INTRODUCTION TO GEOSTATISTICS

an ER&P program short course

by

Abani Ranjan Samal (Student, ER&P)

Courses in Geostatistics:

1. MTech (Mineral Exploration, I.S.M., India):
   - Basic Geostatistics and Advanced Geostatistics
2. MSc (Mineral Exploration, Imperial College, London)
   - Geostatistics
   - Short Course in MDE (Margaret Armstrong)

Work Experience:

Mineral Resource Evaluation using Micromine

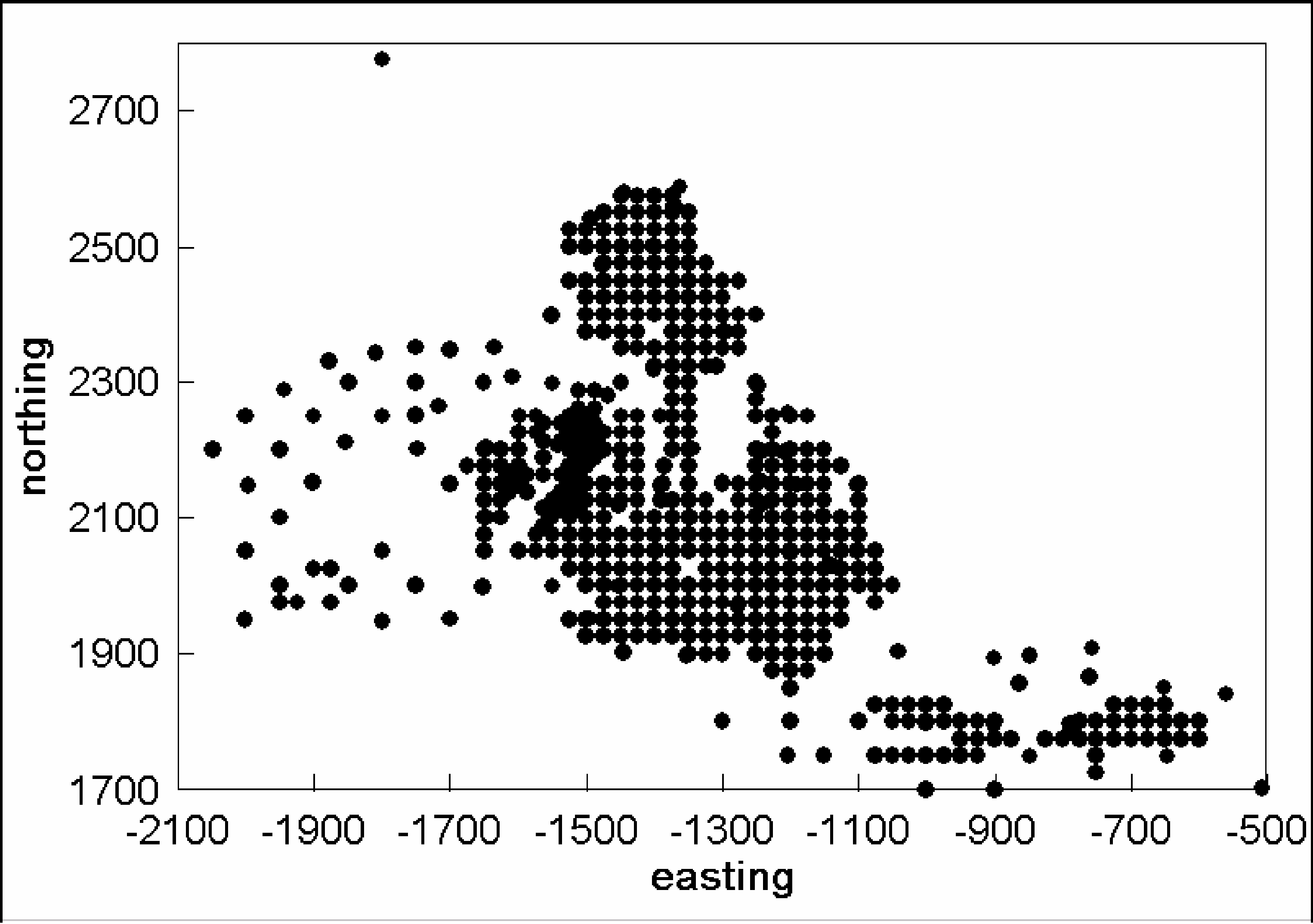
Outline:

- Introduction:
    - Classical statistics Vs Geostatistics,
    - Popular Estimation Methods Vs Kriging
- Semivariogram
    - Introduction,
    - Types and formulae
- Kriging
    - Formula(e),
    - Types,
    - B.L.U.E.
- Application

Geology/Mineral Exploration/Mining,

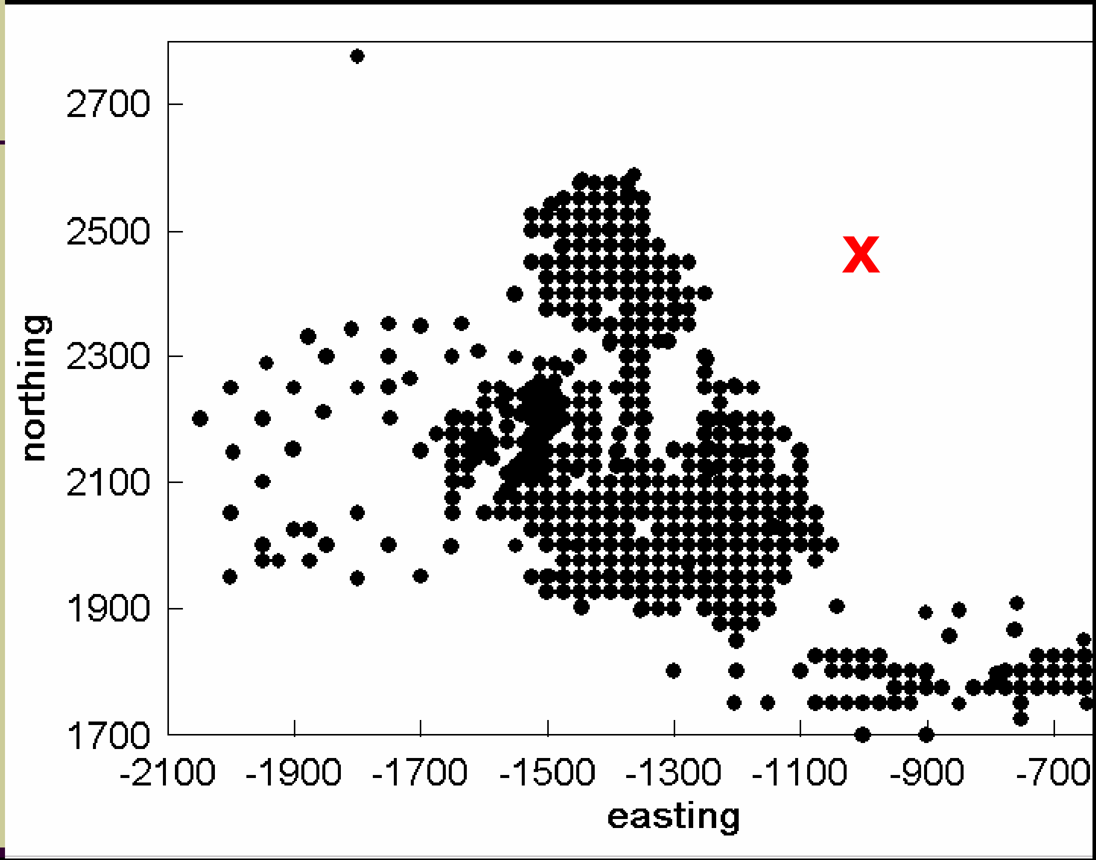
- GIS
- Conclusion

Conventional Statistics



Au		Ag	
Mean	1.475018315	Mean	67.34758242
Median	1.48	Median	64.65
Mode	1.8	Mode	31
Standard Deviation	0.62121581	Standard Deviation	36.33928301
Sample Variance	0.385909082	Sample Variance	1320.54349
Kurtosis	4.369420709	Kurtosis	0.41431362
Skewness	0.769522334	Skewness	0.653851135
Range	5.8	Range	222.67
Minimum	0.05	Minimum	2.33
Maximum	5.85	Maximum	225
Sum	805.36	Sum	36771.78
Count	546	Count	546
Confidence Level(95.0%)	0.052222768	Confidence Level(95.0%)	3.054877102

Conventional Statistics



Au:
1.475±0.621215

Ag:67.34758±36.33928

(from statistics)

Correct?

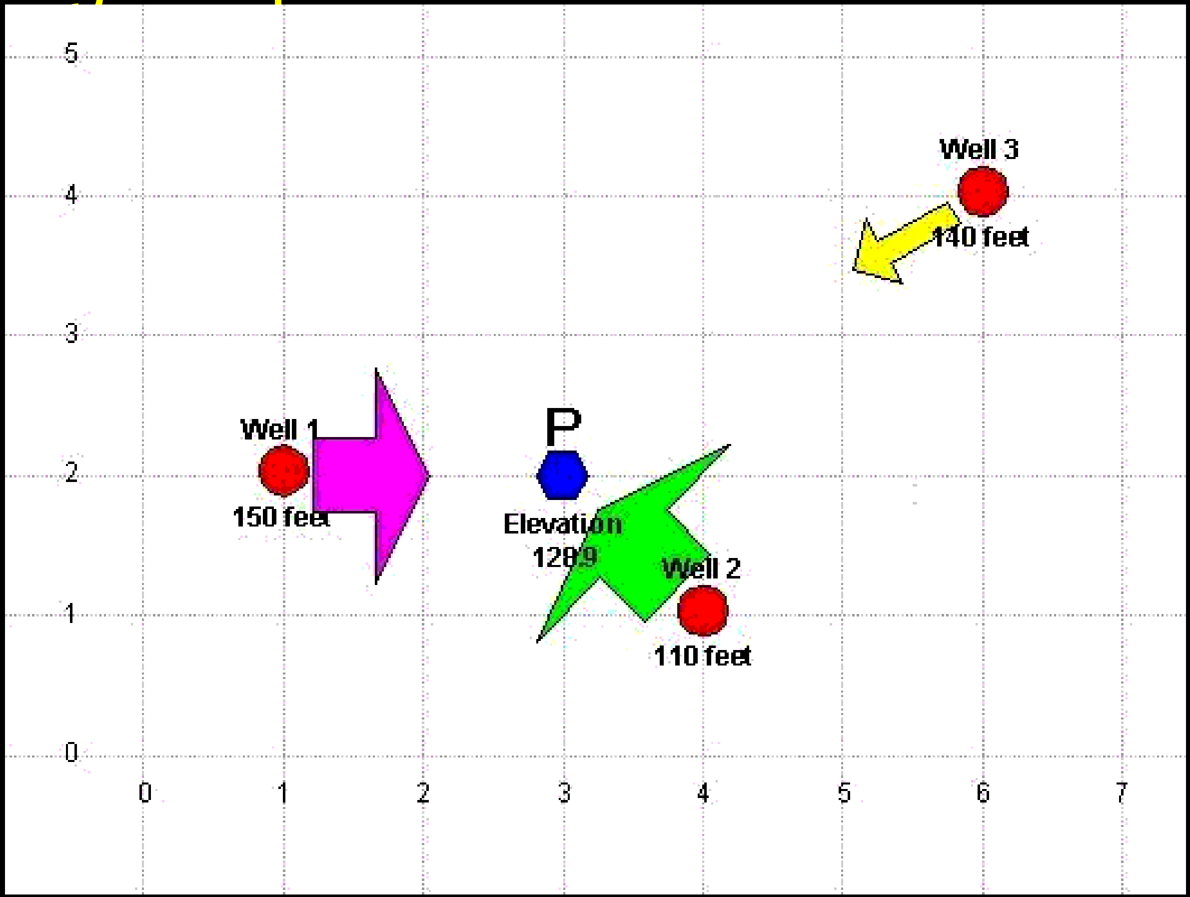
If Correct,
Error?

What is the value of the variable (say Au grade, Ag grade) at "x"?

Whats Next?

INTERPOLATION

- Estimating values at un-sampled location using neighboring samples



Source:http://www.cee.vt.edu/program_areas/environmental/teach/smprimer/kriging/dt-arrow.gif

INTERPOLATION

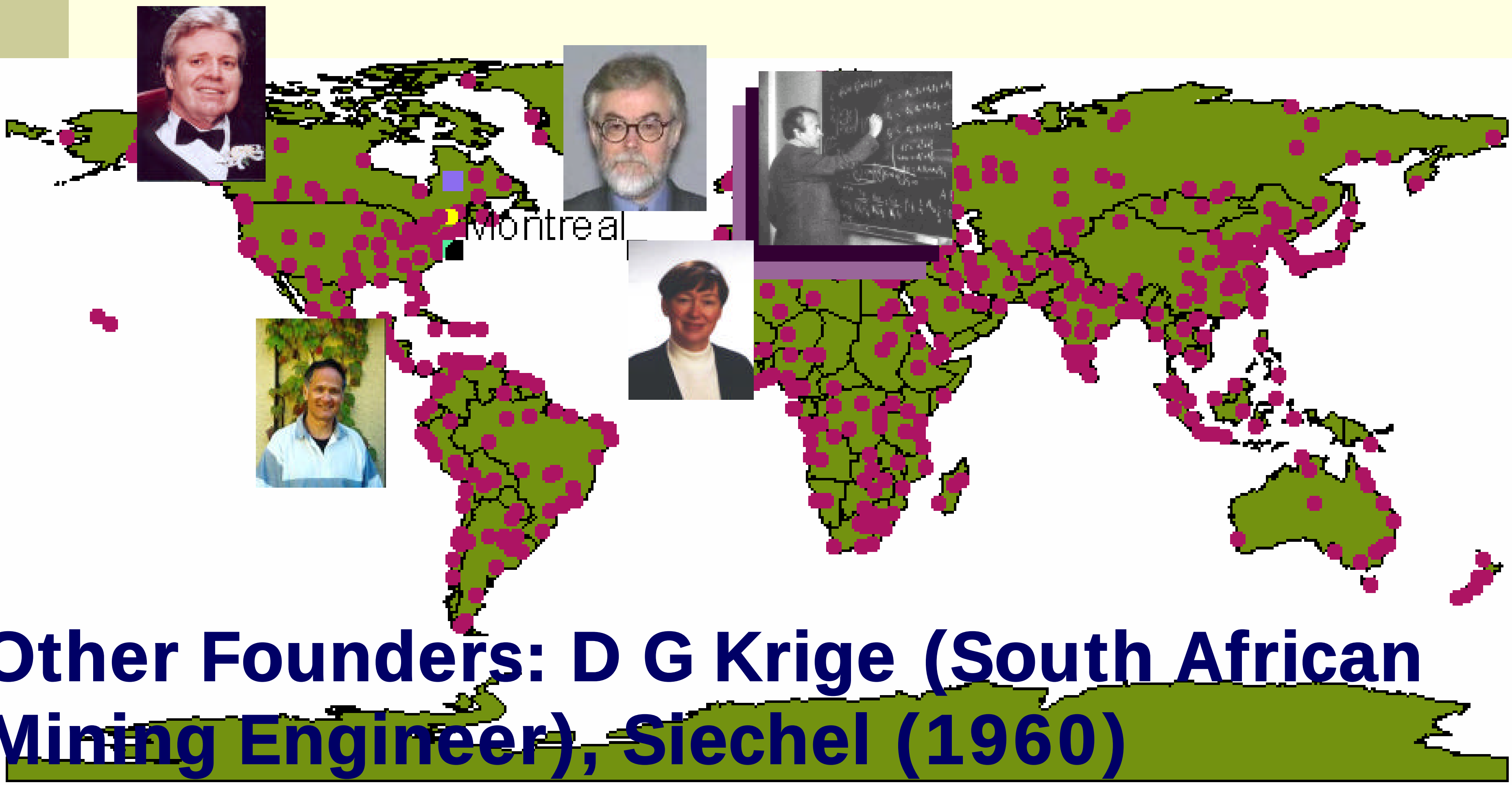
- What Interpolation Techniques we have?
- Deterministic
  - Exact Interpolator
    - Inverse Distance Weighted (IDW); popularly known as IDS and Radial Basis Functions
  - Inexact interpolator
    - Global Polynomial and Local Polynomial
- Geoststistical: Incorporates statistics of data in interpolation

$$\mathbf{Z}^{*}{}_{0} = \sum_{i=1}^{n} w_i z_i$$

HISTORY:

- Started in 1960s
- Began with D G Krige and Siechel
- Developed by G Matheron with the concept of Regionalized Variable (late 1960s)
- Present Status:
  - Highly developed
  - Applied in almost all areas of regionalized variable
  - In almost all parts of the world
  - Lots of software tools available
    - Free and Commercial

HISTORY



Other Founders: D G Krige (South African Mining Engineer), Siechel (1960)

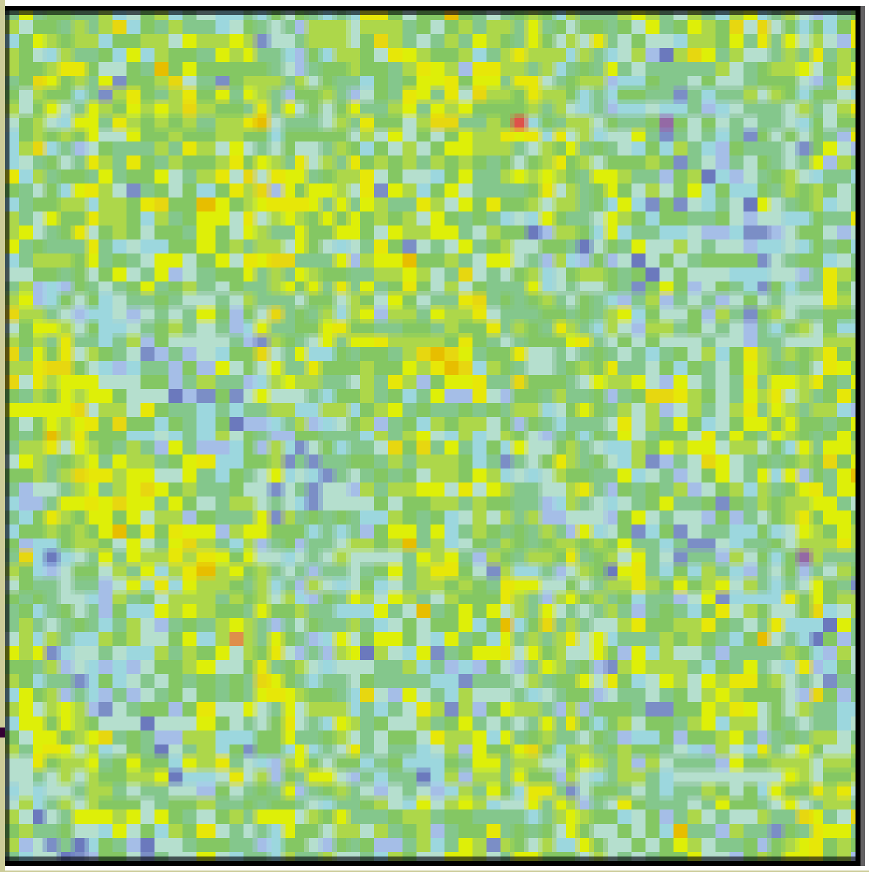
Other Famous Geostatisticians: Snowden (Australia), Goovaerts (Belgium), E H Isaaks (USA), R M Srivastav(USA) etc.

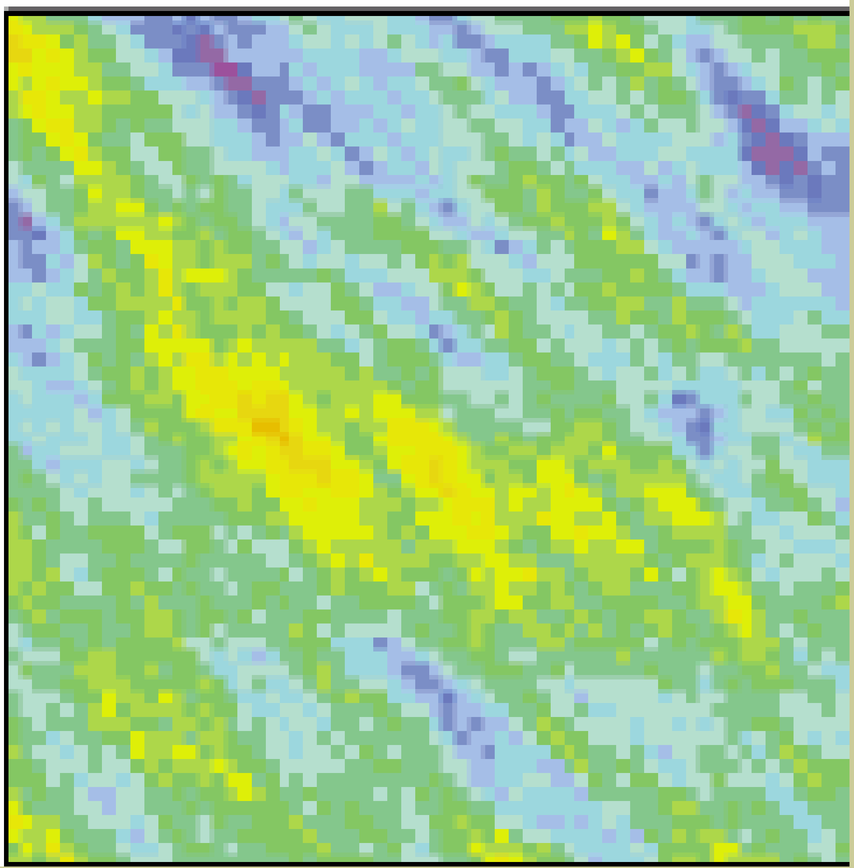
IDS Vs KRIGING

- IDS:An Exact Interpolator
  - Least Variance?
    - Not as compared to Kriging
  - Unbiased?
    - NO
  - Sum of all weight factors (w_i)=1 ?
    - NO
  - Should we discard IDS?
  - **NO**: Because where data is *not good for kriging,* IDS is the best interpolation method.

$$Z^* = \frac{\sum \frac{1}{d_i^2} Z_i}{\sum \frac{1}{d_i^2}}$$

IDS Vs Kriging:

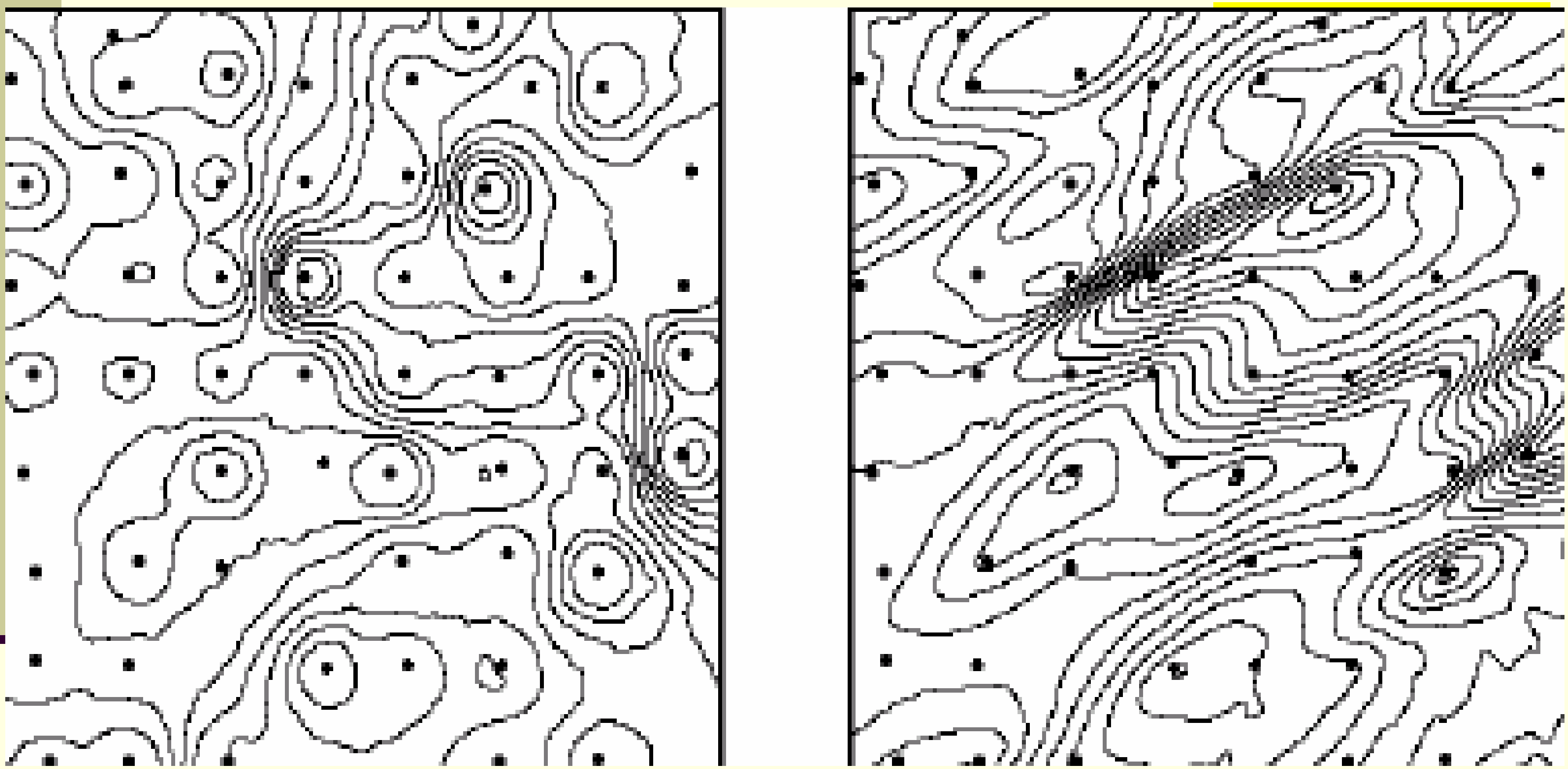




Though these two look different in their spatial distribution, but their mean and variances are identical.

Ref: Petroleum geostatistics for nongeostatisticians, *THE LEADING EDGE MAY 2000*

IDS Vs Kriging:



IDS (left) and kriging (right) using a 3:1 anisotropic variogram model orientedN60E. The neighborhood search ellipse is identical for both.

Why Geostatistics?

- Incorporates the position of sample in space
- Weights the sample size : *support*
    - Theory of **Regionalized Variable** (Matheron)
    - Support: Shape and Volume of the Sample (David)
- Uses statistical parameters of the samples for estimation (*kriging*)
- Two important terms (Estimators ?)
    - **VARIOGRAM:** for structural analysis of data, ststistically models the data in space
    - **KRIGING:** for interpolation
    - *Kriging* uses the information from a variogram to find an optimal set of weights (w_i) that are used in estimating a surface at unsampled locations.
        - *Variogram is not a part of Kriging rather a pre requirement*

Random Variable
Vs
Regionalized Variable

- **Random Variable**
  - **whose values are randomly generated according to some probabilistic mechanism**
    - **Example: the result of casting a die can take one of six equally probable values.**
- **Regionalized Variable**
  - **distributed in space**
  - **Support: Shape and Volume**
    - **Example: mostly natural phenomena. For instance, the heavy metal content of the top layer of a soil**

REGIONALIZED VARIABLE

Two aspects: A structured and an erratic aspect

- **The structured aspect**
  - distribution of the natural phenomenon.
- **The erratic aspect**
  - the local behavior of the natural phenomenon.

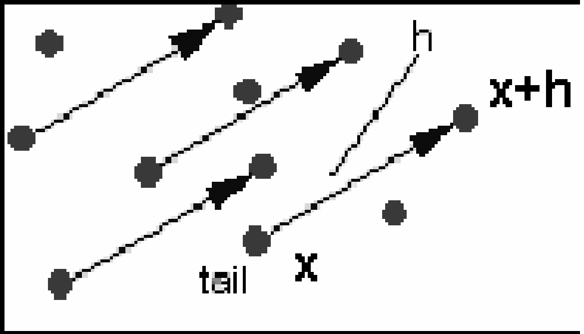
  Distribution of the heavy metal content seems to fluctuate randomly.

Spatial Stationarity and Intrinsic Hypothesis

- Strict Stationarity
  - invariance of joint probability density function under spatial shift ("translation")
    - { $z(x_1)$, ... , $z(x_k)$ } and { $z(x_{1+h})$, ... , $z(x_{k+h})$ }
    - information about process the same no matter where it is obtained
- Moment Stationarity
  - moments invariant under shift
    - constant mean and constant variance
    - covariance only function of spatial separation h
- Intrinsic Hypothesis
  - The Mean, Variance of the increments $z(x_{k+h})$-$z(x_k)$ exists and independent of the point X_k.

Spatial distribution of data

- Location (X,Y,Z)
- Support (Volume and Shape)



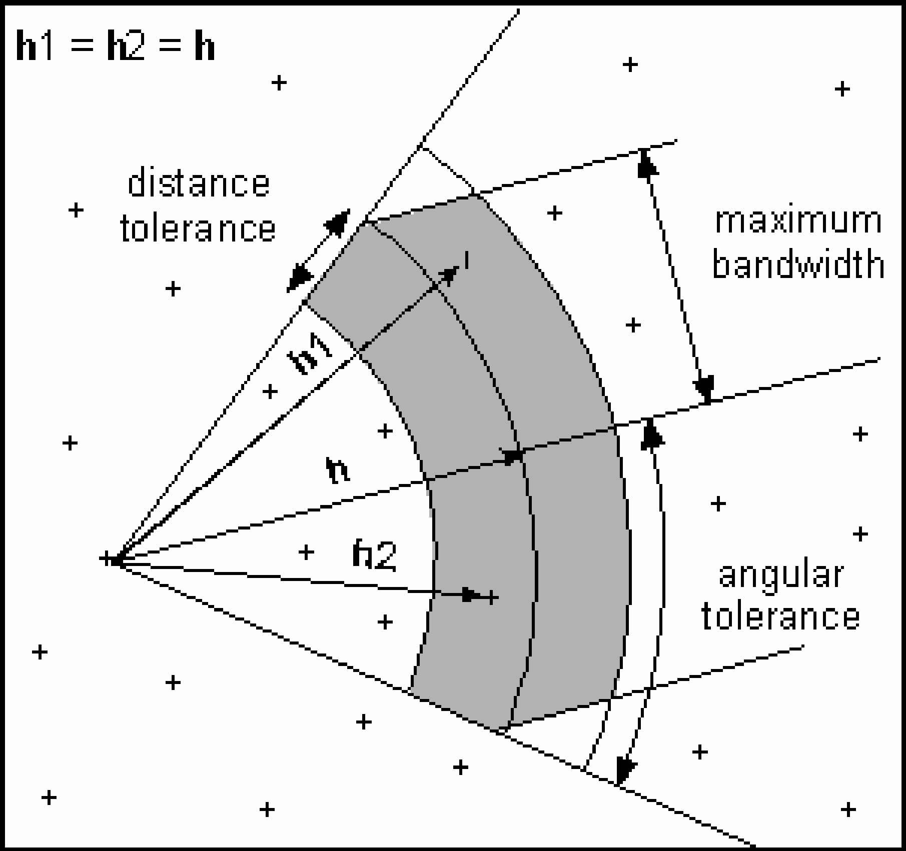
H Scatter plot:

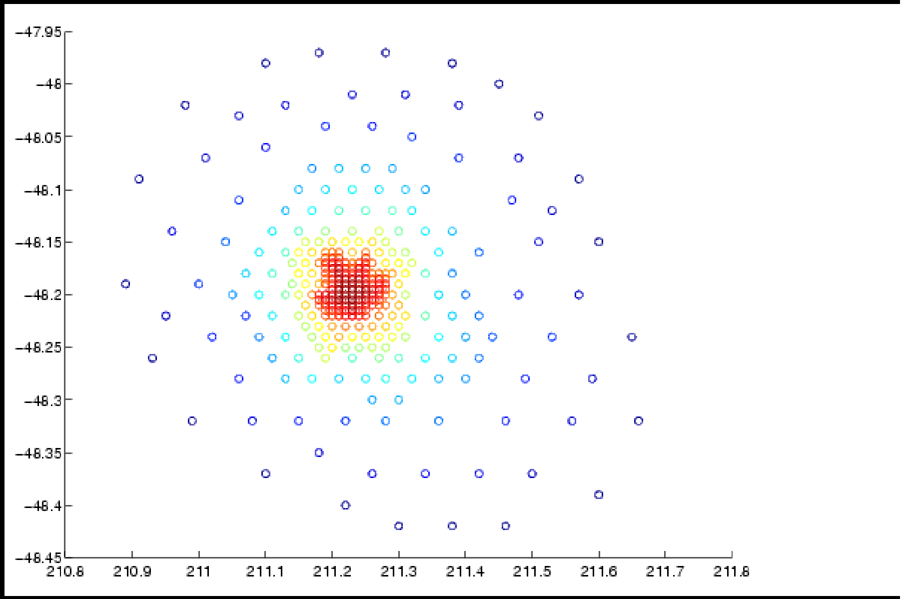
- Describes “Spatial Continuity” (& Spatial dependence)

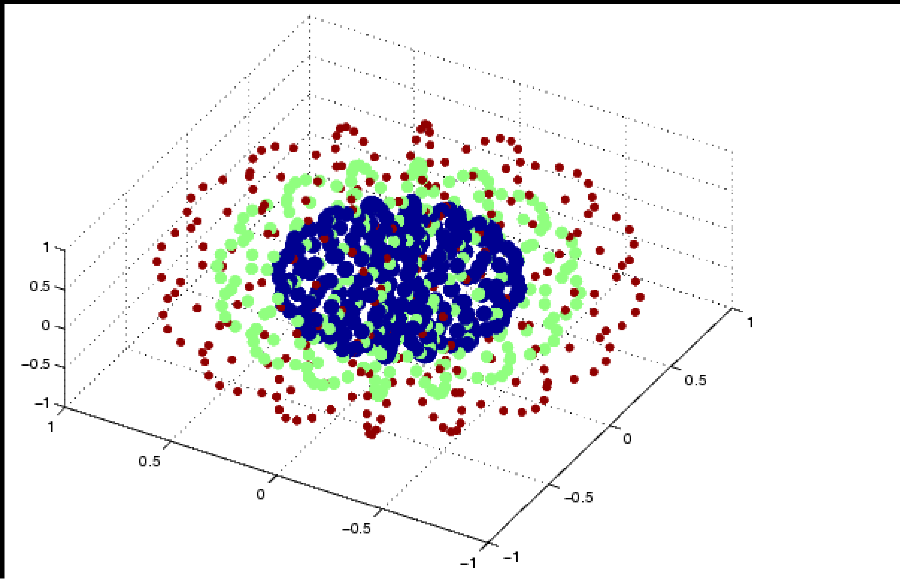
- Plot of value of one variable at position **x** against the value of the same variable at position **x+h**, **h** being a separation vector

H-Scatter plot

- Tolerance
  - Direction and magnitude





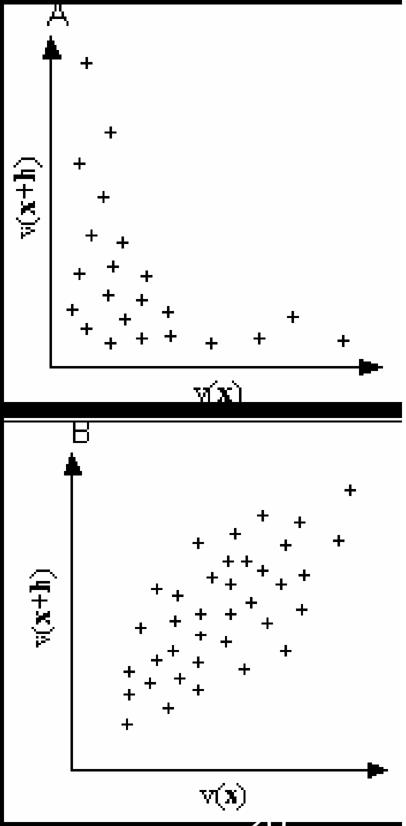


11/12/2002

H Scatter plot

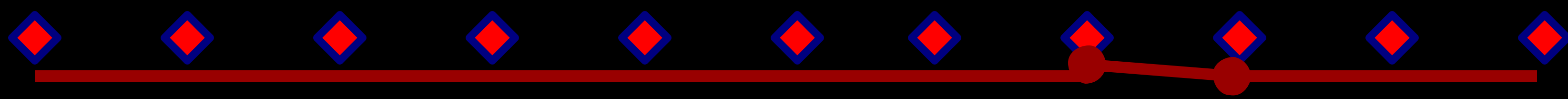
- Associated with each g**(h)** value of the experimental semivariogram
- Spatial Continuity
  - Better continuity in H-Scatter plot = better meaning of directional variogram
- Pattern:
  - Butterfly pattern: Skewed distribution, requires transformation
  - Groups of pairs: mixture of different populations
    - A way to separate different populations
  - Visual check of multigausian hypothesis

  (multigausian = limit distribution of central limit theorems. elliptical cloud around the diagonal)



Experimental Semivariogram

Lag (h) = 1unit

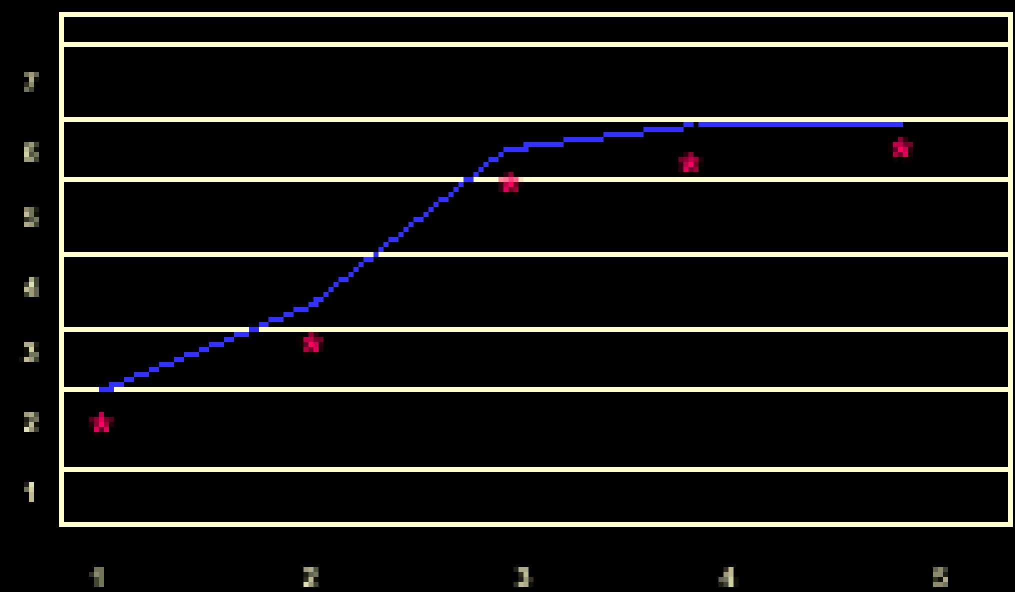


Lag (h) = 2unit

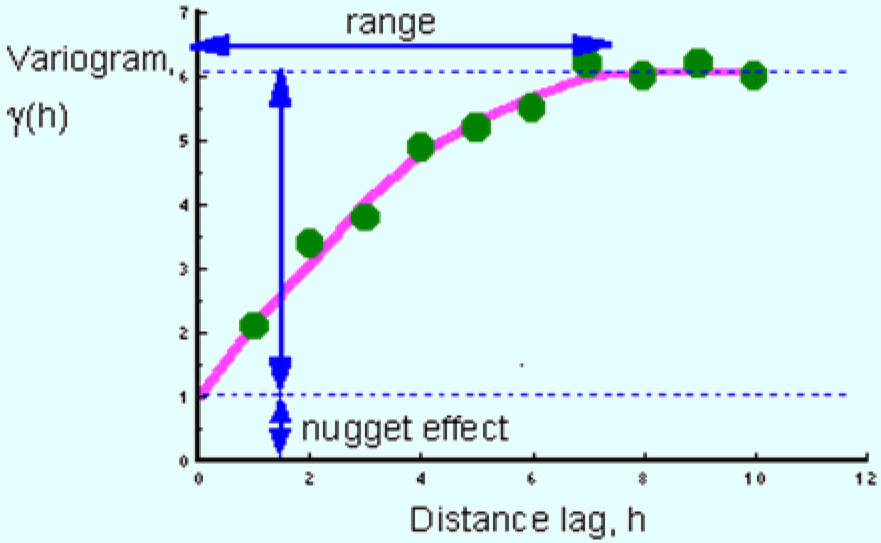
Lag (h) = 3unit

Experimental Semivariogram

- Sample points: 3 4 6 5 7 7 6 4 3 5 5 6 5 7
- N=14, 2
- <u>Distance = h</u> <u>2?(h)= {z(x) - z(x+h)}2</u>
- 1 26/13 = 2.0
- 2 41/12 = 3.4
- 3 55/11 = 5.1
- 4 50/10 = 5.8
- 5 53/9 = 5.9



Experimental Semivariogram



COMPONENTS

•**Sill :** Total Variance

•**Range:** The (max.) distance over which samples are correlated (zone of influence)

•**Nugget:** γ*(h)* @ h→0 : magnitude of the discontinuity

- A result of human error
- inability to sample at the very same sample location more than once
- Laboratory error

Semivariogram Formula

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

$= E[\{(Z_{(x+h)}-m)^2 + (Z_{(x)}-m)^2\} -$

$2(Z_{(x+h)}-m)(Z_{(x)}-m)\}$

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

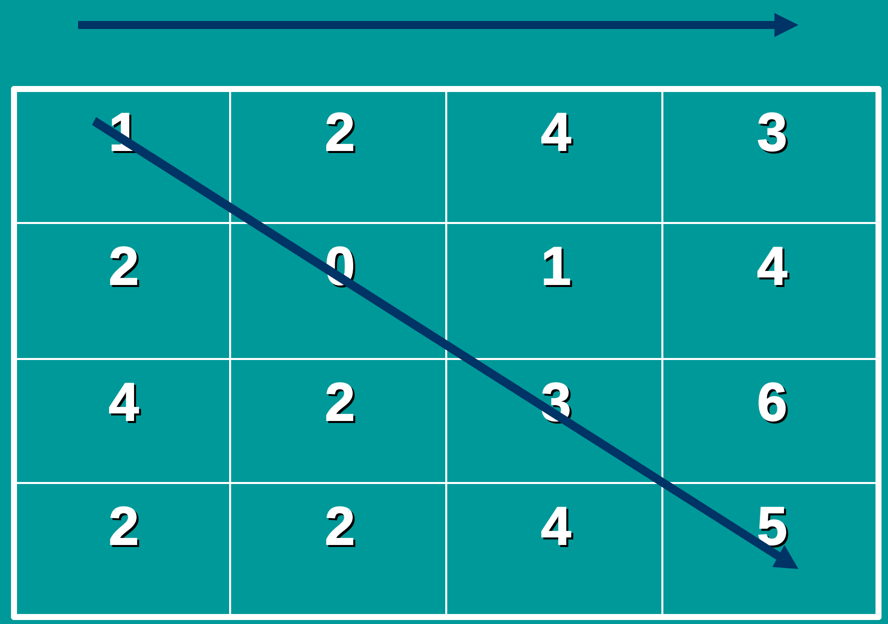
$\Rightarrow \gamma(h) = C(0) - C(h)$

$\Rightarrow \gamma(h) = \text{Sill} - C(h)$

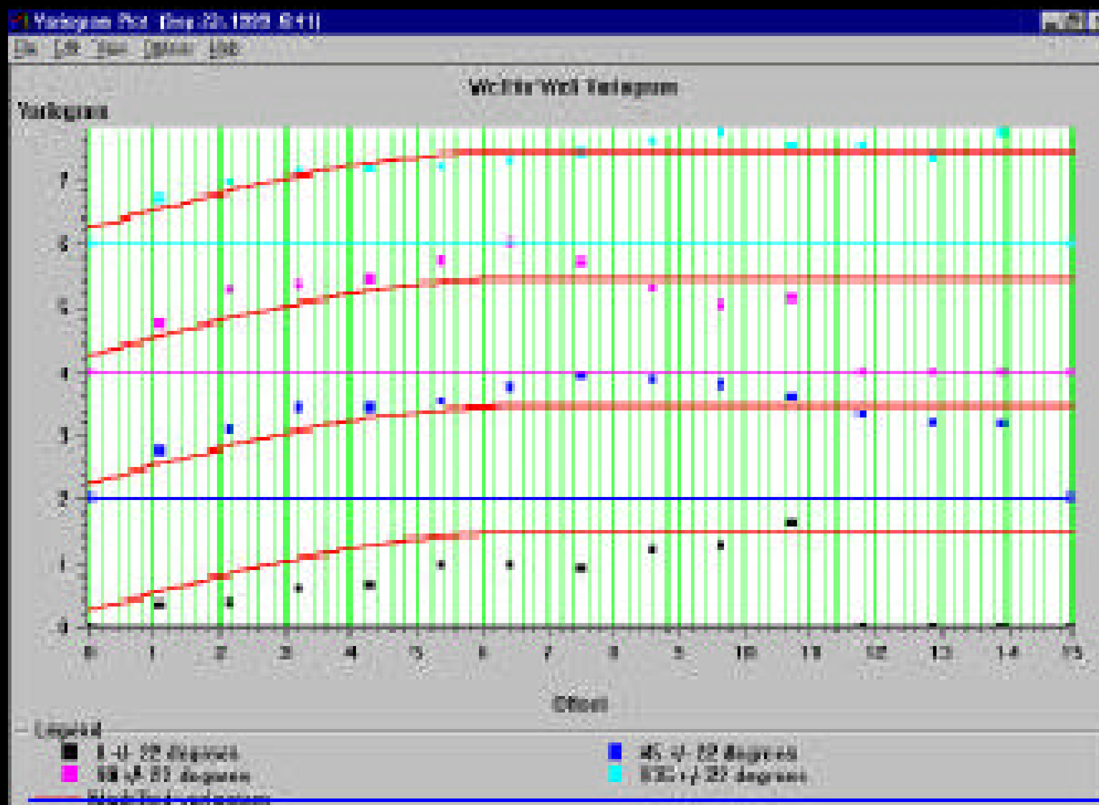
$\Rightarrow \gamma(h) = \sigma^2 - C(h)$

Directional Semivariogram: Two Dimensional Example

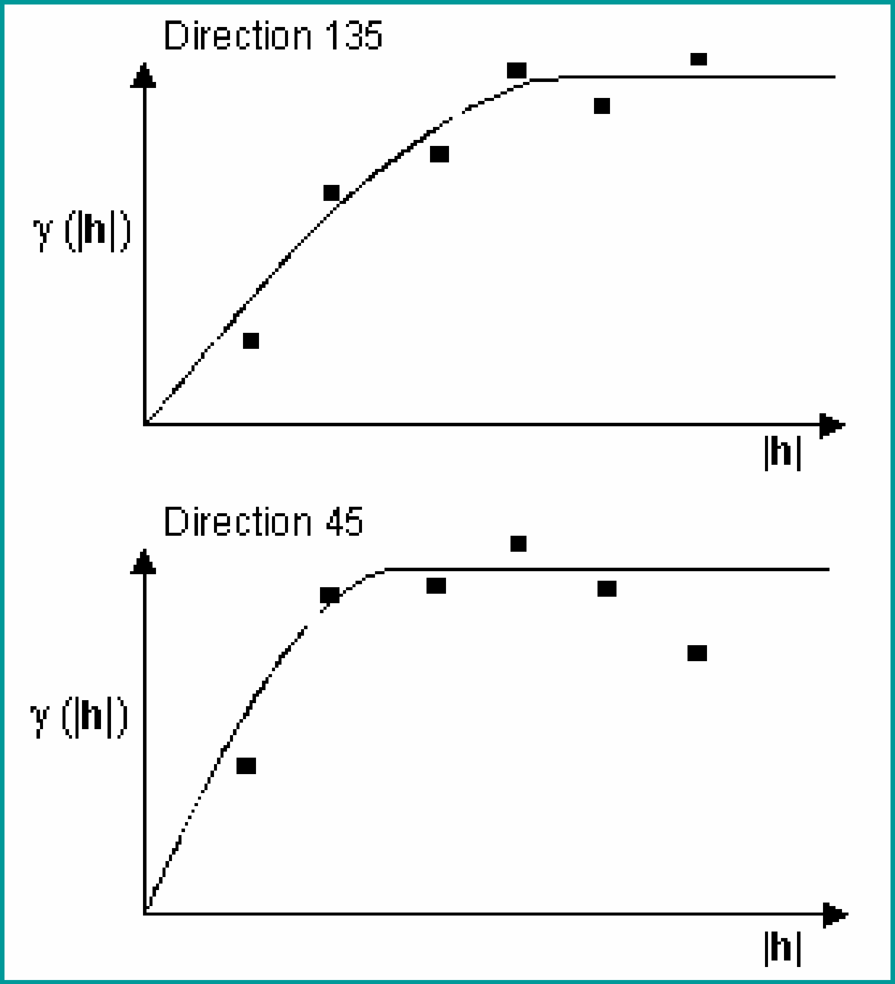
Distance	N	Semivariance
1.000	24	1.538
1.414	18	2.528
1.690	34	2.662
2.000	16	2.813
2.384	32	3.110
2.236	24	3.000
3.097	20	2.000
2.828	8	3.438
3.733	10	3.950
3.000	8	1.625
3.162	12	2.250
3.309	30	2.650
3.606	8	3.875
4.243	2	4.250



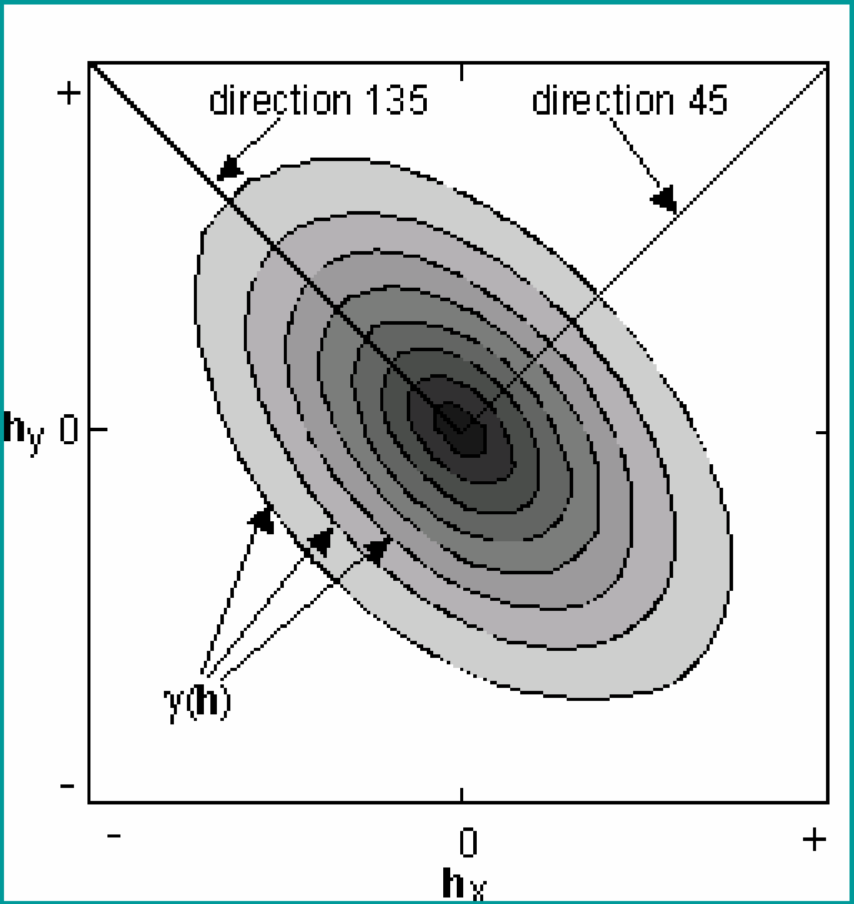
Directional Semivariogram: Two Dimensional Example



Directional Semivariogram:

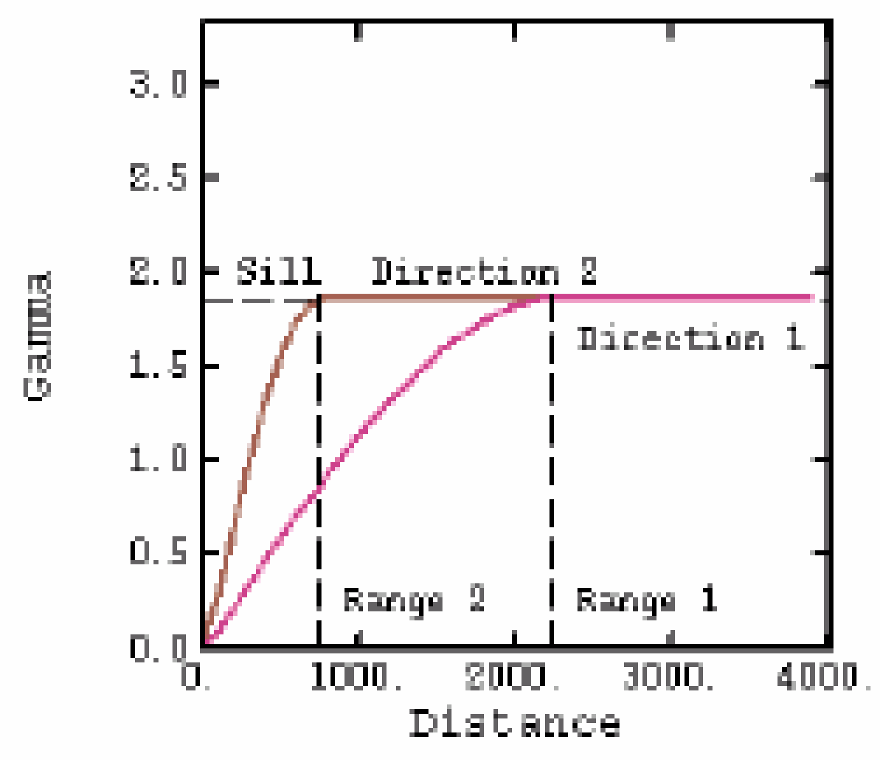


Variograms

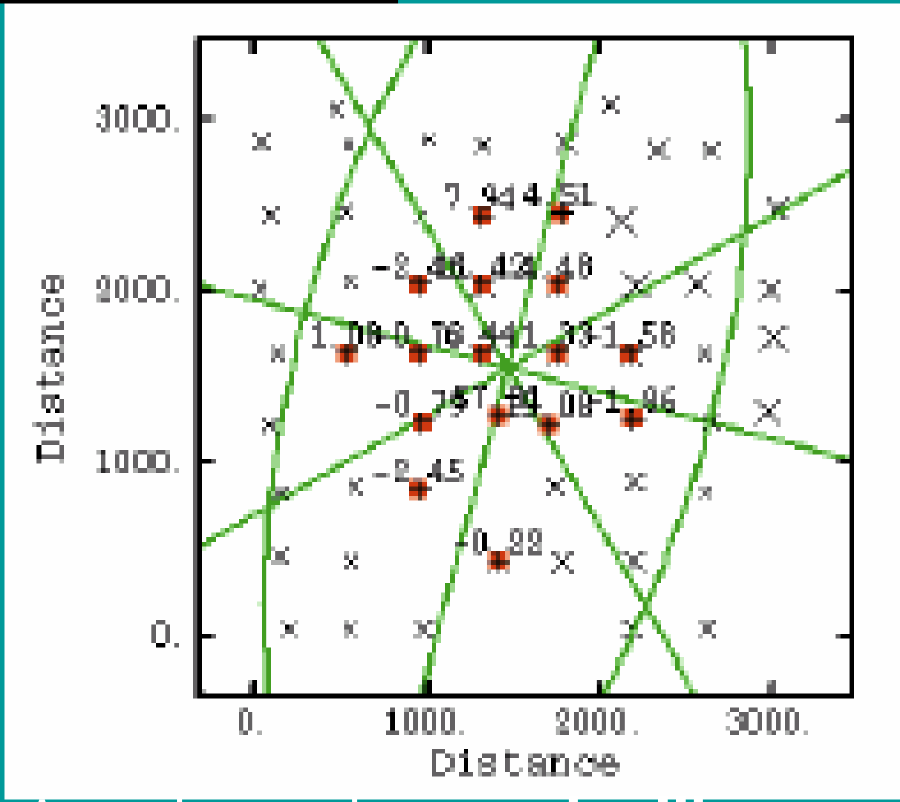


Variogram Surface

Directional Semivariogram: Another Example



Anisotropic variogram short-scale range of 800 m and long-scale range of 2200 m.



An anisotropic **search ellipse** with eight sectors and a maximum of two data points per sector. The minor axis has a length of 1000 m.

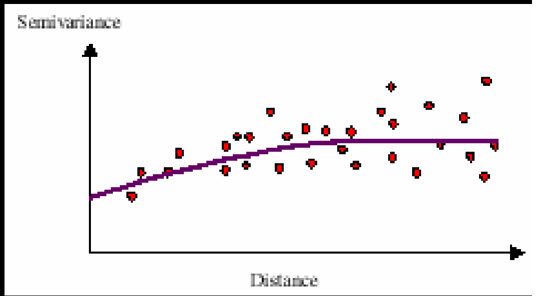
Ref: Petroleum geostatistics for nongeostatisticians, *THE LEADING EDGE MAY 2000*

Anisotropism: Directional Semivariogram

- Spatial structure may not be the same in all directions :"Anisotropic"
  - Spatial Structure same in all directions: Isotropic
- Develop a semi-variogram for each (major) direction
- Major Directions: $0^0, 45^0, 90^0, 135^0$
- Angle of tolerance ($\leq 22.5^0$)

Semivariogram Modeling

$$g(h) = \frac{1}{2N(h)} \sum_{i=1}^{n} [Z(x) - Z(x+h)]^2$$

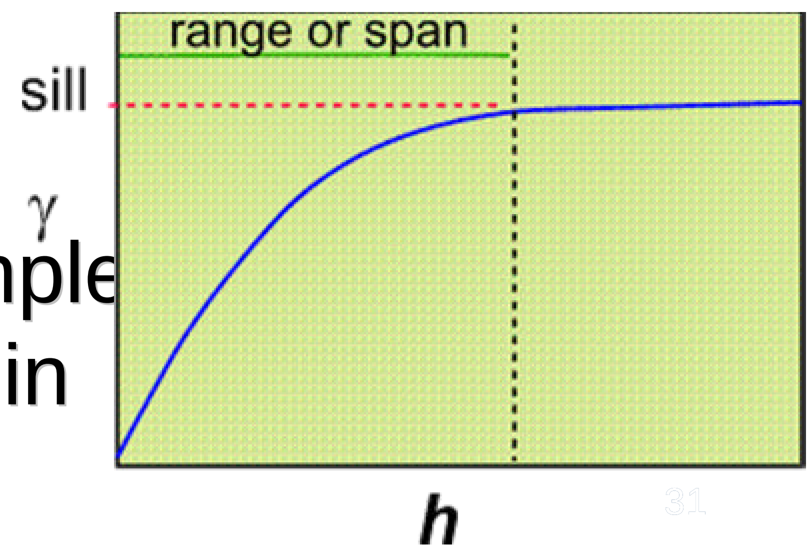


1. **Experimental Semivariogram**
2. **Model fitting:**
   - Hand fitting
   - Weighted Least Square method *(avoided always)* ***Why?***
     1. ***Model must be positive definitive function***
        - ***Otherwise variance <0 (impossible)***
        - ***Rarely satisfied by LS method***
     2. ***LS assumes : samples are independent***
        - ***NOT true in case of spatially distributed data (natural data)***
     3. ***Behavior @ origin at i.e., distance shorter than first lag (h)***
        - ***IGNORED by LS method***

SEMIVARIOGRAM FITTING

Rules of Thumb

- Fit the model well at the origin
- Range = 3/2 of the distance where the tangent to the experimental semivariogram meets Sill.
- Take 3-4 points with comparatively more sample pairs starting at the origin

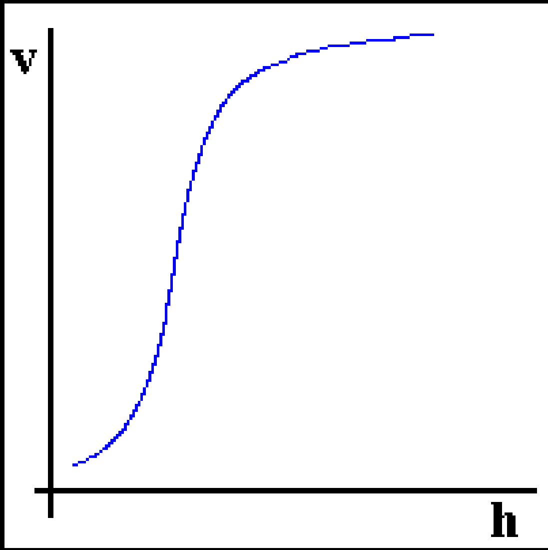


$$g(h) = \frac{1}{2} E[Z(x) - Z(x+h)]^2$$

Semivariogram Models

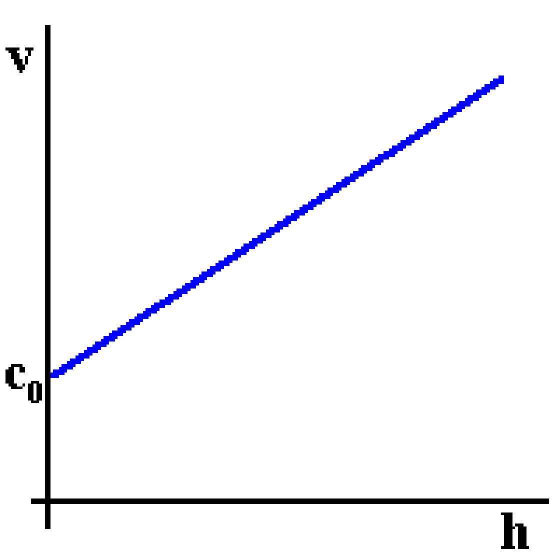
- **Spherical**

$$g(h) = \begin{cases} C\left(\frac{3}{2}\frac{|h|}{a} - \frac{1}{2}\left(\frac{|h|^3}{a^3}\right)\right) & |h| < a \\ C & |h| \geq a \end{cases}$$



- **Linear**

$$g(h) = Ah + B$$

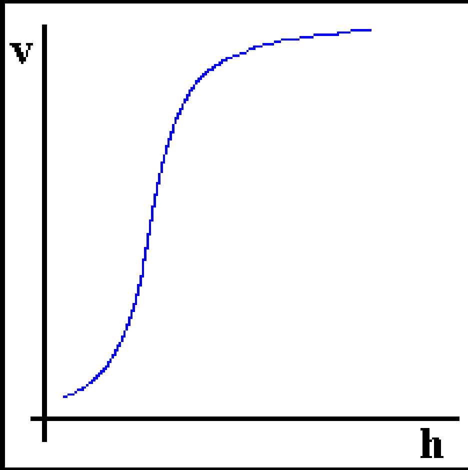


$$g(h) = \frac{1}{2} E[Z(x) - Z(x+h)]^2$$

Semivariogram Models

- **Gausian**

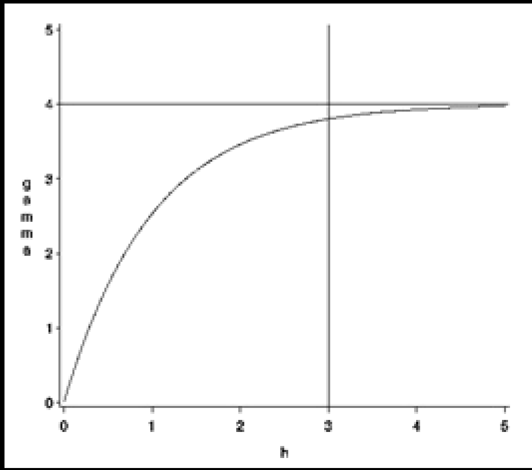
$$g(h) = C\left(1 - \exp(-\frac{|h|^2}{a^2})\right)$$



-

- **Exponential**

$$g(h) = C(1 - \exp(-|h|/a))$$

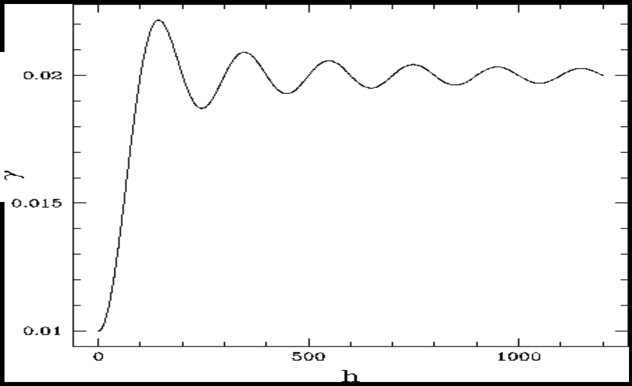


$$g(h) = \frac{1}{2} E[Z(x) - Z(x+h)]^2$$

Semivariogram Models

- **Hole effect**

$$g(h) = C\left(1 - \frac{\sin ah}{ah}\right)$$



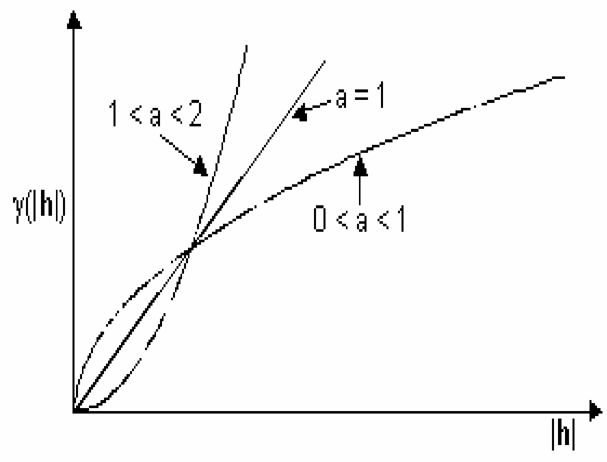
- **De Wijsian**

$$\ln g(h) = l \ln h$$

- **(No model available in this slide)**

- **Power**

$$g(h) = C|h|^a, with\, 0 < a \leq 2$$



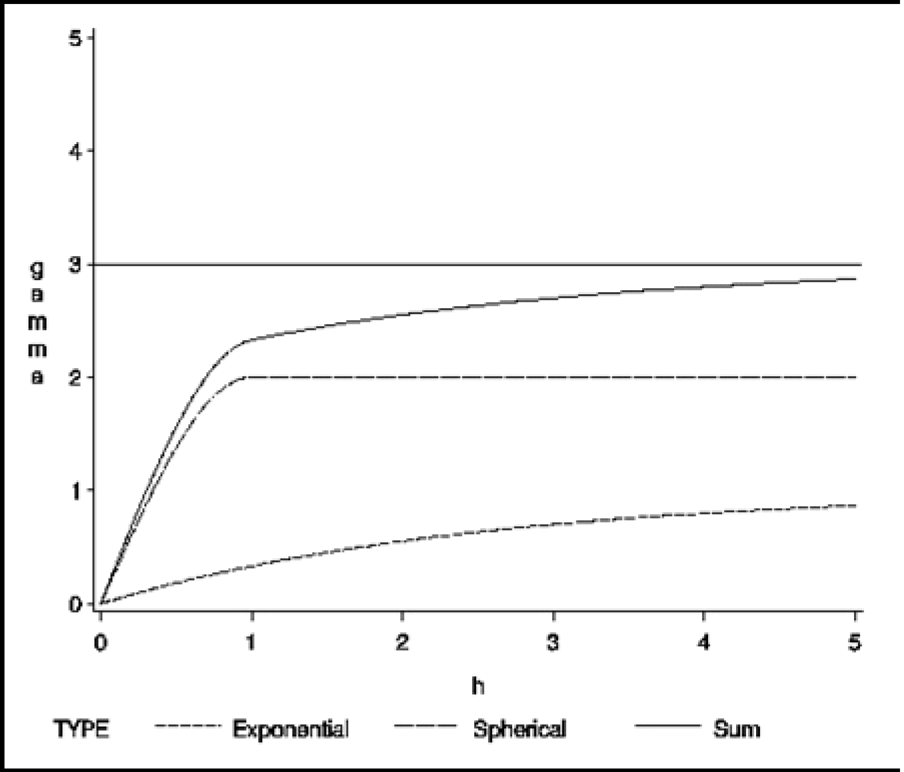
- **Nugget**

$$g(h) = \begin{array}{ll} 0 & h = 0 \\ C & |h| > 0 \end{array}$$

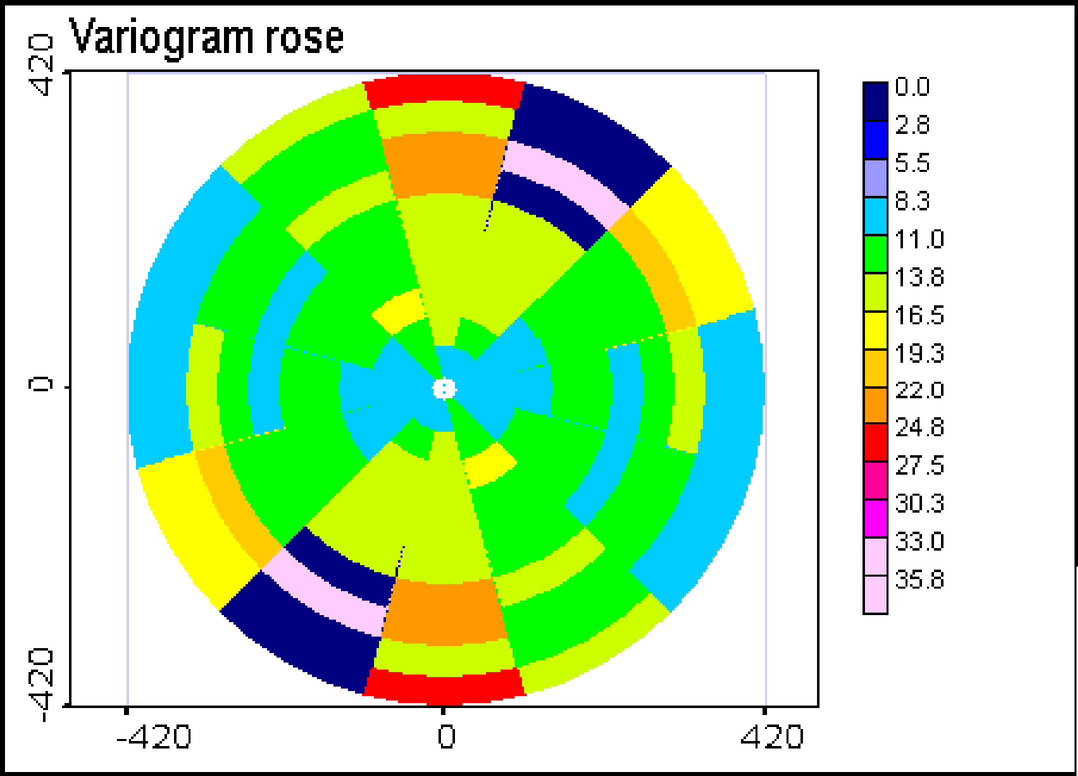
Pure nugget-effect; no spatial correlation

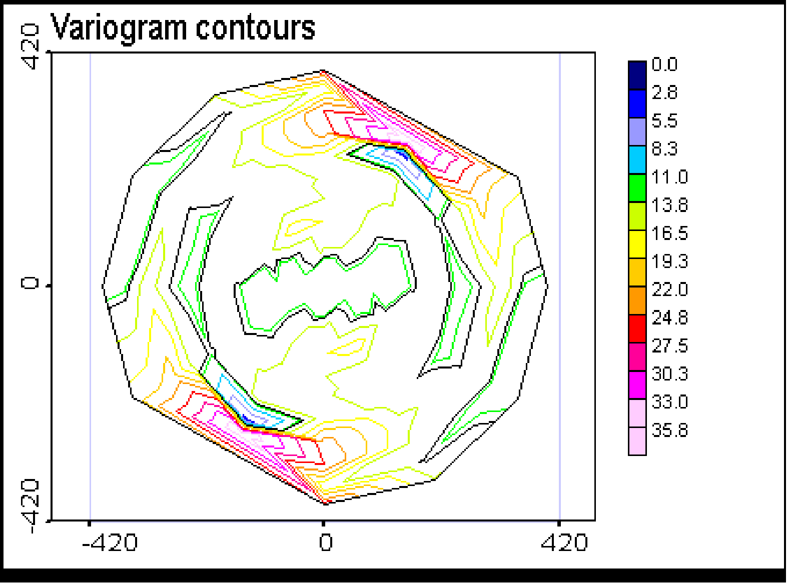
NESTED Semivariogram

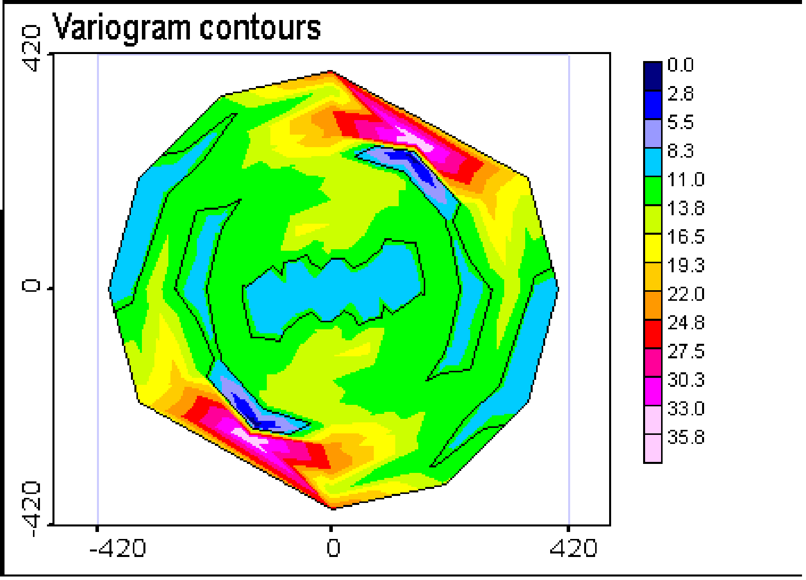
- **Sum of Exponential and Spherical Structures at Different Scales**
- **does not resemble any *single* -semivariogram**
- **sill value is the sum of the individual sills $C_{0,1}=1$ and $C_{0,2}=2$.**



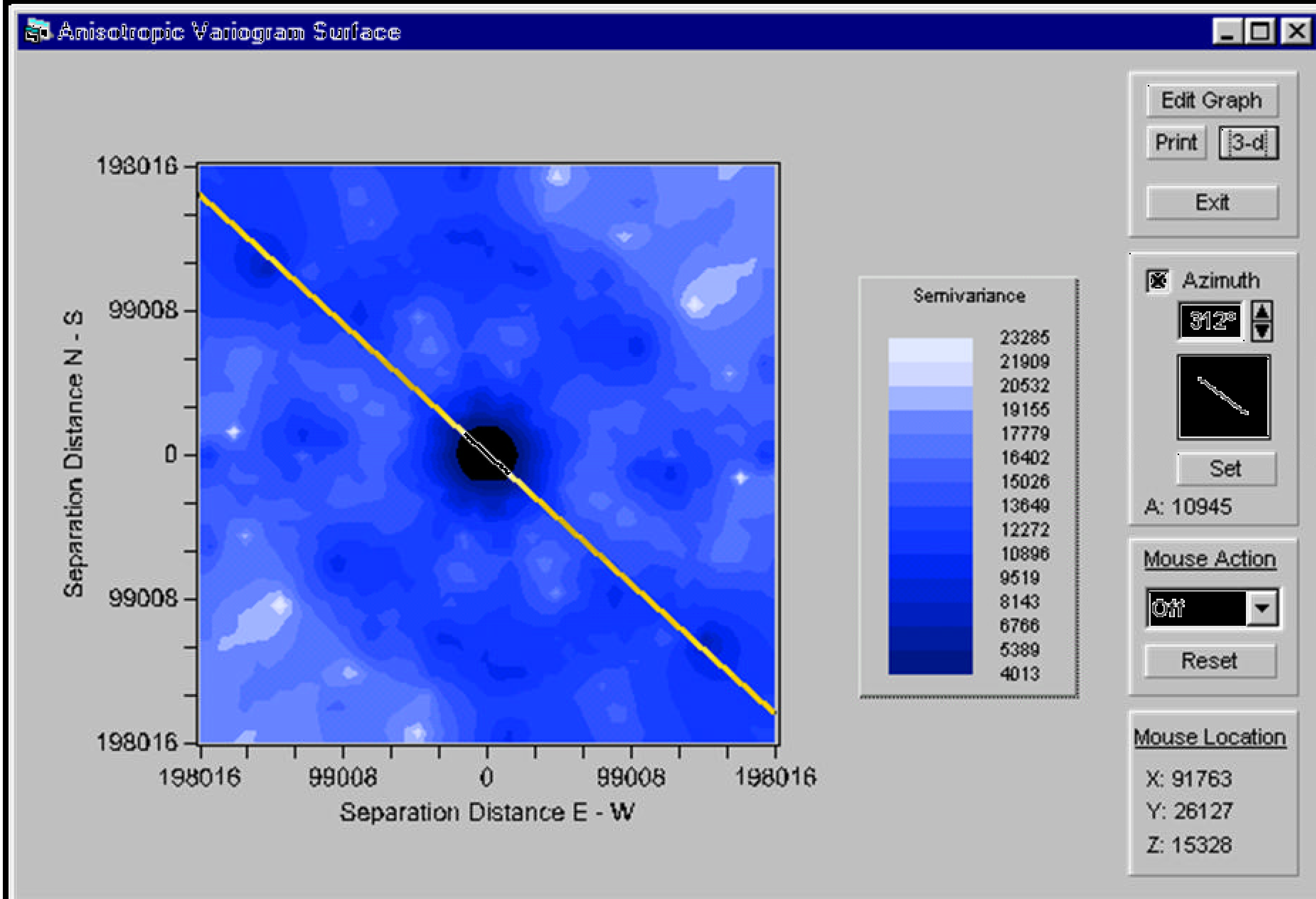
Presenting and interpreting Directional Semivariogram:







11/12/2002

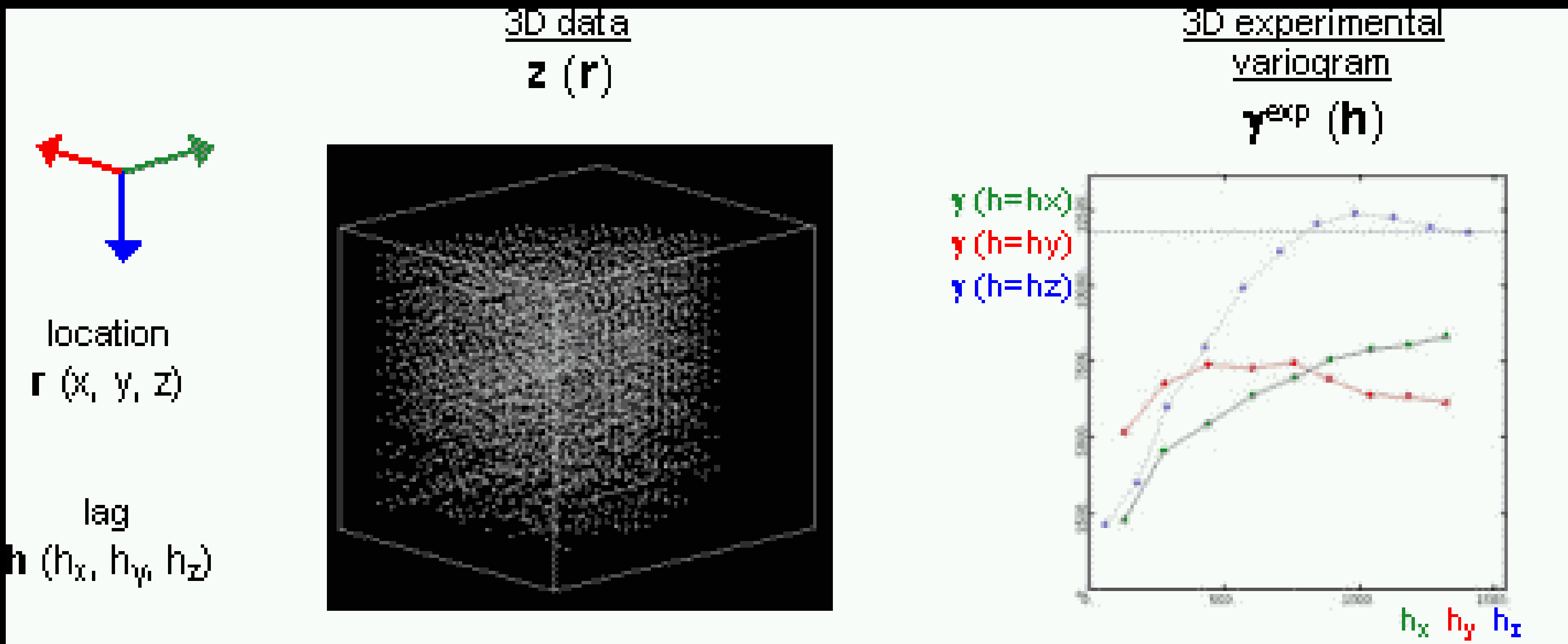


Anisotropic Variogram Surface
Edit Graph
Print
3-d
Exit
198016
99008
0
99008
198016
Separation Distance N - S
198016
99008
0
99008
198016
Separation Distance E - W
Semivariance
23285
21909
20532
19155
17779
16402
15026
13649
12272
10896
9519
8143
6766
5389
4013
Azimuth
312°
Set
A: 10945
Mouse Action
Off
Reset
Mouse Location
X: 91763
Y: 26127
Z: 15328

VALIDATION OF SEMIVARIOGRAM MODEL

- PKCV
- Goodness of fit
  - It gives a measure of how well the model adjusts the directional variograms
  - a number without units and a value close to zero indicates a good fit
  - a standardized measure of fit

3D variography



$$\gamma^{\exp}(\mathrm{h}) = \frac{1}{2n_c}\sum_{\alpha=1}^{n_c}\left(z(\mathrm{r}_\alpha + \mathrm{h}) - z(\mathrm{r}_\alpha)\right)^2$$

Source:http://www.erms.fr/english/applied_geostatistics.html#vario

Goodness of Fit:

$$GF = \frac{1}{N}\sum_{k=1}^{N}\sum_{i=0}^{n(k)} \frac{hmax(k)}{\sum_{j=0}^{n(k)} P(j)} \cdot \frac{P(i)}{h(i)} \cdot \left[\frac{\gamma(i) - \gamma^{*}(i)}{\sigma^{2}}\right]^{2}$$

- **N = number of directional variogram measures used for the model**
- **n(k) = number of lag for the kth variogram measure**
- **P(i) = number of pair for lag i**
- **h(i) = mean distance for lag i**
- **hmax(k) = maximum distance for the kth variogram measure**
- g ***(i)* = experimental variogram measure for lag i**
- g ***(i)* = modeled variogram measure for the mean distance of lag i**
- **s^2 = variance of the data for the semivariogram and the non ergodic covariance, maximum experimental value of all variogram measures for the madogram, 1 for the correlogram and the standardized inverted covariance.**

Search Criteria

	2D	3D
Isotropic	**Circle**	**Sphere**
Anisotropic	**Ellipse**	**Ellipsoid**

Kriging:

- **Interpolation:** *what measured values tell us about the properties at un-sampled locations*.
- a geostatistical inter-polation technique.
- a linear weighted-averaging method
- Kriging weights depend on a model of spatial correlation.

Kriging

- Why Kriging:

- Best : Minimum Kriging Variance (σ^2_k)
- Linear:

$$Z_i^* = \sum l_i Z(X_i)$$
$$\sum l_i = 1$$

- l_i: Weights, $Z(x_i)$ are sample values

- Unbiased : Error → 0 i.e, $E[Z(x) - Z^*(X)] = 0$
- Estimator

Common Types of Kriging:

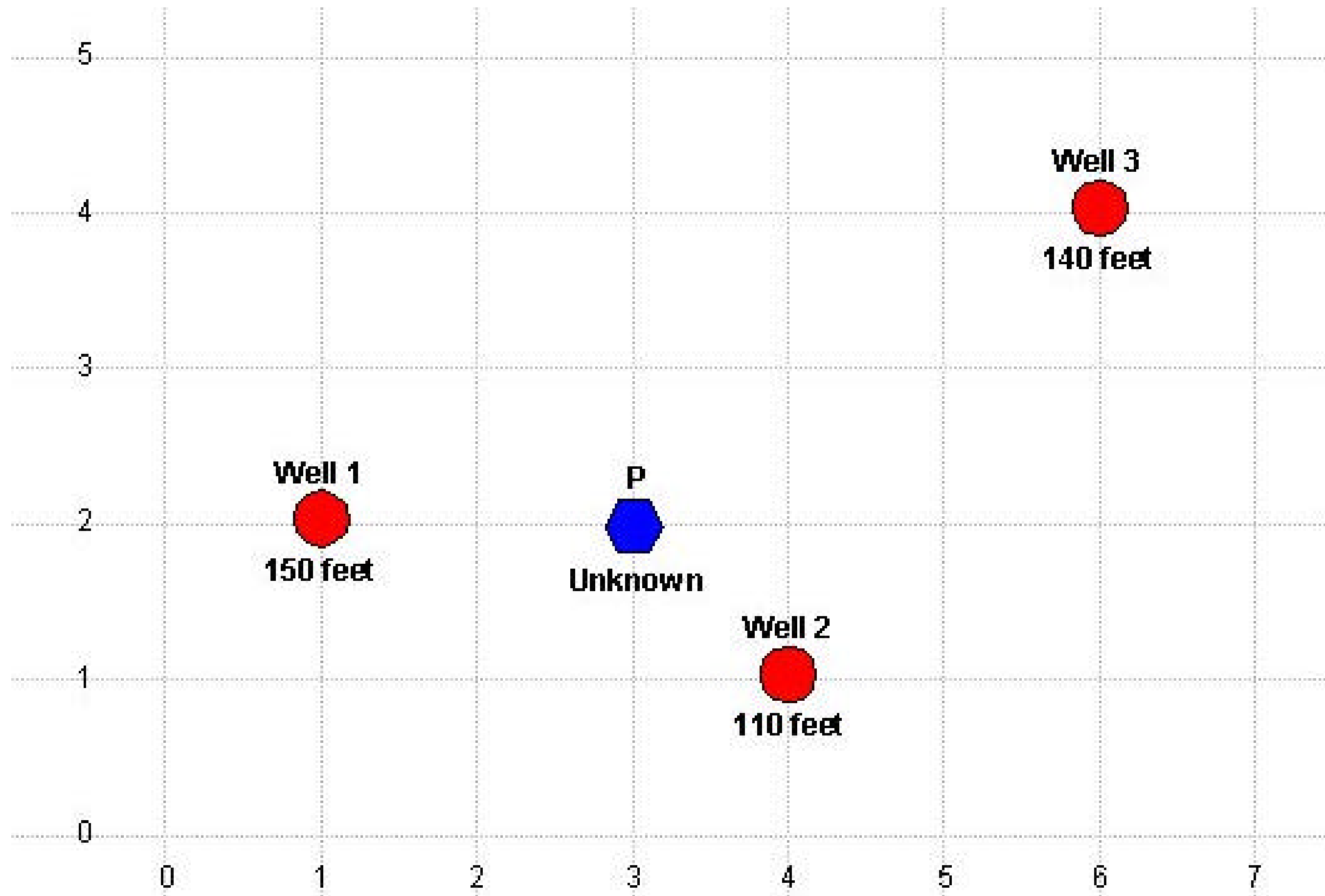
- **Punctual Kriging (also termed Ordinary Kriging):** uses only the samples in the local neighborhood for the estimate. most common method used in environmental engineering.
- **Block Kriging:** Estimating the value of a block from a set of nearby sample values using kriging.
- **Point Kriging:** Estimating the value of a point from a set of nearby sample values using kriging.

  When a kriged point happens to coincide with a sample location, the kriged estimate = the sample value.
- **Universal Kriging :** used when a trend, or slow change in average values, in the samples exists
- **Co-kriging:** for more than one variable

Kriging Formula:

$$\begin{bmatrix} \gamma(x_1 - x_1) & \gamma(x_1 - x_2) & \cdots & \gamma(x_1 - x_N) & 1 \\ \gamma(x_2 - x_1) & \gamma(x_2 - x_2) & \cdots & \gamma(x_2 - x_N) & 1 \\ \vdots & \vdots & \ddots & \vdots & 1 \\ \gamma(x_N - x_1) & \gamma(x_N - x_2) & \cdots & \gamma(x_N - x_N) & 1 \\ 1 & 1 & \cdots & 1 & 0 \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_N \\ \mu \end{bmatrix} = \begin{bmatrix} \gamma(x_1 - x) \\ \gamma(x_2 - x) \\ \vdots \\ \gamma(x_N - x) \\ 1 \end{bmatrix},$$

An Example:



1	2	3	p	
1	0	3.16	5.39	2
2	-	0	3.61	1.41
3	-	-	0	3.61

Distance Matrix

	1	2	3	p
1	0	12.65	21.54	8
2	-	0	14.42	5.66
3	-	-	0	14.42

Semivariogram Matrix

λ

The complete set of simultaneous equations are:

$$W_1\gamma(h_{11}) + W_2\gamma(h_{12}) + W_3\gamma(h_{13}) + \lambda = \gamma(h_{1p})$$

$$W_1\gamma(h_{21}) + W_2\gamma(h_{22}) + W_3\gamma(h_{23}) + \lambda = \gamma(h_{2p})$$

$$W_1\gamma(h_{31}) + W_2\gamma(h_{32}) + W_3\gamma(h_{33}) + \lambda = \gamma(h_{3p})$$

$$W_1 + W_2 + W_3 + 0 = 1$$

s_k^2

Wi are weights, and S Wi = 1

λIs Lagrange multiplier (to ensure minimum possible kriging variance is obtained)

In matrix form:

$$\begin{bmatrix} \gamma(h_{11}) & \gamma(h_{12}) & \gamma(h_{13}) & 1 \\ \gamma(h_{21}) & \gamma(h_{22}) & \gamma(h_{23}) & 1 \\ \gamma(h_{31}) & \gamma(h_{32}) & \gamma(h_{33}) & 1 \\ 1 & 1 & 1 & 0 \end{bmatrix} \bullet \begin{bmatrix} W_1 \\ W_2 \\ W_3 \\ \lambda \end{bmatrix} = \begin{bmatrix} \gamma(h_{1p}) \\ \gamma(h_{2p}) \\ \gamma(h_{3p}) \\ 1 \end{bmatrix}$$

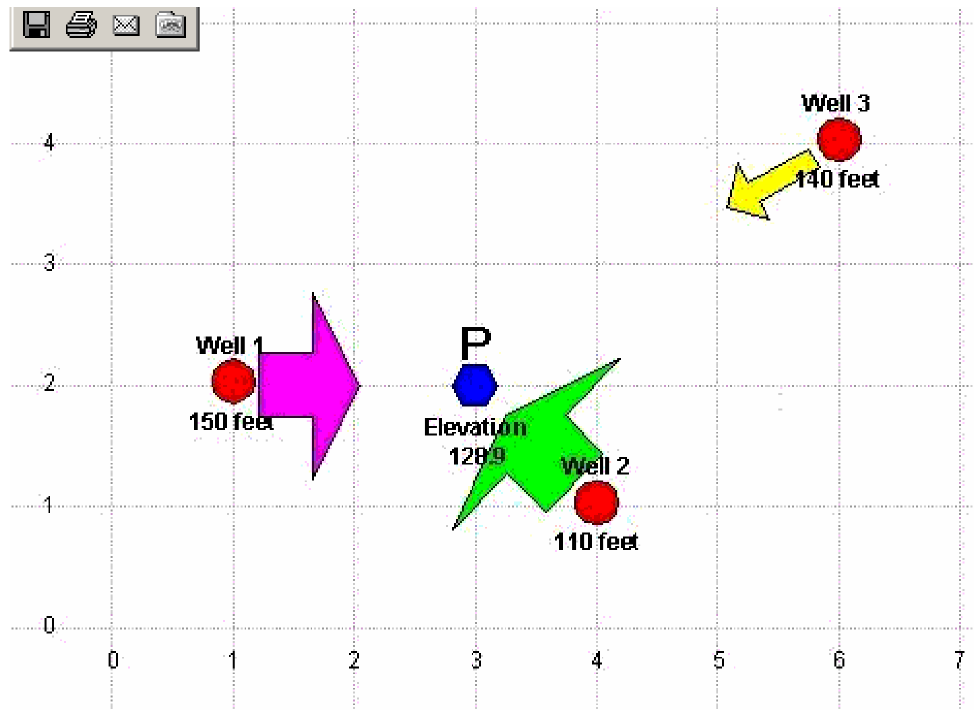
$$\left[g_{ij} \right] * \left[W_i \right] = \left[\overline{g}_{ip} \right]$$

OR,

$$\left[W_i \right] = \left[\overline{g}_{ip} \right] * \left[g_{ij} \right]^{-1}$$

Solution

$$\begin{bmatrix} W_1 \\ W_2 \\ W_3 \\ \lambda \end{bmatrix} = \begin{bmatrix} 0.3805 \\ 0.4964 \\ 0.1232 \\ -0.9319 \end{bmatrix}$$



Estimation of *Unknown* value:

$Y_{E,P}$ **= 0.3805(150) + 0.4964(110) + 0.1232(140) = <u>128.9 meters</u>**

	1	2	3	p
1	0	12.65	21.54	8
2	-	0	14.42	5.66
3	-	-	0	14.42

Semivariogram Matrix

Estimation of variance:

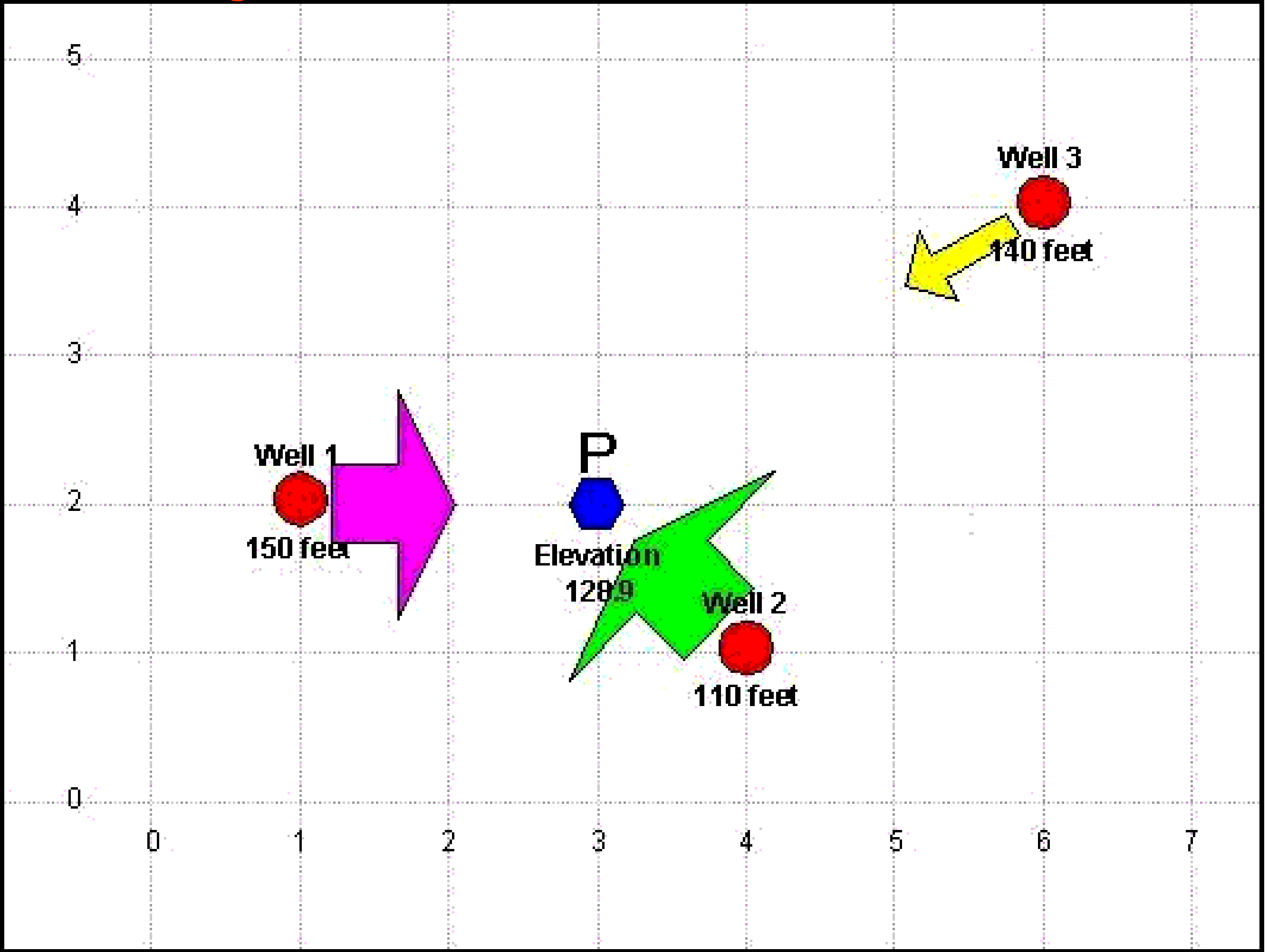
$$Y_{E,P} = 0.3805(8.0) + 0.4964(5.66) + 0.1232(14.42) - 0.9319(1.0) = 6.70\ m^2$$

Standard error : $(6.70)^{1/2} = 2.59$ meters

Assuming, errors of estimation are normally distributed:

$Y_{E,P} =$ 128.9 +/- 5.18 meters, with 95% probability

$Y_{E,Q}$ = 138.6 + / - 6.45 meters, with 95 % probability

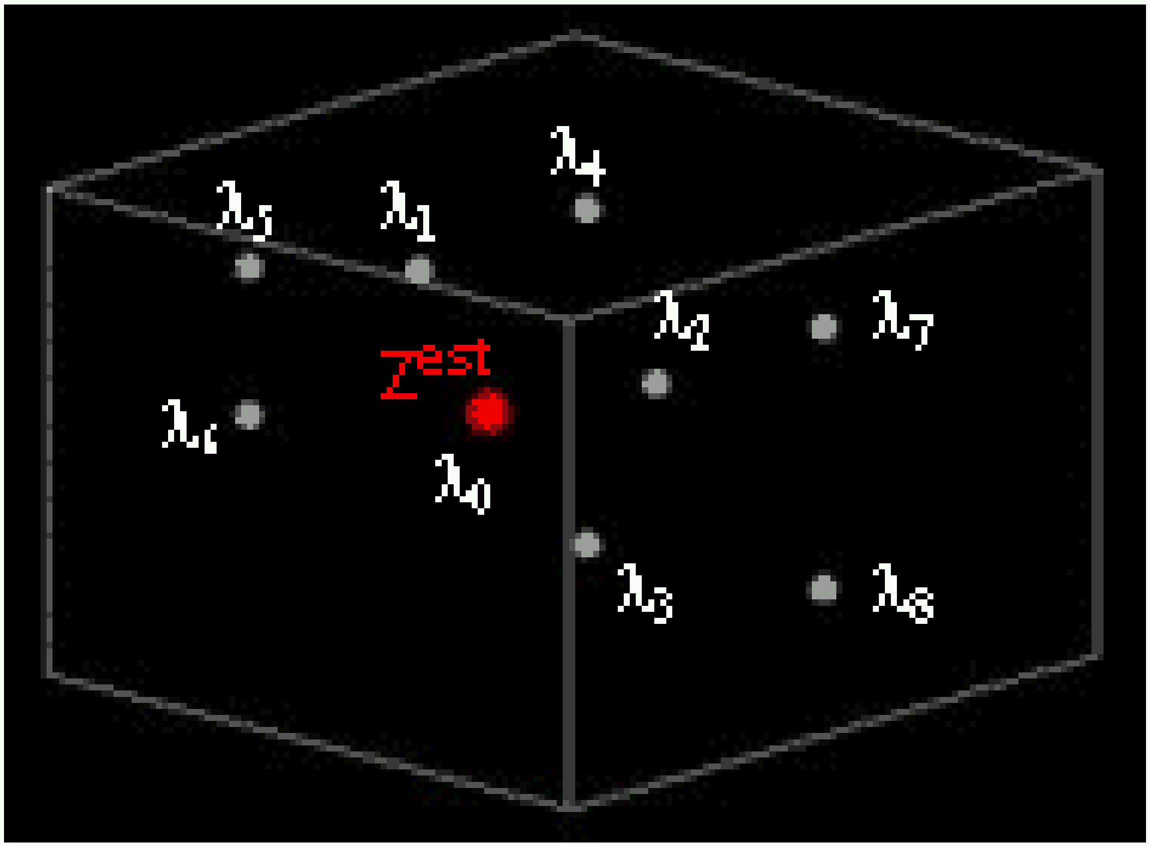


Finally

3D Data analysis

3D data

$\mathbf{z}\ (\mathbf{r})$



$$\mathbf{z}^{est}(\mathbf{r}_0) = \Sigma_i \lambda_i . \mathbf{z}(\mathbf{r}_i)$$

Free Software Systems:

- VARIOWIN: Software for Spatial Data Analysis in 2D

Download from: http://www-sst.unil.ch/research/variowin/

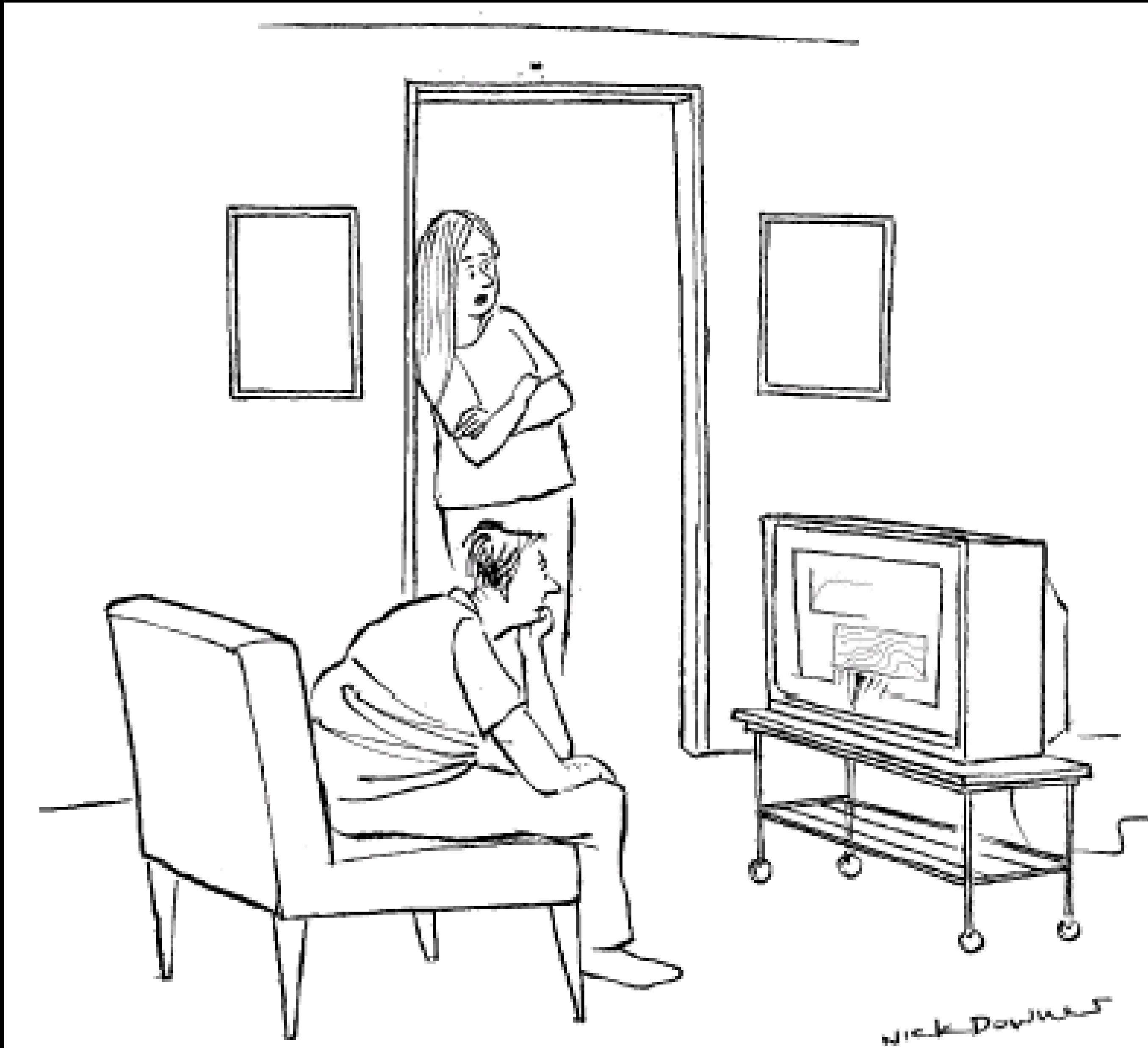
- Geo-EAS 1.2.1
- Geostatistical Toolbox 1.30
- GEOSTAT OFFICE Software
- Geostat etc
- Krigame: Software and data Download from http://geoecosse.bizland.com/softwares/index.htm

Commercial Packages:

- GS+ Version 5
- ISATIS
- Snowden ANALYSOR, Snowden VISOR etc

Applications:

- **Mining and Geology (Geochemistry, Geophysics, hydrology, Paleontology, Mine Planning, Production and development etc.)**
- **Mineral exploration**
- **GIS and Mapping**
- **Remote sensing**
- **Fishery**
- **Environment**
- **Forestry**
- **Public Administration (Crime analysis, Services etc.)**
- **Waste management**



"I didn't even know there was an 'all geostatistics' cable channel."

American Scientist (1998)