An Introduction to Geostatistics

András Bárdossy

Universität Stuttgart

Institut für Wasser- und Umweltsystemmodellierung
Lehrstuhl für Hydrologie und Geohydrologie
Prof. Dr. rer. nat. Dr.-Ing. András Bárdossy
Pfaffenwaldring 61, 70569 Stuttgart, Deutschland www.iws.uni-stuttgart.de

Introduction

- **This is not:**
  - An introduction to a software package
  - Nor a set of recipes how to …

- **It is an introduction to**
  - Why we use geostatistics
  - What is important what is not
  - What is good and what is not

Books

An Introduction to Applied Geostatistics

(Isaaks and Srivastava)

Mining Geostatistics

(Journel and Huijbregts)

Geostatistics for natural resources evaluation

(Goovaerts)

Software

SgeMS

(Stanford Geostatistical Modelling Software)

GEOEAS

(Geostatistical **Environmental Assessment** ***Software*****)**

GSLIB

(Geostatistical Software Library)

ArcGIS **Geostatistical Analyst**

God does not play dice. (Albert Einstein)

Quantum mechanics

Draw of lottery numbers ?

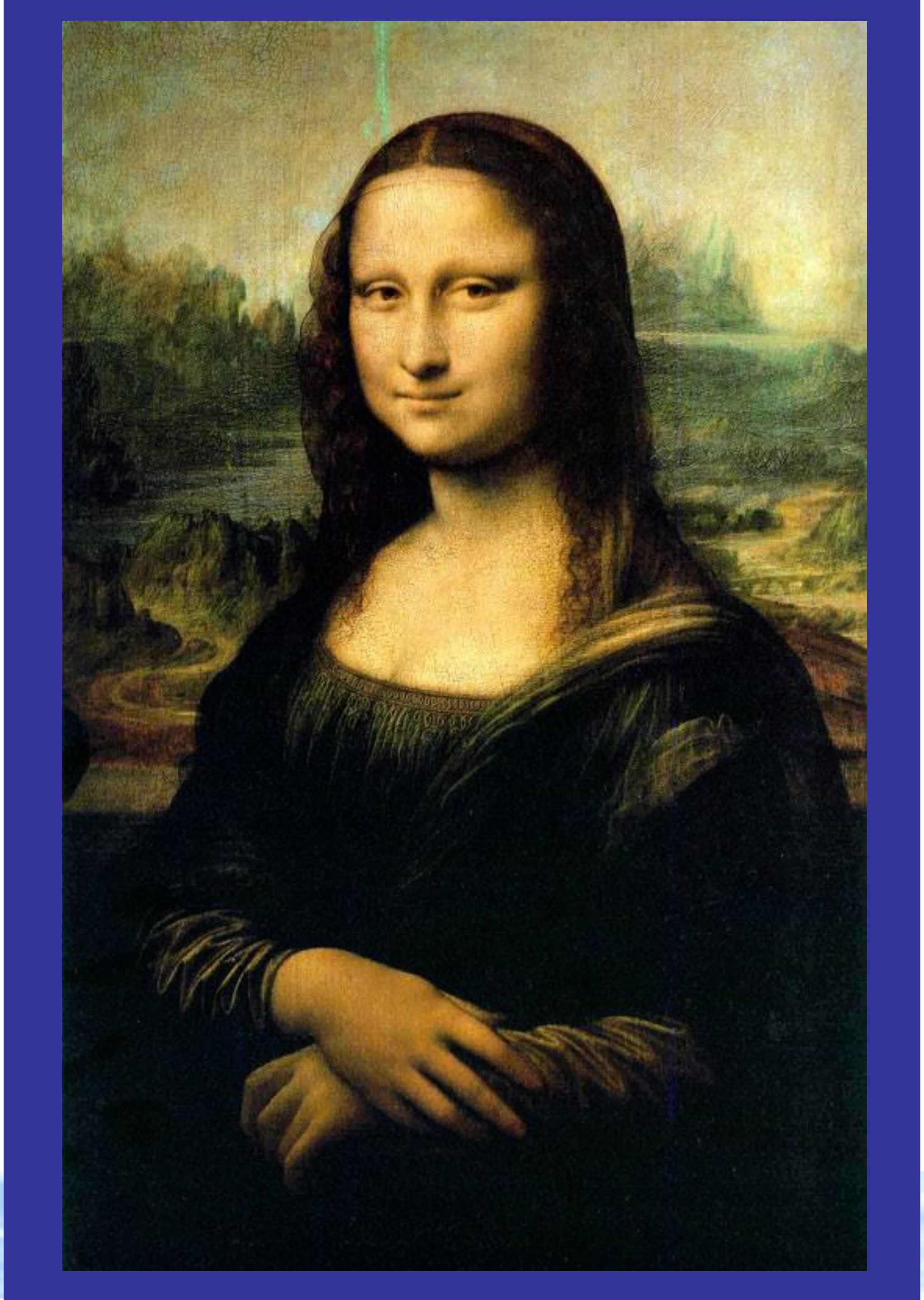
Our case – we do not even know the exact circumstances of the processes.

The problem

- **Discrete observations (points and blocks)**
- **Unknown in between**

- **Generating processes**
  - Physical, chemical, biological
  - Circumstances, inputs … unknown

- **Uncertainty – assumptions related to uncertainty**



The problem

- **We do not know reality**
- **But:**
  - We can use methods of statistics
  - Deriving certain measures from observations
  - Applying a stochastic “analogue”

- **We assume that what is observed is like the realization of a stochastic process**

Mixture of structure and randomness

Intrinsic hypothesis

1. **The expected value of the random function Z(u) is constant all over the domain D**
2. **The variance of the increment corresponding to two different locations depends only on the vector separating them**

Intrinsic hypothesis

- **These conditions can be formulated as:**

$$E\left[Z(u)\right]=m \qquad \text{for all } u\in D$$

- **And for the increments**

$$\frac{1}{2}Var\left[Z(u+h)-Z(u)\right]=$$

$$=\frac{1}{2}E\left[\left(Z(u+h)-Z(u)\right)^2\right]=\gamma(h)$$


BETHLEHEM
MANGER

The variogram

1. $\gamma(0)=0$
2. $\gamma(h) > 0$, for all vectors h
3. $\gamma(h) = \gamma(-h)$, for all vectors h
4. variance of the increments is supposed to increase with the length of the vector h
5. limit in the continuity of the parameter, vector separating two points exceeds a certain limit the variance of the increment will not increase any more
6. The variogram is often discontinuous near the origin. For any $h>0$ we have $\gamma(h)>C_0>0$, nugget effect.

Experimental variogram

- **Variogram can be estimated with the help of the following formula**

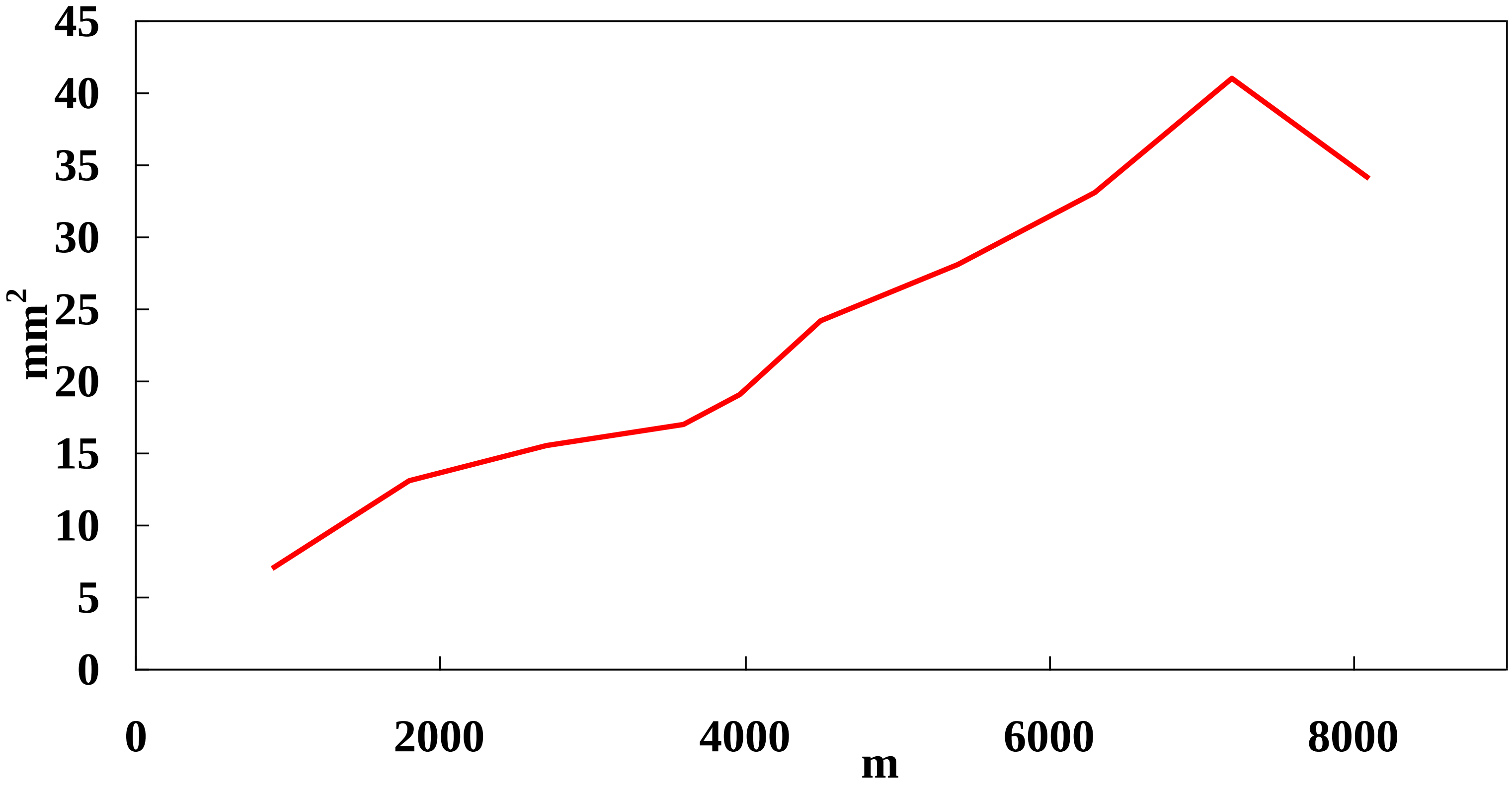
$$\gamma^*(h) = \frac{1}{2N(h)} \sum_{u_i - u_j = h} \left(Z(u_i) - Z(u_j) \right)^2$$

Allowing a certain difference in both the angle and the length of vector

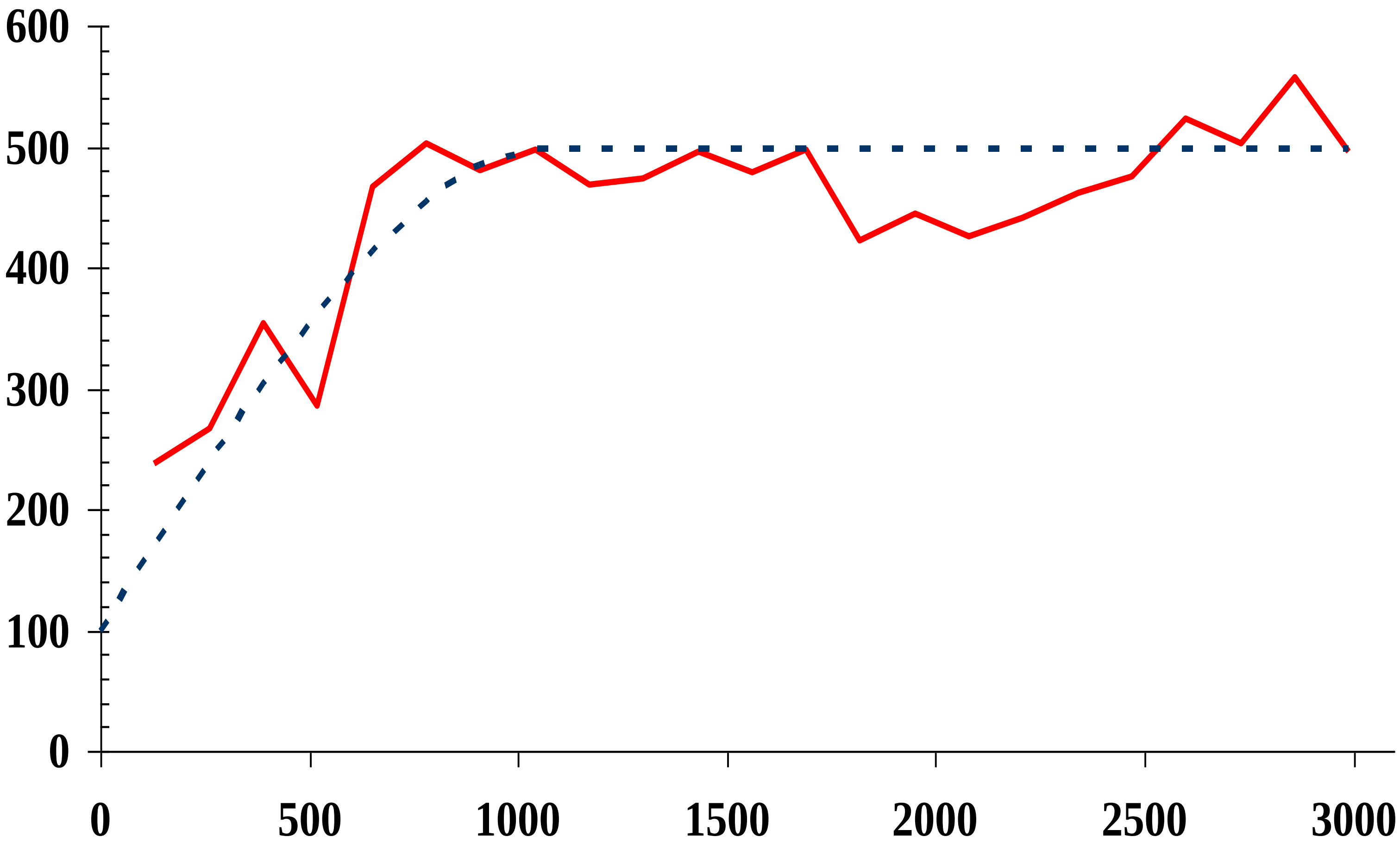
$$\left| \left| u_i - u_j \right| - |h| \right| \leq \varepsilon$$

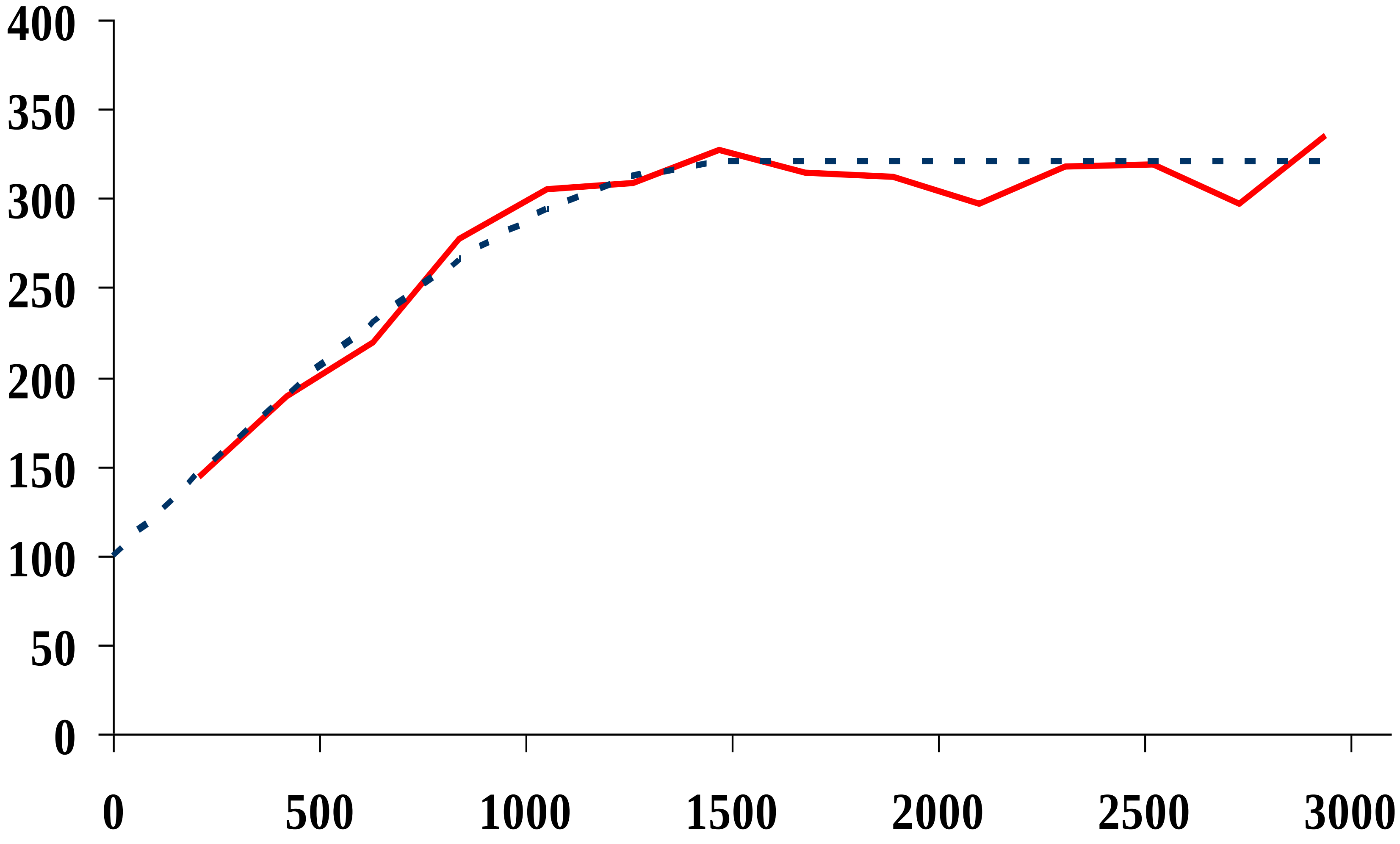
$$Angle\left(u_i - u_j, h\right) \leq \delta$$

Experimental variogram



Estimation variogram Cl




400
350
300
250
200
150
100
50
0
0
500
1000
1500
2000
2500
3000

Point kriging

Unbiasedness

$$Z^*(u) = \sum_{i=i}^{n} \lambda_i Z(u_i)$$

$$E[Z(u)] = m \quad for\, all \; u \in D$$

$$E[Z^*(u)] = \sum_{i=i}^{n} \lambda_i E[Z(u_i)] = m$$

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

Estimation variance using the variogram

- **Minimize estimation variance:**

$$\sigma^2(u) = Var\left[Z(u) - Z^*(u)\right] =$$

$$= -\sum_{j=1}^{n}\sum_{i=1}^{n}\lambda_i\lambda_j\gamma\left(u_i - u_j\right) + 2\sum_{i=1}^{n}\lambda_i\gamma\left(u_i - u\right)$$

Kriging equations using variogram

$$\sum_{j=1}^{n} \lambda_j \gamma\left(u_i - u_j\right) + \mu = \gamma\left(u_i - u\right) \qquad i=1,\ldots,n$$

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

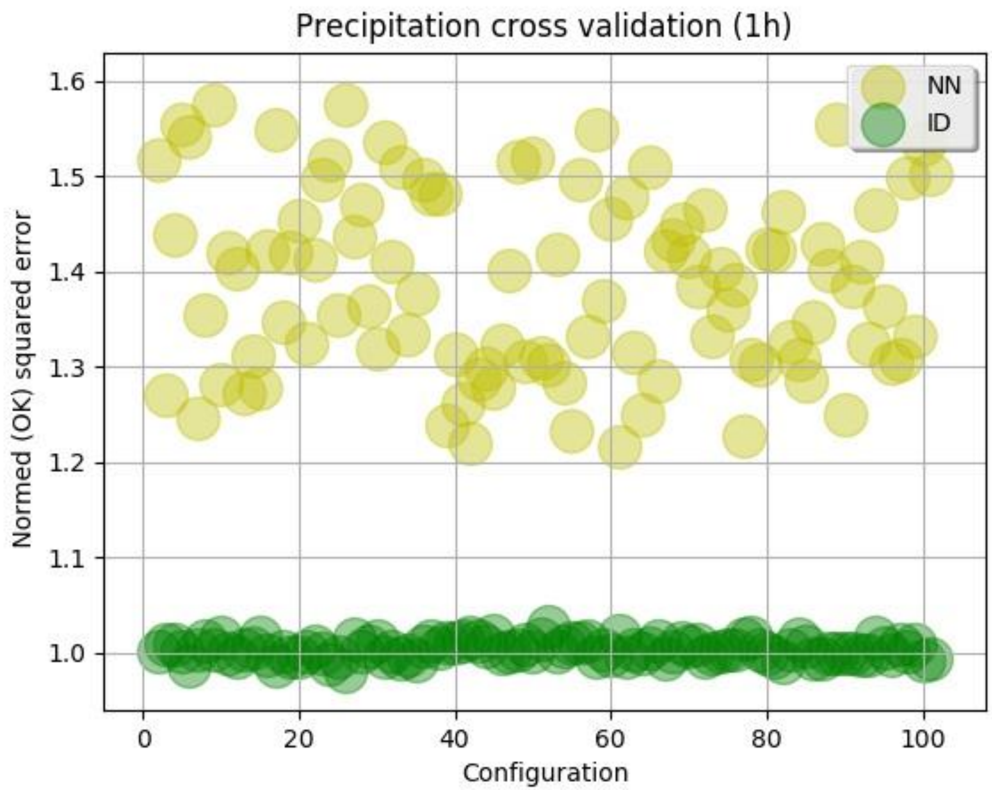
$$\boxed{\sigma_K^2(u) = \sum_{j=1}^{n} \lambda_j \gamma\left(u_i - u_j\right) + \mu}$$

This is an analogy

- **Is it better than others? (NN,ID)**
  - Rational:
    - Distinguishes between variables (variograms)
    - Good properties
      - Data configurations are reflected
  - Testing:
    - Estimates using cross validation:
      - Leave some out and estimate them using the rest
      - Compare estimated and observed
    - Uncertainty using confidence intervals (CI)
      - Leave some out and estimate them using the rest
      - Where is the observed in the estimated CI

Precipitation cross validation

Normed squared error for unused stations for each method and different time aggregations:



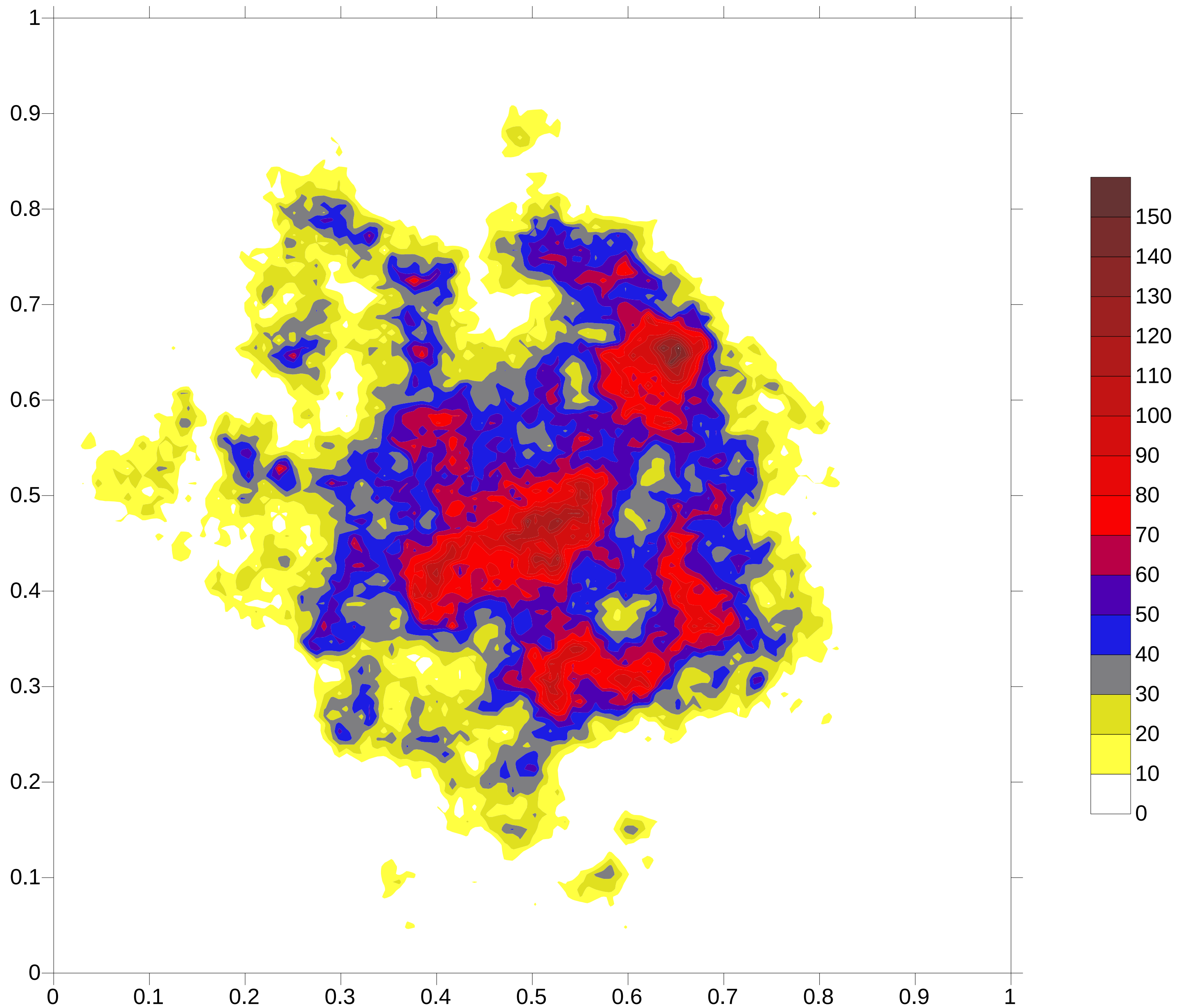
Is good also true?

- **Interpolation is a good estimator**
  - Is the obtained map the truth?
    - **NO NO NO NO NO ….**
    - It is the “best” estimate but
    - Impossible as it has different properties as observations
      - Smooth – lower variance and variogram
      - Consequence → Problems for risk assessment

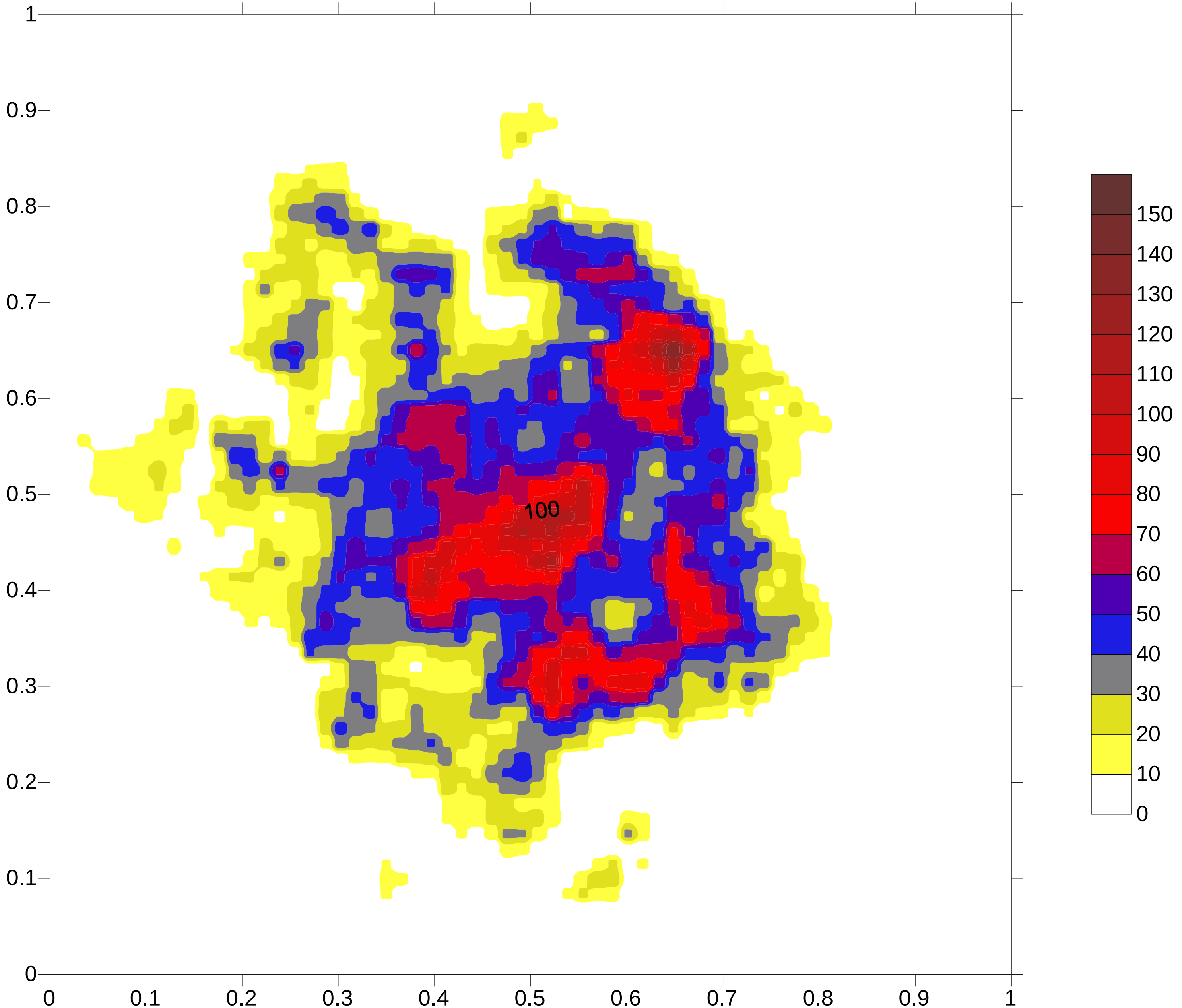
Support

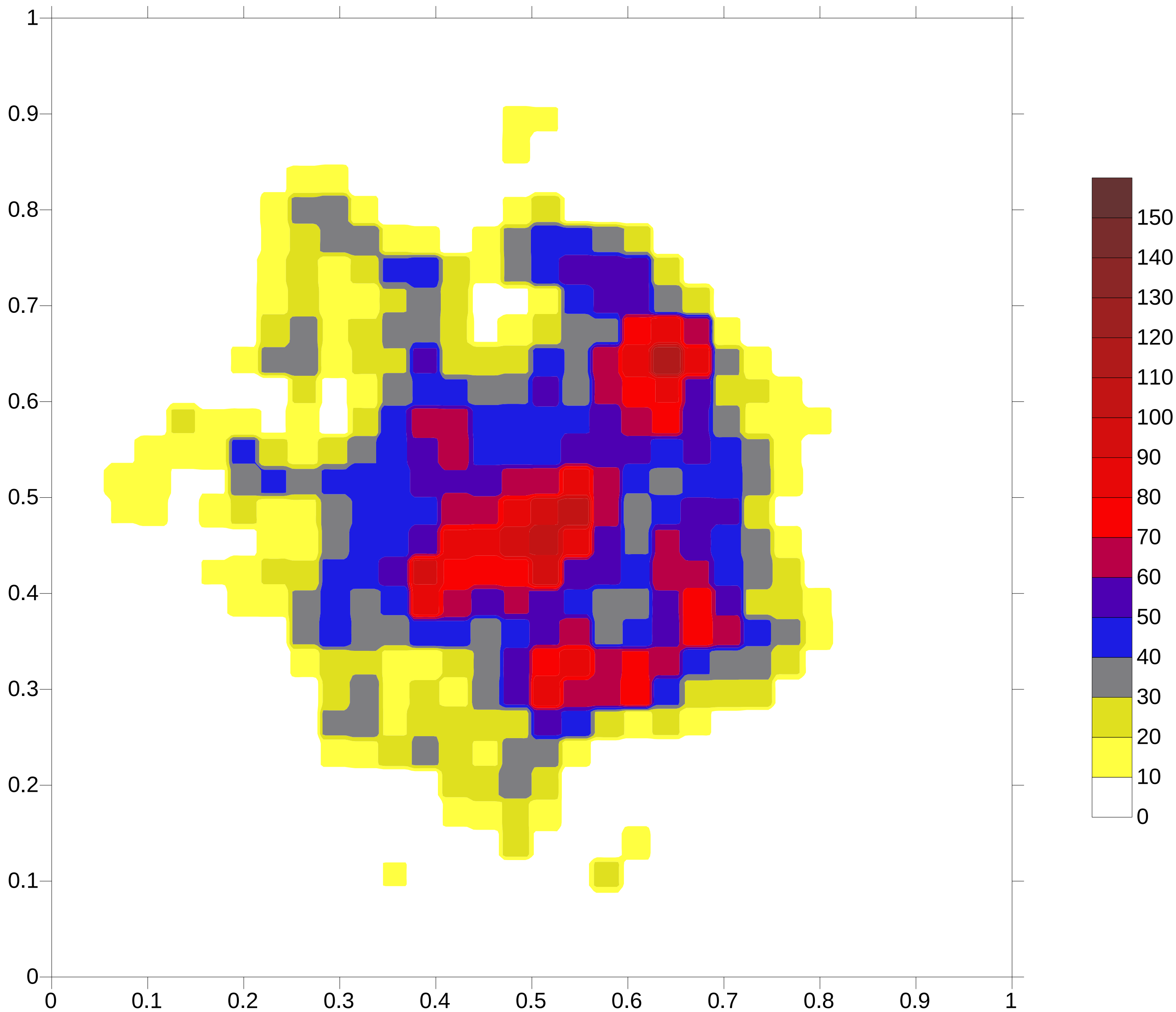
- **Observations correspond to a certain area or volume**
- **These may differ**
  - Measurement devices
  - Sampling techniques

- **Distribution of values change when support changes !!**

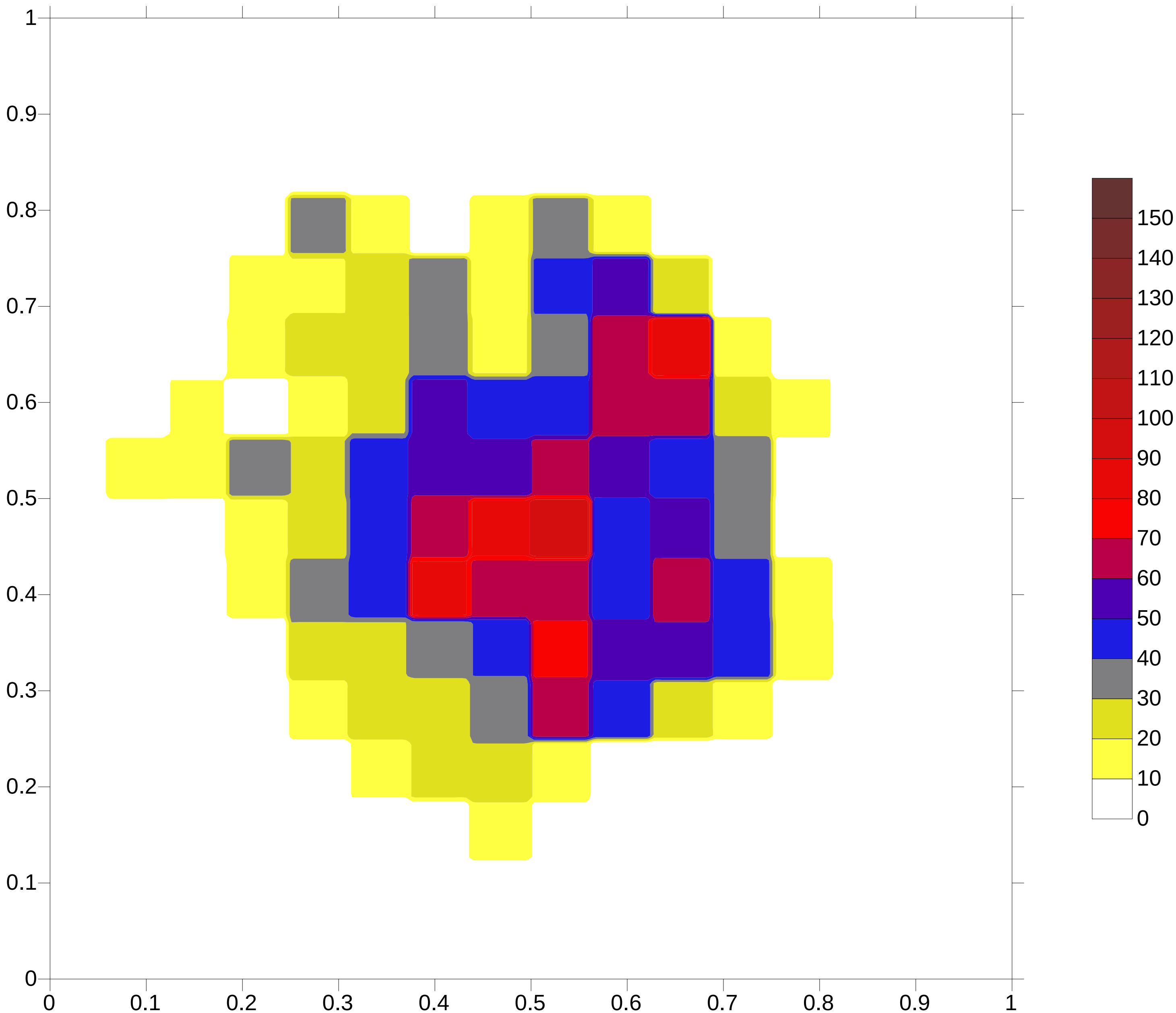


1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
150
140
130
120
110
100
90
80
70
60
50
40
30
20
10
0

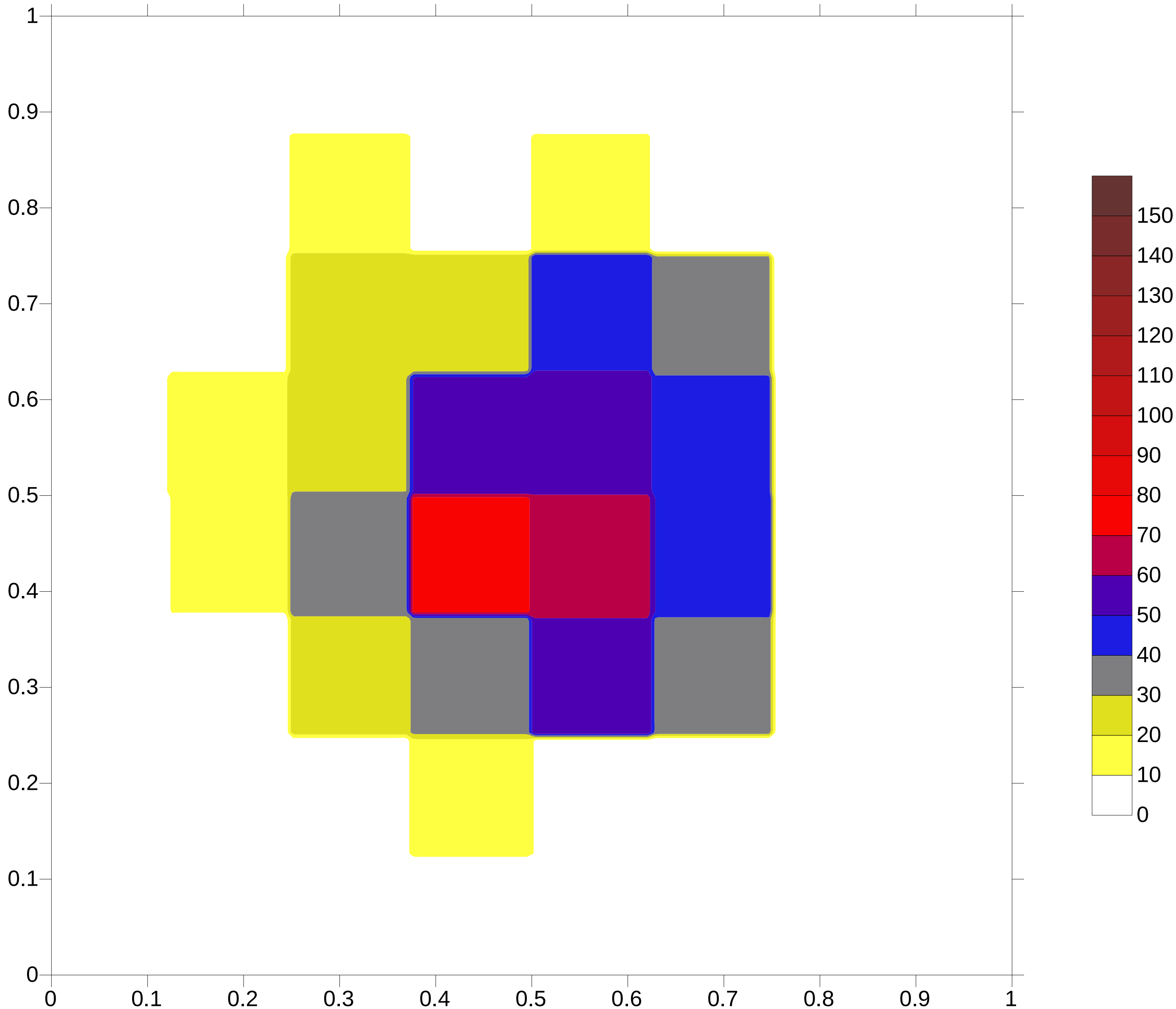

1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
100
150
140
130
120
110
100
90
80
70
60
50
40
30
20
10
0



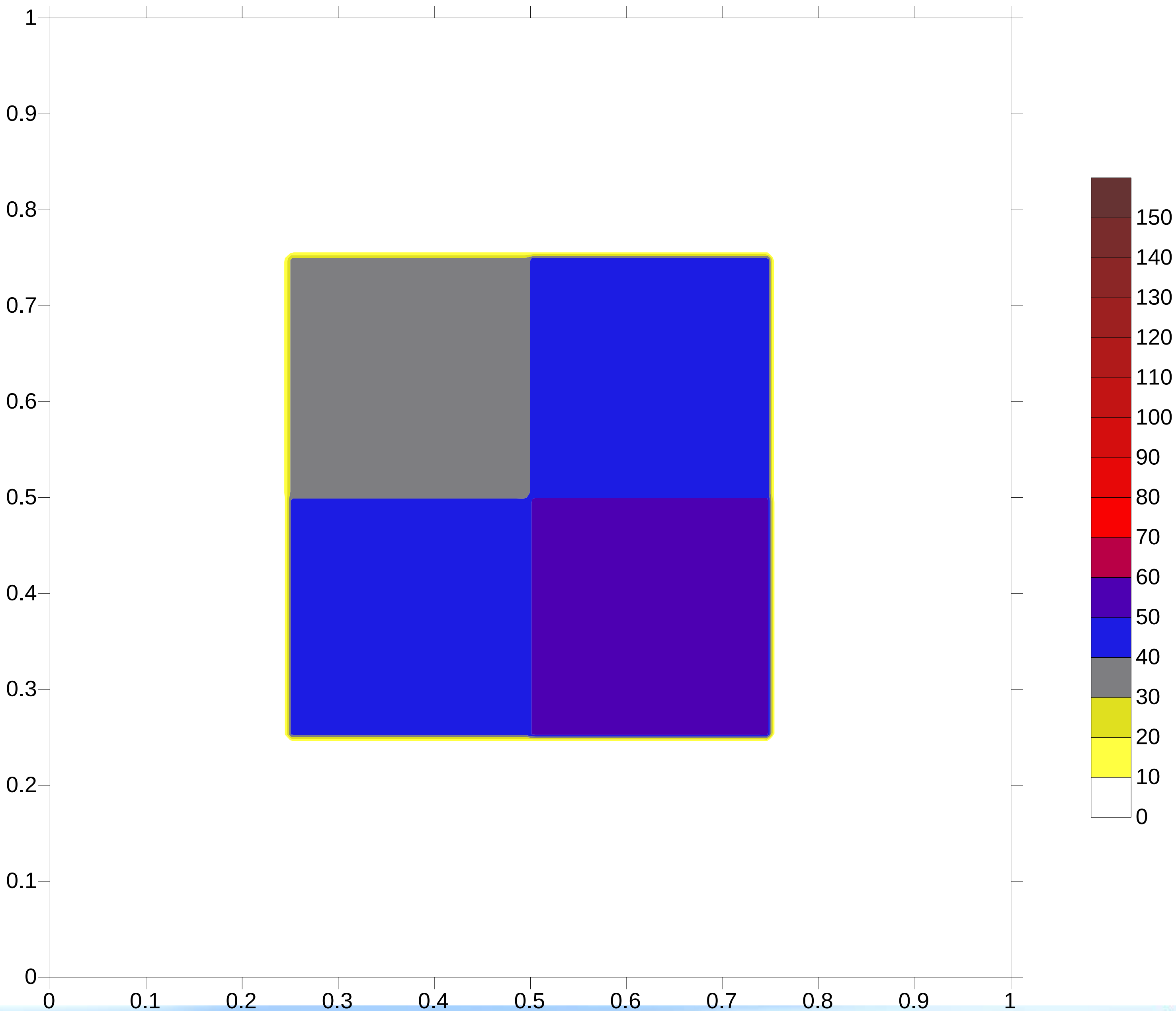
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
150
140
130
120
110
100
90
80
70
60
50
40
30
20
10
0



1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
150
140
130
120
110
100
90
80
70
60
50
40
30
20
10
0



0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1
0 10 20 30 40 50 60 70 80 90 100 110 120 130 140 150



0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
150
140
130
120
110
100
90
80
70
60
50
40
30
20
10
0

- **Different supports:**
  - – Different marginal distributions
  - – Increase of support → decrease of variance
  - – Different variongrams

- **Do not mix them for:**
  - – Variogram calculation
- **You can mix for interpolation**
  - – Block Kriging

Alternate realities

- **Interpolation is not a possible reality**
- **Generate realities:**
  - Same variogram
  - Same observations
  - Matching other statistics
- **Consequence**
  - Simulation error is higher than interpolation error
  - Variability is realistic (?) thus good for nonlinear cases (Risk assessment)

Simulation methods

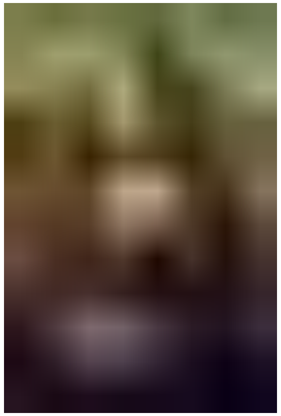
- **Using different assumptions**
  - LR Cholewsky decomposition
  - Turning bands
  - Fast Fourier Transform
  - Sequential Simulations
  - Simulated Annealing

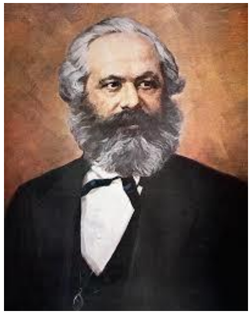
  …

  Uncertainty validation is necessary!

Goals

Interpolation
Simulation





Thank you for your attention!