Geostatistics - Uncertainty quantification: the MultiGaussian framework

Julián M. Ortiz - https://julianmortiz.com/

August 31, 2024

Summary

We have already discussed the kriging approach to determine the best linear unbiased estimate at every location. The kriging estimate has an associated kriging variance which measures the variance of the error between the estimate and the true value. However, this measure of uncertainty does not depend on the actual value of the variable, but only on the spatial configuration of the sample data used in the estimation. Thus, it does not capture the "proportional effect", that is, the fact that the variability depends on the local value of the variable.

In this chapter, we introduce the multiGaussian framework, which allows characterizing the uncertainty in a simple, yet effective manner.

As most variables are not normally distributed, the approach requires a quantile transformation and a back transformation at the end of the analysis. We show that the simple kriging estimate and its corresponding variance identify the conditional expectation and variance in a multiGaussian framework. Thus, uncertainty quantification is simply achieved by transforming the data to a Gaussian distribution, performing simple kriging of the normally transformed values and back transforming the resulting distribution to original units.

1 Introduction

The random function model provides the means for inference. We can assume properties (statistical properties, that is) of the variable at every location, and from there, derive an estimate that complies with some imposed properties, such as unbiasedness and optimality.

Once the estimate is obtained, the next question arises: **what is the uncertainty around that estimate?**

The kriging estimate provides the best unbiased linear estimator and it comes with an associated kriging variance. Let us write the estimate and estimation variance in simple and ordinary kriging:

$$Z^*_{SK}(\mathbf{u}_0) = \left(1 - \sum_{i=1}^{n} \lambda_i^{SK}\right) m + \sum_{i=1}^{n} \lambda_i^{SK} Z(\mathbf{u}_i) \quad (1)$$

$$\sigma^2_{SK}(\mathbf{u}_0) = \sigma_0^2 - \sum_{i=1}^{n} \lambda_i^{SK} C_{i0} \quad (2)$$

$$Z^*_{OK}(\mathbf{u}_0) = \sum_{i=1}^{n} \lambda_i^{OK} Z(\mathbf{u}_i) \quad (3)$$

$$\sigma^2_{OK}(\mathbf{u}_0) = \sigma_0^2 - \sum_{i=1}^{n} \lambda_i^{OK} C_{i0} - \mu \quad (4)$$

Both expressions for the variances (**Equations 2** and **4**), do not depend on the Z values. The variances only depend on the spatial arrangement of the samples with respect to the location estimated through the covariance. They also depend on the prior variance of the random variable σ_0^2.

Variables showing a lognormal distribution have what is called a "proportional effect". This is, a variability that depends on the actual values. Typically low values are less variable (hence easier to estimate) than high values. The variance is therefore a function of the value at the estimation location.

The kriging variance is therefore a poor tool to infer the uncertainty expected around the estimate.

We introduce next the multiGaussian model. Under this model the kriging variance is an appropriate measure of uncertainty, and this can be useful to assess uncertainty in our case because using a transformation of the data, we can re-

cover an uncertainty that accounts for the proportional effect.

2 The multiGaussian distribution

2.1 Univariate case

Let us start by recalling the univariate Gaussian distribution. We say a random variable Y (upper case) follows a Gaussian (or normal) distribution, if its probability density function (pdf) is:

$$f_Y(y) = \frac{1}{\sqrt{2\pi\sigma^2}} exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right) \tag{5}$$

Notice that the density is defined for each value $Y = y$. The probability density function depends only on two parameters: the mean μ and the variance σ^2. We note:

$$Y \sim \mathcal{N}(\mu, \sigma^2) \tag{6}$$

We say that "Y follows a Gaussian or normal distribution with mean μ and variance σ^2".

The variable can be standardized by subtracting the mean and dividing it by the standard deviation, leading to a standard Gaussian variable:

$$Y' = \frac{Y-\mu}{\sigma} \tag{7}$$

$$Y' \sim \mathcal{N}(0, 1) \tag{8}$$

2.2 Bivariate case

Now, we extend this definition to the bivariate case. The bivariate Gaussian pdf for two random variables Y_1 and Y_2 is:

$$f_{Y_1Y_2}(y_1, y_2) = \frac{1}{2\pi\sigma_{Y_1}\sigma_{Y_2}\sqrt{1-\rho^2_{Y_1Y_2}}} exp\left(-\frac{1}{2(1-\rho^2_{Y_1Y_2})}\left\{\frac{(y_1-\mu_{Y_1})^2}{\sigma^2_{Y_1}} + \frac{(y_2-\mu_{Y_2})^2}{\sigma^2_{Y_2}} - \frac{2\rho_{Y_1Y_2}(y_1-\mu_{Y_1})(y_2-\mu_{Y_2})}{\sigma_{Y_1}\sigma_{Y_2}}\right\}\right) \tag{9}$$

Again, the joint (bivariate) pdf is defined for all values taken by the two random variables $Y_1 = y_1$ and $Y_2 = y_2$.

And we note:

$$\boldsymbol{Y} = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} \sim \mathcal{N}_2\left[\begin{pmatrix} \mu_{Y_1} \\ \mu_{Y_2} \end{pmatrix}, \begin{pmatrix} \sigma^2_{Y_1} & C_{Y_1Y_2} \\ C_{Y_2Y_1} & \sigma^2_{Y_2} \end{pmatrix}\right] \tag{10}$$

In this case, we have the joint pdf of two Gaussian variables, each parameterized by their corresponding means and variances: $\mu_{Y_1}, \mu_{Y_2}, \sigma^2_{Y_1}, \sigma^2_{Y_2}$. The relationship between the two univariate Gaussian distributions is controlled by the correlation coefficient $\rho_{Y_1Y_2}$, which is related to the covariances as follows:

$$C_{Y_1Y_2} = C_{Y_2Y_1} = \rho_{Y_1Y_2}\sigma_{Y_1}\sigma_{Y_2} \tag{11}$$

Notice that if both variables are standard Gaussian, the joint distribution simplifies to:

$$\boldsymbol{Y'} = \begin{pmatrix} Y'_1 \\ Y'_2 \end{pmatrix} \sim \mathcal{N}_2 \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right] \tag{12}$$

where $\rho_{Y'_1 Y'_2} = \rho_{Y'_2 Y'_1} = \rho$ is the correlation coefficient between both variables. The joint pdf can be written as:

$$f_{Y'_1 Y'_2}(y_1, y_2) = \frac{exp\left(-\frac{1}{2(1-\rho^2)}\left\{y_1^2 + y_2^2 - 2\rho y_1 y_2\right\}\right)}{2\pi\sqrt{1-\rho^2}} \tag{13}$$

2.3 Multivariate case

Finally, let us write the probability density function for a multivariate Gaussian case. This means we have n random variables with means and variances μ_{Y_i} and $\sigma^2_{Y_i}$. Furthermore, these random variables are correlated and their pairwise relationships are quantified by the correlation coefficients $\rho_{Y_i Y_j}$ with $i, j = 1, \ldots, n$:

$$f_{Y_1 \ldots Y_n}(y_1, \ldots, y_n) = \frac{exp\left(-\frac{1}{2}(\boldsymbol{y}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{y}-\boldsymbol{\mu})\right)}{\sqrt{(2\pi)^n |\boldsymbol{\Sigma}|}} \tag{14}$$

where we use bold characters for vector and matrix notation:

$$\boldsymbol{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \qquad \boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{pmatrix} \qquad \boldsymbol{\Sigma} = \begin{pmatrix} \sigma^2_{Y_1} & C_{Y_1Y_2} & \cdots & C_{Y_1Y_n} \\ C_{Y_2Y_1} & \sigma^2_{y_2} & \cdots & C_{Y_2Y_n} \\ \vdots & \vdots & \ddots & \vdots \\ C_{Y_nY_1} & C_{Y_nY_2} & \cdots & \sigma^2_{Y_n} \end{pmatrix}$$

Again, each covariance term in the variance-covariance matrix (or, simply, the covariance matrix) can be written as a function of the pairwise correlations and the standard deviations:

$$C_{Y_iY_j} = C_{Y_jY_i} = \rho_{Y_iY_j}\sigma_{Y_i}\sigma_{Y_j}$$

2.4 Conditional distribution

Multivariate case

When some of the arguments of a multivariate Gaussian distribution are known, we can compute the conditional distribution. The conditional distribution can be derived by first partitioning the multiGaussian vector $\boldsymbol{Y}$ into $\boldsymbol{Y_1}$ (the variables to be predicted) and $\boldsymbol{Y_2}$ (the variables known, which are used as conditioning), vectors of size n_1 and n_2, such that $n_1 + n_2 = n$. The corresponding means and covariance matrices are:

$$\boldsymbol{Y} = \begin{pmatrix} \boldsymbol{Y_1} \\ \boldsymbol{Y_2} \end{pmatrix} \qquad \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu_1} \\ \boldsymbol{\mu_2} \end{pmatrix} \qquad \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma_{11}} & \boldsymbol{\Sigma_{12}} \\ \boldsymbol{\Sigma_{21}} & \boldsymbol{\Sigma_{22}} \end{pmatrix} \tag{15}$$

One of the features of a multivariate Gaussian distribution is that any subset of its variables, conditioned to known

values of another subset of its variables, is still multivariate Gaussian. Therefore, we only need to know the corresponding means and covariance matrix to fully know the conditional distribution. When conditioning to $\boldsymbol{Y_2} = \boldsymbol{y_2}$, the means and covariance matrix of $\boldsymbol{Y_1}|\boldsymbol{Y_2} = \boldsymbol{y_2}$ (we say $\boldsymbol{Y_1}$ given $\boldsymbol{Y_2} = \boldsymbol{y_2}$) can be easily computed (by applying Bayes' law). The resulting conditional moments are (recall that the mean and variance are called moments of the distribution):

$$\begin{aligned} \boldsymbol{\mu_{1|2}} &= \boldsymbol{\mu_1} + \boldsymbol{\Sigma_{12}}\boldsymbol{\Sigma_{22}^{-1}}(\boldsymbol{y_2} - \boldsymbol{\mu_2}) \\ \boldsymbol{\Sigma_{11|2}} &= \boldsymbol{\Sigma_{11}} - \boldsymbol{\Sigma_{12}}\boldsymbol{\Sigma_{22}^{-1}}\boldsymbol{\Sigma_{21}} \end{aligned} \tag{16}$$

Bivariate case

In the bivariate case, these translate into:

$$\begin{aligned} \mu_{Y_1|Y_2=y_2} &= \mu_{Y_1} + \frac{C_{Y_1Y_2}}{\sigma^2_{Y_2}}(y_2 - \mu_{Y_2}) \\ \sigma^2_{Y_1|Y_2=y_2} &= \sigma^2_{Y_1} - \frac{C^2_{Y_1Y_2}}{\sigma^2_{Y_2}} \end{aligned} \tag{17}$$

This means that if the random variable Y_2 is known to take a specific value y_2, we can "update" the prior distribution of the random variable Y_1 and obtain what we call a posterior (or conditional) distribution. The conditional distribution is a univariate Gaussian distribution with parameters provided in **Equation 17**, or:

$$Y_1|Y_2 = y_2 \sim \mathcal{N}(\mu_{Y_1|Y_2=y_2}, \sigma^2_{Y_1|Y_2=y_2}) \tag{18}$$

We will now introduce the concept of multiGaussian kriging and see how it relates to the conditional expectation and conditional variance shown above.

3 MultiGaussian kriging

3.1 Introduction

It was mentioned before that there are many variants of kriging. They all build from the same principles, but depending on the assumptions, they yield different results.

One of such approaches is to work with a transform of the data. In this case, a quantile transformation is performed over the original distribution to convert it into a standard Gaussian distribution. Kriging is then performed over the transformed variable, also called normal scores. And this is where things become interesting. If simple kriging is used, the simple kriging estimate and the simple kriging variance (of the normally transformed values) is identical to the expression for the conditional expectation and conditional variance presented in **Equation 16**.

3.2 The conditional mean and variance in the multiGaussian case

We can rewrite **Equation 16** to find the conditional expectation and conditional variance of one of the Gaussian random variables (the “estimated variable”) conditioned to the remaining random variables (the “samples”).

Let us write the vector of random variables and their means as:

$$\boldsymbol{Y} = \begin{pmatrix} Y_0 \\ \hline \boldsymbol{Y_1} \end{pmatrix} = \begin{pmatrix} Y_0 \\ \hline Y_1 \\ \vdots \\ Y_n \end{pmatrix} \qquad \boldsymbol{\mu} = \begin{pmatrix} \mu_0 \\ \hline \boldsymbol{\mu_1} \end{pmatrix} = \begin{pmatrix} \mu_0 \\ \hline \mu_1 \\ \vdots \\ \mu_n \end{pmatrix} \tag{19}$$

And the variance-covariance matrix as:

$$\boldsymbol{\Sigma} = \left[\begin{array}{c|c} \sigma_0^2 & \boldsymbol{\Sigma_{01}} \\ \hline \boldsymbol{\Sigma_{10}} & \boldsymbol{\Sigma_{11}} \end{array}\right] = \left[\begin{array}{c|ccc} \sigma_0^2 & \sigma_{01} & \cdots & \sigma_{0n} \\ \hline \sigma_{10} & \sigma_1^2 & \cdots & \sigma_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n0} & \sigma_{n1} & \cdots & \sigma_n^2 \end{array}\right] \tag{20}$$

We can recognize the matrix $\boldsymbol{\Sigma_{11}}$ as the matrix of covariances between the conditioning variables, and the vector $\boldsymbol{\Sigma_{01}}$ as the vector of covariances between the n conditioning variables and the variable we are trying to estimate Y_0.

Recall the regression expression for the conditional mean and variance in **Equation 16**, now written in terms of $\boldsymbol{Y_0}$ and $\boldsymbol{Y_1}$:

$$\boldsymbol{\mu_{0|1}} = \boldsymbol{\mu_0} + \boldsymbol{\Sigma_{01}}\boldsymbol{\Sigma_{11}^{-1}}(\boldsymbol{y_1} - \boldsymbol{\mu_1})$$
$$\boldsymbol{\Sigma_{00|1}} = \boldsymbol{\Sigma_{00}} - \boldsymbol{\Sigma_{01}}\boldsymbol{\Sigma_{11}^{-1}}\boldsymbol{\Sigma_{10}}$$

And assume that the mean for all the random variables is m, per the stationarity assumption. We can rewrite the conditional mean and variance as:

$$Y^*_{0|1} = m + \boldsymbol{\Sigma_{01}}\boldsymbol{\Sigma_{11}^{-1}}(\boldsymbol{Y_1} - m) \tag{21}$$
$$\sigma^2_{0|1} = \sigma_0^2 - \boldsymbol{\Sigma_{01}}\boldsymbol{\Sigma_{11}^{-1}}\boldsymbol{\Sigma_{10}} \tag{22}$$

Equation 21 is exactly the same as the simple kriging estimate, and **Equation 22** is the simple kriging variance. The vector $\mathbf{\Sigma_{01}\Sigma_{11}^{-1}}$ is no other than the vector $\mathbf{\Gamma}$ of simple kriging weights $\lambda_i, i = 1, \ldots, n$:

$$Y^*_{0|1} = \mathbf{\Gamma Y_1} + (1 - \mathbf{\Gamma})m \tag{23}$$

$$\sigma^2_{0|1} = \sigma^2_0 - \mathbf{\Gamma \Sigma_{10}} \tag{24}$$

In summary, the conditional mean and variance in the multivariate multiGaussian regression coincide with the simple kriging estimate and simple kriging variance.

3.3 Practical implementation

What does it mean for a random function Y distributed in space over a domain D to be multiGaussian?

This condition requires that each random variable $Y(\boldsymbol{u_i}) \in D$ is univariate Gaussian, and that the joint distribution between variables $Y(\boldsymbol{u_i})$ and $Y(\boldsymbol{u_j})$ for all $i, j = 1, \ldots, n$ be bivariate Gaussian, and between variables $Y(\boldsymbol{u_i})$, $Y(\boldsymbol{u_j})$, and $Y(\boldsymbol{u_k})$, for all $i, j, k = 1, \ldots, n$ be trivariate Gaussian, and, in general, that any combination of n random variables follow a n-variate multiGaussian distribution. Under the assumption of stationarity, this is the equivalent to say that the bivariate relationships between $Y(\boldsymbol{u})$ and $Y(\boldsymbol{u} + \boldsymbol{h})$ must be biGaussian and, in general, between $Y(\boldsymbol{u})$, $Y(\boldsymbol{u} + \boldsymbol{h_1})$, ..., $Y(\boldsymbol{u} + \boldsymbol{h_k})$ must be a $(k + 1)$ multiGaussian distribution. Notice that the relationships are no longer a function of location, but depend on the separation between locations.

In practice, the variable Z is not distributed as a multi-Gaussian variable. Therefore, we will need to transform it. There is not a direct way to do this (that is, to transform the joint distribution of multiple random variables, each at a different location, into a multiGaussian distribution), so we proceed by transforming the univariate histogram into a univariate Gaussian distribution. To keep things simple, we use the standard Gaussian distribution (which has a mean of 0 and a variance of 1). Then, the high-order distribution is usually assumed and sometimes checked to verify the multiGaussian assumption is reasonable.

Gaussian transformation

The Gaussian transformation or normal score transformation is a quantile transformation that assigns to any value in an original distribution Z, a corresponding Gaussian value or normal score. The resulting distribution of normal scores follows a standard normal distribution:

$$Z = \varphi(Y) \tag{25}$$
$$Y = \varphi^{-1}(Z) \tag{26}$$
$$Y \sim \mathcal{N}(0, 1) \tag{27}$$

Notice that the transformation function φ links the cumulative distribution functions of Z and Y. If Z follows a distribution $F(z)$, and we call the standard Gaussian cdf $G(y)$, then $\varphi = F^{-1} \circ G$.

In order to implement this Gaussian transformation, we need the representative distribution of Z, therefore, declustering weights may need to be accounted for, to ensure the

distribution is corrected for spatial bias. The process is illustrated in **Figures 1** and **2**.

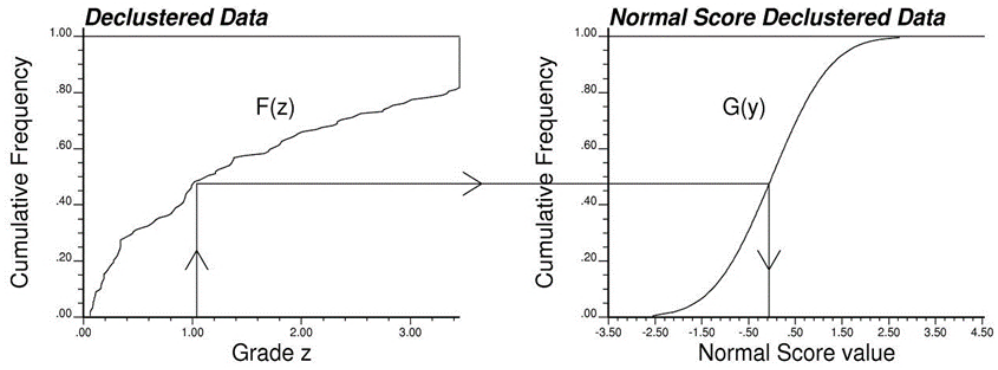

Figure 1: The graphical representation of the normal score transformation. Corresponding quantiles of the original declustered distribution and from a standard Gaussian distribution are matched.

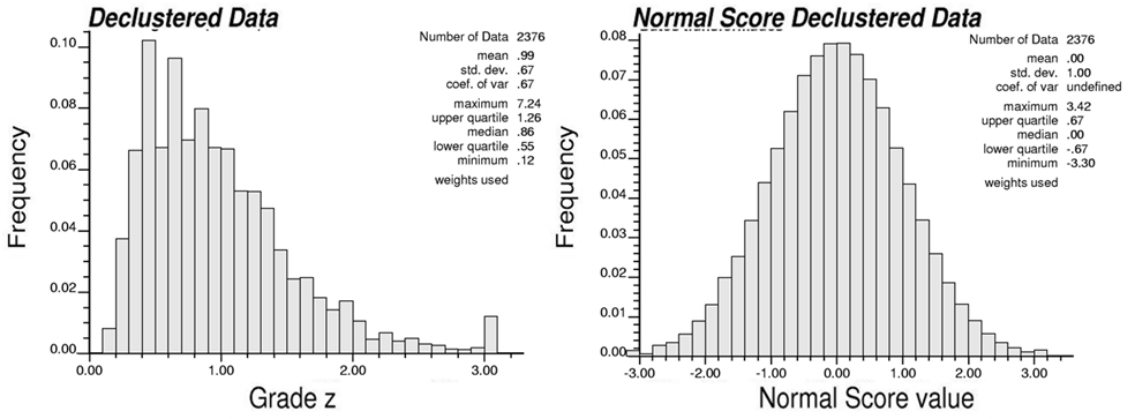

Figure 2: The original and transformed histograms.

Back-transformation

One of the nice properties of the normal score transformation is that results of operations over Gaussian values can be back-transformed into original units, by simply inverting the transformation:

$$Z^r = \varphi(Y^r) \tag{28}$$

where the superscript ^{r} indicates the result of a mathematical operation of Gaussian values. Furthermore, the transformation is a bijection, that is, a one-to-one correspondence exists between the two distributions. Each value has a unique transformed value, hence one can go back and forth from one variable to the other.

MultiGaussian kriging to determine local conditional distributions

Now, let us put everything together.

- We have a variable Z over a domain D. A set of samples $\{z(\boldsymbol{u_i}), i = 1, \ldots, N\}$ are available within the domain.
- Through a declustering technique, we determine that these samples must be weighted to compensate for clusters and spatial bias. These weights $\{w(\boldsymbol{u_i}), i = 1, \ldots, N\}$ are assigned to the corresponding samples.
- The cdf of the variable Z is built by sorting the values $\{z(\boldsymbol{u_i}), i = 1, \ldots, N\}$ from low to high, obtaining the sorted set $\{z^{(I)}(\boldsymbol{u_i}), i = 1, \ldots, N, I = 1, \ldots, N\}$, and assigning the corresponding weights $\{w(\boldsymbol{u_i}), i = 1, \ldots, N\}$. Interpolation between sample values and extrapolation

beyond the minimum $z^{(1)}$ and maximum $z^{(N)}$ are necessary to complete the experimental cumulative distribution. For each sample value $z^{(I)}(\boldsymbol{u_i})$, the corresponding cumulative probability can be computed as:

$$F\left(z^{(I)}(\boldsymbol{u_i})\right) = \sum_{j=1}^{I} w^{(j)}$$

where $w^{(j)}$ corresponds to the weight of the sample $z^{(j)}(\boldsymbol{u_i})$.

- For each sample $\{z(\boldsymbol{u_i}), i = 1, \ldots, N\}$, the cumulative probability is computed and the corresponding Gaussian score is assigned as:

$$y(\boldsymbol{u_i}) = G^{-1}\left(F(z(\boldsymbol{u_i}))\right)$$

  The histogram of Y (accounting for declustering weights, since the normal scores may also be clustered spatially) is Gaussian with mean 0 and variance 1.
- Perform a variogram analysis of the normal score data, to obtain a variogram model of the normal scores, in 3D.
- Perform **simple kriging** of the normal scores to estimate the normal score at unsampled locations:
  - Search for samples in a local neighborhood.
  - Select n samples according to the constraints imposed by the kriging plan.
  - Compute the normal score covariances between samples and between the samples and the estimation location.

- Solve the simple kriging system of equations to obtain the weights.

$$\left(\lambda_{\mathbf{Y}}^{\mathbf{SK}}\right) = \left[\mathbf{C}_Y^{-1}\right]\left(\mathbf{k}_Y\right)$$

- The resulting kriging estimate and variance are (recall that the global mean $m = 0$ and the variance is $\sigma_0^2 = 1$):

$$Y_{SK}^*(\mathbf{u}_0) = \sum_{i=1}^{n} \lambda_{Y,i}^{SK} Y(\mathbf{u}_i) \tag{29}$$

$$\sigma_{Y,SK}^2(\mathbf{u}_0) = 1 - \sum_{i=1}^{n} \lambda_{Y,i}^{SK} C_{Y,i0} \tag{30}$$

- Since the distribution of Y is assumed to be multiGaussian, then these expressions provide the conditional expectation and conditional variance of the distribution of $Y(\boldsymbol{u_0})$, and we know the local distribution has to be Gaussian in shape. Then, we can back transform any quantile of the distribution to original units. We back transform the entire local conditional distribution, as shown in **Figure 3**.
- Finally, from the numerically obtained local distribution in original units, we can compute any desired statistics (such as the mean and variance) by numerical integration.

4 Final comments

One common mistake is to back-transform the kriging estimate of the normal scores and think that it represents the

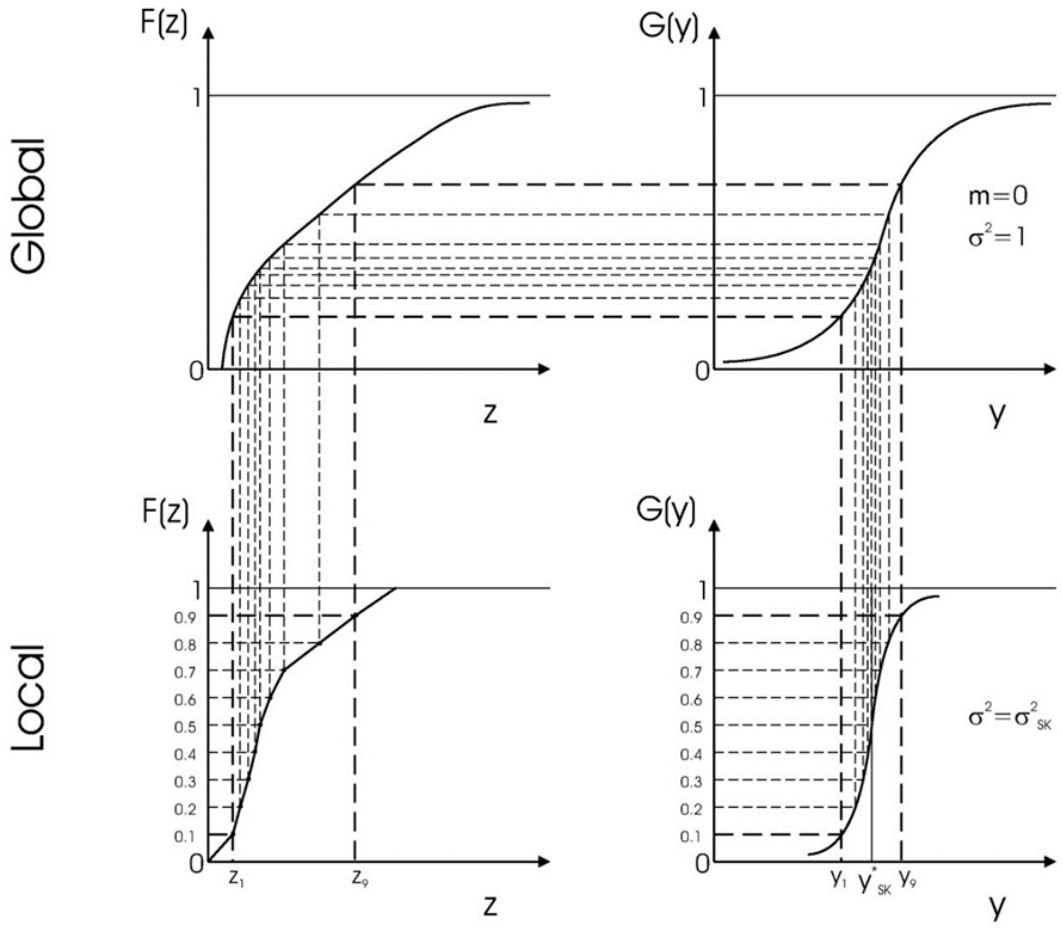

Figure 3: The back-transformation from the conditional Gaussian distribution to the original units. Here, only the 9 deciles are back-transformed, but we could use as many quantiles as needed.

estimate in original units. This is wrong! this estimate is the 50^{th} percentile of the local distribution, that is, it represents the median. However if the local distribution is not perfectly symmetric, the median is different than the mean.

The mean (the expected value) of the Z variable in original units must be obtained by numerical integration of the local conditional distribution. One way of doing this is drawing a large number of uniform values in [0,1] and back transforming each as shown in **Figure 3**. The average of these back-transformed values is an estimate of the mean of the local distribution in original units.

MultiGaussian kriging is an elegant approach to infer local uncertainty, however, it only does so at point support. We cannot do block kriging of normal scores, because the transformation function is not linear, therefore, the average normal scores do not match the average original values.

Any change of support must be done over the original variable. Notice that, if the variable is additive, then we can simply average the point estimates obtained over a discretization of the block we are trying to estimate. However, computing the variance at block support requires knowing the covariances between the blocks.

All the estimation with multiGaussian kriging makes a strong assumption of stationarity. As soon as we transform the data, we are "locking" the mean and variance of the global distribution. The conditional expectation and variance of a multiGaussian variable coincides with the simple kriging estimate and variance. If ordinary kriging is used, there is no guarantee of the accuracy of the prediction of the conditional mean and variance in the multiGaussian context. This leads to an inflated variance that may cause problems when back-transformed. In theory, only simple kriging is allowed.

MultiGaussian kriging is not commonly applied (although some academic and industrial applications exist). However, as we will see, it is the basis of conditional simulation.

Index