GEOG574 Introduction to Geostatistics

Simple Kriging

Daoqin Tong

School of Geography & Development
408 Harvill Building
Email: daoqin@email.arizona.edu

Interpolation Methods

- Deterministic
  - Function fitting
  - A priori knowledge of underlying behavior needed
    - ⇒ Function form
    - ⇒ Search window size
- Kriging
  - No a priori knowledge
  - Statistical derivation of structure
  - Can deal with uncertain measurements
  - Interpolation includes error estimate
  - Local, exact, stochastic

Kriging Interpolation

- The procedures involved in kriging incorporate measures of error and uncertainty when determining estimations.
- In the kriging method, every known data value and every missing data value has an associated variance.
- Based on the semivariogram used, optimal weights are assigned to known values in order to calculate unknown ones. Since the variogram changes with distance, the weights depend on the known sample distribution.

Generalized Least Squares

- Ordinary Least squares (OLS)

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

$$E(\varepsilon(s)) = 0;\ Cov(\varepsilon(s_i), \varepsilon(s_j)) = 0$$

$$\hat{\beta} = (X^T X)^{-1} X^T Y$$

- Generalized Least Squares (GLS)

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

$$E(U(s)) = 0;\ Cov(U(s_i), U(s_j)) = C(s_i - s_j)$$

$$\hat{\beta} = (X^T C^{-1} X)^{-1} X^T C^{-1} Y$$

GLS in Practice

- Generally, the following steps are implemented
  - Use OLS to fit the model
  - Estimate a variogram model using the residuals form the OLS estimation
  - Derive the covariogram model from the variogram model
  - Refit the model using GLS and the estimated covariance matrix based on covariogram model
  - Iterate the above steps if necessary

Reliability

- The validity of the final model depends on
  - Appropriate trend surface model
  - Appropriate variogram model

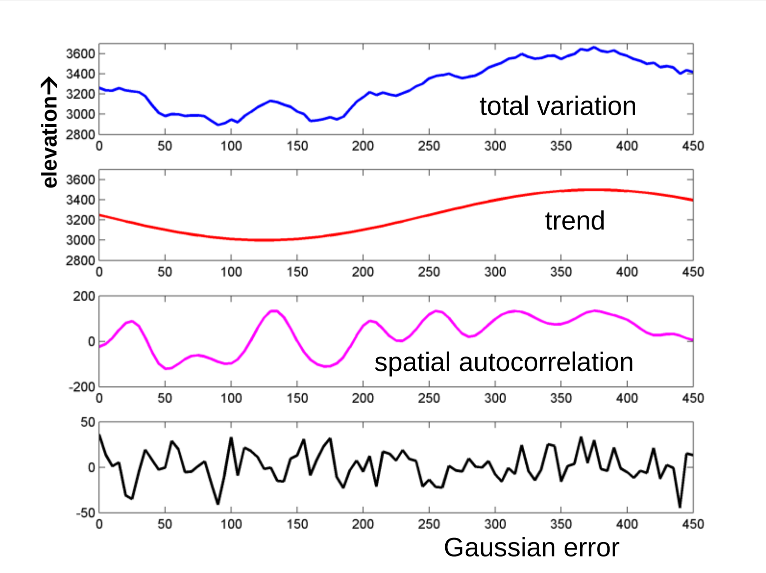
Two Different Objectives

- Understand and describe the nature of the spatial variation in the observed data and isolate any systematic large scale trend
  - Knowledge of the trend component and the covariance structure is needed
- Predict or interpolate the unobserved points
  - Objective of Kriging

Kriging

- Named after South African mining geologist D.G. Krige
- Given a model, $Y(s) = \mu(s) + U(s)$, where covariance structure of *U(s)* is known through a variogram model, we can predict *Y(s)* by adding the prediction of a local component of *U(s)* to the trend surface $\hat{\mu}(s)$
  - An optimal spatial linear prediction method
  - It is unbiased and minimizes the mean squared prediction error (MSPE)
  - Prediction weights are based on the spatial dependence between observations modeled by the variogram
  - Based primarily on the second order properties of the process Y

kriging principle



Kriging Models

- Simple Kriging
  - The first order component $\mu(s)$ is known a priori and does not have be estimated from the observed data
- Ordinary Kriging
  - The first order component $\mu(s)$ is constant but unknown and needs to be estimated
- Universal Kriging
  - The first order component $\mu(s)$ is not constant but an unknown linear combination of known functions

Linear Predictor

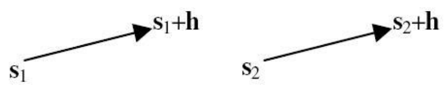
- Intuitively sensible to predict U(s) by a weighted linear combination of observed residuals

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

- Many smoothing and interpolation methods (moving average, nearest neighbor methods, kernel methods) are such predictors
- Kriging allows for choosing weights according to the degree of spatial correlation, which can be estimated from the data

Random Error U(S)

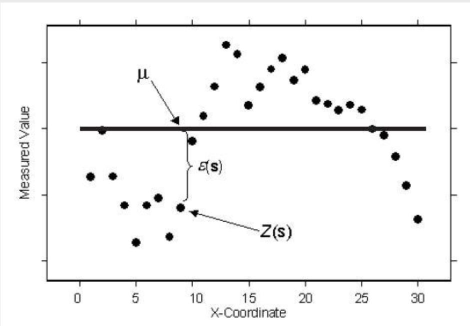
- Second-order stationarity assumption
  - Expected to be 0
  - Autocorrelation between U(s) and U(s+h) does not depend on the actual location s
  - i.e. spatial structure of the variable is consistent over the entire domain of the dataset.



Simple Kriging

- $\mu(s)$ is known
- Given sample data y(s_1),…,y(s_n), the observed values of residuals can be calculated by $u(s_i) = y(s_i) - \mu(s)$
- Assume stationarity for these residuals: zero mean, known variance, and known covariance function C(h)
- Find an estimate of U(s) at location s given observed u(si) of the random variable U(si) using

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



Known

$$z(s) = \mu + U(s)$$

- **Because you assume to know $\mu(s)$ then you also know $U(s)$**

- **This assumption (knowing $\mu(s)$) is often unrealistic, unless working with physical based models.**

Simple Kriging

- Linear predictor $\hat{U}(s) = \sum_{i=1}^{n} \lambda_i(s) U(s_i)$
- $\hat{U}(s)$ has zero mean
- How to determine the weights

$$
\begin{aligned}
E(\hat{U}(s) - U(s))^2 &= E(\hat{U}^2(s)) + E(U^2(s)) - 2E(\hat{U}(s)U(s)) \\
&= \sum_i \sum_j \lambda_i(s)\lambda_j(s) C(s_i, s_j) + \sigma^2 - 2\sum_i \lambda_i(s) C(s, s_i) \\
&= \lambda^T(s) C \lambda(s) + \sigma^2 - 2\lambda^T(s) c(s)
\end{aligned}
$$

where **C** is the $n \times n$ matrix of covariances between all possible pairs of sample points, $c(s)$ is an $n \times 1$ column vector of covariances between prediction point s and each of the sample points

Simple kriging

- Determines the optimal weights as those minimizing the mean square error (MSE) of the predictor

$$E\left(\hat{U}(s) - U(s)\right)^2 = \lambda^T(s) \mathrm{C} \lambda(s) + \sigma^2 - 2\lambda^T(s) c(s)$$

- Differentiating the MSE with respect to $\lambda(s)$

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

- The simple Kriging predictor is $\hat{U}(s) = \lambda^T(s) U = c^T(s) C^{-1} U$
- Corresponding MSE is $E\left(\hat{U}(s) - U(s)\right)^2 = \sigma^2 - c^T(s) C^{-1} c(s)$

which is referred to as *mean square prediction error* (*MSPE*), or *kriging variance*, denoted by σ_e^2

Simple Kriging

- With the optimal predictor $\hat{u}(s) = \lambda^T(s)u = c^T(s)C^{-1}u$
- The optimal predictor of y(s) is $\hat{y}(s) = \mu(s) + \hat{u}(s)$
- As simple kriging also gives the associated variance, the 95% confidence interval for our prediction can be constructed

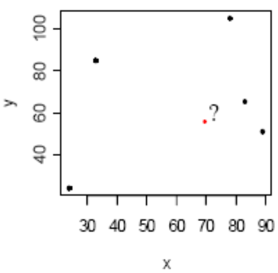
$$\hat{y}(s) \pm 1.96\sigma_e$$

  – Note that the prediction error for $y(s)$ is only from the predictor error for $u(s)$ because $\mu(s)$ is known

Example

- Data
  – Constant mean 160.37
  – Known covariogram $C(h) = 20e^{-3h/100}$

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



Example (Cont'd)

- Calculate the distances (h)

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

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

C

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

$c(s)$

6.896
5.032
10.15
8.083
4.079

Example (Cont'd)

- Calculate the inverse of C

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

- Calculate the weights

$$\lambda(s) = C^{-1}c(s) = \underbrace{\begin{bmatrix} 0.054914 & -0.01004 & -0.00646 & -0.00095 & -0.00746 \\ -0.01004 & 0.056751 & -0.01508 & 0.000168 & 0.000252 \\ -0.00646 & -0.01508 & 0.087403 & -0.05044 & -0.00194 \\ -0.00095 & 0.000168 & -0.05044 & 0.082023 & -0.00422 \\ -0.00746 & 0.000252 & -0.00194 & -0.00422 & 0.051936 \end{bmatrix}}_{C^{-1}} \times \underbrace{\begin{bmatrix} 6.896 \\ 5.032 \\ 10.15 \\ 8.083 \\ 4.079 \end{bmatrix}}_{c(s)} = \underbrace{\begin{bmatrix} 0.2225 \\ 0.066 \\ 0.351 \\ 0.128 \\ 0.107 \\ 0.877 \end{bmatrix}}_{\lambda(s)}$$

Example (Cont'd)

- The estimated residual

$$\hat{u}(s) = \lambda^T(s)u = \sum_{i=1}^{n} \lambda_i(s)u(s_i)$$
$$= 0.225\times(-38.27) + 0.066\times 22.63 + 0.351\times(-12.37) + 0.128\times(-0.37) + 0.107\times 15.63$$
$$= -9.86$$

- The estimated y(s)

$$\hat{y}(s) = \mu(s) + \hat{u}(s) = 160.37 - 9.86 = 150.5$$

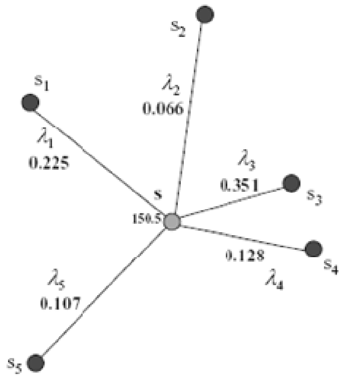
- The kriging variance

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

- 95% confidence interval

$$\hat{y}(s) \pm 1.96\sigma_e = 150.5 \pm 7.09$$

Check the Weights



Properties of Simple Kriging

- Simple kriging weights may be negative for some sample point I
- Simple kriging weights do not necessarily decrease with distance; clustered sites are down weighted
- Simple kriging variance does not depend on y