Introduction

Data: set of n attribute measurements $\{z(\mathbf{s}_i), i = 1, \ldots, n\}$, available at n sample locations $\{\mathbf{s}_i, i = 1, \ldots, n\}$

Objectives: (i) predict or interpolate unknown attribute value $z(\mathbf{s}_p)$ at location $\mathbf{s}_p$ from the n sample data, and (ii) assess reliability of predicted value

Slide 1

Geostatistical spatial interpolation:

- predicted attribute value = weighted linear combination of sample data values + attribute mean, if known (non-linear methods also exist)
- a semivariogram model is used to determine the weights, which account for:
  - spatial auto-correlation between sample data and unknown value
  - spatial auto-correlation between sample data themselves (data redundancy)
- in addition, and contrary to most interpolation algorithms, geostatistics offers a measure of reliability (prediction error variance) regarding the attribute prediction

Simple Kriging (SK)

SK prediction: $$\hat{z}(\mathbf{s}_p) = m + \sum_{i=1}^{n} w_p(\mathbf{s}_i)[z(\mathbf{s}_i) - m] = \mathbf{w}_p^T \mathbf{r}$$

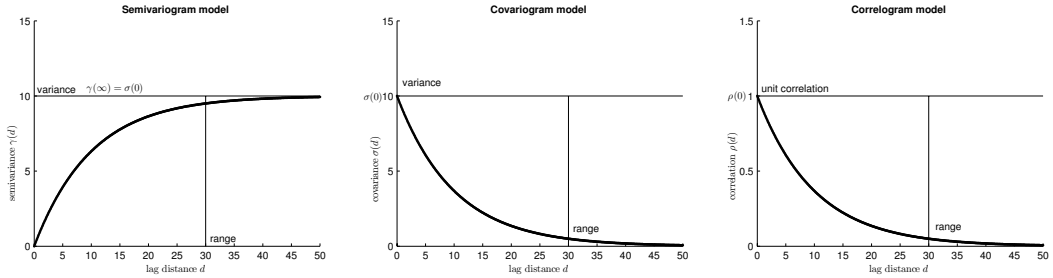
- $\mathbf{w}_p = [w_p(\mathbf{s}_i), i = 1, \ldots, n]^T$: $(n \times 1)$ vector of SK-weights assigned to n sample data for prediction at location $\mathbf{s}_p$; superscript T denotes transposition
- $\mathbf{r} = [z(\mathbf{s}_i) - m, i = 1, \ldots, n]^T$: $(n \times 1)$ vector of residual data from *known* mean m

Slide 2

$$\hat{z}(\mathbf{s}_p) = m + \underbrace{\begin{bmatrix} w_p(\mathbf{s}_1) & \cdots & w_p(\mathbf{s}_i) & \cdots & w_p(\mathbf{s}_n) \end{bmatrix}}_{\mathbf{w}_p^T} \underbrace{\begin{bmatrix} z(\mathbf{s}_1) - m \\ \vdots \\ z(\mathbf{s}_i) - m \\ \vdots \\ z(\mathbf{s}_n) - m \end{bmatrix}}_{\mathbf{r}}$$

use semivariogram model to determine weights at each prediction location; typically, it is the covariogram model that is used due to computational reasons

Semivariogram / Covariogram / Correlogram Model



Slide 3

Conversion between models, with $\sigma(0) = \gamma(\infty)$ being the sill of the semivariogram model:

- Semivariogram $\rightarrow$ covariogram: $\boxed{\sigma(d) = \sigma(0) - \gamma(d)}$
- Covariogram $\rightarrow$ correlogram: $\boxed{\rho(d) = \dfrac{\sigma(d)}{\sigma(0)}}$
- Semivariogram $\rightarrow$ correlogram: $\boxed{\rho(d) = 1 - \dfrac{\gamma(d)}{\sigma(0)}}$
- Covariogram $\rightarrow$ semivariogram: $\boxed{\gamma(d) = \sigma(0) - \sigma(d)}$

Requisites for Geostatistical Interpolation I

Data-to-data and data-to-unknown distances:

$$\mathbf{D} = \begin{bmatrix} 0 & \cdots & d_{1j} & \cdots & d_{1n} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ d_{i1} & \cdots & 0 & \cdots & d_{in} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ d_{n1} & \cdots & d_{nj} & \cdots & 0 \end{bmatrix} \quad \text{and} \quad \mathbf{d}_p = \begin{bmatrix} d_{1p} \\ \vdots \\ d_{ip} \\ \vdots \\ d_{np} \end{bmatrix}$$

Slide 4

Comments:

- as any other interpolation method, one accounts for the proximity of the n sample locations to the prediction location $\mathbf{s}_p$
  Note: Vector $\mathbf{d}_p$ changes from one prediction location $\mathbf{s}_p$ to another, hence the subscript p
- unlike other interpolation methods, one also accounts for the proximity between sample locations themselves (sample configuration or data layout)
  Note: Matrix $\mathbf{D}$ of sample-to-sample distances is the same for all prediction locations

Requisites for Geostatistical Interpolation II

From distance matrices to model covariance matrices: Take any distance value d_{ij} and d_{ip}, i.e., any entry in $\mathbf{D}$ and $\mathbf{d}_p$, and transform it, via the covariogram model, to a covariance value $\sigma(d_{ij})$ and $\sigma(d_{ip})$

Data-to-data and data-to-unknown model covariances:

Slide 5

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma(0) & \cdots & \sigma(d_{1j}) & \cdots & \sigma(d_{1n}) \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \sigma(d_{i1}) & \cdots & \sigma(0) & \cdots & \sigma(d_{in}) \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \sigma(d_{n1}) & \cdots & \sigma(d_{nj}) & \cdots & \sigma(0) \end{bmatrix} \quad \text{and} \quad \boldsymbol{\sigma}_p = \begin{bmatrix} \sigma(d_{1p}) \\ \vdots \\ \sigma(d_{ip}) \\ \vdots \\ \sigma(d_{np}) \end{bmatrix}$$

- data-to-data covariance matrix $\boldsymbol{\Sigma}$: $(n \times n)$ matrix with *model* covariance values $\sigma(d_{ij})$ between any two sample locations separated by distance d_{ij}
- data-to-unknown covariance vector $\boldsymbol{\sigma}_p$: $(n \times 1)$ vector with *model* covariance values $\sigma(d_{ip})$ between the n sample locations and the prediction location $\mathbf{s}_p$
  Note: Vector $\boldsymbol{\sigma}_p$ changes from one prediction location $\mathbf{s}_p$ to another, hence the subscript p

Requisites for Geostatistical Interpolation III

Data-to-data and data-to-unknown model covariances:

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma(0) & \cdots & \sigma(d_{1j}) & \cdots & \sigma(d_{1n}) \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \sigma(d_{i1}) & \cdots & \sigma(0) & \cdots & \sigma(d_{in}) \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \sigma(d_{n1}) & \cdots & \sigma(d_{nj}) & \cdots & \sigma(0) \end{bmatrix} \quad \text{and} \quad \boldsymbol{\sigma}_p = \begin{bmatrix} \sigma(d_{1p}) \\ \vdots \\ \sigma(d_{ip}) \\ \vdots \\ \sigma(d_{np}) \end{bmatrix}$$

Slide 6

Comments:

- data-to-data covariance matrix $\boldsymbol{\Sigma}$: encapsulates the redundancy between the sample data; for positive spatial auto-correlation, the more clustered is the sample layout, the more redundant are the sample data (less information content);
  a clustered sample layout typically translates into larger entries in $\boldsymbol{\Sigma}$
- data-to-unknown covariance vector $\boldsymbol{\sigma}_p$: encapsulates the statistical proximity (correlation) between the sample data and the unknown attribute value $z(\mathbf{s}_p)$ at the prediction location $\mathbf{s}_p$; that correlation is a function of distance between sample and prediction locations, not of the actual (unknown) value $z(\mathbf{s}_p)$;
  The larger the entries of vector $\boldsymbol{\sigma}_p$, the stronger the predictive power of sample data

Simple Kriging (SK) System & Weights

$$\begin{bmatrix} \sigma(0) & \cdots & \sigma(d_{1n}) \\ \vdots & \ddots & \vdots \\ \sigma(d_{n1}) & \cdots & \sigma(0) \end{bmatrix} \begin{bmatrix} w_p(\mathbf{s}_1) \\ \vdots \\ w_p(\mathbf{s}_n) \end{bmatrix} = \begin{bmatrix} \sigma(d_{1p}) \\ \vdots \\ \sigma(d_{np}) \end{bmatrix}$$

$$\boxed{\boldsymbol{\Sigma}\mathbf{w}_p = \boldsymbol{\sigma}_p}$$

Comments:

Slide 7

- the SK system is a (disguised) version of the normal equations for the case of regression with no intercept term: $\mathbf{X}^T\mathbf{X}\mathbf{b} = \mathbf{X}^T\mathbf{y}$, where $\mathbf{X}$ is the design matrix and $\mathbf{y}$ is the vector of data on the dependent variable; in regression, the data-to-data covariance is estimated as $\mathbf{X}^T\mathbf{X}/n$, and the data-to-unknown covariance as $\mathbf{X}^T\mathbf{y}/n$
- the weights vector $\mathbf{w}_p$ is obtained by solving the SK system, as $\mathbf{w}_p = \boldsymbol{\Sigma}^{-1}\boldsymbol{\sigma}_p$, *anew* at each prediction location $\mathbf{s}_p$ since the entries of $\boldsymbol{\sigma}_p$ change
- entries of $\mathbf{w}_p$ *do not depend* on data values or on sill, $\sigma(0)$, of covariogram model:

$$\sigma(0)\begin{bmatrix} \rho(0) & \cdots & \rho(d_{1n}) \\ \vdots & \ddots & \vdots \\ \rho(d_{n1}) & \cdots & 1 \end{bmatrix} \begin{bmatrix} w_p(\mathbf{s}_1) \\ \vdots \\ w_p(\mathbf{s}_n) \end{bmatrix} = \sigma(0)\begin{bmatrix} \rho(d_{1p}) \\ \vdots \\ \rho(d_{np}) \end{bmatrix}$$

Interpreting the Simple Kriging Weights

$$\begin{bmatrix} w_p(\mathbf{s}_1) \\ \vdots \\ w_p(\mathbf{s}_n) \end{bmatrix} = \frac{1}{\sigma(0)} \begin{bmatrix} 1 & \cdots & \rho(d_{1n}) \\ \vdots & \ddots & \vdots \\ \rho(d_{n1}) & \cdots & 1 \end{bmatrix}^{-1} \sigma(0) \begin{bmatrix} \rho(d_{1p}) \\ \vdots \\ \rho(d_{np}) \end{bmatrix} \Rightarrow \boxed{\mathbf{w}_p = \boldsymbol{\Sigma}^{-1}\boldsymbol{\sigma}_p}$$

- if sample interdistances d_{ij} are larger than correlogram range, then $\rho(d_{ij}) = 0$, and $\boldsymbol{\Sigma} = \sigma(0)\mathbf{I}$, the $(n \times n)$ identity matrix; this entails that $w_p(s_i) = \rho(d_{ip})$, i.e., weights are equal to correlogram values

Slide 8

- but in general, $\boldsymbol{\Sigma} \neq \sigma(0)\mathbf{I}$, i.e., sample interdistances are within correlation range, in which case $\boldsymbol{\Sigma}^{-1}$ modulates $\boldsymbol{\sigma}_p$: influence of samples in clusters is downplayed
- the closer the sample data to the prediction location, and the more spread out the data over the study region, the better the SK prediction is expected to be
- for sample data far away (beyond correlation range) from the prediction location $\mathbf{s}_p$, $\rho(d_{ip}) = 0$ and $w_p(\mathbf{s}_i) = 0$: all weighs are equal to 0
- for prediction at a sample location $\mathbf{s}_p \equiv \mathbf{s}_i$, data-to-unknown covariance vector $\boldsymbol{\sigma}_p = \boldsymbol{\sigma}_i$ is same as i-th column of $\boldsymbol{\Sigma}$; this yields $w_p(\mathbf{s}_i) = 1$ if $\mathbf{s}_i = \mathbf{s}_p$, 0 otherwise: only sample co-located with prediction location receives non-zero $(= 1)$ weight

Simple Kriging Prediction and Error Variance

Once the SK weights are computed as $\mathbf{w}_p = \mathbf{\Sigma}^{-1}\boldsymbol{\sigma}_p$, they are substituted in the following equations to compute the SK prediction $\hat{z}(\mathbf{s}_p)$ and associated error variance $\hat{\sigma}(\mathbf{s}_p)$

SK prediction does not depend on sill $\sigma(0)$ of covariogram model:

$$\hat{z}(\mathbf{s}_p) = m + \mathbf{w}_p^T \mathbf{r} = m + [w_p(\mathbf{s}_1) \ \cdots \ w_p(\mathbf{s}_n)] \begin{bmatrix} z(\mathbf{s}_1) - m \\ \vdots \\ z(\mathbf{s}_n) - m \end{bmatrix} = m + \sum_{i=1}^{n} w_p(\mathbf{s}_i)[z(\mathbf{s}_i) - m]$$

Slide 9

SK prediction error variance does depend on covariogram model sill $\sigma(0)$:

$$\hat{\sigma}(\mathbf{s}_p) = \sigma(0) - \mathbf{w}_p^T \boldsymbol{\sigma}_p = \sigma(0) - [w_p(\mathbf{s}_1) \ \cdots \ w_p(\mathbf{s}_n)] \begin{bmatrix} \sigma(d_{1p}) \\ \vdots \\ \sigma(d_{np}) \end{bmatrix} = \sigma(0) - \sum_{n=1}^{n} w_p(\mathbf{s}_i)\sigma(d_{ip})$$

which can also be written as: $\hat{\sigma}(\mathbf{s}_p) = \sigma(0)\left[1 - \sum_{i=1}^{n} w_p(\mathbf{s}_i)\rho(d_{ip})\right]$

Interpreting the SK Prediction and Error Variance

$$\hat{z}(\mathbf{s}_p) = m + \sum_{i=1}^{n} w_p(\mathbf{s}_i)[z(\mathbf{s}_i) - m] \qquad \hat{\sigma}(\mathbf{s}_p) = \sigma(0) - \sum_{i=1}^{n} w_p(\mathbf{s}_i)\sigma(d_{ip})$$

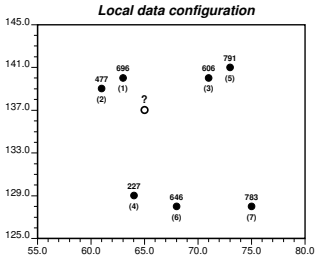
Comments:

- for sample data far away (beyond correlation range) from the prediction location $\mathbf{s}_p$, $w_p(\mathbf{s}_i) = 0, \forall i$: *all weighs are equal to* 0. In this case, the SK prediction equals the known mean m and the SK error variance equals the known covariogram sill: $\hat{z}(\mathbf{s}_p) = m$ and $\hat{\sigma}(\mathbf{s}_p) = \sigma(0)$; *away from the sample data, SK yields back the (assumed known) attribute overall mean and variance*

Slide 10

- for prediction at a sample location $\mathbf{s}_p \equiv \mathbf{s}_i$, $w_p(\mathbf{s}_i) = 1$ if $\mathbf{s}_i = \mathbf{s}_p$, 0 otherwise: the SK prediction identifies the known sample datum and the SK error variance is zero: $\hat{z}(\mathbf{s}_i) = z(\mathbf{s}_i)$ and $\hat{\sigma}(\mathbf{s}_i) = 0$; *SK is an exact interpolation algorithm*
- for all other prediction locations, the SK predictions depend on the sample data configuration and their values, while the SK error variances depend only on the sample data configuration; *both SK predictions and error variances depend on the covariogram model $\sigma(d)$ adopted*

Determining the SK Weights: Step 1



(n x n) matrix of data-to-data inter-distances:

Slide 11

$$\mathbf{D} = \begin{bmatrix} 0.00 & 2.24 & 8.00 & 11.05 & 10.05 & 13.00 & 16.97 \\ 2.24 & 0.00 & 10.05 & 10.44 & 12.17 & 13.04 & 17.80 \\ 8.00 & 10.05 & 0.00 & 13.04 & 2.24 & 12.37 & 12.65 \\ 11.05 & 10.44 & 13.04 & 0.00 & 15.00 & 4.12 & 11.05 \\ 10.05 & 12.17 & 2.24 & 15.00 & 0.00 & 13.93 & 13.15 \\ 13.00 & 13.04 & 12.37 & 4.12 & 13.93 & 0.00 & 7.00 \\ 16.97 & 17.80 & 12.65 & 11.05 & 13.15 & 7.00 & 0.00 \end{bmatrix}$$

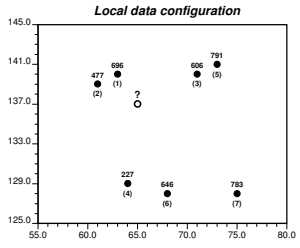
i, j-th element of $\mathbf{D}$: $d_{ij} = ||\mathbf{s}_i - \mathbf{s}_j||$

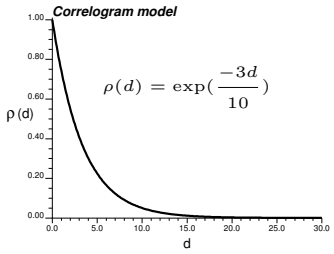
(n x 1) vector of prediction-to-data-location distances:

$$\mathbf{d}_p = \begin{bmatrix} 3.61 & 4.47 & 6.71 & 8.06 & 8.94 & 9.49 & 13.45 \end{bmatrix}^T$$

i-th element of $\mathbf{d}_p$: $d_{ip} = ||\mathbf{s}_i - \mathbf{s}_p||$

Determining the SK Weights: Step 2



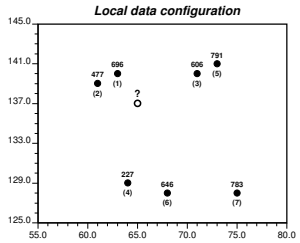


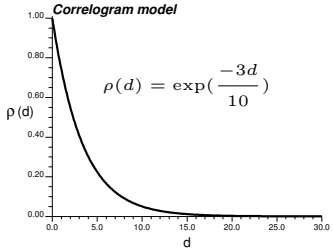
Slide 12

$$\underbrace{\begin{bmatrix} 3.61 \\ 4.47 \\ 6.71 \\ 8.06 \\ 8.94 \\ 9.49 \\ 13.45 \end{bmatrix}}_{\mathbf{d}_p} \rightarrow 1 \underbrace{\begin{bmatrix} \exp(-3 \times 3.61/10) \\ \exp(-3 \times 4.47/10) \\ \exp(-3 \times 6.71/10) \\ \exp(-3 \times 8.06/10) \\ \exp(-3 \times 8.94/10) \\ \exp(-3 \times 9.49/10) \\ \exp(-3 \times 13.45/10) \end{bmatrix}}_{\boldsymbol{\sigma}_p = \text{sill} \exp(-3\mathbf{d}_p/\text{range})} = \begin{bmatrix} 0.34 \\ 0.26 \\ 0.13 \\ 0.09 \\ 0.07 \\ 0.06 \\ 0.02 \end{bmatrix}$$

These would be the weights if one ignored auto-correlation between sample data

Determining the SK Weights: Step 3





SK system:

Slide 13

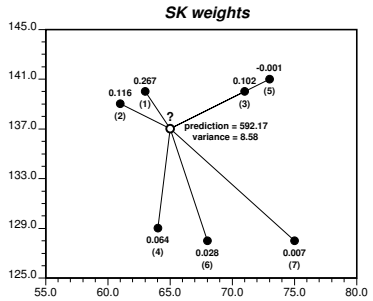
$$\underbrace{\begin{bmatrix} 1.00 & 0.51 & 0.09 & 0.04 & 0.05 & 0.02 & 0.01 \\ 0.51 & 1.00 & 0.05 & 0.04 & 0.03 & 0.02 & 0.00 \\ 0.09 & 0.05 & 1.00 & 0.02 & 0.51 & 0.02 & 0.02 \\ 0.04 & 0.04 & 0.02 & 1.00 & 0.01 & 0.29 & 0.04 \\ 0.05 & 0.03 & 0.51 & 0.01 & 1.00 & 0.02 & 0.02 \\ 0.02 & 0.02 & 0.02 & 0.29 & 0.02 & 1.00 & 0.12 \\ 0.01 & 0.00 & 0.02 & 0.04 & 0.02 & 0.12 & 1.00 \end{bmatrix}}_{\boldsymbol{\Sigma}} \underbrace{\begin{bmatrix} w_p(\mathbf{s}_1) \\ w_p(\mathbf{s}_2) \\ w_p(\mathbf{s}_3) \\ w_p(\mathbf{s}_4) \\ w_p(\mathbf{s}_5) \\ w_p(\mathbf{s}_6) \\ w_p(\mathbf{s}_7) \end{bmatrix}}_{\mathbf{w}_p} = \underbrace{\begin{bmatrix} 0.34 \\ 0.26 \\ 0.13 \\ 0.09 \\ 0.07 \\ 0.06 \\ 0.02 \end{bmatrix}}_{\boldsymbol{\sigma}_p}$$

i, j-th element of matrix $\boldsymbol{\Sigma}$: $\sigma_{ij} = 1 \times \exp(-3 \times d_{ij}/10)$

Determining the SK Weights: Step 4

$$\underbrace{\begin{bmatrix} w_p(\mathbf{s}_1) \\ w_p(\mathbf{s}_2) \\ w_p(\mathbf{s}_3) \\ w_p(\mathbf{s}_4) \\ w_p(\mathbf{s}_5) \\ w_p(\mathbf{s}_6) \\ w_p(\mathbf{s}_7) \end{bmatrix}}_{\mathbf{w}_p} = \underbrace{\begin{bmatrix} 1.36 & -0.69 & -0.09 & -0.02 & 0.00 & 0.00 & -0.01 \\ -0.69 & 1.35 & 0.00 & -0.02 & 0.00 & -0.01 & 0.01 \\ -0.09 & 0.00 & 1.36 & -0.01 & -0.69 & -0.01 & -0.01 \\ -0.02 & -0.02 & -0.01 & 1.09 & 0.00 & -0.32 & -0.01 \\ 0.00 & 0.00 & -0.69 & 0.00 & 1.35 & -0.01 & -0.01 \\ 0.00 & -0.01 & -0.01 & -0.32 & -0.01 & 1.11 & -0.12 \\ -0.01 & 0.01 & -0.01 & -0.01 & -0.01 & -0.12 & 1.02 \end{bmatrix}}_{\boldsymbol{\Sigma}^{-1}} \underbrace{\begin{bmatrix} 0.34 \\ 0.26 \\ 0.13 \\ 0.09 \\ 0.07 \\ 0.06 \\ 0.02 \end{bmatrix}}_{\boldsymbol{\sigma}_p}$$

Slide 14

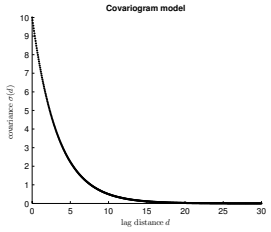


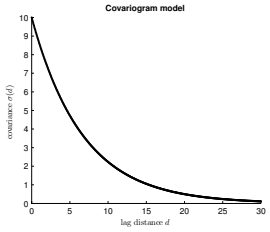
original weights vector ($\mathbf{w}_p = \boldsymbol{\sigma}_p$) modified by $\boldsymbol{\Sigma}^{-1}$ to account for sample redundancy;
e.g., $w_p(\mathbf{s}_1) = 0.27$ instead of $\rho(d_{1p}) = 0.34$

Covariogram Models and SK Weights (1)

Two covariogram models (with different ranges):

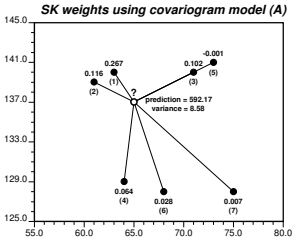
A: $\sigma(d) = 10\exp(\frac{-3d}{10})$ **B:** $\sigma(d) = 10\exp(\frac{-3d}{20})$

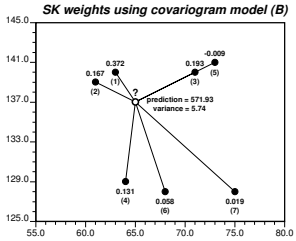




Slide 15

Two sets of Simple Kriging weights:



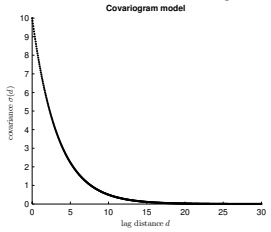


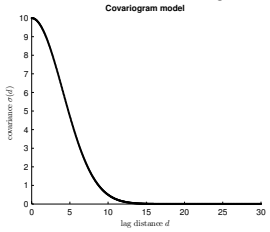
shorter range (left) tends to decrease the SK weights towards 0

Covariogram Models and SK Weights (2)

Two covariogram models (with different shapes):

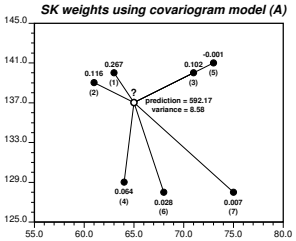
A: $\sigma(d) = 10\exp(\frac{-3d}{10})$ **B:** $\sigma(d) = 10\exp(\frac{-3d^2}{10^2})$

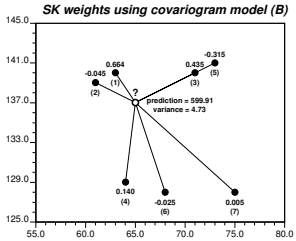




Slide 16

Two sets of Simple Kriging weights:





Gaussian covariogram yields larger weights for nearby data; negative weights possible

Spatial Prediction via Simple Kriging

SK characteristics:

- known attribute mean m
- prediction = weighted linear combination of sample attribute residuals (from known attribute mean m)
- weights account for:

Slide 17

  - redundancy (inverse correlation) between sample data
  - correlation between sample data and unknown values
  - functional form (e.g., exponential or Gaussian) of covariogram model
- SK variance = reliability of prediction = overall attribute variance (covariogram model sill $\sigma(0)$) reduced by "weighted" influence of "nearby" samples:
  - "nearby" = statistical proximity (correlogram)
  - "weighted" = sample redundancy (SK weights)
  - independent of sample data values, hence useful in sampling design

Ordinary Kriging (1)

OK prediction: $$\hat{z}(\mathbf{s}_p) = \mathbf{w}_p^T \mathbf{z} = [w_p(\mathbf{s}_1) \ \cdots \ w_p(\mathbf{s}_n)] \begin{bmatrix} z(\mathbf{s}_1) \\ \vdots \\ z(\mathbf{s}_n) \end{bmatrix}$$

attribute mean m is unknown $\Rightarrow$ constrain weights to sum to 1: $\sum_{i=1}^{n} w_p(\mathbf{s}_i) = \mathbf{w}_p^T \mathbf{1} = 1$

Slide 18

OK system: $$\begin{bmatrix} \sigma(0) & \cdots & \sigma(d_{1n}) & 1 \\ \vdots & \ddots & \vdots & \vdots \\ \sigma(d_{n1}) & \cdots & \sigma(0) & 1 \\ 1 & \cdots & 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} w_p(\mathbf{s}_1) \\ \vdots \\ w_p(\mathbf{s}_n) \\ -\xi_p \end{bmatrix} = \begin{bmatrix} \sigma(d_{1p}) \\ \vdots \\ \sigma(d_{np}) \\ 1 \end{bmatrix}$$

$\boldsymbol{\Sigma}^+$: $((n+1) \times (n+1))$ augmented data-to-data covariance matrix
$\boldsymbol{\sigma}_p^+$: $((n+1) \times 1)$ augmented data-to-unknown covariance vector
$\mathbf{w}_p^+ = [\mathbf{w}_p^T \ \ -\xi_p]^T$: $((n+1) \times 1)$ augmented vector of weights
with $-\xi_p$ being a Lagrange parameter accounting for constraint on weights

$$\boxed{\boldsymbol{\Sigma}^+ \mathbf{w}_p^+ = \boldsymbol{\sigma}_p^+}$$

Ordinary Kriging (2)

$$\hat{z}(\mathbf{s}_p) = \sum_{i=1}^{n} w_p(\mathbf{s}_i) z(\mathbf{s}_i) \qquad \hat{\sigma}(\mathbf{s}_p) = \sigma(0) - \sum_{i=1}^{n} w_p(\mathbf{s}_i)\sigma(d_{ip}) + \xi_p$$

Comments:

- OK weights, and consequently OK predictions and error variances, are **not** the same as those obtained via SK

Slide 19

- similar, however, spatial characteristics between SK and OK predictions and variances
- unknown attribute mean m is implicitly estimated as an unequally weighted average of sample data, based again on the covariogram model
- far away from data locations, OK weights tend to $1/n$ and OK predictions equal the sample mean, at data locations they reproduce the known attribute measurements
- SK variance is smaller than OK variance; extra penalty for not knowing the attribute mean is encapsulated in the Lagrange parameter ξ_p

Spatial Prediction with Ordinary Kriging

1. compute sample semivariogram from n sample data values $\{z(\mathbf{s}_i), i = 1, \ldots, n\}$
2. fit theoretical model to sample variogram, and convert it to a covariogram model
3. compute $(n \times n)$ data-to-data interdistance matrix $\mathbf{D}$
4. use covariogram model to transform distance matrix $\mathbf{D}$ to matrix of data-to-data covariance values $\mathbf{\Sigma}$; then, augment to $\mathbf{\Sigma}^+$ by appending row and column of 1s

Slide 20

5. consider a set of P prediction locations $\{\mathbf{s}_p, p = 1, \ldots, P\}$ on a regular grid or not
6. visit the p-th prediction location $\mathbf{s}_p$:
   (a) compute $(n \times 1)$ vector of distances $\mathbf{d}_p$ between $\mathbf{s}_p$ and all n data locations
   (b) use covariogram model to transform $\mathbf{d}_p$ into a $(n \times 1)$ vector of covariance values $\boldsymbol{\sigma}_p$; then, augment to $\boldsymbol{\sigma}_p^+$ by appending a 1
   (c) solve OK system for n weights and Lagrange parameter; compute OK prediction $\hat{z}(\mathbf{s}_p)$ and OK prediction error variance $\hat{\sigma}(\mathbf{s}_p)$
7. move to another prediction location $\mathbf{s}_{p'}$, and repeat steps 6a-6c

Ordinary Kriging with Sliding Search Neighborhoods

same procedure as before BUT:

- at each prediction location $\mathbf{s}_p$, use only closest $n_p << n$ sample data $\{z(\mathbf{s}_i), i = 1, \ldots, n_p\}$ within neighborhood $W(\mathbf{s}_p)$ centered at $\mathbf{s}_p$ to compute the $(n_p \times n_p)$ data-to-data covariance matrix $\mathbf{\Sigma}^+$ and the $(n_p \times 1)$ data-to-unknown covariance vector $\boldsymbol{\sigma}_p^+$

Pros:

Slide 21

- no need to invert a large $\mathbf{\Sigma}^+$ matrix in the case of many sample data
- local (constant within each neighborhood $W(\mathbf{s}_p)$) estimation of attribute mean allows capturing departures from global mean m

Cons:

- need for definition of search neighborhood (not a big issue); use circle with radius $\sim$ range of covariogram model
- no clear-cut decomposition in (and interpretation of) first- and second-order effects

OK with moving neighborhoods = most frequently used geostatistical prediction algorithm

Summary

Geostatistical spatial prediction accounts for:

- statistical proximity (correlation) between data and unknowns
- redundancy (inverse correlation) between sample data themselves, distributing weights among nearby samples

Fundamental differences from other methods:

Slide 22

- everything depends on the model covariogram (which is typically built from a sample semivariogram), thus reflecting the notion that weights should account for the nature of spatial variability (smooth versus rough) of a particular phenomenon
- an output of Kriging is a set of reliability measures for the predicted values. Such measures are encoded in the prediction error variances, which are independent of the actual data values and only depend on the sample configuration and covariogram model; Kriging error variances are often used in sampling design applications

Outlook: There are other more sophisticated (and more complex) methods for spatial prediction, but we won't be looking into them in this class