Note on Working with Trends in Geostatistics

John G. Manchuk and Clayton V. Deutsch

Non-stationary approaches in Geostatistics are being utilized more frequently. Considering a trend model for the mean has been used for some time. Working with trends can be done in several ways, this paper focuses on some of the explicit techniques including the traditional use of residuals, locally varying mean, and a conditional transformation approach. Some discussion on the presence of bias when using residuals is provided and a relationship between the variogram of residuals and the variogram of the original variable and trend function is developed. The result indicates two important factors for the variogram of residuals: 1 – the nugget effect of the residual variogram is the same as that of the original variable; 2 – the short range structure of the residuals variogram is the same as well.

Introduction

A common limitation of geostatistical methods is their dependence on a decision of stationarity over a finite domain of interest. In general, the assumption of stationarity indicates a constant mean, homoscedastic variance, and fixed distribution function within the domain. This case is rarely encountered with real data. It is common to segregate the domain into different zones that could be defined based on geology, lithology, texture, or other identifiers, with each zone honouring the assumption of stationarity. If such segregation is possible, statistics such as the mean must vary in a discontinuous fashion between zones. In cases where statistics vary continuously, segregation does not aid in the removal of non-stationary features and other techniques are needed. Trend modeling can be used to address situations where non-stationarity of the mean and variance is suspected. In this work, use of the word trend is interchangeable with drift.

Two approaches for trend modeling exist: implicit and explicit techniques. With implicit techniques, only the functional form of a trend is specified, rather than the trend itself. These techniques focus on trends that represent the mean of the random function. The variance may still exhibit heteroscedastic character. Ordinary kriging (Matheron, 1963; Journel and Huijbregts, 1978; Isaaks and Srivastava, 1989) assumes an unknown mean that is a zero order polynomial, that is, a constant dependent on spatial location. More general polynomial forms are possible with universal kriging (Huijbregts and Matheron, 1971), also referred to as kriging with a trend (Journel and Rossi, 1989). Another class of implicit approach are k-order intrinsic random functions (IRF-k) (Matheron, 1963; Chiles and Delfiner, 1999); however, they are restricted to a specific form of the trend function. The use of generalized covariance functions makes IRF-k uncommon in practice.

Explicit techniques model the suspected trend directly. Often, the trend is removed from the original data and we are left with a set of residuals. A known issue with this approach is the presence of a bias in the covariance matrix of the residuals regardless of the form of the trend function (Cressie, 1993; Kitanidis, 1993; Beckers and Bogaert, 1998). The bias always leads to an underestimation of the variance. Another issue is that there is no control on the bounds of the predicted random function when the residuals are used. It is possible to predict negative values for strictly non-negative random functions, or to predict fractions less than zero or greater than one for random functions with units of concentration or percent. An approach to alleviate this issue was proposed by Leuangthong and Deutsch (2004) and involved transforming the random function conditional to the trend via a stepwise conditional transform (Rosenblatt, 1952). Another explicit technique is to utilize the trend as a locally varying mean (LVM), although the LVM approach to first-order non-stationary applications is typically not associated with geostatistical modeling with a trend.

This paper focuses primarily on explicit techniques for incorporating a trend into models. First, the concept of bias discussed by Cressie (1993), Kitanidis (1993) and Beckers and Bogaert (1998) is revisited to identify a few potential correlations that may be overlooked as well as identify the case where the bias does not exist. Another form of bias explored by Starks and Fang (1982) that is caused by underlying trends is also covered and a relation between the variogram of the original variable, trend, and residuals is developed. A few of the approaches to incorporate trends into Gaussian simulation including working with residuals, locally varying mean, and conditional transformation (Gooverts, 1997; Journel and Rossi, 1989; Gringarten, Deutsch, 2001; Leuangthong and Deutsch, 2004) are covered, particularly to identify the correct variogram to use.

Background

The idea that a bias exists when making predictions using the residuals obtained from explicit trend models is revisited. In papers where the bias term is formulated (Kitanidis, 1993; Beckers and Bogaert, 1998), equations for the bias term are developed for generalized regression or ordinary least squares; however, an issue is that the typical assumptions involved in application of either of these models do not apply to the modeling of trends for spatial geostatistical problems. A few of the assumptions from linear regression that are reasonable include:

1. The model is linear: we must conclude that if linear regression is chosen to model a trend, the data have indicated that the response variable can be represented as a linear combination of the spatial coordinates as well as any other exhaustive spatial data.
2. The expected value of the residuals is zero if we assume the response variable is first order stationary after removing the trend.
3. Homoscedasticity: this assumption is viable for normally distributed variables; however, cases such as lognormal variables will exhibit heteroscedasticity in the residuals.
4. Normality of residuals: the distribution of residuals is usually not known, but we often assume it is normal, or we transform the residuals to follow a normal distribution.

Assumptions that do not hold relate to the covariance of the response variable, predictors, and residuals. The following assumptions do not necessarily hold:

5. Independent errors: one test to indicate if a regression model is appropriate is to check the residuals for any correlated behaviour. Application of regression for trends in geostatistics rarely has this property because we design trends to remove only part of the spatial dependence of the response variable. Short range spatial correlation is not removed and residuals are not independent.
6. Independence of residuals and predictors: trends modeled using regression will result in a correlation of zero, but the variables will not be independent due to 5.
7. Independence of residuals and response: since some spatial correlation is retained in the residuals, the response and residuals must be correlated at least up to the correlation length of the residuals.

Considering that there is correlation between residuals and predictors as well as with the response, the - equations that develop the bias are revisited starting with the representation of the random function decomposed into deterministic and stochastic components. Equation (1) expresses a random function, $z(\mathbf{u})$, as the sum of a mean, $X\beta$, and random function (residuals), $\varepsilon(\mathbf{u})$, where X are design variables, β are weights, and $\mathbf{u}$ are spatial coordinates.

$$z(\mathbf{u}) = X\beta + \varepsilon(\mathbf{u}) \tag{1}$$

It is common in geostatistics for X to contains polynomial functions of $\mathbf{u}$. The bias developed in Beckers and Bogaert (1998) suggests that the covariance of the residuals, Σ_ε, is always underestimated, even when the true form of $X\beta$ or Σ_ε is known. Assuming the trend component is unknown and predicted using ordinary least squares, the residuals are expressed as Equation (2), with $\hat{\beta}$ estimated from Equation (3).

$$\varepsilon(\mathbf{u}) = z(\mathbf{u}) - X\hat{\beta} \tag{2}$$

$$\hat{\beta} = \left(X'X\right)^{-1} X'z(\mathbf{u}) \tag{3}$$

Rearranging (2) the residuals are expressed as a function of the projection matrix, P:

$$\varepsilon(\mathbf{u}) = \left(I - X\left(X'X\right)^{-1} X'\right) z(\mathbf{u}) = \left(I - P\right) z(\mathbf{u}) \tag{4}$$

Evaluating the variance of the residuals based on (4) reveals the bias term:

$$\begin{aligned} Var\left\{\varepsilon(\mathbf{u})\right\} &= E\left\{\left(I - P\right) z(\mathbf{u}) z'(\mathbf{u}) \left(I - P\right)'\right\} \\ &= \left(I - P\right)\Sigma_\varepsilon \left(I - P\right)' \\ &= \Sigma_\varepsilon - \left(\Sigma_\varepsilon P' + P\Sigma_\varepsilon - P\Sigma_\varepsilon P'\right) \\ &= \Sigma_\varepsilon - B \neq \Sigma_\varepsilon \end{aligned} \tag{5}$$

The strong assumption that the covariance of $z(\mathbf{u})$ is equal to the covariance of $\varepsilon(\mathbf{u})$ must be made to obtain this result; however, this is unreasonable. Expanding the variance of $z(u)$ we obtain Equation (6).

$$\begin{aligned} Var\left\{z(\mathbf{u})\right\} &= E\left\{\left(X\beta + \varepsilon(\mathbf{u})\right)\left(X\beta + \varepsilon(\mathbf{u})\right)'\right\} - E\left\{X\beta + \varepsilon(\mathbf{u})\right\}^2 \\ &= E\left\{X\beta\beta'X' + X\beta\varepsilon'(\mathbf{u}) + \varepsilon(\mathbf{u})\beta'X' + \varepsilon(\mathbf{u})\varepsilon'(\mathbf{u})\right\} - E\left\{X\beta\beta'X'\right\} \\ &= E\left\{X\beta\varepsilon'(\mathbf{u})\right\} + E\left\{\varepsilon(\mathbf{u})\beta'X'\right\} + \Sigma_\varepsilon \end{aligned} \tag{6}$$

If and only if $X\beta$ and $\varepsilon(\mathbf{u})$ are independent can we obtain the result:

$$Var\left\{z(\mathbf{u})\right\} = Var\left\{\varepsilon(\mathbf{u})\right\} = \Sigma_\varepsilon \tag{7}$$

Independence is only obtained if the residuals are independent (assumption 5), that is, if the spatial correlation is zero; however, when modeling trends we usually obtain residuals that retain structure. An interesting case to consider is when the trend is a constant equal to zero, or $X\beta = 0$. Initially, it would seem the bias still exists based on Equation (5), but this is not the case due to the idempotent structure of $(I - P)$ and knowing that $\varepsilon(\mathbf{u}) = z(\mathbf{u})$ when $X\beta = 0$:

$$\begin{aligned} Var\left\{\varepsilon(\mathbf{u})\right\} &= E\left\{\left(I - P\right) z(\mathbf{u}) z'(\mathbf{u}) \left(I - P\right)'\right\} \\ &= E\left\{z(\mathbf{u}) z'(\mathbf{u})\right\} \qquad \text{since } \left(I - P\right) z(\mathbf{u}) = z(\mathbf{u}) \\ &= \Sigma_\varepsilon \end{aligned} \tag{8}$$

The presence of a trend in a random function leads to a biased estimate of the variogram because it influences the squared differences calculated for a particular lag (Starks and Fang, 1982). Typically, the bias is observed in the experimental variogram as a significant deviation above the sill. Observation of such a feature may indicate that trend analysis is warranted. The form of the bias is essentially the variogram of the trend itself; however, it is more complex when there is correlation between the original variable and the trend.

When a trend exists, it is possible to formulate the variogram of the residuals, γ_ε, as a function of the variogram of the original variable, γ_z, and of the trend, γ_m, where $X\beta = m$. Using Equation (2) for the residuals and evaluating for the variogram we obtain the following expression

$$\begin{aligned}\gamma_\varepsilon &= E\left\{\left(\varepsilon - \varepsilon_{\mathbf{h}}\right)^2\right\} \\ &= E\left\{\left((z-m)-(z_{\mathbf{h}}-m_{\mathbf{h}})\right)^2\right\} \\ &= E\left\{z^2 - 2zm + m^2 + z_{\mathbf{h}}^2 - 2z_{\mathbf{h}}m_{\mathbf{h}} + m_{\mathbf{h}}^2\right. \\ &\quad \left. -2zz_{\mathbf{h}} + 2zm_{\mathbf{h}} + 2z_{\mathbf{h}}m - 2mm_{\mathbf{h}}\right\}\end{aligned}$$

where the following are implied: $x = x_u$ and $x_{\mathbf{h}} = x_{\mathbf{u+h}}$with $x = \varepsilon, m, z$. Rearranging the terms, the expressions for the variogram of z and m are found:

$$\begin{aligned}&= E\left\{z^2 - 2zz_{\mathbf{h}} + z_{\mathbf{h}}^2 + m^2 - 2mm_{\mathbf{h}} + m_{\mathbf{h}}^2\right. \\ &\quad \left. -2zm + 2zm_{\mathbf{h}} + 2z_{\mathbf{h}}m - 2z_{\mathbf{h}}m_{\mathbf{h}}\right\} \\ &= E\left\{z^2 - 2zz_{\mathbf{h}} + z_{\mathbf{h}}^2\right\} + E\left\{m^2 - 2mm_{\mathbf{h}} + m_{\mathbf{h}}^2\right\} \\ &\quad -2E\left\{(z-z_h)(m-m_h)\right\} \\ &= \gamma_z + \gamma_m - 2E\left\{(z-z_h)(m-m_h)\right\}\end{aligned}$$

The last term is the covariance based cross variogram (Myers, 1982; Cressie and Wilke, 1998) denoted ν_{zm} in Equation (9).

$$\gamma_\varepsilon = \gamma_z + \gamma_m - 2\nu_{zm} \tag{9}$$

This equation is very similar to that for the additivity of variance for two correlated random variables. If we assume that the trend is a continuous function, then both γ_m and ν_{zm} will exhibit a parabolic behaviour near the origin and the short range structure of γ_ε will be very similar to that of γ_z and it will be identical at the origin. This implies that the nugget effect of z and ε are equal.

Gaussian Simulation with Trends

Several techniques exist for incorporating trends into geostatistical simulation such as working with residuals or transformation of a random function conditional to the trend. A few issues exist in working with residuals including the bias in variance already discussed and the possibility of simulating values that are beyond the range of feasibility of a random function. For example, simulation of residuals could lead to negative values for a non-negative random function due to the decomposition in Equation (1).

A few approaches exist to avoid simulating values that are beyond physical limits of a variable: instead of using the trend explicitly, it could be used as an LVM, or; the original variable could be transformed to an infinite range, for example using a normal score transformation prior to trend modeling. Both of these options typically involves a normal score transform in practice. There are two approaches for simulation with an LVM: 1 – the local mean is assumed stationary in the space spanned by the data to be used in kriging (see Equation (10), where y^* is the estimate, $\boldsymbol{\lambda}$ is the vector of kriging weights, $\mathbf{y}$ is the vector of conditioning data, and m is the LVM); 2 – local residuals are calculated from the LVM and used in kriging (see Equation (11), where $\mathbf{m}$ is a vector of LVM values co-located with $\mathbf{y}$).

$$y^*(\mathbf{u}) = \boldsymbol{\lambda}'\mathbf{y} + (1 - \mathbf{1}'\boldsymbol{\lambda})m(\mathbf{u}) \tag{10}$$

$$y^*(\mathbf{u}) = \boldsymbol{\lambda}'(\mathbf{y} - \mathbf{m}) + m(\mathbf{u}) \tag{11}$$

The second and most common approach is similar to explicit handling of a trend where residuals are calculated and used. The difference is that in the explicit case, residuals are usually computed using the original variable and trend via Equation (12), whereas in the LVM case of Equation (11), the original variable and trend are normal score transformed independently in Equation (13), where the trend is transformed using z as a reference distribution.

$$\begin{aligned} \varepsilon_g &= \text{nscore}(z - m) \\ &= \text{nscore}(\varepsilon) \end{aligned} \tag{12}$$

$$\begin{aligned} \varepsilon_g' &= \text{nscore}(z) - \text{nscore}(m) \\ &= y - m_g \neq \varepsilon_g \end{aligned} \tag{13}$$

Resulting residuals, ε_g and ε'_g, are not equal and there is no guarantee that they retain the same rank order. Moreover, the bivariate distribution between z and m is not necessarily Gaussian, so the residuals, ε'_g, are not guaranteed to be normally distributed even though the SGS algorithm assumes they are. The normal score transforms in Equation (13) are applied independently because often the LVM is not a trend that was modeled directly from z, but rather it is an exhaustive secondary variable that may not be in the same unit space as z. During simulation, values of y are simulated after adding the trend to the estimate; therefore, the resulting back-transformed values are guaranteed to be in the physical range of z. This tends to result in variance inflation because the variance of y is increased by the variance of m_g.

For Equations (12) and (13), the correct variogram to use should be the variogram of the corresponding residuals, possibly debiased based on Equation (5). It is unclear the form of the bias when the residuals are normal score transformed in Equation (12) because the trend and stochastic components cannot be decoupled. In Equation (13), these two components remain decoupled; however, the trend is not necessarily derived from linear regression and there is no expression for the projection matrix. The bias can be written in terms of expected values in Equation (14).

$$\begin{aligned} Var\left\{\varepsilon_g'\right\} &= E\left\{(y - m_g)(y - m_g)'\right\} \\ &= E\left\{yy'\right\} - E\left\{ym_g'\right\} - E\left\{m_g y'\right\} + E\left\{m_g m_g'\right\} \\ &= \Sigma_{\varepsilon_g'} - \left(E\left\{ym_g'\right\} + E\left\{m_g y'\right\} - E\left\{m_g m_g'\right\}\right) \\ &= \Sigma_{\varepsilon_g'} - B \end{aligned} \tag{14}$$

Incorporating an LVM using the first approach (Equation (10)) also involves normal score transformation of z and m independently. Since kriging is based on y, the variogram of y should be the correct one; however, by admitting that simulation should involve an LVM this indicates the variable is influenced by a trend or is non-stationary. However, the presence of a trend influences variogram interpretation and modeling (Starks and Fang, 1982).

Another approach for incorporating trends into simulation was discussed by Leuangthong and Deutsch (2004) and involves transforming the original variable conditional to the trend. A normal score transform would be performed on $z|m$ defined in Equation (15) producing yet another form of residual variable for modeling and simulation.

$$\varepsilon = \text{nscore}(z \mid m) \tag{15}$$

With this approach it is not possible to simulate grades that are beyond the physical limits of the original variable since it is involved directly in the transformation. Back-transformation will not yield negative grades for a non-negative random function for example. What is unclear is if this transformation process yields any biases in terms of the variance or variogram. Naturally, one would perform the transformation, which results in a random function with zero mean and unit variance, and use the variogram of ε.

The conditional transformation approach and the traditional approach that involves a normal score transform of the residuals are very similar in that they result in distributions with similar statistics. Both result in a Gaussian distribution with zero mean and unit variance; the correlation between the resulting random functions is high; the rank order within any given class of m involved in the transformation is the same, and; the variograms are similar. A small example is used to show these similarities and involves a 2D set of data with a linear trend in one direction (X) and zero-trend in the perpendicular direction (Y), see Figure 1. There is a significant amount of data (10,000 samples) so that the conditional transformation is well informed.

The samples were uniformly sampled from a 200 by 200 cell regular grid so that choosing the bins for the conditional transformation was straightforward. Each column of cells with constant x forms a bin since the trend is constant in y. All the data in each column is normal score transformed according to Equation (15) . There were nominally 50 samples in each class. Statistics of $z|m$ and traditional residuals are shown on Figure 2. Both distributions are Gaussian; however, due to the limited number of samples in each bin of the conditional transform, the tails are not fully informed. This is due to the numerical nature of the normal score transform and to the uniformity of the number of samples in each bin.

Because the correlation coefficient between the two random functions is high, 0.973, the variograms must be similar. These are shown in Figure 3 in the x and y directions for the original variable, trend, normal scored residuals, and conditional transformed values. In the y direction, the variograms of z, the normal scores of $z - m$ and of $z|m$ are very similar. In the x direction, the variograms of $z - m$ and $z|m$ are similar, whereas the variogram of z clearly shows the effect of the trend. The only notable difference between the variograms of $z - m$ and $z|m$ is the apparent nugget effect of the conditional transformed variable. This is caused by binning z based on m and by the normal score transform. The normal score transform is usually applied using the empirical cumulative distribution function of a set of data, which provides the probabilities required to compute the equivalent Gaussian values. As the number of data decreases, the differences between nearby pairs of Gaussian values increases, hence the variogram increases. Binning also has an effect because it leads to a loss in continuity across the bin boundaries, hence the increased nugget effect. A conditional transformation that does not utilize bins may eliminate this source of short range variability.

Conclusion

One approach to account for first order non-stationarity in Geostatistics is to use a trend model. There are several techniques available for incorporating a trend, this paper discussed three explicit ones including the traditional approach that uses residuals, using the trend as an LVM, and using a conditional transformation approach. In all cases, the correct variogram to use is the variogram of the residuals. This result was clear for the traditional approach, but has not been done in practice with LVM's. For the conditional transformation approach, it seems that the variogram of the normal score residuals and of the conditional transformed variable are nearly identical and either variogram could be used in practice; however, more complex examples may yield a different result.

References

Chiles, J.-P., Delfiner, P., 1999. Geostatistics: modeling spatial uncertainty. John Wiley & Sons, Inc., 720pp.

Matheron, G., 1963. Principles of Geostatistics. Economic Geology, Vol. 58, 1246-1266

Matheron, G., 1963. The intrinsic random functions and their applications. Advances in Applied Probability, Vol. 5, No. 3, 439-468

Journel, A.G., Huijbregts, C.J., 1978. Mining Geostatistics. Academic Press, London, 600pp.

Isaaks, E.H., Srivastava, M.R., 1989. An introduction to applied Geostatistics. Oxford University Press, New York, 561pp.

Huijbregts, C.J., Matheron, G., 1971. Universal kriging – an optimal approach to trend surface analysis. In Canadian Institute of Mining and Metallurgy, Decision Making in the Mineral Industry, Vol. 12, 159-169

Beckers, F., Bogaert, P., 1998. Nonstationarity of the mean and unbiased variogram estimation: extension of the weighted least-squares method. Mathematical Geology, Vol. 30, No. 2, 223-240

Cressie, N.A.C., 1993. Statistics for spatial data. John Wiley & Sons, Inc., 900pp.

Cressie, N., Wilke, C.K., 1998. The variance-based cross-variogram: you can add apples and oranges. Mathematical Geology, 30(7), 789-799

Kitanidis, P.K., 1993. Generalized covariance functions in estimation. Mathematical Geology, Vol. 25, No. 5, 525-540

Leuangthong, O., Deutsch, C.V., 2004. Transformation of residuals to avoid artifacts in geostatistical modeling with a trend. Mathematical Geology, Vol. 36, No. 3, 287-305

Rosenblatt, M., 1952. Remarks on a multivariate transformation. Ann. Math. Stat., Vol. 23, No. 3, 470-472

Journel, A.G., Rossi, M.E., 1989. When do we need a trend model in kriging? Mathematical Geology, Vol. 21, No. 7, 715-739

Gringarten, E., Deutsch, C.V., 2001. Teacher's aide: variogram interpretation and modeling. Mathematical Geology, Vol. 33, No. 4, 507-534

Gooverts, P., 1997. Geostatistics for natural resource evaluation. Oxford University Press, New York, 483pp.

Myers, D., 1982. Matrix formulation of co-kriging. Mathematical Geology, 14(3), 249-257

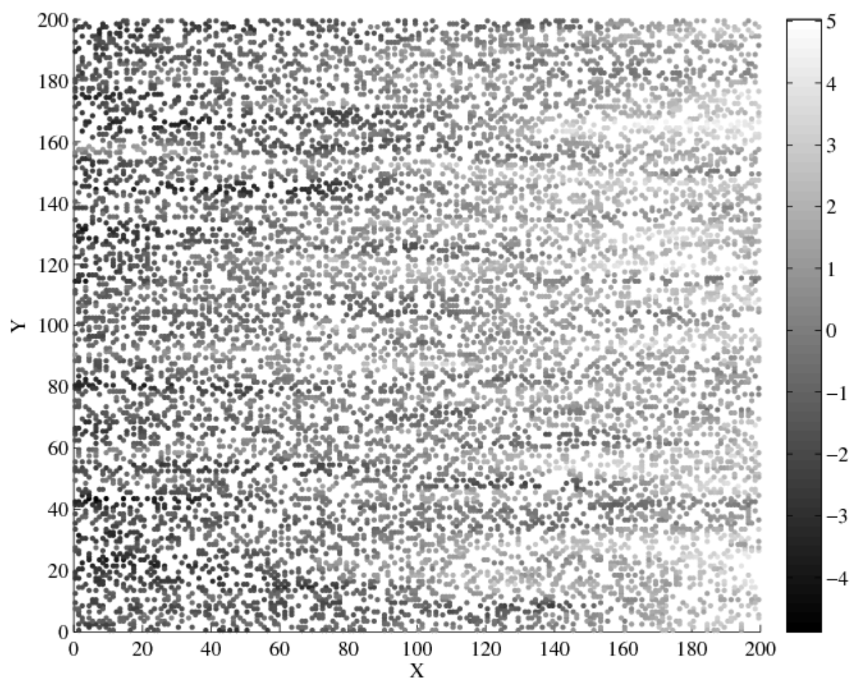

Figure 1: Sample data for assessing the conditional transformation approach.

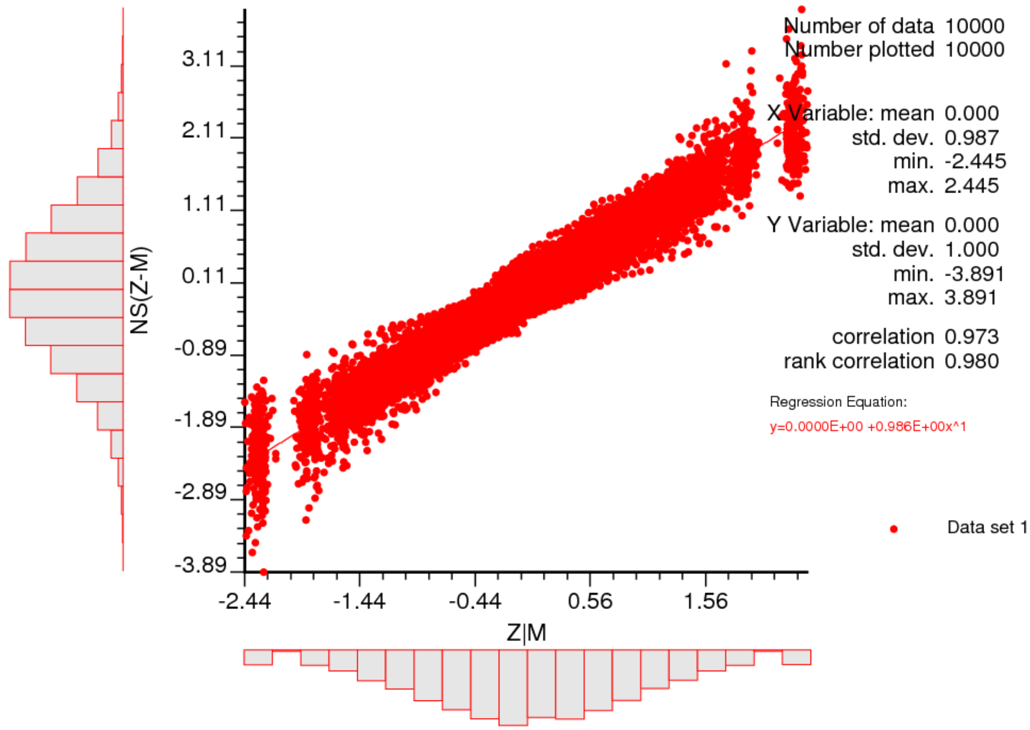

Figure 2: Cross plot of variables from the conditional transform and from the normal score transform of traditional residuals.

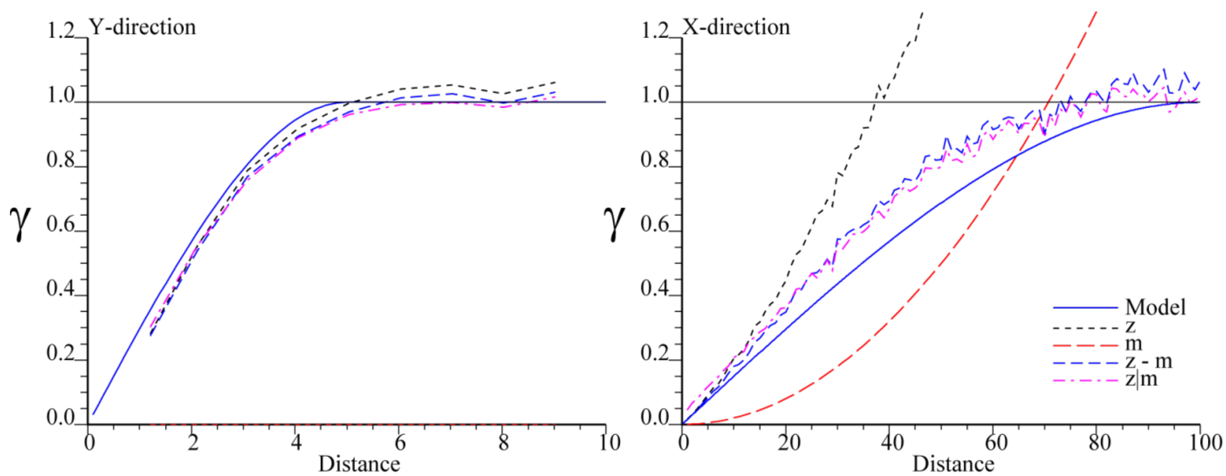

Figure 3: Variograms of the input model, original variable, trend and residuals for the example.