# Modeling Spatially Continuous Data

Bailey and Gatrell – Chapter 5

Lecture 15

October 27, 2009



## Generalized Least Squares

- First fit the model to the observed data using ordinary least squares
- Estimate a variogram model using the residuals from the ordinary least squares estimation
- Compute covariogram using the variogram model
- From the covariogram estimate the covariance matrix and refit the model using generalized least squares

The validity of the final model depends on:

- appropriate trend surface model choice
- appropriate variogram model choice



## Generalized Least Squares

### Objectives

- Understand and describe the nature of the spatial variation in an attribute

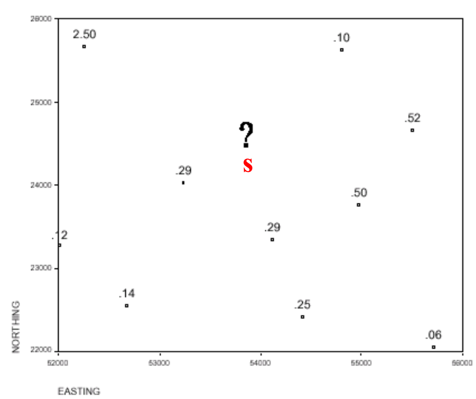
  Knowledge of trend and covariance structure are sufficient

- Predict an attribute value at locations where it has not been sampled

  Consider how to use derived model for prediction purposes



**Want to predict a value for the attribute at location s**



## Developing Models for Prediction

Assume we have the model

$$Y(s) = x^T(s)\beta + U(s)$$

*U(s)* is a zero mean process with covariance function C()

From this model we can do better than just a prediction based on mean value $\mu = x^T(s)\hat{\beta}$

We add a local component to the mean based on knowledge of the covariance structure and the observed values at sampled points

**This approach is known as kriging**

5

## Kriging

*Matheron coined the term in honor of D. G. Krige, a South African mining engineer*

- An optimal spatial linear prediction method
- Optimal in the sense that it is unbiased and minimizes the mean squared prediction error
- Based primarily on the second order properties of the process Y
- Prediction weights are based on the spatial dependence between observations modeled by the variogram

6

## Kriging Models

- **Simple Kriging**

  Mean is known and constant throughout the study region

- **Ordinary Kriging**

  Mean is fixed but unknown and needs to be estimated

- **Universal Kriging**

  Mean varies, is an unknown linear combination of known functions

7

## Simple Kriging

Assumes mean is known and does not need to be estimated

- Subtract known mean from sample observations to obtain a set of residuals $u_i$ $\quad u_i = y_i - \mu(s)$
- We assume these residuals have zero mean, a known variance $\sigma^2$ and covariance function C(s)
- Find an estimate $\hat{u}(s)$ for a value $u(s)$ of the random variable *U(s)* at the location *s* given observed values $u_i$ of the random variable $U(s_i)$ at n sample locations $s_i$
- With an estimate $\hat{u}(s)$, the predicted value of the random variable $\hat{y}(s)$ is derived by adding $\hat{u}(s)$ to the known trend at the point s.

8

## Simple Kriging

Estimates are weighted linear combinations of the observed residuals

Weighted sum of *n* random variables at sample sites $s_i$

$$\hat{U}(s) = \sum_{i=1}^{n} \lambda_i(s) U(s_i)$$

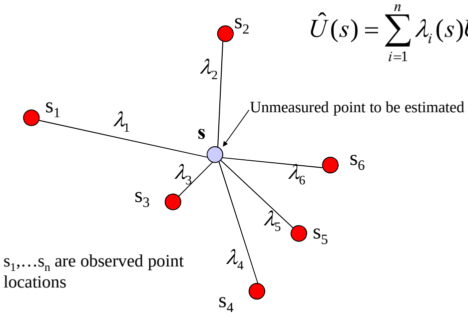
$\lambda_i(s)$ different weights applied to different locations **s**

$\hat{U}(s)$ as the sum of random variables is a random variable itself

$\hat{U}(s)$ should have a mean value of zero for any set of weights since by assumption, the mean of $U(s)$ is zero and weights are constants

9

$$\hat{U}(s) = \sum_{i=1}^{n} \lambda_i(s) U(s_i)$$




$s_1,\ldots s_n$ are observed point locations

On average the value of $\hat{U}(s)$ will be zero

10

## Simple Kriging

The semivariance is

$$2\gamma(h) = E\{[U(s) - U(s+h)]^2\}$$

The covariance is

$$C(h) = E\{U(s) * U(s+h)\}$$

$$2\gamma(h) = E\{U^2(s)\} + E\{U^2(s+h)\} - 2 * E\{U(s) * U(s+h)\}$$

$$= Var\{U(s)\} + Var\{U(s+h)\} - 2 * C(h)$$

$$= 2[C(0) - C(h)]$$

$$C(h) = C(0) - \gamma(h)$$

11

## Simple Kriging

We define the error of an estimated value as the difference between the estimated value $\hat{U}(s)$ and the "true" value $U(s)$ at the same location.

Assuming $\hat{U}(s)$ and $U(s)$ both have zero mean the expected mean square error is:

$$E((\hat{U}(s) - U(s))^2 = E(\hat{U}^2(s)) + E(U^2(s)) - 2E(U(s)\hat{U}(s))$$

12

## Simple Kriging

$$\hat{U}(s) = \sum_{i=1}^{n} \lambda_i(s) U(s_i)$$

$E((\hat{U}(s) - U(s))^2$ expected mean square error

$$= E\{\hat{U}^2(s)\} - 2E\{U(s)\hat{U}(s)\} + E\{U^2(s)\}$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{n} \lambda_i \lambda_j E\{U(s_i) * U(s_j)\} - 2\sum_{i=1}^{n} \lambda_i E\{U(s) * U(s_i)\} + C(0)$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{n} \lambda_i \lambda_j C(s_i, s_j) - 2\sum_{i=1}^{n} \lambda_i C(s, s_i) + C(0)$$

$$= \lambda^T(s)\mathbf{C}\lambda(s) + \sigma^2 - 2\lambda^T(s)\mathbf{c}(s)$$

13

## Simple Kriging

$$= \lambda^T(s)\mathbf{C}\lambda(s) + \sigma^2 - 2\lambda^T(s)\mathbf{c}(s)$$

***C*** is an (n x n) matrix of covariances, between all possible pairs of the n sample points

**c**(s) is an (n x 1) column vector of covariances between the prediction point **s** and each of the n sample points

14

## Simple Kriging

$$E((\hat{U}(s) - U(s))^2 = \sum_{i=1}^{n}\sum_{j=1}^{n} \lambda_i \lambda_j C(s_i, s_j) - 2\sum_{i=1}^{n} \lambda_i C(s, s_i) + C(0)$$

Expression of the error variance is a function of n variables namely the weights λ

The minimization of a function of n variables is accomplished by setting the n partial first derivatives to 0

$$\frac{\partial(\sum_{i=1}^{n}\sum_{j=1}^{n} \lambda_i \lambda_j C_{ij})}{\partial \lambda_1} = \frac{\partial(\lambda_1^2 C_{11} + 2\lambda_1 \sum_{j=2}^{n} \lambda_j C_{1j})}{\partial \lambda_1}$$

$$= 2\lambda_1 C_{11} + 2\lambda_1 \sum_{j=2}^{n} \lambda_j C_{1j}$$

$$= 2\sum_{j=1}^{n} \lambda_j C_{1j}$$

$$\frac{\partial(\sum_{i=1}^{n} \lambda_i C_{i0})}{\partial \lambda_1} = \frac{\partial(\lambda_1 C_{10})}{\partial \lambda_1}$$

$$= C_{10}$$

$$2\sum_{j=1}^{n} \lambda_j C_{1j} - 2C_{10} = 0$$

15

## Simple Kriging

$$2\sum_{j=1}^{n} \lambda_j C_{1j} - 2C_{10} = 0$$

$$\sum_{j=1}^{n} \lambda_j C_{1j} = C_{10}$$

$$\sum_{j=1}^{n} \lambda_j C(s_i, s_j) = C(s, s_i) \quad i = 1, \ldots, n$$

$$\lambda(s) = C^{-1} c(s)$$

$$\hat{U}(s) = \lambda^T(s) U = c^T(s) C^{-1} U$$

16

## Simple Kriging

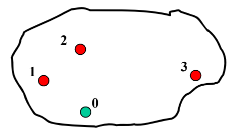
The minimized expected mean square error corresponding to the choice of weights is

$$E((\hat{U}(s)-U(s))^2)=\sigma^2-c^T(s)C^{-1}c(s)$$

Mean square prediction error or kriging variance $\sigma_e^2$

17

## Simple Kriging




$$\sum_{j=1}^{3}\lambda_j C(s_i,s_j)=C(s,s_i) \quad i=1,\ldots,3$$

$$\lambda_1 * C(1,1)+\lambda_2 * C(1,2)+\lambda_3 * C(1,3)=C(0,1)$$
$$\lambda_1 * C(2,1)+\lambda_2 * C(2,2)+\lambda_3 * C(2,3)=C(0,2)$$
$$\lambda_1 * C(3,1)+\lambda_2 * C(3,2)+\lambda_3 * C(3,3)=C(0,3)$$

$$\begin{bmatrix} C(1,1) & C(1,2) & C(1,3) \\ C(2,1) & C(2,2) & C(2,3) \\ C(3,1) & C(3,2) & C(3,3) \end{bmatrix} * \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \end{bmatrix} = \begin{bmatrix} C(0,1) \\ C(0,2) \\ C(0,3) \end{bmatrix} \quad \lambda(s)=C^{-1}c(s)$$

18

## Worked Example –Simple Kriging

| ID | X | Y | Z | u |
|---|---|---|---|---|
| 1 | 33 | 85 | 122 | -38.37 |
| 2 | 78 | 105 | 183 | 22.63 |
| 3 | 83 | 65 | 148 | -12.37 |
| 4 | 89 | 51 | 160 | -0.37 |
| 5 | 24 | 24 | 176 | 15.63 |

**Known mean is 160.37**

| Distances | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| 1 | 0 | 49.2 | 53.9 | 65.2 | 60.795 |
| 2 | | 0 | 40.3 | 55.9 | 97.1 |
| 3 | | | 0 | 15.8 | 72.09 |
| 4 | | | | 0 | 69.6 |
| 5 | | | | | 0 |

19

## Worked Example –Simple Kriging

| Distances | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| 1 | 0 | 49.2 | 53.9 | 65.2 | 60.795 |
| 2 | | 0 | 40.3 | 55.9 | 97.1 |
| 3 | | | 0 | 15.8 | 72.09 |
| 4 | | | | 0 | 69.6 |
| 5 | | | | | 0 |

$C(h)=20e^{-3h/100}$ Exponential model with sill = 20 and range = 100

| Covariances | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| 1 | 20.000 | 4.571 | 3.970 | 2.828 | 3.228 |
| 2 | 4.571 | 20.000 | 5.970 | 3.739 | 1.086 |
| 3 | 3.970 | 5.970 | 20.000 | 12.450 | 2.300 |
| 4 | 2.828 | 3.739 | 12.450 | 20.000 | 2.479 |
| 5 | 3.228 | 1.086 | 2.300 | 2.479 | 20.000 |

20

## Worked Example –Simple Kriging

| Covariances | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| 1 | 20.000 | 4.571 | 3.970 | 2.828 | 3.228 |
| 2 | 4.571 | 20.000 | 5.970 | 3.739 | 1.086 |
| 3 | 3.970 | 5.970 | 20.000 | 12.450 | 2.300 |
| 4 | 2.828 | 3.739 | 12.450 | 20.000 | 2.479 |
| 5 | 3.228 | 1.086 | 2.300 | 2.479 | 20.000 |

$$\lambda(s) = C^{-1}c(s)$$

$C^{-1}$

| | | | | |
|---|---|---|---|---|
| 0.054914 | -0.01004 | -0.00646 | -0.00095 | -0.00746 |
| -0.01004 | 0.056751 | -0.01508 | 0.000168 | 0.000252 |
| -0.00646 | -0.01508 | 0.087403 | -0.05044 | -0.00194 |
| -0.00095 | 0.000168 | -0.05044 | 0.082023 | -0.00422 |
| -0.00746 | 0.000252 | -0.00194 | -0.00422 | 0.051936 |

*c(s)*

| | |
|---|---|
| s1 | 6.895 |
| s2 | 5.032 |
| s3 | 10.15 |
| s4 | 8.083 |
| s5 | 4.079 |

21

## Worked Example –Simple Kriging

$C^{-1}$

| | | | | |
|---|---|---|---|---|
| 0.054914 | -0.01004 | -0.00646 | -0.00095 | -0.00746 |
| -0.01004 | 0.056751 | -0.01508 | 0.000168 | 0.000252 |
| -0.00646 | -0.01508 | 0.087403 | -0.05044 | -0.00194 |
| -0.00095 | 0.000168 | -0.05044 | 0.082023 | -0.00422 |
| -0.00746 | 0.000252 | -0.00194 | -0.00422 | 0.051936 |

*

| c(s) |
|---|
| 6.895 |
| 5.032 |
| 10.15 |
| 8.083 |
| 4.079 |

$$\lambda(s) = C^{-1}c(s)$$

| **λ(s)** |
|---|
| 0.225 |
| 0.066 |
| 0.351 |
| 0.128 |
| 0.107 |
| **Sum** 0.877 |

Simple kriging weights need not sum to 1

22

## Worked Example –Simple Kriging

$$\hat{U}(s) = \sum_{i=1}^{n} \lambda_i(s) U(s_i)$$

$$\hat{u}(s) = 0.225 * -38.37 + 0.066 * 22.63 + 0.351 * -12.37 + 0.128 * -0.37 + 0.107 * 15.63$$

$$\cong -9.86$$

$$\hat{y}(s) = 160.37 - 9.86 \cong 150.5$$

$$\sigma_e^2 = \sigma^2 - c^T(s) C^{-1} c(s)$$

$= 20.0\text{-}6.92$

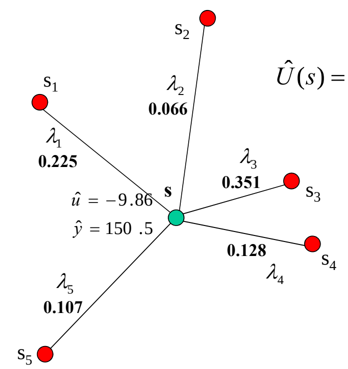
$= 13.08$

| $\lambda_i$(s) | u(s) |
|---|---|
| 0.225 | -38.37 |
| 0.066 | 22.63 |
| 0.351 | -12.37 |
| 0.128 | -0.37 |
| 0.107 | 15.63 |

95 percent confidence interval is $\hat{y}(s) \pm 1.96\sigma_e$

95 percent CI = $150.5 \pm 7.09$

23




$$\hat{U}(s) = \sum_{i=1}^{n} \lambda_i(s) U(s_i)$$

$s_1, \ldots s_n$ are observed point locations

24

## Ordinary Kriging

- First order effects are implicitly estimated as part of the prediction process
- Prediction of $y_i$ occurs in one step using a weighted linear combination of the observed values $y_i$

$$\hat{Y}(s_0) = \sum_{i=1}^{n} \omega_i Y(s_i)$$

In this case we chose weights $\omega$ so that the mean value of $\hat{Y}(s)$ is constrained to be the mean



## Ordinary Kriging

$$\hat{Y}(s_0) = \sum_{i=1}^{n} \omega_i Y(s_i)$$

Prediction is unbiased when:

$$E[Y(s_0) - \hat{Y}(s_0)] = 0$$

This is achieved if and only if:

$$\sum_{i=1}^{n} \omega_i = 1$$



## Ordinary Kriging

- The mean of $Y(s)$ and each of the $Y(s_i)$ are all $\mu$
- The mean of $\hat{Y}(s)$ will also be $\mu$ as long as the weights sum to one
- With this constraint we minimize the mean squared error between values of $Y(s)$ and $\hat{Y}(s)$



## Derivation of kriging weights

Minimize the expected squared prediction error, in other words choose the $\omega_i$ such that:

$$E[(\hat{Y}(s) - Y(s))^2]$$

is as small as possible

## Ordinary Kriging

$$\hat{Y}(s_0) = \sum_{i=1}^{n} \omega_i Y(s_i)$$

The expected mean square error is

$$E((\hat{Y}(s) - Y(s))^2) = \omega^T(s) C \omega(s) + \sigma^2 - 2\omega^T(s) c(s)$$

***C*** is an (n x n) matrix of covariances, ***C****$(s_i, s_j)$* between all possible pairs of the *n* sample sites and *c(s)* is an (n x 1) column vector of covariances, ***C****$(s,s_i)$* between the prediction point ***s*** and each of the *n* sample sites



## Ordinary Kriging

$$E((\hat{Y}(s) - Y(s))^2) = \omega^T(s) C \omega(s) + \sigma^2 - 2\omega^T(s) c(s)$$

We want to minimize this mean square error but subject to a constraint

To carry out optimization with constraints use the method of Lagrange multiplier *v(s)* – subject to matrix constraint $\omega^T \mathbf{1} = 1$

$$\omega^T(s) C \omega(s) + \sigma^2 - 2\omega^T(s) c(s) + 2v(s)(\omega^T \mathbf{1} - 1)$$

The Lagrange parameter converts an constrained minimization problem to an unconstrained one

Setting n partial derivatives to zero produces n equations and n unknowns – the unbiasedness condition adds an equation without an unknown



## Ordinary Kriging

Minimize the following:

$$E[(\sum_{i=1}^{n} \omega_i Y(s_i) - Y(s_0))^2] - 2v(\sum_{i=1}^{n} \omega_i - 1)$$

Setting the derivative with respect to the Lagrange multiplier ***v*** to zero yields the unbiasedness constraint

$$\frac{\partial \sigma_e^2}{\partial v} = \frac{\partial(2v(\sum_{i=1}^{n} \omega_i - 1))}{\partial v} \qquad = 2\sum_{i=1}^{n} \omega_i - 2 \qquad 2\sum_{i=1}^{n} \omega_i = 2$$

$$\sum_{i=1}^{n} \omega_i = 1$$



## Ordinary Kriging

$$\omega^T(s) C \omega(s) + \sigma^2 - 2\omega^T(s) c(s) + 2v(s)(\omega^T \mathbf{1} - 1)$$

Leads to 2 simultaneous equations

$$\omega^T \mathbf{1} = 1$$

$$\mathbf{C}\omega(s) + \mathbf{1}v(s) = c(s)$$

## Ordinary Kriging

These equations are expressed as modified matrix $\boldsymbol{C}_+$ and vectors $\omega_+$ and $c_+$

$$\omega^T \mathbf{1} = 1$$

$$\mathbf{C}\omega(s) + \mathbf{1}v(s) = c(s)$$

$$\mathbf{C}_+ \quad * \quad \omega_+(s) \quad = \quad c_+(s)$$

$$\begin{pmatrix} C(s_1,s_1) & \cdots & C(s_1,s_n) & 1 \\ \vdots & \ddots & \vdots & 1 \\ \vdots & & \vdots & \vdots \\ C(s_n,s_1) & \cdots & C(s_n,s_n) & 1 \\ 1 & \cdots & 1 & 0 \end{pmatrix} \begin{pmatrix} \omega_1(s) \\ \vdots \\ \omega_n(s) \\ v(s) \end{pmatrix} = \begin{pmatrix} C(s,s_1) \\ \vdots \\ C(s,s_n) \\ 1 \end{pmatrix}$$

33

## Ordinary Kriging

To obtain the prediction $\hat{y}(s)$ $\qquad \mathbf{C}_+ * \boldsymbol{\omega}_+(s) = \mathbf{c}_+(s)$

Solve the equation for $\omega_+(s)$ $\qquad \mathbf{C}_+^{-1}\mathbf{C}_+ * \boldsymbol{\omega}_+(s) = \mathbf{C}_+^{-1}\mathbf{c}_+(s)$

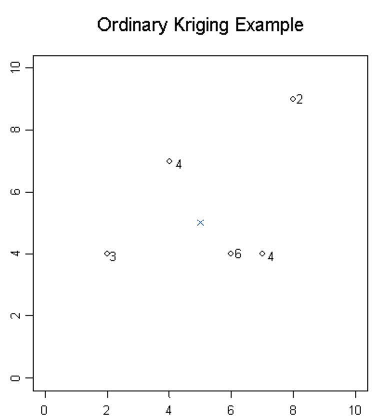
Extract from this the vector $\omega(s)$ $\qquad \mathbf{I} * \boldsymbol{\omega}_+(s) = \mathbf{C}_+^{-1}\mathbf{c}_+(s)$

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

Then $\hat{y}(s) = \omega^T(s)y$

$y$ is the original set of observations $y_i$

34




Source http://www.msu.edu/~ashton/466/2002_notes/3-18-02/ok_ill.html

35

## Worked Example – Ordinary Kriging

| | x | y | z |
|---|---|---|---|
| 1 | 2 | 4 | 3 |
| 2 | 4 | 7 | 4 |
| 3 | 8 | 9 | 2 |
| 4 | 7 | 4 | 4 |
| 5 | 6 | 4 | 6 |

The distance matrix for the 5 data points is:

| | [1] | [2] | [3] | [4] | [5] |
|---|---|---|---|---|---|
| [1] | 0.000 | 3.605 | 7.810 | 5.000 | 4.000 |
| [2] | 3.605 | 0.000 | 4.472 | 4.243 | 3.605 |
| [3] | 7.810 | 4.472 | 0.000 | 5.099 | 5.385 |
| [4] | 5.000 | 4.243 | 5.099 | 0.000 | 1.000 |
| [5] | 4.000 | 3.605 | 5.385 | 1.000 | 0.000 |

The distance vector between the data points and the unknown point is:

[1] 3.162 2.236 5.000 2.236 1.414

36

## Worked Example – Ordinary Kriging

Variogram model is a spherical model with nugget=2.5, sill=7.5, range=10.0

$$\gamma(h)=\begin{cases} 2.5+(7.5-2.5)\left(\frac{3h}{20}-\frac{h^3}{2000}\right) & 0\le h\le 10 \\ 0 & h=0 \\ 7.5 & otherwise \end{cases}$$

$$\gamma(h)=\sigma^2-C(h)$$

$$C(h)=7.5-2.5+(7.5-2.5)\left(\frac{3h}{20}-\frac{h^3}{2000}\right)$$

The covariance matrix adjusted for the Lagrange Multiplier

| | | | | | |
|---|---|---|---|---|---|
| 10 | 3.620065 | 0.500173 | 2.34375 | 3.24 | 1 |
| 3.620065 | 10 | 2.80438 | 3.013076 | 3.620065 | 1 |
| 0.5001733 | 2.80438 | 10 | 2.260774 | 2.027458 | 1 |
| 2.34375 | 3.013076 | 2.260774 | 10 | 6.37875 | 1 |
| 3.24 | 3.620065 | 2.027458 | 6.37875 | 10 | 1 |
| 1 | 1 | 1 | 1 | 1 | 0 |



## Worked Example – Ordinary Kriging

The vector of covariances for the points to the unknown point is:

[4.061023 5.026350 2.343750 5.026350 5.919616 1.000000]

$$\omega_+(s)=\mathbf{C}_+^{-1}c_+(s)$$

A vector of weights along with the LaGrange multiplier

| [1] | [z] |
|---|---|
| [1] 0.17289193 | 3 |
| [2] 0.26523729 | 4 |
| [3] 0.05887157 | 2 |
| [4] 0.16986833 | 4 |
| [5] 0.33313088 | 6 |
| [6] -0.13471033 | |

$$\hat{y}(s)=\omega^T(s)y$$

Multiply the weight for each data point by the attribute value of that point to determine the ordinary kriging estimate:

4.375627



## Worked Example – Ordinary Kriging

Calculate the variance as the transpose of c times the inverse of C times c, subtracted from the sill $\sigma_e^2=\sigma^2-c_+{}^T(s)C_+{}^{-1}c_+(s)$

$\mathbf{c}^T$ [4.061023 5.026350 2.343750 5.026350 5.919616 1.000000]

**c**

4.061023
5.026350
2.343750
5.026350
5.919616
1.000000

*

$\mathbf{C}^{-1}$

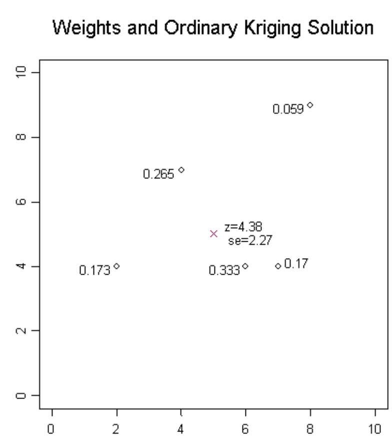
| | | | | | |
|---|---|---|---|---|---|
| 0.103710846 | -0.04595 | -0.00987 | -0.01464 | -0.03325 | 0.276208 |
| -0.045948564 | 0.128043 | -0.03866 | -0.01542 | -0.02801 | 0.142191 |
| -0.009870999 | -0.03866 | 0.090115 | -0.02815 | -0.01343 | 0.303088 |
| -0.014644484 | -0.01542 | -0.02815 | 0.164988 | -0.10676 | 0.17559 |
| -0.033246799 | -0.02801 | -0.01343 | -0.10676 | 0.181442 | 0.102923 |
| 0.276208178 | 0.142191 | 0.303088 | 0.17559 | 0.102923 | -4.17343 |

4.864229

The kriging standard error is the square root of this:

$\sigma_e$ = 2.2055





## Ordinary versus Simple Kriging

Primarily a local neighborhood estimator

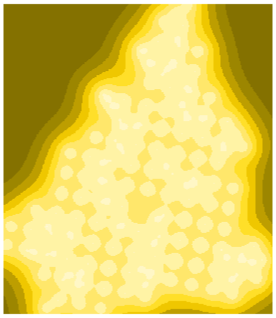
There is no need to estimate a first-order trend; instead, the mean is estimated from nearby data only.

Estimates are not as sensitive to non-stationarity (though the covariance model may be affected, it is not as strongly affected at short ranges as at long ones).

41

Predicted values for Nickel using Ordinary Kriging

Kriging Variance for Nickel



spherical model with nugget=2.5, sill=7.5, and range=10.0

42

## Ordinary Kriging Implementation

Ordinary kriging is often applied for a local moving window

A window approximating the size of the range is typically used

Kriging weights can be estimated directly from the semivariance rather than covariances

$$\omega_+(s) = \Gamma_+^{-1}\gamma_+(s)$$

When a local search window is used a subset of the observation points is used for potentially every point estimate

This requires that a new $\Gamma$ is used for each estimate and must be inverted for each estimate – such a matrix is usually much smaller however.

43

## Ordinary Kriging Implementation

When a local search neighborhood is used the mean is assumed constant for the local neighborhood.

Since any trend is not explicitly removed or modeled in ordinary kriging, a locally constant mean may be a reasonable assumption

However the variogram model is always globally estimated

44

## Universal Kriging

Includes a first order trend component

$$Y(s) = \beta_0 + \beta_1 x_1(s) + \beta_2 x_2(s) + \cdots + \beta_n x_n(s) + \varepsilon(s)$$

Forms a prediction for y in one step as a linear combination of observed values

$$\hat{Y}(s) = \sum_{i=1}^{n} \omega_i Y(s_i)$$

As for ordinary kriging, the weights are chosen to minimize mean squared error

$$MSE = E\left\{ | Y(s) - \hat{Y}(s) |^2 \right\}$$

Subject to the constraint that $\hat{Y}(s)$ is unbiased for *Y(s)*

that is $E\{\hat{Y}(s)\} = E\{(Y(s)\}$ for all $\beta_0, \beta_1, \beta_2, \cdots \beta_p$

45

## Universal Kriging

$E\{\hat{Y}(s)\} = E\{(Y(s)\}$ Unbiasedness condition

$\hat{Y}(s)$ is unbiased if and only if $\sum_{i=1}^{n} \omega_i = 1$

and $\sum_{i=1}^{n} \omega_i x_j(s_i) = x_j(s_0) \;\; j = 1 \cdots p$

For universal kriging the explanatory variables x1,...., xp have to be known at each "$s_0$", that is, at each location for which the value of Y is to be predicted

46

## Universal Kriging

To obtain the weights that minimize the mean square error subject to these constraints, again use the method of Lagrange multipliers

Taking the derivatives with respect to ω and *v*, setting the expressions to zero and re-arranging terms we get the kriging equations (using variogram)

$$\sum_{i=1}^{n} \omega_i \gamma(s_i - s_k) + v_0 + \sum_{j=1}^{p} v_j x_j(s_k) = \gamma(s_k - s_0); k = 1, 2 \cdots, n$$

$$\sum_{i=1}^{n} \omega_i = 1 \qquad \sum_{i=1}^{n} \omega_i x_k(s_i) = x_k(s_0); k = 1, 2 \cdots, p$$

47

## Universal Kriging

In matrix notation

$$\begin{pmatrix} 0 & \gamma(s_1,s_2) & \cdots & \gamma(s_1,s_n) & 1 & x_1(s_1) & \cdots & x_p(s_1) \\ \gamma(s_2,s_1) & 0 & \cdots & \gamma(s_2,s_n) & 1 & x_1(s_2) & \cdots & x_p(s_2) \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \gamma(s_n,s_1) & \gamma(s_1,s_2) & \cdots & 0 & 1 & x_1(s_n) & \cdots & x_p(s_n) \\ 1 & 1 & \cdots & 1 & 0 & 0 & \cdots & 0 \\ x_1(s_1) & x_1(s_2) & \cdots & x_1(s_n) & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ x_p(s_1) & x_p(s_2) & \cdots & x_p(s_n) & 0 & 0 & \cdots & 0 \end{pmatrix} * \begin{pmatrix} \omega_1 \\ \omega_2 \\ \vdots \\ \omega_n \\ v_0 \\ v_1 \\ \vdots \\ v_p \end{pmatrix} = \begin{pmatrix} \gamma(s_1 - s_0) \\ \gamma(s_2 - s_0) \\ \vdots \\ \gamma(s_n - s_0) \\ 1 \\ x_1(s_0) \\ \vdots \\ x_p(s_0) \end{pmatrix}$$

$$\Gamma_+ \omega_+(s) = \boldsymbol{\gamma}_+(s)$$

$$\omega_+(s) = \Gamma_+^{-1} \boldsymbol{\gamma}_+(s)$$

Prediction = $\hat{y}(s) = \omega^T(s)\mathbf{y}$

$$\sigma_e^2 = \boldsymbol{\gamma}_+^T(s) \Gamma_+^{-1} \boldsymbol{\gamma}_+(s)$$

48

## Universal Kriging

- May make more sense to estimate trend explicitly
- Need to estimate trend to derive residuals for variogram modeling in any case since it is only safe to estimate variogram model from y when the mean is assumed constant
- Does not make sense to use it for local neighborhoods

Best to use generalized least squares approach and remove trend explicitly

49

## Evaluation of Kriging Results

How do you check the adequacy of kriged map?

Kriging is an exact interpolator so no residuals to check

Can take additional observations on the variable

Can set aside some portion of the data to validate the spatial predictor

Often all data are used and there is no opportunity for further data collection

50

## Cross Validation

Validation method using sample data

Used to evaluate adequacy of a spatial correlation model

Can also be used to evaluate choice of lag and angle tolerances in estimating the variogram

Method uses tests at the location of existing samples

**Procedure**

For location $s_i$, omit the observation $y_i$ temporarily

Estimate $y_i$ using the remaining points by the estimation method or model

Compare $\hat{y}_i$ to $y_i$

Repeat steps for all i=1,…,n data points

51

## Cross Validation

**Summary Statistics**

$$(1/n)\sum_{i=1}^{n}(\hat{y}_i - y_i)$$ **Mean Error**

**Should be approximately zero**

$$\left[(1/n)\sum_{i=1}^{n}(\hat{y}_i - y_i)^2\right]^{1/2}$$

**RMSE - Should be small**

52

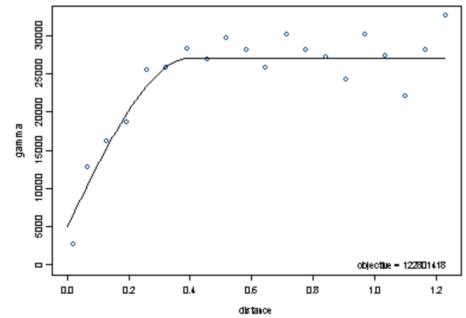
## Cross Validation

Scatterplots of the sample versus cross validated predicted values can be viewed to see how closely the values agree. Pairs of points deviating from the 45 degree line are not well modeled

Spatial analysis (map) of cross validation residuals is helpful – are there clusters of positive or negative residuals



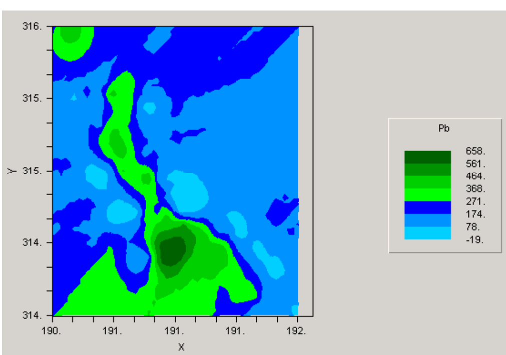
## Cross validation example




spherical model with range=.4, sill=22000, nugget=5000)

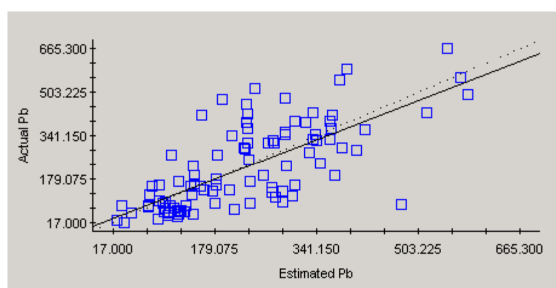


## Cross validation example

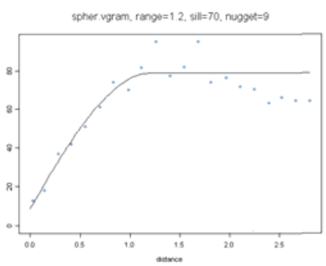


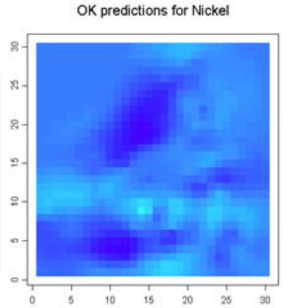



## Cross validation example



## Cross validation example

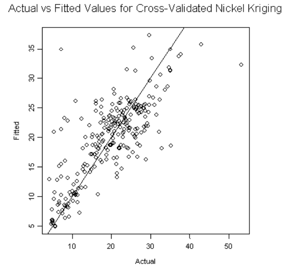


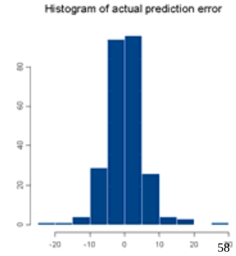




57

## Cross validation example

Nickel range 4.2- 53.2,
m = 19.7, sd = 8.2




58

## Cross Validation

If data are clustered, points within clusters are more heavily validated than point outside the clusters

A model chosen based on cross validation may model cluster areas better than other areas

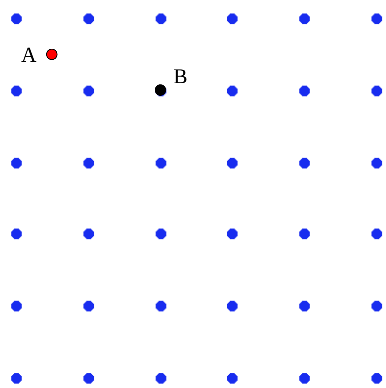
Estimation of cross validation points involves longer distances than estimation points that were never sampled

As a consequence cross validation may be less helpful in model selection for correlation at small lag distances

59

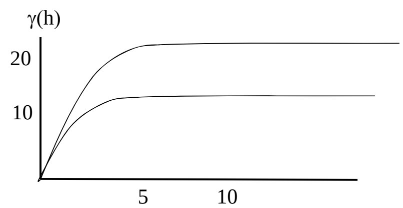
### Estimation of value at point A versus B

In this case any unsampled point will be closer than the measured sites being validated




60

## Effect of scale



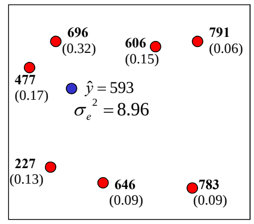

Two variograms that differ only in scale (sill)

$$\gamma_1(h) = 10(1 - e^{-.3|h|})$$

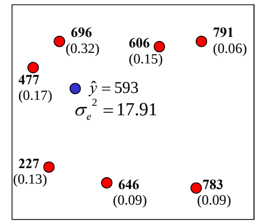
$$\gamma_2(h) = 20(1 - e^{-.3|h|}) = 2\gamma_1(h)$$



## Effect of scale



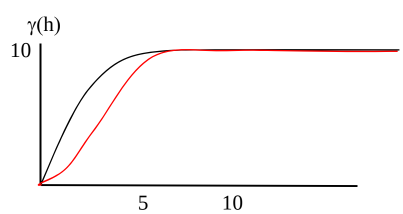

Kriging weights for variogram with sill = 10




Kriging weights for variogram with sill = 20.



## Effect of shape



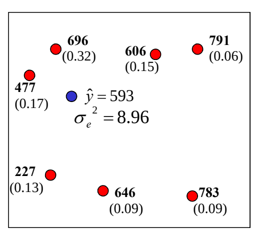

Two variograms that differ in shape

$$\gamma_1(h) = 10(1 - \exp - 3\frac{|h|}{10})$$

$$\gamma_2(h) = 10(1 - \exp(-3\frac{|h|}{10})^2$$

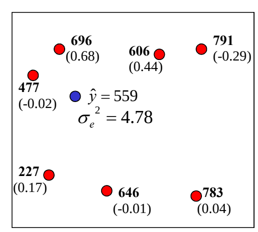


## Effect of shape




Kriging weights for variogram with exponential model

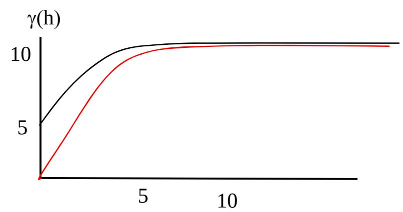
$$\gamma_1(h) = 10(1 - \exp - 3\frac{|h|}{10})$$




Kriging weights for variogram with Gaussian model.

$$\gamma_2(h) = 10(1 - \exp(-3\frac{|h|}{10})^2$$

# Nugget effect



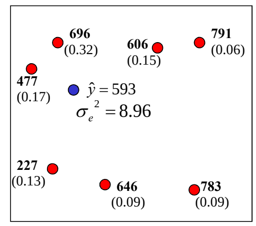

Two variograms that differ in nugget (nugget is 50 percent of sill)

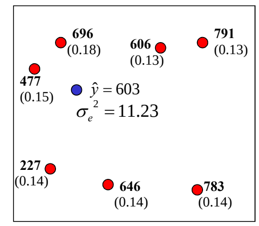
$$\gamma_1(h) = 10(1 - e^{-.3|h|})$$

$$\gamma_2(h) = \begin{cases} 0 & if\ h = 0 \\ 5 + 5(1 - e^{-.3|h|}) & if\ h > 0 \end{cases}$$



# Nugget effect



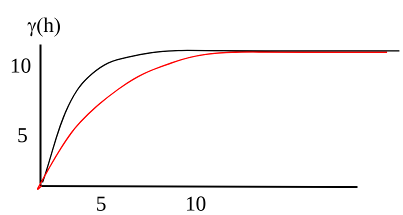

Kriging weights for variogram no nugget

Kriging weights for variogram with nugget effect 50% of sill

Nugget weights are more similar to each other, kriging variance is higher



# Range effect



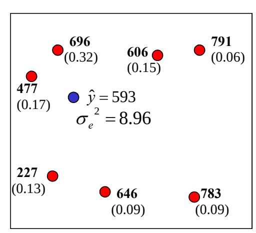

Two variograms that differ in their range

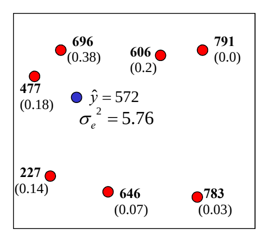
$$\gamma_1(h) = 10(1 - e^{-.3|h|})$$

$$\gamma_2(h) = 10(1 - e^{-.15|h|}) = \gamma_1(\frac{1}{2}h)$$



# Range effect




Kriging weights for variogram with range 10

Kriging weights for variogram with range 20

Doubling the range makes samples appear twice as close – this lowers the kriging variance