Geostatistics: Kriging

8.10.2015

Konetekniikka 1, Otakaari 4, 150

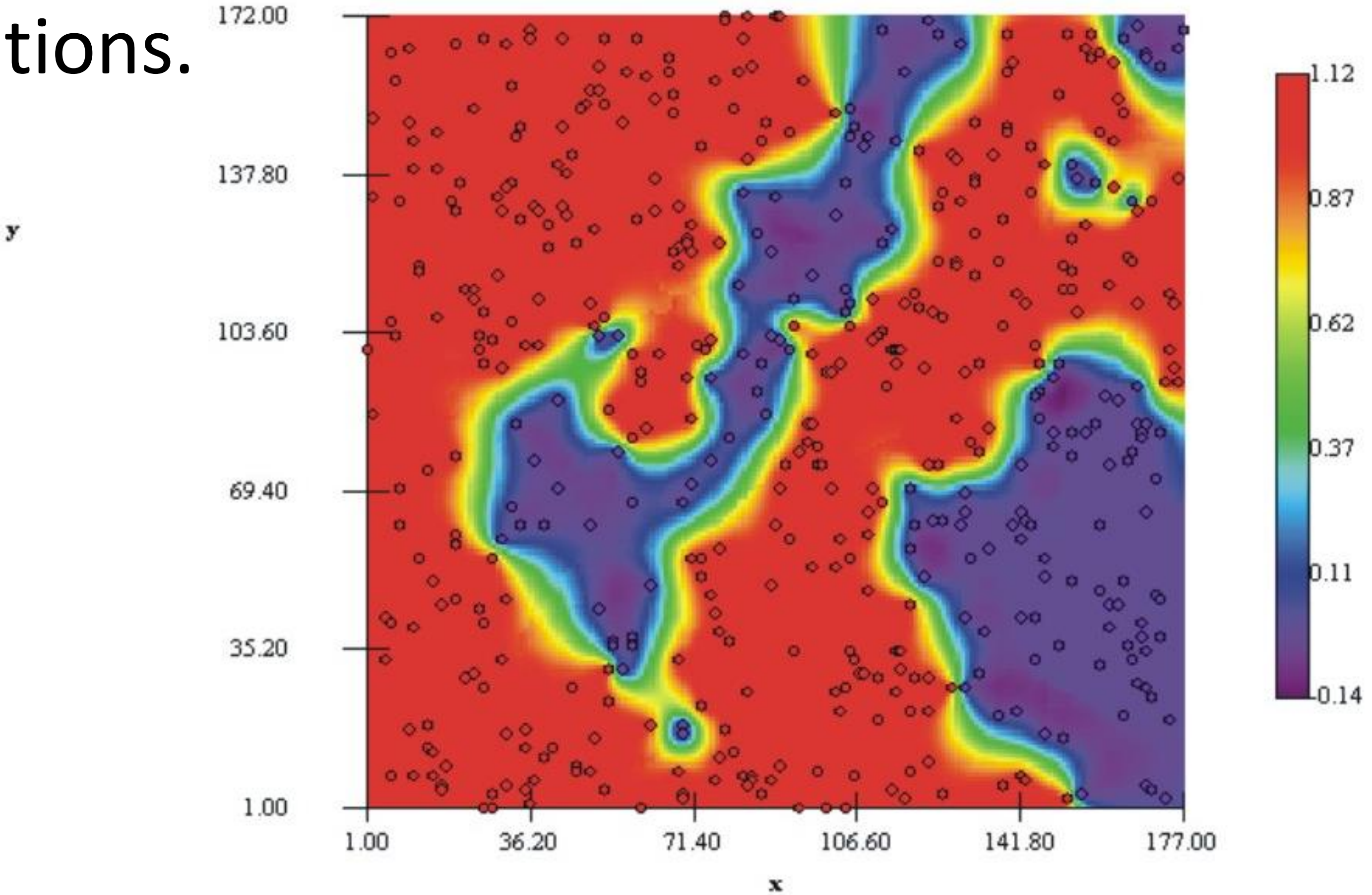
10-12

Rangsima Sunila, D.Sc.

- Background
- What is Geostatitics
- Concepts
- Variogram: experimental, theoretical
- Anisotropy, Isotropy
- Lag, Sill, Range, Nugget
- Types of Kriging
- Example of kriging interpolation

Interpolation

- How to estimate unknown values at specific locations.

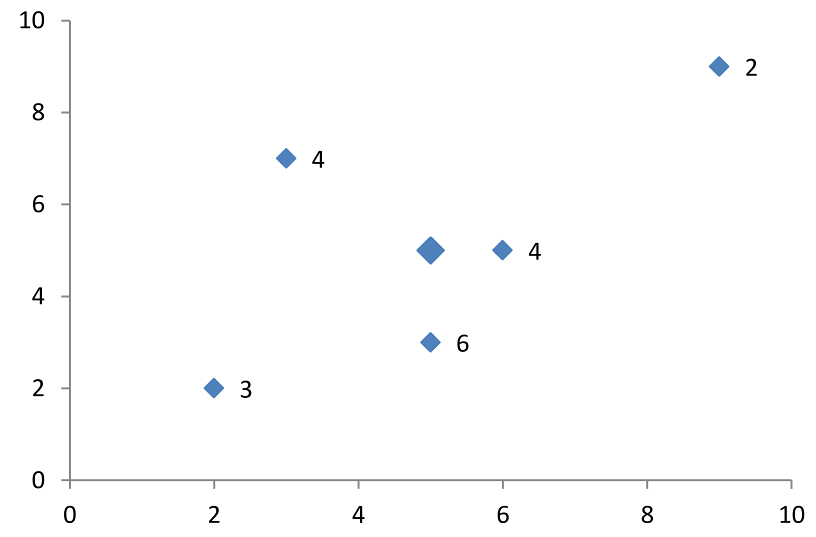


Spatial Interpolation

- Examples
  - Trend surfaces
  - Nearest neighbours: Thiessen(voronoi)
  - Inverse distance weighting (IDW)
  - Splines
  - Kriging

Example:

Site	x	y	z	D to (5,5)
1	2	2	3	4.2426
2	3	7	4	2.8284
3	9	9	2	5.6569
4	6	5	4	1.0000
5	5	3	6	2.0000



We would like to estimate the variable value at (5,5)

Example: IDW

- Value of z(x) is estimated from all known values of z at all n points. (Weighted Moving Average technique)

$$z(x) = \sum_{i=1}^{n} w_i z_i$$

Weights usually add to 1, $\sum_{i=1}^{n} w_i = 1$

Example: IDW

- In IDW, the weights weights are based on the distance from each of the known points (i) to the point we are trying to estimate (k): d_{ik}. In IDW, we consider the inverse distance, $1/d_{ik}$

$$w_i = \frac{\frac{1}{d_{ik}}}{\sum_{1=1}^{n} \frac{1}{d_{ik}}}$$

Example: IDW

Location (x,y)	z	D to (5,5)	ID	Weights
(2,2)	3	4.2426	0.2357	0.1040
(3,7)	4	2.8284	0.3536	0.1560
(9,9)	2	5.6569	0.1768	0.0780
(6,5)	4	1.0000	1	0.4413
(5,3)	6	2.0000	0.5	0.2207
sum			2.2661	1

Z(5,5) = 0.1040(3) + 0.1560(4) + 0.0780(2) + 0.4413(4) + 0.2207(6)
= 4.1814

Historical background

- Geostatistics, first developed by Georges Matheron (1930-2000), the French geomathematician. The major concepts and theory were discovered during 1954-1963 while he was working with the French Geological Survey in Algeria and France.
- In 1963, he defined the linear geostatistics and concepts of variography, varaiances of estimation and kriging (named after Danie Krige) in the *Traité de géostatistique appliquée.* The principles of geostatistics was published in Economic Geology Vol. 58, 1246-1266.
- Kriging was named in honour of Danie Krige (1919-2013), the South African mining engineer who developed the methods of interpolation.

What is Geostatistics

- Techniques which are used for mapping of surfaces from limited sample data and the estimation of values at unsampled locations
- Geostatistics is used for:
  - spatial data modelling
  - characterizing the spatial variation
  - spatial interpolation
  - simulation
  - optimization of sampling
  - characterizing the uncertainty
- The idea of geostatistics is <u>the points which are close to each other in the space should be likely close in values.</u>

- Mining
- Geography
- Geology
- Geophysics
- Oceanography
- Hydrography
- Meterology
- Biotechnology
- Enviromental studies
- Agriculture

Geostatistical methods provide

- How to deal with the limitations of deterministic interpolation
- The prediction of attribute values at unvisited points is optimal
- BLUE (Best Linear Unbiased Estimate)

Geostatistical method for interpolation

- Reconigtion that the spatial variation of any continuous attribute is often too irregular to be modelled by a simple mathematical function.
- The variation can be described better by a stochastic surface.
- The interpolation with geostatistics is known as kriging.

Variogram cloud

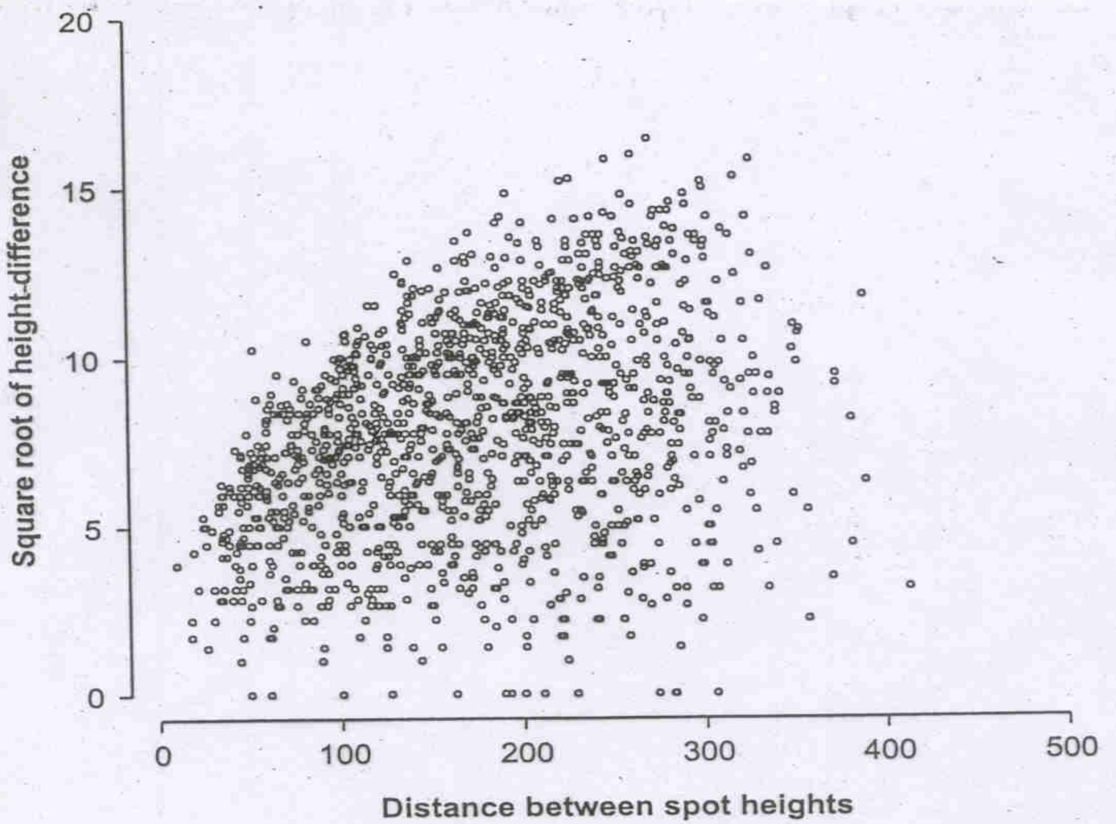


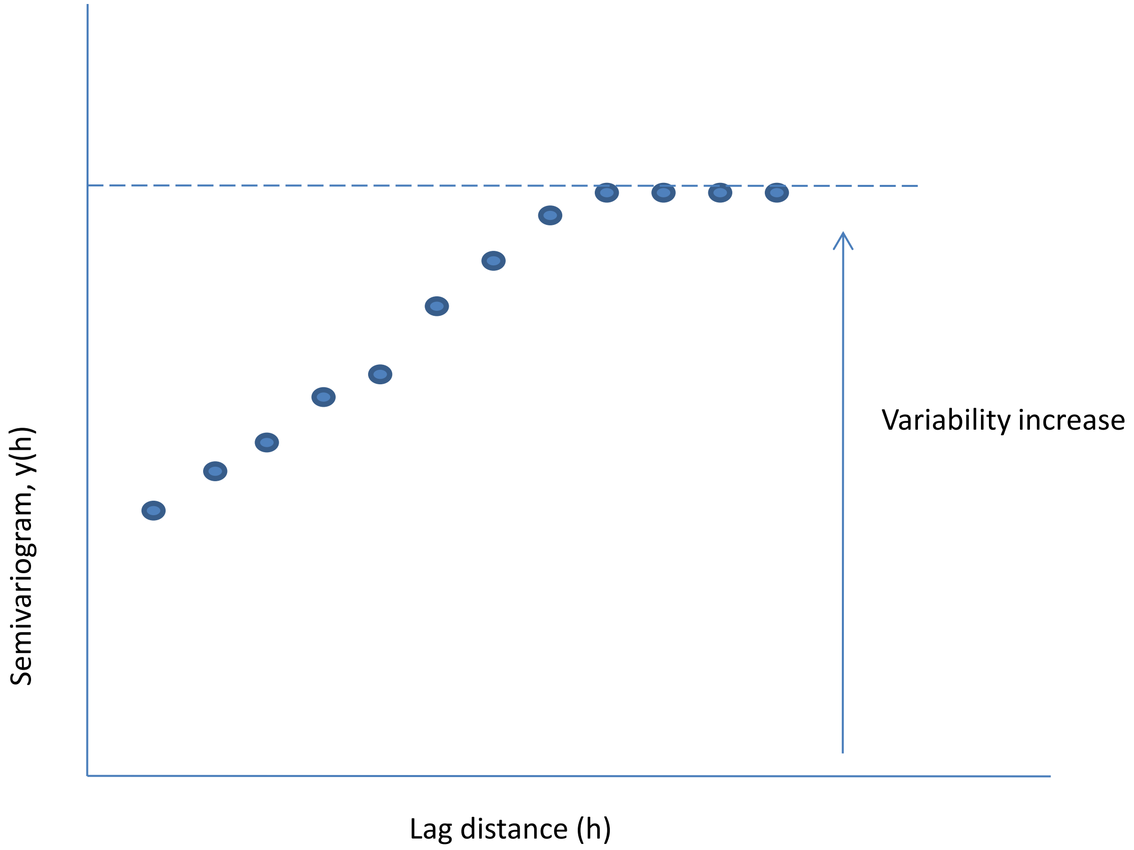
Figure 2.7 Variogram cloud for the spot height data in Figure 2.6.

O´Sullivan & Unwin, 2002)

Variogram/Semivariogram

- Variogram is the variance of the difference random variables at two locations
- To examine the spatial continuity of a regionalized variable and how this continuity changes as a function of distance and direction.
- The computation of a variogram involves plotting the relationship between the semivariance and the lag distance
- Measure the strength of correlation as a function of distance
- Quantify the spatial autocorrelation

Variogram



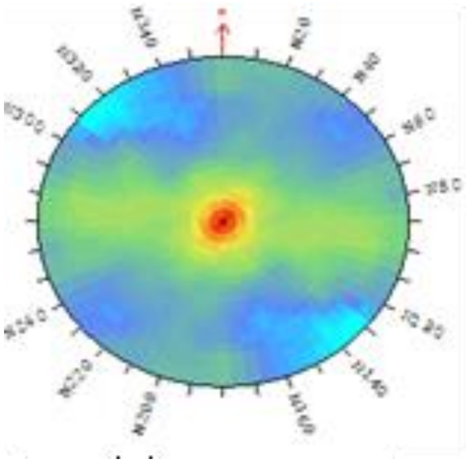
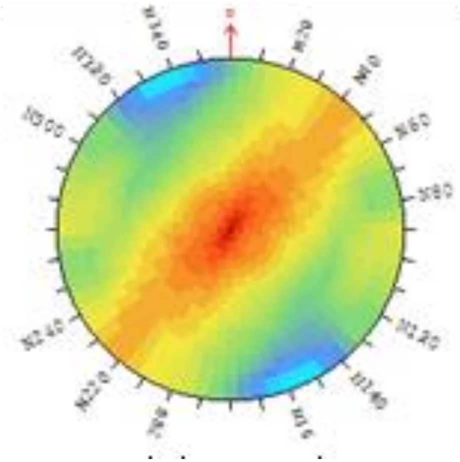
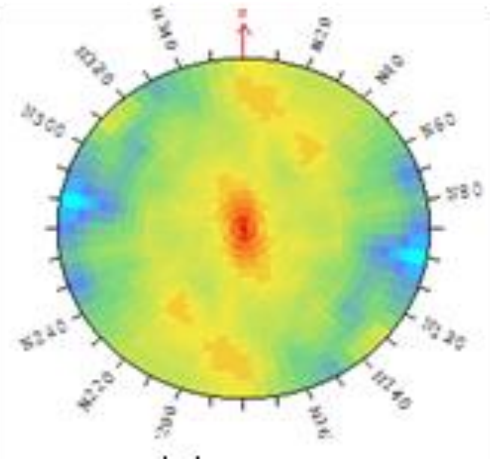
Variograms

- Half of average squared difference between the paired data values.
- The variogram calculated by

$$\gamma(h) = \frac{1}{2N(h)} \sum_{(i,j)|h_{ij}=h} (x_i - x_j)^2$$

Variogram

- Experimental variogram (sample or observed variogram) :
  - when variogram is computed from sampled data.
  - The first step towards a quantitative description of the regionalized variation.
- Theoretical variogram or variogram model:
  - when it is modelled to fit the experimental variogram.

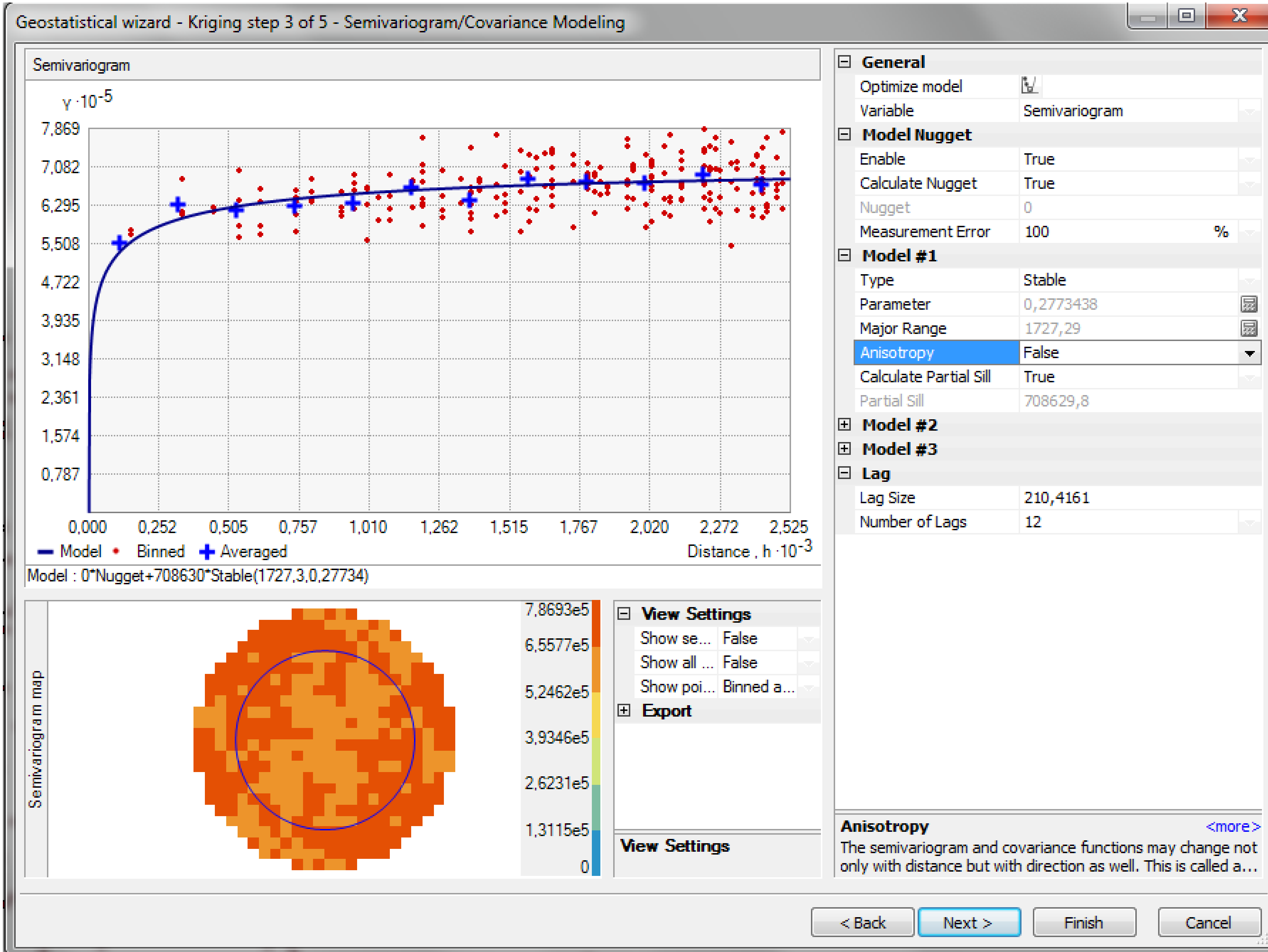




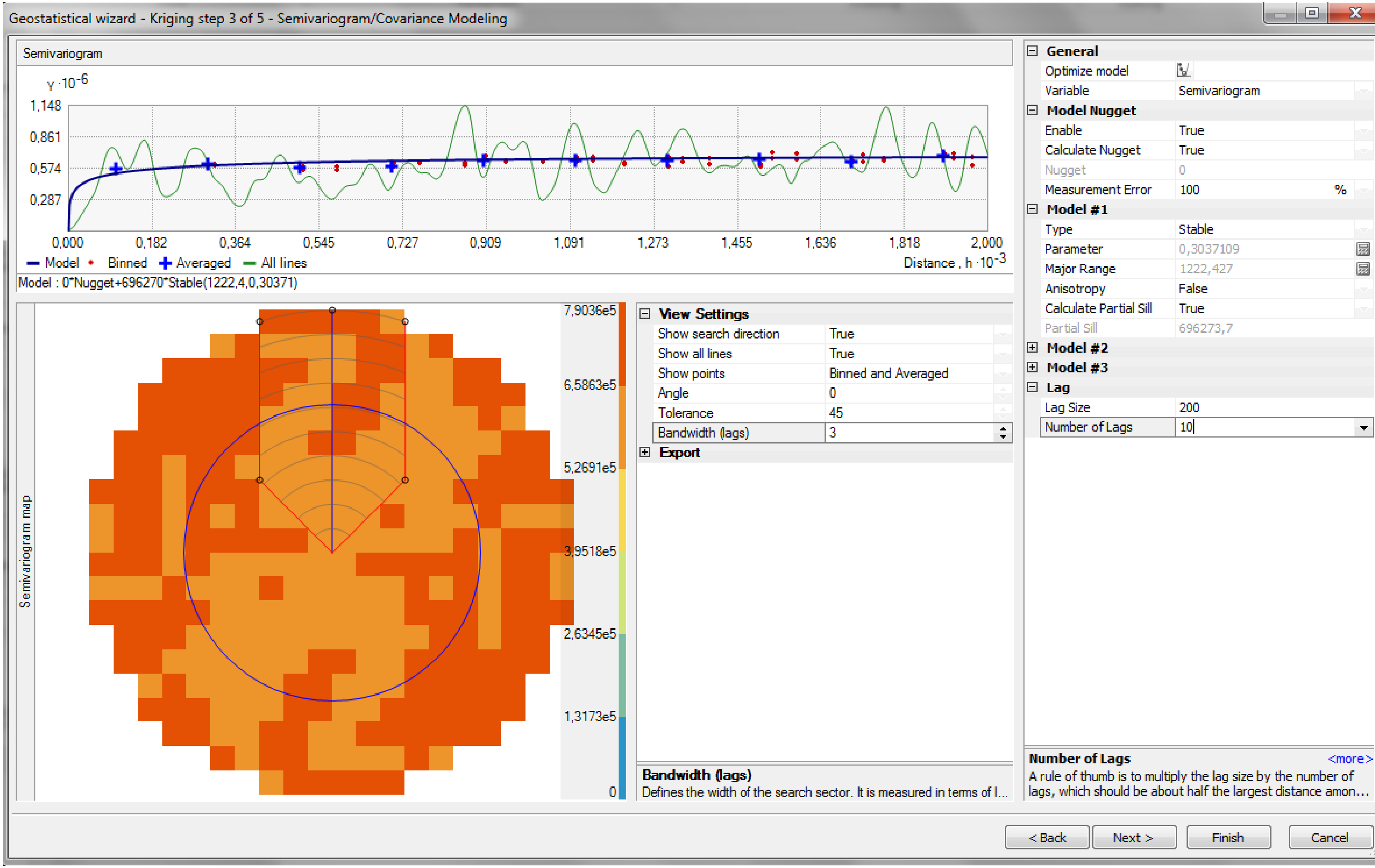
Omnidirectional variogram

- Omnidirectional variogram is a test for erratic directional variograms
- The omnidirectional variogram contains more sample pairs than any directional variogram so it is more likely to show a clearly interpretable structure.
- If the omnidirectional variogram is messy, then try to discover the reasons for the erraticness, e.g. Examine the h-scatterplots may reveal that a single sample value shows large influence on the calculations.

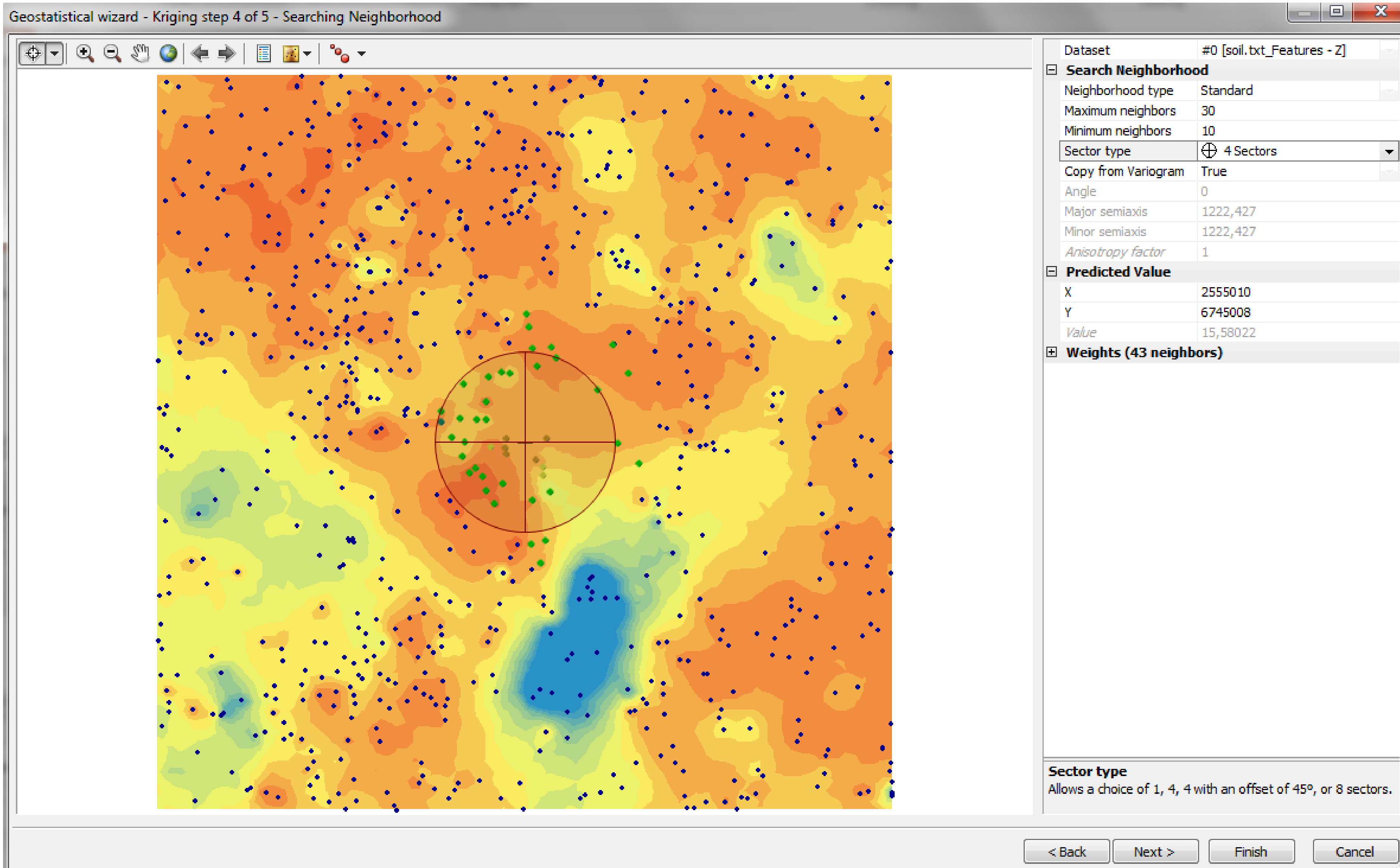
Isotropy

- The spatial correlation structure has no directional effects, the resulting variogram averages the variogram over all directions.
- The covariance function, correlogram, and semivariogram depend only on the magnitude of the lag vector h and not the direction
- The empirical semivariogram can be computed by pooling data pairs separated by the appropriate distances, regardless of direction.
- The semevariogram describes omnidirectional.


Geostatistical wizard - Kriging step 3 of 5 - Semivariogram/Covariance Modeling
Semivariogram
γ ·10^-5
7,869
7,082
6,295
5,508
4,722
3,935
3,148
2,361
1,574
0,787
0,000
0,252
0,505
0,757
1,010
1,262
1,515
1,767
2,020
2,272
2,525
Model
Binned
Averaged
Distance , h ·10^-3
Model : 0*Nugget+708630*Stable(1727,3,0,27734)
Semivariogram map
7,8693e5
6,5577e5
5,2462e5
3,9346e5
2,6231e5
1,3115e5
0
View Settings
Show se... False
Show all ... False
Show poi... Binned a...
Export
View Settings
General
Optimize model
Variable Semivariogram
Model Nugget
Enable True
Calculate Nugget True
Nugget 0
Measurement Error 100 %
Model #1
Type Stable
Parameter 0,2773438
Major Range 1727,29
Anisotropy False
Calculate Partial Sill True
Partial Sill 708629,8
Model #2
Model #3
Lag
Lag Size 210,4161
Number of Lags 12
Anisotropy <more>
The semivariogram and covariance functions may change not only with distance but with direction as well. This is called a...
< Back
Next >
Finish
Cancel

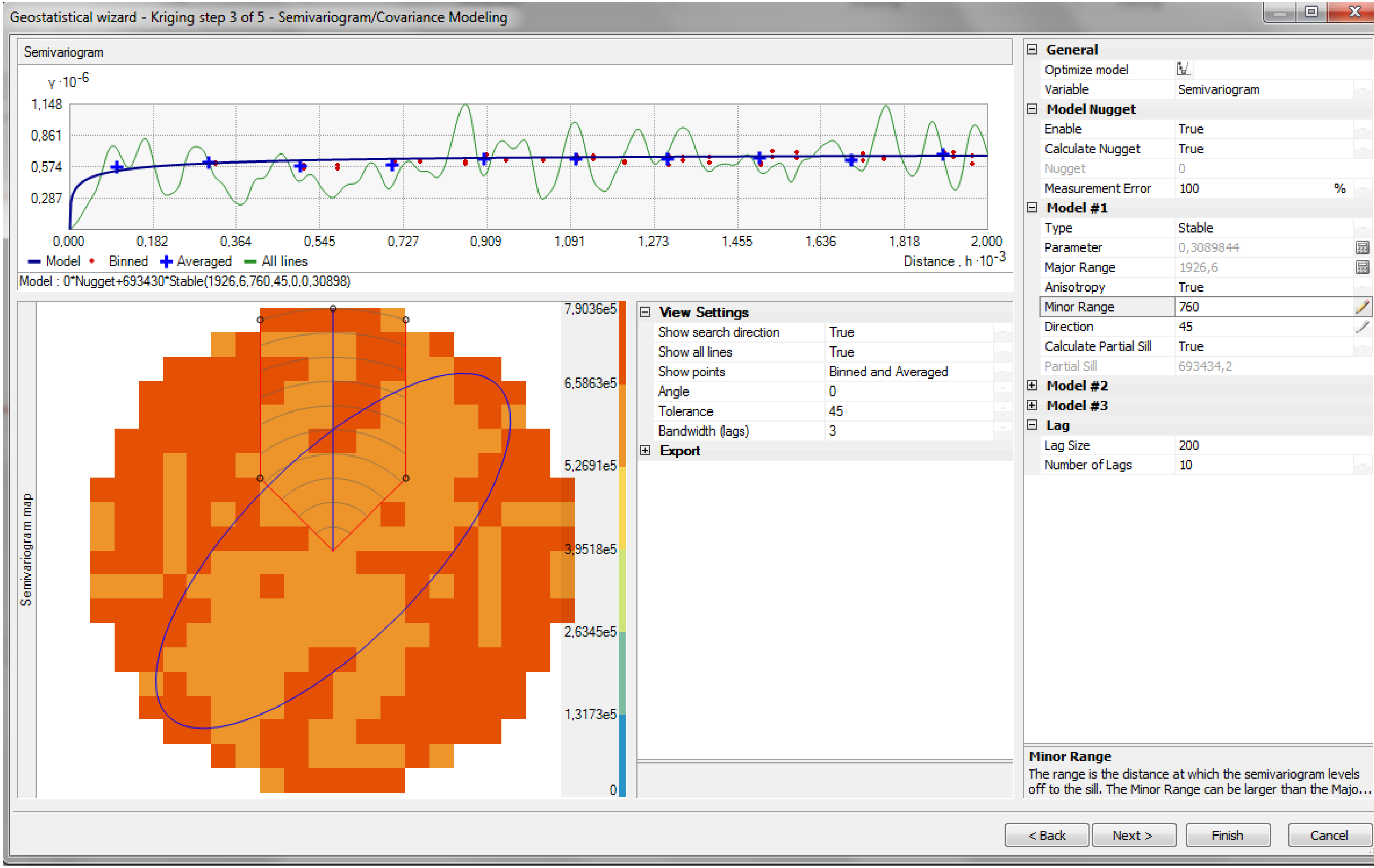


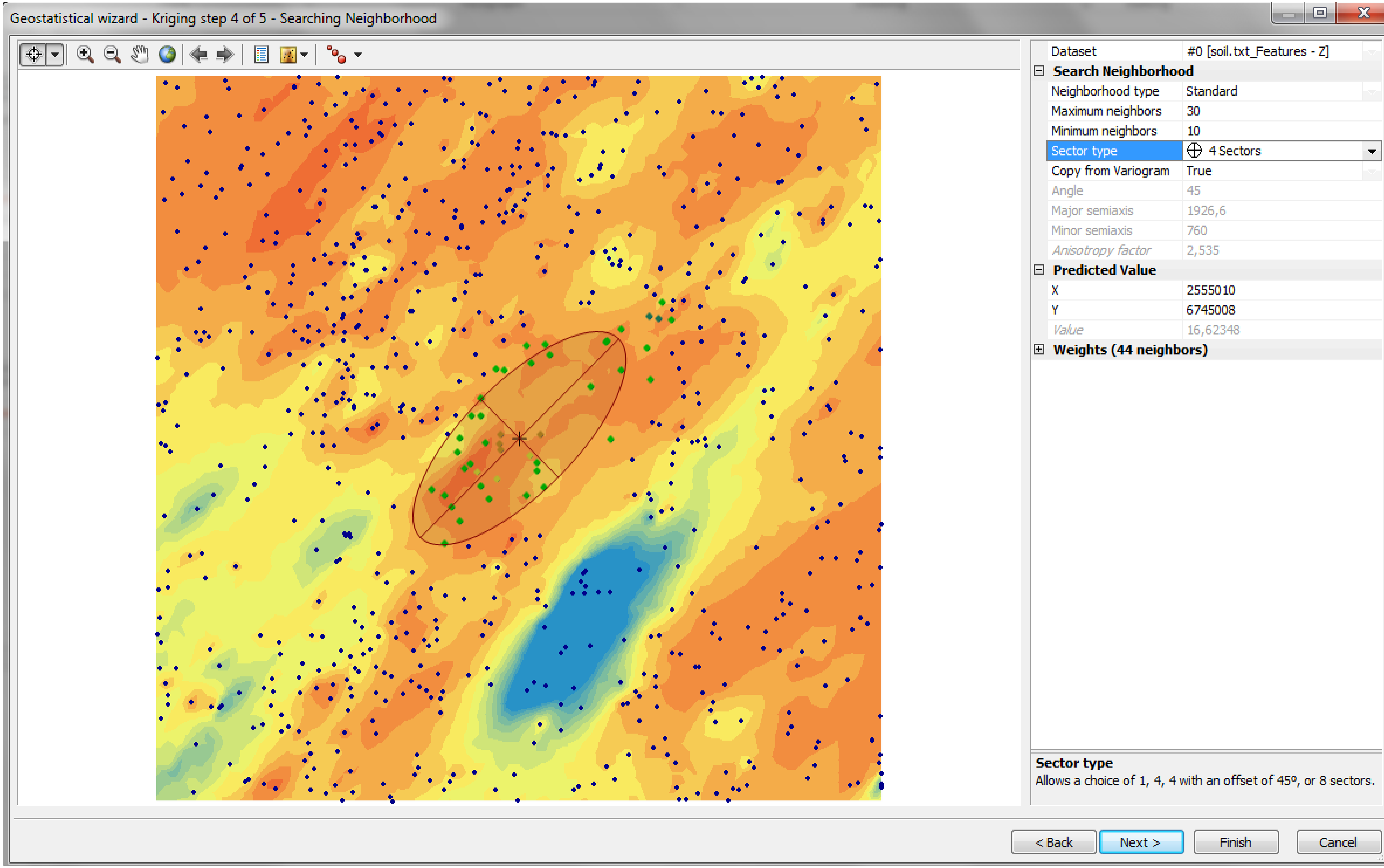
Geostatistical wizard - Kriging step 3 of 5 - Semivariogram/Covariance Modeling
Semivariogram
γ ·10^-6
1,148
0,861
0,574
0,287
0,000 0,182 0,364 0,545 0,727 0,909 1,091 1,273 1,455 1,636 1,818 2,000
Distance , h ·10^-3
Model Binned Averaged All lines
Model : 0*Nugget+696270*Stable(1222,4,0,30371)
Semivariogram map
7,9036e5
6,5863e5
5,2691e5
3,9518e5
2,6345e5
1,3173e5
0
View Settings
Show search direction True
Show all lines True
Show points Binned and Averaged
Angle 0
Tolerance 45
Bandwidth (lags) 3
Export
Bandwidth (lags)
Defines the width of the search sector. It is measured in terms of l...
General
Optimize model
Variable Semivariogram
Model Nugget
Enable True
Calculate Nugget True
Nugget 0
Measurement Error 100 %
Model #1
Type Stable
Parameter 0,3037109
Major Range 1222,427
Anisotropy False
Calculate Partial Sill True
Partial Sill 696273,7
Model #2
Model #3
Lag
Lag Size 200
Number of Lags 10
Number of Lags
<more>
A rule of thumb is to multiply the lag size by the number of lags, which should be about half the largest distance amon...
< Back
Next >
Finish
Cancel


Geostatistical wizard - Kriging step 4 of 5 - Searching Neighborhood
Dataset
#0 [soil.txt_Features - Z]
Search Neighborhood
Neighborhood type
Standard
Maximum neighbors
30
Minimum neighbors
10
Sector type
4 Sectors
Copy from Variogram
True
Angle
0
Major semiaxis
1222,427
Minor semiaxis
1222,427
Anisotropy factor
1
Predicted Value
X
2555010
Y
6745008
Value
15,58022
Weights (43 neighbors)
Sector type
Allows a choice of 1, 4, 4 with an offset of 45º, or 8 sectors.
< Back
Next >
Finish
Cancel

Anisotropy

- Spatial variation is not the same in all directions
- The variogram is computed for specific directions
- If the process is anisotropy, then so is the variogram


Geostatistical wizard - Kriging step 3 of 5 - Semivariogram/Covariance Modeling
Semivariogram
γ ·10^-6
1,148
0,861
0,574
0,287
0,000
0,182
0,364
0,545
0,727
0,909
1,091
1,273
1,455
1,636
1,818
2,000
Model
Binned
Averaged
All lines
Distance , h ·10^-3
Model : 0*Nugget+693430*Stable(1926,6,760,45,0,0,30898)
Semivariogram map
7,9036e5
6,5863e5
5,2691e5
3,9518e5
2,6345e5
1,3173e5
0
View Settings
Show search direction True
Show all lines True
Show points Binned and Averaged
Angle 0
Tolerance 45
Bandwidth (lags) 3
Export
General
Optimize model
Variable Semivariogram
Model Nugget
Enable True
Calculate Nugget True
Nugget 0
Measurement Error 100 %
Model #1
Type Stable
Parameter 0,3089844
Major Range 1926,6
Anisotropy True
Minor Range 760
Direction 45
Calculate Partial Sill True
Partial Sill 693434,2
Model #2
Model #3
Lag
Lag Size 200
Number of Lags 10
Minor Range
The range is the distance at which the semivariogram levels off to the sill. The Minor Range can be larger than the Majo...
< Back
Next >
Finish
Cancel


Geostatistical wizard - Kriging step 4 of 5 - Searching Neighborhood
Dataset
#0 [soil.txt_Features - Z]
Search Neighborhood
Neighborhood type
Standard
Maximum neighbors
30
Minimum neighbors
10
Sector type
4 Sectors
Copy from Variogram
True
Angle
45
Major semiaxis
1926,6
Minor semiaxis
760
Anisotropy factor
2,535
Predicted Value
X
2555010
Y
6745008
Value
16,62348
Weights (44 neighbors)
Sector type
Allows a choice of 1, 4, 4 with an offset of 45º, or 8 sectors.
< Back
Next >
Finish
Cancel

Fitting variogram models

- Why do we need a variogram model?
  - We need a variogram value for some distance or direction for which we do not have a sample variogram value.

Fitting variogram models

- Fitting variogram models can be difficult
  - The accuracy of the observed semivariance is not constant
  - The variation may be anisotropic
  - The experimental variogram may contain much point-to-point fluctuation
  - Most models are non-linear in one or more parameters
- Both visual inspection and statistical fitting are recommended

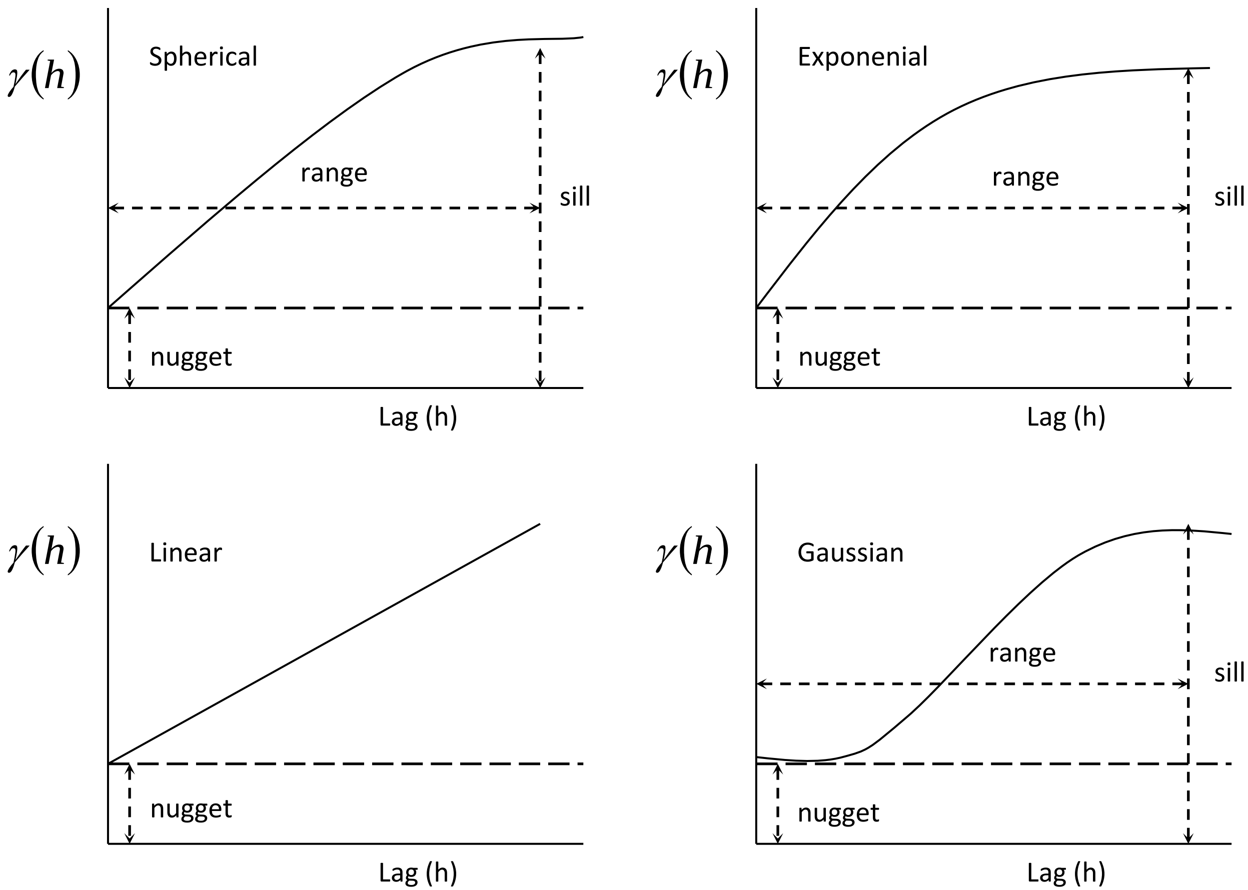
Fitting variogram models

- Fitting variogram models may be poor
  - Chose unsuitable model in the first place
  - Give poor estimates of the parameters at the start of the iteration
  - A lot of scatter in the expertimental variogram
  - The computer program was faulty

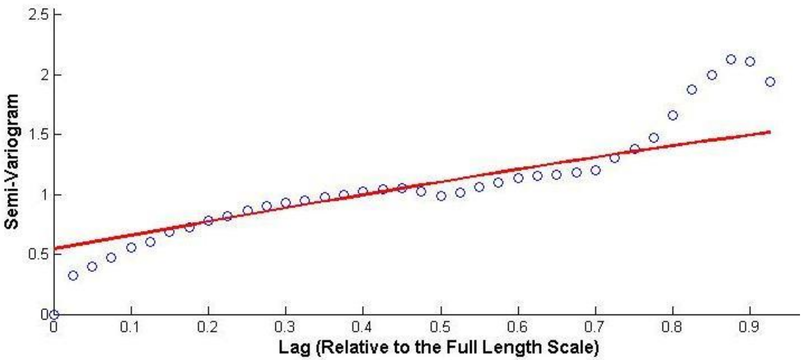
Description

- Lag – The distance between sampling pairs.
- Range – The point where the semivariogram reaches the sill on the lag h axis. Sample points that are farther apart than range are not spatially autocorrelated.
- Nugget – The point where semivariance intercepts the ordinate.
- Sill – The value where the semivariogram first flattens off, the maximum level of semivariance. The points above the sill indicate negative spatial correlation and vice versa.

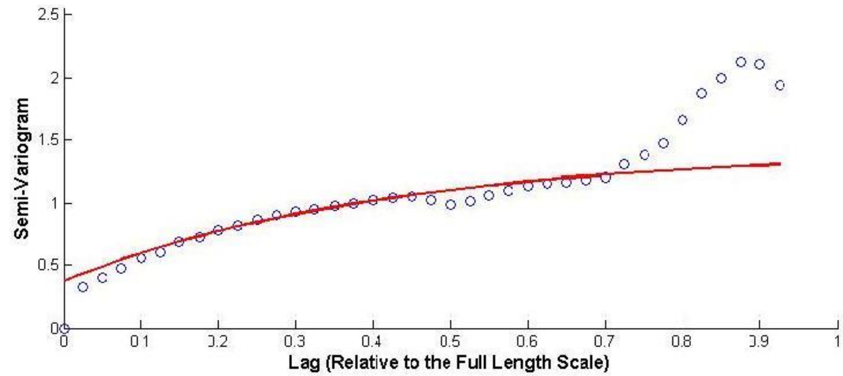
Fitting variogram models



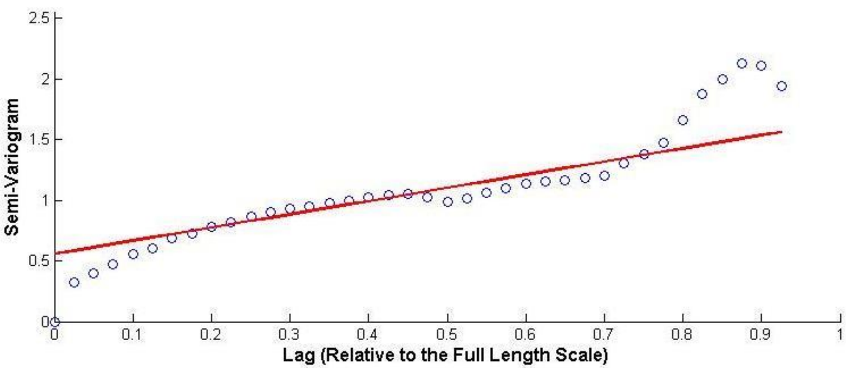
Fitting variogram models



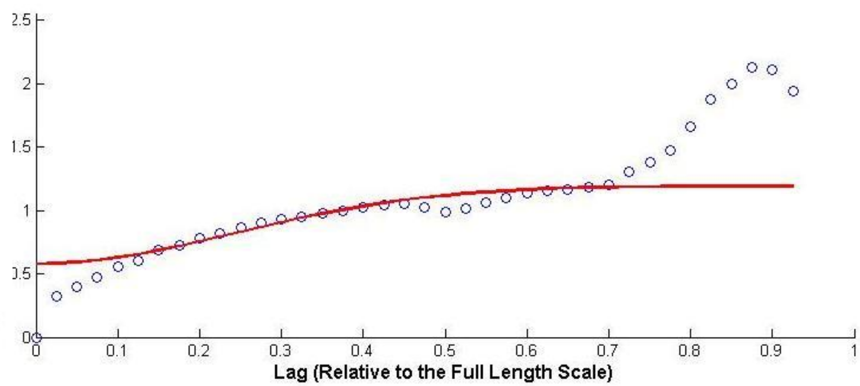
Spherical



Exponential



Linear



Gaussian

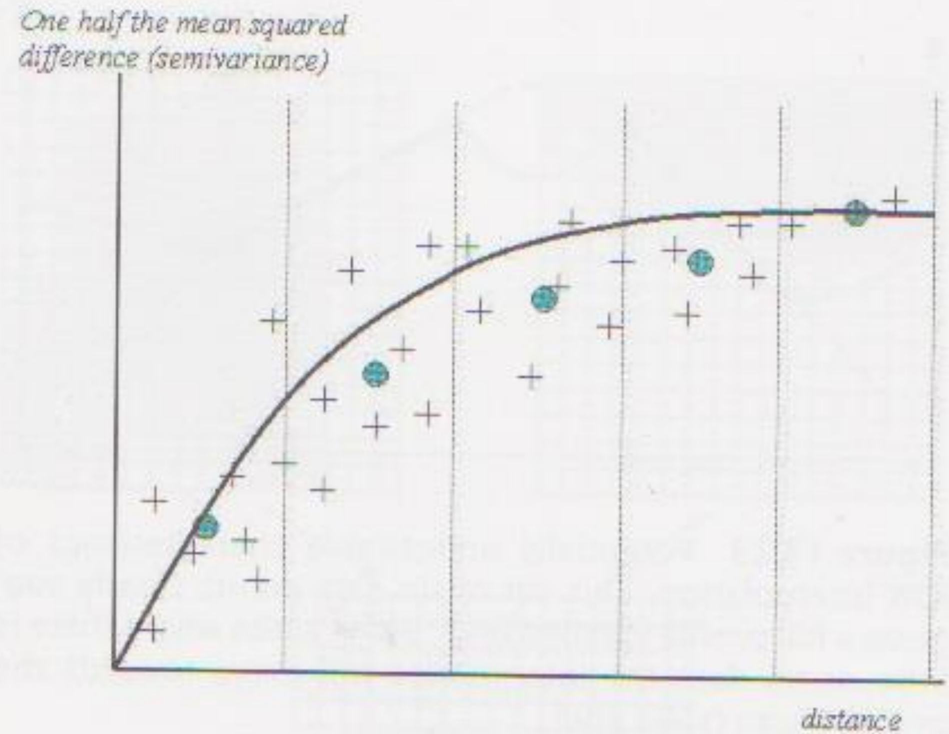


Figure 13.24 A semivariogram. Each cross represents a pair of points. The solid circles are obtained by averaging within the ranges or *bins* of the distance axis. The solid line is the best fit to these five points, using one of a small number of standard mathematical functions

Source: Longley, P.A., Goodchild, M.F., Maguire, D.J. And Rhind, D.W., 2001, Geographic Information Systems and Science

Spherical model

$$\gamma(h) = \begin{cases} c_0 + c_1 \left\{ \dfrac{3h}{2a} - \dfrac{1}{2} \left(\dfrac{h}{a} \right)^3 \right\} & \textit{for} \;\; 0 < h < a \\ c_0 + c_1 & \textit{for} \;\; h \geq a \end{cases}$$

where $a = \text{range}$

$c_0 = \text{nugget variance}$

$c_0 + c_1 = \text{sill}$

Exponential model

$$\gamma(h) = c_0 + c_1\left\{1 - \exp\left(-\frac{h}{a}\right)\right\}$$
$$\gamma(0) = 0$$

Gaussian model

$$\gamma(h) = c_0 + c_1\left\{1 - \exp\left(-\frac{h^2}{a^2}\right)\right\}$$
$$\gamma(0) = 0$$

Linear model

$$\gamma(h) = c_0 + bh$$
$$\gamma(0) = 0$$

where b = the slope of the line

Nested structure

- One variogram model can be created by several variogram models

$$\gamma(h) = \sum_{i=1}^{n} |\omega_i| \gamma_i(h)$$

$$and$$

$$\gamma_t(h) = \gamma_1(h) + \gamma_2(h) + \ldots$$

Nested Structure

- Example: the nested spherical or double spherical

$$
\gamma(h) = \begin{cases} c_1 \left\{ \dfrac{3h}{2a_1} - \dfrac{1}{2}\left(\dfrac{h}{a_1}\right)^3 \right\} + c_2 \left\{ \dfrac{3h}{2a_2} - \dfrac{1}{2}\left(\dfrac{h}{a_2}\right)^3 \right\} & for\ 0 < h \leq a_1 \\ c_1 + c_2 \left\{ \dfrac{3h}{2a_2} - \dfrac{1}{2}\left(\dfrac{h}{a_2}\right)^3 \right\} & for\ a_1 < h \leq a_2 \\ c_1 + c_2 & for\ h > a_2 \end{cases}
$$

Ordinary kriging

- In ordinary kriging, a probability model is used in which the bias and error variance can be computed and select weights for the neighbour sample locations that the everage error for the model is 0 and the error variance is minimized.
- The procedure of ordinary kriging is similar to weighted moving average except the weights are derived from geostatistical analysis.

Ordinary kriging

- The estimationi by ordinary kringing can be expressed by:

$$z(x_0) = \sum_{i=1}^{n} \lambda_i \cdot z(x_i) \text{ where } \sum_{i=1}^{n} \lambda_i = 1$$

The minimum variance of $z(x_0)$ is

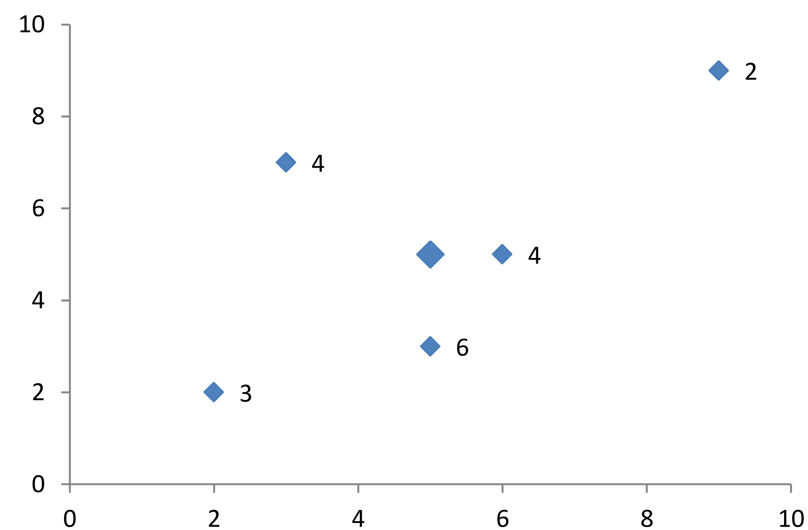
$$\sigma^2 = \sum_{i=1}^{n} \lambda_i \gamma(x_i, x_0) + \phi$$

And it is obtained when

$$\sum_{i=1}^{n} \lambda_i \gamma\left(x_i, x_j\right) + \phi = \gamma(x_j, x_0) \qquad \textit{for all } j$$

Example:

Site	x	y	z	D to (5,5)
1	2	2	3	4.2426
2	3	7	4	2.8284
3	9	9	2	5.6569
4	6	5	4	1.0000
5	5	3	6	2.0000



We would like to estimate the variable value at (5,5)

Example:Ordinary kriging

- Computing kriging weights for the unsampled point x = 5, y = 5. Let the spatial variation of the attribute sampled at the five points be modelled by a spherical variogram with parameters c_0=2.5, c_1=7.5 and range a = 10. The data at the five sampled points are:

x	y	z
2	2	3
3	7	4
9	9	2
6	5	4
5	3	6

- We need to solve

$$A^{-1} \cdot b = \begin{bmatrix} \lambda \\ \phi \end{bmatrix}$$

Where A is the matrix of semivarainces between of pairs of data points, b is the vector of semivariances between each data point and the point to be predicted, λ is the vector of weights and ϕ is a lagrangian for solving the equations.

Example from Principles of Geographical Information Systems by Burrough and McDonnell, 1998, Oxford University Press

Example:Ordinary kriging

- Value at (5,5) = weights * z

  = 4.3985

- With estimation variance = (weights*b)+ Φ

  = 4.2177 + (-0.1544)

  = 4.0628

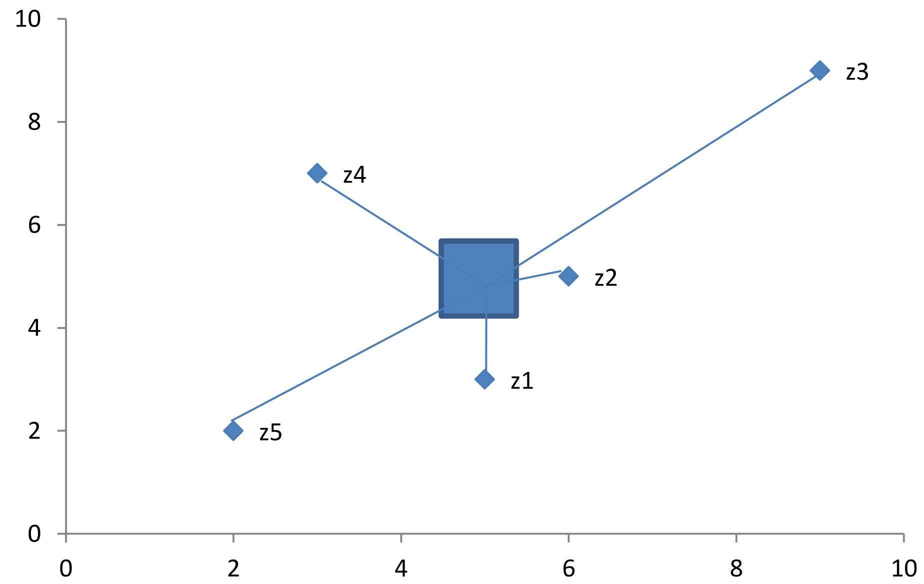
Note: The estimation error variance is also known as kriging variance.

Comparison of the results

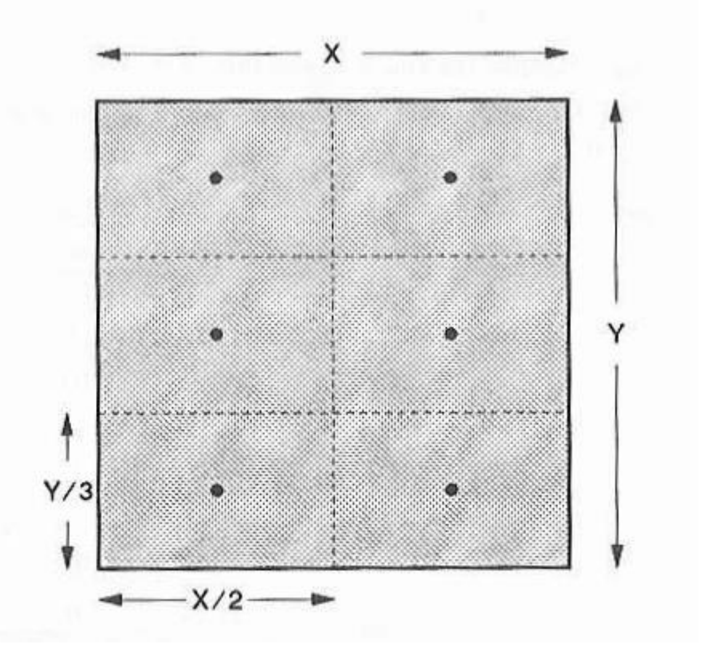
Method	Estimate value z(5,5)
IDW	4.1814
Ordinary Kriging (Kriging variance)	4.3985 (4.0628)

Block kriging

- The modification of kriging equations to estimate an average value z(B) of the variable z over a block of area B.



Block kriging



Example showing a regular 2x3 grid of point locations within a block. Each discretizing point accounts for the same area.

Example from An introduction to applied Geostatistics by Edward H. Isaaks and R.Mohan Srivastava

Block kriging

- The average value of z(B) over the block B is given by

$$z(B) = \int_B \frac{z(x)dx}{area B}$$

is estimated by

$$\hat{z}(B) = \sum_{i=1}^{n} \lambda_i \cdot z(x_i)$$

with $\sum_{i=1}^{n} \lambda_i = 1$

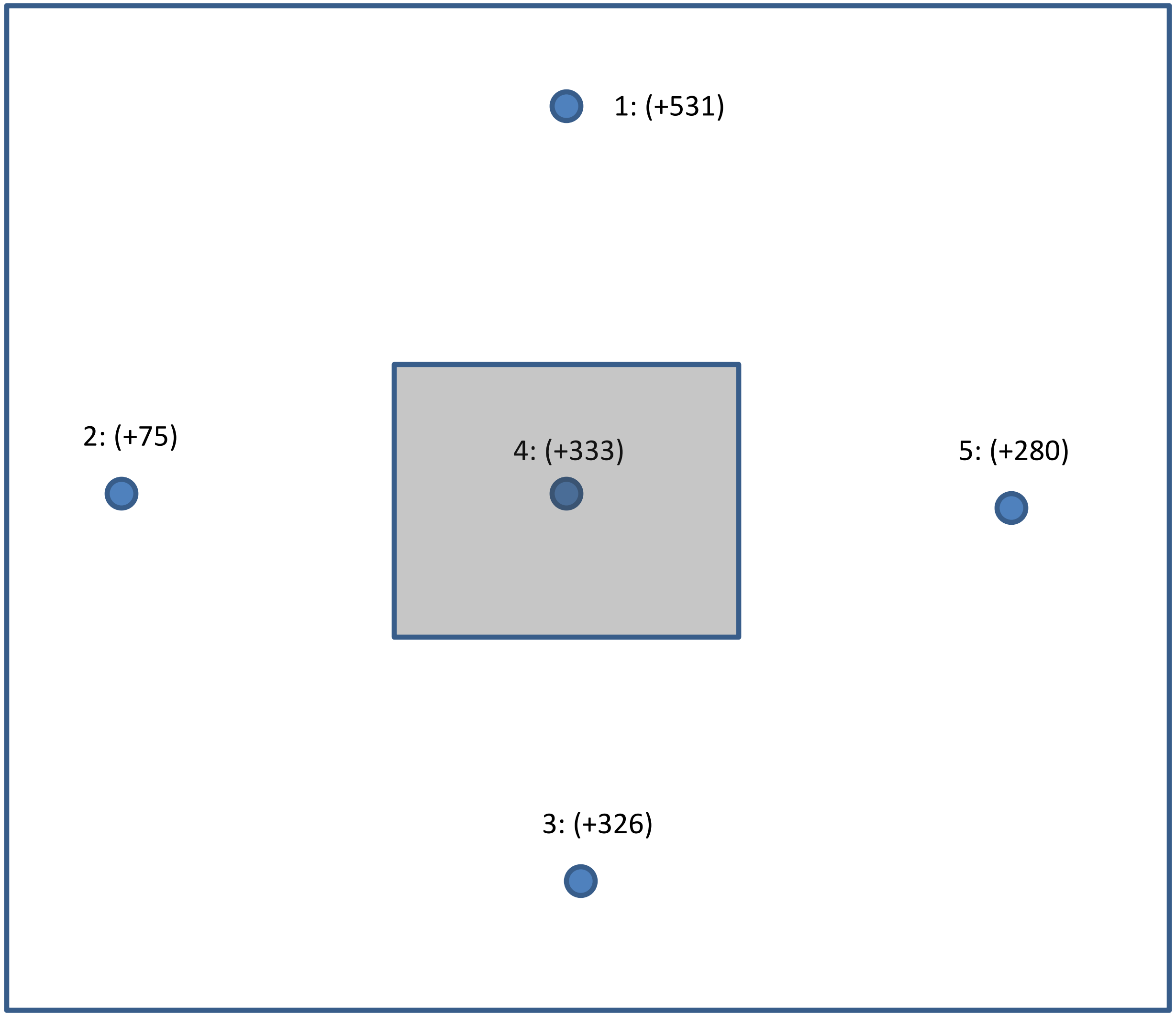
Block kriging

- The minimum variance is

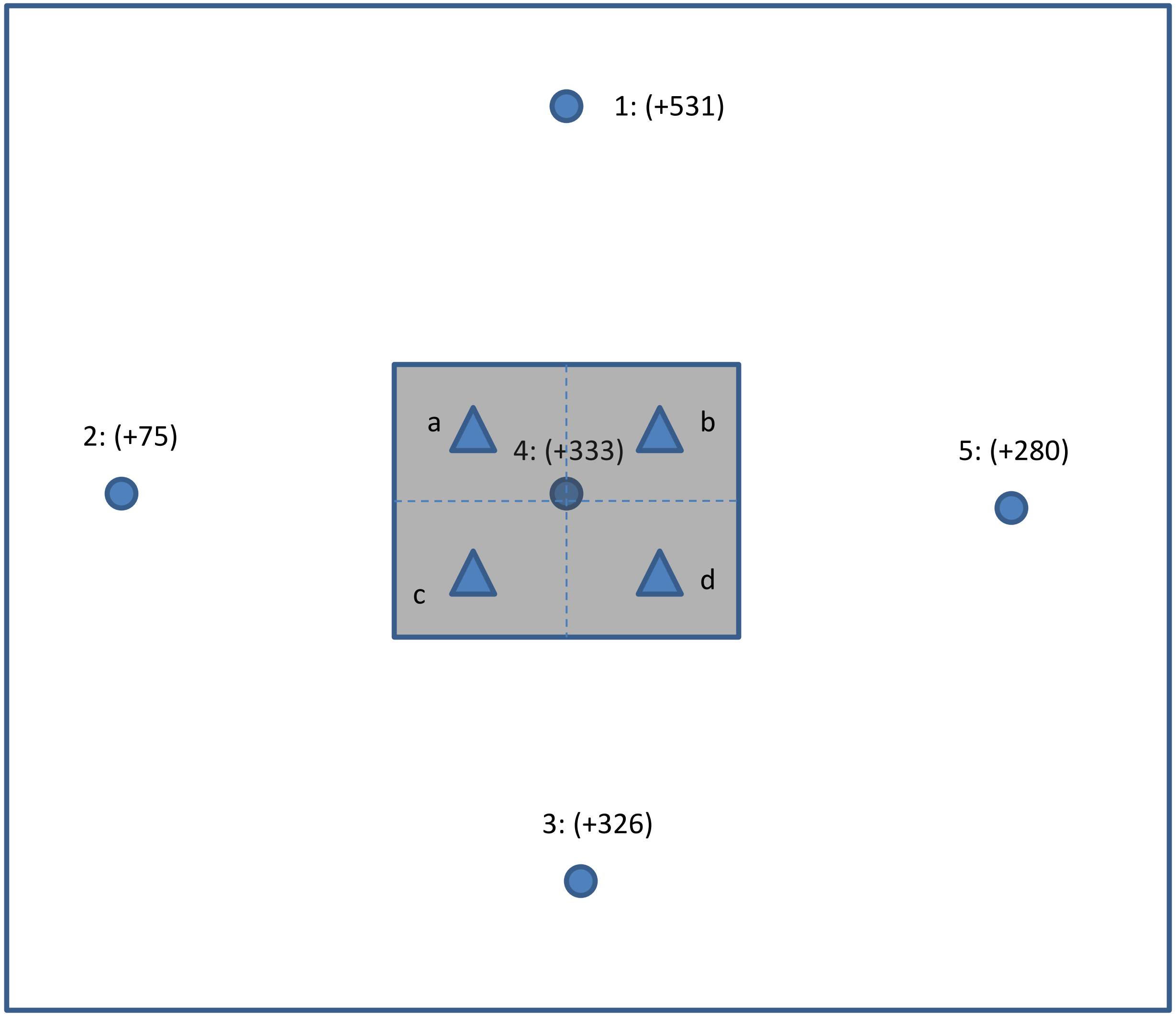
$$\sigma^2(B) = \sum_{i=1}^{n} \lambda_i \bar{\gamma}(x_i, B) + \phi - \bar{\gamma}(B, B)$$

and is obtained with

$$\sum_{i=1}^{n} \lambda_i \gamma(x_i, x_j) + \phi = \bar{\gamma}(x_j, B) \qquad for\, all\ j$$



Example from An introduction to applied Geostatistics by Edward H. Isaaks and R.Mohan Srivastava


1: (+531)
2: (+75)
a
b
4: (+333)
5: (+280)
c
d
3: (+326)

Example: Block kriging

Point	Estimate	Kriging weights for samples				
		1	2	3	4	5
a	336	0.17	0.11	0.09	0.60	0.03
b	361	0.22	0.03	0.05	0.56	0.14
c	313	0.07	0.12	0.17	0.61	0.03
d	339	0.11	0.03	0.12	0.62	0.12
Average	337	0.14	0.07	0.11	0.60	0.08

Simple kriging

- It is similar to ordinary kriging except that the weights sum equation (=1) is not added.
- The mean is a known constant.
- It uses the average of the entire data set. (ordinary kriging uses local average : the average of the points in the subset for a particular interpolation point)

Cokriging

- It is an extension of ordinary kriging where two or more variables are interdependent.
- How:
  - U and V are spatial correlated
  - Variable U can be used to predict variable V that is information about spatial variation of U can help to map V.
  - Why:
    - V data may be expensive to measure or collect or have some limitations in data collection process so the data may be infrequent.
    - U data, on the other hand, may be cheap to measure and possible to collect more observations.

Indicator kriging

- Binary value
- From a continuous variable z(x), an indicator can be created by indicating it 1 for z(x) is less than or equal to a cut-off value, z_c, and 0 otherwise

$$\omega(x) = \begin{cases} 1 & if\ z(x) \leq z_c \\ 0 & otherwise \end{cases}$$

Kriging: Step by step

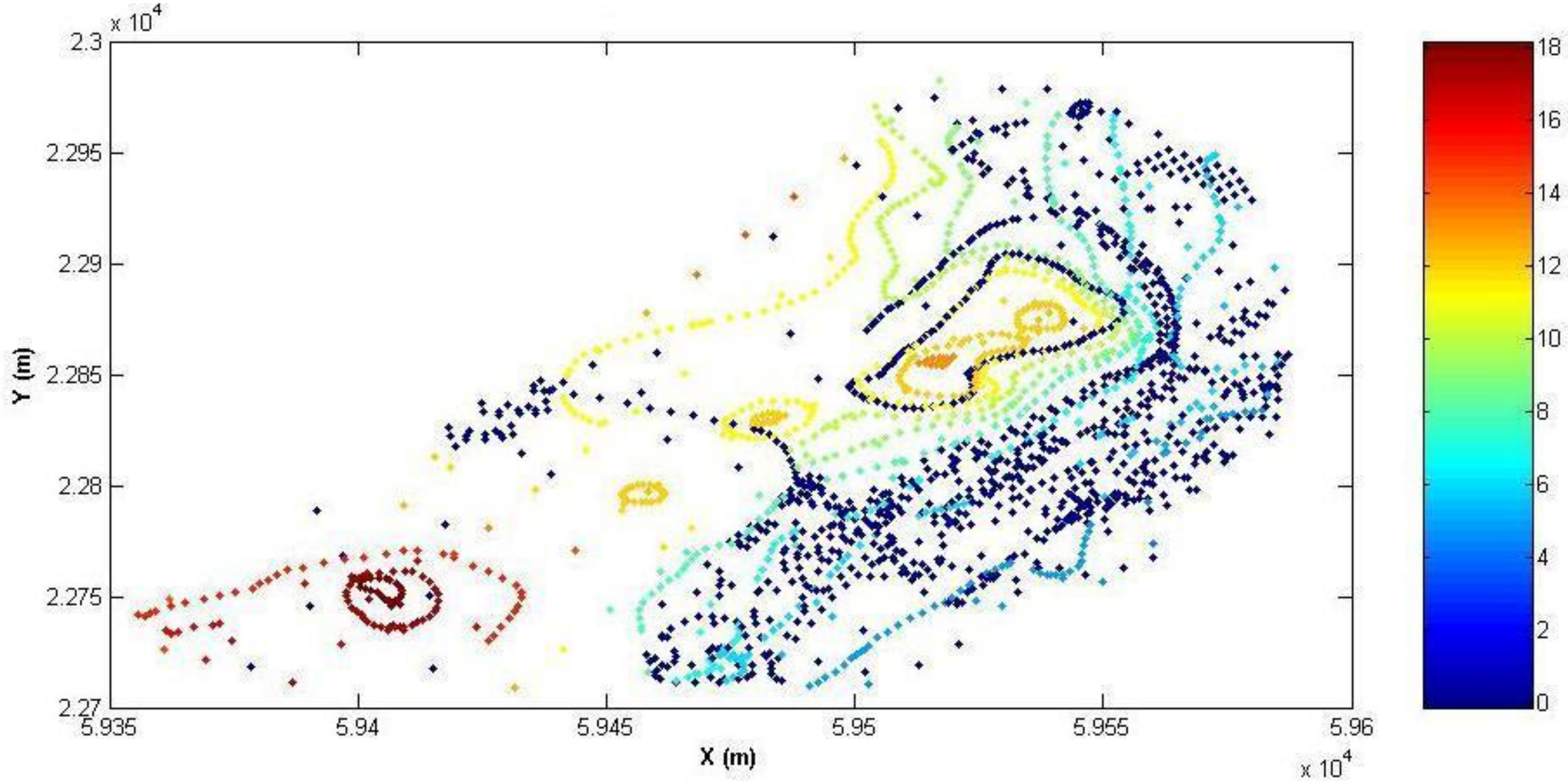
- Studying the gathered data: data analysis
- Fitting variogram models: experimental variogram and theoretical variogram models
- Estimating values at those locations which have not been sampled (kriging) e.g. ordinary kriging, simple kriging, indicator kriging and so on
- Examining standard error which may be used to quantify confidence levels
- Kriging interpolation

Example

- Elevation data set in Rastila
- 2000 laser scanning points

Minimum	0
Maximum	18
Mean	4.7247
Median	0
Skewness	0.64651
Kurtosis	2.0344
Standard deviation	5.543
Variance	30.725

Example


x 10^4
2.3
2.295
2.29
2.285
2.28
2.275
2.27
Y (m)
5.935
5.94
5.945
5.95
5.955
5.96
X (m)
x 10^4
18
16
14
12
10
8
6
4
2
0

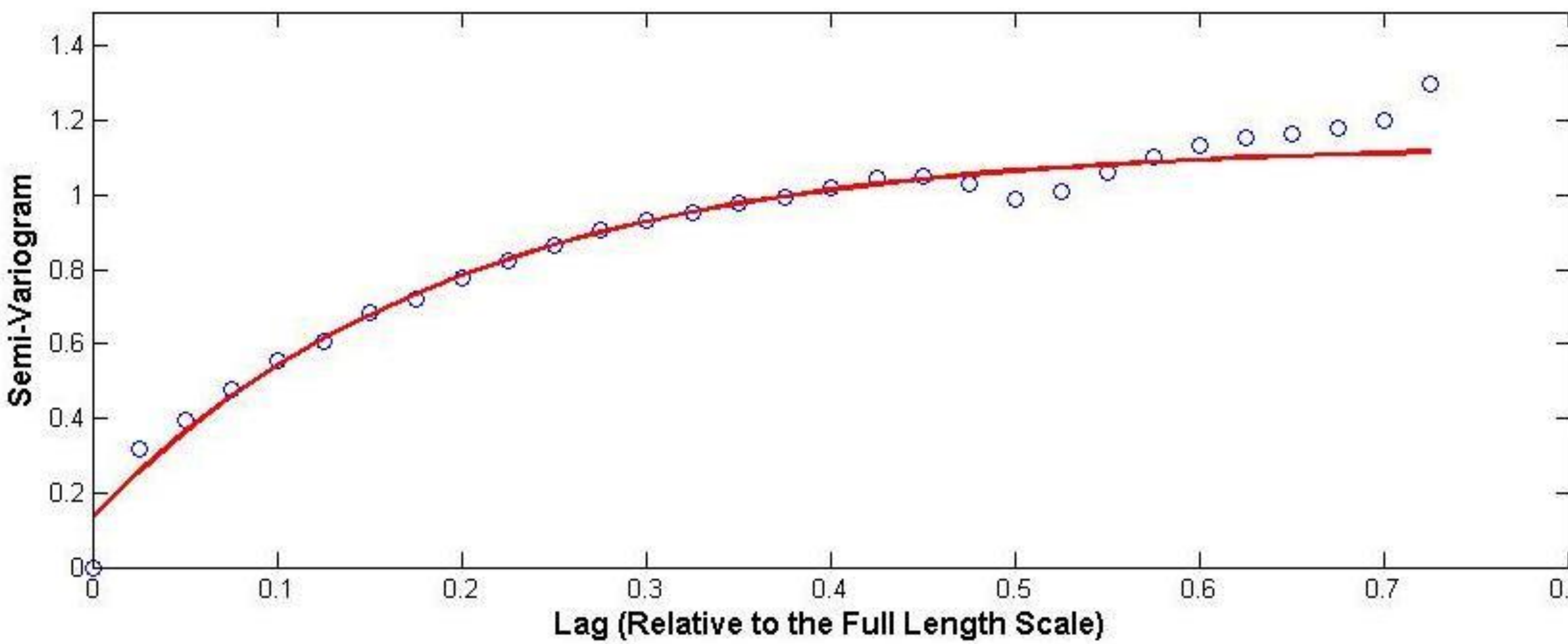
Example

Variogram model	range	nugget	sill	length
Exponential	0.5	0.10847	1.1232	0.18064
Linear	0.5	0.50754	1.806	-
Gaussian	0.5	0.4546	1.0447	0.22811
Spherical	0.75	0.35265	1.0666	0.48576
Exponential	0.75	0.13161	1.1388	0.19144
Linear	0.75	0.58671	1.5711	-
Gaussian	0.75	0.48089	1.0728	0.24717
Spherical	0.95	0.54285	1.9597	1.8513
Exponential	0.95	0.36872	1.4082	0.40794
Linear	0.95	0.55676	1.6424	-
Gaussian	0.95	0.57831	1.1922	0.34484

Example

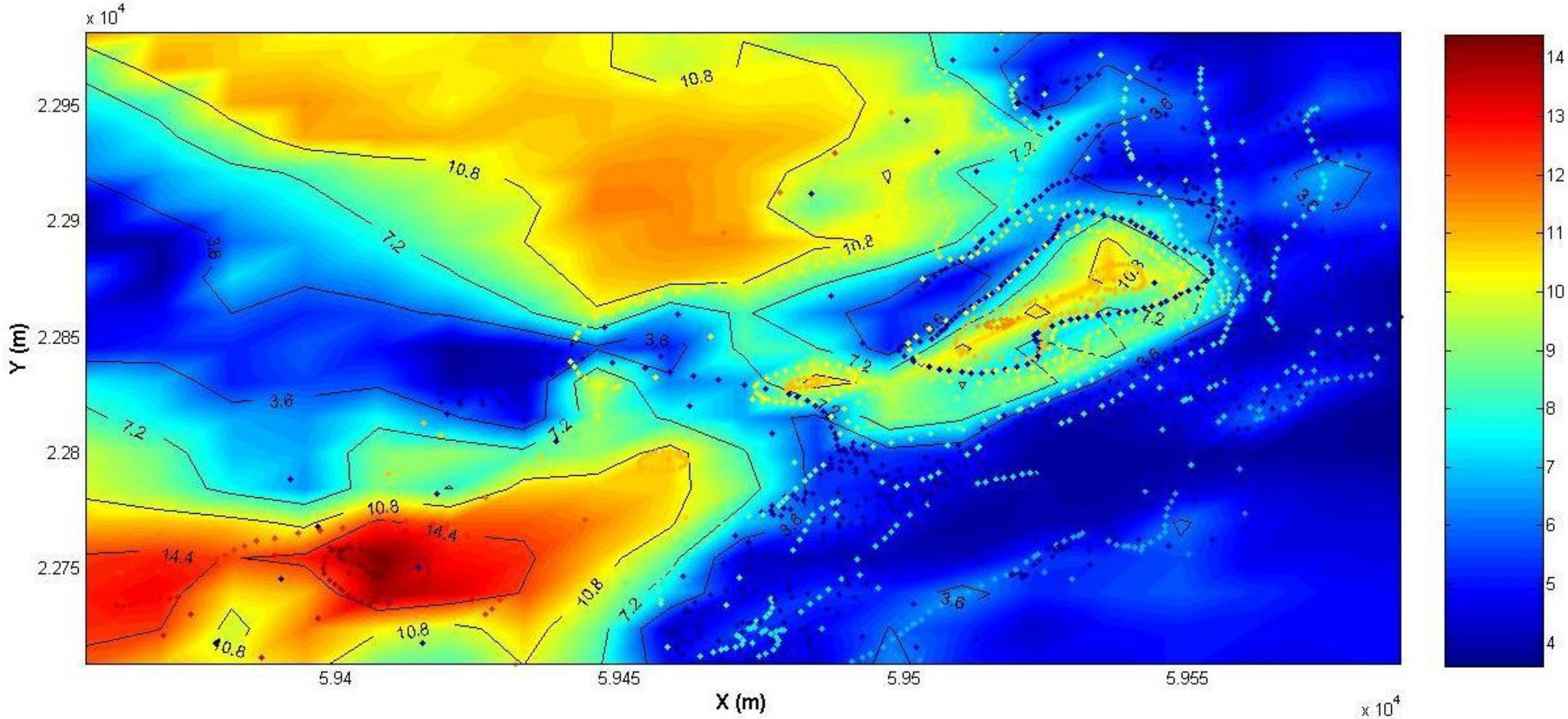
Variogram model	RMSE
Exponential (0.5)	3.281
Linear (0.5)	3.367
Gaussian (0.5)	3.431
Spherical (0.75)	3.307
Exponential (0.75)	3.272
Linear (0.75)	3.397
Gaussian (0.75)	3.439
Spherical (0.95)	3.381
Exponential (0.95)	3.302
Linear (0.95)	3.386
Gaussian (0.95)	3.466

Example



Semi-Variogram
Lag (Relative to the Full Length Scale)
0
0.2
0.4
0.6
0.8
1
1.2
1.4
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8

Example



x 10^4
2.295
2.29
2.285
2.28
2.275
Y (m)
5.94
5.945
5.95
5.955
X (m)
x 10^4
10.8
7.2
3.6
14.4
14
13
12
11
10
9
8
7
6
5
4

References:

- Geographic information Analysis by Sullivan, D. And Unwin, D.
- Geostatistics for Environmental Scientists by Richard Webster and Margaret Oliver
- Principle of Geographical Information Systems, Chapter 5 and 6 by Peter Burrough and Rachael McDonnell
- An Introduction to Applied Geostatistics by Edward Isaaks and Mohan Srivastava
- Quality Aspects in Spatial Data Mining by Alfred Stein, Wenzhong Shi and Wietske Bijker