Unit 5:
Spatial Data Analysis

H.P. Nachtnebel

IWHW-BOKU

Goals and Structure

- Goal: development of tools for spatio-temporal data analysis

- Introduction
- Methodology (spatial analysis)
- Application
- Uncertainty
- Summary and conclusions

Introduction

- Environmental data exhibit often a spatial and a temporal correlation

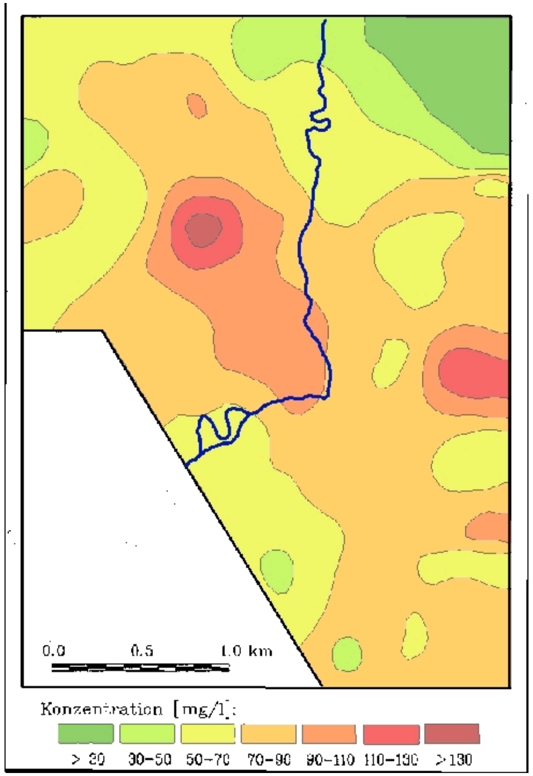
  e.g. the spreading of a pollution plume in a groundwater system

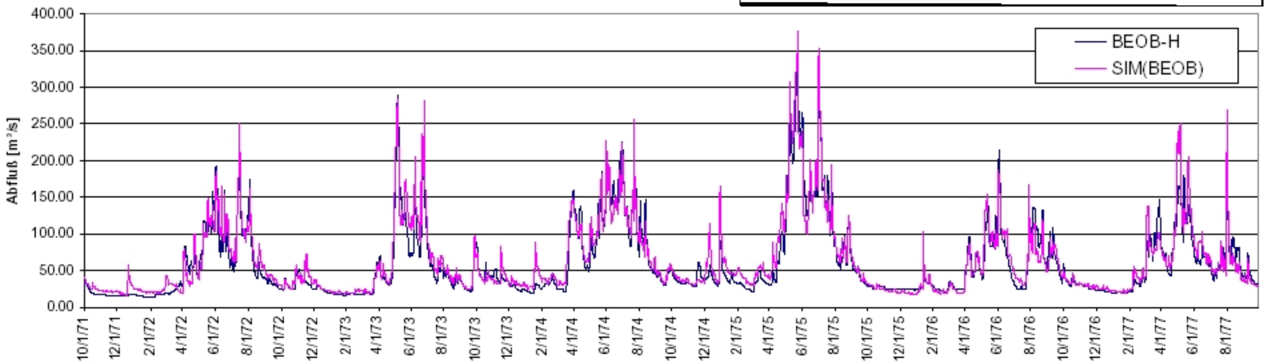
  the movement of a thunderstorm over a basin

  the leaching of pesticides through the soil to the groundwater

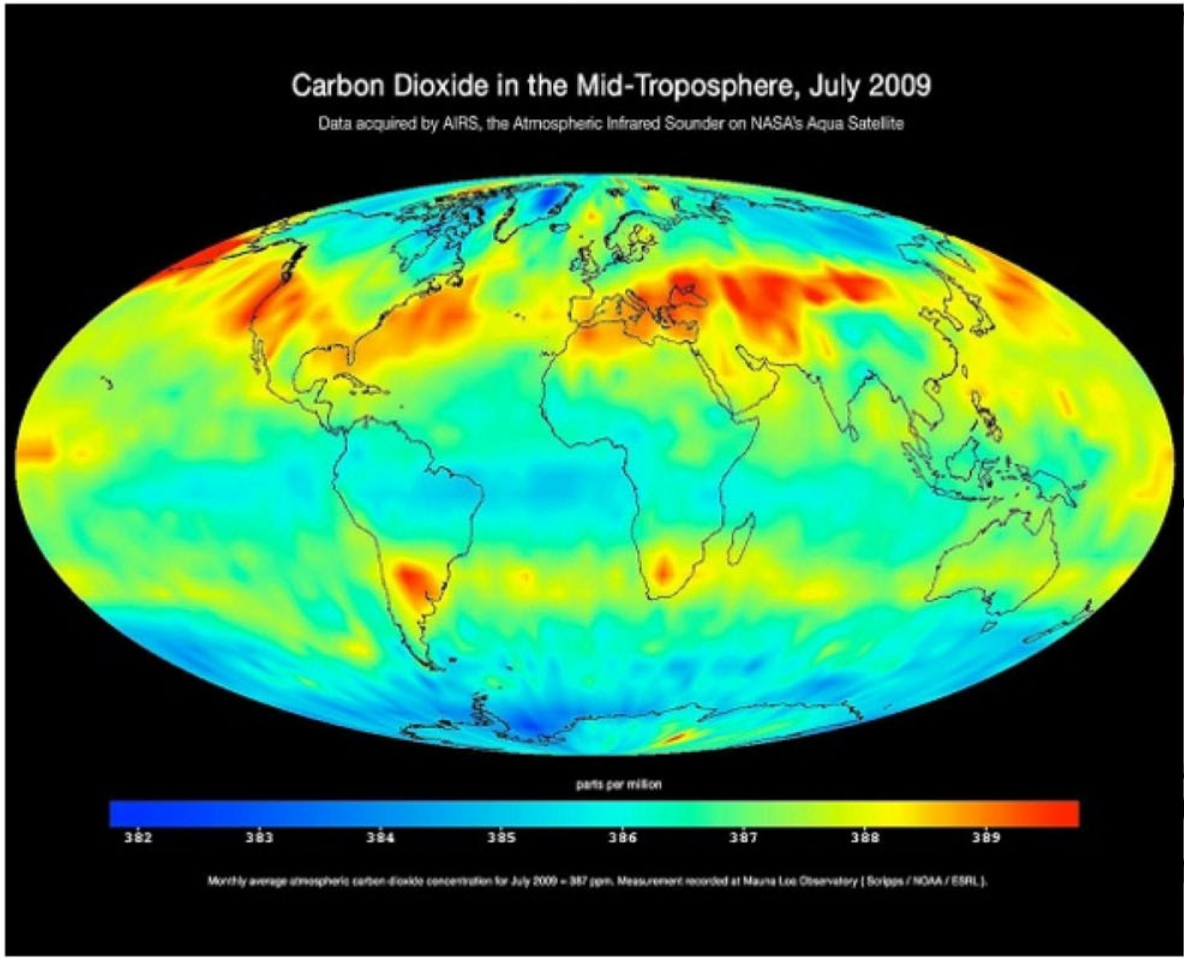
Introduction

- Spatial variability can be detected by a monitoring network
- Temporal variability can be detected by frequent sampling at a location

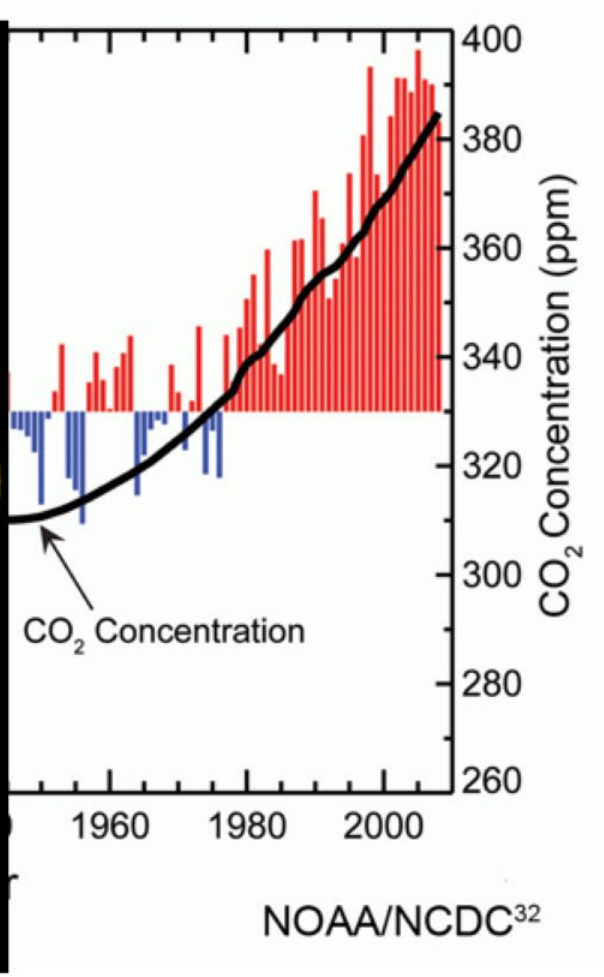




Examples



Carbon Dioxide in the Mid-Troposphere, July 2009
Data acquired by AIRS, the Atmospheric Infrared Sounder on NASA's Aqua Satellite
parts per million
382
383
384
385
386
387
388
389
Monthly average atmospheric carbon dioxide concentration for July 2009 = 387 ppm. Measurement recorded at Mauna Loa Observatory (Scripps / NOAA / ESRL).



400
380
360
340
320
300
280
260
CO2 Concentration (ppm)
CO2 Concentration
1960
1980
2000
NOAA/NCDC32

Examples

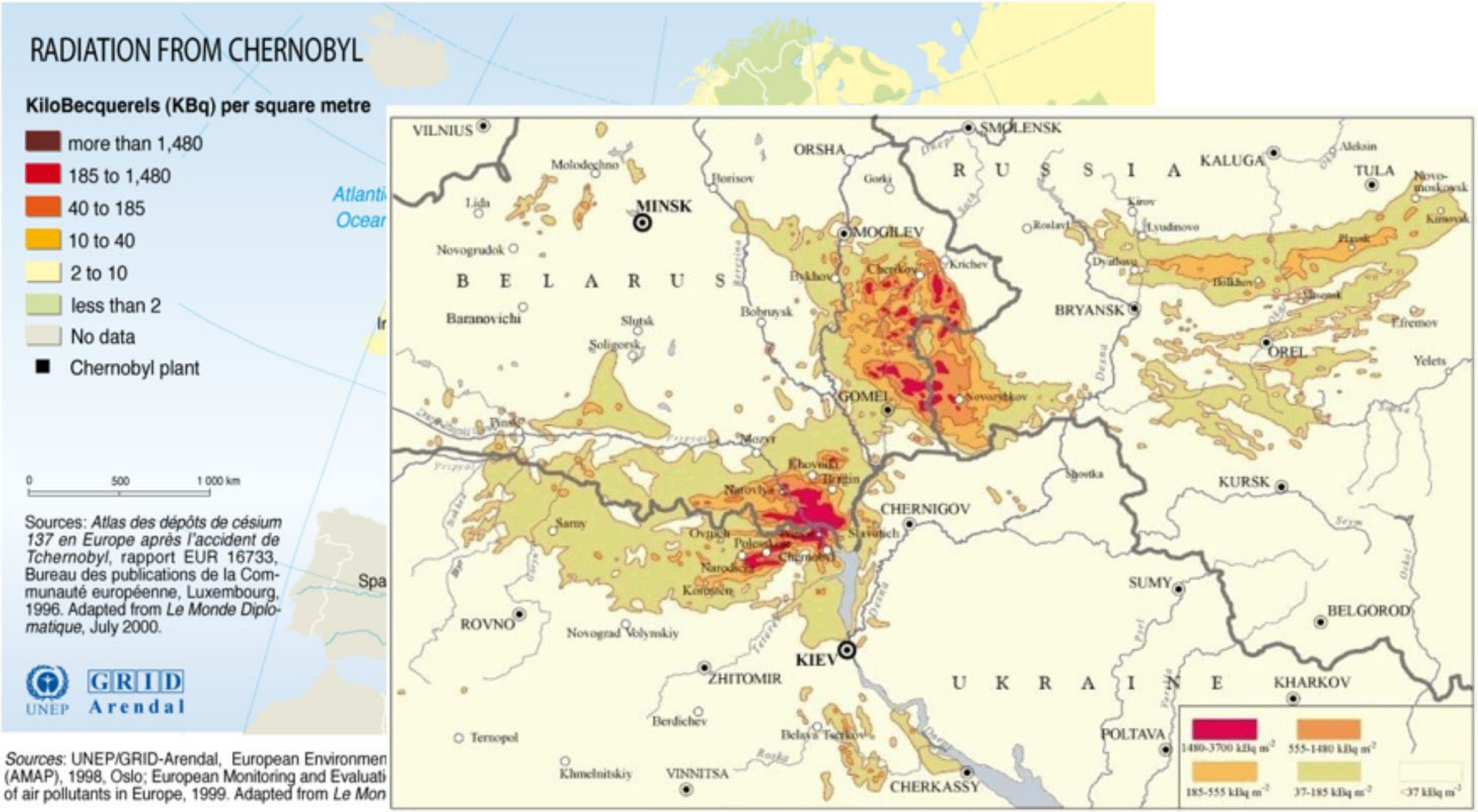
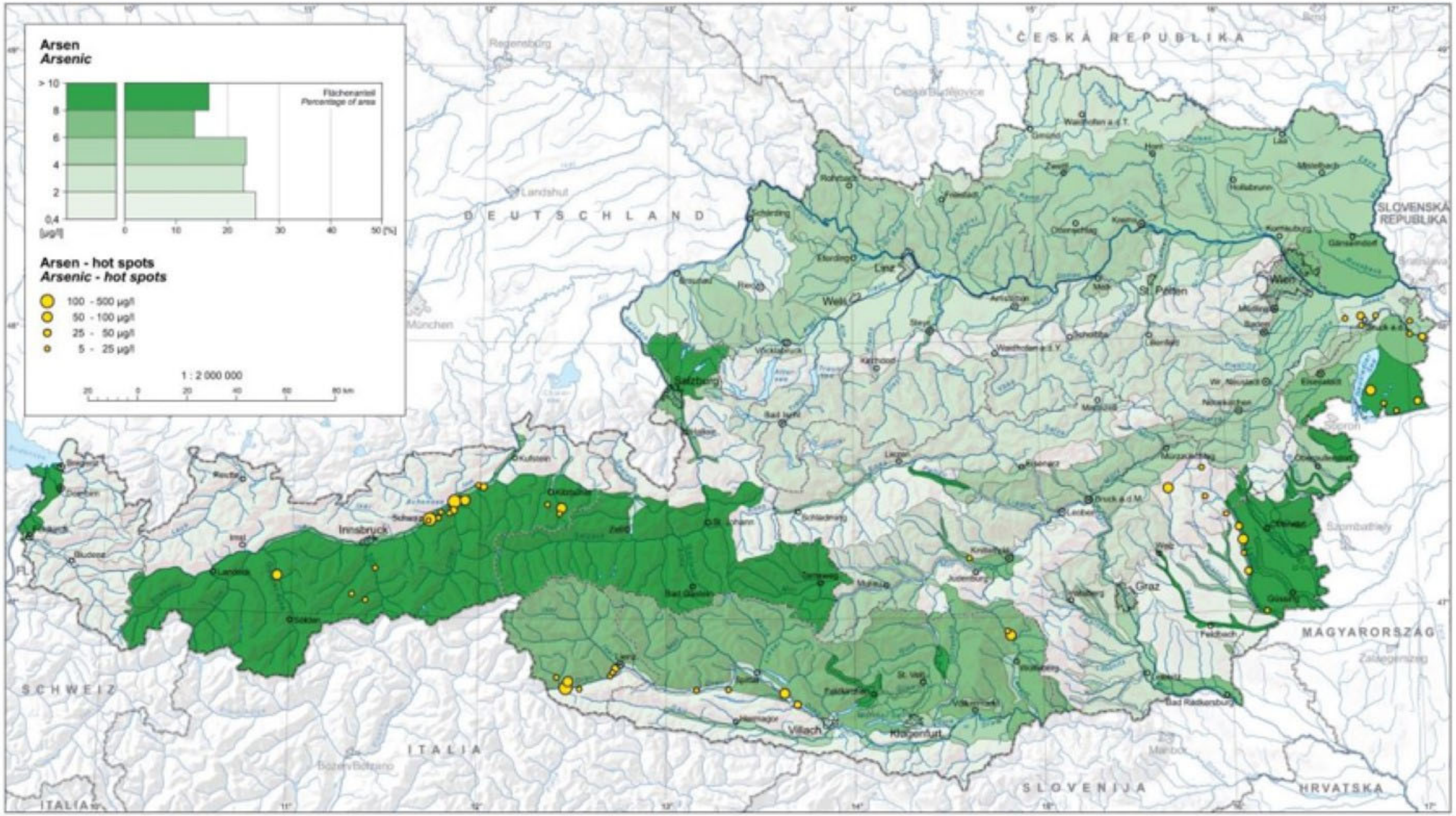


Figure VI. Surface ground deposition of caesium-137 released in the Chernobyl accident [I1, I3].

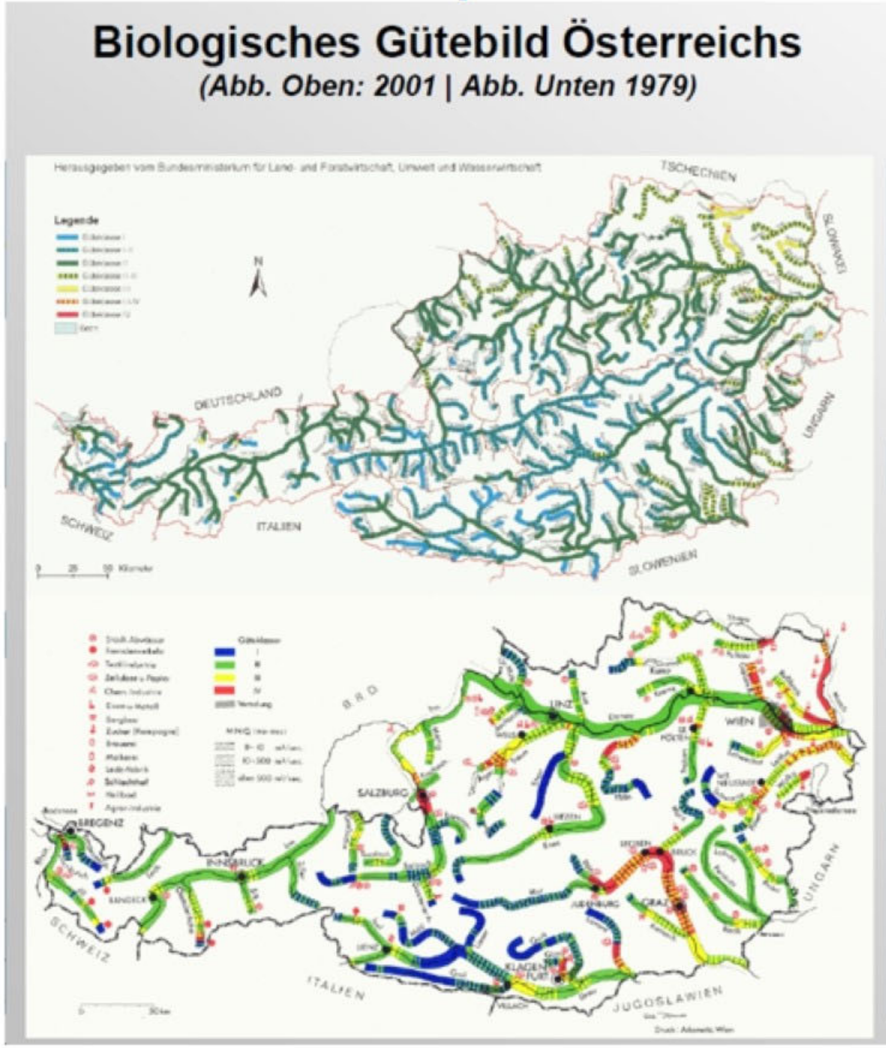
Hydrochemische geogene Hintergrundwerte
in oberflächennahen Grundwasserkörpern
*Hydrochemical geogenic background values
of near-surface groundwater bodies*

Wissenschaftliche Bearbeitung:
Scientific evaluation:
2004

Datengrundlagen:
Data basis:
2004



Examples: 1-D


Biologisches Gütebild Österreichs
(Abb. Oben: 2001 | Abb. Unten 1979)
TSCHECHIEN
DEUTSCHLAND
SCHWEIZ
ITALIEN
SLOWENIEN
UNGARN
SALZBURG
WIEN
LINZ
GRAZ
BREGENZ
SCHWEIZ
ITALIEN
JUGOSLAWIEN
UNGARN

Spatial data analysis

- Is scale dependent
- Quality depends on monitoring system
- Analysis is based on statistical models
- Analysis can be based on physical models

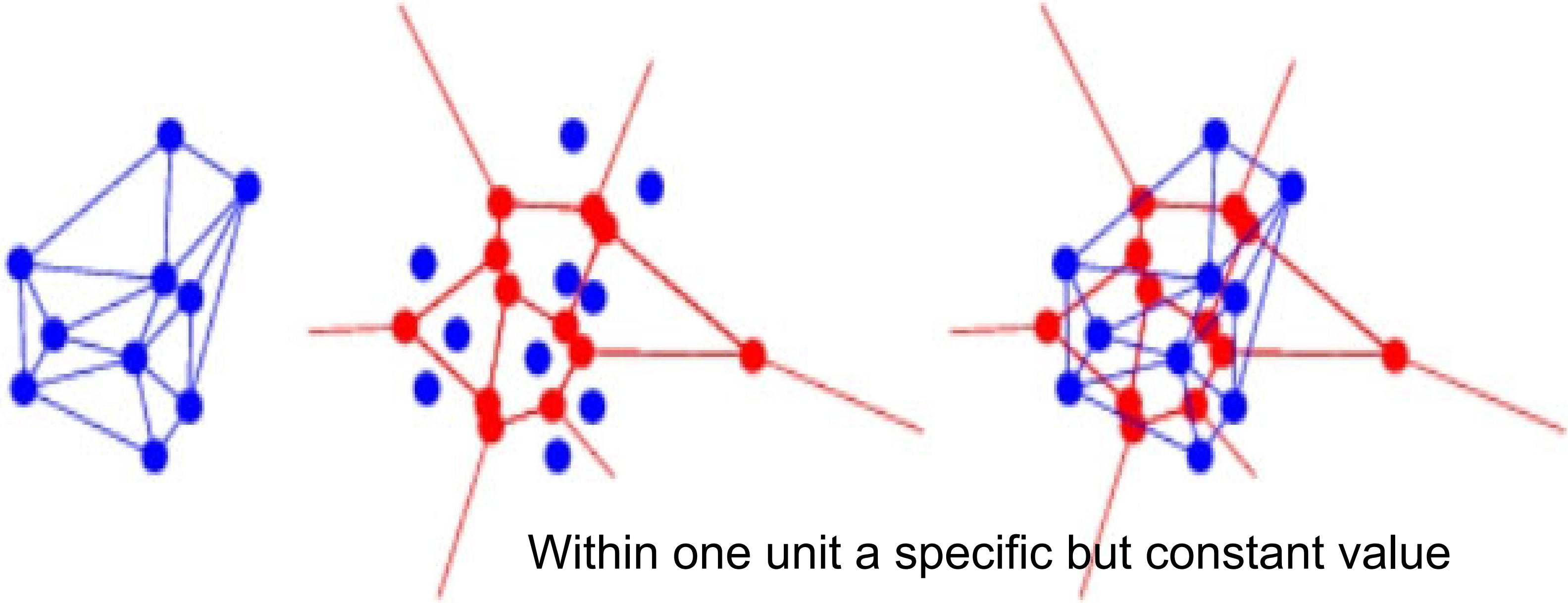
Introduction: Spatial data analysis

- Several concepts are applicable to both spatial and temporal data sets
- Some definitions:
  - **interpolation:** estimation of a value within a domain covered by observations
  - **extrapolation:** estimation of a value outside the domain
  - **regionalisation:** identifying properties within a domain (could be also achieved by transferring information from another region to the domain of interest)

Methods and concepts
(1) Thiessen or Voronoi diagrams

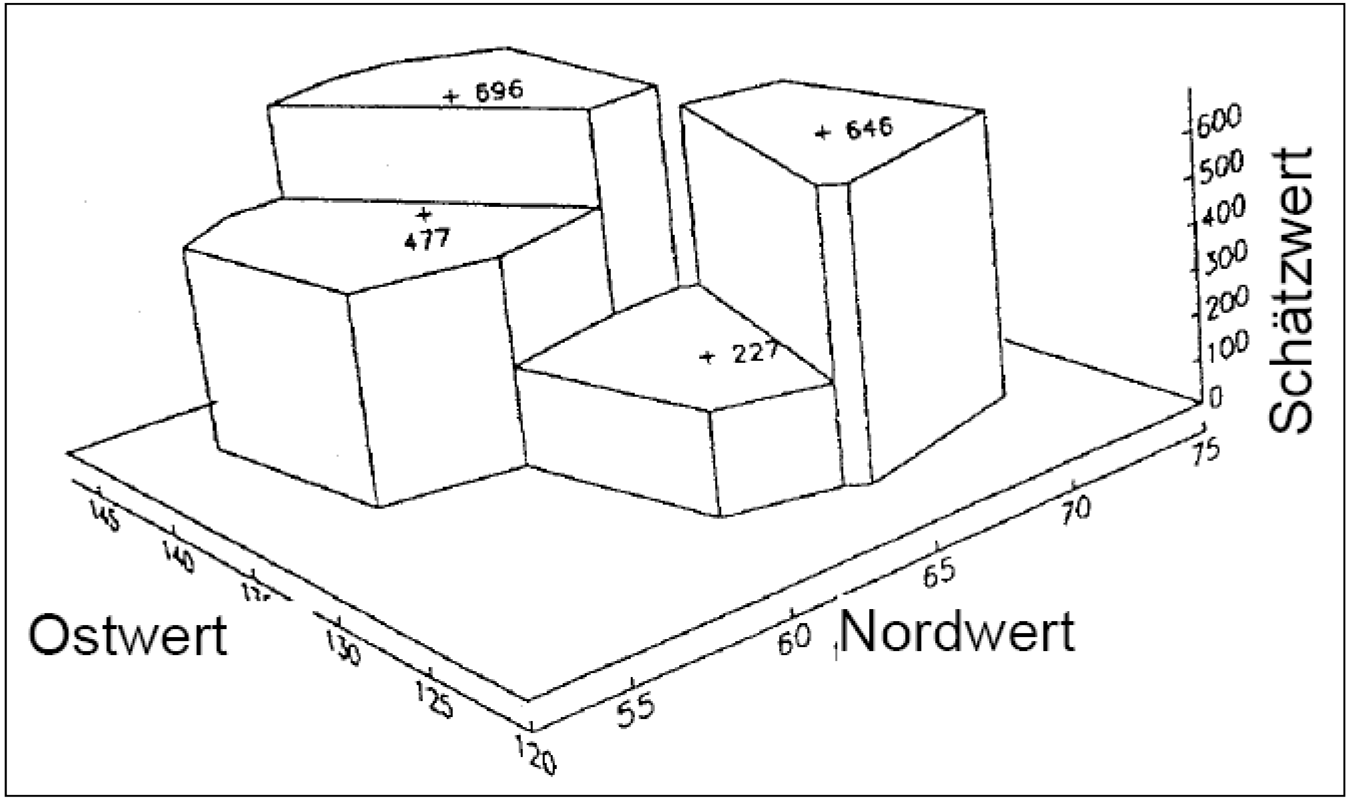
monitoring points

transects



Within one unit a specific but constant value

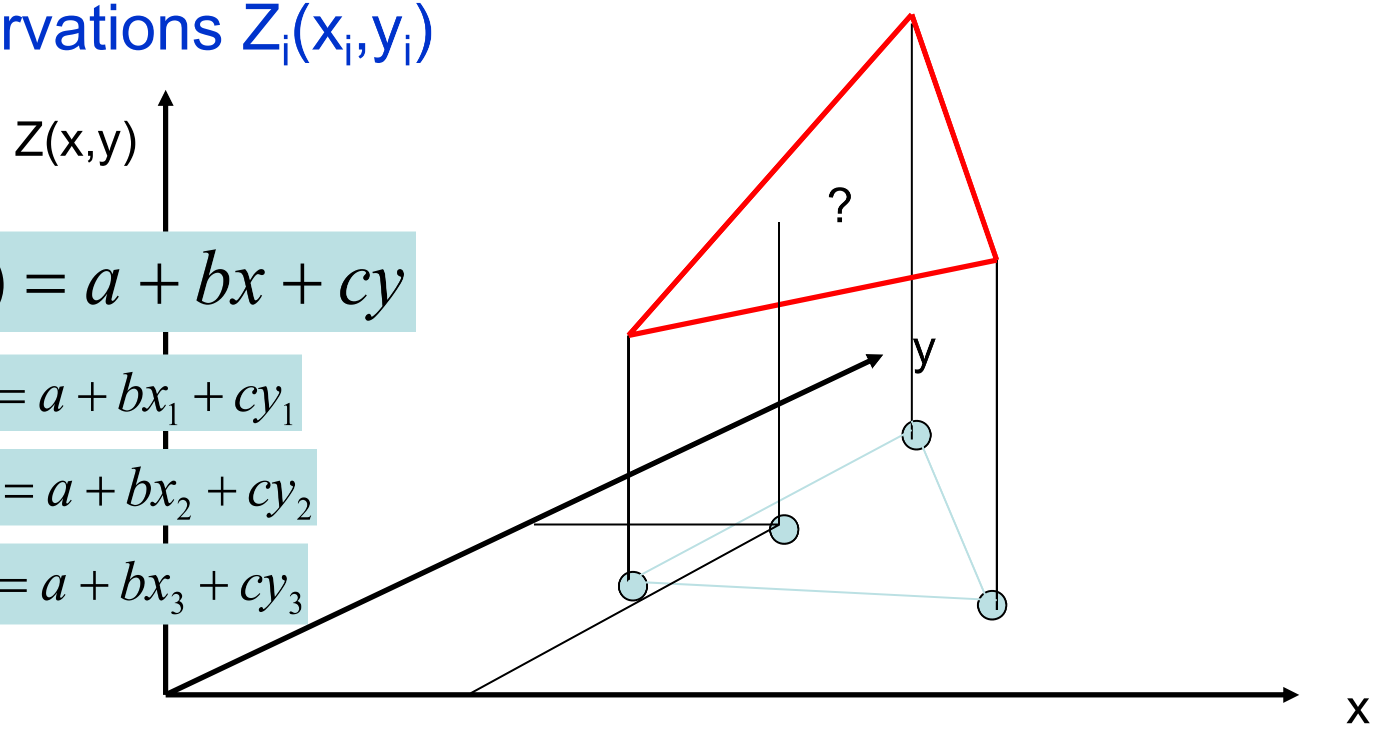
Thiessen or Voronoi diagram



Methods and concepts
(2) Linear interpolation

Monitoring points (x_i, y_i)

and observations $Z_i(x_i, y_i)$



$$Z(x,y) = a + bx + cy$$

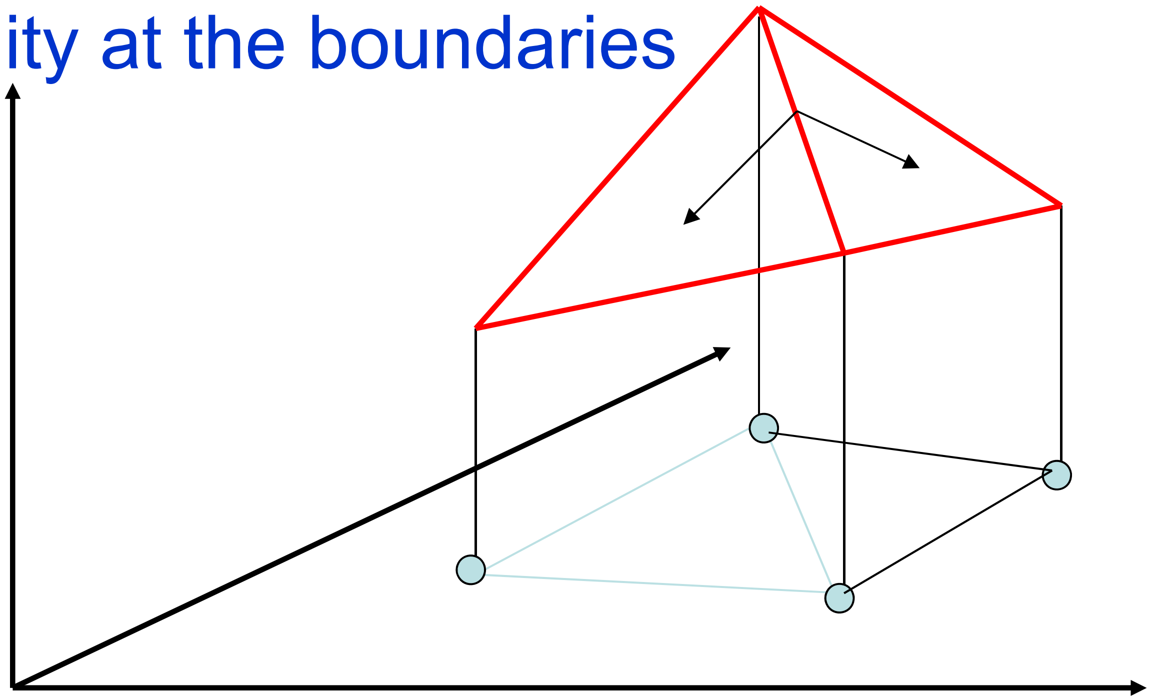
$$Z_1(x_1,y_1) = a + bx_1 + cy_1$$

$$Z_2(x_2,y_2) = a + bx_2 + cy_2$$

$$Z_3(x_3,y_3) = a + bx_3 + cy_3$$

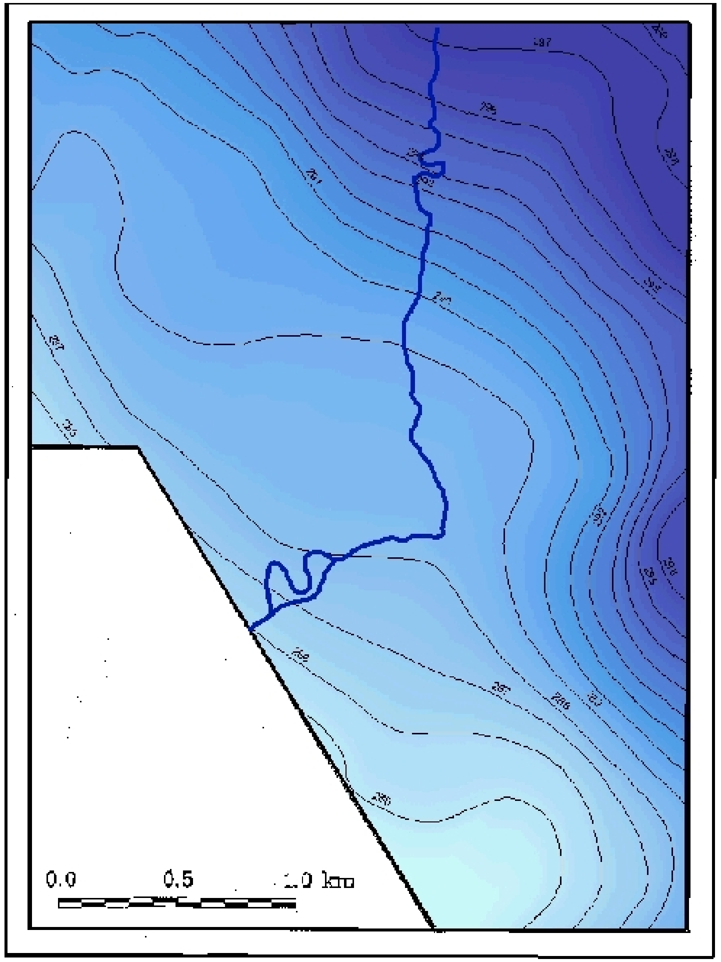
Methods and concepts
(2) Linear interpolation: Isolines

- Linear changes of Z within each element
- Discontinuity at the boundaries



Methods and concepts
(3) nonlinear interpolation

Isolines are generated by linear interpolation at several grid points and then the connecting line is smoothed



Methods and concepts
(3) nonlinear interpolation: Inverse distance

$$Z_k(x_k, y_k) = \sum w_j * Z_j(x_j, y_j)$$

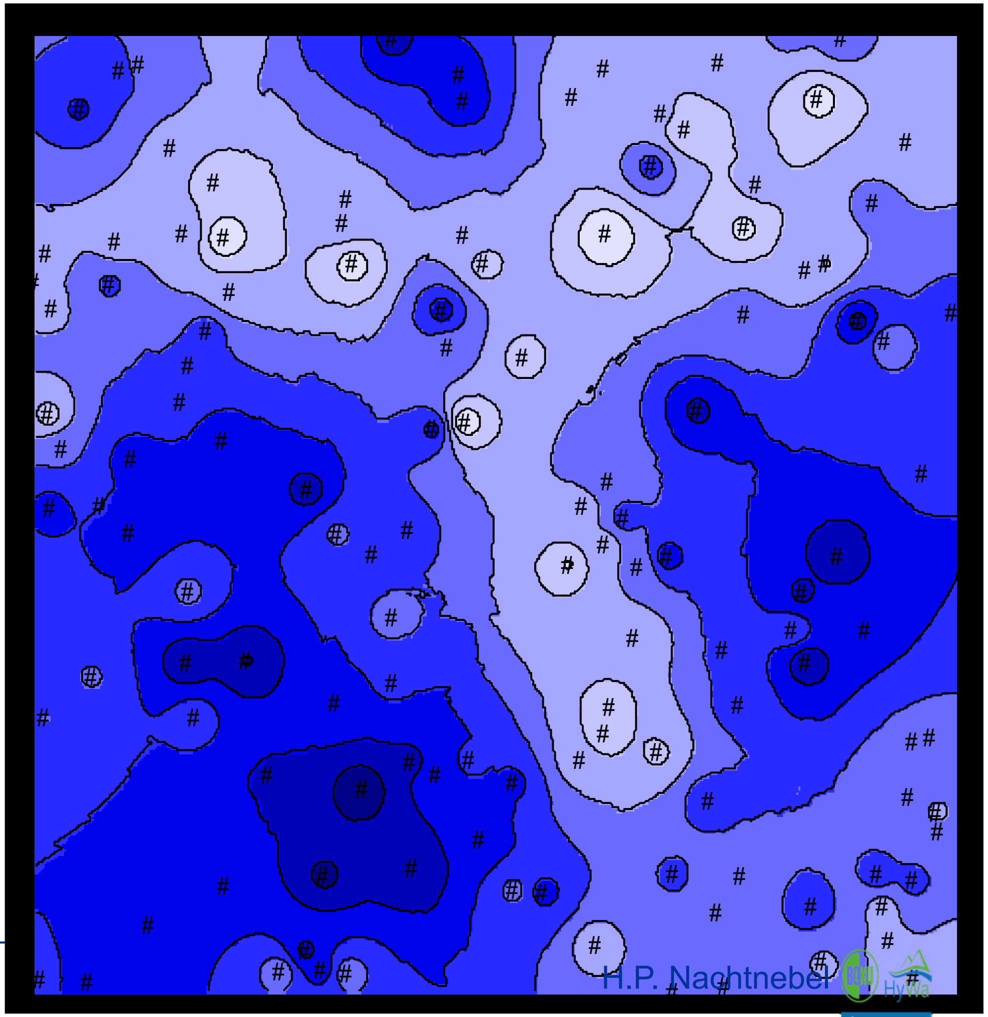
$$w_j = \frac{c}{d_{jk}{}^{\alpha}}$$

α may range from 2-4

$$\Sigma w_j = \Sigma \frac{c}{d_{jk}{}^{\alpha}} = c\Sigma \frac{1}{d_{jk}{}^{\alpha}} = 1$$

Inverse Distance Weighting (IDW)

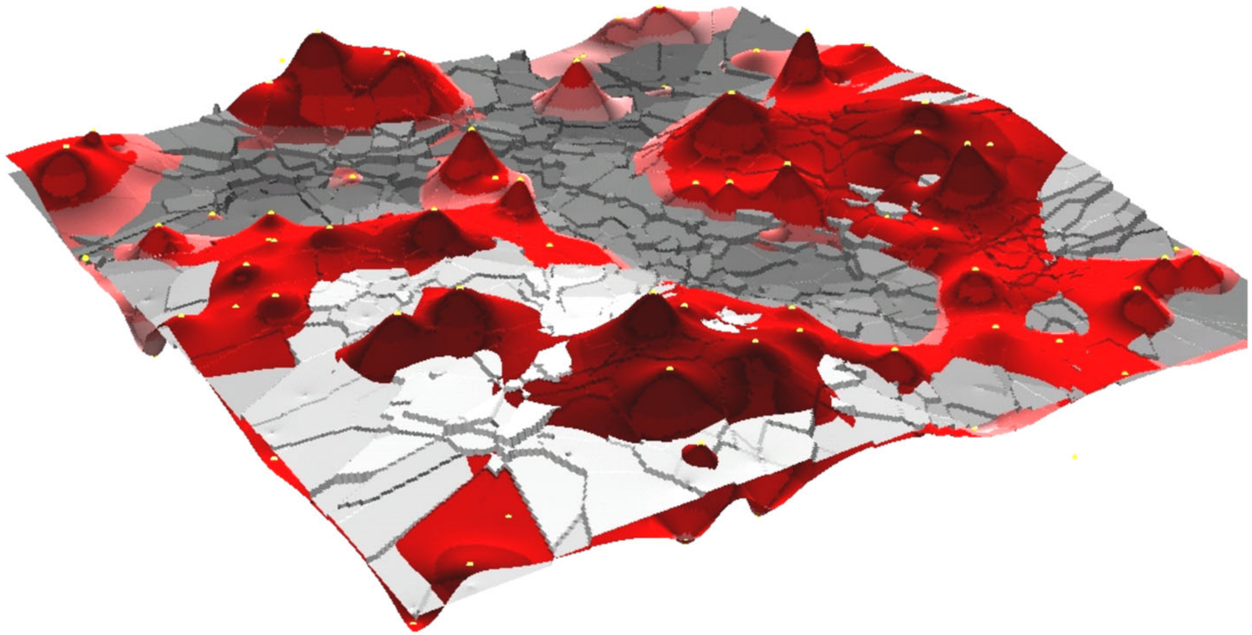
- Bull's eye effect $\alpha = 2$



Inverse Distance Weighting (IDW)

grey:
$\alpha = 0.1$

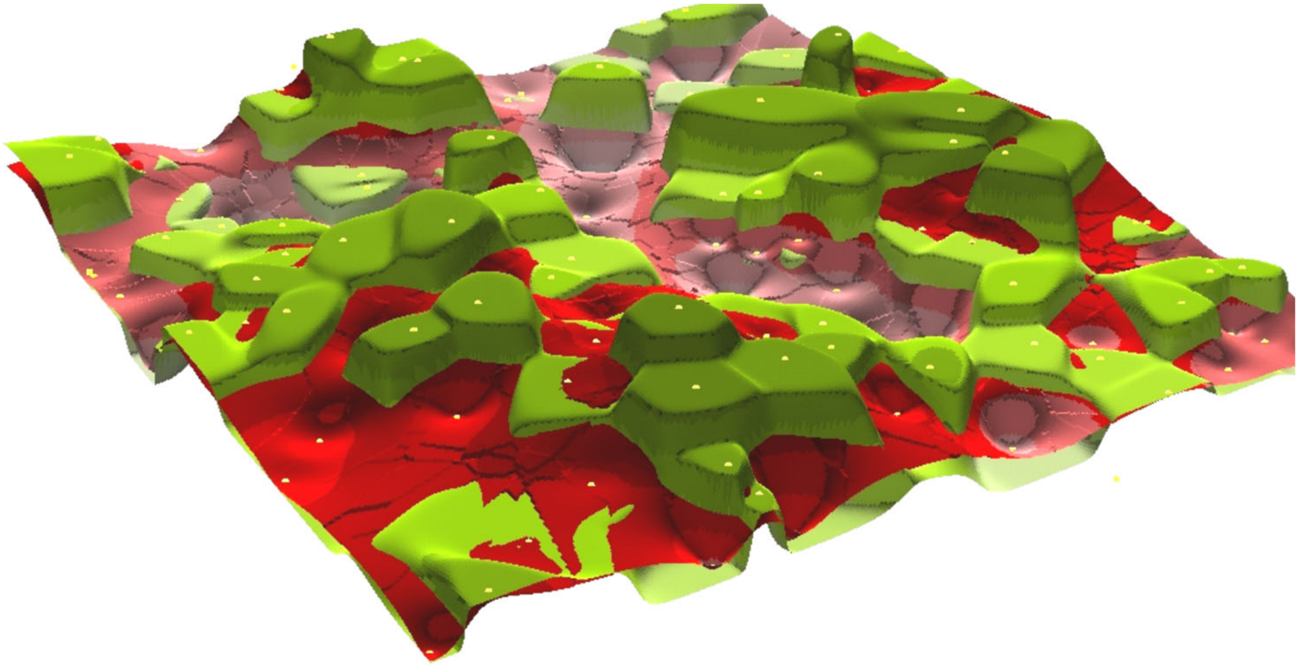
red:
$\alpha = 2$



Inverse Distance Weighting (IDW)

green:
$\alpha = 10$

red:
$\alpha = 2$



Methods and concepts
(3) Kriging

- Was developed by D. Krige (1951) in mining application in South Africa
- Methodological improvements by Matheron (1960) and subsequently applied by Delhomme (1978), Journel (1978)…..
- It provides an estimate and the estimation variance (uncertainty)
- Several extensions from Ordinary Kriging: indicator kriging, external drift kr., universal kr., co-kriging, fuzzy kriging,…

Methods and concepts
(3) Ordinary Kriging

- Basic assumptions: stationarity in space

  mean and co-variance are constant within a given domain

  what we observe in nature is the realisation of a random field

  complete information is included in the co-variance or variogram

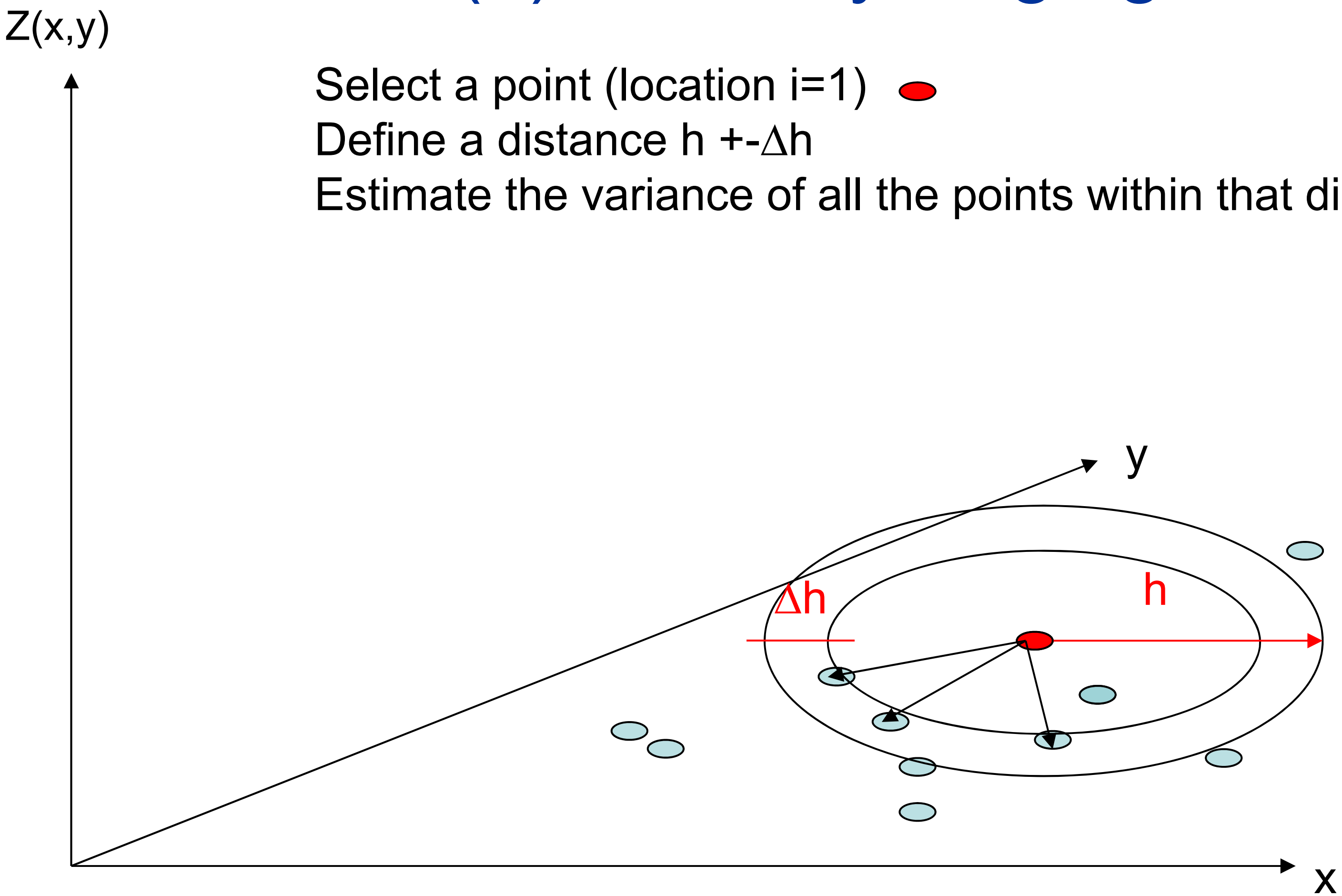
  Kriging is a BLUE estimator (Best Linear Unbiased Estimator)

Methods and concepts
(3) Ordinary Kriging

Select a point (location i=1)

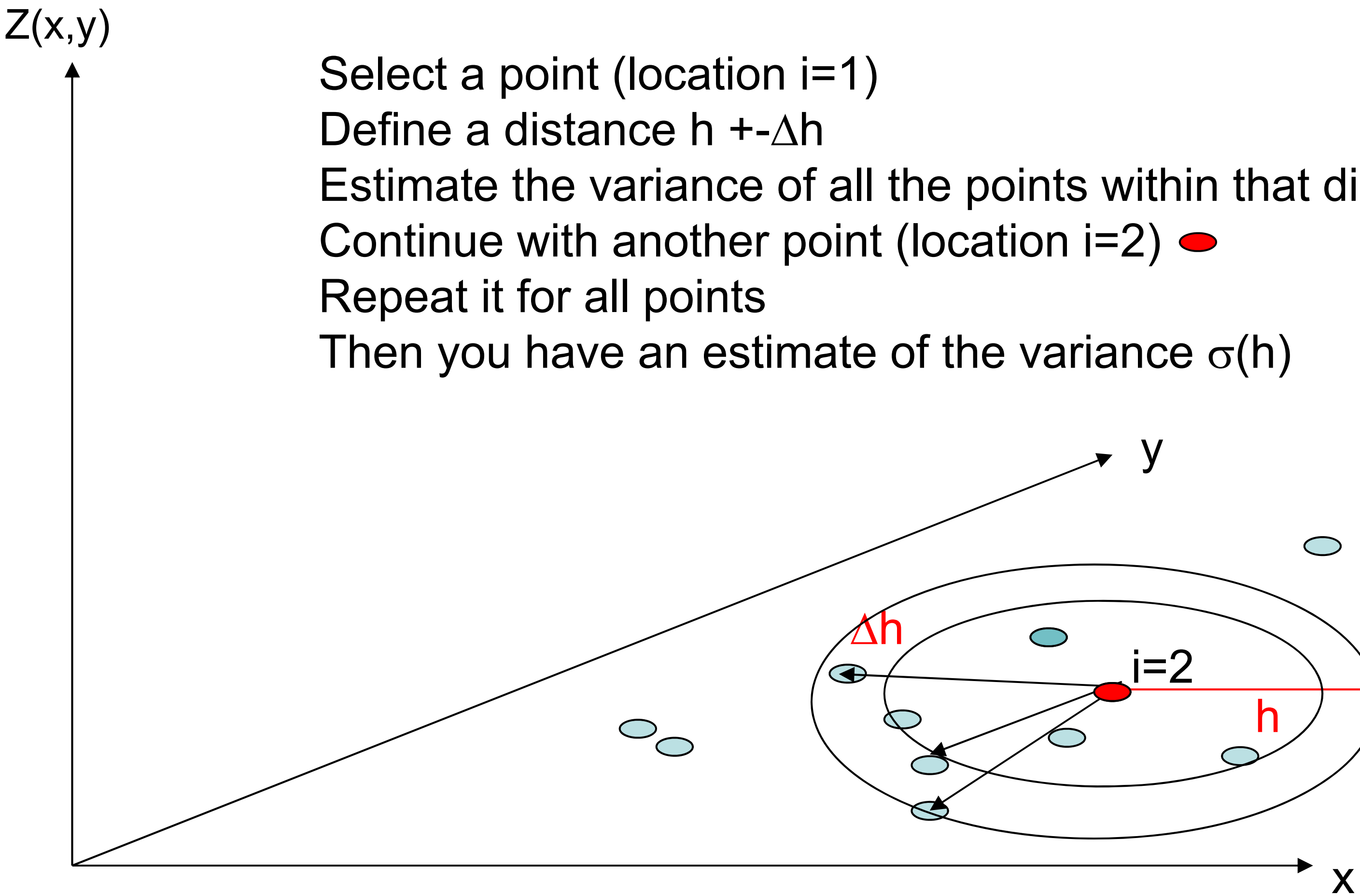
Define a distance h +-Δh

Estimate the variance of all the points within that distance



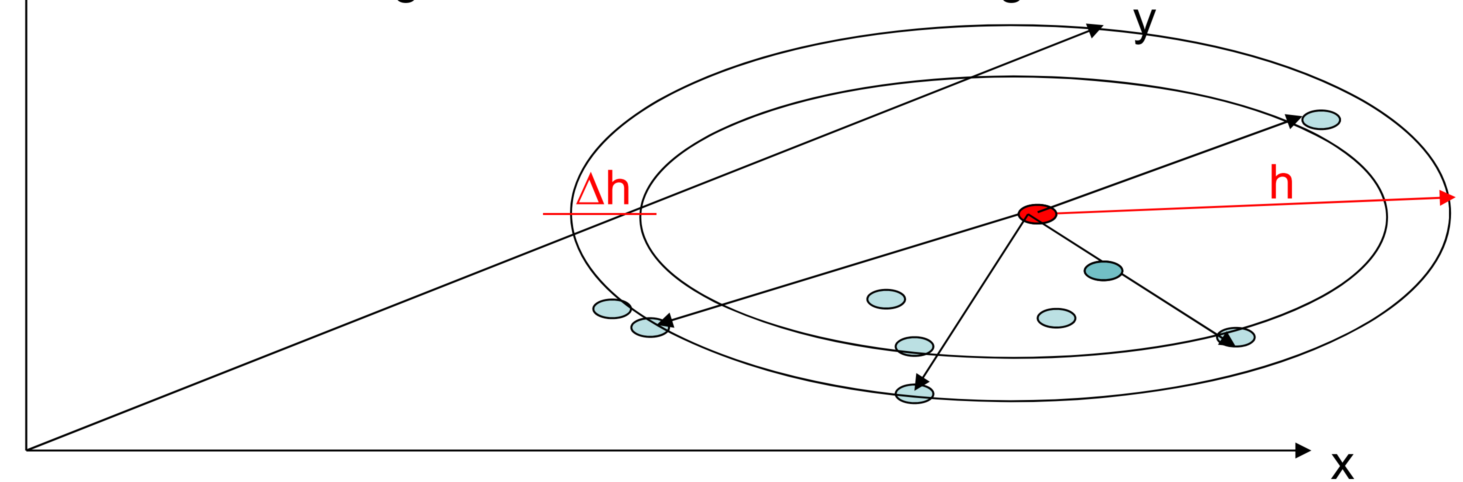
Methods and concepts
(3) Ordinary Kriging

Select a point (location i=1)
Define a distance h +-Δh
Estimate the variance of all the points within that distance
Continue with another point (location i=2)
Repeat it for all points
Then you have an estimate of the variance σ(h)



Methods and concepts
(3) Ordinary Kriging

Select a point (location i=1)
Define a distance h +-Δh
Estimate the variance of all the points within that distance
Continue with another point (location i=2)
Repeat it for all points
Then you have an estimate of the variance σ(h)
Enlarge the distance h and start again



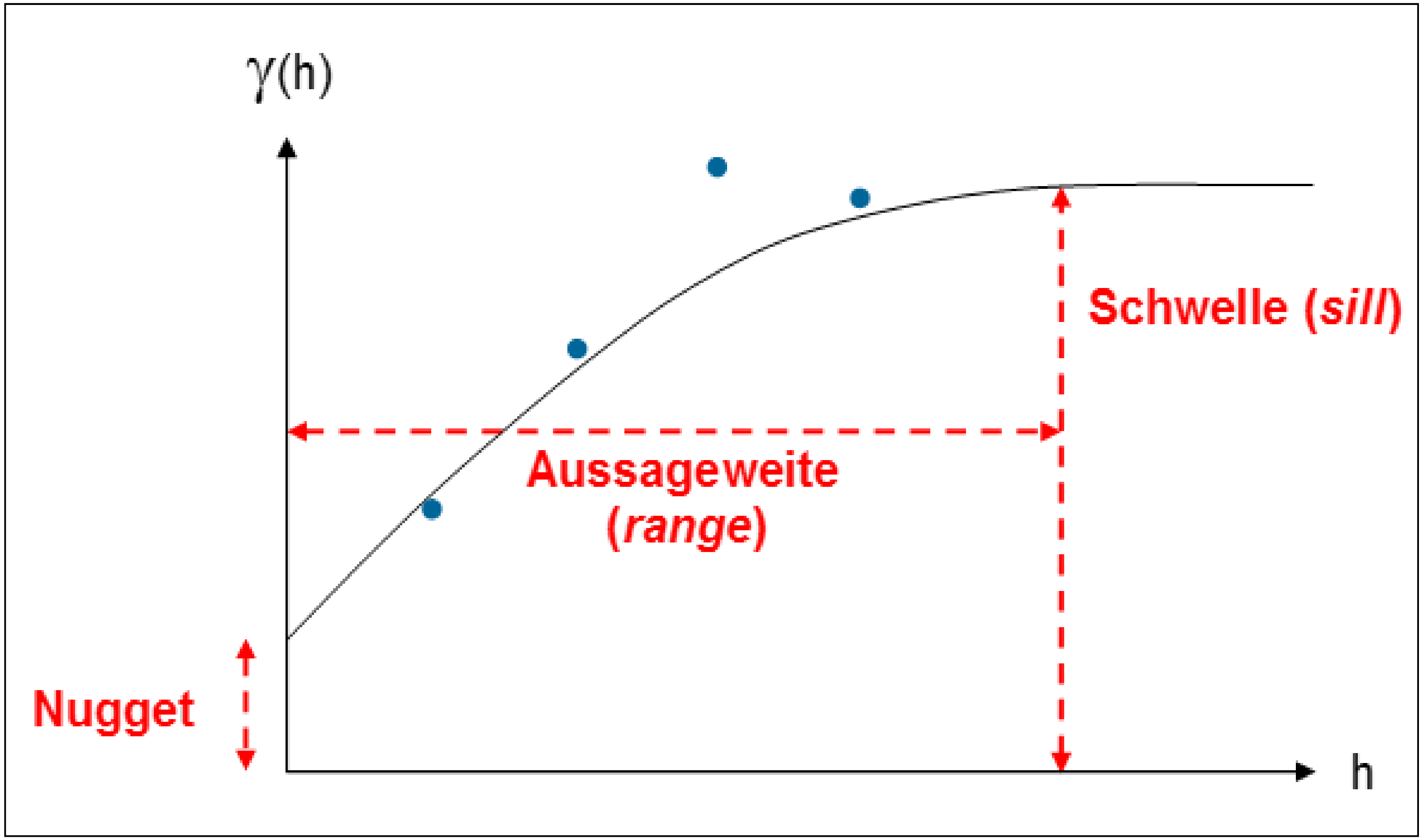
Estimation with Kriging

- Establish an empiral variogram

$$\frac{1}{2N(h)} \sum_{x_i - x_j = h} [Z(x_i) - Z(x_j)]^2 = \gamma^*(h)$$

- Fit a theoretical variogram

Methods and concepts
(3)linear interpolation: Ordinary Kriging



Estimation and its variance

$$Z(x) = \sum_{i=1}^{n} \lambda_i \cdot Z(x_i)$$

$$\sum_{j=1}^{n} \lambda_j \cdot \gamma(x_i - x_j) + \mu = \gamma(x_i - x) \qquad i = 1,\ldots,n \quad \text{(system of equations)}$$

$$\sum_{j=1}^{n} \lambda_j = 1$$

$$\sigma^2(x) = -\sum_{j=1}^{n} \sum_{i=1}^{n} \lambda_i \cdot \lambda_j \cdot \gamma(x_i - x_j) + 2 \cdot \sum_{i=1}^{n} \lambda_i \cdot \gamma(x_i - x)$$

Simple example

- 2 observation: $Z_1(x_1=1)=2$, $Z_2(x_2=-2)=4$
- A linear variogram $\gamma(h)=|h|$
- Estimate $Z(x=0)$ and the estimation variance

$$0\lambda_1 + 3\lambda_2 + \mu = 1$$
$$3\lambda_1 + 0\lambda_2 + \mu = 2$$
$$\lambda_1 + \lambda_2 = 1$$

$$\sum_{j=1}^{n} \lambda_j \cdot \gamma(x_i - x_j) + \mu = \gamma(x_i - x)$$

$\lambda_1 = 0{,}6667$, $\lambda_2 = 0{,}3333$ $\mu = 0$

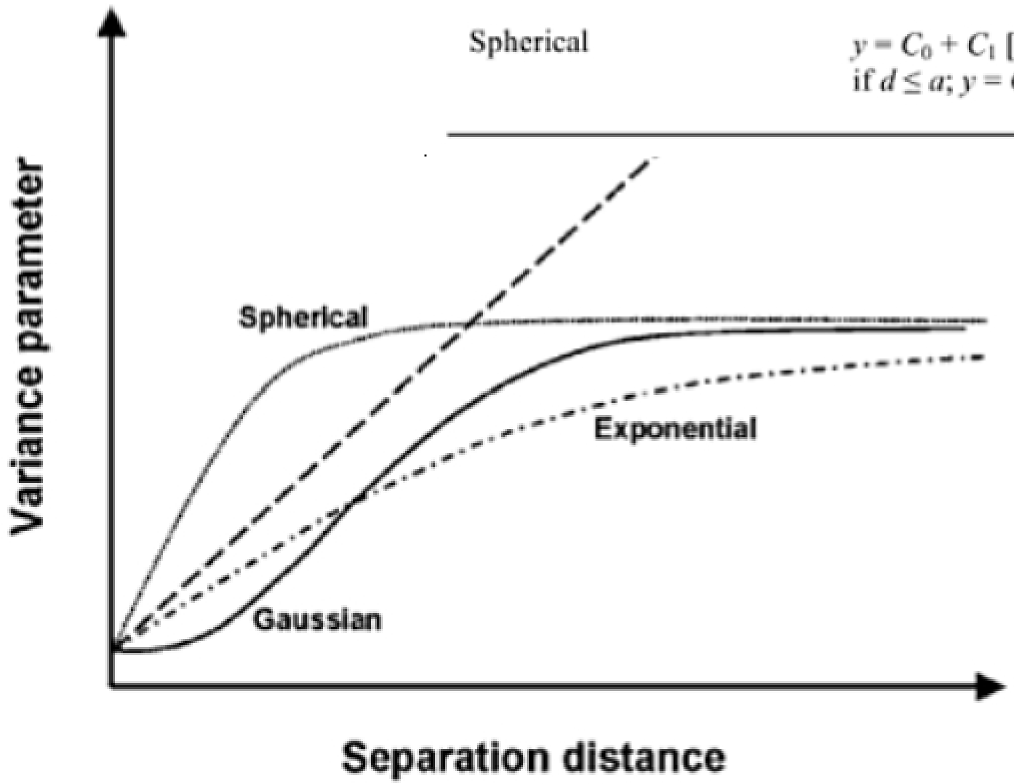
$$Z(x) = \sum_{i=1}^{n} \lambda_i \cdot Z(x_i)$$

$Z^*(x=0) = 2{,}6667$ $\sigma^2 = 1.3333$

$$\sigma^2(x) = -\sum_{j=1}^{n} \sum_{i=1}^{n} \lambda_i \cdot \lambda_j \cdot \gamma(x_i - x_j) + 2 \cdot \sum_{i=1}^{n} \lambda_i \cdot \gamma(x_i - x)$$

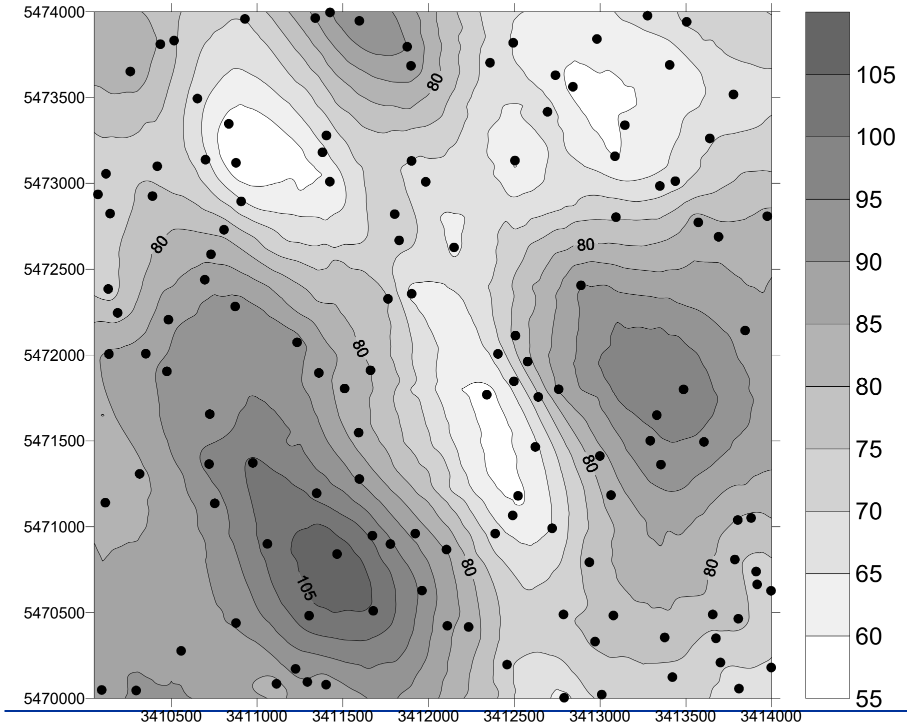
Table 2-3. Summary of model types, plots of y (variance parameter) versus d (distance).

Model type	Equation
Nugget	$y = C_0$
Linear	$y = C_0 + bd$ where b is the slope and C_0 is the intercept
Exponential	$y = C_0 + C_1\,[1 - \exp(-3\ d/a)]$
Gaussian	$y = C_0 + C_1\,[1 - \exp(-3\ d^2/a^2)]$
Spherical	$y = C_0 + C_1\,[1.5\ d/a - 0.5\ (d/a)^3]$ if $d \leq a$; $y = C$ if $d > a$

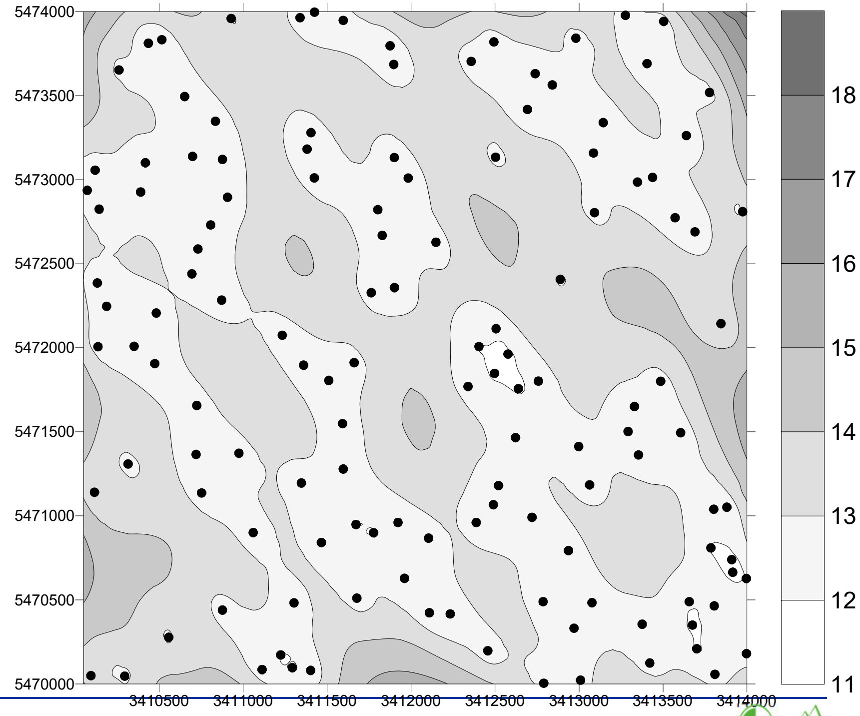


Applications

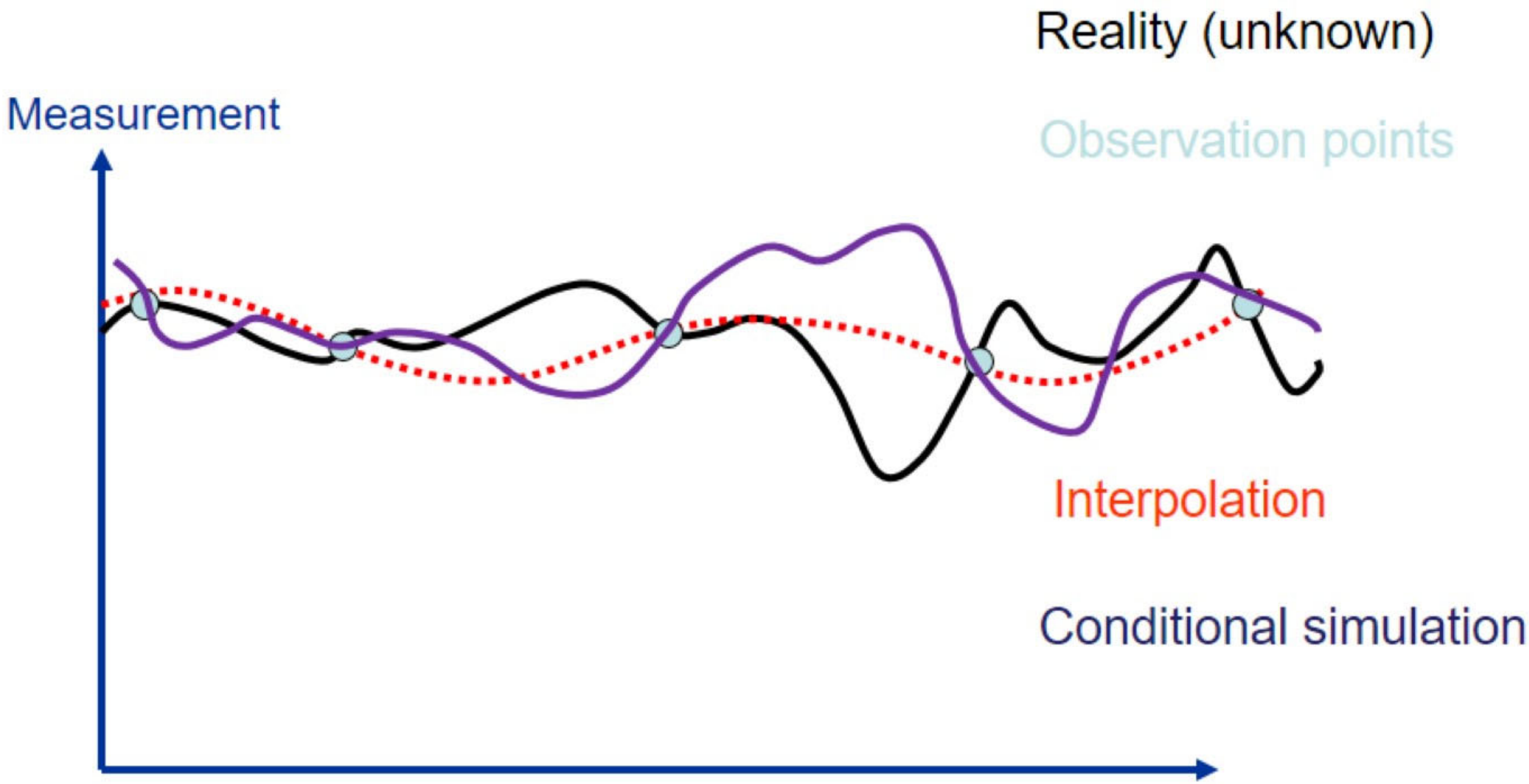
Estimated conductivity



Standard deviation of estimated conductivity



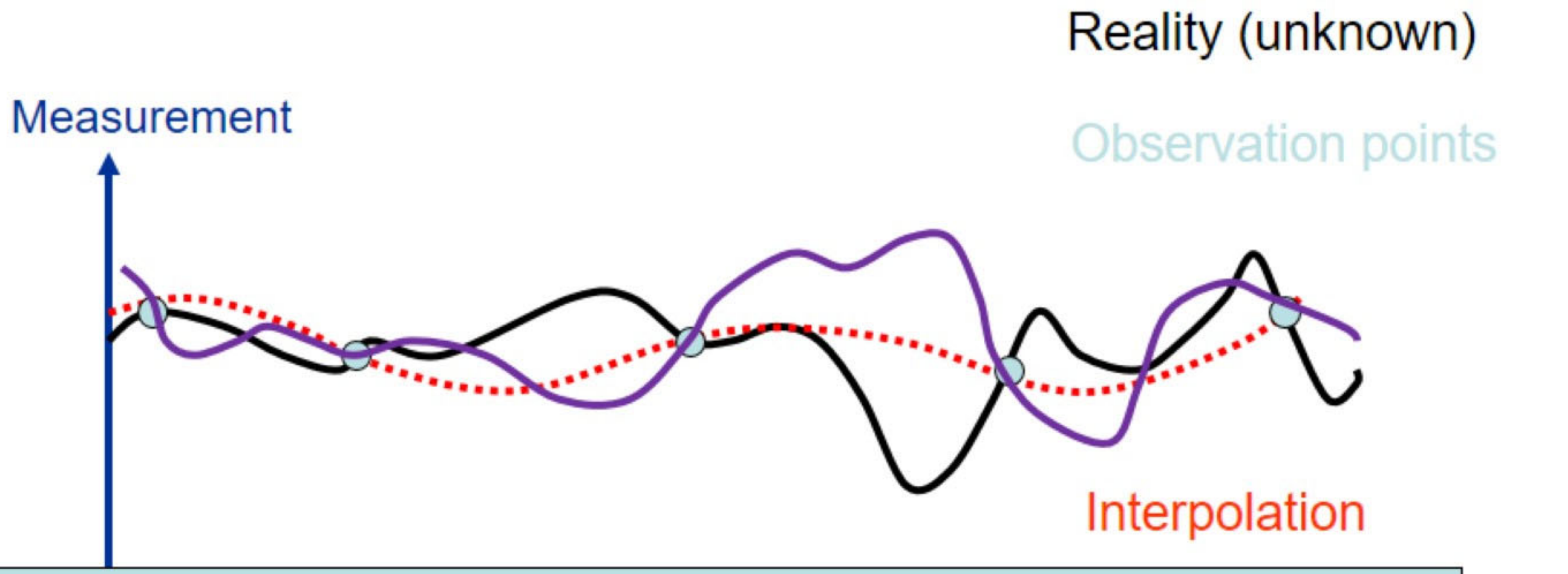
Uncertainty in model parameters



Observation and interpolation

- Observed values have certain spatial variance
- Interpolated values are smoother (they are expected values)
- To simulate a random field (in our case trajectory) conditional simulation is required to generate a field with same variance as in reality

Uncertainty in model parameters



We assume that measurements are perfect and no nugget effect

Summary and conclusions

- Several interpolation/extrapolation techniques were presented
- Linear and nonlinear
- Deterministic and stochastic
- Kriging is a BLUE estimator
  it provides an estimate and the estimation variance
- Can be used to develop a monitoring system
- Can be used to simulate spatial structures