

Histogram and Variogram Inference in the Multigaussian Model

Xavier Emery and Julián M. Ortiz

Department of Mining Engineering,
University of Chile

Abstract

Several iterative algorithms are proposed to improve the histogram and variogram inference in the framework of the multigaussian model. The starting point is the variogram obtained after a traditional normal score transform. The subsequent step consists in simulating many sets of Gaussian values with this variogram at the data locations, so that the ranking of the original values is honored. The expected gaussian transformation and the expected variogram are computed by an averaging operation over the simulated datasets. The variogram model is then updated and the procedure is repeated until convergence. Such an iterative algorithm can adapt to the case of tied data and despike the histogram. Two additional issues are also examined, referred to the modeling of the empirical transformation function and to the optimal pair weighting when computing the sample variogram.

Introduction

Geostatistical simulation is commonly used in environmental applications, groundwater hydrology, petroleum and mining engineering, as a tool for risk assessment and decision-making. Alternatively, the local distributions of unsampled values can be estimated with the conditional expectation estimator. Now, both approaches (conditional expectation and simulation) require modeling the spatial attribute at hand by a random function and defining a spatial distribution model. In this respect, the multigaussian model is the most widely used, for its analytical simplicity and convenience (Verly, 1986; Goovaerts, 1997, p. 271; Chilès and Delfiner, 1999, p. 381).

Of course, this model has its limitations. In particular, for its specific properties, the multigaussian spatial distribution is not always suitable to the available data. Other methods have been developed to overcome this lack of flexibility, such as the indicator approach (Journel, 1983; Goovaerts, 1997, p. 290 and 401) which does not need to define a *spatial* distribution, but only part of the *bivariate* distributions between samples. These methods are indeed more flexible, however their practical implementation requires many approximations and suffers from consistency problems.

This work deals with the aforementioned multigaussian model and focuses on the difficulties related to the inference of the univariate and bivariate distributions. It is organized as follows. The first section stresses the weaknesses of the traditional approach and proposes an iterative algorithm to improve the histogram and variogram inference. The second section deals with the fitting of a model to an empirical transformation function. In the last section, a declustering method is presented to account for the geometrical configuration of the data pairs and for their spatial correlation when computing a sample variogram.

On the empirical normal score transform

Practical limitations of the traditional approach

Suppose we are analyzing an attribute distributed over a bounded domain of an Euclidean space \mathbb{R}^d (usually, $d = 2$ or 3). Let $\{Z(\mathbf{x}), \mathbf{x} \in \mathbb{R}^d\}$ be the random field that describes this attribute and $\{Z(\mathbf{x}_\alpha), \alpha = 1 \dots n\}$ the available dataset. The normal score transform is a quantile-to-quantile transformation that turns the original data into a set of standard gaussian values $\{Y(\mathbf{x}_\alpha), \alpha = 1 \dots n\}$ (Goovaerts, 1997, p. 273):

$$Y(\mathbf{x}) = \psi[Z(\mathbf{x})] \quad (1)$$

where \square is a non-decreasing function called **gaussian transformation**. The multigaussian model is fully determined once the experimental variogram of the normal score data is fitted. However, many practitioners have found that such a sample variogram does not always reach a unit sill, which contradicts the assumption that the transformed data have a standard gaussian distribution with a unit variance. Several causes can be identified:

- 1) stationarity defects: such a situation may occur if the values show a spatial “trend”;
- 2) number and configuration of the data: the sample variogram is not robust when few data are available. Moreover, experience has shown that an irregular or a preferential sampling is likely to deteriorate the quality of this variogram (Rivoirard, 2001, p. 149). Such statements prove the need for a declustering technique when calculating a sample variogram, a point that will be examined in a subsequent section of this work;
- 3) sampled domain too small, which has two different effects on the variogram:
 - a) “chance fluctuations”: given a dataset with a known variogram model, the sample variogram values are more dispersed when the distance increases. The word “fluctuation” refers to the deviation between the sample variogram and the underlying model. In the probabilistic approach, it is a random variable with zero mean, but its variance is not negligible as soon as the lag distance is not small with respect to the domain size (Matheron, 1965, p. 227; Chilès and Delfiner, 1999, p. 137);
 - b) the *empirical* distribution of a dataset differs from its *theoretical* distribution, which refers to the marginal distribution over an infinite domain. A striking illustration of this assertion concerns the second-order moment of the distribution: if the sampled domain is not large enough with respect to the range of the variogram, then the dispersion of the histogram (*empirical* variance) tends to be smaller than the variogram sill (*a priori* variance). Such an inequality is a consequence of Krige’s relationship, as the empirical variance has the meaning of a dispersion variance (Matheron, 1971, p. 52; Chilès and Delfiner, 1999, p. 132):

$$\underbrace{D^2(\circ | \infty)}_{\text{a priori variance}} = \underbrace{D^2(\circ | D)}_{\text{dispersion variance}} + \underbrace{D^2(D | \infty)}_{>0 \text{ if the sampled field } D \text{ is not infinite}} \quad (2)$$

By construction, the gaussian transformation sets the empirical variance of the normal score data to one, so that their *a priori* variance is likely to be greater (Chilès and Delfiner, 1999, p. 471). In brief, in addition to the chance fluctuations, a bias is introduced in the variogram sill when performing the classical normal score transform.

Point a) can be related to point c) as the definition of a “trend” usually depends on the scale of observation. The spatial correlations of random values located in a small domain do not vanish and may produce an apparent trend, although they stem from a stochastic process with constant mean (Chilès and Delfiner, 1999, p. 236). Points b) and c) cover two conceptually very different problems. The former proceeds from the quality and quantity of available information and refers to an “estimation” error due to the use of a finite dataset instead of an exhaustive sampling. In contrast, the latter would not disappear even if the attribute were known over the whole region (bounded domain of \mathbb{R}^d). In this work, we are interested in assessing the theoretical variogram (e.g. for constructing stochastic simulations), hence the study is oriented to the fluctuation problem: this amounts to considering that the “region” where the attribute is defined is the set of sample locations $\{\mathbf{x}_\alpha, \alpha = 1 \dots n\}$.

In practice, if the normal score variogram reaches a sill greater than one, a first solution is to fit a model with the experimental sill, then compute the dispersion variance in the sampled field and check whether this variance is close to one or not. If so, the model is accepted, otherwise the variogram has to be fitted again. Now, this procedure assumes that the theoretical distribution has a non-unit variance, hence it differs from the empirical distribution of the normal score data (which, by construction, have a unit variance) and may not be gaussian.

An alternative solution is to rescale the experimental variogram to the correct unit sill (Fig. 1A), or to model it up to a unit sill and omit the experimental values above this sill (Fig. 1B). As shown in Figure 1, the second option leads to a higher relative nugget effect and a shorter range than the first one: both options are clearly not equivalent. A last solution consists in defining a *dispersion covariance*, which is the dispersion variance (set to one) minus the variogram (Matheron, 1965, p. 138; Matheron, 1971, p. 69), and using it as a structural tool (Fig. 1C). Now, if the variogram reaches a sill greater than one, the dispersion covariance takes negative values for the large distances. Therefore, this description cannot be extended to the cases that require the covariance to be positive, e.g. the gamma isofactorial model (Chilès and Delfiner, 1999, p. 403).

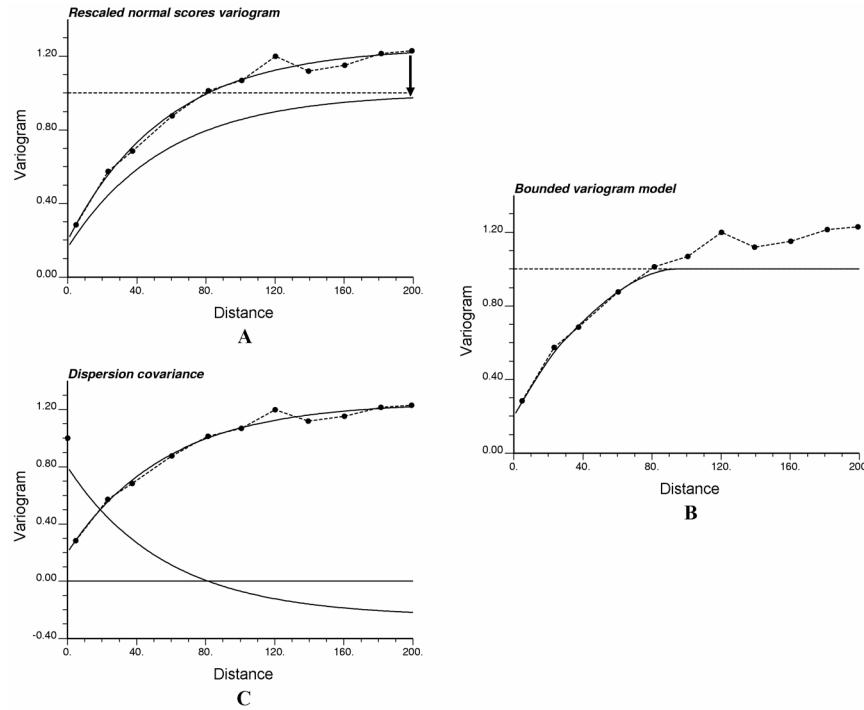


Figure 1. Three options for modeling the variogram of the normal score data

An iterative algorithm to perform the normal score transform

In the following, let us assume that the gaussian values are ranked in ascending order:

$$Y(x_1) < Y(x_2) < \dots < Y(x_n) \quad (3)$$

As stated earlier, one should distinguish between the *empirical distribution* of these values and their *theoretical distribution*. The multigaussian model requires the latter to be gaussian-shaped, but the former is likely to be slightly different. Consequently, the empirical distribution of the gaussian values (which is the target distribution for the normal score transform) is unknown.

To break the deadlock and obtain a dataset consistent with a theoretical multigaussian distribution, an iterative algorithm is proposed. This algorithm also brings a solution to two additional problems that are often critical in geostatistical studies: the declustering and despiking of the data histogram. It can be split into the following steps.

- 1) Perform a classical normal score transform and fit a variogram model $\gamma(\mathbf{h})$.
- 2) Simulate many independent sets of multigaussian values with variogram $\gamma(\mathbf{h})$ at the data locations $\{\mathbf{x}_\alpha, \alpha = 1 \dots n\}$ (a standardization like in Fig. 1A may be considered to correct the variogram sill). Each simulated set should account for the ordering of the original data [Eq. (3)], which is the only constraint the gaussian transformation must honor. To do so, a Gibbs sampler can be used (Geman and Geman, 1984).

Initialization: consider the values $\{Y(x_\alpha), \alpha = 1 \dots n\}$ obtained with the classical normal score transform as the first state.

Iteration

- Choose an index α at random in $\{1 \dots n\}$.
- Replace $Y(x_\alpha)$ by a gaussian value conditioned by all the other values $\{Y(x_\beta), \beta \neq \alpha\}$ and by the inequality constraints given by Eq. (3). This amounts to sampling a truncated gaussian distribution (Fig. 2); an acceptance-rejection criterion can be used (Freulon and de Fouquet, 1993).
- Go back to 2a) and loop many times.

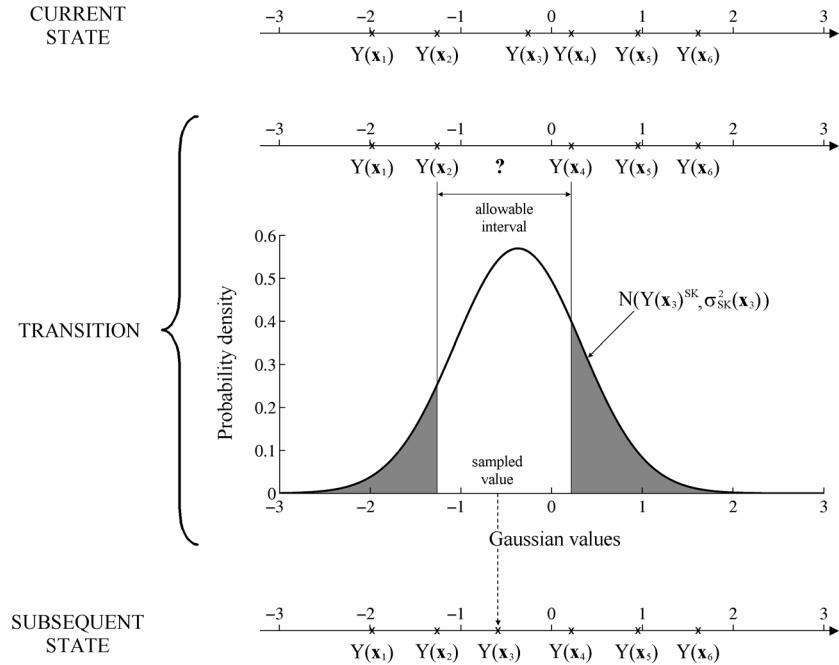


Figure 2. An elementary transition in the Gibbs sampler, corresponding to the replacement of $Y(x_3)$

- For each simulated set, the plot of the gaussian values as a function of the raw values gives an unbiased estimate of the gaussian transformation [Eq. (1)], but the dispersion of the error on a single simulation may be high. The error variance is reduced by averaging the gaussian transformations over the set of simulations, which defines the *expected gaussian transformation* (conditional mean curve of the cloud of points).
- Similarly, compute the sample variograms of all the simulated datasets. Their average gives the *expected variogram*, which is an approximation of the theoretical model (free from statistical fluctuations as the average estimates the expected value).
- Update the normal values $\{Y(x_\alpha), \alpha = 1 \dots n\}$ using the expected gaussian transformation.
- Fit the expected variogram and update $\gamma(\mathbf{h})$. If the gaussian transformation or the updated variogram differ substantially from the former models, go back to step 2).

The authors have observed that in practice the algorithm converges after a few iterations (see the application hereafter), although the convergence is not proven theoretically. A big advantage of this algorithm is that the simulations account for the correlations and redundancies between the

data values, via a variogram model. Hence, the gaussian transformation function (equivalent to the histogram modeling) includes a declustering procedure suited to the spatial distribution model. Such a declustering should improve the traditional techniques that 1) omit the spatial correlation of the data, like the polygonal and cell methods (Journel, 1983; Goovaerts, 1997, p. 81), 2) do not use a multiple-point distribution, e.g. by weighting the data after their kriging weights or after their redundancy matrix (Bourgault, 1997), or 3) build a cdf estimator among a restricted class of functions, like weighted averages of indicator values at data locations (Switzer, 1977; Bogaert, 1999). However, traditional declustering approaches are more general, since they can be applied not only in the multigaussian framework; they are also useful to provide an initial guess for the Gibbs sampler.

Application

To illustrate the previous concepts, a nonconditional simulation is drawn on a 15×15 grid with a lognormal spatial distribution. The variogram of the corresponding gaussian values is the sum of an isotropic spherical model with range 5 and sill 0.8 and a nugget effect with sill 0.2. The purpose of this exercise is to perform the gaussian transformation of the lognormal dataset and obtain a variogram model for the normal score data. For variogram calculations, the lag separation distance is set to 0.5 unit with a tolerance of 0.25.

Figures 3A and 3B provide the gaussian transformation and variogram corresponding to the traditional procedure. Although the histogram of the transformed data is perfectly gaussian with a unit variance, the sample variogram reaches a sill greater than one and is not fully consistent with the standard multigaussian hypothesis. The variogram model for the initial state is a spherical model with range 7 and sill 0.8 plus a nugget effect with sill 0.35; this model is standardized to a unit sill (Fig. 1A). In the next step, 25 realizations are drawn (conditioned by the ordering of the original values) with the previous variogram model, providing a set of 25 gaussian transformations (Fig. 3C) and 25 variograms (Fig. 3D). The average gaussian transformation and average variogram are retained for the subsequent iteration. The variogram model is updated into a spherical model with range 7 and sill 0.73 plus a nugget effect with sill 0.27. The second iteration considers 25 new simulated datasets. The average gaussian transformation and average variogram are close to the ones of the previous iteration, so that there is no need to update them any more. Hence the final gaussian transformation and variogram model are obtained (Fig. 3E and 3F).

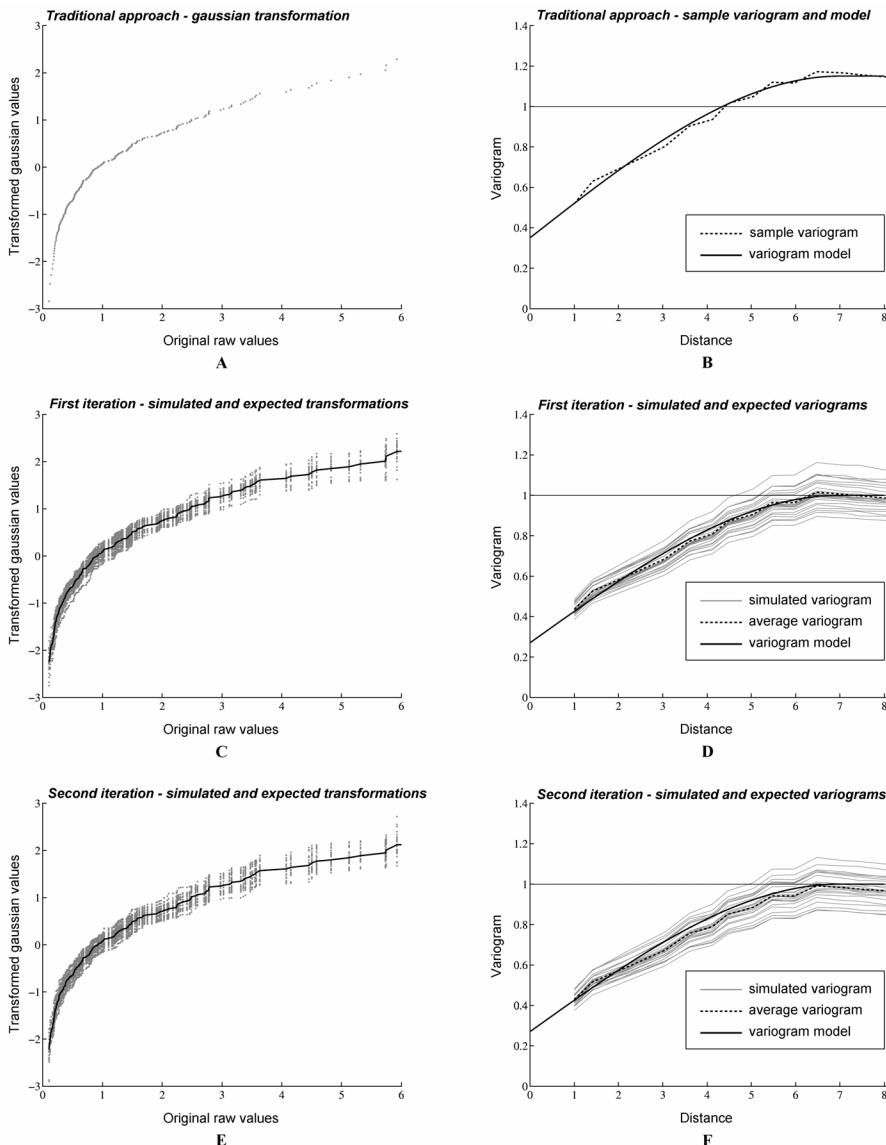


Figure 3. Expected gaussian transformations and variograms for the first steps of the iterative algorithm

Histogram despiking

The available raw data may present tied values, because of the discrete nature of the attribute, of the rounding of the values to decimal numbers or of a “zero effect”, e.g. concentrations below the detection limit and set to zero. Now, in theory, ties cannot happen with the normal score data, since two gaussian values are almost surely different. Usually, ties are broken randomly or according to the neighboring data averages (Verly, 1986). However, this is an arbitrary procedure, for instance in what refers to the definition of the domain for averaging the values.

The previous algorithm can be generalized to account for tied values. Indeed, let us assume that two raw values are equal, say $Z(\mathbf{x}_i)$ and $Z(\mathbf{x}_{i+1})$. Then condition (3) has to be split into:

$$\begin{aligned} Y(\mathbf{x}_1) < \dots < Y(\mathbf{x}_{i-1}) < Y(\mathbf{x}_i) < Y(\mathbf{x}_{i+1}) < \dots < Y(\mathbf{x}_n) \\ Y(\mathbf{x}_1) < \dots < Y(\mathbf{x}_{i-1}) < Y(\mathbf{x}_{i+1}) < Y(\mathbf{x}_{i+2}) < \dots < Y(\mathbf{x}_n) \end{aligned} \quad (4)$$

The other steps of the iterative algorithm remain unchanged, in particular a Gibbs sampler is used to honor the previous inequalities. The simulated gaussian values corresponding to both tied raw data will probably not be equal; neither will be their averages over several simulations. In other words, in general, the gaussian transformation will no longer be a one-to-one function and the algorithm will “despike” the normal score histogram.

As an example, consider the set of 25 gridded data shown in Figure 4A and assume that the variogram is a spherical model with range 3 and no nugget effect. After applying the iterative algorithm, the gaussian values associated with the zeroes are significantly different (they rank from -1.86 to -1.55) (Fig. 4B). Note that the untying depends on the variogram model: for instance, a pure nugget effect (instead of a spherical model) would have led to equal gaussian values for all the zeroes. Hence the algorithm not only accounts for the information surrounding the tied values, but also for the structural patterns of the attribute (anisotropy, nugget effect, etc.).

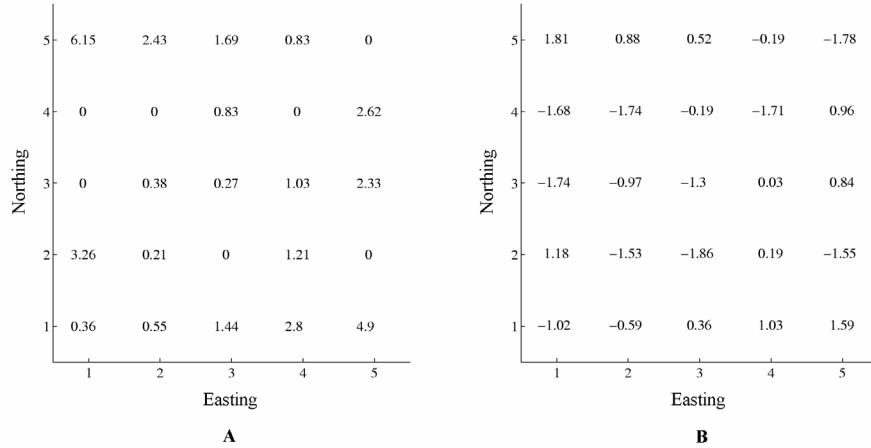


Figure 4. A, raw data with tied values and B, expected gaussian values

A similar despiking procedure was introduced by Freulon (1992, p. 22): the tied data are first associated with intervals (their exact gaussian values remain unknown), whereas the non-tied data are transformed into numerical values and allow determining the variogram model. Finally, the missing gaussian values are obtained thanks to a simulation conditioned by inequality constraints and the already known gaussian data. The approach proposed here is an extension of Freulon’s algorithm to the case when we are unsure of the transformation function for any value and only the ordering of the initial data can be used as a conditioning information.

Modeling the transformation function

When resorting to analytical techniques such as disjunctive kriging or conditional expectation, it is useful to work with the *inverse* transformation function (Φ^{-1}) rather than the direct function defined in Eq. (1). An *empirical* inverse transformation function is obtained by plotting the normal score values versus the raw values. In practice, this function needs to be known for any gaussian value, hence a model is required; this can be done by fitting a polynomial expansion (Chilès and Delfiner, 1999, p. 408). However the modeling by a polynomial is quite sensitive to the presence of important discontinuities in the empirical inverse transformation function (Fig. 5D) that deteriorate the fitting of the smaller discontinuities (Fig. 5C). As a result, the distribution of the transformed values is no longer gaussian-shaped once the modeled transformation function is applied to the raw data (Fig. 5B).

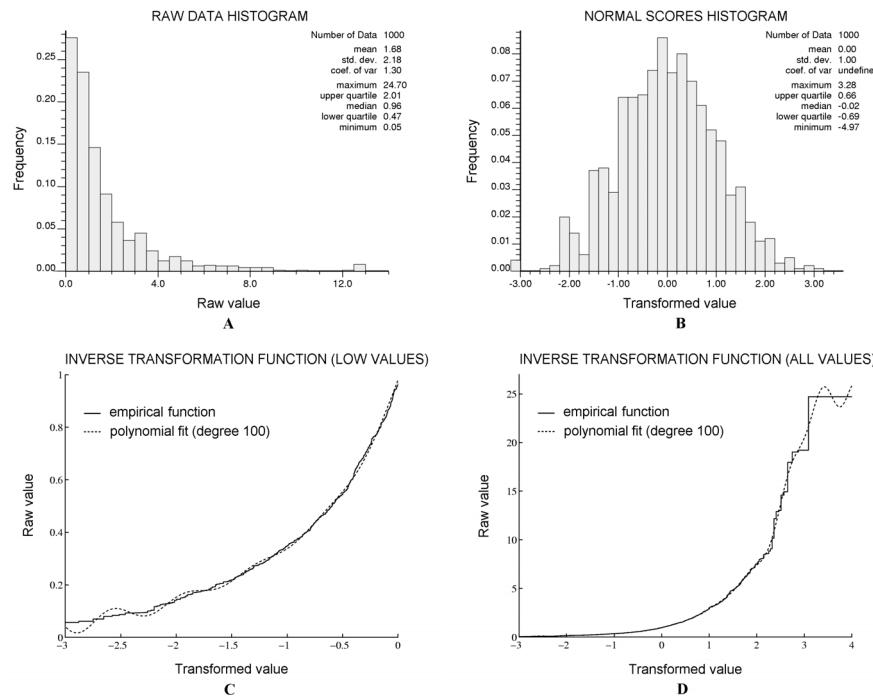


Figure 5. Polynomial fitting of the inverse transformation function for a simulated lognormal dataset

To get round this difficulty, the empirical inverse transformation function can be seen as the cloud of points $\{(Y(x_\alpha), Z(x_\alpha)), \alpha = 1 \dots n\}$, instead of a stair function (Fig. 6C), and completed by a piecewise interpolation between the empirical points. The *global* fit by a polynomial is now traded for a *local* fit with simple functions such as line segments, monomials and / or exponential functions (Fig. 6D). Although the model for Φ^{-1} is not necessarily differentiable at the data values, its polynomial expansion can still be expressed and used for disjunctive kriging or conditional expectation (Appendix A). The main advantage of this approach is that the gaussian values obtained after the iterative algorithm are the same than the ones obtained by applying the modeled transformation function to the raw data, as Φ^{-1} honors the empirical points $\{(Y(x_\alpha), Z(x_\alpha)), \alpha = 1 \dots n\}$.

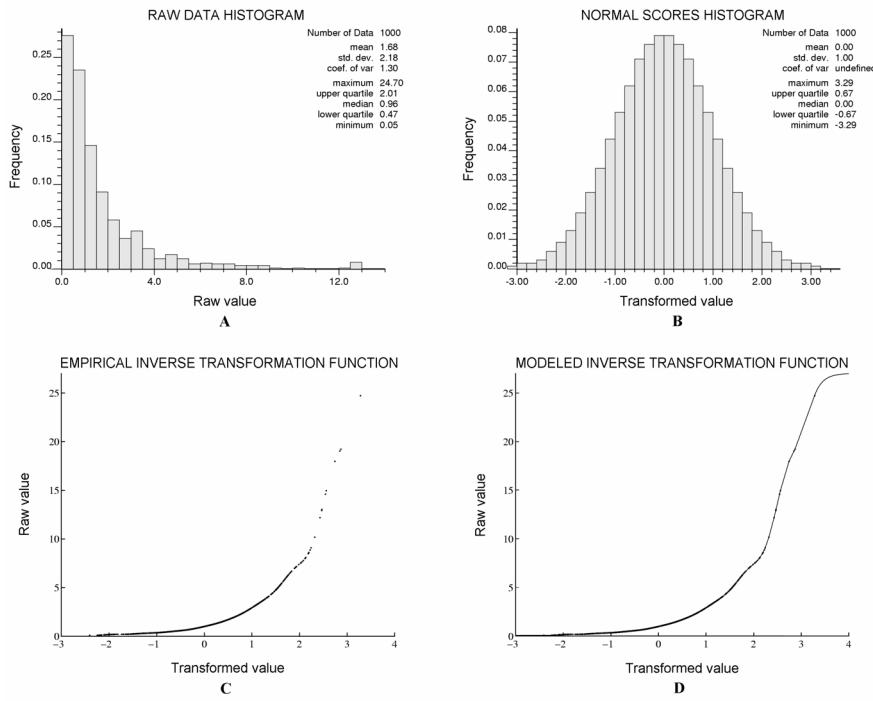


Figure 6. Piecewise fitting of the inverse transformation function for the previous dataset

Variogram declustering

At the beginning of this work, we mentioned that the histogram inference accounts for the correlations and redundancies between the data values. However, nothing has been said about the redundancies between the data pairs involved in variogram analysis. Several authors already looked into the problem of declustering the variogram (Omre, 1984; Rivoirard, 2001; Richmond, 2002), but the algorithms they suggest usually rely on geometrical considerations rather than the structural features of the data under study. Some additional elements are given hereafter in the specific case of the multigaussian model.

Optimal weighted variogram

Suppose we want to estimate the variogram of a set of values $\{Y(x_\alpha), \alpha = 1 \dots n\}$ for a given separation vector \mathbf{h} and assume that these values are distributed according to a multigaussian distribution with zero mean and unit variance. In practice, as soon as the sampling is not regular, some tolerances on the separation vector are defined, which means that the data pairs to use in the variogram estimator will not have exactly the same separation. A similar situation occurs when calculating an omnidirectional variogram: only the length of the separation vector matters, notwithstanding its orientation. So, let us define the variogram estimator as a weighted sum of the squared differences of all the pairs whose separation is equal to \mathbf{h} up to some tolerance factors:

$$\hat{\gamma}(\mathbf{h}) = \frac{1}{2} \sum_{i=1}^{N(\mathbf{h})} \omega_i [Y(x_i + \mathbf{h}_i) - Y(x_i)]^2 \quad (5)$$

One wants this estimator to be unbiased whatever the true underlying variogram (denoted by $\gamma(\mathbf{h})$ in the following):

$$E[\hat{\gamma}(\mathbf{h})] = \gamma(\mathbf{h}) = \frac{1}{2} E\{[Y(\mathbf{x} + \mathbf{h}) - Y(\mathbf{x})]^2\} \quad (6)$$

which leads to the conditions:

$$\sum_{i=1}^{N(\mathbf{h})} \omega_i = 1 \quad (7)$$

$$\forall i \in \{1, \dots, N(\mathbf{h})\}, E\{[Y(\mathbf{x}_i + \mathbf{h}_i) - Y(\mathbf{x}_i)]^2\} \approx 2\gamma(\mathbf{h}) \quad (8)$$

To honor the second condition, the tolerance factor should not be too large. To find the optimal weights, a quality criterion is chosen, which consists in minimizing the variance of the estimation error:

$$\begin{aligned} & \text{var}[\hat{\gamma}(\mathbf{h}) - \gamma(\mathbf{h})] \\ &= \frac{1}{4} \sum_{i=1}^{N(\mathbf{h})} \sum_{j=1}^{N(\mathbf{h})} \omega_i \omega_j \text{cov}\{[Y(\mathbf{x}_i + \mathbf{h}_i) - Y(\mathbf{x}_i)]^2, [Y(\mathbf{x}_j + \mathbf{h}_j) - Y(\mathbf{x}_j)]^2\} \\ &= \frac{1}{2} \sum_{i=1}^{N(\mathbf{h})} \sum_{j=1}^{N(\mathbf{h})} \omega_i \omega_j [\gamma(\mathbf{x}_i - \mathbf{x}_j - \mathbf{h}_j) + \gamma(\mathbf{x}_i - \mathbf{x}_j + \mathbf{h}_i) - \gamma(\mathbf{x}_i - \mathbf{x}_j) - \gamma(\mathbf{x}_i - \mathbf{x}_j + \mathbf{h}_i - \mathbf{h}_j)]^2 \end{aligned} \quad (9)$$

The last equality relies on the expression of the fourth order moments of a multigaussian field in function of the second order moments. It is based on the following identity (Matheron, 1965, p. 221): if $\{Y_1, Y_2, Y_3, Y_4\}$ are four gaussian random variables with mean zero, then

$$E(Y_1 Y_2 Y_3 Y_4) = E(Y_1 Y_2) E(Y_3 Y_4) + E(Y_1 Y_3) E(Y_2 Y_4) + E(Y_1 Y_4) E(Y_2 Y_3) \quad (10)$$

The variance [Eq. (9)] has to be minimized under the unbias constraint given in Eq. (7). This is similar to the estimation of an unknown mean by ordinary kriging (Matheron, 1971, p. 128; Goovaerts, 1997, p. 138), except that the opposite of the variogram between two samples is now replaced by the covariance between squared increments:

$[\gamma(\mathbf{x}_i - \mathbf{x}_j - \mathbf{h}_j) + \gamma(\mathbf{x}_i - \mathbf{x}_j + \mathbf{h}_i) - \gamma(\mathbf{x}_i - \mathbf{x}_j) - \gamma(\mathbf{x}_i - \mathbf{x}_j + \mathbf{h}_i - \mathbf{h}_j)]^2$ replaces
 $-\gamma(\mathbf{x}_i - \mathbf{x}_j)$ in the ordinary kriging system.

Hence the following system of linear equations is obtained, in which μ stands for a Lagrange multiplier:

$$\begin{cases} \sum_{j=1}^{N(\mathbf{h})} \omega_j [\gamma(\mathbf{x}_i - \mathbf{x}_j - \mathbf{h}_j) + \gamma(\mathbf{x}_i - \mathbf{x}_j + \mathbf{h}_i) - \gamma(\mathbf{x}_i - \mathbf{x}_j) - \gamma(\mathbf{x}_i - \mathbf{x}_j + \mathbf{h}_i - \mathbf{h}_j)]^2 \\ + \mu = 0 \quad \forall i \in \{1, \dots, N(\mathbf{h})\} \\ \sum_{j=1}^{N(\mathbf{h})} \omega_j = 1 \end{cases} \quad (11)$$

An interpretation of this system in terms of disjunctive kriging is given in Appendix B. The variance of the estimation error [Eq. (9)] is:

$$\text{var}[\hat{\gamma}(\mathbf{h}) - \gamma(\mathbf{h})] = -\frac{\mu}{2} \quad (12)$$

Such a variance indicates the degree of reliability of the declustered variogram and can be used when fitting the theoretical model (Pardo-Igúzquiza and Dowd, 2001). Solving Eq. (11) requires knowing the variogram model, which is precisely the parameter that is sought after. An initial model has to be chosen, for instance the variogram of the normal score data obtained by the traditional approach: once the declustered sample variogram is obtained, the model is updated and the procedure is repeated. In practice, a few iterations (say, less than ten) should be sufficient to ensure the convergence of the algorithm.

Discussion

The optimal weights account for both the geometrical configuration of the data pairs and the spatial structure of the attribute, through its variogram. Besides, they differ from one lag distance to another (i.e. the solution depends on \mathbf{h}). In general, for a given lag, the data pairs are not equally weighted, even if the variogram is a pure nugget effect and the samples are regularly spaced: hence the declustered variogram never coincides with the traditional variogram estimator. Figure 7 shows an example for a set of 25 gridded data in R^2 and a pure nugget effect model. The lag distance is set to 1, so that each pair corresponds to two adjacent data. The pairs located in the center of the sampled area receive less weight than the pairs located on the edges.

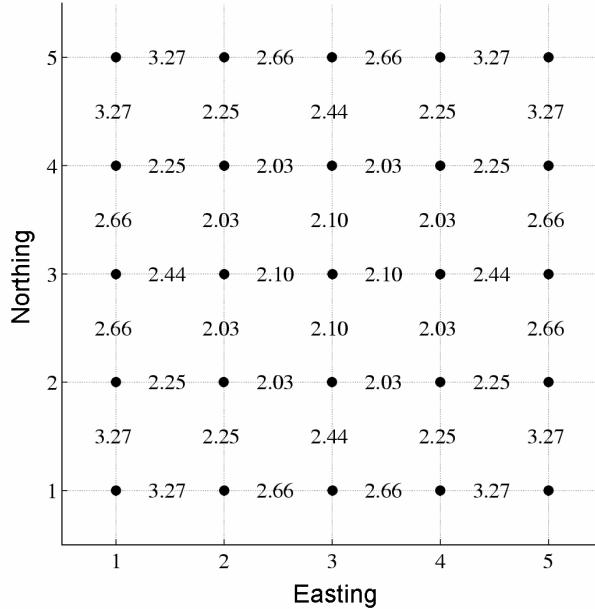


Figure 7. Optimal variogram pair weighting for gridded samples with pure nugget effect variogram. The sample locations are indicated by a black dot; the weight of each pair (in percent) is indicated in its gravity center.

A drawback of the optimal weighting is that the solution of system (11) is CPU-intensive if the estimator involves several thousands of pairs. Furthermore it is based on a strict multigaussian assumption and uses the variogram model for lags close to the domain diameter, while the model is purely conventional for such distances. Note that the latter problem also arises when estimating the first order moment (expected value) by ordinary kriging (Matheron, 1971, p. 128) or when seeking an optimal estimate of the marginal distribution with a prior variogram model (Switzer, 1977; Matheron, 1977, p. 2).

To avoid misinterpretations, the traditional and weighted variograms should be considered as complementary tools for variogram analysis. Although the declustering procedure has been established in a multigaussian framework, it may be useful in other cases as a way to complement the traditional structural analysis.

Conclusions

Iterative algorithms have been proposed to perform the histogram inference and variogram declustering in the scope of the multigaussian model. The whole methodology can be summarized as follows:

- 1) perform a traditional normal score transform;
- 2) compute the non-weighted variogram and fit a first model $\square(\mathbf{h})$;
- 3) compute the weighted variogram using this model in Eq. (11); if needed, update $\square(\mathbf{h})$ and loop until convergence;
- 4) simulate many sets of multigaussian values at the sample locations with $\square(\mathbf{h})$ as the variogram model, so that the ordering of the original data is honored (Gibbs sampler);

- 5) calculate the expected gaussian transformation (average over the simulations);
- 6) calculate the (weighted) expected variogram and update $\square(\mathbf{h})$;
- 7) go back to 4).

In general, such iterative algorithms are computationally intensive. In return, they ensure the consistency of the model parameters with the data, and provide a solution to several critical problems: declustering and despiking of the data histogram, declustering of the sample variogram. The approach can be extended to nongaussian models, provided that a conditional simulation algorithm is available and that the fourth-order moments can be expressed. Besides, although it is established in a multigaussian framework, the pair weighting for the sample variogram calculation may be useful in other situations as a way to complement the traditional structural analysis.

Acknowledgements

The authors would like to acknowledge the Chair in Ore Reserve Evaluation at the University of Chile and the sponsoring by Codelco-Chile for supporting this research.

Appendix A

The polynomial fitting of the inverse transformation function (\square^{-1}) usually resorts to the normalized Hermite polynomials $\{H_p, p \in \mathbb{N}\}$, which are defined by Rodrigues' formula (Chilès and Delfiner, 1999, p. 640):

$$\forall p \in \mathbb{N}, \forall y \in \mathbb{R}, H_p(y) = \frac{1}{\sqrt{p!} g(y)} \frac{d^p g(y)}{dy^p} \quad (A1)$$

where $g(.)$ is the standard gaussian pdf. The inverse transformation function can be expanded as follows (Chilès and Delfiner, 1999, p. 395):

$$\psi^{-1}(y) = \sum_{p=0}^{+\infty} \phi_p H_p(y) \text{ with } \forall p \in \mathbb{N}, \phi_p = \int_{-\infty}^{+\infty} \psi^{-1}(y) H_p(y) g(y) dy \quad (A2)$$

Suppose that \square^{-1} is modeled by interpolating the experimental points $\{(Y(x_\square), Z(x_\square)), \square = 1 \dots n\}$, using line segments, monomials and/or exponential functions. Then, the coefficients $\{\phi_p, p \in \mathbb{N}\}$ in Eq. (A2) can be computed once the following quantities are known for any value of $p \in \mathbb{N}, q \in \mathbb{N}^*$ and $u, v \in \mathbb{R}$:

$$\begin{aligned} A(p, u, v) &= \int_u^v H_p(y) g(y) dy \\ B(p, q, u, v) &= \int_u^v y^q H_p(y) g(y) dy \\ C(p, \lambda, u, v) &= \int_u^v \exp(-\lambda y) H_p(y) g(y) dy \end{aligned} \quad (A3)$$

Using Rodrigues' formula [Eq. (A1)] and the fact that H_0 is constant and equal to 1:

$$A(0, u, v) = G(v) - G(u) \text{ and } \forall p > 0, A(p, u, v) = \frac{1}{\sqrt{p}} [H_{p-1}(v)g(v) - H_{p-1}(u)g(u)] \quad (\text{A4})$$

Similarly, one finds

$$\begin{aligned} B(0, 1, u, v) &= g(u) - g(v) \\ \forall p > 0, B(p, 1, u, v) &= \frac{1}{\sqrt{p}} \{ [y H_{p-1}(y) g(y)]_u^v - \int_u^v H_{p-1}(y) g(y) dy \} \\ &= \frac{1}{\sqrt{p}} [v H_{p-1}(v) g(v) - u H_{p-1}(u) g(u) - A(p-1, u, v)] \end{aligned} \quad (\text{A5})$$

The values of $B(p, q, u, v)$ for $q \geq 2$ can be calculated recursively:

$$\begin{aligned} B(0, q, u, v) &= u^{q-1} g(u) - v^{q-1} g(v) + (q-1) B(0, q-2, u, v) \\ \forall p \geq 1, B(p, q, u, v) &= \frac{1}{\sqrt{p}} \{ [y^q H_{p-1}(y) g(y)]_u^v - q \int_u^v y^{q-1} H_{p-1}(y) g(y) dy \} \\ &= \frac{1}{\sqrt{p}} [v^q H_{p-1}(v) g(v) - u^q H_{p-1}(u) g(u) - q B(p-1, q-1, u, v)] \end{aligned} \quad (\text{A6})$$

Finally, the expression of $C(p, \lambda, u, v)$ requires expanding the exponential function into a series of monomials:

$$C(p, \lambda, u, v) = \sum_{q=0}^{+\infty} \frac{(-\lambda)^q}{q!} \int_u^v y^q H_p(y) g(y) dy = \sum_{q=0}^{+\infty} \frac{(-\lambda)^q}{q!} B(p, q, u, v) \quad (\text{A7})$$

An approximate value is obtained by truncating the expansion to a high order.

Appendix B

This appendix aims at interpreting the variogram estimator in Eq. (5) in terms of disjunctive kriging. To do so, let us define the increment function associated with lag \mathbf{h} :

$$\forall \mathbf{x} \in \mathbb{R}^d, D_{\mathbf{h}}(\mathbf{x}) = Y(\mathbf{x} + \mathbf{h}) - Y(\mathbf{x}) \quad (\text{B1})$$

Since $\{Y(\mathbf{x}), \mathbf{x} \in \mathbb{R}^d\}$ is a multigaussian random field, so is $\{D_{\mathbf{h}}(\mathbf{x}), \mathbf{x} \in \mathbb{R}^d\}$ for any value of \mathbf{h} . The variogram $\gamma(\mathbf{h})$ is the expected value of half the square of $D_{\mathbf{h}}(\mathbf{x})$, which can be expressed as follows:

$$\frac{1}{2} D_{\mathbf{h}}(\mathbf{x})^2 = a_{\mathbf{h}} + b_{\mathbf{h}} H_2[c_{\mathbf{h}} D_{\mathbf{h}}(\mathbf{x})] \quad (\text{B2})$$

where

- H_2 is the normalized Hermite polynomial of the second degree;
- $a_{\mathbf{h}} = \gamma(\mathbf{h})$, $b_{\mathbf{h}} = \gamma(\mathbf{h})\sqrt{2}$ and $c_{\mathbf{h}} = [2\gamma(\mathbf{h})]^{-1/2}$
- $\{c_{\mathbf{h}} D_{\mathbf{h}}(\mathbf{x}), \mathbf{x} \in \mathbb{R}^d\}$ is a standardized multigaussian field.

With the same notations, the estimator [Eq. (5)] becomes

$$\hat{\gamma}(\mathbf{h}) = \frac{1}{2} \sum_{i=1}^{N(\mathbf{h})} \omega_i D_{\mathbf{h}_i}(\mathbf{x}_i)^2 = \sum_{i=1}^{N(\mathbf{h})} \omega_i \{a_{\mathbf{h}_i} + b_{\mathbf{h}_i} H_2[c_{\mathbf{h}_i} D_{\mathbf{h}_i}(\mathbf{x}_i)]\} \quad (\text{B3})$$

Equation (B3) is nothing but an estimator of the expected value of the Hermite polynomial of a gaussian field by a weighted average of this polynomial applied to correlated gaussian data. In other words, system (11) can be seen as a disjunctive kriging system, except that an unbias condition is added (Chilès and Delfiner, 1999, p. 417). In this respect, the weighted variogram appears as a transposition to a two-point statistics of the “disjunctive-type” estimator proposed by Matheron (1977, p. 2) to assess a marginal distribution.

References

- Bogaert P** (1999) On the optimal estimation of the cumulative distribution function in presence of spatial dependence. *Math Geol* 31(2): 213-239
- Bourgault G** (1997) Spatial declustering weights. *Math Geol* 29(2): 277-290
- Chilès JP, Delfiner P** (1999) *Geostatistics: Modeling spatial uncertainty*. Wiley, New York, pp 696
- Freulon X** (1992) Conditionnement du modèle gaussien par des inégalités ou des randomisées, Doctoral thesis, Centre de Géostatistique, Ecole des Mines de Paris, Fontainebleau, pp 168
- Freulon X, de Fouquet C** (1993) Conditioning a gaussian model with inequalities. In: Soares A (ed) *Geostatistics Tróia'92*. Kluwer Academic, Dordrecht, pp 201-212

- Geman S, Geman D** (1984) Stochastic relaxation, Gibbs distribution and the Bayesian restoration of images. *IEEE Transac Pattern Anal Mach Intell* 6: 721-741
- Goovaerts P** (1997) Geostatistics for natural resources evaluation. Oxford University Press, New York, pp 480
- Journel AG** (1983) Nonparametric estimation of spatial distributions. *Math Geol* 15(3): 445-468
- Matheron G** (1965) Les variables régionalisées et leur estimation. Masson, Paris, pp 305
- Matheron G** (1971) The theory of regionalized variables and its applications. Les cahiers du Centre de Morphologie Mathématique de Fontainebleau, Fascicule 5, Ecole des Mines de Paris, Fontainebleau, pp 212
- Matheron G** (1977) Peut-on imposer des conditions d'universalité au krigeage disjonctif ?, internal report N-539, Centre de Géostatistique, Ecole des Mines de Paris, Fontainebleau, pp 24
- Omre H** (1984) The variogram and its estimation. In: Verly G, David M, Journel AG, Maréchal A (eds) *Geostatistics for Natural Resources Characterization*. Reidel, Dordrecht, pp 107-125
- Pardo-Igúzquiza E, Dowd PA** (2001) The variance-covariance matrix of the experimental variogram: assessing variogram uncertainty. *Math. Geol* 33(4): 397-419
- Richmond A** (2002) Two-point declustering for weighting data pairs in experimental variogram calculation. *Computers & Geosciences* 28(2): 231-241
- Rivoirard J** (2001) Weighted variograms. In: Kleingeld WJ, Krige DG (eds) *Proceedings of the 6th International Geostatistics Congress*. Cape Town, South Africa, pp 145-155
- Switzer P** (1977) Estimation of spatial distributions from point sources with application to air pollution measurement. *Bulletin of the International Statistical Institute* 47(2): 123-137
- Verly G** (1986) Multigaussian kriging - a complete case study. In: Ramani RV (ed) *Proceedings of the 19th APCOM International Symposium*. Society of Mining Engineers, Littleton, Colorado, pp 283-298