Comparison of kriging with external drift and regression-kriging*

July 3, 2003

TOMISLAV HENGL
Department of Earth Systems Analysis, International Institute for Geo-information Science and Earth Observation (ITC), P.O. Box 6, 7500 AA Enschede, Netherlands, hengl@itc.nl;

GERARD B.M. HEUVELINK
Wageningen University, Laboratory of Soil Science and Geology, P.O. Box 37, 6700 AA Wageningen, Netherlands, gerard.heuvelink@wur.nl;

ALFRED STEIN
Department of Earth Observation Science, International Institute for Geo-information Science and Earth Observation (ITC), P.O. Box 6, 7500 AA Enschede, Netherlands, stein@itc.nl;

*Please cite as: Hengl T., Geuvelink, G.B.M. and Stein A. 2003. Comparison of kriging with external drift and regression-kriging. Technical note, ITC, Available on-line at http://www.itc.nl/library/Academic_output/

This technical note was written as a supplementary material to a paper accepted for publication in Geoderma journal: Hengl T., Heuvelink, G.B.M. and Stein A. 2003. *A generic framework for spatial prediction of soil variables based on regression-kriging.* In further text, theory behind kriging with external drift and regression-kriging and differences between them are explained in more detail. We focus mainly on practical issues, i.e. how to derive predictions and prediction uncertainty. We also give a small case study where you can follow calculation of different elements. Use footnotes to obtain additional explanations. For an introduction to matrix algebra, read the general introductions in classical statistical books (Neter et al., 1996, §5). A detailed introduction to matrix algebra for KED can be found in Wackernagel (1998, appendix).

1 Introduction

Universal kriging, kriging with external drift and regression-kriging belong to the group of the so-called 'hybrid' (McBratney et al., 2000), i.e. non-stationary geostatistical methods (Wackernagel, 1998). Unfortunately, there has been quite some confusion between different authors with what are the computational differences between the three: different authors use the same names for different approaches and different names for the same approach. The most probable cause is that similar applications have been developed among different professions and with different goals. The second important cause of this confusion is that some authors, more involved in the practice of kriging ('geostatisticians'), consider these techniques a special interpolation technique, while the other group ('statisticians') consider kriging to be only a case of regression analysis with spatially correlated data. Both views are correct in fact. The non-stationary geostatistical techniques, where auxiliary information (e.g. slope map) is used to improve spatial prediction[1], can be in general classified depending on the properties of input data:

- (**A**) if number of auxiliary variables is low and they are not available at all grid-nodes, **co-kriging** (CK) should be used to improve the prediction (co-kriging requires estimation of cross-variograms);
- (**B**) if auxiliary information is available at all grid-nodes and correlated with the target variable, kriging with unknown mean (Chiles and Delfiner, 1999) or kriging with a trend model (Deutsch and Journel, 1992), also called "external drift" (Bourennane et al., 1996; Hudson and Wackernagel, 1994; Bourennane et al., 2000), should be used. Here, at least three, computationally different approaches can be recognised:
  - (B.1) "**Universal kriging**" (UK), first introduced by Matheron (1969), is as a special case of kriging with changing mean where the trend is modelled as a function of coordinates. Other authors (Deutsch and Journel, 1992; Wackernagel, 1998; Papritz and Stein, 1999) also agree that the term *Universal kriging* should be reserved for the case where only the coordinates are used.
  - (B.2) If, instead of using monomials of the coordinates in the UK equations, the drift is defined externally through some auxiliary variables, the term "**Kriging with external drift**" (KED) or external trend is used (Wackernagel, 1998; Chiles and Delfiner, 1999, ,p. 355). This is probably the most preferred name used for kriging with auxiliary information.

[1] Spatial prediction is process of estimating the target quantity (z) at a new, unvisited location (s_0), given its coordinates and interpolation data set ($z(s_1), z(s_2), ..., z(s_n)$). In GIS terms, predictions are made at all raster nodes or pixels in a new map. Hence, spatial prediction is in fact interpolation, i.e. mapping process.

– (B.3) The drift and residuals can also be fitted separately and then summed. Similar procedure was first time suggested by Ahmed and de Marsily (1987). Odeh et al. (1994, 1995) named it "**Regression-kriging**" (RK), while Goovaerts (1999) uses term *kriging after detrending*.

The advantage of RK is that it can be easily combined with stratification, GAM, regression trees etc. (McBratney et al., 2000). In this note, we concentrate on differences between the KED and RK in order to find out do they give different predictions and which approach is more optimal?

2 Kriging with external drift

In the case of KED, predictions at new locations are made by:

$$\hat{z}_{\texttt{KED}}(s_0) = \sum_{i=1}^{n} w_i^{\texttt{KED}}(s_0) \cdot z(s_i) \tag{1}$$

for

$$\sum_{i=1}^{n} w_i^{\texttt{KED}}(s_0) \cdot q_k(s_i) = q_k(s_0); \quad k = 1, ..., p \tag{2}$$

or in matrix notation:

$$\hat{z}_{\texttt{KED}}(s_0) = \delta_{\mathbf{0}}^{\mathbf{T}} \cdot \mathbf{z} \tag{3}$$

where z is the target variable[2], q_k's are the predictor variables i.e. values at a new location (s_0), $\delta_{\mathbf{0}}$ is the vector of KED weights ($w_i^{\texttt{KED}}$), p is the number of predictors and $\mathbf{z}$ is the vector of n observations at primary locations. The KED weights are solved using the extended matrices:

$$\begin{aligned} \lambda_{\mathbf{0}}^{\texttt{KED}} &= \left\{ w_1^{\texttt{KED}}(s_0), ..., w_n^{\texttt{KED}}(s_0), \varphi_0(s_0), ..., \varphi_p(s_0) \right\}^{\mathbf{T}} \\ &= \mathbf{C}^{\texttt{KED}-1} \cdot \mathbf{c}_{\mathbf{0}}^{\texttt{KED}} \end{aligned} \tag{4}$$

[2] A typical data-set consists of n observations of the target variable ($z(s_i)$; $i = 1, ..., n$; $s_i \in G$) with xy coordinates of sampling locations (so called 'primary' locations), where G denotes the study area. The auxiliary variables, also called 'predictors', ($q_1(s_i), ..., q_k(s_i)$; $k = 1, ..., p$) are measured at the same locations and are also available at all new locations. Sign q is used instead of more common x, to avoid confusion with geographical coordinates.

where $\lambda_{\mathbf{0}}^{\texttt{KED}}$ is the vector of solved weights, φ_p are the Lagrange multipliers, $\mathbf{C}^{\texttt{KED}}$ is the extended covariance matrix of residuals and $\mathbf{c}_{\mathbf{0}}^{\texttt{KED}}$ is the extended vector of covariances at new location.

In the case of KED, the extended covariance matrix of residuals looks like this (Webster and Oliver, 2001):

$$\mathbf{C}^{\texttt{KED}} = \begin{bmatrix} C(s_1,s_1) & \cdots & C(s_1,s_n) & 1 & q_1(s_1) & \cdots & q_p(s_1) \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ C(s_n,s_1) & \cdots & C(s_n,s_n) & 1 & q_1(s_n) & \cdots & q_p(s_n) \\ 1 & \cdots & 1 & 0 & 0 & \cdots & 0 \\ q_1(s_1) & \cdots & q_1(s_n) & 0 & 0 & \cdots & 0 \\ \vdots & & \vdots & 0 & \vdots & & \vdots \\ q_p(s_1) & \cdots & q_p(s_n) & 0 & 0 & \cdots & 0 \end{bmatrix} \tag{5}$$

and $\mathbf{c}_{\mathbf{0}}^{\texttt{KED}}$ like this:

$$\mathbf{c}_{\mathbf{0}}^{\texttt{KED}} = \{C(s_0,s_1),...,C(s_0,s_n),q_0(s_0),q_1(s_0),...,q_p(s_0)\}^{\mathbf{T}}; \quad q_0(s_0) = 1 \tag{6}$$

Note that we use the covariances since a common practice in geostatistics is to model the variogram using a semivariance function and then, for the reasons of computational efficiency, use the covariances[3]. The relation between the covariances and semivariances is (Isaaks and Srivastava, 1989, p. 289):

$$C(\mathbf{h}) = C_0 + C_1 - \gamma(\mathbf{h}) \tag{7}$$

where $C(\mathbf{h})$ is the covariance, and $\gamma(\mathbf{h})$ is the semivariance function. So for example, in the case of exponential model:

$$\gamma(\mathbf{h}) = \begin{cases} 0 & \text{if} \quad |\mathbf{h}| = 0 \\ C_0 + C_1 \cdot \left[1 - e^{-\left(\frac{\mathbf{h}}{R}\right)}\right] & \text{if} \quad |\mathbf{h}| > 0 \end{cases} \tag{8}$$

$$C(\mathbf{h}) = \begin{cases} C_0 + C_1 & \text{if} \quad |\mathbf{h}| = 0 \\ C_1 \cdot \left[e^{-\left(\frac{\mathbf{h}}{R}\right)}\right] & \text{if} \quad |\mathbf{h}| > 0 \end{cases} \tag{9}$$

where $|\mathbf{h}|$ is the Euclidian distance[4] between the point pairs and C_0, C_1, R are the estimated parameters. Note that the covariance at zero distance ($C(0)$) is by definition equal to the mean residual error (Cressie, 1993) — $C(h_{11})$ also written

[3] In the case of solving the kriging weights or drift model coefficients, both the matrix of semivariances and covariances give the same results. The same works for the calculation of GLS estimate of drift model coefficients (Eq. 13). In the case of calculating the drift prediction error, however, the matrix of covariances needs to be used.

[4] $\mathbf{h}$ is a vector of distances in x and y coordinates.

as $C(s_1, s_1)$ is equal to $C(0) = C_0 + C_1 = Var\{z(s)\}$. Note that, at zero distances (diagonal), the covariance (seminvariance) function needs to be replaced with $C_0 + C_1$[5].

The variance of the prediction error[6], i.e. KED variance is then calculated as the weighted average of covariances from the new point (s_0) to all calibration points ($s_1, .., s_n$), plus the Lagrange multipliers (Webster and Oliver, 2001, p. 183):

$$\begin{aligned}\sigma^2_{\mathtt{KED}}(s_0) &= (C_0 + C_1) - \mathbf{c_0^{KED T}} \cdot \lambda_{\mathbf{0}}^{\mathtt{KED}} \\ &= C_0 + C_1 - \sum_{i=1}^{n} w_i(s_0) \cdot C(s_0, s_n) + \sum_{k=0}^{p} \varphi_k(s_0) \cdot q_k(s_0)\end{aligned} \tag{10}$$

3 Regression kriging

In the case of RK, the predictions are made separately for the drift and residuals and then added back together[7]:

$$\hat{z}_{\mathtt{RK}}(s_0) = \hat{m}(s_0) + \hat{e}(s_0) \tag{11}$$

$$\hat{z}_{\mathtt{RK}}(s_0) = \sum_{k=0}^{p} \hat{\beta}_k \cdot q_k(s_0) + \sum_{i=1}^{n} w_i(s_0) \cdot e(s_i); \qquad q_0(s_0) = 1; \quad i = 1, ..., n \tag{12}$$

where $\hat{\beta}_k$ are estimated drift model coefficients, w_i are weights determined by the semivariance function and e are the regression residuals. The (global) drift model coefficients are optimally estimated[8] using the generalized least squares (GLS) to account for the spatial correlation of residuals (Cressie, 1993, p. 166):

$$\hat{\beta}_{gls} = \left(\mathbf{q^T} \cdot \mathbf{C^{-1}} \cdot \mathbf{q}\right)^{\mathbf{-1}} \cdot \mathbf{q^T} \cdot \mathbf{C^{-1}} \cdot \mathbf{z} \tag{13}$$

[5](Burrough and McDonnell, 1998, p. 140) in their example use the actual nugget at zero distances (instead of $C(0)$). This gives a slightly different, suboptimal result.

[6]Prediction error is the statistical estimate of the model uncertainty. Note that the 'true' prediction power can only be assessed by using the independent (control) data set. The prediction error is therefore often referred to as the *precision of prediction*.

[7]This inexpensive in a GIS: first the drift model coefficients are solved, then residuals interpolated using kriging and summed back to the drift estimates using a calculation with maps. For example, a GIS calculation for mapping the topsoil thickness (*DEPTH*) predicted from the slope map (*SLOPE*) and krigged residuals (*RES_OK*) might look like: *DEPTH* = 21.31 - 0.2628 * *SLOPE* + *RES_OK*

[8]This is equivalent to weighted linear regression, where the covariance matrix is used as the matrix of weights.

where $\mathbf{C}$ is the covariance matrix of residuals[9]:

$$\mathbf{C} = \begin{bmatrix} C(s_1, s_1) & \cdots & C(s_1, s_n) \\ \vdots & \ddots & \vdots \\ C(s_n, s_1) & \cdots & C(s_n, s_n) \end{bmatrix} \tag{14}$$

Thus, RK in matrix notation is:

$$\hat{z}(s_0) = \mathbf{q_0^T} \cdot \hat{\beta} + \lambda_\mathbf{0}^\mathbf{T} \cdot \mathbf{e} \tag{15}$$

where $\hat{z}(s_0)$ is the predicted value, $\mathbf{q_0}$ is vector of $p+1$ predictors at new location, $\hat{\beta}$ is vector of $p+1$ estimated drift model coefficients, $\lambda_\mathbf{0}$ is vector of n kriging weights and $\mathbf{e}$ is vector of n residuals. The Eq. (15), in statistical terms, is the best linear unbiased predictor or BLUP (Christensen, 1990, p. 268) and gives exactly the same predictions as KED in Eq. (1). Note that the estimation of the residuals is an iterative process[10]: first the drift model is estimated using ordinary least squares (OLS); then the covariance function of the residuals is used to obtain the GLS coefficients; these can be used to re-compute residuals and so on.

The additivity relationship from Eq. (12) extends to variances as well. Hence, the prediction error is sum of error of predicting the drift and kriging error of residuals. The summary error is then (Chiles and Delfiner, 1999, p. 183):

$$\sigma^2_{\mathtt{RK}}(s) = \sigma^2\left\{\hat{m}(s)\right\} + \sigma^2\left\{\hat{e}(s)\right\} \tag{16}$$

where $\sigma^2\left\{\hat{m}(s)\right\}$ is the drift prediction error and $\sigma^2\left\{\hat{e}(s)\right\}$ is the kriging variance of residuals. The Eq. (16) can be also referred to as the *composite variance*. If the drift model coefficients are estimated using OLS, the covariance between residuals and estimated drift is assumed to be zero[11]. Hence, the composite variance can be derived using:

$$\sigma^2_{\mathtt{RK}}(s_0) = \underbrace{(C_0 + C_1) - \mathbf{c_0^T} \cdot \mathbf{C^{-1}} \cdot \mathbf{c_0}}_{\hat{\sigma}^2\{\hat{e}(s_0)\}} + \underbrace{\mathbf{q_0^T} \cdot \left(\mathbf{q^T} \cdot \mathbf{C^{-1}} \cdot \mathbf{q}\right)^{-1} \cdot \mathbf{q_0}}_{\hat{\sigma}^2\{\hat{m}(s_0)\}} \tag{17}$$

[9]Note that the $\mathbf{C}$ is in fact $(n+1) \times (n+1)$ matrix if it is used to derive kriging weights. One extra row and column are used to ensure that the sum of weights is equal to one. Elsewhere $n \times n$ part of matrix is used only.

[10]The major dissatisfaction of using KED or RK is that both the regression model parameters and semivariance function parameters need to be estimated simultaneously. However, in order to estimate coefficients we need to know covariance function of residuals, which can only be estimated after the coefficients (the "chicken-egg" problem). One solution is the iterative calculation of residuals and their covariances. However, Cressie (1993) thinks that not even the iterative estimation of both will necessarily lead to the optimal solution. Moreover, Kitanidis (1994) showed that multiple iterations are not needed. Therefore, in practice a single iteration can be considered satisfactory solution.

[11]This is a reasonable assumption since the OLS regression residuals are by definition orthogonal, i.e. normal random variables with 0 mean and constant variance (Neter et al., 1996, p. 97).

where $\mathbf{q_0}$ is the vector of $p+1$ predictors at unvisited location and $\mathbf{c_0}$ is the vector of covariances at new location:

$$\mathbf{c_0} = \{C(s_0, s_1), ..., C(s_0, s_n)\}^{\mathbf{T}} \tag{18}$$

This is a sub-optimal solution since the OLS estimation should be replaced with GLS estimation of the drift to get an unbiased estimate of coefficients. In the case of GLS estimation, we use the residuals to estimate the drift coefficient and therefore the covariance between the estimated drift and residuals is different from zero. A commonly used formula to calculate variance of the prediction error for both the drift and residuals, and account for covariance between the drift estimation and residuals is the UK variance (Cressie, 1993, p. 154):

$$\begin{aligned}\sigma^2_{\mathrm{E}}(s_0) = (C_0 + C_1) - \mathbf{c_0^T} \cdot \mathbf{C^{-1}} \cdot \mathbf{c_0} \\ + \left(\mathbf{q_0} - \mathbf{q^T} \cdot \mathbf{C^{-1}} \cdot \mathbf{c_0}\right)^{\mathbf{T}} \cdot \left(\mathbf{q^T} \cdot \mathbf{C^{-1}} \cdot \mathbf{q}\right)^{-1} \cdot \left(\mathbf{q_0} - \mathbf{q^T} \cdot \mathbf{C^{-1}} \cdot \mathbf{c_0}\right)\end{aligned} \tag{19}$$

Moreover, Cressie (1993) shows that the UK variance is equivalent to the KED variance (Eq. 10). Also note that the Eq. (19) looks very much like (17), except it will give slightly lower values.

The UK variance is, in fact, derived as the variance of prediction error (Papritz and Stein, 1999, p. 94):

$$\begin{aligned}\sigma^2_{\mathrm{E}}(s_0) = E\left\langle\{\hat{z}(s_0) - z(s_0)\}^2\right\rangle = Var\{z(s_0)\} - 2 \cdot Cov\{\hat{z}(s_0), z(s_0)\} + Var\{\hat{z}(s_0)\} \\ \sigma^2_{\mathrm{E}}(s_0) = C(0) - 2 \cdot \delta_{\mathbf{0}}^{\mathbf{T}} \cdot \mathbf{c_0} + \delta_{\mathbf{0}}^{\mathbf{T}} \cdot \mathbf{C} \cdot \delta_{\mathbf{0}}\end{aligned} \tag{20}$$

where $\delta_{\mathbf{0}}$ is:

$$\delta_{\mathbf{0}} = \left[\lambda_{\mathbf{0}}^{\mathbf{T}} + \mathbf{q} \cdot \left(\mathbf{q^T} \cdot \mathbf{C^{-1}} \cdot \mathbf{q}\right)^{-1} \cdot \left(\mathbf{q_0} - \mathbf{q^T} \cdot \mathbf{C^{-1}} \cdot \lambda_{\mathbf{0}}^{\mathbf{T}}\right)\right]^{\mathbf{T}} \cdot \mathbf{C}^{-1} \tag{21}$$

Note that the key assumption of the KED or RK is that there is no spatial dependence between the auxiliary variable and the residual of the linear regression of target variable on auxiliary variable at same point (Rivoirard, 2002). Further on, it can be shown that the UK variance is equal to variance of the prediction error of regression modelling if there is no spatial correlation between the residuals (*pure nugget effect*[12]). Hence, $\mathbf{C}$ can be reduced to identity matrix:

$$\mathbf{C} = \begin{bmatrix} C_0 + C_1 & \cdots & 0 \\ \vdots & C_0 + C_1 & 0 \\ 0 & 0 & C_0 + C_1 \end{bmatrix} = (C_0 + C_1) \cdot \mathbf{I} \tag{22}$$

[12] $C_1 \approx 0$ in variogram modelling.

and $\mathbf{c_0}$ is the zero vector[13], so that UK variance reduces to:

$$\sigma^2_{\mathtt{UK}}(s_0) = (C_0 + C_1) - 0 + \mathbf{q_0^T} \cdot \left(\mathbf{q^T} \cdot \frac{1}{(C_0 + C_1)} \cdot \mathbf{q}\right)^{\mathbf{-1}} \cdot \mathbf{q_0}$$

$$\sigma^2_{\mathtt{UK}}(s_0) = (C_0 + C_1) + (C_0 + C_1) \cdot \mathbf{q_0^T} \cdot \left(\mathbf{q^T} \cdot \mathbf{q}\right)^{\mathbf{-1}} \cdot \mathbf{q_0}$$

and since $(C_0 + C_1) = C(0) = MSE$, the UK variance reduces to:

$$\sigma^2_{\mathtt{UK}}(s_0) = MSE \cdot \left[1 + \mathbf{q_0^T} \cdot \left(\mathbf{q^T} \cdot \mathbf{q}\right)^{\mathbf{-1}} \cdot \mathbf{q_0}\right] \tag{23}$$

which is equal to the prediction error around the regression line (Neter et al., 1996, p. 210).

The expression $\sigma^2\{\hat{m}(s)\}$ in Eq. 17, in linear regression terms, is equivalent to the weighted curvature of the confidence bands around the regression hyperplane. In multivariate case, the regression error around the regression hyperplane would look like this[14]:

$$\sigma^2\{\hat{m}(s_0)\} = \frac{1}{n} + \frac{\sum_{k=1}^{p} (q_0 - \bar{q}_{wk})^2}{\sum_{k=1}^{p}\sum_{i=1}^{n} w_i (q_i - \bar{q}_{wk})^2} \tag{24}$$

where $\bar{q}_{wk}$ is the weighted average of the predictor values. Hence, from Eq. (19) it can be inferred that the prediction uncertainty will increase as the new point gets further away from observation points geographically and further away from the centre of the feature space[15].

In practice, if the correlation between the target and predictors is significant, the KED or RK prediction error ($\sigma^2_{\mathtt{E}}(s)$) will be in overall lower[16] than the OK error. At some locations, the composite prediction error is more sensitive to the predictor values than OK, i.e. how well is the feature space covered. For example, if the areas of very high slopes were overlooked during the sampling, they will appear in the prediction error map as high values. In fact, at locations of high extrapolation in both respective spaces, the KED/RK will result in higher prediction errors than OK.

[13] This is not the case only for predictions at primary locations $\hat{z}(s_i)$.

[14] This is just an approximation. The prediction error in multivariate case is derived using matrix algebra.

[15] Feature space is the physical range of predictor variables ($\mathbf{q}$).

[16] Papritz and Stein (1999) show that, the formulas, the KED variance will in general be larger than the OK variance, due to the general parameterisation of the drift. In practice, the RK/KED will in general account for more variation than OK and improve the prediction. In fact, the same authors (Papritz and Stein, 1999, p. 113) further on show that the auxiliary information can considerably improve the precision of the prediction, when compared to OK.

4 Summary remarks

At the end, it is important to emphasize that all three terms (UK, KED and RK), practically describe the same generic method that should give the same predictions and prediction error if the same input parameters are used (C_0, C_1, R, $\hat{\beta}_{gls}$). They differ, however, in the methodological steps used:

- KED solves kriging weights by extending the covariance matrix with auxiliary variables so that the universality conditions are integrated into the kriging system; here, the difficulty is obtaining satisfactory residual variogram in the presence of drift (Webster and Oliver, 2001);
- RK solves the drift model coefficients, while the residuals are interpolated using OK and added back to drift model; here, the difficulty is obtaining unbiased regression coefficients in the presence of spatial auto-correlation of residuals (Cressie, 1993);

Note that, although KED technique seems to be computationally more straightforward, it needs variogram parameters of the GLS regression residuals, and therefore the GLS regression coefficients (Eq. 13) as with RK. However, some authors make different assumptions and skip some computational step, such as estimation of GLS coefficients[17] or estimation of variogram of residuals[18]. These short-cuts might be more attractive for practical applications, but are sub-optimal statistically. This technical note serves specifically to minimise confusion and help understanding the differences when the short-cuts are made.

It is important to emphasize that, before applying KED or RK, some general requirements need to be fulfilled (Goovaerts, 1997):

- relation between the target and predictors must be linear (residuals show normal distribution);
- value of predictors must be known at all primary data locations (s_i) and all new locations (s_0) where the predictions are made;
- in the case of KED (weights are solved together), secondary variable should vary smoothly in space to avoid instability of the KED system[19];

[17] Hudson and Wackernagel (1994), for example, make an assumption that the variogram of residuals (e) is equal to the variogram of target variable (z), which is of course often simplification. In this case, higher importance will be given to the kriging weights and the KED prediction map will look more similar to the OK map than to the RK map.

[18] Odeh et al. (1994, 1995), for example, used only the OLS estimate of the drift and OLS residuals, which is a sub-optimal but shorter solution. In this case, the drift and residuals are assumably uncorrelated and Eq. (17) can be used instead of Eq. (19) to derive prediction error.

[19] If one predictor column/row has the same value at all primary locations, the matrix is unsolvable.

The fact that KED and RK are equal can also be shown mathematically. First, start from the RK (Eq. 15) and input the formula for GLS estimation of regression coefficients (Eq. 13):

$$\begin{aligned}\hat{z}_{\texttt{RK}}(s_0) &= \mathbf{q_0^T} \cdot \hat{\beta}_{gls} + \lambda_\mathbf{0}^\mathbf{T} \cdot (\mathbf{z} - \mathbf{q} \cdot \hat{\beta}_{gls}) \\ &= \mathbf{q_0^T} \cdot \left(\mathbf{q^T} \cdot \mathbf{C^{-1}} \cdot \mathbf{q}\right)^{\mathbf{-1}} \cdot \mathbf{q^T} \cdot \mathbf{C^{-1}} \cdot \mathbf{z} + \lambda_\mathbf{0}^\mathbf{T} \cdot \left[\mathbf{z} - \mathbf{q} \cdot \left(\mathbf{q^T} \cdot \mathbf{C^{-1}} \cdot \mathbf{q}\right)^{\mathbf{-1}} \cdot \mathbf{q^T} \cdot \mathbf{C^{-1}} \cdot \mathbf{z}\right] \\ &= \left[\mathbf{q_0^T} \cdot \left(\mathbf{q^T} \cdot \mathbf{C^{-1}} \cdot \mathbf{q}\right)^{\mathbf{-1}} \cdot \mathbf{q^T} \cdot \mathbf{C^{-1}} + \lambda_\mathbf{0}^\mathbf{T} \cdot \left(\mathbf{I} - \mathbf{q} \cdot \left(\mathbf{q^T} \cdot \mathbf{C^{-1}} \cdot \mathbf{q}\right)^{\mathbf{-1}} \cdot \mathbf{q^T} \cdot \mathbf{C^{-1}}\right)\right]^\mathbf{T} \cdot \mathbf{z}\end{aligned}$$

if you take a closer look, you will note that the first part of the equation equals (Papritz and Stein, 1999):

$$\mathbf{q_0^T} \cdot \left(\mathbf{q^T} \cdot \mathbf{C^{-1}} \cdot \mathbf{q}\right)^{\mathbf{-1}} \cdot \mathbf{q^T} \cdot \mathbf{C^{-1}} + \mathbf{c_0^T} \cdot \mathbf{C^{-1}} \cdot \left(\mathbf{I} - \mathbf{q} \cdot \left(\mathbf{q^T} \cdot \mathbf{C^{-1}} \cdot \mathbf{q}\right)^{\mathbf{-1}} \cdot \mathbf{q^T} \cdot \mathbf{C^{-1}}\right)$$

with some regrouping, you get:

$$= \left[\lambda_\mathbf{0}^\mathbf{T} + \mathbf{q} \cdot \left(\mathbf{q^T} \cdot \mathbf{C^{-1}} \cdot \mathbf{q}\right)^{\mathbf{-1}} \cdot \left(\mathbf{q_0} - \mathbf{q^T} \cdot \mathbf{C^{-1}} \cdot \lambda_\mathbf{0}^\mathbf{T}\right)\right]^\mathbf{T} \cdot \mathbf{C}^{-1}$$

which equals the $\delta_\mathbf{0}$, as explained in the Eq. (21). Hence,

$$\hat{z}_{\texttt{RK}}(s_0) = \delta_\mathbf{0}^\mathbf{T} \cdot \mathbf{z} = \sum_{i=1}^{n} w_i^{\texttt{KED}}(s_0) \cdot z(s_i) = \hat{z}_{\texttt{KED}}(s_0) \tag{25}$$

This means that we can use either Eq. (1) or Eq. (12) and should come up to the same predictions. Similarly, Eq. (10) and Eq. (19) should also give the same values for the prediction uncertainty (UK variance). The UK (with coordinates) is implemented in several geostatistical packages where the only input is the type of drift (xy transforms). The KED equations are implemented in GSTAT linked with `GRASS` GIS for example (Pebesma, 1999, see universal kriging). Moreover, in `GSTAT`, user can select number of maps for predictors and number of target variables at the same time.

5 Example

We applied described theory using a simple example with 20 points from an existing data set[20] in Croatia. The response or target variable is thickness of the topsoil in cm (*DEPTH*) and the auxiliary predictor is slope in % (*SLOPE*). The matrix calculations are applied in `MS Excel` (version 2000) using the following commands:

$$\begin{aligned} \mathbf{A} \cdot \mathbf{B} &= MMULT(\mathbf{A}; \mathbf{B}) \\ \mathbf{A^T} &= TRANSPOSE(\mathbf{A}) \\ \mathbf{A^{-1}} &= INVERSE(\mathbf{A}) \end{aligned}$$

`MS Excel` is limited to 100×100 cells matrices. Optionally, matrix calculations are possible in command line of a statistical package such as `S-PLUS` (MathSoft Inc., 1999). Here, the following functions are equivalent:

$$\begin{aligned} \mathbf{A} \cdot \mathbf{B} &= \mathbf{A}\%*\%\mathbf{B} \\ \mathbf{A^T} &= T(\mathbf{A}) \\ \mathbf{A^{-1}} &= SOLVE(\mathbf{A}) \end{aligned}$$

Note that in S-PLUS, the covariance matrices and vectors of target and predictors have to be first imported in the memory. There is no limit on the size of matrices, although the computations with >500×500 matrices and with >5000 points, might need lots of memory and processing time[21]. The advantage of using `Excel` is that the values are linked and can be changed manually, which gives a better insight into the principles of the method.

The data set can be seen in Table 1 and Fig. 1. Use `VESPER` (Minasny et al., 2002) or similar software to fit the variograms. Make predictions at a single new point (s_0, $X = 2415474$, $Y = 4972080$), where the predictor (*SLOPE*) value is $q_1(s_0) = 12.4\%$. From the target variable (z_i) and predictors (q_0, q_1), first derive the OLS regression coefficients ($\hat{\beta}_{ols}$). You get:

$$DEPTH = 21.35 - 0.2817 \cdot SLOPE \qquad (R^2 = 0.48)$$

See the respective prediction (z') and residuals (e_{ols}) in Table 1. The OLS residuals are now inspected for spatial autocorrelation. Due to a fairly small data set, the variogram modelling is difficult. Instead, use the exponential model with

[20] download the Excel spreadsheet from `http://www.itc.nl/library/Academic_output/`

[21] In the case of larger size of datasets, the KED estimation and KED variance calculated with Eq. (19) might not be feasible in standard statistical software or table calculator. For example, imagine a case with 200 observations, 10 auxiliary variables and a 1000×1000 new locations. Predictions and prediction error needs to be derived at 10^6 points, which means that the system has to solve 200×200 size matrices and multiply them with 10^6 size vectors. This is almost impossible in `S-PLUS` on a standard office PC.

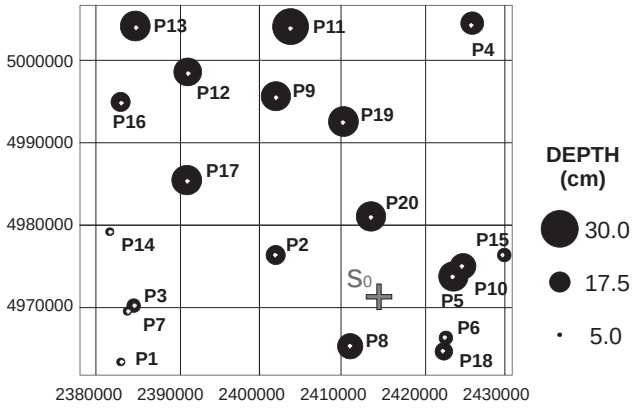

Figure 1: Observed points (20) and values of target variable ($DEPTH$); s_0 indicates the point for which the predictions are made.

the following parameters $C_0 = 0$, $C_1 = 16.2$ and $R = 1.907$ km (Fig. 2). Note that the residuals should typically show lower sill and shorter range of spatial dependence than the original variable.

Use the C_0, C_1 and R parameters to solve the GLS coefficients ($\hat{\beta}_{gls}$). First calculate the covariance matrix (size 20×20) using the exponential function (Eq. 9), then derive the GLS coefficients using Eq. 13. This gives somewhat different (more realistic) model (Fig. 3):

$$DEPTH = 21.31 - 0.2628 \cdot SLOPE \qquad (R^2 = 0.47)$$

Note that for GLS estimation, residuals are somewhat biased (compare the average OLS and GLS residuals in Table 1) around the regression line and so is the goodness of fit lower. In this case, GLS predictions (z'') are always higher than the OLS predictions (z').

To make predictions at the new points, first derive a vector of covariances at new location using the given variogram parameters and distances from the new point to all primary locations (Table 2). Note that the KED weights ($w_i^{\texttt{KED}}$) are quite different from the OK weights (w_i^{OK}). For example, the highest OK weights are at P8, P6, P5 (closest points), while the highest KED weights are differently distributed and with smaller difference (Table 2). Now, make predictions at the new point using Eqs. (1) and (12). You get the same result:

$$\hat{z}_{\texttt{KED}}(s_0) = 18.09$$
$$\hat{z}_{\texttt{RK}}(s_0) = 17.86 + 0.23 = 18.09$$

Table 1: Coordinates, target variable, predictors and residuals.

Point ID	Coordinates		Predictors		Target	OLS		GLS	
	Easting	Northing	Intercept	*SLOPE*	*DEPTH*				
	X	Y	q_0	q_1	z	z'	e_{ols}	z''	e_{gls}
P1	2382641	4965281	1	51.1	4	7.0	-3.0	7.9	-3.9
P2	2401577	4978420	1	15.9	13.3	16.9	-3.6	17.1	-3.8
P3	2384193	4972198	1	17.3	9	16.5	-7.5	16.8	-7.8
P4	2425685	5006765	1	31.7	16.1	12.4	3.7	13.0	3.1
P5	2423415	4975745	1	2.1	21.1	20.8	0.3	20.8	0.3
P6	2422427	4968288	1	37	9	10.9	-1.9	11.6	-2.6
P7	2383457	4971502	1	42	5	9.5	-4.5	10.3	-5.3
P8	2410757	4967205	1	19.7	18	15.8	2.2	16.1	1.9
P9	2401657	4997833	1	12.3	21.1	17.9	3.2	18.1	3.0
P10	2424570	4977027	1	3.4	18.5	20.4	-1.9	20.4	-1.9
P11	2403448	5006355	1	4.2	26.1	20.2	5.9	20.2	5.9
P12	2390827	5000819	1	0.9	20	21.1	-1.1	21.1	-1.1
P13	2384422	5006488	1	21.5	21.6	15.3	6.3	15.7	5.9
P14	2381288	4981203	1	37.5	5	10.8	-5.8	11.5	-6.5
P15	2429583	4978421	1	38.1	9	10.6	-1.6	11.3	-2.3
P16	2382591	4997168	1	13.3	13.3	17.6	-4.3	17.8	-4.5
P17	2390709	4987547	1	3.6	21.6	20.3	1.3	20.4	1.2
P18	2422209	4966624	1	27.3	12.1	13.7	-1.6	14.1	-2.0
P19	2409895	4994779	1	2.1	21.5	20.8	0.7	20.8	0.7
P20	2413321	4983096	1	47.3	21.1	8.0	13.1	8.9	12.2
avg.				21.42	15.32	15.32	0.00	15.68	-0.36
std.				16.54	6.70	4.66	4.81	4.35	4.82

Calculate the prediction error at the new location using both Eqs. (10) and (19). The value differs slightly, which might be due to the rounding of numbers in `Excel`:

$$\sigma^2_{\texttt{KED}}(s_0) = 17.23$$
$$\sigma^2_{\texttt{RK}}(s_0) = 16.13 + 1.05 = 17.18$$

Note that the estimation error of the drift is much smaller than the kriging variance of residuals. Optionally, you can change values of predictor (*SLOPE*) or coordinates at new location in order to see how will it influence the prediction error. You can see that, if you increase the *SLOPE* to say 100%, the estimation error of the drift will be higher than the kriging variance of residuals. Similarly, if you change coordinates so that they are fairly close to the primary locations, the kriging variance of residuals will decrease and become insignificant. Note that we make predictions at only one point. To calculate predictions and prediction error at all locations, you will need more processing power.

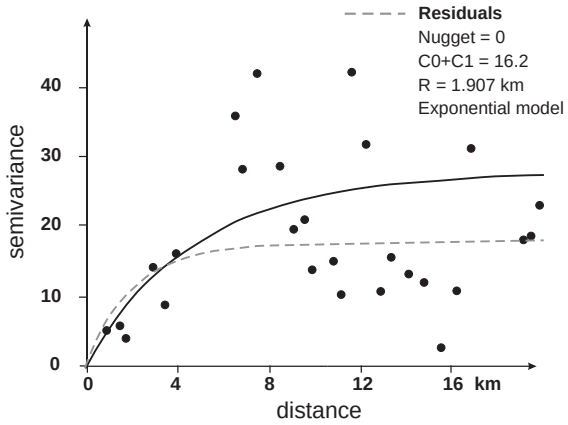

Figure 2: Variogram modelling of the target variable (solid line) and the OLS residuals (dashed line).

Table 2: Distances, covariances and derived weights at the new point.

Point ID	Distances / Covariances h_{0i}	$C(h_{0i})$	OK weights w_i^{OK}	KED weights $w_i^{\texttt{KED}}$
P1	33,530	0.000	0.026	0.002
P2	15,275	0.005	0.048	0.062
P3	31,281	0.000	0.013	0.075
P4	36,157	0.000	0.032	0.036
P5	8,746	0.165	0.125	0.071
P6	7,920	0.255	0.125	0.022
P7	32,022	0.000	0.011	-0.026
P8	6,783	0.462	0.244	0.084
P9	29,225	0.000	0.020	0.066
P10	10,354	0.071	0.013	0.057
P11	36,324	0.000	0.026	0.081
P12	37,860	0.000	0.020	0.086
P13	46,348	0.000	0.025	0.051
P14	35,382	0.000	0.025	0.025
P15	15,468	0.005	0.019	0.019
P16	41,361	0.000	0.023	0.065
P17	29,198	0.000	0.023	0.082
P18	8,668	0.172	0.061	0.043
P19	23,375	0.000	0.023	0.084
P20	11,224	0.045	0.097	0.012
Σ			1.000	1.000
φ_0			0.937	1.437
φ_1				-0.027

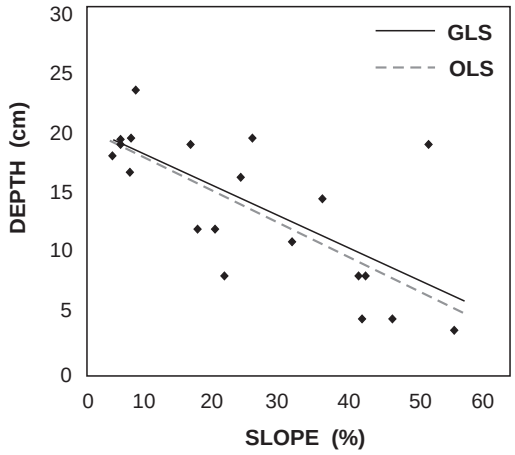

Figure 3: Correlation analysis between the target (*DEPTH*) and predictor (*SLOPE*): ordinary least squares estimation (OLS) and generalized least squares (GLS).

References

Ahmed, S., de Marsily, G., 1987. Comparison of geostatistical methods for estimating transmissivity using data on transmissivity and specific capacity. Water Resources Research 23 (9), 1717–1737.

Bourennane, H., King, D., Chery, P., Bruand, A., 1996. Improving the kriging of a soil variable using slope gradient as external drift. European journal of Soil Science 47 (4), 473–483.

Bourennane, H., King, D., Couturier, A., 2000. Comparison of kriging with external drift and simple linear regression for predicting soil horizon thickness with different sample densities. Geoderma 97 (3–4), 255–271.

Burrough, P., McDonnell, R., 1998. Principles of geographical information systems. Oxford University Press, Oxford.

Chiles, J., Delfiner, P., 1999. Geostatistics: modeling spatial uncertainty. John Wiley & Sons, New York.

Christensen, R., 1990. Linear Models for Multivariate, Time Series, and Spatial Data. Springer Verlag, New York.

Cressie, N., 1993. Statistics for Spatial Data, revised edition. John Wiley & Sons, New York.

Deutsch, C., Journel, A., 1992. Geostatistical Software Library and User's Guide. Oxford University Press, New York.

Goovaerts, P., 1997. Geostatistics for Natural Resources Evaluation. Oxford University Press, New York.

Goovaerts, P., 1999. Using elevation to aid the geostatistical mapping of rainfall erosivity. Catena 34 (3–4), 227–242.

Hudson, G., Wackernagel, H., 1994. Mapping temperature using kriging with external drift: theory and an example from scotland. International journal of Climatology 14 (1), 77–91.

Isaaks, E., Srivastava, R., 1989. Applied Geostatistics. Oxford University Press, New York.

Matheron, G., 1969. Le krigeage universel. Vol. 1. Cahiers du Centre de Morphologie Mathematique, Ecole des Mines de Paris, Fontainebleau.

MathSoft Inc., 1999. S-PLUS 4 Guide to Statistics. Vol. 1 and 2. MathSoft Inc., Seattle.

McBratney, A., Odeh, I., Bishop, T., Dunbar, M., Shatar, T., 2000. An overview of pedometric techniques for use in soil survey. Geoderma 97 (3–4), 293–327.

Minasny, B., McBratney, A., Whelan, B., 2002. VESPER version 1.5. Australian Centre for Precision Agriculture, McMillan Building A05, The University of Sydney, NSW 2006.
URL `http://www.usyd.edu.au/su/agric/acpa`

Neter, J., Kutner, M., Nachtsheim, C., Wasserman, W. (Eds.), 1996. Applied Linear Statistical Models, 4th Edition. The McGraw-Hill Companies.

Odeh, I., McBratney, A., Chittleborough, D., 1994. Spatial prediction of soil properties from landform attributes derived from a digital elevation model. Geoderma 63 (3–4), 197–214.

Odeh, I., McBratney, A., Chittleborough, D., 1995. Further results on prediction of soil properties from terrain attributes: heterotopic cokriging and regression-kriging. Geoderma 67 (3–4), 215–226.

Papritz, A., Stein, A., 1999. Spatial prediction by linear kriging. In: Stein, A., van der Meer, F., Gorte, B. (Eds.), Spatial statistics for remote sensing. Kluwer Academic publishers, Dodrecht, pp. 83–113.

Pebesma, E. J., 1999. Gstat user's manual. Dept. of Physical Geography, Utrecht University, `http://www.gstat.org`.

Rivoirard, J., 2002. On the structural link between variables in kriging with external drift. Mathematical Geology 34 (7), 797–808.

Wackernagel, H., 1998. Multivariate geostatistics: an introduction with applications, 2nd Edition. Springer-Verlag.

Webster, R., Oliver, M., 2001. Geostatistics for Environmental Scientists. Statistics in Practice. John Wiley & Sons, Chichester.