tag{A.6}$$

The required local tail and head variances are estimated by

$$\hat{\sigma}^2_{-\mathbf{h}}(\mathbf{o}) = \sum_{\alpha=1}^{N(\mathbf{h})} \omega'(\mathbf{u}_\alpha, \mathbf{u}_\alpha + \mathbf{h}; \mathbf{o}) \cdot \left[z(\mathbf{u}_\alpha) - \hat{m}_{-\mathbf{h}}(\mathbf{o})\right]^2,$$
$$\hat{\sigma}^2_{+\mathbf{h}}(\mathbf{o}) = \sum_{\alpha=1}^{N(\mathbf{h})} \omega'(\mathbf{u}_\alpha, \mathbf{u}_\alpha + \mathbf{h}; \mathbf{o}) \cdot \left[z(\mathbf{u}_\alpha + \mathbf{h}) - \hat{m}_{+\mathbf{h}}(\mathbf{o})\right]^2. \tag{A.7}$$

Appendix B: Hermite Model of Local Gaussian Transformation

Given a number of quantiles $z_p(\mathbf{o})$ of the local distribution, the local Gaussian transformation function is approximated by (Journel and Huijbregts 1978; Wackernagel 2003)

$$z_p(\mathbf{o}) = \varphi_Z(y_p; \mathbf{o}) \approx \sum_{q=0}^{Q} \phi_q(\mathbf{o}) H_q[y_p]. \tag{B.1}$$

$H_q[y_p]$ is the Hermite polynomial of order q, and the corresponding local coefficients $\phi_q(\mathbf{o})$ are obtained by

$$\phi_q(\mathbf{o}) = \sum_{p=2}^{P} (z_{p-1}(\mathbf{o}) - z_p(\mathbf{o})) \frac{1}{\sqrt{q}} H_{q-1}(y_p) g(y_p). \tag{B.2}$$

Depending on the number of samples and the complexity of the cdf, between 50 to 200 local quantiles may suffice for the modeling of the local Gaussian transformation function. The expansion into Hermite polynomials allows for the approximation of the local variance (Rivoirard 1994)

$$\hat{\sigma}^2_Z(\mathbf{o}) \approx \sum_{q=1}^{Q} \phi_q^2(\mathbf{o}). \tag{B.3}$$

In practice, expansions of degree 20 and 40 are commonly used (Vann and Sans 1996; Wackernagel 2003, p. 247).

Appendix C: Locally Stationary Simple Kriging Estimator and Variance

The locally-stationary simple kriging variance is expressed as

$$\sigma^2_{LSSK}(\mathbf{o}) = C_Z(0; \mathbf{o}) \left[1 - \sum_{\alpha=1}^{n(\mathbf{o})} \lambda_\alpha^{(LSSK)}(\mathbf{o}) \rho_Z(\mathbf{o} - \mathbf{u}_\alpha; \mathbf{o}) \right]. \tag{C.1}$$

$C_Z(0; \mathbf{o})$ is equivalent to the location-dependent variance. The locally stationary simple kriging estimator is

$$Z_{LSSK}^*(\mathbf{o}) = \sum_{\alpha=1}^{n(\mathbf{o})} \lambda_\alpha^{(LSSK)}(\mathbf{o}) \cdot Z(\mathbf{u}_\alpha) + \left[1 - \sum_{\alpha=1}^{n(\mathbf{o})} \lambda_\alpha^{(LSSK)}(\mathbf{o})\right] m(\mathbf{o}), \qquad \text{(C.2)}$$

with $m(\mathbf{o})$ as the local mean.

References

Alabert F (1987) The practice of fast conditional simulations through the LU decomposition of the covariance matrix. Math Geol 19:369–386. doi:10.1007/bf00897191

Boisvert J, Manchuk J, Deutsch C (2009) Kriging in the presence of locally varying anisotropy using non-Euclidean distances. Math Geosci 41:585–601. doi:10.1007/s11004-009-9229-1

Borradaile G (2003) Statistics of earth science data: their distribution in time, space, and orientation. Springer, Berlin Heidelberg

Bourgault G (1997) Spatial declustering weights. Math Geol 29:277–290. doi:10.1007/bf02769633

Brunsdon C, Fotheringham A, Charlton M (2002) Geographically weighted summary statistics—a framework for localised exploratory data analysis. Comput Environ Urban 26:501–524. doi:10.1016/s0198-9715(01)00009-6

Chilès JP, Delfiner P (1999) Geostatistics: modeling spatial uncertainty. Wiley, New York

Cressie N (1985) Fitting variogram models by weighted least squares. Math Geol 17:563–586. doi:10.1007/bf01032109

Davis J (2002) Statistics and data analysis in geology. Wiley, New York

Deutsch C (2002) Geostatistical reservoir modeling. Oxford University Press, New York

Deutsch C, Journel A (1998) GSLIB: geostatistical software library and user's guide. Oxford University Press, New York

Dimitrakopoulos R, Luo X (2004) Generalized sequential Gaussian simulation on group size v and screen-effect approximations for large field simulations. Math Geol 36:567–591. doi:10.1023/B:MATG.0000037737.11615.df

Emery X (2010) Iterative algorithms for fitting a linear model of coregionalization. Comput Geosci 36:1150–1160. doi:10.1016/j.cageo.2009.10.007

Fotheringham A (1997) Trends in quantitative methods I: stressing the local. Prog Hum Geogr 21:88–96. doi:10.1191/030913297676693207

Fotheringham A, Brunsdon C, Charlton M (2002) Geographically weighted regression: the analysis of spatially varying relationships. Wiley, New York

Gómez-Hernández JJ, Cassiraga EF (1993) Theory and practice of sequential simulation. In: Armstrong M, Dowd PA (eds) Geostatistical simulations: proceedings of the geostatistical simulation workshop. Kluwer Academic, Fontainebleau, France

Goovaerts P (1997) Geostatistics for natural resources evaluation. Oxford University Press, New York

Gringarten E, Deutsch C (2001) Teacher's aide variogram interpretation and modeling. Math Geol 33:507–534. doi:10.1023/a:1011093014141

Haas T (1990) Kriging and automated variogram modeling within a moving window. Atmos Environ, A Gen Top 24:1759–1769. doi:10.1016/0960-1686(90)90508-k

Harris P, Charlton M, Fotheringham A (2010) Moving window kriging with geographically weighted variograms. Stoch Environ Res Risk Assess 24:1193–1209. doi:10.1007/s00477-010-0391-2

Isaaks E, Srivastava R (1989) An introduction to applied geostatistics. Oxford University Press, New York

Journel A (1986) Geostatistics: models and tools for the earth sciences. Math Geol 18:119–140. doi:10.1007/bf00897658

Journel A, Huijbregts C (1978) Mining geostatistics. Academic Press, New York

Korvin G (1982) Axiomatic characterization of the general mixture rule. Geoexploration 19:267–276. doi:10.1016/0016-7142(82)90031-x

Machuca-Mory D, Deutsch C (2008) Location dependent variograms. In: Ortiz JM, Emery X (eds) Geostats 2008: proceedings of the eighth international geostatistics congress. University of Chile, Santiago, Chile, pp 497–506

Marchant B, Lark R (2004) Estimating variogram uncertainty. Math Geol 36:867–898. doi:10.1023/B:MATG.0000048797.08986.a7

Matheron G (1970) La théorie des variables régionalisées et ses applications. Cahiers du Centre de Morphologie Mathématique de Fontainebleau. Ecole des Mines de Paris

McLennan J (2007) The decision of stationarity. Dissertation, University of Alberta, Canada, 191 pp

Monestiez P, Switzer P (1991) Semiparametric estimation of nonstationary spatial covariance models by metric multidimensional scaling. Technical report, Stanford University, USA, 29 pp

Muñoz-Pardo JF (1987) Approche géostatistique de la variabilité spatiale des milieux géophysiques: application à l'échantillonnage de phénomènes bidimensionnels par simulation d'une fonction aléatoire. Dissertation, University Grenoble-1, France, 200 pp

Myers D (1989) To be or not to be... stationary? That is the question. Math Geol 21:347–362. doi:0882-8121/89/0400-0347506.00/1

Omre H (1984) The variogram and its estimation. In: Verly G, David M, Marechal A, Journel A (eds) Geostatistics for natural resources characterization proceedings of NATO ASI conference at Lake Tahoe, vol 1. Reidel, Dordrecht, pp 107–125

Pardo-Igúzquiza E (1999) VARFIT: a Fortran-77 program for fitting variogram models by weighted least squares. Comput Geosci 25:251–261. doi:10.1016/s0098-3004(98)00128-9

Rivoirard J (1994) Introduction to disjunctive kriging and non-linear geostatistics. Clarendon Press, Oxford

Sampson PD, Guttorp PT (1992) Nonparametric estimation of nonstationary spatial covariance structure. J Am Stat Assoc 87:108–119

Vann J, Sans H (1996) Global resource estimation and change of support at the enterprise gold mine, pine creek, northern territory-application of the geostatistical discrete Gaussian model. In: Proceedings of APCOM XXV 1995: application of computers and operations in the minerals industries. Australasian Institute of Mining and Metallurgy, Brisbane

Verly G (1983) The multiGaussian approach and its applications to the estimation of local reserves. Math Geol 15:259–286. doi:10.1007/bf01036070

Wackernagel H (2003) Multivariate geostatistics: an introduction with applications. Springer, Berlin Heidelberg

Wasserman L (2006) All of nonparametric statistics. Springer, New York

Webster R, Oliver MA (1992) Sample adequately to estimate variograms of soil properties. J Soil Sci 43:177–192. doi:10.1111/j.1365-2389.1992.tb00128.x

Xu W (1996) Conditional curvilinear stochastic simulation using pixel-based algorithms. Math Geol 28:937–949. doi:10.1007/bf02066010