School of Geography and Geology
McMaster University

Advanced Topics in Spatial Statistics

Spatially Continuous Data
V & VI

This session

- Trend surface analysis
- Generalized least squares
- Spatial prediction
- Kriging

Modeling Spatially Continuous Data

- Trend Surface Analysis
  - A method to model the first order component of a spatial process
  - Multiple regression analysis using the coordinates of the points as “explanatory” variables

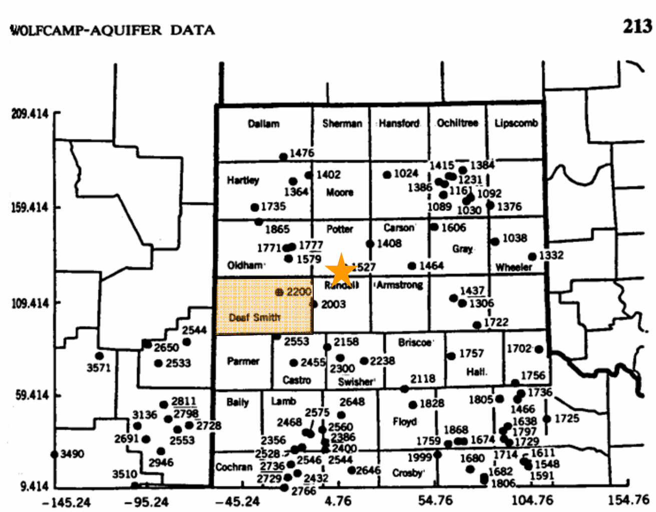
Example

- Wolfcamp aquifer (Texas)
  - High level nuclear waste depository (U.S.)
  - Candidate locations: Texas, Washington state, Nevada
  - 68,000 waste canisters placed underground, surrounded by salt
  - Covering an area of about 2 square miles
  - US Department of Energy stipulates that canisters must be isolated by 10,000 years

Example: Wolfcamp Aquifer

- Wolfcamp aquifer (Texas)
  - Potential issues
    - Leaks
    - Tiny quantities of water in the salt could migrate towards the canisters
    - Salt+water → hydrocloric acid could corrode canisters
    - Contamination of aquifer

Example: Wolfcamp Aquifer



Example: Wolfcamp Aquifer

- Data
  - Piezometric head at 85 locations in Texas panhandle (h)
- Geostatistical problems
  - Determine sites at risk
  - Interpolate surface
  - Quantify uncertainty (location of monitoring stations)

Example: Wolfcamp Aquifer

- Spatial distribution of observations

Trend Surface Analysis

- Quadratic trend surface

$$h_i = b_1 + x_i^2 b_2 + x_i b_3 + x_i y_i b_4 + y_i b_5 + y_i^2 b_6 + e_i$$

  - h: head
  - b_1 , b_2 ,…, b_6 : regression parameters
  - x , y: coordinates of point I
- Higher order polynomials
- Cubic
  - Quartic
  - …

Trend Surface Analysis

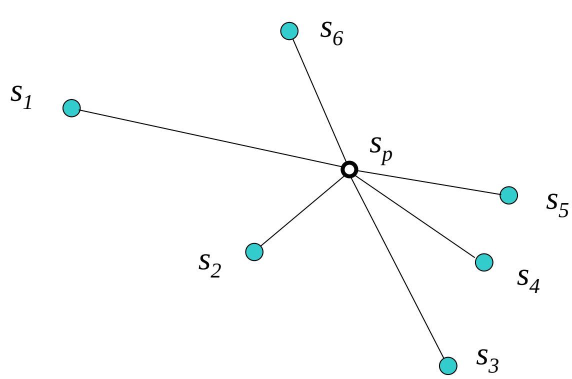
- Explanation vs. prediction?
- Potential issues
  - Multicolinearity
  - Trend surfaces tend to “curl” around the edges
- This example: fairly good fit

Spatial Prediction

- Q: What are the implications of spatial structure for prediction?

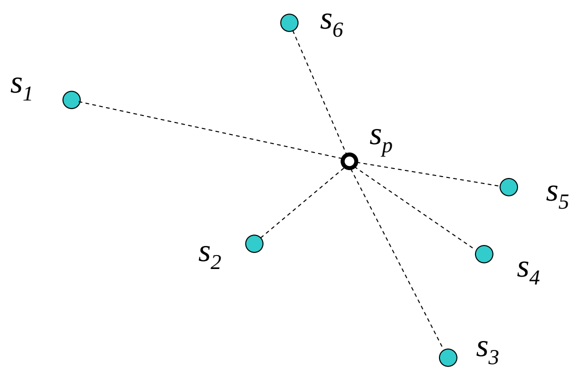
Spatial Prediction

- Error terms are not completely random, i.e., unpredictable



Spatial Independence

- Error terms are completely random, i.e., unpredictable



Generalized Least Squares (GSL)

- OLS
  - Variance is constant
  - Error terms are independent (i.e. random)

Generalized Least Squares

- More general regression model
  - Relaxes some of the assumptions imposed on the error terms
  - Non-constant variance
  - **Spatial dependency** (spatial structure)
- Incorporate residual spatial structure

Regression Analysis

- Error terms (Ordinary Least Squares)
  - Constant variance
  - Independent

$$E[\boldsymbol{ee}'] = \mathbf{C} = \begin{bmatrix} \sigma^2 & & 0 & 0 \\ & \sigma^2 & & 0 \\ 0 & & \sigma^2 & \\ 0 & 0 & & \sigma^2 \end{bmatrix} = \sigma^2 \begin{bmatrix} 1 & & 0 & 0 \\ & 1 & & 0 \\ 0 & & 1 & \\ 0 & 0 & & 1 \end{bmatrix} = \sigma^2 \mathbf{I}$$

Regression Analysis

- Error terms (Generalized Least Squares)
  - Constant variance
  - NOT Independent: covariance between e_i and $e_j \neq 0$

$$E[\boldsymbol{ee}'] = \mathbf{C} = \begin{bmatrix} \sigma^2 & \sigma^2_{21} & \cdots & \sigma^2_{1n} \\ \sigma^2_{21} & \sigma^2 & & \sigma^2_{2n} \\ \vdots & & \ddots & \vdots \\ \sigma^2_{n1} & \sigma^2_{n2} & \cdots & \sigma^2 \end{bmatrix}$$

Regression Analysis: GSL

- Parameters

$$\hat{\boldsymbol{b}} = \left(\mathbf{X}^T\mathbf{C}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{C}^{-1}\boldsymbol{Z}$$

- Variance

$$\hat{\sigma}^2 = \frac{1}{n-k}\left(\boldsymbol{Z} - \mathbf{X}^T\hat{\boldsymbol{b}}\right)^T\left(\boldsymbol{Z} - \mathbf{X}^T\hat{\boldsymbol{b}}\right)$$

Spatial Prediction

- Kriging
  - So called after South African mining geologist D.G. Krige
  - Method for spatial prediction
  - Makes use of autocorrelation information
  - Optimal spatial prediction!

Simple Kriging

- A method for optimal prediction when there is no trend (first order effects), or when the trend is known

$$\boldsymbol{Z} = \boldsymbol{\mu} + \boldsymbol{e}$$

Model with constant or known trend μ

$$Z_p = \mu + e_p$$

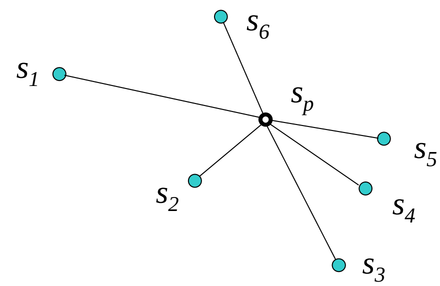
Spatial prediction

Simple Kriging

- When errors are autocorrelated

$$\hat{e}_p = \sum_{j=1}^{n} \lambda_{jp} e_j = \boldsymbol{\lambda}_p^T \mathbf{e}$$

λ_{jp}: linear weights



Simple Kriging

- $\hat{e}_p$ is a random variable
  - Expected (mean) value = 0
  - How close (on average) is it to $\boldsymbol{e}$?

Simple Kriging

- How close (on average) is $\hat{e}_p$ to $\boldsymbol{e}$?
  - *Expected mean square error* (prediction variance):

$$E\left[\left(\hat{e}_p - e_i\right)^2\right] = E\left[\hat{e}_p^2\right] + E\left[e_i^2\right] - 2E\left[e_i e_p\right]$$

$$= \boldsymbol{\lambda}_p^T \mathbf{C} \boldsymbol{\lambda}_p + \sigma^2 - 2\boldsymbol{\lambda}_p^T \boldsymbol{c}_p$$

$\mathbf{C}$: covariance matrix

$\boldsymbol{c}_p$: covariance vector between $\boldsymbol{e}$ and $\hat{e}_p$

Simple Kriging

- Minimizing the expected mean square error gives weight λ:

$$\boldsymbol{\lambda}_p = \mathbf{C}^{-1} \boldsymbol{c}_p$$

$\mathbf{C}$: covariance matrix

$\boldsymbol{c}_p$: covariance vector between $\boldsymbol{e}$ and $\hat{e}_p$

Simple Kriging

- This leads to:

$$\hat{e}_p = \boldsymbol{\lambda}_p^T \boldsymbol{e} = \boldsymbol{c}_p^T \mathbf{C}^{-1} \boldsymbol{e}$$

- Which can be added to the known mean to give:

$$Z_p = \mu + e_p = \mu + \boldsymbol{c}_p^T \mathbf{C}^{-1} \boldsymbol{e}$$

Simple Kriging

- Prediction variance:

$$\sigma_p^2 = E\left[\left(\hat{e}_p - e_i\right)^2\right] = \sigma^2 - \boldsymbol{c}_p^T \mathbf{C}^{-1} \boldsymbol{c}_p$$

- Interval of confidence for prediction (95% level of confidence):

$$Z_p \pm 1.96\sigma_p$$

Simple Kriging

- Limitations
  - Trend is not always constant or known
  - Covariance structure is not known

Defining Spatial Structure

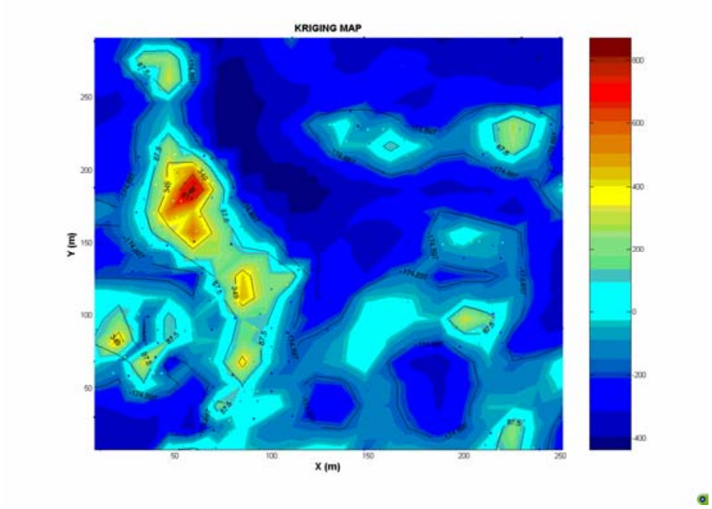
- Variogram ← → Covariance

- Generalized Least Squares
  - Parameter estimation
  - Inference
- Spatial prediction
  - Simple kriging
  - Generalized spatial prediction
  - Ordinary and universal kriging

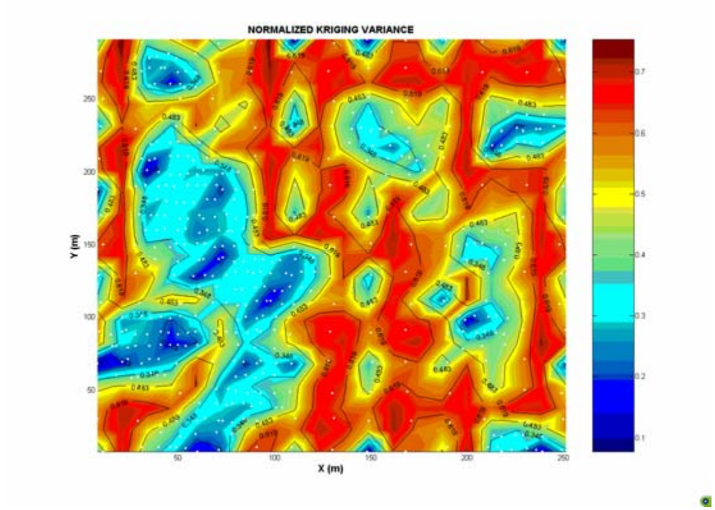
Simple Kriging

- Example: Walker Lake data
  - Very weak trend – no trend
  - Linear trend: $R^2 = 0.063$
  - Quadratic trend: $R^2 = 0.126$
  - Cubic trend: $R^2 = 0.172$
- Perfect application of simple kriging
  - "Known" trend: mean

Example: Kriging Map (Exponential)



Example: Kriging Variance (Exponential)



General Spatial Prediction

- In simple kriging there is no trend or the trend is known
  - Example: Walker Lake dataset

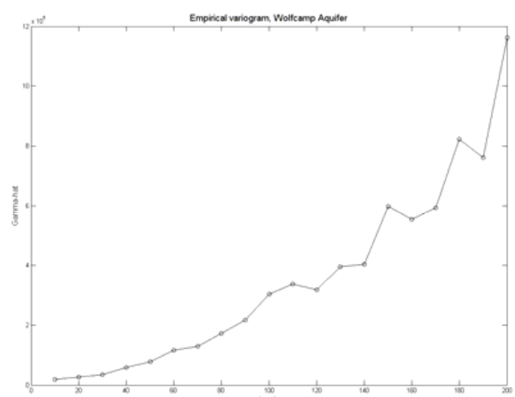
- In many other cases there is a trend that must be estimated from the data

General Spatial Prediction

- Evidence of a trend
  - Trend surface analysis
  - Empirical variogram (absolute values or deviation from the mean)

General Spatial Prediction

- Evidence of a trend



General Spatial Prediction

- It becomes necessary to estimate the trend
  - Trend surface analysis
  - Other covariates
  - A combination of trend surface analysis and other covariates

General Spatial Prediction

- Modeler's dilemma!
  - The trend is estimated by OLS (We don't know if there is autocorrelation!)

$$\hat{\boldsymbol{b}} = \left(\mathbf{X}^T \mathbf{X}\right)^{-1} \mathbf{X}^T \boldsymbol{Z}$$

  - However, if there is autocorrelation, the estimators should be:

$$\hat{\boldsymbol{b}} = \left(\mathbf{X}^T \mathbf{C}^{-1} \mathbf{X}\right)^{-1} \mathbf{X}^T \mathbf{C}^{-1} \boldsymbol{Z}$$

General Spatial Prediction

- Iterative procedure:
  - Step 1. Fit a trend surface model by OLS and obtain the residuals:

$$\hat{\boldsymbol{e}} = \boldsymbol{Z} - \mathbf{X}\hat{\boldsymbol{b}} = \boldsymbol{Z} - \mathbf{X}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\boldsymbol{Z}$$

  - Step 2. Describe spatial structure of residuals using a variogram and derive a covariance matrix $\mathbf{C}$

General Spatial Prediction

- Iterative procedure:
  - Step 3. Fit a trend surface model by **GLS** and obtain revised residuals:

$$\hat{\boldsymbol{u}} = \boldsymbol{Z} - \mathbf{X}\hat{\boldsymbol{b}} = \boldsymbol{Z} - \mathbf{X}\left(\mathbf{X}^T\hat{\mathbf{C}}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}^T\hat{\mathbf{C}}^{-1}\boldsymbol{Z}$$

  - Step 4. Iterate steps 2 and 3 until stability of the parameters is achieved

General Spatial Prediction

- Iterative procedure:
  - Step 5. Predict Z_p as:

$$Z_p = \boldsymbol{X}_p\hat{\boldsymbol{b}} + \hat{u}_p = \boldsymbol{X}_p\hat{\boldsymbol{b}} + \hat{\boldsymbol{c}}_p^T\hat{\mathbf{C}}^{-1}\hat{\boldsymbol{u}}$$

  - Step 6. Calculate the prediction variance:

$$\sigma_p^2 = \left\{\left(\boldsymbol{X}_p - \mathbf{X}^T\hat{\mathbf{C}}^{-1}\hat{\boldsymbol{c}}_p\right)^T\left(\mathbf{X}^T\hat{\mathbf{C}}^{-1}\mathbf{X}\right)\left(\boldsymbol{X}_p - \mathbf{X}^T\hat{\mathbf{C}}^{-1}\hat{\boldsymbol{c}}_p\right)\right\}$$

$$+\left(\hat{\sigma}^2 - \hat{\boldsymbol{c}}_p^T\hat{\mathbf{C}}^{-1}\hat{\boldsymbol{c}}_p\right)$$

General Spatial Prediction

- Example: Wolfcamp Aquifer data
  - Strong trend
  - Linear trend: $R^2 = 0.892$
  - Quadratic trend: $R^2 = 0.913$
  - Cubic trend: $R^2 = 0.923$
- General spatial prediction
  - Use trend surface

Variogram Selection

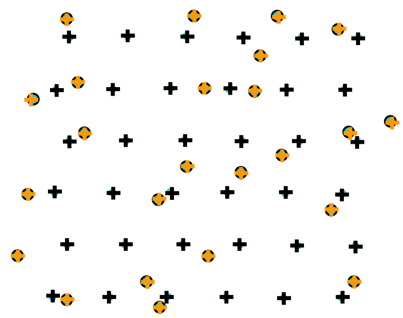
- What is the best variogram for my application?
- Variogram selection

Variogram Selection

- Validation sample
  - Use a sub-sample of data for model estimation
  - Use model to predict those observations not used for estimation

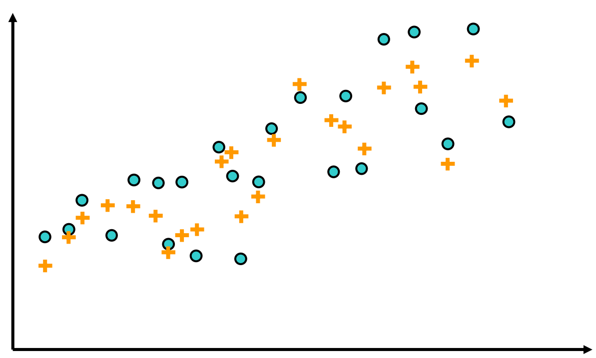
Variogram Selection

- Cross-validation
  - Double kriging



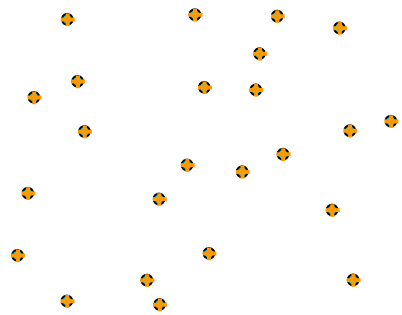
Variogram Selection

- Double kriging



Variogram Selection

- Cross-validation
  - Leave-one-out



Variogram Selection

- Cross-validation scores

$$CVS = \frac{1}{n}\sum_{i=1}^{n}\frac{Z_i - \hat{Z}_{(i)}}{\hat{\sigma}_{(i)}}$$

- Z_i : observed value at location i
- $\hat{Z}_{(i)}$: predicted value after removing observation at i
- $\hat{\sigma}_{(i)}$: standard error estimated after removing observation i

Variogram Selection

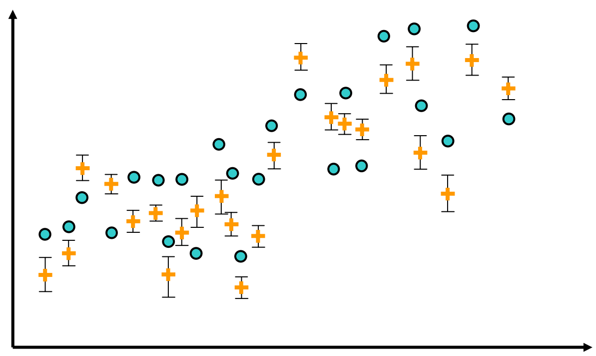
- Theoretical soundness
  - What do we know about the generating process?

Variogram Selection

- Cross-validation
  - Accuracy
- Prediction (kriging) variance
  - Precision

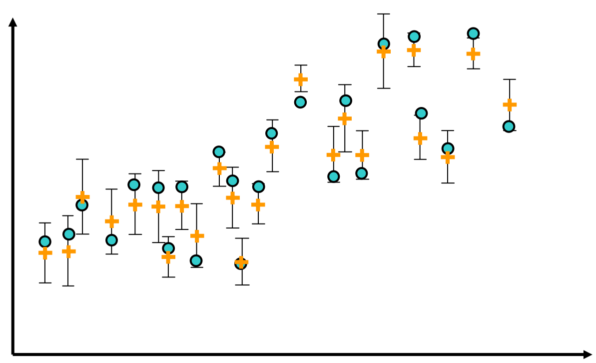
Variogram Selection

- Precise…inaccurate



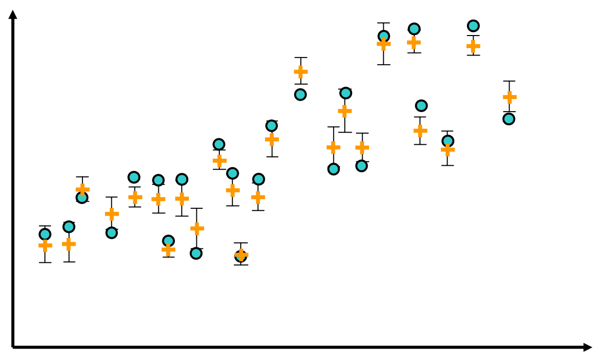
Variogram Selection

- Accurate…imprecise



Variogram Selection

- Accurate…precise



Example

- Variogram Selection: Wolfcamp Aquifer
  - Linear trend

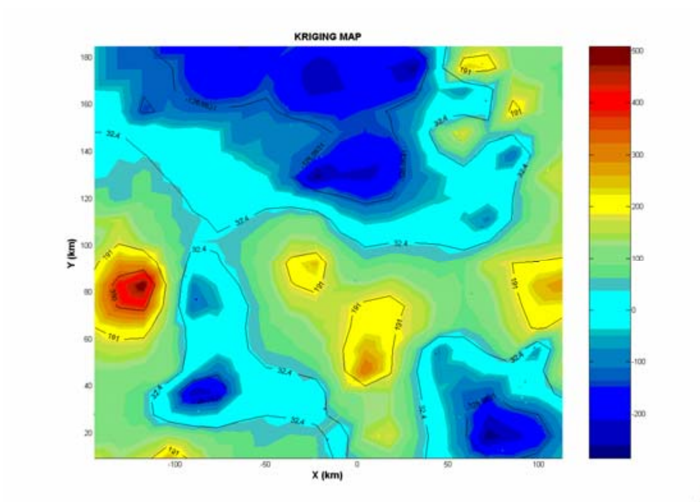
	Cross Validation			
	Double Kriging		Leave-one-out	
Variogram:	CVS	s	CVS	s
Linear	-6.176 (1)	194.001 (4)	0.782 (1)	201.766 (4)
Spherical	-25.674 (4)	164.331 (1)	9.050 (4)	176.892 (2)
Exponential	-22.163 (3)	167.810 (3)	6.083 (3)	176.696 (1)
Gaussian	-15.499 (2)	166.129 (2)	3.477 (2)	178.466 (3)

Example

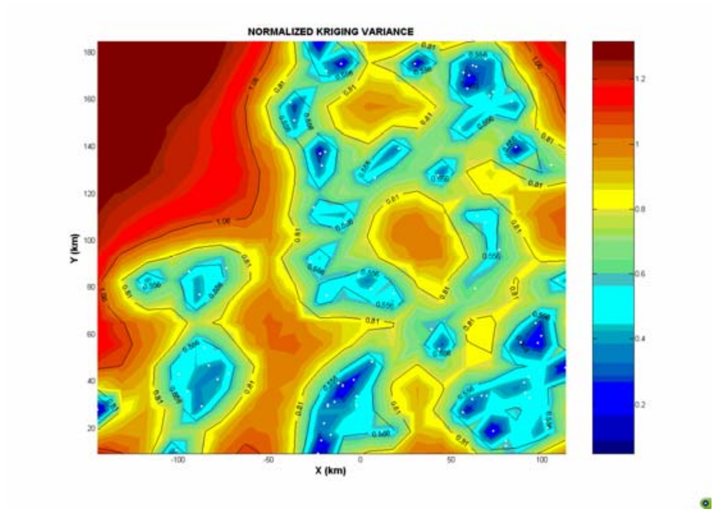
- Variogram Selection: Wolfcamp Aquifer
  - Quadratic trend

	Cross Validation			
	Double Kriging		Leave-one-out	
Variogram:	CVS	s	CVS	s
Linear	-3.952 (1)	182.683 (4)	3.512 (1)	193.237 (4)
Spherical	-15.648 (2)	175.816 (2)	7.150 (2)	174.749 (2)
Exponential	-22.821 (4)	170.063 (1)	12.064 (4)	174.563 (1)
Gaussian	-16.062 (3)	175.853 (3)	7.359 (3)	176.692 (3)

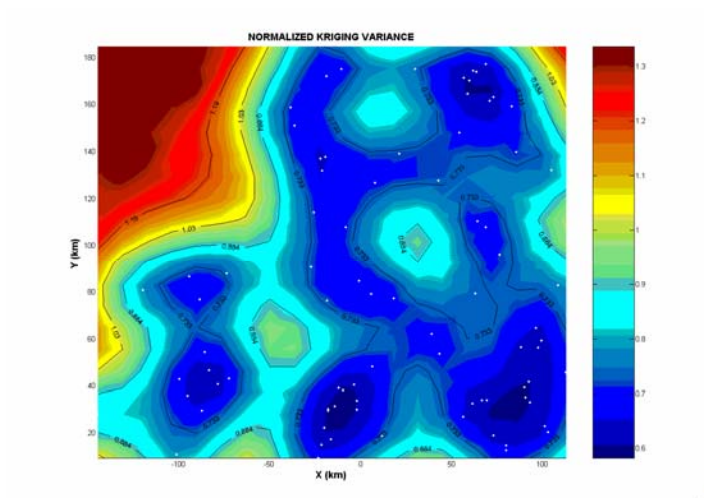
Example: Kriging Map (Exponential)



Example: Kriging Variance (Exponential)



Example: Kriging Variance (Gaussian)



Ordinary and Universal Kriging

- Mathematically equivalent to simple kriging and general spatial prediction
  - Simple kriging = ordinary kriging
  - Universal kriging = general spatial prediction
- Different emphasis: purely predictive methods

Ordinary and Universal Kriging

- Take a “one-step” approach with the trend implicit in the prediction process
  - Not so relevant in the case of simple kriging
  - May simplify the process in the case of processes with a trend
  - Although it is possible to retrieve the trend, Bailey and Gatrell recommend explicitly modeling it by using general spatial prediction

Next

- Area Data I & II