Kriging

Overview

- ***Simple kriging*** assumes that the ***constant mean level*** of the surface is ***known externally***
  Potential extension: In combination with Feasible General Least Squares we know that the expectation of the residuals.
  Thus we can apply simple kriging on the residuals of a GLS model[1].
  GLS separates the first order component $\mu_i(\mathbf{s})$ from the stochastic second order component $\varepsilon(\mathbf{s}_i) \equiv \varepsilon'(\mathbf{s}_i) + \varepsilon''(\mathbf{s}_i)$ of the spatial field variable $Y(\mathbf{s}_i) = \mu(\mathbf{s}_i) + \varepsilon(\mathbf{s}_i)$.
    - The *first order component* is $\mu(\mathbf{s}_i)$ modeled by a GLS trend surface model assuming a given autocorrelation structure of the observations.
    - The residuals $\varepsilon(\mathbf{s}_i) = Y(\mathbf{s}_i) - \mu(\mathbf{s}_i)$ are used to estimate the semi-variogram $\gamma(h)$ as a function of distance h.
    - The spatially dependent error term $\varepsilon(\mathbf{s}_i)$ at site $\mathbf{s}_i$ is assumed to have a mean of zero at any location and it is modeled by simple kriging.

[1] This is done in the example by Bailey and Gatrell and this is the approach geo-statisticians prefer.

- ***Ordinary kriging*** focuses on the prediction of the surface $Y(\mathbf{s})$ at any point ignoring additional explanatory variables. However, it estimates a ***constant mean level*** of the surfaces from the sample data.
  Remember that the semi-variogram can be robustly estimated as long as the expected value of the surface is constant.

use semi-viodiagram do not need stationary in universal kriging

- ***Universal kriging*** incorporates explanatory variables into the prediction process. These explanatory variables are usually establishing a trend surface because a trend surface can also ***predict first order values*** of the surface at ***non-sampling locations***.
  However, universal kriging does not give us any information about the estimated trend surface model and the underlying polynomial expression in the *x*- and *y*-coordinates, which prohibits any interpretation and model identification.

Simple Kriging

- Simple kriging starts from the model $Y(\mathbf{s}_i) = \mu_i(\mathbf{s}) + \varepsilon(\mathbf{s}_i)$, where $\mu_i(\mathbf{s}) = const$ is ***given externally*** and is assumed to be constant for all locations $\mathbf{s}$. It attempts to model the surface in the error component $\varepsilon(\mathbf{s}_i) = Y(\mathbf{s}_i) - const$.
- Because the error components $\varepsilon(\mathbf{s}_i)$ are no longer statistically independent, ***knowledge of the autocorrelation structure*** in error components at the ***sample location*** can be used to estimate the error component at any ***prediction location***.
- We also assume that the error component $\varepsilon(\mathbf{s}_i)$ has a constant variance at each location and preferably a joint normal distribution.

$\Rightarrow$ we may need to transform the data so that $\varepsilon(\mathbf{s}_i)$ follow approximately a normal distribution or at least are symmetrically distributed.

- The semi-variogram of the error component $\varepsilon(\mathbf{s}_i)$ allows estimating the ***covariances*** at any distance h, which can be derived from the estimated semi-variogram function (see Bailey & Gatrell, p 178).
- For any location **s** we estimate the error $\hat{\varepsilon}(\mathbf{s})$ from the given sample error components $\varepsilon(\mathbf{s}_i), i \in \{1,\ldots,n\}$, by using a set of estimated weights $\hat{\lambda}_i(\mathbf{s}), i \in \{1,\ldots,n\}$ and $\mathbf{s} \in \mathbb{R}$.

  These weights *depend*:
  [a] on the prediction location **s** and
  [b] on the spatial arrangements of the sampling locations $\mathbf{s}_i, i \in \{1,\ldots,n\}$ (see Bailey Fig 5.18).

  - This leads to $\hat{\varepsilon}(\mathbf{s}) = \sum_{i=1}^{n} \hat{\lambda}_i(\mathbf{s}) \cdot \varepsilon(\mathbf{s}_i)$ which is structurally similar to the inverse distance weighted interpolator (Bailey & Gatrell p 156).
  - However, inverse distance weighted interpolator uses an externally given weights functions $\lambda_i(\mathbf{s})$.
  - In contrast, for Kriging the weights $\hat{\lambda}_i(\mathbf{s})$ are recalculated at each prediction location **s** based on the configuration of the sampling points.
- *Excurse (not test relevant):* For those interested in technical details, the coefficients $\hat{\lambda}_i(\mathbf{s})$ are estimated by minimizing the expected mean square error (OLS technique), which leads to the optimal BLUE estimator:

  $$E\left[(\hat{\varepsilon}(\mathbf{s}) - \varepsilon(\mathbf{s}))^2\right] = E\left[\hat{\varepsilon}^2(\mathbf{s})\right] + E\left[\varepsilon^2(\mathbf{s})\right] - 2 \cdot E\left[\hat{\varepsilon}(\mathbf{s}) \cdot \varepsilon(\mathbf{s})\right]$$

where

$$E\left[\hat{\varepsilon}^2(\mathbf{s})\right] = E\left[\sum_{i=1}^{n}\hat{\lambda}_i(\mathbf{s})\cdot\varepsilon(\mathbf{s}_i)\cdot\sum_{j=1}^{n}\hat{\lambda}_j(\mathbf{s})\cdot\varepsilon(\mathbf{s}_j)\right]$$

[a]
$$= \sum_{i=1}^{n}\sum_{j=1}^{n}\hat{\lambda}_i(\mathbf{s})\cdot\underbrace{C(\mathbf{s}_i,\mathbf{s}_j)}_{=E\left[\mathbf{s}_i\cdot\mathbf{s}_j\right]}\cdot\hat{\lambda}_j(\mathbf{s})$$

$$= \hat{\boldsymbol{\lambda}}^T(\mathbf{s})\cdot\mathbf{C}\cdot\hat{\boldsymbol{\lambda}}(\mathbf{s})$$

[b] $E\left[\varepsilon^2(\mathbf{s})\right] = \sigma^2$

$$E\left[\hat{\varepsilon}(\mathbf{s})\cdot\varepsilon(\mathbf{s})\right] = E\left[\left(\sum_{i=1}^{n}\hat{\lambda}_i(\mathbf{s})\cdot\varepsilon(\mathbf{s}_i)\right)\cdot\varepsilon(\mathbf{s})\right]$$

[c]
$$= \sum_{i=1}^{n}\hat{\lambda}_i(\mathbf{s})\cdot\underbrace{c(\mathbf{s}_i,\mathbf{s})}_{=E\left[\mathbf{s}_i\cdot\mathbf{s}\right]}$$

$$= \hat{\boldsymbol{\lambda}}^T(\mathbf{s})\cdot\mathbf{c}(\mathbf{s})$$

Thus $E\left[(\hat{u}(\mathbf{s})-u(\mathbf{s}))^2\right] = \hat{\boldsymbol{\lambda}}^T(\mathbf{s})\cdot\mathbf{C}\cdot\hat{\boldsymbol{\lambda}}(\mathbf{s}) + \sigma^2 - 2\cdot\hat{\boldsymbol{\lambda}}^T(\mathbf{s})\cdot\mathbf{c}(\mathbf{s})$

Setting the first derivative with regards to $\hat{\boldsymbol{\lambda}}(\mathbf{s})$ to zero gives

$$\frac{\partial E\left[(\hat{u}(\mathbf{s})-u(\mathbf{s}))^2\right]}{\partial\hat{\boldsymbol{\lambda}}(\mathbf{s})} = 2\cdot\mathbf{C}\cdot\hat{\boldsymbol{\lambda}}(\mathbf{s}) - 2\cdot\mathbf{c}(\mathbf{s}) \Rightarrow 0 = 2\cdot\mathbf{C}\cdot\hat{\boldsymbol{\lambda}}(\mathbf{s}) - 2\cdot\mathbf{c}(\mathbf{s})$$

$$\hat{\boldsymbol{\lambda}}(\mathbf{s}) = \mathbf{C}^{-1}\cdot\mathbf{c}(\mathbf{s})$$

(+++ End of Excurse +++)

- The minimum of this equation with respect to $\hat{\boldsymbol{\lambda}}(\mathbf{s})$ becomes $\hat{\boldsymbol{\lambda}}(\mathbf{s}) = \mathbf{C}^{-1} \cdot \mathbf{c}(\mathbf{s})$

  Consequently, in order to estimate the weights vector $\hat{\boldsymbol{\lambda}}(\mathbf{s})$ for any prediction location $\mathbf{s}$, only the covariance $\mathbf{c}(\mathbf{s}) = \left(c(\mathbf{s}_1, \mathbf{s}), c(\mathbf{s}_2, \mathbf{s}), \ldots, c(\mathbf{s}_n, \mathbf{s})\right)^T$ between that prediction location $\mathbf{s}$ and all the sample locations $\mathbf{s}_i, i \in \{1, \ldots, n\}$ needs to be updated using the semi-variogram function $\hat{\gamma}(h)$ at the observed distances $h = d(\mathbf{s}_i, \mathbf{s})$.

  The inverse covariance matrix $\mathbf{C}^{-1}$ remains fixed because it is solely based on the given sample locations $\mathbf{s}_i$.

- Note ArcGIS (also mentioned in Bailey & Gatrell, p 191) allows us to work with a subset of sampling points around the prediction location. Ideally this subset is defined by the range of the variogram.
  - Advantage: This reduces the dimensionality of the covariance matrix substantially. Thus the calculation of its inverse is readily performed.
  - Drawback: The inverse covariance matrix must be re-evaluated for each prediction point $\mathbf{s}$.

- Jointly [a] the covariance matrix among the sampling locations $\mathbf{s}_i$ and [b] the covariance vector of the sampling locations $\mathbf{s}_i$ with the prediction location $\mathbf{s}$ allows to calculate the predicted error component at location $\mathbf{s}$ by $\hat{\varepsilon}(\mathbf{s}) = \hat{\boldsymbol{\lambda}}^T(\mathbf{s}) \cdot \boldsymbol{\varepsilon} = \underbrace{\mathbf{c}^T(\mathbf{s}) \cdot \mathbf{C}^{-1}}_{=\hat{\boldsymbol{\lambda}}^T} \cdot \boldsymbol{\varepsilon}$

- By using the estimated weights vector $\hat{\boldsymbol{\lambda}}(\mathbf{s})$ we also get the prediction variance (or standard error) for each predicted value $\hat{\varepsilon}(\mathbf{s})$ at any locations $\mathbf{s}$

$$Var\left[\hat{\varepsilon}(\mathbf{s})\right] = E\left[(\hat{\varepsilon}(\mathbf{s}) - \varepsilon(\mathbf{s}))^2\right]$$
$$= \sigma^2 - \mathbf{c}^T(\mathbf{s}) \cdot \mathbf{C}^{-1} \cdot \mathbf{c}(\mathbf{s})$$

- Thus, in contrast to global spatial smoothing procedures, we have locally adaptive weights $\hat{\boldsymbol{\lambda}}(\mathbf{s})$ which are specific to each prediction location $\mathbf{s}$:
    - that are adjusted for redundant information at clustered sample locations and not just by the distance between the sampling points and the prediction location

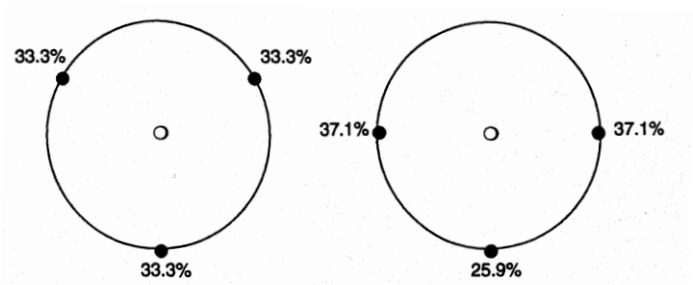


Figure 12.5: Three samples on a circle around the estimation location (on the left $\sigma^2_{OK} = .45\sigma^2$ and on the right $\sigma^2_{OK} = .48\sigma^2$).

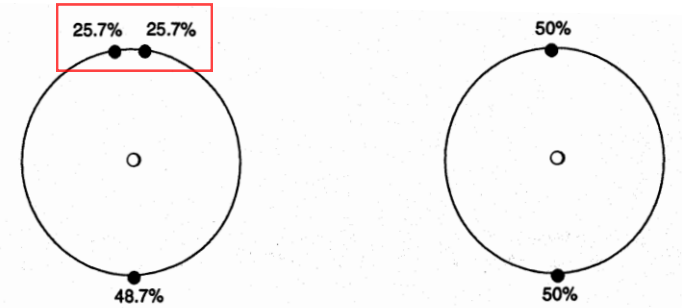


Figure 12.6: On the left, two samples have been set near to each other at the top ($\sigma^2_{OK} = .526\sigma^2$). On the right, the two upper samples have been merged into one single top sample ($\sigma^2_{OK} = .537\sigma^2$).

    - that take spatial autocorrelation in the neighborhood of the prediction location $\mathbf{s}$ explicitly into account.

If spatial autocorrelation is not present in the sample data the weights would reduce to zero, i.e., $\hat{\boldsymbol{\lambda}}(\mathbf{s})$=0 because $\mathbf{c}(\mathbf{s}) = 0$ and the covariance matrix $\mathbf{C}$ becomes diagonal.

- that leads to *statistical* estimates that provide *prediction standard errors* and thus give an indication of the uncertainty of prediction at the prediction locations.
- For prediction locations outside the "range" to any sampling location, autocorrelation is not present.
  $\Rightarrow$ the weights will be zero and the predicted error term will be $\hat{\varepsilon}(\mathbf{s}) = 0$.
  Consequently, for extrapolation predictions, autocorrelation in the stochastic error terms becomes irrelevant and only the first order component matters.

- Example of simple Kriging in Bailey and Gatrell p 186-188
  - Recall the relationship $C(h) = \sigma^2 - \gamma(h)$.
    An exponential semi-variogram has been assumed with given parameters
    $\gamma(h) = 20 \cdot \left(1 - \exp(-3 \cdot h/100)\right) \Leftrightarrow c(h) = 20 \cdot \exp(-3 \cdot h/100)$
  - Discuss the values in $\mathbf{C}$ and $\mathbf{c}(\mathbf{s})$ in dependence of the distance *h*.
  - Discuss the weights $\hat{\boldsymbol{\lambda}}(\mathbf{s})$ in terms of the redundant information in adjacent sampling sites $\mathbf{s}_i$.
    Note: The weights are not just a simple function of distance!

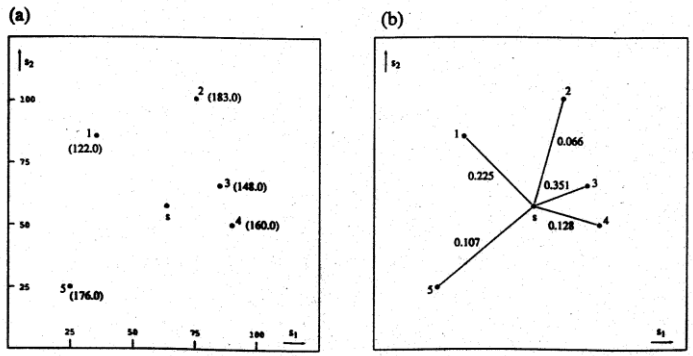


Fig. 5.19 An illustration of simple kriging

$$C(h) = 20e^{-3h/100}$$

C

Sites	s_1	s_2	s_3	s_4	s_5
s_1	20.00	4.571	3.970	2.828	3.228
s_2	4.571	20.00	5.970	3.739	1.086
s_3	3.970	5.970	20.00	12.45	2.300
s_4	2.828	3.739	12.45	20.00	2.479
s_5	3.228	1.086	2.300	2.479	20.00

$c(s)$

Sites	s
s_1	6.895
s_2	5.032
s_3	10.15
s_4	8.083
s_5	4.079

Sites	$\lambda(s)$
s_1	0.225
s_2	0.066
s_3	0.351
s_4	0.128
s_5	0.107

Sites	y
s_1	122.0
s_2	183.0
s_3	148.0
s_4	160.0
s_5	176.0

Sites	u
s_1	−38.37
s_2	22.63
s_3	−12.37
s_4	−0.37
s_5	15.63

book: $\mu(s) = 160.37$

data: $\mu(s) = 157.8$

Our estimate $\hat{u}(s)$ is now:

$$\begin{aligned}\hat{u}(s) &= 0.225 \times -38.37 + 0.066 \times 22.63 + 0.351 \times -12.37 \\ &\quad + 0.128 \times -0.37 + 0.107 \times 15.63 \\ &\simeq -9.86\end{aligned}$$

$$\hat{y}(s) = 160.37 - 9.86 \simeq 150$$

standard deviation

$$\begin{aligned}\sigma_e^2 &= \sigma^2 - c^T(s)C^{-1}c(s) \\ &= 20.0 - 6.92 \\ &= 13.08\end{aligned}$$

$\sqrt{13.08} = 3.62$

$1.96 \cdot 3.62 = 7.09$

approximate 95% confidence interval of 150.5 ± 7.09.

upper band of normal distribution

Regression Simple Kriging Spatial Prediction Procedure with Exogenous Trend

- Estimation, prediction and mean square prediction error are based on the (iterative) feasible general least estimator.
- The basic estimates for $\hat{\boldsymbol{\beta}}_{gls}$ and $\hat{\mathbf{C}}$ are obtained interactively in several steps:
  1. Estimate $\hat{\boldsymbol{\beta}}$ by OLS
  2. Estimate the error components $\hat{\varepsilon}(\mathbf{s}_i)$ (residuals) at the sample locations from previous $\hat{\boldsymbol{\beta}}$ estimates (either step [1] or step [5]).
  3. Calibrate the semi-variogram $\hat{\gamma}(h)$ on these residuals $\hat{\varepsilon}(\mathbf{s}_i)$
  4. Use the semi-variogram $\hat{\gamma}(h)$ to estimate the covariance matrix $\hat{\mathbf{C}}$.
  5. Re-estimate $\hat{\boldsymbol{\beta}}_{gls}$ using general least squares $\hat{\boldsymbol{\beta}}_{gls} \equiv (\mathbf{X}^T \cdot \hat{\mathbf{C}}^{-1} \cdot \mathbf{X})^{-1} \cdot \mathbf{X}^T \cdot \hat{\mathbf{C}}^{-1} \cdot \mathbf{y}(\mathbf{s})$
- Steps 2 through 5 can be repeated by constantly updating $\hat{\boldsymbol{\beta}}_{gls}$, $\hat{\varepsilon}(\mathbf{s}_i)$, $\hat{\gamma}(h)$, and $\hat{\mathbf{C}}$ until these estimators converge at a stable value.
- The predicted value at any location is (Bailey p 189)

$$\hat{Y}(\mathbf{s}) = \mathbf{x}^T(\mathbf{s}) \cdot \hat{\boldsymbol{\beta}}_{gls} + \hat{\varepsilon}(\mathbf{s})$$
$$= \mathbf{x}^T(\mathbf{s}) \cdot \hat{\boldsymbol{\beta}}_{gls} + \mathbf{c}^T(\mathbf{s}) \cdot \mathbf{C}^{-1} \cdot \hat{\boldsymbol{\varepsilon}}(\mathbf{s})$$

- The prediction standard error is now composed of two additively linked terms (see **Bailey bottom of p 189**):

- The first expression is the *error variance* associated with the prediction of the *first order term* through GLS (e.g., the GLS trend surface estimate)
  Under *spatial independence* it would have been simply the prediction error of OLS
  $Var(\hat{Y}(\mathbf{s})) = \hat{\sigma}^2 \cdot \left(1 + \mathbf{x}^T(\mathbf{s}) \cdot (\mathbf{X}^T \cdot \mathbf{X})^{-1} \cdot \mathbf{x}(\mathbf{s})\right)$
- The second term is the variance of the spatially autocorrelated second order error term.
  $Var\left[\hat{\varepsilon}(\mathbf{s})\right] = \hat{\sigma}^2 - \hat{\mathbf{c}}^T(\mathbf{s}) \cdot \hat{\mathbf{C}}^{-1} \cdot \hat{\mathbf{c}}(\mathbf{s})$

- Example of simple Kriging in Bailey and Gatrell p 190 (continued)
  - Note that $\mu_{ols} = 157.8$ but $\mu_{gls} = 160.4$
  - The predicted value is at location $\mathbf{s}$ is $\hat{Y}(\mathbf{s}) = 160.4 - 9.86 = 150.5$
  - The variance of the predicted value is now the variance of the first order term and the variance of the second term of the two components $Var(\hat{Y}(\mathbf{s})) = 0.107 + 13.08 = 13.19$

variance = variance from OLS and error term and sum them together, then square root from them.

Ordinary kriging

- ***Focuses not*** on the errors $\mathbf{u}$ but on the regionalized variable $\mathbf{Y}$ itself.
- Assumes a ***constant but unknown*** first order component μ_{gls}.
- ***Intrinsic stationarity*** based on constant mean levels μ_{gls} still allows to estimate the semi-variogram.
- The estimation equation of ordinary kriging is $\hat{Y}(\mathbf{s}) = \sum_{i=1}^{n} \omega_i(\mathbf{s}) \cdot Y(\mathbf{s}_i)$.

- Note: we change for ordinary and universal Kriging the notation for the weights from $\lambda_i(\mathbf{s})$ to $\omega_i(\mathbf{s})$.
- Because the ***expected value*** μ_{gls} for $\hat{Y}(\mathbf{s})$ any prediction location has to be equal to the expected values for any observed $Y_i(\mathbf{s})$, the estimated weights need to satisfy the constraint $\sum_{i=1}^{n}\hat{\omega}_i(\mathbf{s})=1$, that is,

$$\mu_{gls}=\sum_{i=1}^{n}\omega_i(\mathbf{s})\cdot\underbrace{E\left[Y(\mathbf{s}_i)\right]}_{\mu_{gls}}=\sum_{i=1}^{n}\omega_i(\mathbf{s})\cdot\mu_{gls} \text{ only if } \sum_{i=1}^{n}\hat{\omega}_i(\mathbf{s})=1$$

Estimation of the weights under this ***constraint*** $\sum_{i=1}^{n}\hat{\omega}_i(\mathbf{s})=1$ is performed by the method of ***Lagrange multipliers*** based on the augmented covariance matrix $\mathbf{C}_+$ and augmented covariance vector $\mathbf{c}_+(\mathbf{s})$:

$$\begin{pmatrix} c(\mathbf{s}_1,\mathbf{s}_1) & \cdots & c(\mathbf{s}_1,\mathbf{s}_n) & 1 \\ \vdots & \ddots & \vdots & \vdots \\ c(\mathbf{s}_n,\mathbf{s}_1) & \cdots & c(\mathbf{s}_n,\mathbf{s}_n) & 1 \\ 1 & \cdots & 1 & 0 \end{pmatrix}\cdot\begin{pmatrix} \omega_1(\mathbf{s}) \\ \vdots \\ \omega_n(\mathbf{s}) \\ v(\mathbf{s}) \end{pmatrix}=\begin{pmatrix} c(\mathbf{s}_1,\mathbf{s}) \\ \vdots \\ c(\mathbf{s}_1,\mathbf{s}) \\ 1 \end{pmatrix}$$

$$\mathbf{C}_+ \quad\cdot\quad \omega_+(\mathbf{s}) \quad=\quad \mathbf{c}_+(\mathbf{s})$$

- The advantage of ordinary kriging is that we do *not* need to know μ_{gls} *a priori* and thus can skip the step of GLS intercept model estimation, because μ_{gls} is constant for all observations.
- In addition, the estimators can be expressed directly in terms of the semi-variogram matrix instead of using the covariance matrix, that is,

$$\hat{\boldsymbol{\omega}}_+(\mathbf{s})=\mathbf{C}_+^{-1}\cdot\mathbf{c}_+(\mathbf{s})$$
$$=\boldsymbol{\Gamma}_+^{-1}\cdot\boldsymbol{\gamma}_+(\mathbf{s})$$

Therefore, estimation can be performed in terms of the semi-variogram and not just in terms of the covariance.

Universal kriging

- ***Universal kriging*** is a simple extension of Ordinary kriging by incorporating more than just the constant vector of ones into the model covariance matrix $\mathbf{C}_+$.

$$\begin{pmatrix} c(\mathbf{s}_1,\mathbf{s}_1) & \cdots & c(\mathbf{s}_1,\mathbf{s}_n) & 1 & \mathbf{x}^T(\mathbf{s}_1) \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ c(\mathbf{s}_n,\mathbf{s}_1) & \cdots & c(\mathbf{s}_n,\mathbf{s}_n) & 1 & \mathbf{x}^T(\mathbf{s}_n) \\ 1 & \cdots & 1 & 0 & \mathbf{0} \\ \mathbf{x}(\mathbf{s}_1) & \cdots & \mathbf{x}(\mathbf{s}_n) & \mathbf{0} & \mathbf{0} \end{pmatrix} \cdot \begin{pmatrix} \omega_1(\mathbf{s}) \\ \vdots \\ \omega_n(\mathbf{s}) \\ v_1(\mathbf{s}) \\ \boldsymbol{v}_x(\mathbf{s}) \end{pmatrix} = \begin{pmatrix} c(\mathbf{s}_1,\mathbf{s}) \\ \vdots \\ c(\mathbf{s}_1,\mathbf{s}) \\ 1 \\ \mathbf{x}(\mathbf{s}) \end{pmatrix}$$

for slope

- In ordinary kriging the ***constant vector of ones*** models the intercept.
- In universal kriging locational information, such as the coordinates for trend-surface modeling, are included to capture the first order spatial trend component.
- It leads again to location dependent weight estimates $\omega_i(\mathbf{s})$ which allow for predicting the regionalized variable at that location

  $\hat{Y}(\mathbf{s}) = \sum_{i=1}^{n} \omega_i(\mathbf{s}) \cdot Y(\mathbf{s}_i)$
- However, it provides ***no means*** for the interpretation of the trend-surface or the impact of any spatially distributed covariable.

- A ***major problem*** of universal kriging is that the ***underlying semi-variogram*** is estimated from the observed $Y(\mathbf{s})$ that include a ***non-stationary*** trend component $\mu_{gls}(\mathbf{s}_i)$ and not on the stationary regression residuals.

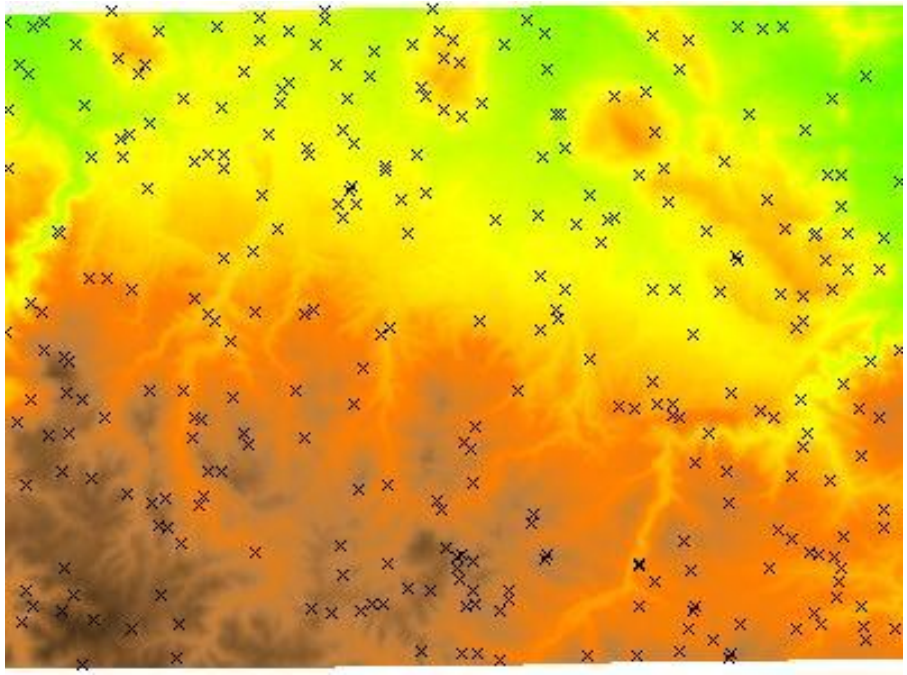
Discussion

- As trend-surface analysis, Kriging depends on applied coordinate system and the ***spherical distortions*** needs to be taken into account for larger study areas.
- ***Extrapolation*** outside the study area is unreliable (one can see this because the prediction standard error increases rapidly).
  However, for simple and ordinary Kriging, the predicted value becomes constant outside the study area.
- ***Semi-variogram calibration*** depends crucially on ***stationarity assumptions***. If these stationarity assumptions are violated the estimates become unreliable.
- In contrast to deterministic spatial smoothing estimators, the estimated kriging weights $\lambda_i(\mathbf{s})$ or $\omega_i(\mathbf{s})$ are optimal in the sense of minimizing the mean square error.
  They accommodate spatial autocorrelation and data redundancies, which arises from their spatial configuration.
- Kriging estimators also provide a measure of ***uncertainty*** for the predictions by means of the Kriging variance (prediction standard error).
  This is an outstanding feature of Kriging and distinguishes it from deterministic spatial interpolators.

do not have more redundant points and pay attention to edge effects

- The Kriging variance depends on the ***spatial configuration*** of the sample points. Thus the design of the selected sampling locations becomes important.
- The **GEOSTATISTICAL ANALYSIS** does not incorporate the ***iterative GLS*** least squares estimation for the first order component that is described in Bailey and Gatrell p. 189.
- For practical purposes, frequently only ***local search neighborhood*** information is used in estimating the weights. (see tutorial of the **GEOSTATISTICAL ANALYST** p 33).
  This circumvents the ***inversion*** of large co-variance or semi-variance matrices but has problems for universal kriging because the first order component is estimated localized.
  The ***semi-variogram***, however, is always ***estimated globally*** by using all available sample observations.
- ***Cross-validation*** helps to evaluate the fit of the Kriging model by comparing the observed regionalized variables with their predicted counterparts, which are based on a model that excludes the prediction location and then estimate the predicted value from a model with the reduced data-set.
- Universal Kriging is a good method for doing ***predictions***. However, it is not good for ***theoretical guided modeling***, because we will not have control over exogenous information and cannot perform tests on model parameters.

R Example with Library gstat



- See -script KrigeWithGStat.r and Elev.csv data

Block kriging (not test relevant)

- Rather than producing point predictions, block Kriging aims at estimating an average of a continuous variable $Y(\mathrm{A})$ either for an area or a volume A .

- Naïve approach: perform many estimates at $\mathbf{s}_i \in \mathrm{A}$ within the block A and then integrate over them to obtain the average
- Drawback of the naïve approach is the large numerical effort because we would need sufficient number of estimates for each block.

Example: Ordinary block kriging

- Under ordinary kriging where we assume that the expectation of $E[Y(s)] = const$ with $\mu(\mathbf{s}) = \mu \quad \forall\, \mathbf{s} \in \mathrm{R}$
- The augmented kriging systems enforcing the constraint $\sum_{i=1}^{n} \omega_i(\mathbf{s}) = 1$ by the Lagrange multiplier $\nu(\mathbf{s})$ is

$$\underset{\mathbf{C}_+}{\begin{pmatrix} C(s_1,s_1) & \cdots & C(s_1,s_n) & 1 \\ \vdots & \ddots & \vdots & \vdots \\ C(s_n,s_1) & \cdots & C(s_1,s_1) & 1 \\ 1 & \cdots & 1 & 0 \end{pmatrix}} \cdot \underset{\boldsymbol{\omega}_+(\mathbf{s})}{\begin{pmatrix} \omega_1(\mathbf{s}) \\ \vdots \\ \omega_n(\mathbf{s}) \\ \nu(\mathbf{s}) \end{pmatrix}} = \underset{c_+(\mathbf{s})}{\begin{pmatrix} c_1(\mathbf{s}) \\ \vdots \\ c_n(\mathbf{s}) \\ 1 \end{pmatrix}}$$

- The prediction at location s is obtained by $\hat{Y}(\mathbf{s}) = \omega^T(\mathbf{s}) \cdot \mathbf{Y}$ with the error variance

$$\sigma^2(\mathbf{s}) = \underbrace{C(\mathbf{s},\mathbf{s})}_{\sigma^2} - \mathbf{c}_+^T(\mathbf{s}) \cdot \mathbf{C}_+^{-1} \cdot \mathbf{c}_+$$

- The key is to replace the estimates $c_i(\mathbf{s})$ and $C(\mathbf{s}_i,\mathbf{s}_j)$ by their block estimates

$$c(\mathsf{A},\mathbf{s}_i)=\frac{\int_{\mathsf{A}} c(\mathbf{s},\mathbf{s}_i)\cdot d\mathbf{s}}{|\mathsf{A}|} \quad i\in\{1,\ldots,n\} \text{ and}$$

$$C(\mathsf{A}_i,\mathsf{A}_j)=\frac{\int_{\mathsf{A}_i}\int_{\mathsf{A}_j} c(\mathbf{s}_i,\mathbf{s}_j)\cdot d\mathbf{s}_i\cdot d\mathbf{s}_j}{|\mathsf{A}_i|\cdot|\mathsf{A}_j|}$$

- The integrals must be evaluated by numerical integration or approximated by the average of sufficient sampling point co-variances within the blocks A .
- While evaluation of the covariances is still resource demanding it, nevertheless, requires less calculations than estimating several predicted values $\hat{y}(\mathbf{s})$ with $\mathbf{s}\in\mathsf{A}$ and the averaging them.

Co-Kriging (not test relevant)

- In simple, ordinary and universal kriging only one ***primary*** variable $Y(\mathbf{s}^Y)$ was observed at several sampling locations $\mathbf{s}_i^Y \; i\in\{1,\ldots,n^Y\}$
- Information of ***secondary*** variables $X(\mathbf{s}^X)$ at identical or different sampling locations $\mathbf{s}_j^X \; j\in\{1,\ldots,m^X\}$ may improve within the neighborhood of the primary variable the prediction
- The cross-variogram must indicate correlation between of the sets $Y(\mathbf{s}^Y)$ and $X(\mathbf{s}^X)$ at some distance bands.

- The cross-covariance between $Y(\mathbf{s}^Y)$ and $X(\mathbf{s}^X)$ under the assumptions that both variables have constant means μ_Y and μ_X, respectively, over the whole study area R is
  $C_{YX}(\mathbf{h}) = E\left[\left(Y(\mathbf{s}+\mathbf{h}) - \mu_Y\right)\left(X(\mathbf{s}) - \mu_X\right)\right]$
- This cross-covariance is not symmetric, i.e., $C_{YX}(\mathbf{h}) \neq C_{XY}(\mathbf{h})$.
- In contrast, its cross-variogram estimator assumes symmetry by its definition. The cross-variogram estimator becomes
  $2\gamma_{YX}(h) = E\left[\left(Y(\mathbf{s}+\mathbf{h}) - Y(\mathbf{s})\right)\left(X(\mathbf{s}+\mathbf{h}) - X(\mathbf{s})\right)\right]$
- The sample cross-variogram can be estimated using those pairs of observations of the primary and secondary variable that are given the same locations $\mathbf{s}_i^X = \mathbf{s}_j^Y$
  $2\hat{\gamma}_{YX}(h) = \frac{1}{n(h)} \sum_{|\mathbf{s}_1 - \mathbf{s}_2| = h} \left(Y(\mathbf{s}_1) - Y(\mathbf{s}_2)\right)\left(X(\mathbf{s}_1) - X(\mathbf{s}_2)\right)$
- Note that
  - this expression can be generalized for primary and secondary variables at different but adjacent locations
  - that the distance h empirical becomes a for anisotropic models at directed distance block and for isotropic model at distance bin.
  - $n(h)$ is the number of pairs in the distance bin.
- Practically the cross-dissimilarity products are plotted against the distances h for isotropic models and then a cross-variogram function is fitted thru these cross-dissimilarity products

- The semi-variogram models for the primary variable $Y(\mathbf{s}^Y)$ and secondary variables $X(\mathbf{s}^X)$ as well as the cross-semivariogram $\gamma_{YX}(\mathbf{h})$ must satisfy the condition that the resulting joint covariance matrix

$$\mathbf{C} = \begin{pmatrix} \mathbf{C}_Y & \mathbf{C}_{YX} \\ \mathbf{C}_{XY} & \mathbf{C}_X \end{pmatrix}$$

  is positive definite, which implies that identical functional forms of the variogram functions, which may however differ in their specification, must be used for all components of the covariance matrix.

- The co-kriging estimator for the ordinary kriging model is
$\hat{Y}(\mathbf{s}) = \sum_i \omega_i^Y(\mathbf{s}) \cdot Y(\mathbf{s}_i^Y) + \sum_j \omega_j^X(\mathbf{s}) \cdot X(\mathbf{s}_j^X)$

- As for ordinary kriging, the expected value at each prediction location must be $E\left[\hat{Y}(\mathbf{s})\right] = \mu_Y$ and thus the weights $\omega_i^Y(\mathbf{s})$ have to satisfy the constraint $\sum_i \omega_i^Y(\mathbf{s}) = 1$.
  Furthermore, the weights $\omega_j^X(\mathbf{s})$ must be constraint to $\sum_j \omega_j^X(\mathbf{s}) = 0$ so that no additional components are added to $E\left[\hat{Y}(\mathbf{s})\right] = \mu_Y$ and $E\left[\hat{Y}(\mathbf{s})\right]$ remains unbiased.

- The ordinary co-kriging estimator becomes $\mathbf{C}_+ \cdot \boldsymbol{\omega}_+(\mathbf{s}) = \mathbf{c}_+(\mathbf{s})$
  where the Lagrange augmented cross-covariance matrix $\mathbf{C}_+$ becomes

$$\mathbf{C}_+ = \begin{pmatrix} c_Y(\mathbf{s}_1^Y,\mathbf{s}_1^Y) & \cdots & c_Y(\mathbf{s}_1^Y,\mathbf{s}_{n^Y}^Y) & c_{YX}(\mathbf{s}_1^Y,\mathbf{s}_1^X) & \cdots & c_{YX}(\mathbf{s}_1^Y,\mathbf{s}_{n^X}^X) & 1 & 0 \\ \vdots & \ddots & \vdots & \vdots & & \vdots & \vdots & \vdots \\ c_Y(\mathbf{s}_{n^Y}^Y,\mathbf{s}_1^Y) & \cdots & c_Y(\mathbf{s}_{n^Y}^Y,\mathbf{s}_{n^Y}^Y) & c_{YX}(\mathbf{s}_{n^Y}^Y,\mathbf{s}_1^X) & \cdots & c_{YX}(\mathbf{s}_{n^Y}^Y,\mathbf{s}_{n^X}^X) & 1 & 0 \\ c_{XY}(\mathbf{s}_1^X,\mathbf{s}_1^Y) & \cdots & c_{XY}(\mathbf{s}_1^X,\mathbf{s}_{n^Y}^Y) & c_X(\mathbf{s}_1^X,\mathbf{s}_1^X) & \cdots & c_X(\mathbf{s}_1^X,\mathbf{s}_{n^X}^X) & 0 & 1 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ c_{XY}(\mathbf{s}_{n^X}^X,\mathbf{s}_1^Y) & \cdots & c_{XY}(\mathbf{s}_{n^X}^X,\mathbf{s}_{n^Y}^Y) & c_X(\mathbf{s}_{n^X}^X,\mathbf{s}_1^X) & \cdots & c_X(\mathbf{s}_{n^X}^X,\mathbf{s}_{n^X}^X) & 0 & 1 \\ 1 & \cdots & 1 & 0 & \cdots & 0 & 0 & 0 \\ 0 & \cdots & 0 & 1 & \cdots & 1 & 0 & 0 \end{pmatrix}$$

the augmented vector $\boldsymbol{\omega}_+(\mathbf{s})$ becomes

$$\boldsymbol{\omega}_+(\mathbf{s}) = \left(\omega_1^Y(\mathbf{s}),\cdots,\omega_{n^Y}^Y(\mathbf{s}),\omega_1^X(\mathbf{s}),\cdots,\omega_{n^X}^X(\mathbf{s}),\nu^Y(\mathbf{s}),\nu^X(\mathbf{s})\right)^T$$

and the covariance vector $\mathbf{c}_+(\mathbf{s})$ becomes

$$\mathbf{c}_+(\mathbf{s}) = \left(c_Y(\mathbf{s},\mathbf{s}_1^Y),\cdots,c_Y(\mathbf{s},\mathbf{s}_{n^Y}^Y),c_{YX}(\mathbf{s},\mathbf{s}_1^X),\cdots,c_{YX}(\mathbf{s},\mathbf{s}_{n^X}^X),1,0\right)^T$$

- The co-kriging weights become $\boldsymbol{\omega}_+(\mathbf{s}) = \mathbf{C}_+^{-1}\cdot\mathbf{c}_+(\mathbf{s})$ and the estimated prediction error is

$$\sigma^2(\mathbf{s}) = \underbrace{c_y(\mathbf{s},\mathbf{s})}_{\sigma_Y^2} - \mathbf{c}_+^T(\mathbf{s})\cdot\mathbf{C}_+^{-1}\cdot\mathbf{c}_+(\mathbf{s})$$

Discussion:

- Extensions are possible for a set of several secondary variables

- A secondary variable $X(\mathbf{s})$ will not improve the prediction if the cross-covariance $c_{YX}(\mathbf{h})$ is zero for all distances.
  In fact, due to the increased model complexity with many additional prior estimations the prediction quality may diminish.
- Co-kriging becomes infeasible if the secondary variable $X(\mathbf{s})$ is based on a large number of sampling locations m^X because the augmented covariance matrix $\mathbf{C}_+$ becomes large and difficult to invert. Local neighborhood approaches should be used here.
- In case the locations of the primary and secondary variable are identical,
  that is, $\mathbf{s}_i^X = \mathbf{s}_j^Y$ for all $\mathbf{s}_i^Y \; i \in \{1,\ldots,n^Y\}$ and $\mathbf{s}_i^X \; i \in \{1,\ldots,n^X\}$ with $n^Y = m^X$,
  and the cross-covariance is proportional to the autocovariance, that is,
  $c_{YX}(\mathbf{h}) = \varphi \cdot c_Y(\mathbf{h})$ for all distances **h**, then nothing is gained by co-kriging.
- Co-kriging can be generalized to universal co-kriging and block co-kriging.