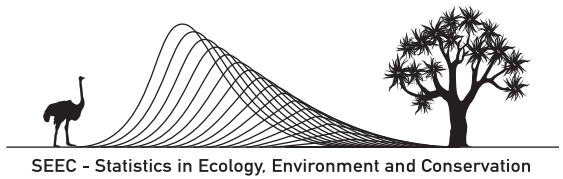
SEEC stats toolbox seminar series:

Spatial Interpolation

Mzabalazo Z. Ngwenya

Centre for Statistics in Ecology, Environment and Conservation (SEEC)



Department of Statistical Sciences
University of Cape Town

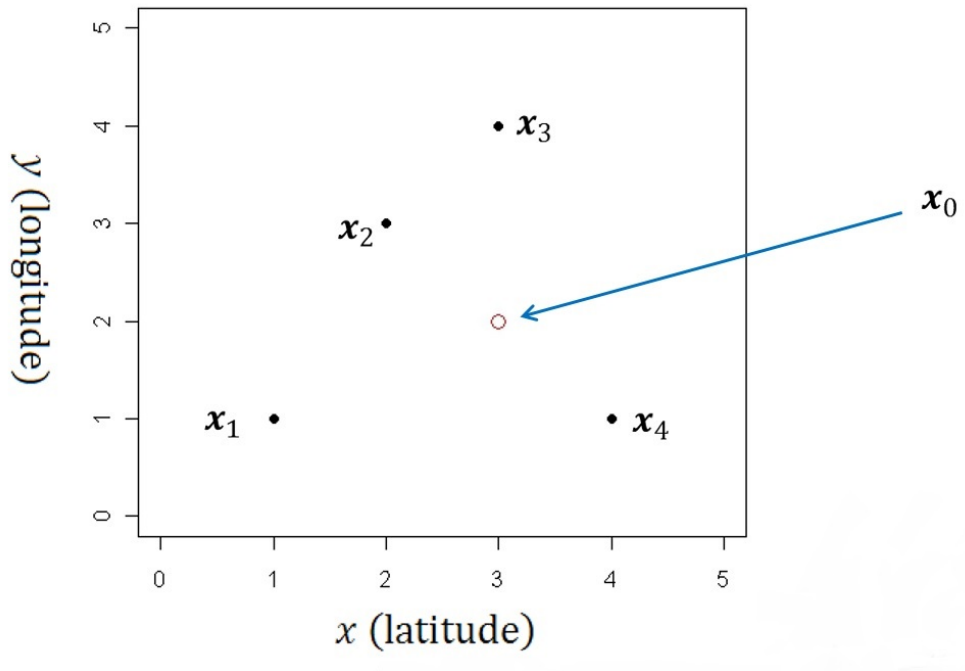
"everything is related to everything else, but near things are more related than distant things."

Tobler's first law of geography

Outline

1.1 Spatial Interpolation Problem

- Predict the value of a physical process at an unobserved location $\mathbf{x}_0$ using n observations $\mathbf{z} = [z(\mathbf{x}_1), \ldots, z(\mathbf{x}_n)]^\top$ obtained at n locations $\mathbf{x}_1, \ldots, \mathbf{x}_n$ within the study area D.



The dimensions over which the process $Z(\cdot)$operates is commonly $d = 2$ but may also be $d = 1$ or $d = 3$

- General approach to interpolation involves expressing a prediction of the unknown value $z(\mathbf{x}_0)$ as a weighted average of the neighbouring observations;

[1] A detailed review of interpolation methods and categories is presented in Lee and Heap (2014).

- General approach to interpolation involves expressing a prediction of the unknown value $z(\mathbf{x}_0)$ as a weighted average of the neighbouring observations;

$$\hat{z}(\mathbf{x}_0) = \sum_{i=1}^{n(\mathbf{x}_0)} \lambda_i z(\mathbf{x}_i)$$

where:
$n(\mathbf{x}_0)$ is the number of points in a search neighbourhood around the prediction point.

λ_i is the weight associated with the i^{th} observation $z(\mathbf{x}_i)$

[1] A detailed review of interpolation methods and categories is presented in Lee and Heap (2014).

- General approach to interpolation involves expressing a prediction of the unknown value $z(\mathbf{x}_0)$ as a weighted average of the neighbouring observations;

$$\hat{z}(\mathbf{x}_0) = \sum_{i=1}^{n(\mathbf{x}_0)} \lambda_i z(\mathbf{x}_i)$$

where:
$n(\mathbf{x}_0)$ is the number of points in a search neighbourhood around the prediction point.

λ_i is the weight associated with the i^{th} observation $z(\mathbf{x}_i)$

- The are two broad classes of interpolation methods.[1]

[1] A detailed review of interpolation methods and categories is presented in Lee and Heap (2014).

1.2 Deterministic methods

- Use mathematical functions used to predict unknown values.

  Examples:
  - Global and local polynomials
  - Inverse distance weighting
  - Splines

- No strict assumptions about the variability of a process are made.

- Arbitrary or empirical model parameters used.

- Prediction accuracy of model determined via use cross-validation to obtain an estimate the Root Mean Square Prediction Error (RMSPE);

$$\mathrm{RMSPE} = \frac{1}{n}\sqrt{\sum_{i=1}^{n}\{z(\mathbf{x}_i) - \hat{z}_{(-i)}(\mathbf{x}_i)\}}$$

  where $z(\mathbf{x}_i)$ is the observed value at $\mathbf{x}_i$ and $\hat{z}_{(-i)}(\mathbf{x}_i)$ is the predicted value obtained by fitting over all locations other than $\mathbf{x}_i$.

- The RMSPE measures the overall accuracy of the predictions (average prediction error).

1.3 Stochastic methods

- Use both mathematical and statistical techniques to predict values at all locations within D.

  Examples:

  - Machine learning based; SVM, RF, NN
  - Bayesian-based models; Bayesian Maximum Entropy
  - Geostatistical; kriging

- Strict assumptions made about the physical process

$$Z(\cdot) = \{Z(\mathbf{x}) : \mathbf{x} \in D \subset \mathbb{R}^d\}.$$

- Model parameters estimated objectively following probability theory.

- For stochastic methods the accuracy of predictions can not only be judged by use of the RMSPE but may also be gauged using the Mean Squared Prediction Error (MSPE)

$$\mathrm{E}\{[z(\mathbf{x}_0)-\hat{z}(\mathbf{x}_0)]^2\} = \mathrm{Var}\{z(\mathbf{x}_0)\}+\mathrm{Var}\{\hat{z}(\mathbf{x}_0)\}-2\cdot\mathrm{Cov}\{z(\mathbf{x}_0),\hat{z}(\mathbf{x}_0)\}$$

- The MSPE provides a measure of the prediction error of each prediction $\hat{z}(\mathbf{x}_0)$.
- This allows one to not only create a prediction surface but also create a surface of prediction errors.
- Can also calculate confidence intervals for predictions.

Other categorizations of interpolation methods

- Exact vs Approximate:
  In exact interpolation methods the predicted values match exactly the observed values at the sample points, $\mathbf{x}_i$, $i = 1, \ldots, n$, that is $\hat{z}(\mathbf{x}_i) = z(\mathbf{x}_i)$. In approximate interpolation measured and predicted values do not have to coincide.

- Convex vs Non-convex:
  In convex interpolation all predicted values $\hat{z}(\mathbf{x}_0)$ lie within the range of the sample values $\mathbf{z} = [z(\mathbf{x}_1), \ldots, z(\mathbf{x}_n)]^\top$. In non-convex interpolation some predictions may be below or above the range of sampled values.

- Local vs Global:
  Global interpolation methods are those methods that use the complete data set in making a prediction at an unobserved locale $\mathbf{x}_0$. Local methods use a subset of points around each prediction $\mathbf{x}_0$ to make a prediction at that point.

2.1 Inverse Distance Weighting

- Exact and convex interpolation method.
- Global method, but in practise neighbourhoods are defined around each prediction point to speed up computation on large datasets.
- The weights are a simple function of the distance between individual locations $\mathbf{x}_i$, $i = 1, \ldots, n$ and the prediction point $\mathbf{x}_0$, usually $1/||\mathbf{x}_i - \mathbf{x}_0||$ such that

$$\lambda_i = \left(\frac{1}{||\mathbf{x}_i - \mathbf{x}_0||} \right)^p$$

where p is some integer.

- The higher the value of p the less importance will be put on distant points.
- Inverse squared distance, which is $p = 2$, is widely used.
- A more objective way of determining p is cross-validation; i.e. trying different values of p and see which one gives you the smallest RMSPE.
- Predictions sensitive to sample location $\mathbf{x}_1, \ldots, \mathbf{x}_n$ and the presence of outliers.
- Very suitable for interpolating surfaces of phenomenon that can change quickly over short spatial scales.

3.1 Splines

- Splines are piecewise polynomial functions
- Local, non-convex and exact.
- Most widely used splines are thin-plate splines and the regularized spline with tension and smoothing, the latter gives one more control over the nature of the fitted surface.
- For regularized spline with tension and smoothing predictions are obtained as

$$\hat{z}(\mathbf{x}_0) = a_1 + \sum_{i=1}^{n} w_i \cdot R(v_i)$$

where a_1 is a constant and $R(v_i)$ is the radial basis function. [2]

[2] Hengl (2007) gives details of the most common implmentation of the regularized spline with tension.

- The coefficients a_1 and w_i are obtained by solving the following system of equations;

$$\sum_{i=1}^{n} w_i = 0$$

$$a_1 + \sum_{i=1}^{n} w_i \cdot \left[R(v_i) + \delta_{ij} \cdot \frac{\vartheta_0}{\vartheta_i} \right] = z(\mathbf{x}_i)$$

  where ϑ_0/ϑ_i is a positive weighting factor representing the smoothing parameter at each prediction point $\mathbf{x}_0$.

- The tension parameter φ controls the distance over which points influence each other. The smoothing ϑ_0/ϑ_i controls the vertical deviation of the surface from the points.

- Smoothing and tension parameters have to be determined by user. Can use cross-validation to determine the best values for these parameters.
- For Regularized spline: The higher the weight, the smoother the surface, weights between 0 and 5 are recommended.

  For tension splines the higher the weight, the coarser the surface and the more predicted values conform to the range of sample data. Weight values must be greater than or equal to zero.
- Works best for interpolating gently varying surfaces.

4.1 Kriging model

- Non-convex interpolator that can be exact or inexact.
- Based on the concept that spatial variables can be considered as partly deterministic and partly stochastic

$$\mathbf{Z} = \boldsymbol{\mu} + \boldsymbol{\delta}; \quad \boldsymbol{\delta} \sim (\mathbf{0}, \mathbf{V}(\boldsymbol{\theta}))$$

  where $\boldsymbol{\mu} = \mathrm{E}\{\mathbf{Z}\}$ is a deterministic mean and $\boldsymbol{\delta}$ is stochastic component which represents spatial variation.
- The deterministic trend $\boldsymbol{\mu}$ is commonly modelled as a function of the observation locations $\mathbf{x}_1, \ldots, \mathbf{x}_n$ by polynomial functions $f_k(\cdot)$ and unknown parameters β_k such that $\boldsymbol{\mu} = \mathbf{F}\boldsymbol{\beta}$.

- Under the kriging model $\boldsymbol{\delta}$ is as a zero mean process with variance-covariance matrix $\mathbf{V}(\boldsymbol{\theta})$ parameterized by q parameters $[\theta_1, \ldots, \theta_q]$.
- $\boldsymbol{\delta}$ can be further decomposed into two components;

$$\boldsymbol{\delta} = \boldsymbol{\eta} + \boldsymbol{\epsilon}$$

.

- The first component is a structured random process and the second is noise.

- The q parameters $\boldsymbol{\theta} = [\theta_1, \ldots, \theta_q]$ of the variance-covariance matrix correspond to some aspect of the correlation structure of $Z(\cdot)$;

  $\theta_1 = \tau^2$: variance of random noise process (nugget)

  $\theta_2 = \sigma_0^2$: variance of the structured stochastic process (partial sill)

  $\theta_3 = \phi$: the maximum distance over which any two points have an effect on each other (range)

  $\theta_4 = \kappa$: a parameter that describes the smoothness of the process.

- The q parameters $\boldsymbol{\theta} = [\theta_1, \ldots, \theta_q]$ of the variance-covariance matrix correspond to some aspect of the correlation structure of $Z(\cdot)$;

  $\theta_1 = \tau^2$: variance of random noise process (nugget)

  $\theta_2 = \sigma_0^2$: variance of the structured stochastic process (partial sill)

  $\theta_3 = \phi$: the maximum distance over which any two points have an effect on each other (range)

  $\theta_4 = \kappa$: a parameter that describes the smoothness of the process.

- These parameters correspond to what is termed the semivariogram $\gamma(h)$ of the process.

- The q parameters $\boldsymbol{\theta} = [\theta_1, \ldots, \theta_q]$ of the variance-covariance matrix correspond to some aspect of the correlation structure of $Z(\cdot)$;

  $\theta_1 = \tau^2$: variance of random noise process (nugget)

  $\theta_2 = \sigma_0^2$: variance of the structured stochastic process (partial sill)

  $\theta_3 = \phi$: the maximum distance over which any two points have an effect on each other (range)

  $\theta_4 = \kappa$: a parameter that describes the smoothness of the process.

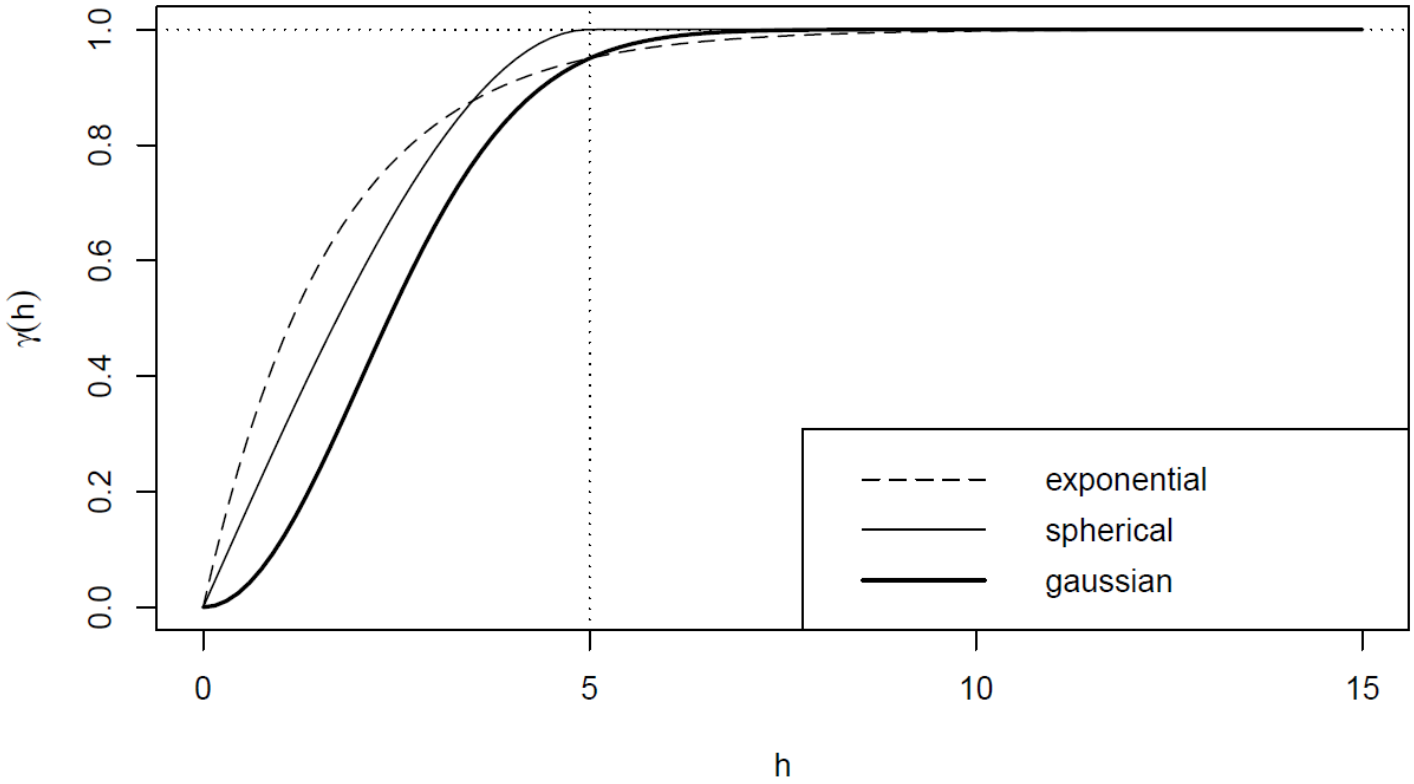
- These parameters correspond to what is termed the semivariogram $\gamma(h)$ of the process.

- The semivariogram of $\gamma(h)$ is commonly modelled through parametric functions of the distance between points h.

Some common parametric semivariogram models:

Model	$\gamma(h)$
Exponential	$\tau^2 + \sigma_0^2\{1 - e^{-3h}/\phi\}$
Gaussian	$\tau^2 + \sigma_0^2\{1 - e^{-3h^2/\phi^2}\}$
Matérn	$\tau^2 + \sigma_0^2\left\{1 - \frac{1}{2^{\kappa-1}\Gamma(\kappa)}\left(\frac{h}{\phi}\right)^{\kappa} K_\kappa\left(\frac{h}{\phi}\right)\right\}$
Cauchy	$\tau^2 + \sigma_0^2\left\{1 - \left[1 + \left(\frac{h}{\phi}\right)^2\right]^{-\kappa}\right\}$
Circular	$\tau^2 + \sigma_0^2\left\{1 - \frac{2}{\pi}\left[\arccos\left(\frac{h}{\phi}\right) - \left(\frac{h}{\phi}\right)\sqrt{1 - \left(\frac{h}{\phi}\right)^2}\right]\right\}$, for $h \leq \phi$, 0, otherwise
Spherical	$\tau^2 + \sigma_0^2\left\{\frac{3h}{2\phi} - \frac{h^3}{2\phi^3}\right\}$, for $0 < h \leq \phi$, $\tau^2 + \sigma_0^2$, for $h \geq \phi$
Cardinal-sine (Wave)	$\tau^2 + \sigma_0^2\left\{1 - \left(\frac{\phi}{h}\right)\sin\left(\frac{h}{\phi}\right)\right\}$

An example of the shape of semivariogram models:



Models have the same range $\phi = 5$, partial sill $\sigma^2 = 1$ and nugget $\tau^2 = 0$.

- In kriging the weights in

$$\hat{z}(\mathbf{x}_0) = \sum_{i=1}^{n(\mathbf{x}_0)} \lambda_i z(\mathbf{x}_i)$$

are determined by minimizing the prediction error

$$\mathrm{E}\{[z(\mathbf{x}_0) - \hat{z}(\mathbf{x}_0)]^2\}.$$

- The main assumption of kriging models is stationarity of the process. Various levels of stationarity may be assumed;
  - Strict stationarity :
    The mean and variance of $Z(\cdot)$ are constant throughout the domain D.
  - Second order stationarity :
    The mean is the same everywhere and for each pair of variables the covariance exists and depends only on the distance between points h.
  - Intrinsic stationarity:
    the mean is not the same everywhere however the covariance exists and depends only on the distance between points h.
- The stationarity assumption made affects the form of the mean.

- $\mathbf{F}\boldsymbol{\beta}$ can be modified to give a variety of models.

Type	Assumption	Mean model
Universal kriging	non-stationary, unknown mean	$\mathbf{F}\boldsymbol{\beta}$
Ordinary kriging	constant, unknown mean	$\mathbf{1}\beta$
Simple kriging	constant, known mean β	$\mathbf{1}\beta$

4.2.1 Least squares approach:

- The q parameters of $\boldsymbol{\theta}$ are commonly obtained via least squares methods.[3]
- This approach is based on fitting theocratical semivariogram models to experimental semivariograms curves in a two step approach.
  1. Estimate the (empirical) semivariogram $\hat{\boldsymbol{\gamma}} = [\hat{\gamma}(h_1), \ldots, \hat{\gamma}(h_M)]^\top$.

[3] Cressie (1993) gives a detailed of account of the implementation different of least squares algorithms to solve this problem.

4.2 Model fitting

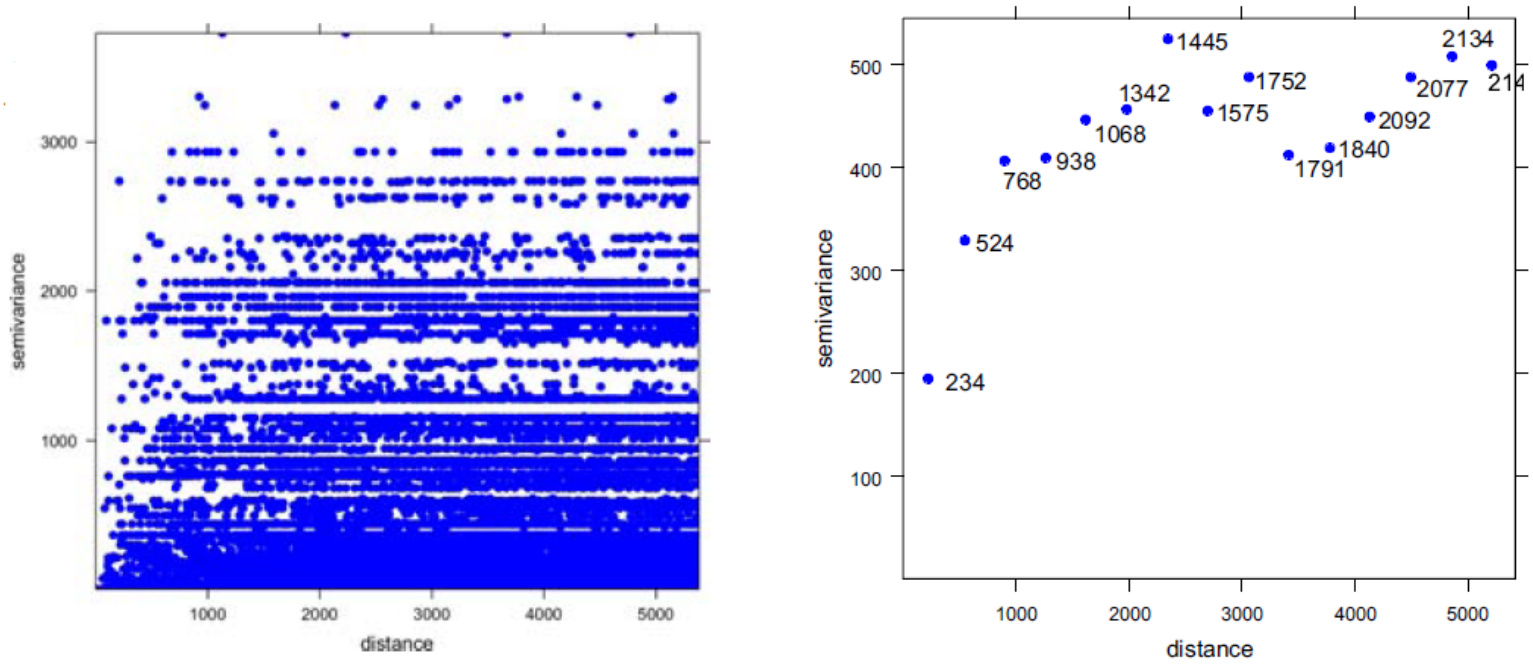
4.2.1 Least squares approach:

- The q parameters of $\boldsymbol{\theta}$ are commonly obtained via least squares methods.[3]
- This approach is based on fitting theocratical semivariogram models to experimental semivariograms curves in a two step approach.
  1. Estimate the (empirical) semivariogram $\hat{\boldsymbol{\gamma}} = [\hat{\gamma}(h_1), \ldots, \hat{\gamma}(h_M)]^\top$.
  2. Fit a theocratical model to experimental semivariogram by minimizing the sum of squares.

[3] Cressie (1993) gives a detailed of account of the implementation different of least squares algorithms to solve this problem.

Step 1:

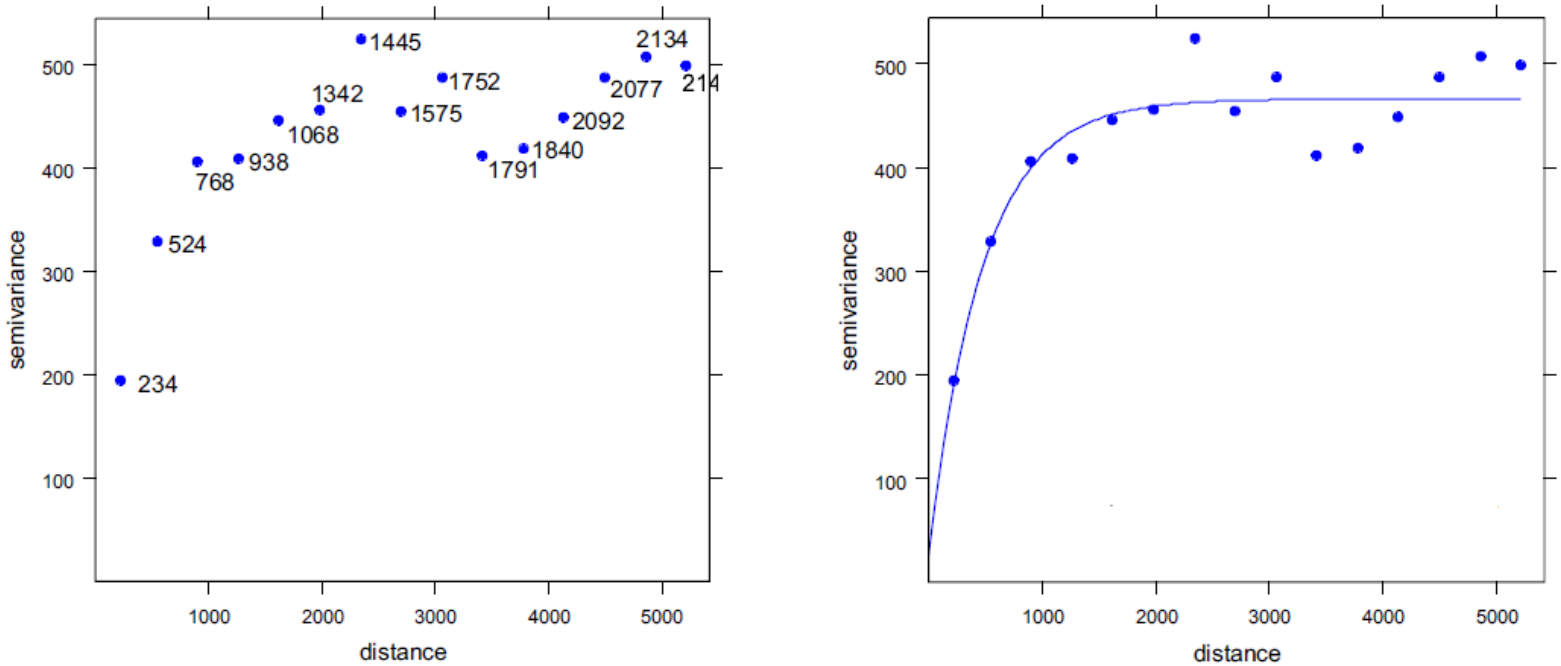
Calculate the semivariance between all points to obtain a semivariogram cloud. Use an appropriate estimator to determine obtain estimates of $\gamma(h)$ *at each of these lags;* $\hat{\gamma}(h_1), \ldots, \hat{\gamma}(h_M)$



Source: Hengl (2007).

$$\hat{\gamma}(h_{ij}) = \frac{1}{2|N(h_{ij})|} \sum_{N(h_{ij})} \{z(\mathbf{x}_i) - z(\mathbf{x}_j)\}^2, \qquad h_{ij} \in \mathbb{R}^d$$

Step 2:
Fit a theocratical semivariogram model to experimental semivariogram by minimizing the sum of squares.



Source: Hengl (2007).

- This is however a subjective method.
- Subjectively arises in the process of constructing the empirical semivariogram $[\hat{\gamma}(h_1), \ldots, \hat{\gamma}(h_M)]^\top$.
- No formal way of determining optimal number of lags M.
- No formal way to determine the number of points $N(h_{ij})$ to use in calculating each $\hat{\gamma}(h_{ij})$.

4.2.2 Likelihood estimation approach

- A more objective method of estimating $\boldsymbol{\theta}$ is to use Likelihood based estimation methods.[4]

- ML estimation of the semivariogram parameters involves maximizing the negative log-likelihood

$$l(\boldsymbol{\theta}; \boldsymbol{\beta}, \mathbf{Z}) = -\frac{1}{2}\Big[n \ \ln\{2\pi\} + \ln\{|\mathbf{V}(\boldsymbol{\theta})|\} + (\mathbf{Z} - \mathbf{F}\boldsymbol{\beta})^{\top}\mathbf{V}(\boldsymbol{\theta})^{-1}(\mathbf{Z} - \mathbf{F}\boldsymbol{\beta})\Big].$$

- REML estimation of $\boldsymbol{\theta}$ involves

$$\begin{aligned} l_r(\boldsymbol{\theta}; \mathbf{KZ}) = -\frac{1}{2}\Big[&(n-K-1)\ln\{2\pi\} + \ln\{|\mathbf{F}^{\top}\mathbf{V}(\boldsymbol{\theta})^{-1}\mathbf{F}| + \ln\{|\mathbf{V}(\boldsymbol{\theta})| \\ &+ \ln\{|\mathbf{F}^{\top}\mathbf{F}| + (\mathbf{Z} - \mathbf{F}\boldsymbol{\beta})^{\top}\mathbf{V}(\boldsymbol{\theta})^{-1}(\mathbf{Z} - \mathbf{F}\boldsymbol{\beta})\}\Big]. \end{aligned}$$

where the matrix of contrasts $\mathbf{K}$ is chosen so that $\mathrm{E}\{\mathbf{KZ}\} = \mathbf{0}$.

[4] Diggle and Riberio (2007) give a good account of this approach

- Can compute goodness-of-fit measures such as *AIC* and *BIC* allowing one to compare the fits of competing models.

$$AIC = -2 \cdot \ln\{\hat{L}\} + 2 \cdot p$$
$$BIC = -2 \cdot \ln\{\hat{L}\} + p \cdot \log\{n\}$$

where $\hat{L}$ is the value of the maximized log likelihood, p is the number of parameters in the model and n is the number of observations.

With information criterion based methods such the *AIC* and *BIC* you can compare nested and/or non-nested models.

4.3 Estimating the prediction error at a point

- As noted earlier, major advantage of stochastic interpolation methods is that they not only provide a method of judging overall prediction accuracy but also the accuracy of individual predictions $\hat{z}(\mathbf{x}_0)$.
- In kriging this prediction error is called the kriging variance.
- There exist several estimators of the kriging variance.[5]

[5] Thiart, C. Ngwenya, MZ and Haines, LM (2014) give details on the different estimators and their sources of bias. Performance of the estimators is also discussed.

- Analytic estimators;
  - Traditional (plug-in) estimator
  - Kacker-Harville estimator
  - Prasad-Rao estimator
- Bootstrap estimators;
  - Unconditional Bootstrap estimator
  - Conditional Bootstrap estimator

- Although the traditional plug in estimator is widely used it is known to underestimate the prediction error.
- The Prasad-Rao estimator is to be preferred.
- REML estimation of the parameters leads to least biased estimates.

Bibliography


Cressie, N. A. C. (1993)
Statistics for Spatial Data (revised edition), John Wiley and Sons, New York.


Diggle, P.J. and Riberio, P.J. (2007)
Model-based geostatistics. Springer, New York.


Goovaerts, P.(1997)
Geostatistics for Natural Resources Evaluation. Oxford University Press, Oxford.


Hengl, T.(2007)
A Practical Guide to Geostatistical Mapping of Environmental Variables. European Commision, Italy.


Li, J. and Heap A.D. (2014).
Spatial interpolation methods applied in environmental sciences: A review.
Environmental Modelling and Software, **53**, 173-189.


Thiart, C Ngwenya, MZ and Haines, LM (2014).
Investigating'optimal'kriging variance estimation using an analytic and a bootstrap approach.
Journal of the Southern African Institute of Mining and Metallurgy, **114**, 613-619.

Take home!!!

- Use kriging for interpolation; (assumptions are as restrictive as they appear).
- Fit models using likelihood methods; REML is to be preferred here.
- Estimate the kriging variance using the Prasad-Rao estimator.