WP5 – Training session: RGeostats

Geostatistics Course & Exercices

INTAROS – General Assembly

Haus der Wissenschaft,
Sandstrasse 4/5
28195 Bremen, Germany
9.00-16.00
January 11th 2019

RENARD Didier, ARMINES (lead)
ORS Fabien, ARMINES (co-lead)

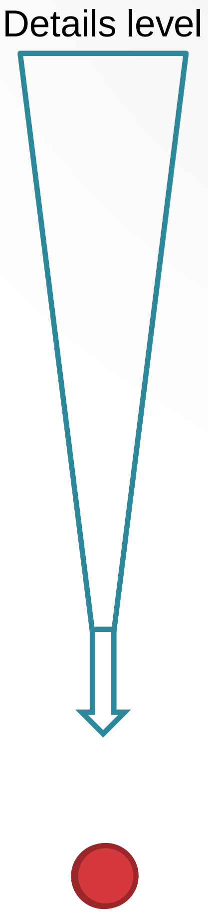
Outline

Thursday 10th

1. Creating iAOS Processing Services
2. Geostatistics and RGeostats
3. Ellip Notebooks using RGeostats
4. Ellip Worflow using RGeostats
5. IMR Case Study - RGeostats in Action!

Friday 11th

Geostatistics Course & Exercices

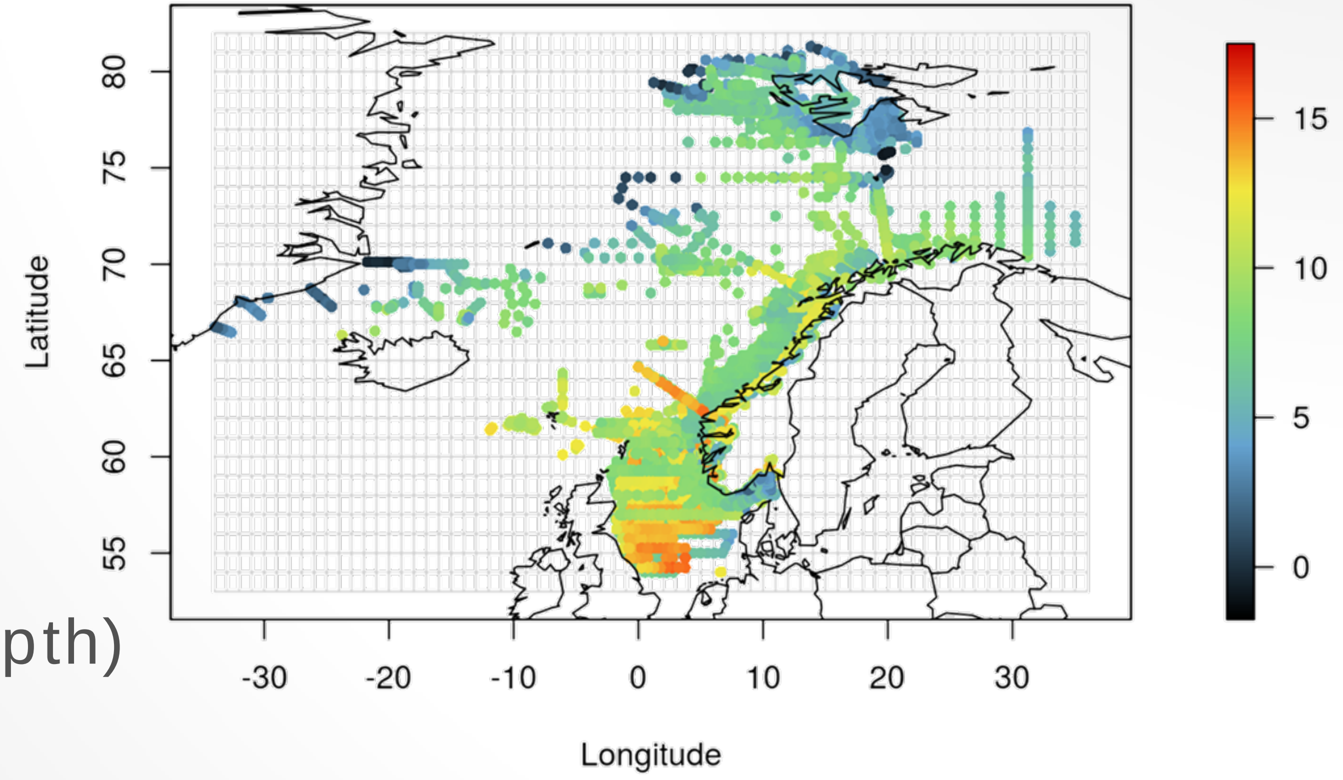


Introduction

IMR dataset global overview

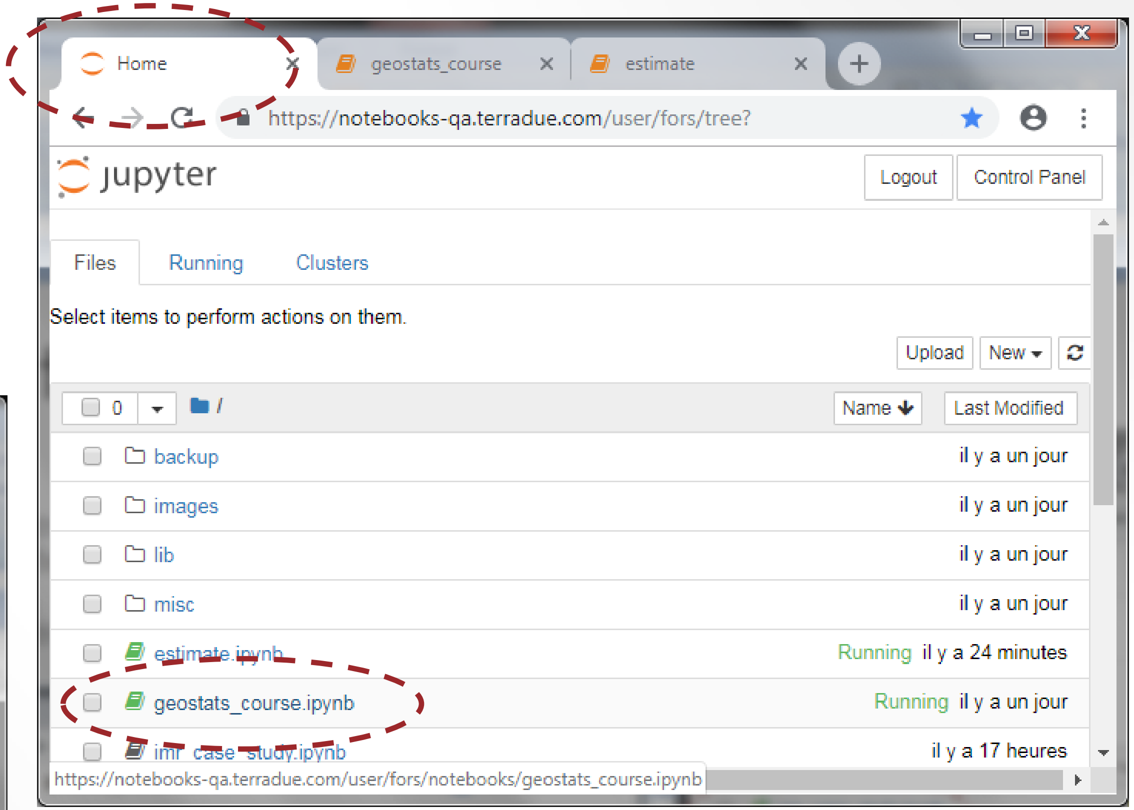
- 7 vessels
- from 1995 to 2016
- 3 variables measured:
    - Temperature
    - Salinity
    - Conductivity

- 63 500 positions {long, lat}
- 63 500 vertical profiles (in depth)
- A few million samples
- 84 NetCDF files (~60 Mb each)



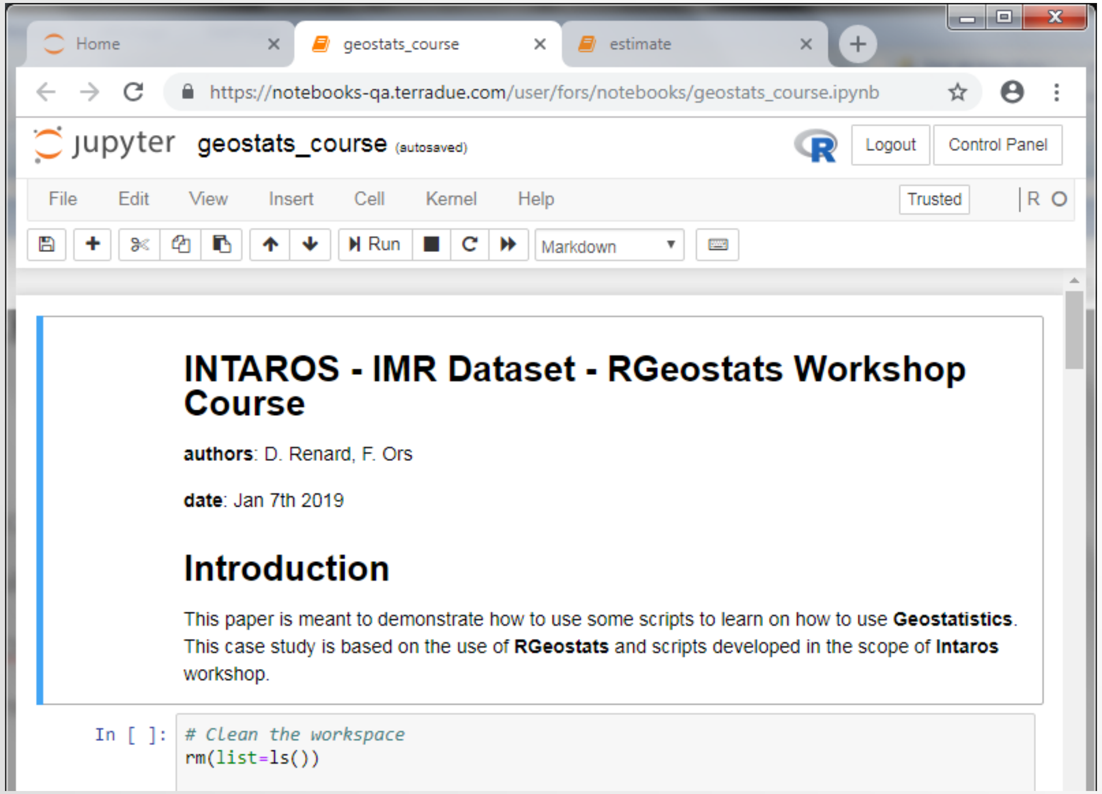
Getting Ready for Exercices!

- Click on the **Home** tab of your Jupyter Notebooks (in Chrome)
- Click on the **geostats_course.ipynb**



Select items to perform actions on them.

backup		il y a un jour
images		il y a un jour
lib		il y a un jour
misc		il y a un jour
estimate.ipynb	Running	il y a 24 minutes
geostats_course.ipynb	Running	il y a un jour
imr_case_study.ipynb		il y a 17 heures



INTAROS - IMR Dataset - RGeostats Workshop Course

authors: D. Renard, F. Ors

date: Jan 7th 2019

Introduction

This paper is meant to demonstrate how to use some scripts to learn on how to use **Geostatistics**. This case study is based on the use of **RGeostats** and scripts developed in the scope of **Intaros** workshop.

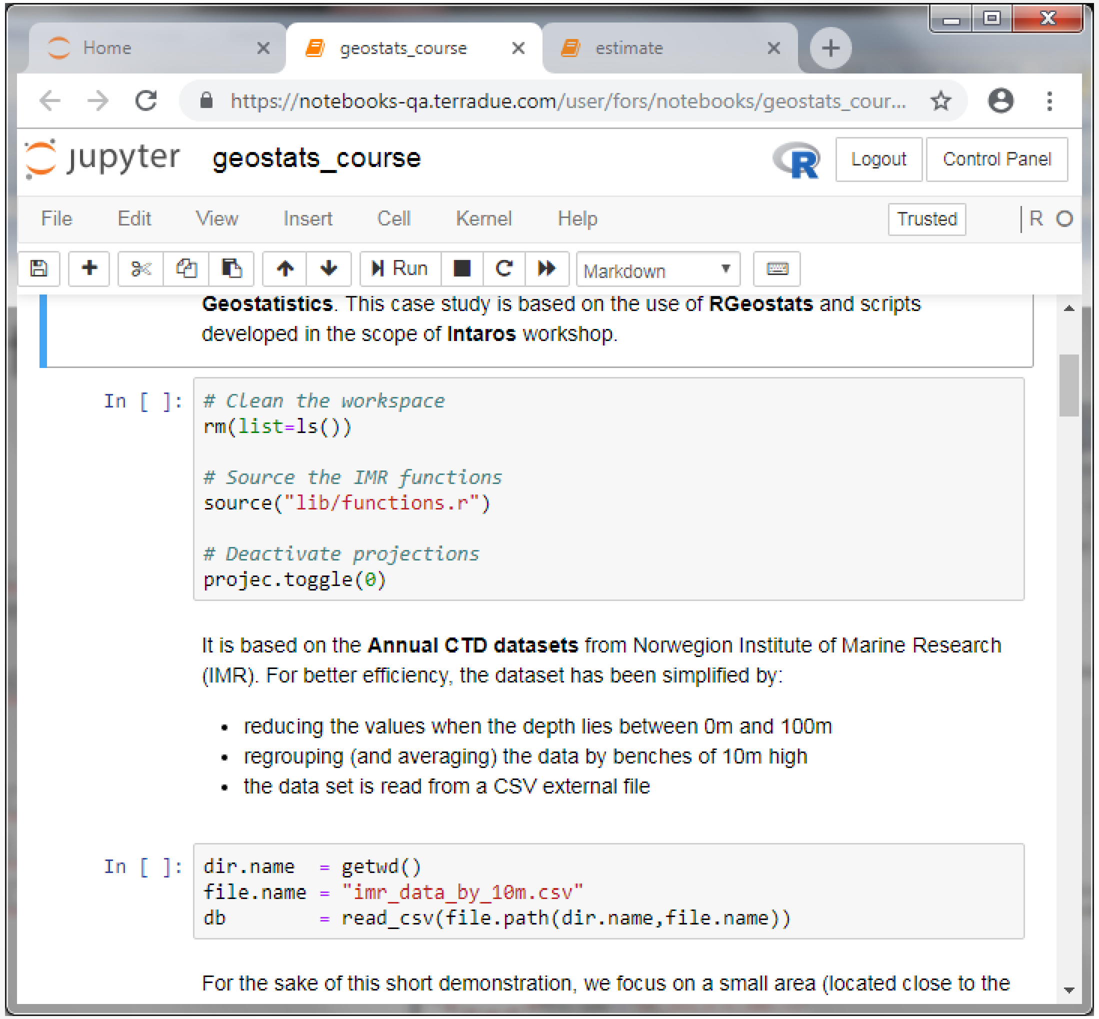
```
In [ ]: # Clean the workspace
        rm(list=ls())
```

Getting Ready for Exercices!

- Execute all code cells in the **Introduction** section:

  => Hitting **Shift + Enter**

  - Definition of R functions
  - Loading Data (slow)
  - Setting global environment



Geostatistics. This case study is based on the use of **RGeostats** and scripts developed in the scope of **Intaros** workshop.

In [ ]:
```
# Clean the workspace
rm(list=ls())

# Source the IMR functions
source("lib/functions.r")

# Deactivate projections
projec.toggle(0)
```

It is based on the **Annual CTD datasets** from Norwegion Institute of Marine Research (IMR). For better efficiency, the dataset has been simplified by:

- reducing the values when the depth lies between 0m and 100m
- regrouping (and averaging) the data by benches of 10m high
- the data set is read from a CSV external file

In [ ]:
```
dir.name  = getwd()
file.name = "imr_data_by_10m.csv"
db        = read_csv(file.path(dir.name,file.name))
```

For the sake of this short demonstration, we focus on a small area (located close to the

From Data to Estimation (Kriging)

Data – Measurements

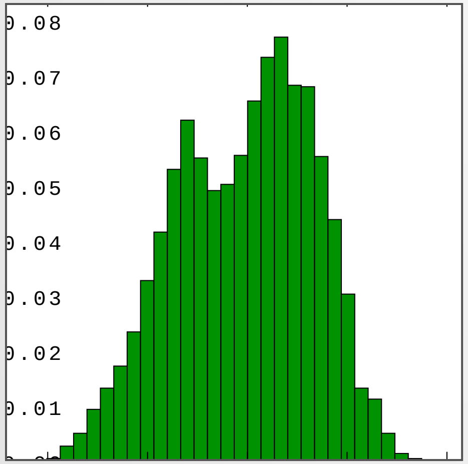
Space: 1-D, 2-D, 3-D or more

Data:

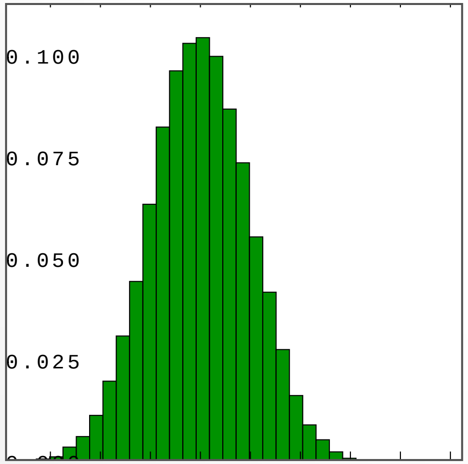
- Number : few to several thousands
- Locations: isolated, along lines, regular
- Time dependency

Statistics

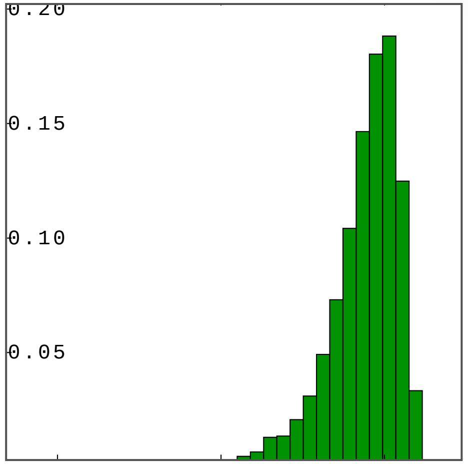
- Mean, variance (or standard deviation)
- Histograms: extremes, mode, median, quantiles



Bimodal



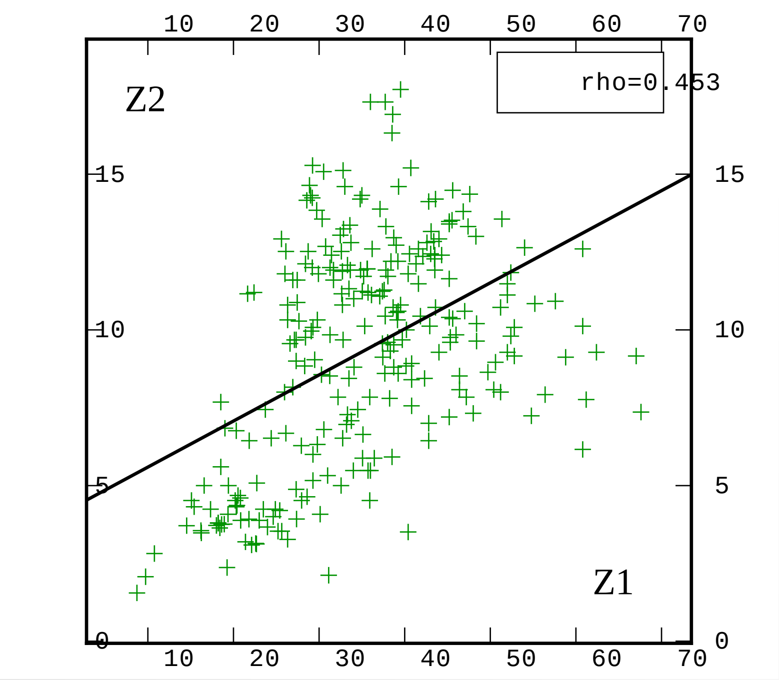
Symmetric

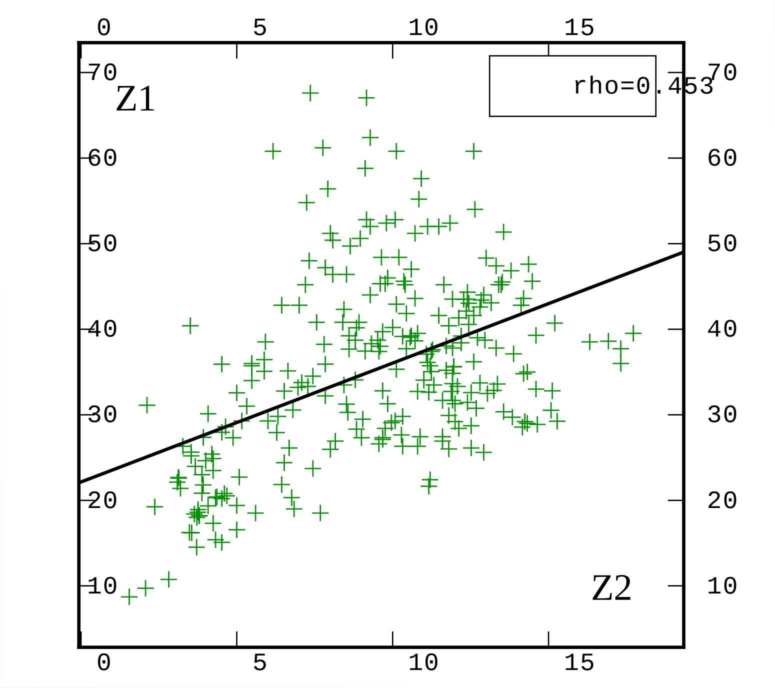


Skewed

Statistics

- Mean and variance of each variable
- Correlation, covariance, linear regression





$$Z_2 = aZ_1 + b$$

$$Z_1 = a'Z_2 + b'$$

Back to Jupyter Notebook!

- We focus on the cells from **Basics statistics** section:
  - Global Database Statistics
  - Display Temperature Variable
  - Temperature Histogram
  - Temperature vs Salinity Correlation
  - Sample Filtering
  - Statistics per Blocks
  - Mean and Variances by Years

Exercise

- Calculate the experimental mean and variance for Z_1

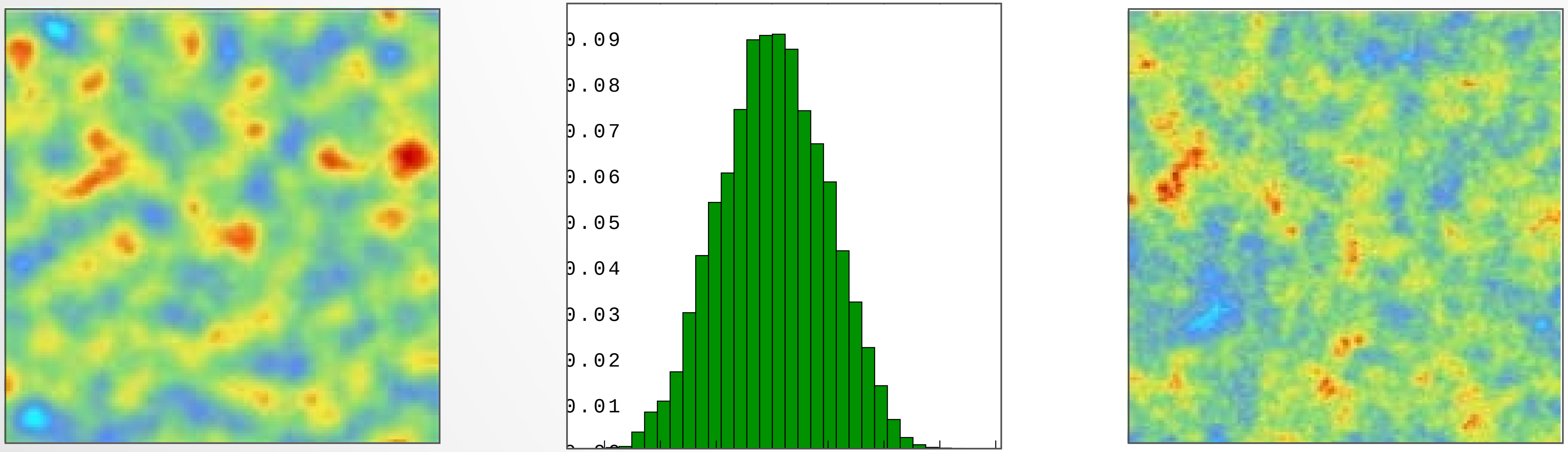
0 1 2 0 2 1 0 1 2 1

- Same calculation for Z_2

1 1 1 0 2 1 0 0 2 2

Statistics

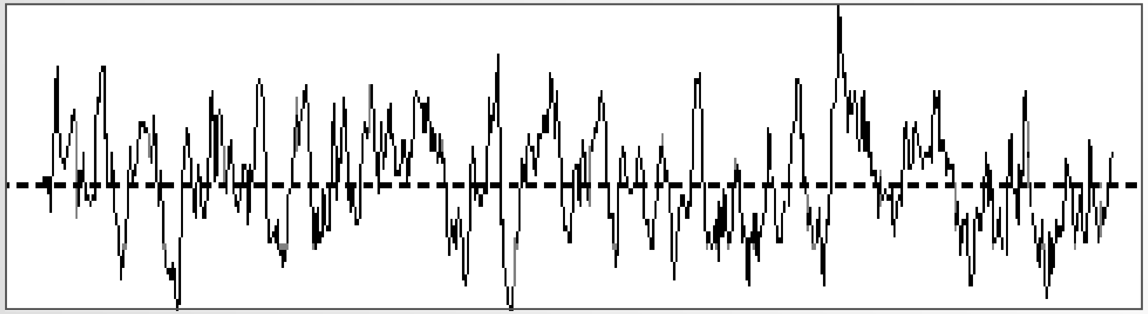
Punctual statistics are not sufficient:



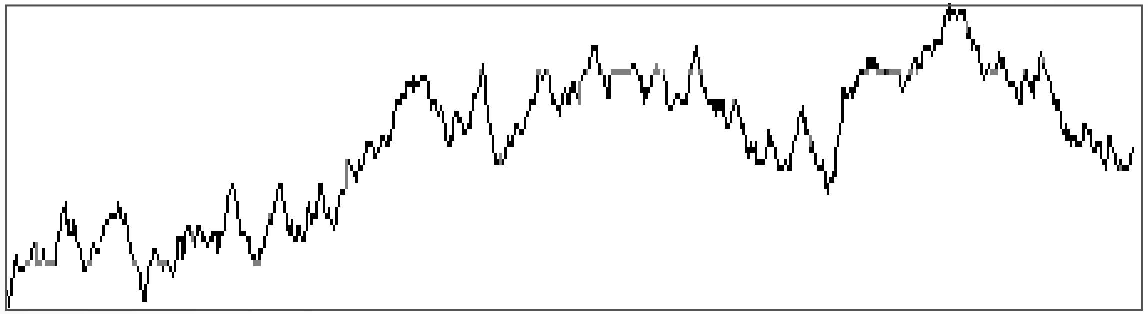
Two images with the same histogram

Hypotheses

- Stationary RF : invariance under translation of the spatial law.
- Order-2 stationary RF: the first two moments exist and are invariant under translation.
- Intrinsic RF (IRF0): the increments are order-2 stationary.



Stationary



Intrinsic

Hypotheses

- An stationary RF is also intrinsic, but an intrinsic RF is not necessarily stationary.

- No attraction by the mean…
  but no systematic behavior (as the depth of the bottom of the sea which increases regularly from the beach)

- The purely intrinsic model stands between the stationary and the non-stationary cases. The choice of the degree of non-stationarity depends upon the field of observation.

Experimental Variogram

- Stationary hypothesis: covariance

$$Cov(h) = E\left[Z(x+h) \times Z(x)\right]$$

- Intrinsic hypothesis

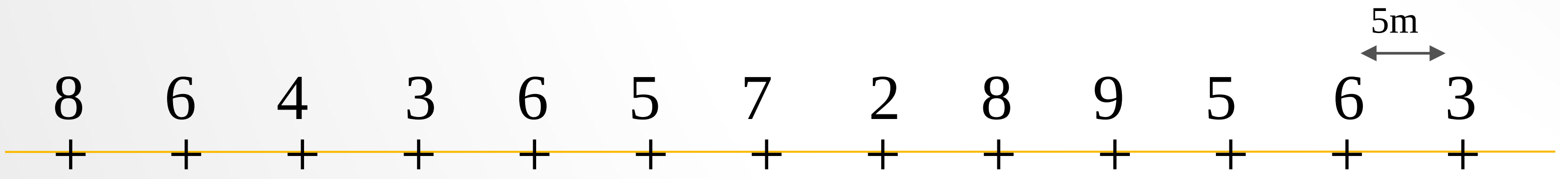
$$\gamma(h) = \frac{1}{2} Var\left[Z(x+h) - Z(x)\right] = \frac{1}{2} E\left[Z(x+h) - Z(x)\right]^2$$

$$\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} \left[Z\left(x_i + h\right) - Z\left(x_i\right)\right]^2$$

Work with variogram rather than covariance (more general framework)

Exercise:
Variogram on regular 1-D sampling

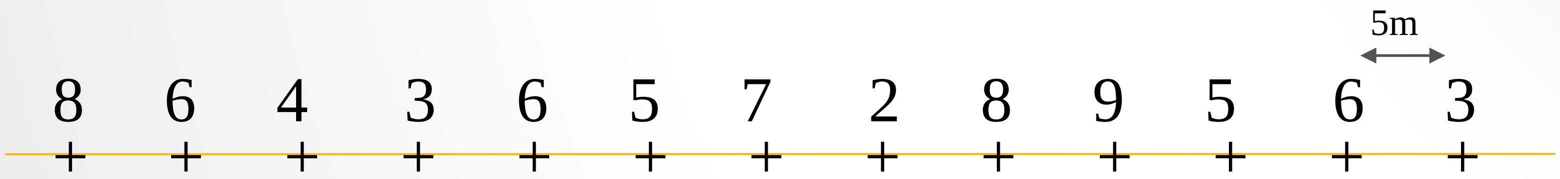
Variable Z defined on a regular grid in 1-D



- Calculate the experimental variogram for the lags: 5m, 10m and 15m.
- Evaluate the experimental variogram of the new variable:

$$Y(x) = Z(x) + 3.2$$

Solution:
Variogram on regular 1-D sampling



Consider the distance 5m:

$$\gamma(5m)=\frac{1}{2}\left[(8-6)^2+(6-4)^2+(4-3)^2+\ldots+(6-3)^2\right]=4.625$$

Consider the distance 10m:

$$\gamma(10m)=\frac{1}{2}\left[(8-4)^2+(6-3)^2+(4-6)^2+\ldots+(5-3)^2\right]=5.227$$

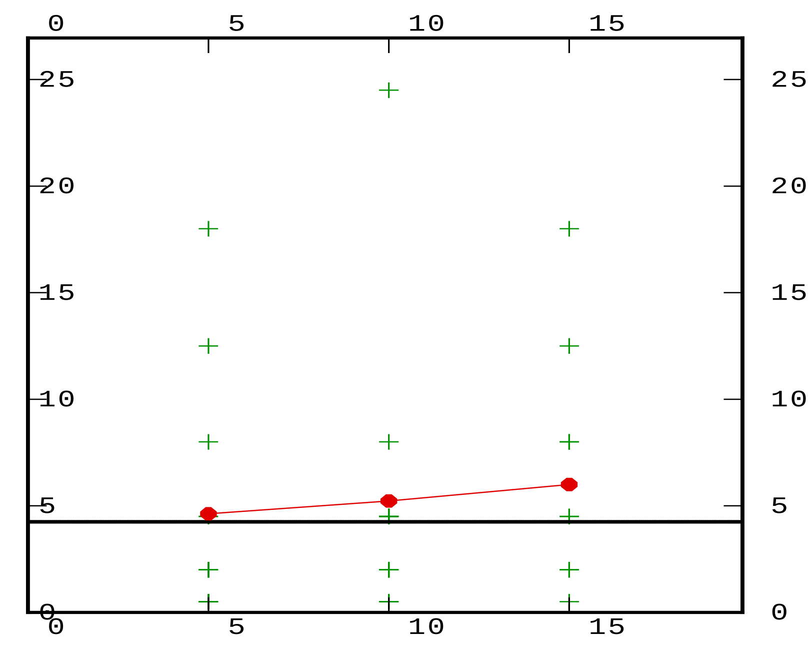
Consider the distance 15m:

$$\gamma(15m)=\frac{1}{2}\left[(8-3)^2+(6-6)^2+(4-5)^2+\ldots+(9-3)^2\right]=6.000$$

Variogram Cloud

Variable Z defined on a regular grid in 1-D

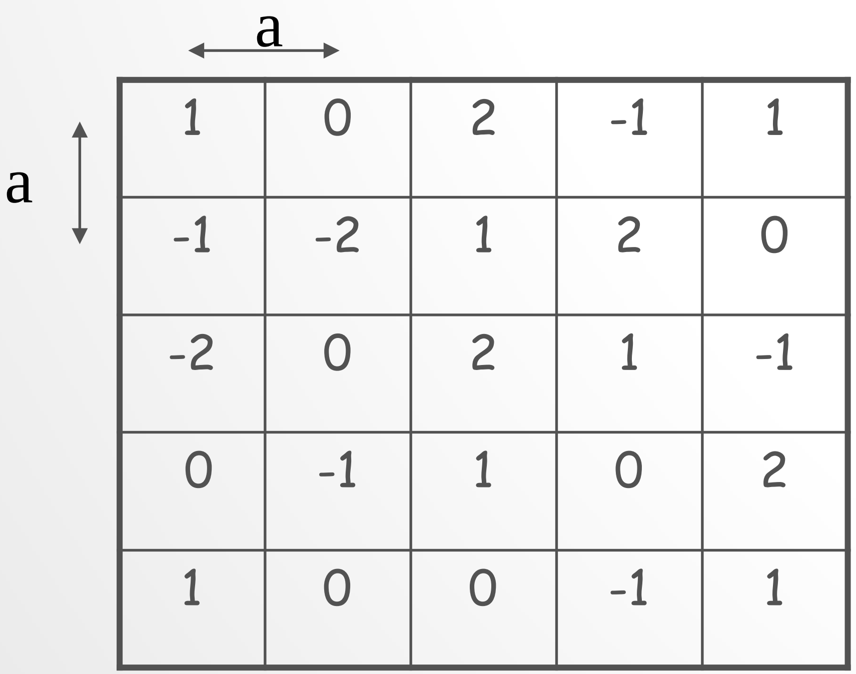
$\frac{1}{2}\,[(Z(x_1)-Z(x_2))^2]$



Distance(x_1,x_2)

Variogram on regular 2-D grid

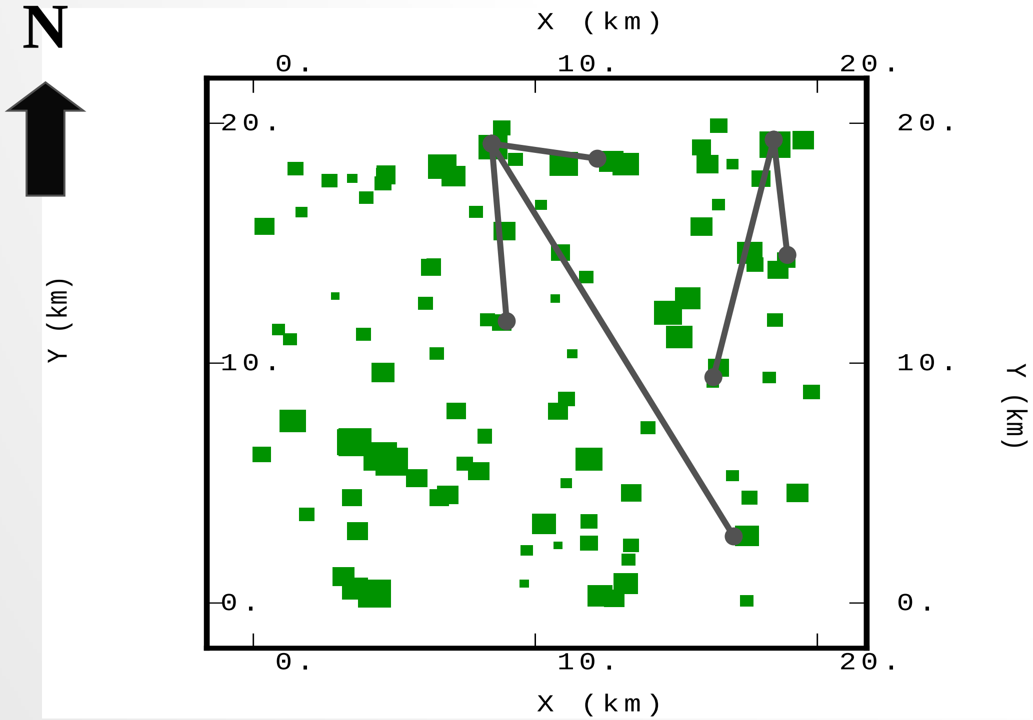
Variable Z defined on a regular grid in 2-D (square mesh = a)



1	0	2	-1	1
-1	-2	1	2	0
-2	0	2	1	-1
0	-1	1	0	2
1	0	0	-1	1

Variogram on 2-D irregular data

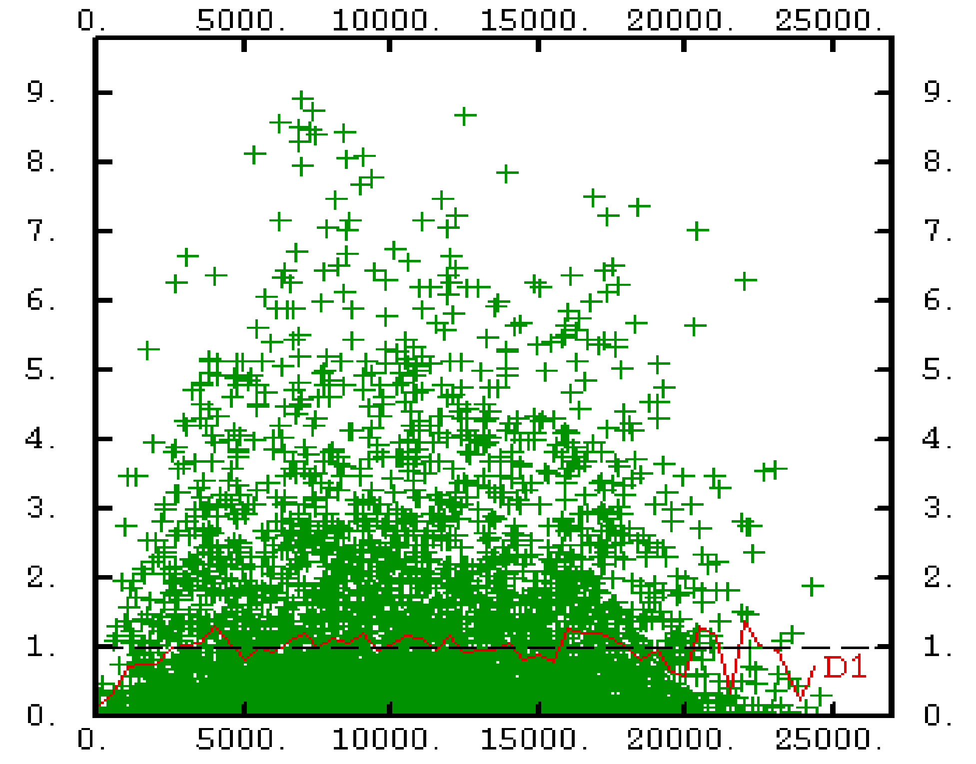
Calculation of the experimental omni-directional variogram



Variogram on 2-D irregular data

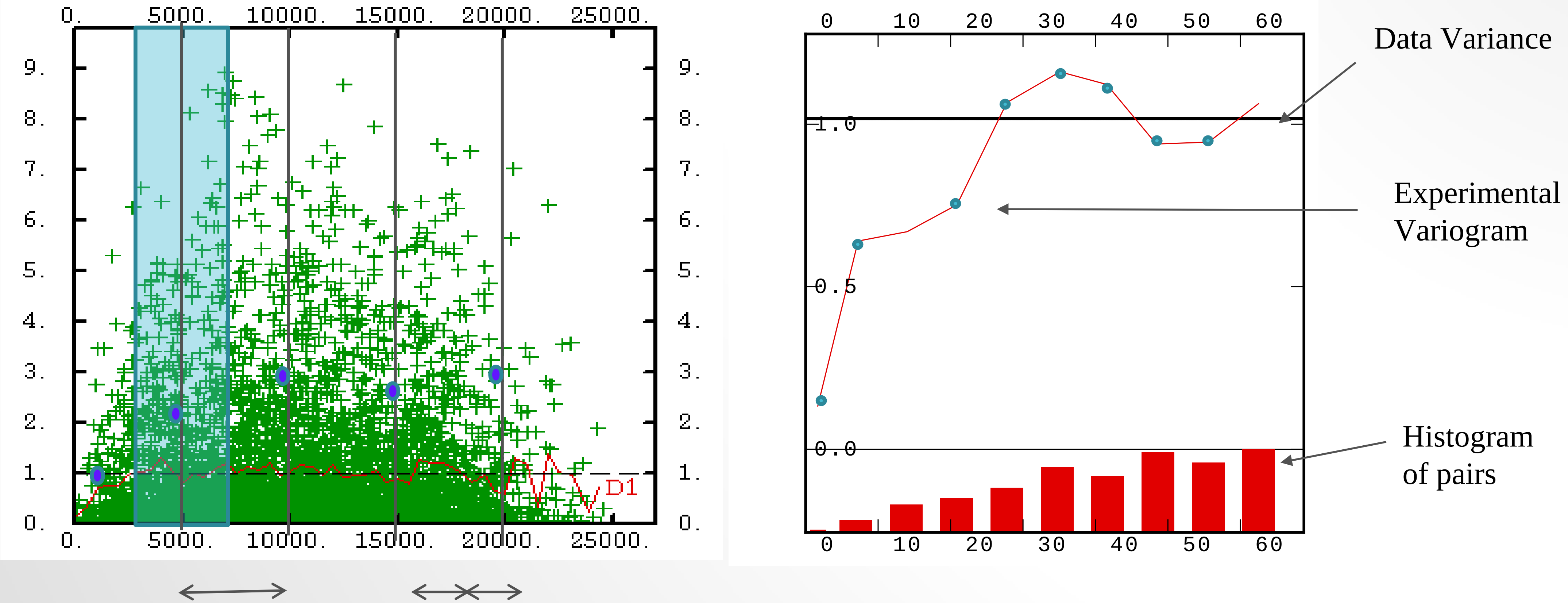
Establish the Variogram Cloud

½ $[(Z(x_1)-Z(x_2))^2]$



Distance(x_1,x_2)

Experimental Variogram



Data Variance

Experimental Variogram

Histogram of pairs

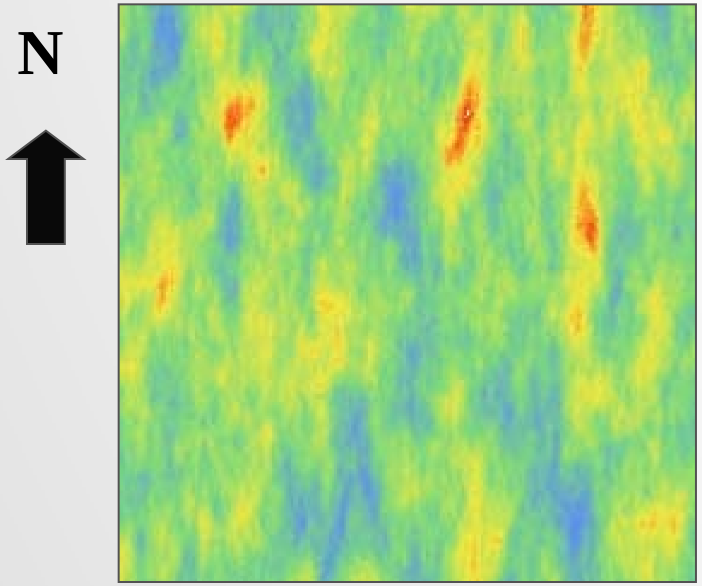
Lag

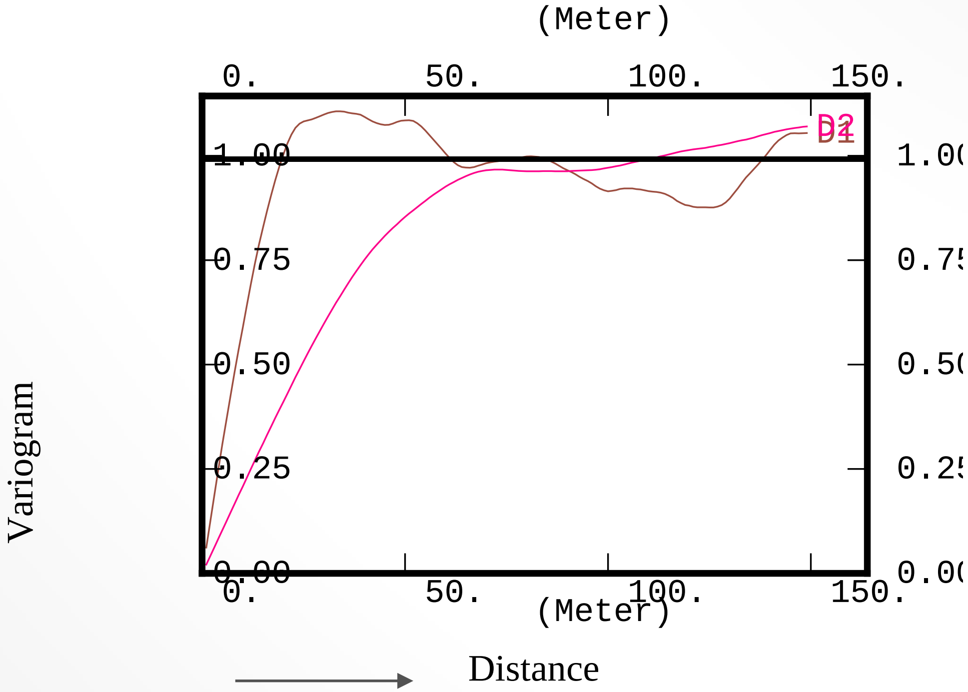
Tolerance

Number of lags

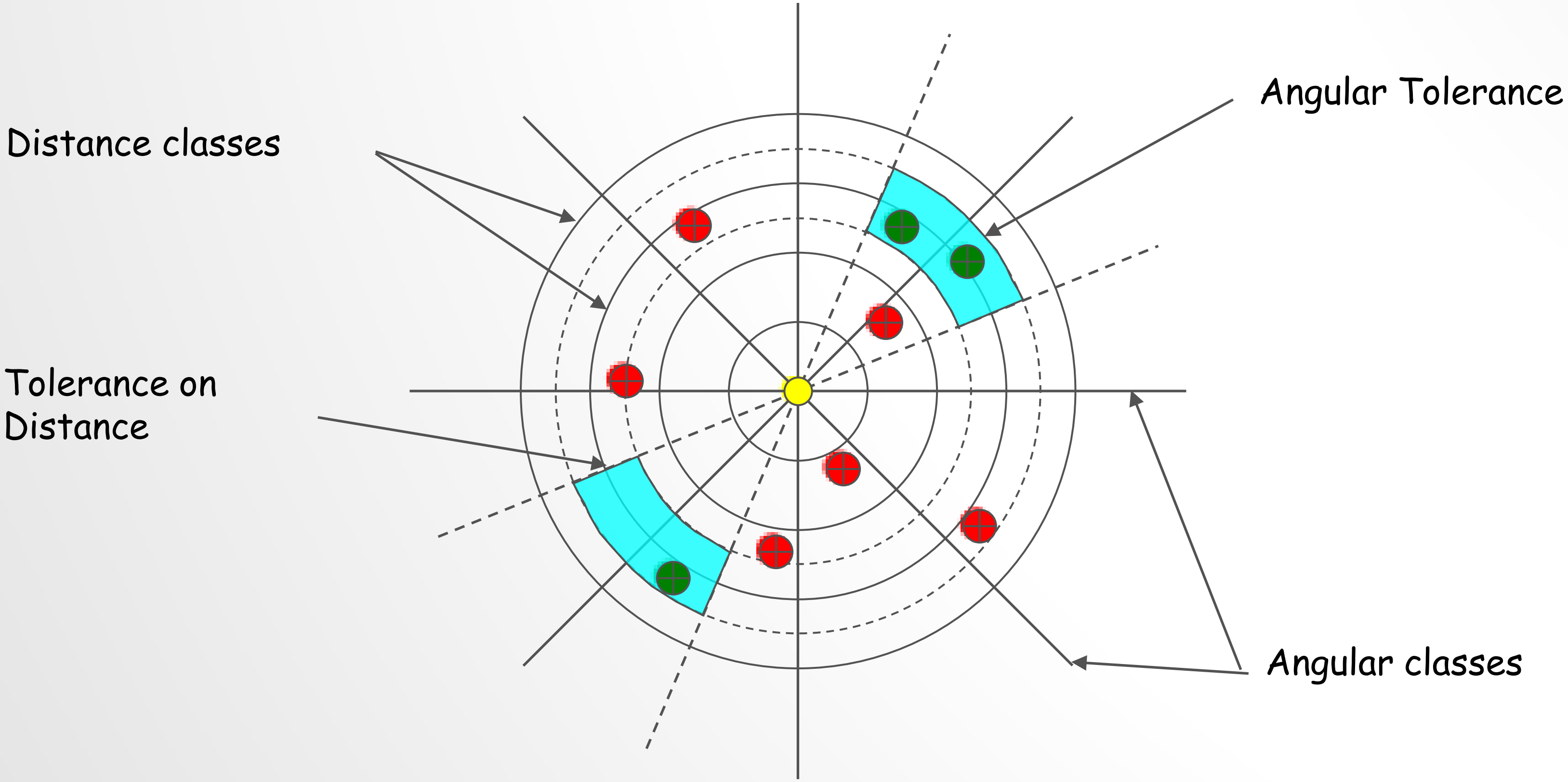
Directional Variograms

Calculation along two directions: E-W and N-S – Looking for **Anisotropy**





Directional Variograms



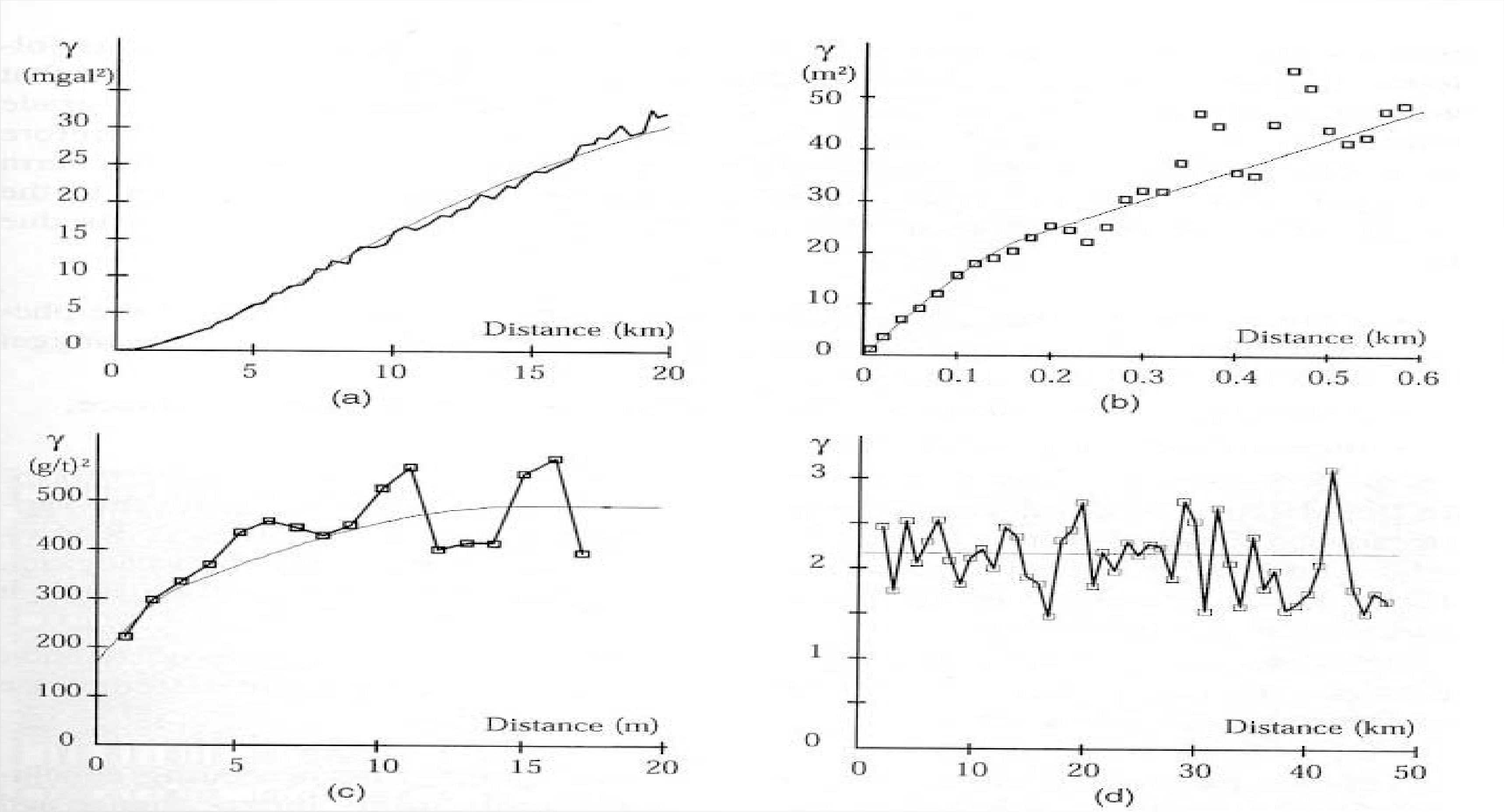
Distance classes

Angular Tolerance

Tolerance on Distance

Angular classes

Examples of Variograms



From "Geostatistics: modeling spatial uncertainty"
J.P. Chilès, P. Delfiner (Wiley 1999)

Hints for Variogram calculation

- Analyze the variogram cloud in order to detect outliers. Pay attention to presence of different areas or faults, different measurement tools.
- Choose the lag and the tolerance on distance. Check the homogeneity of number of pairs for all lags.
- If possible, calculate variograms in several directions (at least 4 in 2D, 5 in 3D) in order to look for possible anisotropy

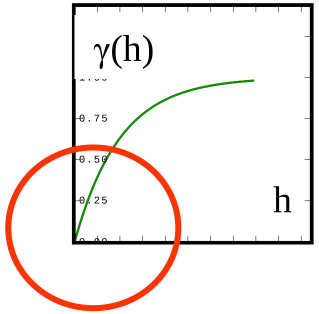
Fitting a Model

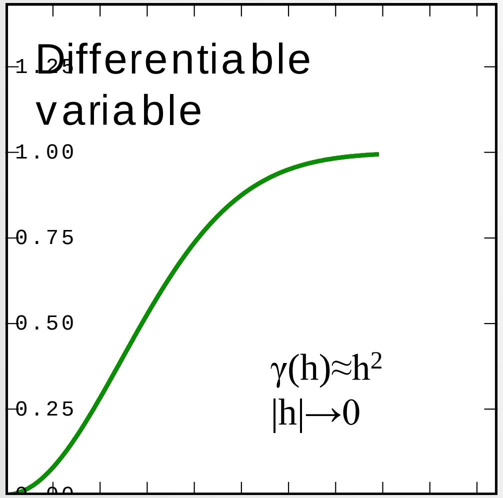
Procedure:

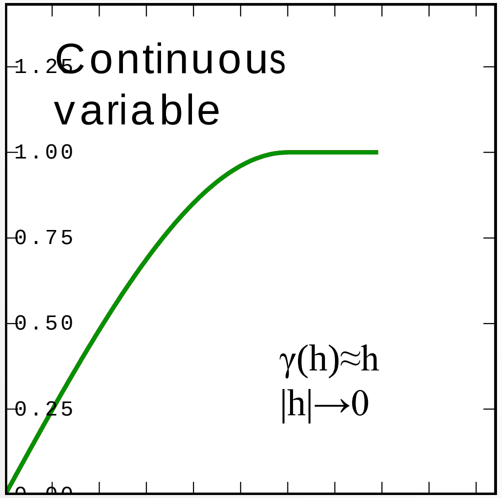
- Choose a single variogram **Model** for all directions and all distances
- Use a valid type of function: authorized Model
- Needed further (for Estimation) as it ensures (conditional) positivity of the variance of any linear combination
- As close as possible to the Experimental Variogram:
    - Behavior near the origin
    - Behavior at large distance
    - Specific features: presence of anisotropy, presence of trend, …

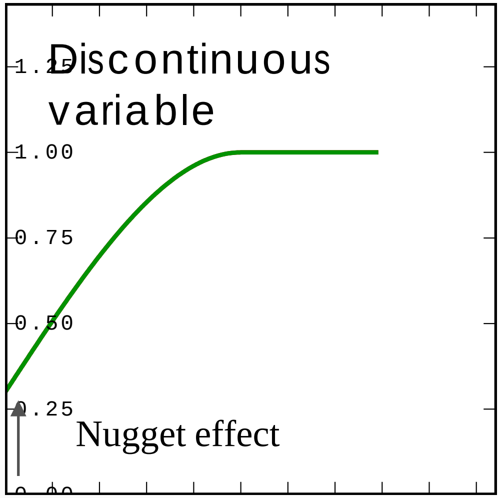
Behavior near the origin

- The behavior of the variogram near the origin is directly linked to the continuity of the variable



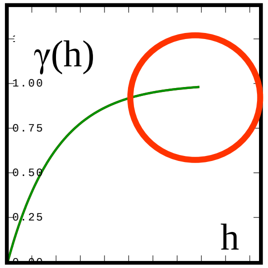

Differentiable variable


Continuous variable

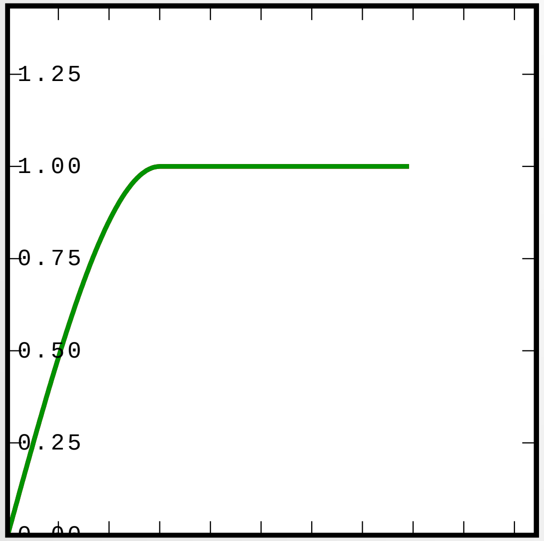

Discontinuous variable

Behavior at large distance

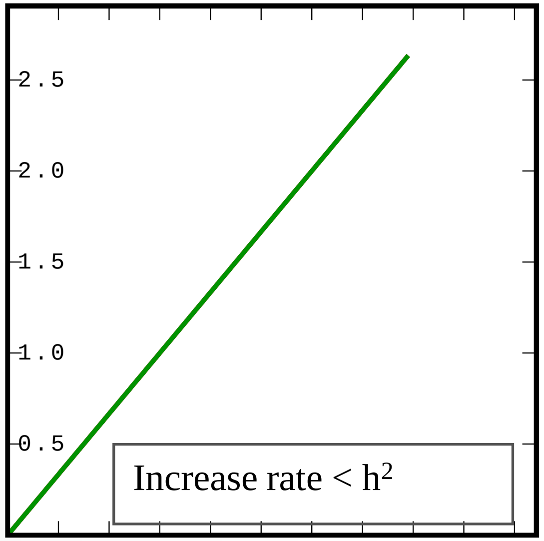
Check for strict stationarity



Bounded variogram

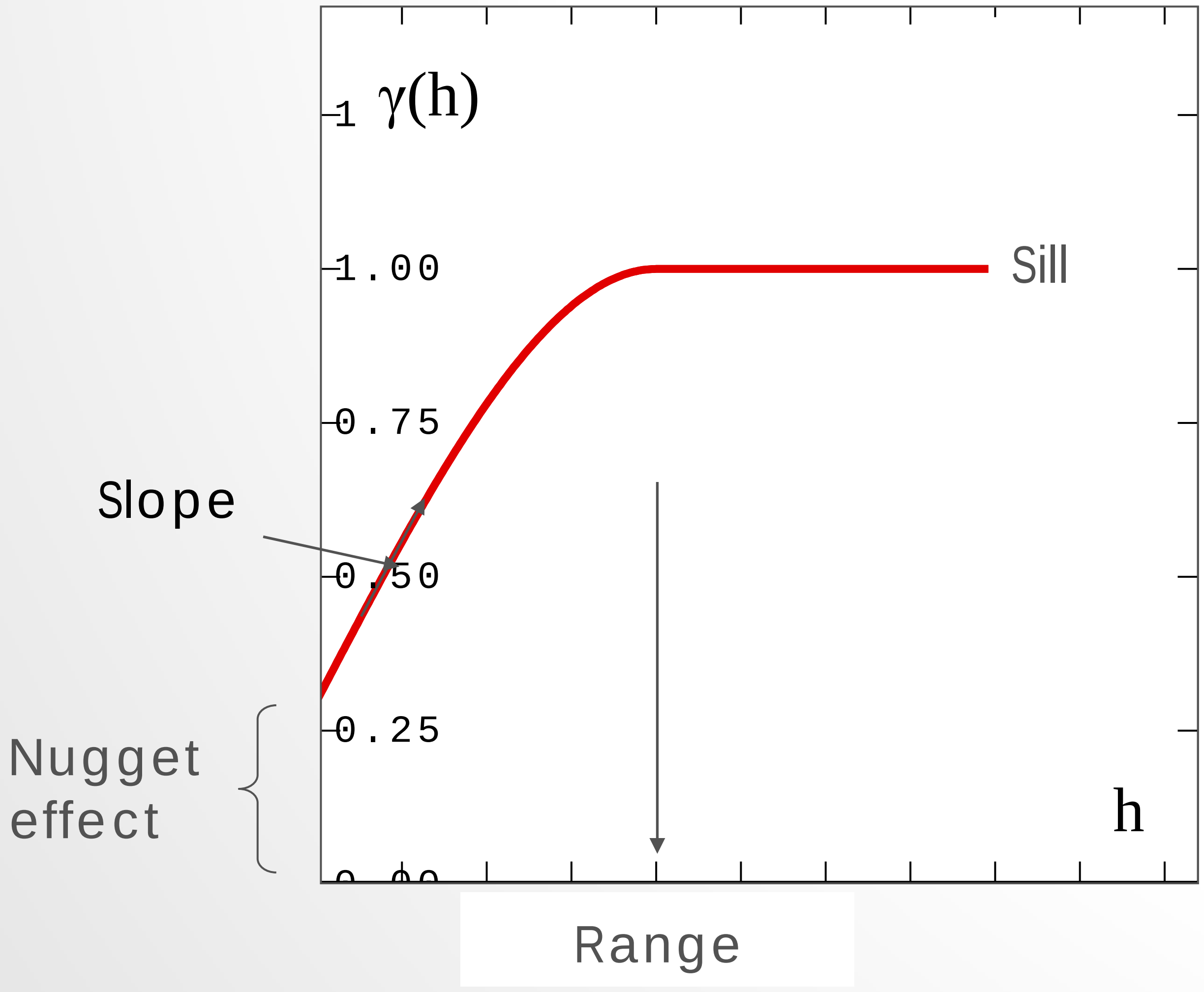


Unbounded variogram

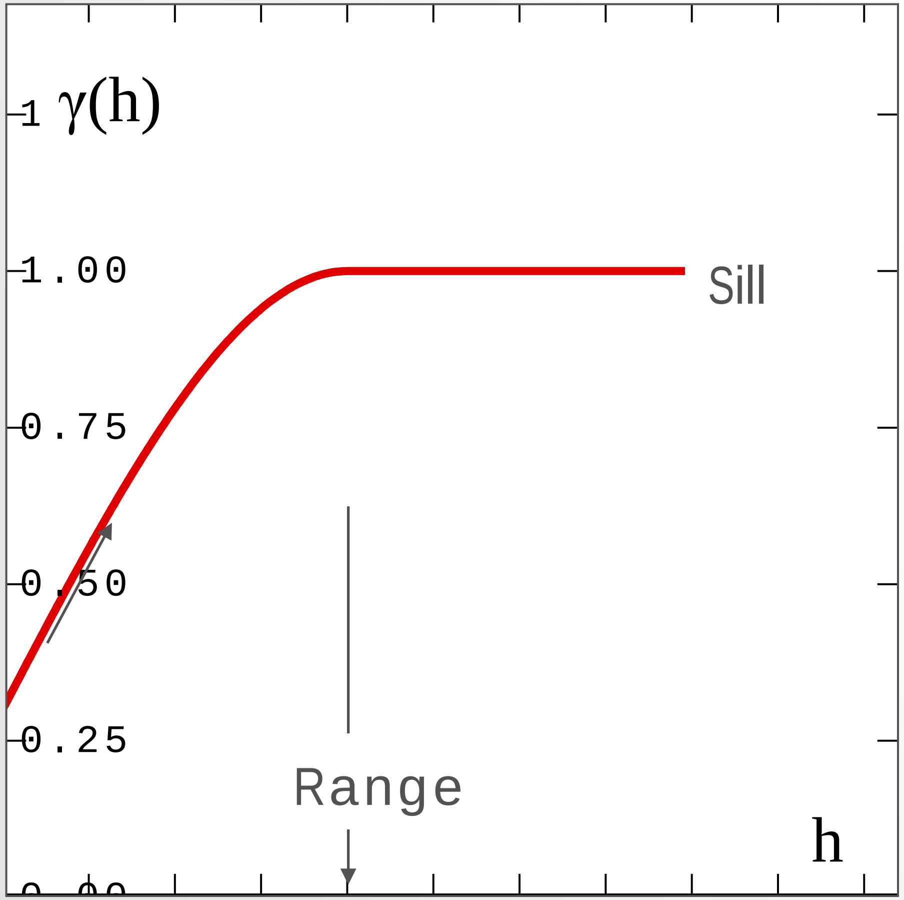


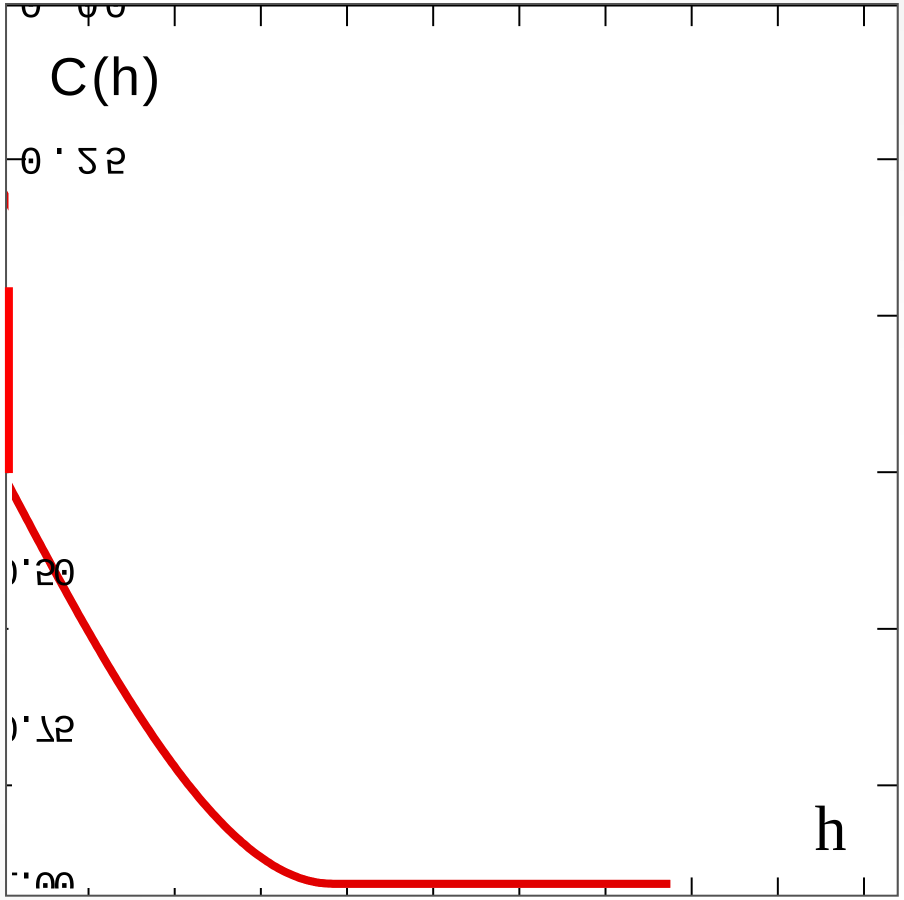
Increase rate < h^2

Model characteristics



Variogram vs. Covariance

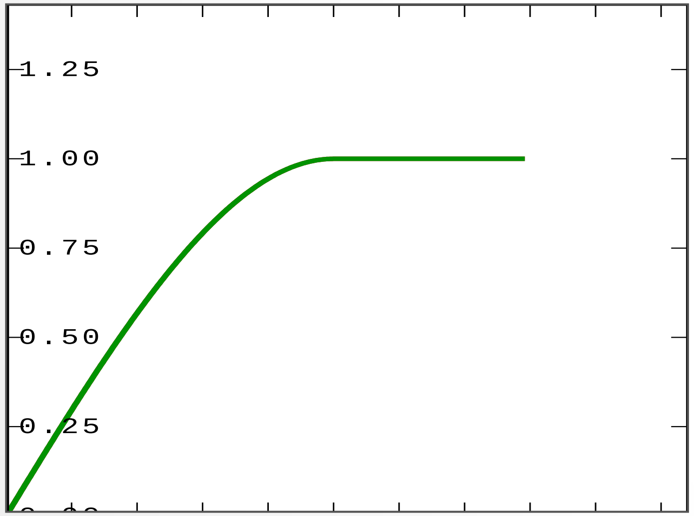




Bounded variogram can be turned into covariance (stationary case)

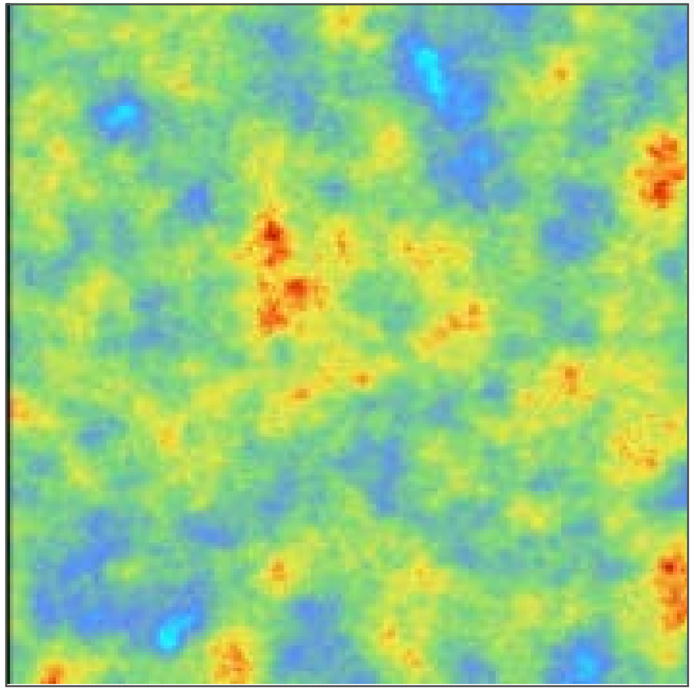
Usual Models

Spherical



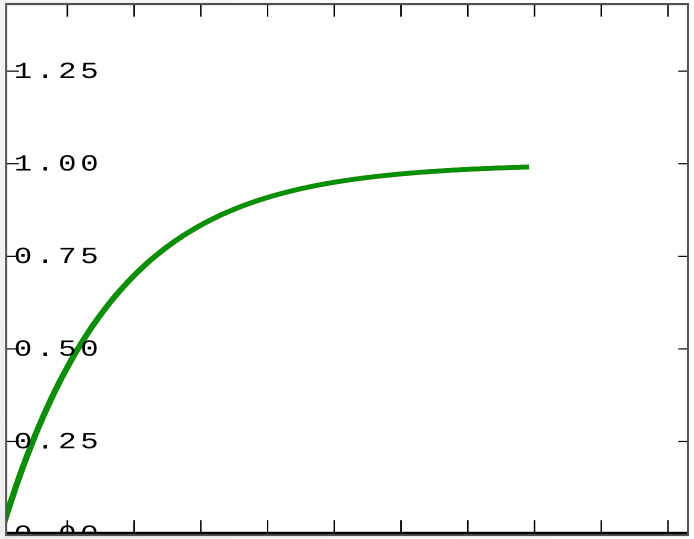
$$\gamma(h) \quad = \frac{c}{2}\left(\frac{3h}{a} - \frac{h^3}{a^3}\right) \quad h \le a$$

$$= c \quad h > a$$

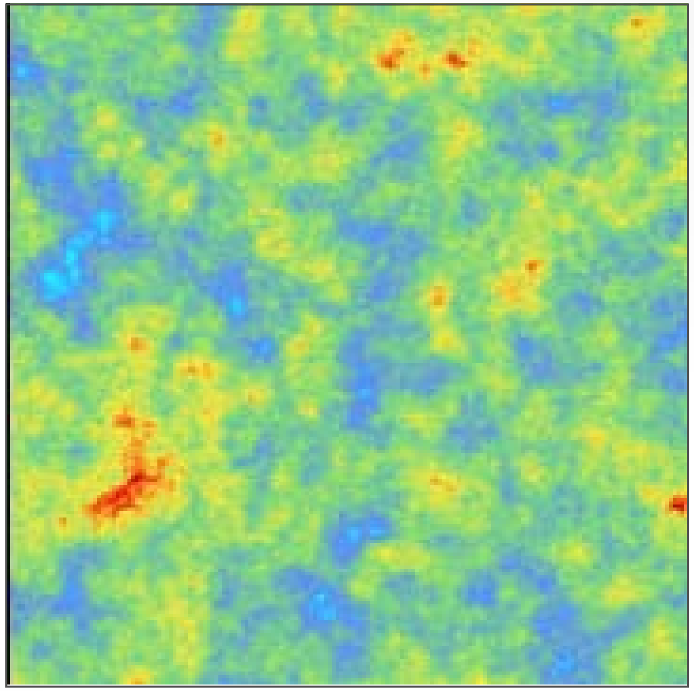


Usual Models

Exponential

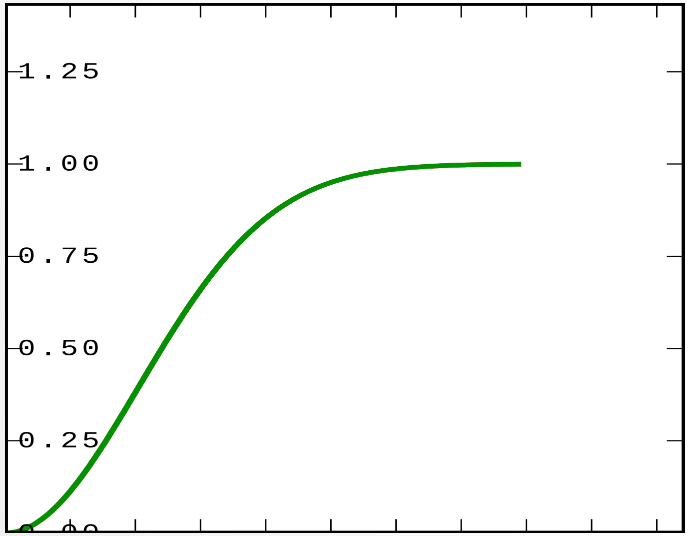


$$\gamma(h) = c\left(1 - \exp^{-\frac{|h|}{a}}\right)$$

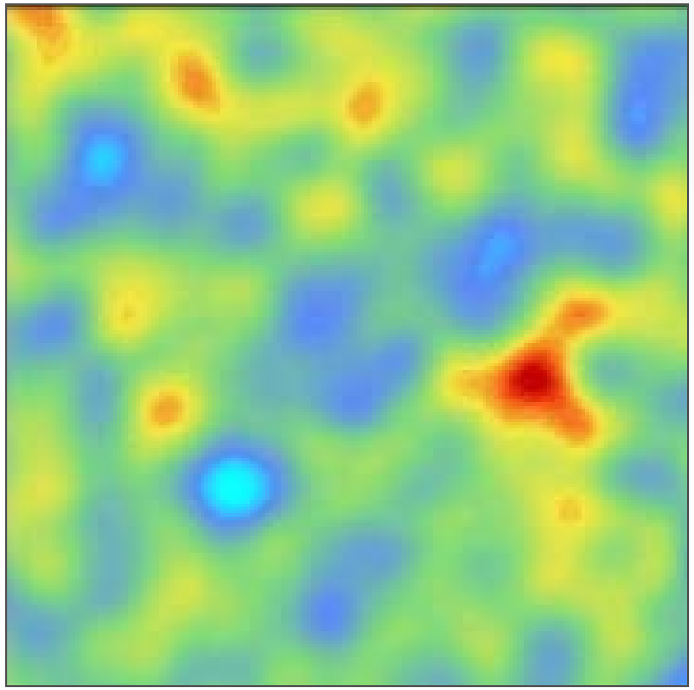


Usual Models

Gaussian

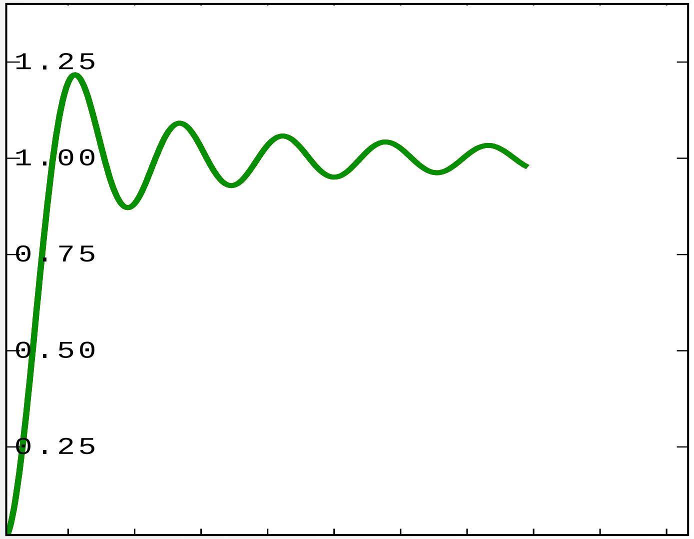


$$\gamma(h) = c\left(1-\exp^{-\left(\frac{h}{a}\right)^2}\right)$$

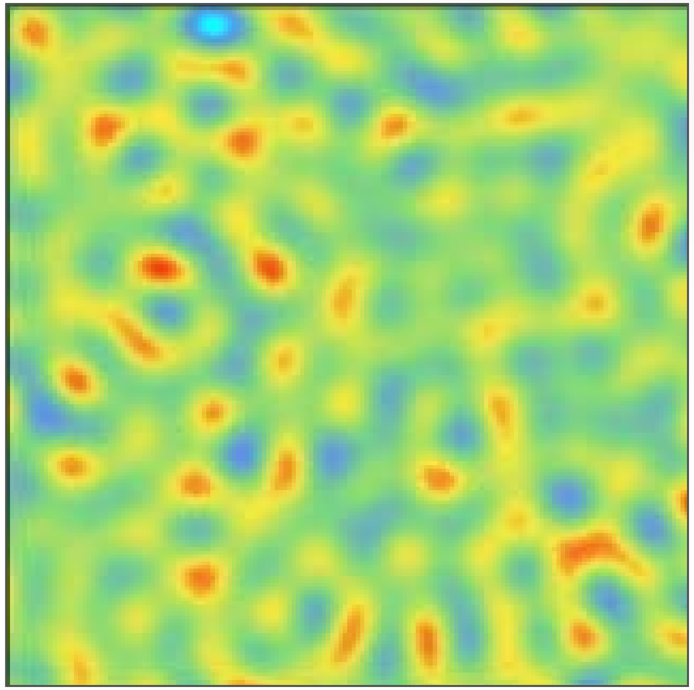


Usual Models

Cardinal Sine

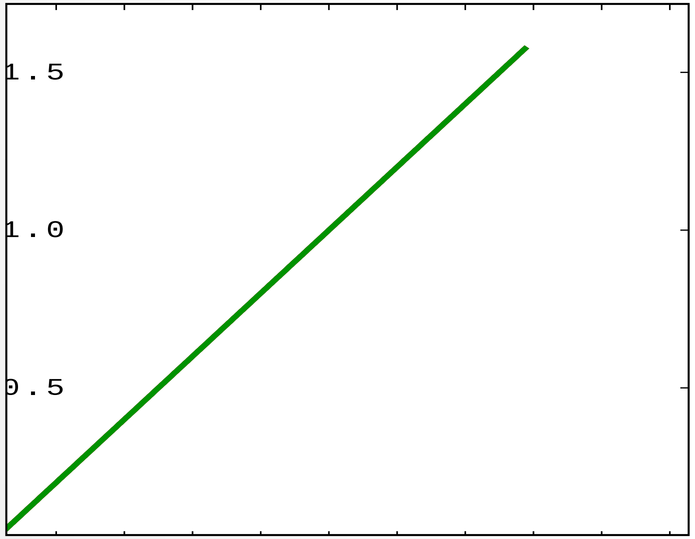


$$\gamma(h) = c\frac{\sin(h/a)}{h/a}$$

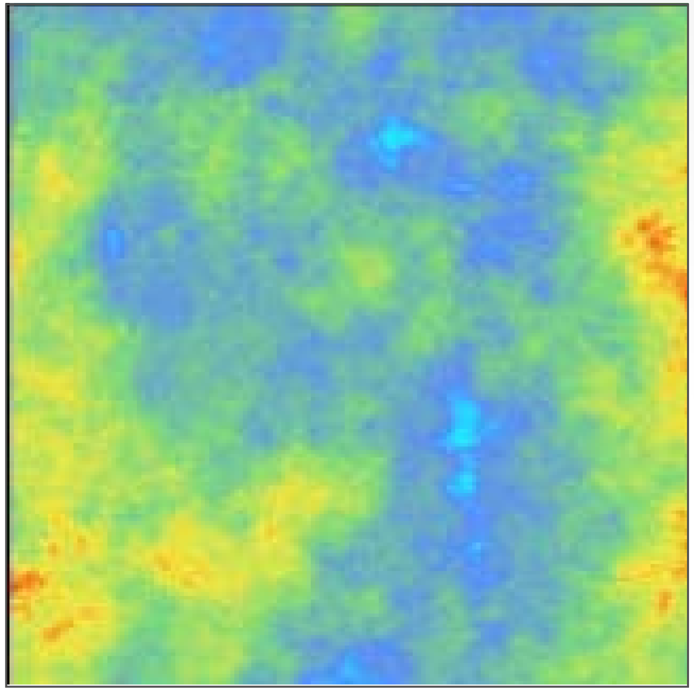


Usual Models

Linear



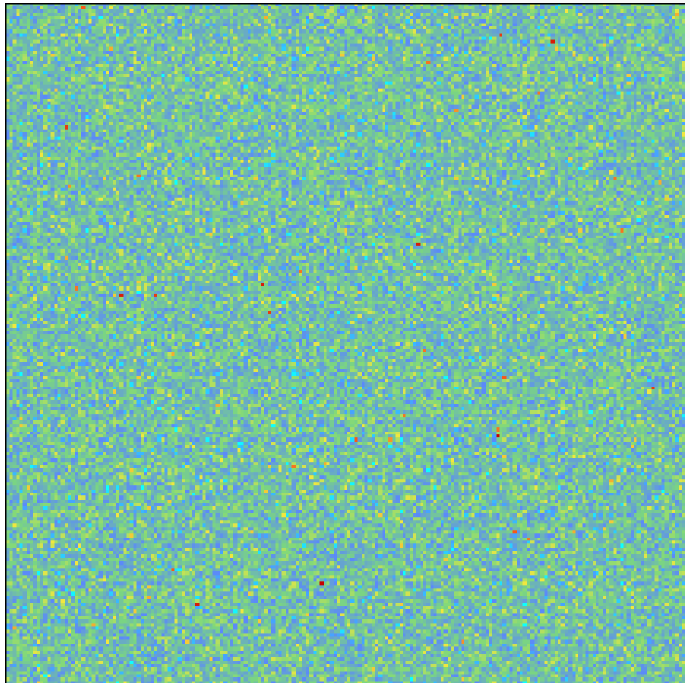
$$\gamma(h) = c\frac{|h|}{a}$$



Usual Models

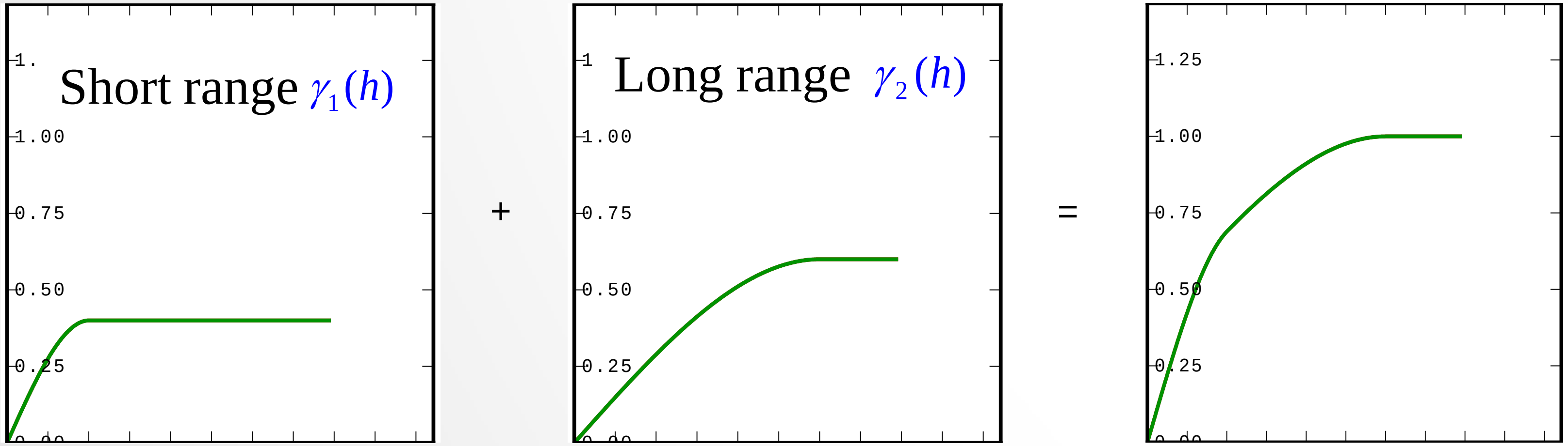
Nugget Effect

$$\gamma(h) \quad = 0 \quad h = 0$$
$$\quad\quad\; = c \quad h > 0$$



Nested Variograms: Definition

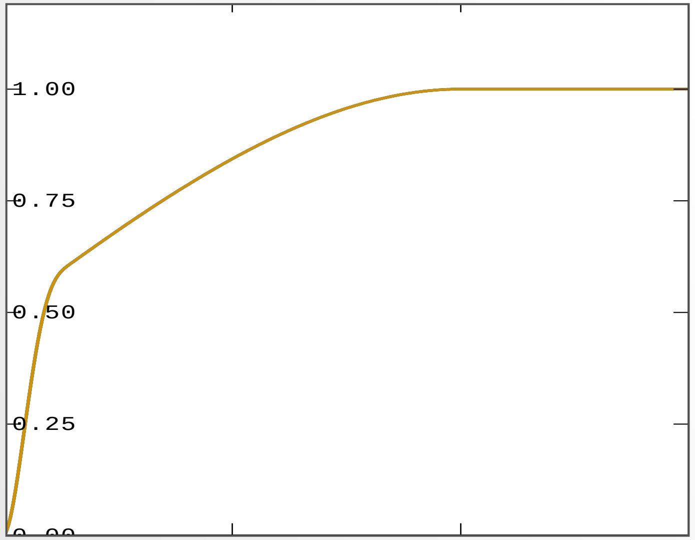
Nesting variograms = Adding values for each distance

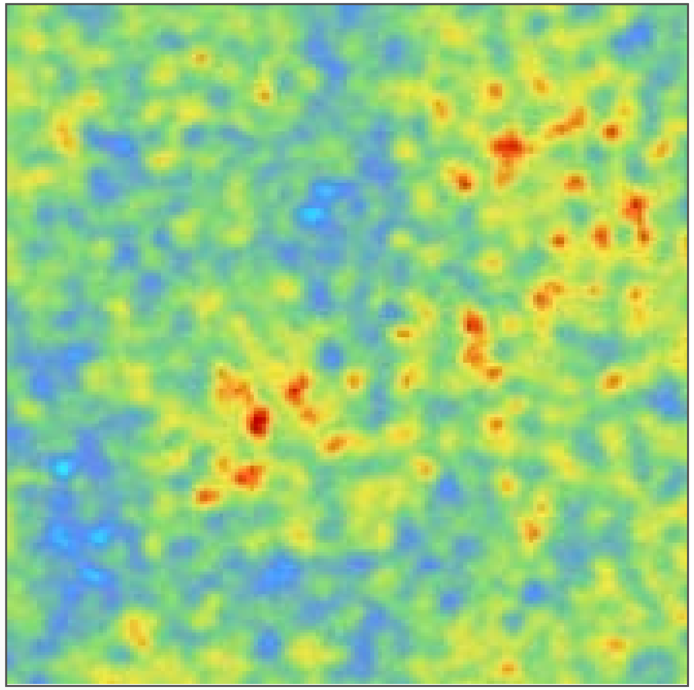


$$\gamma(h) = \gamma_1(h) + \gamma_2(h)$$

Nested Variograms: Example

Cubic (short range) + Spherical (long range)

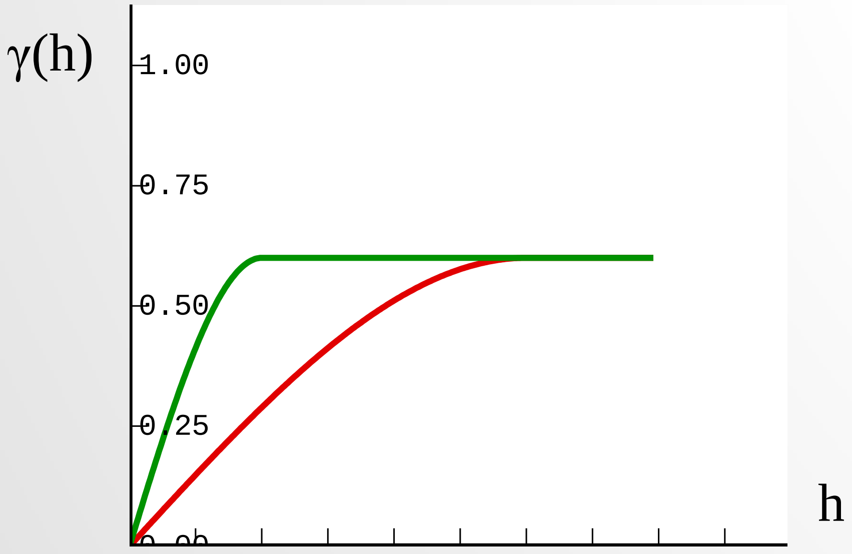




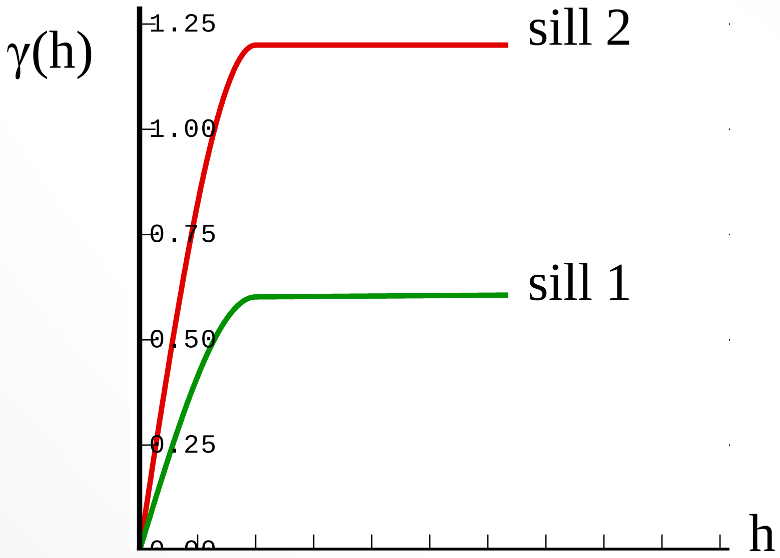
Anisotropy: Definition

Two types of Anisotropies:

Geometrical anisotropy

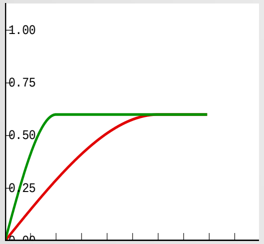


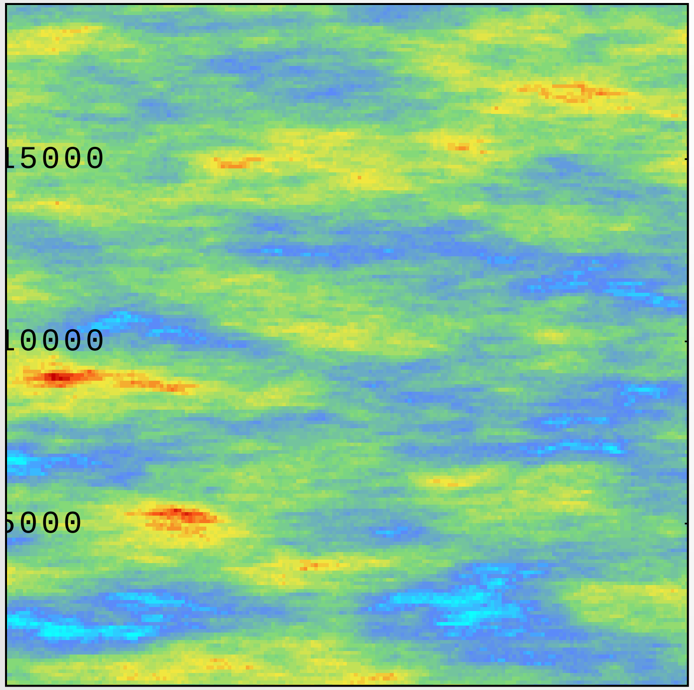
Zonal anisotropy



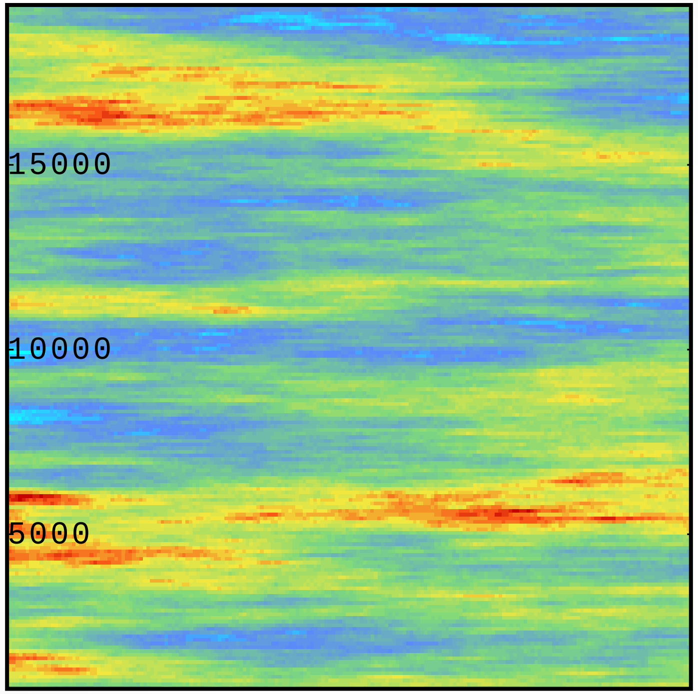
Anisotropy: Examples

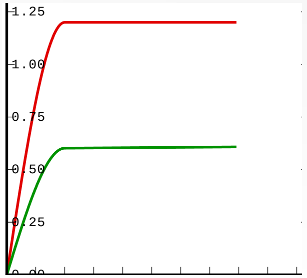
Geometrical Anisotropy





Zonal Anisotropy





Variances

Calculate the **variance** of a linear combination: $\sum_i \lambda_i Z_i$

- Using the Covariance $C(h)$

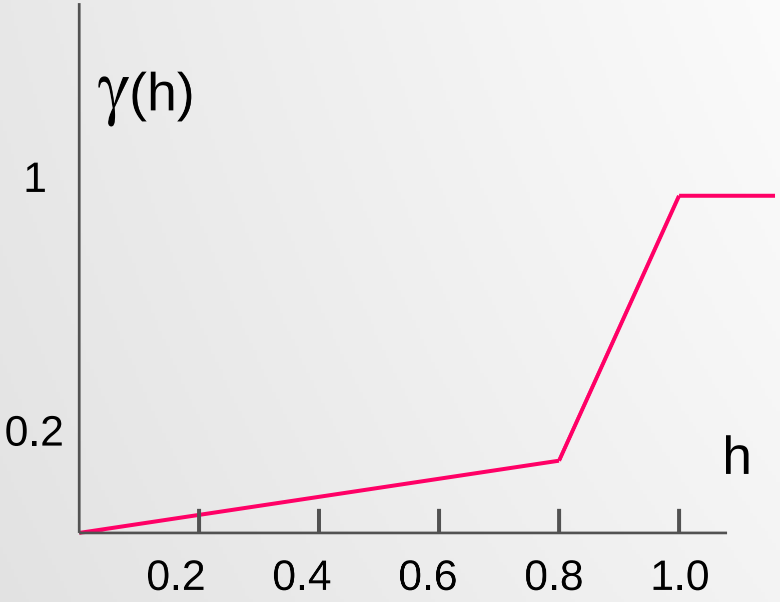
$$Var\left(\sum_i \lambda_i Z_i\right) = \sum_i \sum_j \lambda_i \lambda_j C\left(h_{ij}\right) = \sum_i \sum_j \lambda_i \lambda_j C_{ij} \geq 0$$

- Using the Variogram $\gamma(h)$

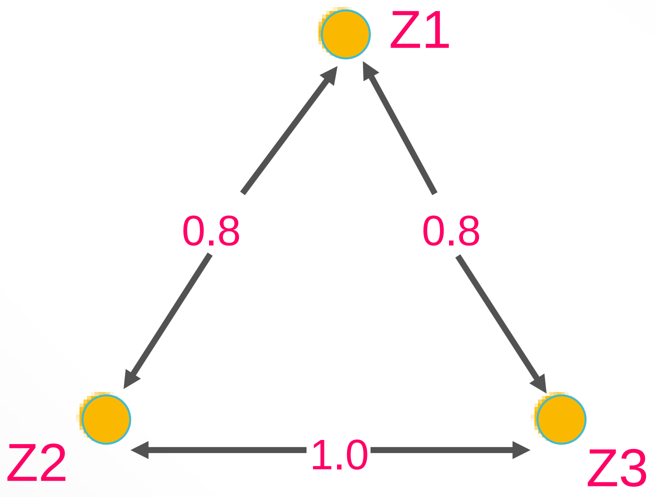
$$Var\left(\sum_i \lambda_i Z_i\right) = -\sum_i \sum_j \lambda_i \lambda_j \gamma\left(h_{ij}\right) = -\sum_i \sum_j \lambda_i \lambda_j \gamma_{ij} \geq 0$$

Why using authorized variogram

Calculate the Variance of $Z_1 - 1/2\, Z_2 - 1/2\, Z_3$



Samples



Why using authorized variogram

Calculate the Variance of $Z_1 - \frac{1}{2}Z_2 - \frac{1}{2}Z_3 = \sum_i \lambda_i Z_i$

- Check the sum of weights: $\lambda_1 + \lambda_2 + \lambda_3 = (1) + \left(-\frac{1}{2}\right) + \left(-\frac{1}{2}\right) = 0$

- Variance:

$$
\begin{aligned}
Var &= -\sum_i \sum_j \lambda_i \lambda_j \gamma_{ij} \\
&= -\left(\lambda_1^2 \gamma_{11} + \lambda_2^2 \gamma_{22} + \lambda_3^2 \gamma_{33} + 2\lambda_1 \lambda_2 \gamma_{12} + 2\lambda_1 \lambda_3 \gamma_{13} + 2\lambda_2 \lambda_3 \gamma_{23}\right) \\
&= -\left(2 \times (1) \times \left(-\frac{1}{2}\right) + 2 \times (1) \times \left(-\frac{1}{2}\right) + 2 \times \left(-\frac{1}{2}\right) \times \left(-\frac{1}{2}\right)\right)
\end{aligned}
$$

$$Var = -0.1$$

Hints for Modeling

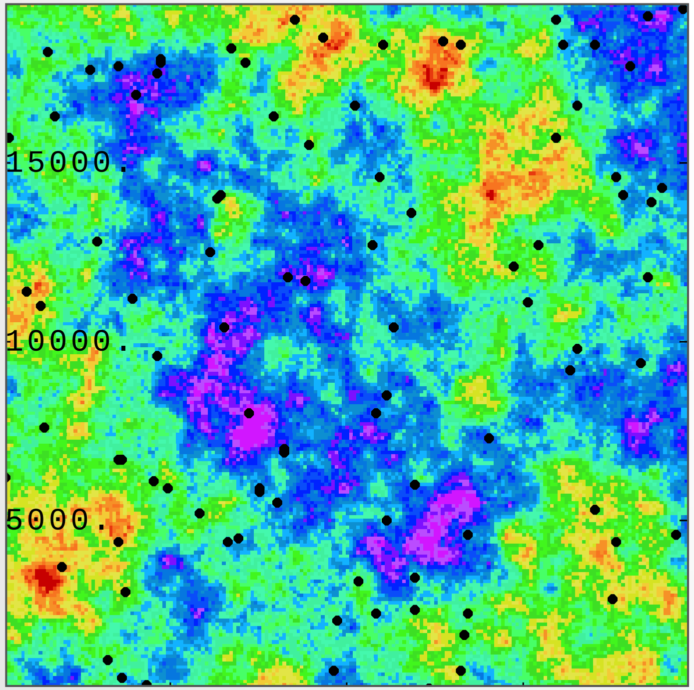
- The Experimental variogram must be fitted by the Model:
  - For any distance
  - For any direction
- The Model (if using authorized basic structures) ensures the positivity of the variance of any linear combination of the data:
  - Compulsory for the next step: Estimation by Kriging

Back to Jupyter Notebook!

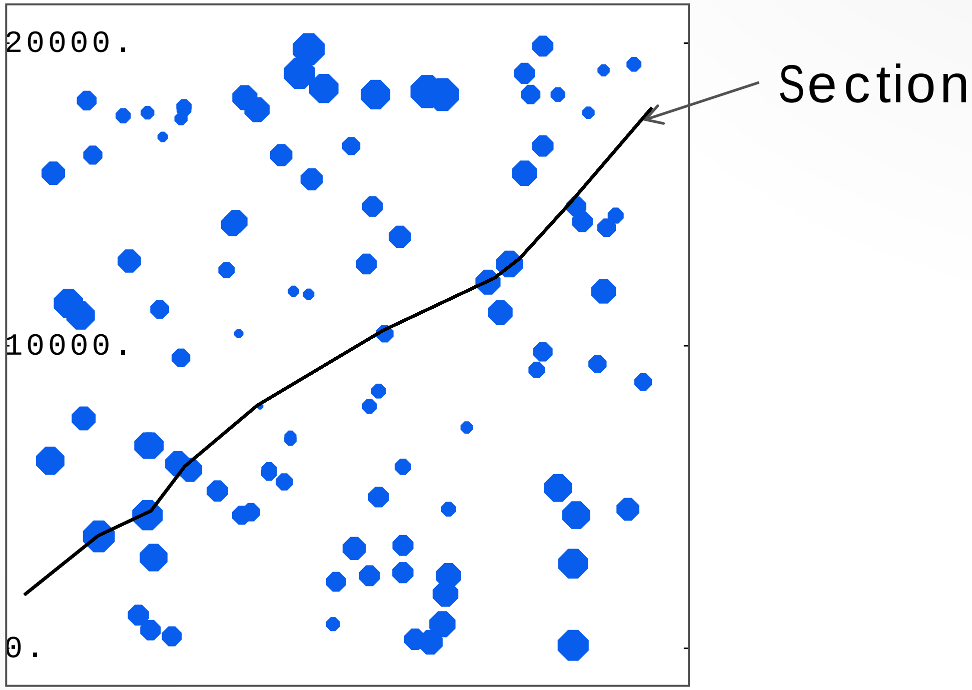
- We focus on the cells from **Variography** section:
    - Temperature 2-D Omni-directional Experimental Variogram
    - 4 Directions Experimental Variograms
    - Variogram Model Fitting

Estimation

Consider an exhaustive data set and extract 100 samples randomly located



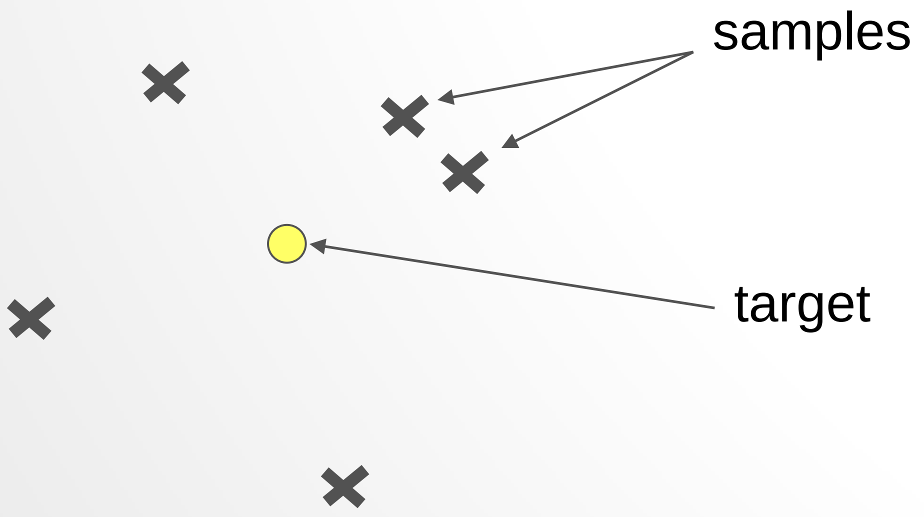
Exhaustive Data Set



Sampled Data Set

Estimation

Estimation at target site as a linear combination of the sample values

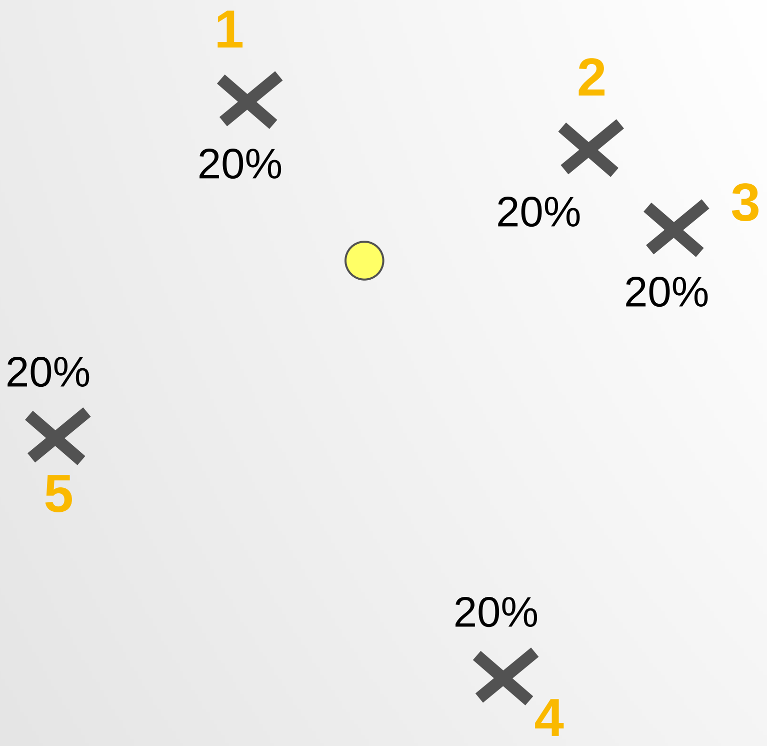


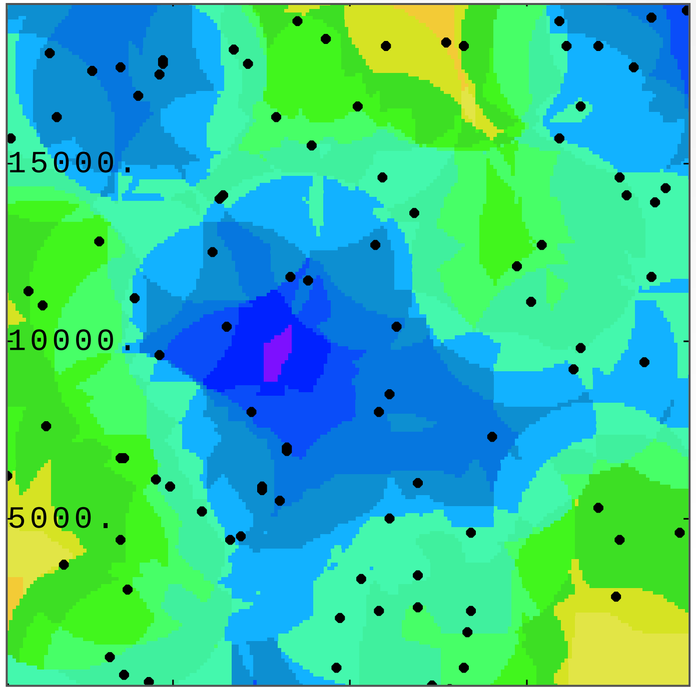
$$Z^* = \sum_i \lambda_i Z_i$$

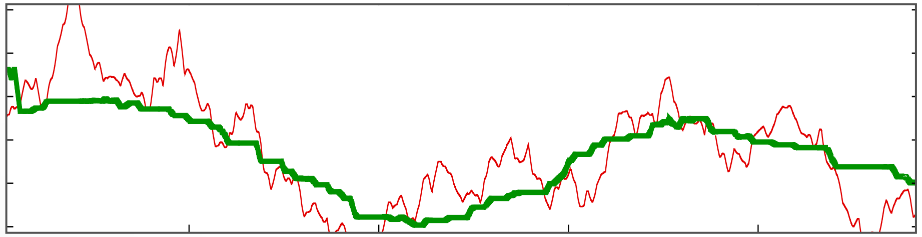
Estimation

Moving Average

$$Z^* = \sum_i \frac{Z_i}{5}$$



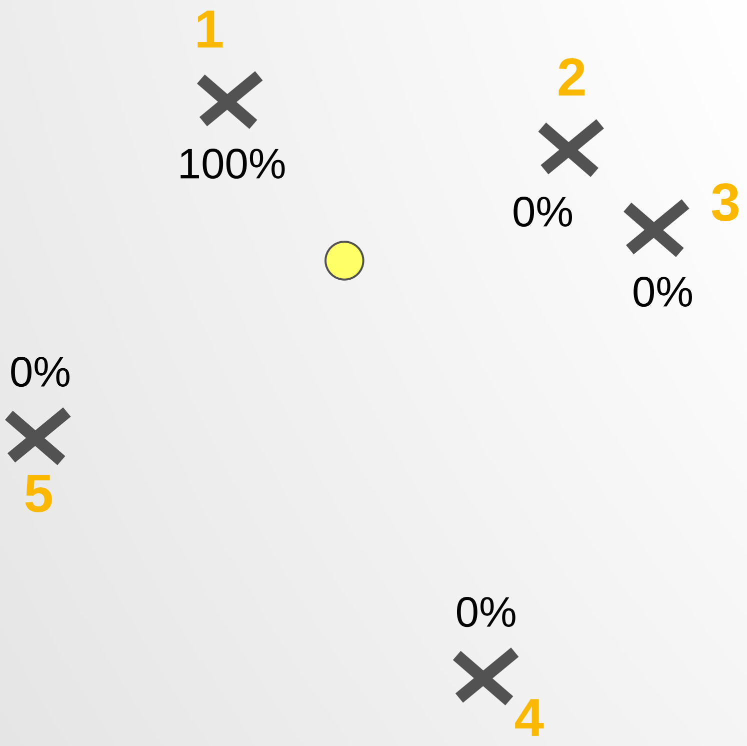


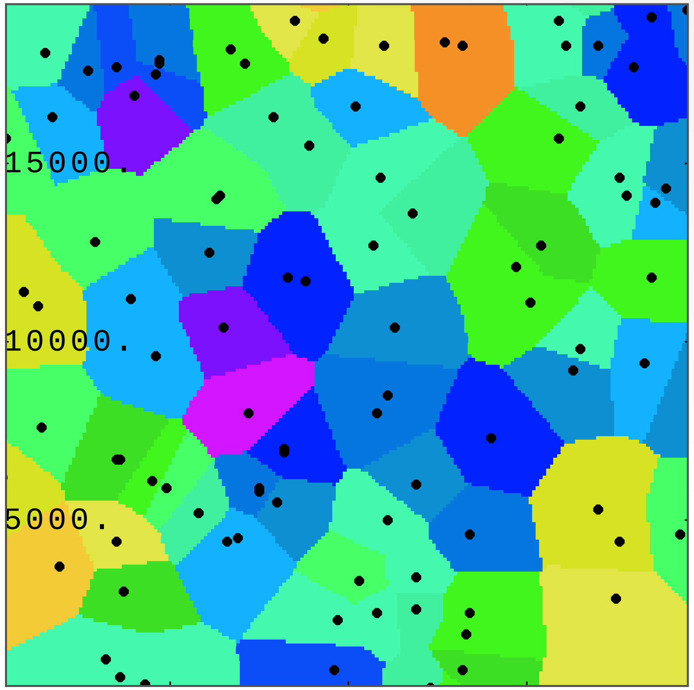


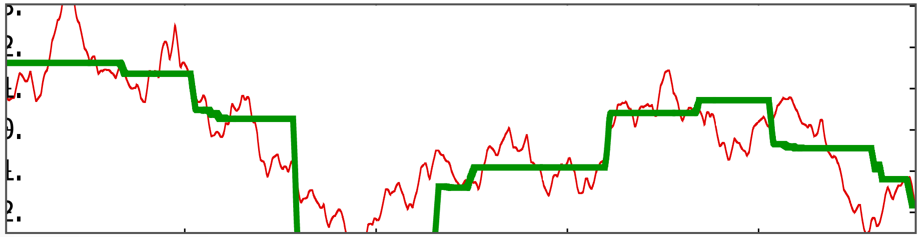
Estimation

Influence Polygon

$$Z^* = Z_1$$



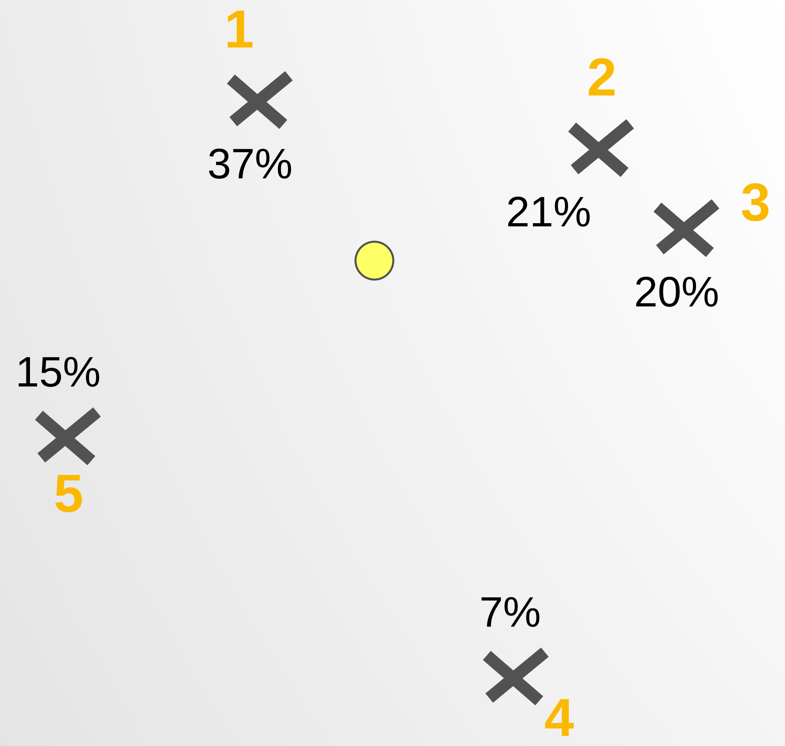


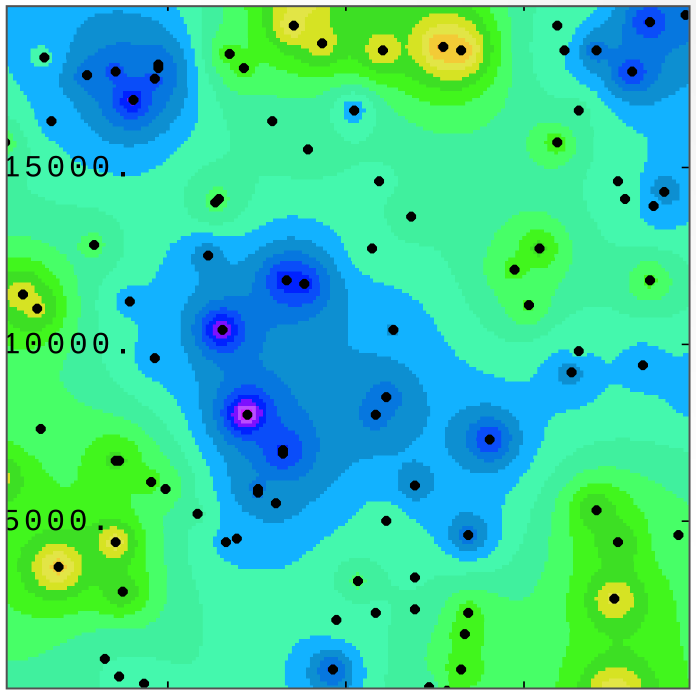


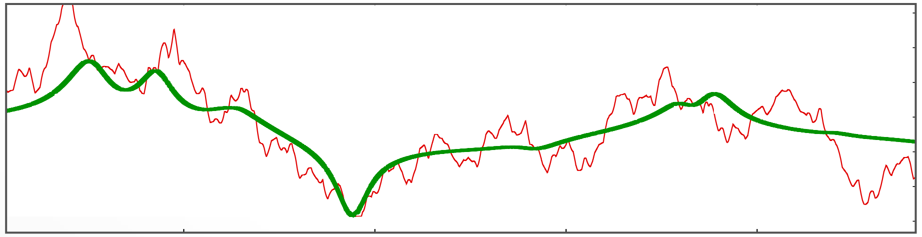
Estimation

Inverse Distance

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







Kriging: Principle

- Kriging produces the estimation Z* of the variable Z on a target site (point, block, polygon) as a linear combination of the sample values

$$Z_0^* = \lambda_0 + \sum_i \lambda_i Z_i$$

- The real unknown value is denoted Z_0
- The estimation error: $\varepsilon = Z_0 - Z_0^* = Z_0 - \sum_i \lambda_i Z_i - \lambda_0$
  - is a linear combination of the data
  - authorized
  - with a zero expectation (non bias)
  - with minimum variance (optimality)
- The constraints are strictly ordered

Simple Kriging

- Stationary hypothesis – **Simple Kriging** - **Known mean**: $E(Z) = m$
- Non bias:

$$E(\varepsilon) = E\left(Z_0 - \sum_i \lambda_i Z_i - \lambda_0 \right) = m - \sum_i \lambda_i \times m - \lambda_0 = 0 \Rightarrow \lambda_0 = m\left(1 - \sum_i \lambda_i \right)$$

- Optimality:

$$Var(\varepsilon) = Var\left(Z_0 - \sum_i \lambda_i Z_i - \lambda_0 \right) = C_{00} - 2\sum_i \lambda_i C_{0i} + \sum_i \sum_j \lambda_i \lambda_j C_{ij} \quad \text{minimum}$$

$$\Rightarrow \frac{\partial Var(\varepsilon)}{\partial \lambda_i} = \sum_j \lambda_j C_{ij} - C_{0i} = 0 \qquad \forall i$$

- Results:

$$\begin{cases} Z_0^* = \sum_i \lambda_i Z_i + m\left(1 - \sum_i \lambda_i \right) \\ Var(\varepsilon) = C_{00} - \sum_i \lambda_i C_{0i} \end{cases}$$

Simple Kriging

In algebraic terms

- Simple Kriging System:

$$\begin{bmatrix} C_{11} & \cdots & C_{1n} \\ \vdots & \ddots & \vdots \\ C_{n1} & \cdots & C_{nn} \end{bmatrix} \times \begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{bmatrix} = \begin{bmatrix} C_{10} \\ \vdots \\ C_{n0} \end{bmatrix}$$

- Estimation:

$$Z_0^* = [\lambda_1 \quad \cdots \quad \lambda_n] \bullet \begin{bmatrix} Z_1 \\ \vdots \\ Z_n \end{bmatrix} + m\left(1 - \sum_i \lambda_i\right)$$

- Variance of Estimation Error:

$$Var(\varepsilon) = C_{00} - [\lambda_1 \quad \cdots \quad \lambda_n] \bullet \begin{bmatrix} C_{10} \\ \vdots \\ C_{n0} \end{bmatrix}$$

Simple Kriging

- Correlation between estimation and estimation error:

$$Cov(\varepsilon, Z_0^*) = Cov(Z_0^* - Z_0, \varepsilon) = Var(Z_0^*) - Cov(Z_0, Z_0^*) = \sum_i \sum_j \lambda_i \lambda_j C_{ij} - \sum_i \lambda_i C_{i0} = 0$$

- Re-writing the definition of the estimation error:

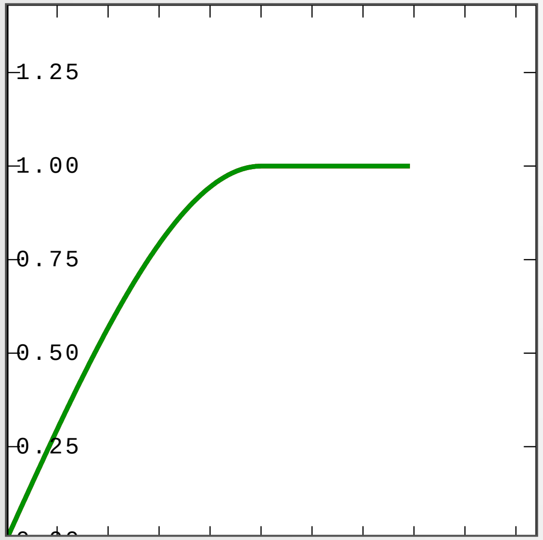
$$\varepsilon = Z_0 - Z_0^* \quad \Leftrightarrow \quad Z_0 = \varepsilon + Z_0^*$$

- In variances: $Var(Z_0) = Var(Z_0^*) + Var(\varepsilon) + 2Cov(\varepsilon, Z_0^*) = Var(Z_0^*) + Var(\varepsilon)$
- Finally: $\begin{cases} Var(\varepsilon) \leq Var(Z_0) \\ Var(Z_0^*) \leq Var(Z_0) \end{cases}$

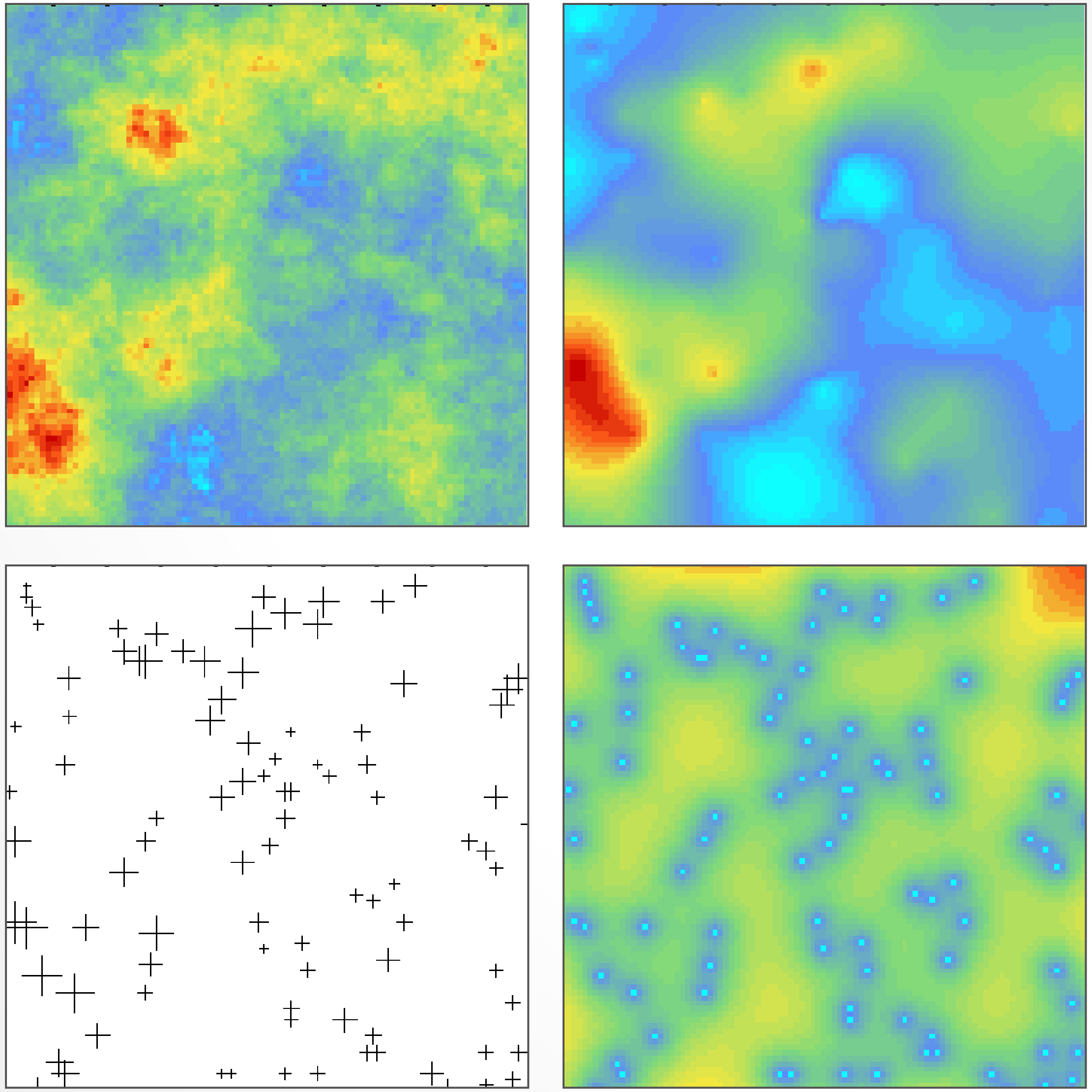
1. Kriging gives more accurate results than Statistics
2. Kriging is **smoothing** reality

Simple Kriging: smoothing

Spherical

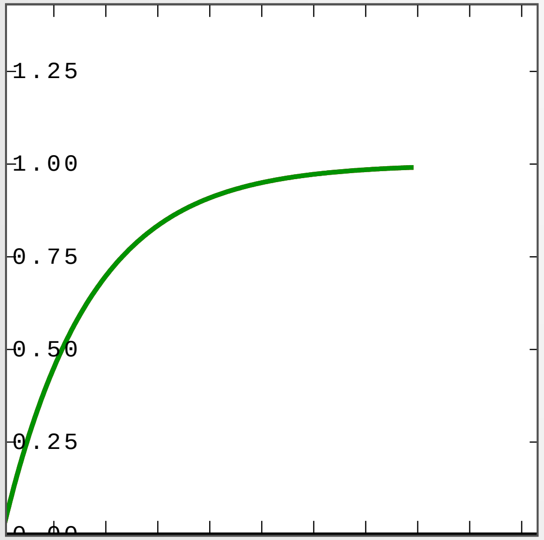


Reality	Kriging
Sampling	St. Dev.

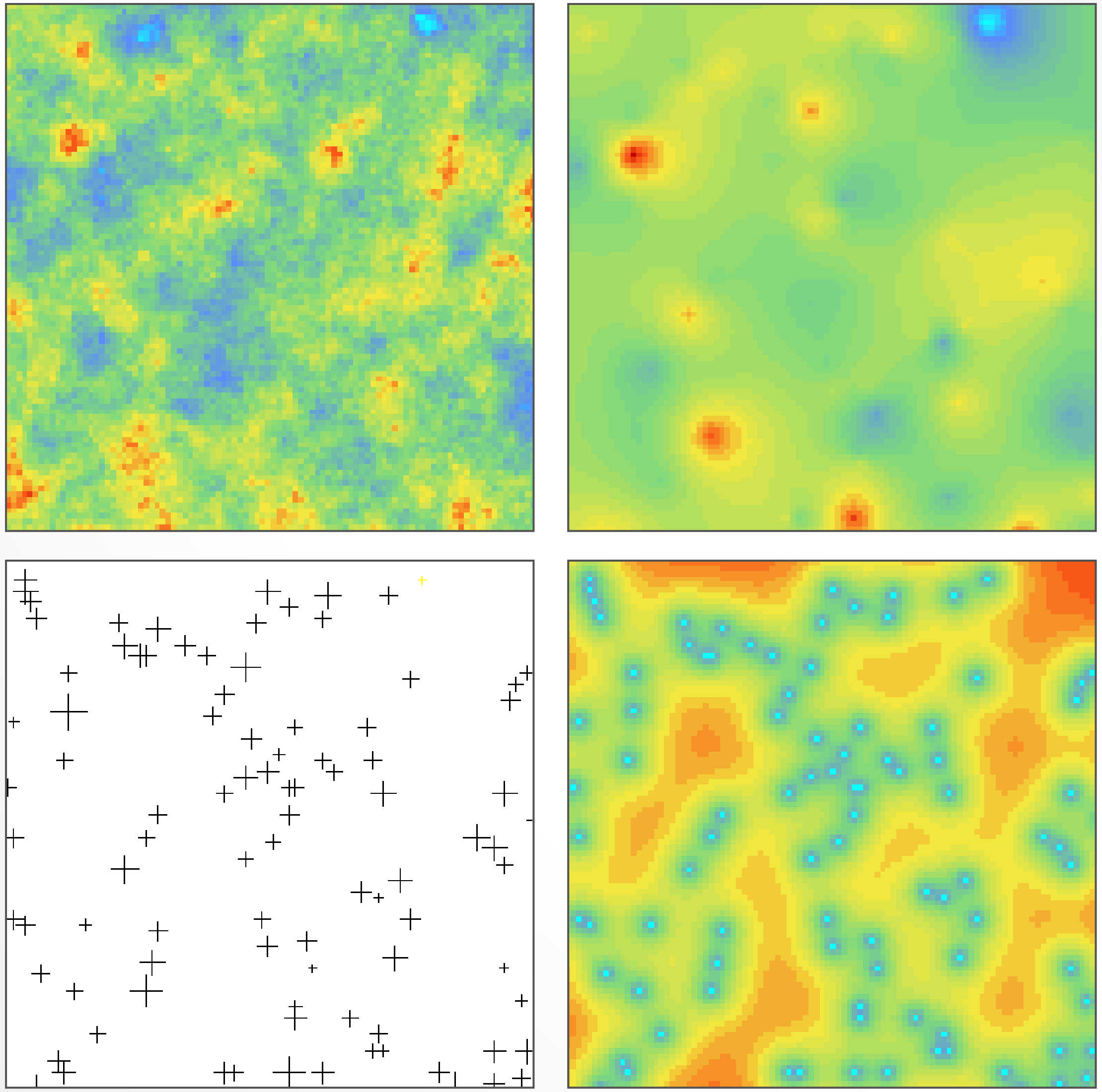


Simple Kriging: smoothing

Exponential

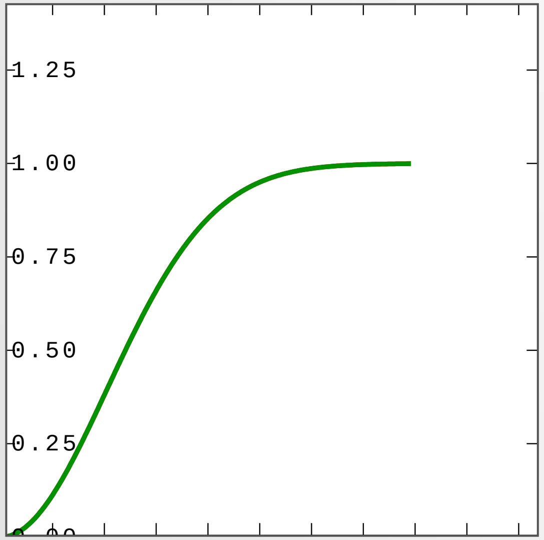


Reality	Kriging
Sampling	St. Dev.

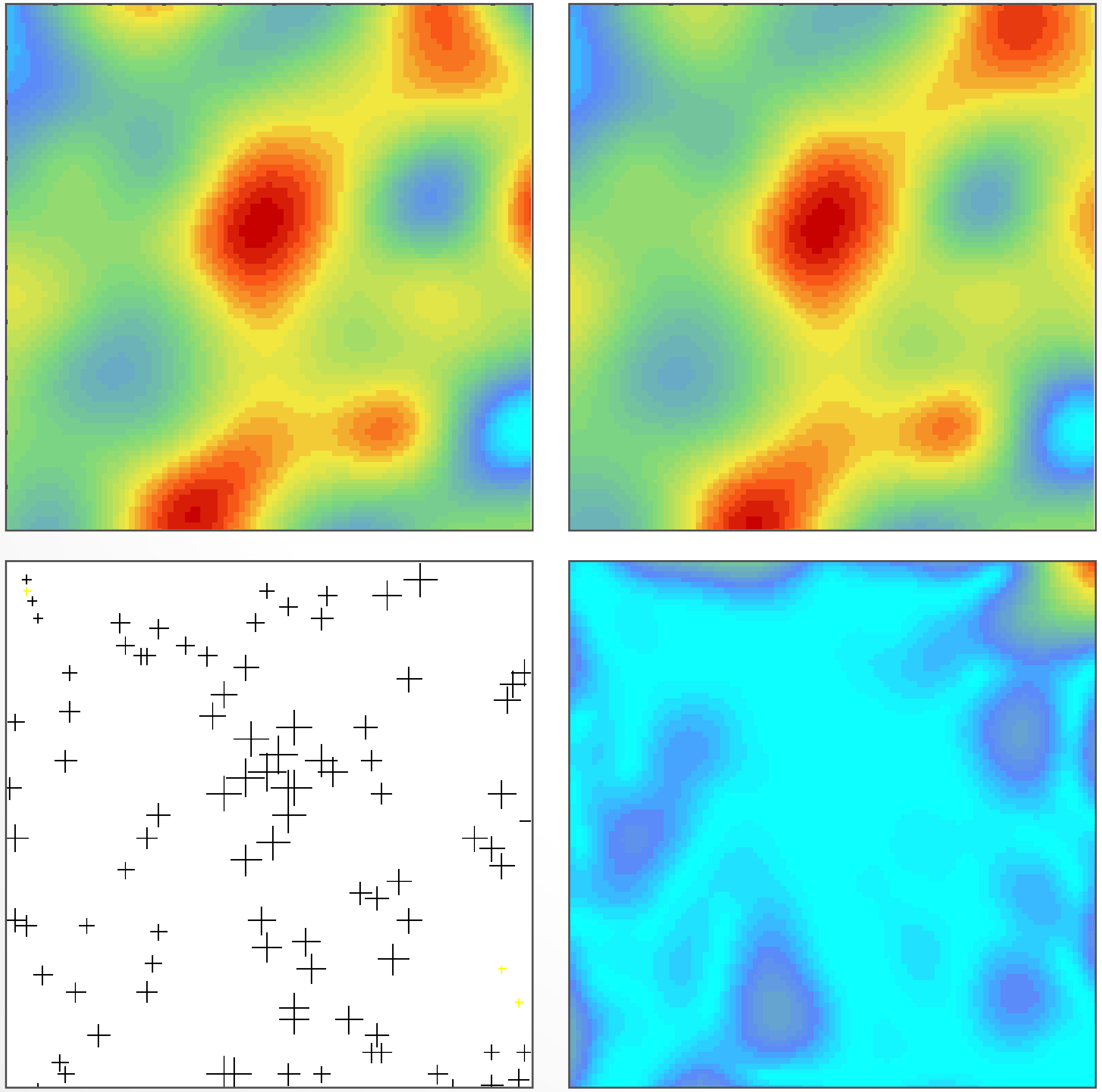


Simple Kriging: smoothing

Gaussian

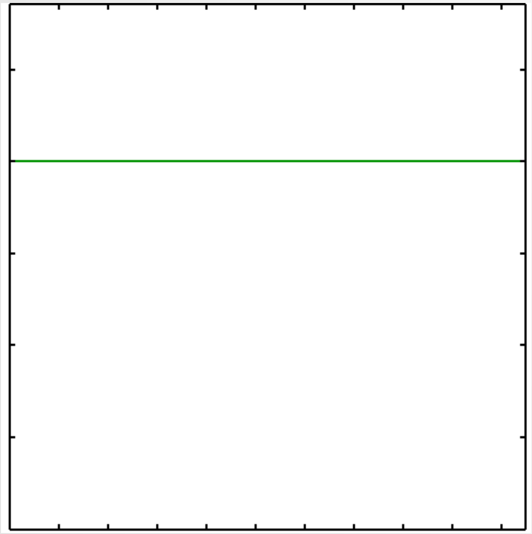


Reality	Kriging
Sampling	St. Dev.

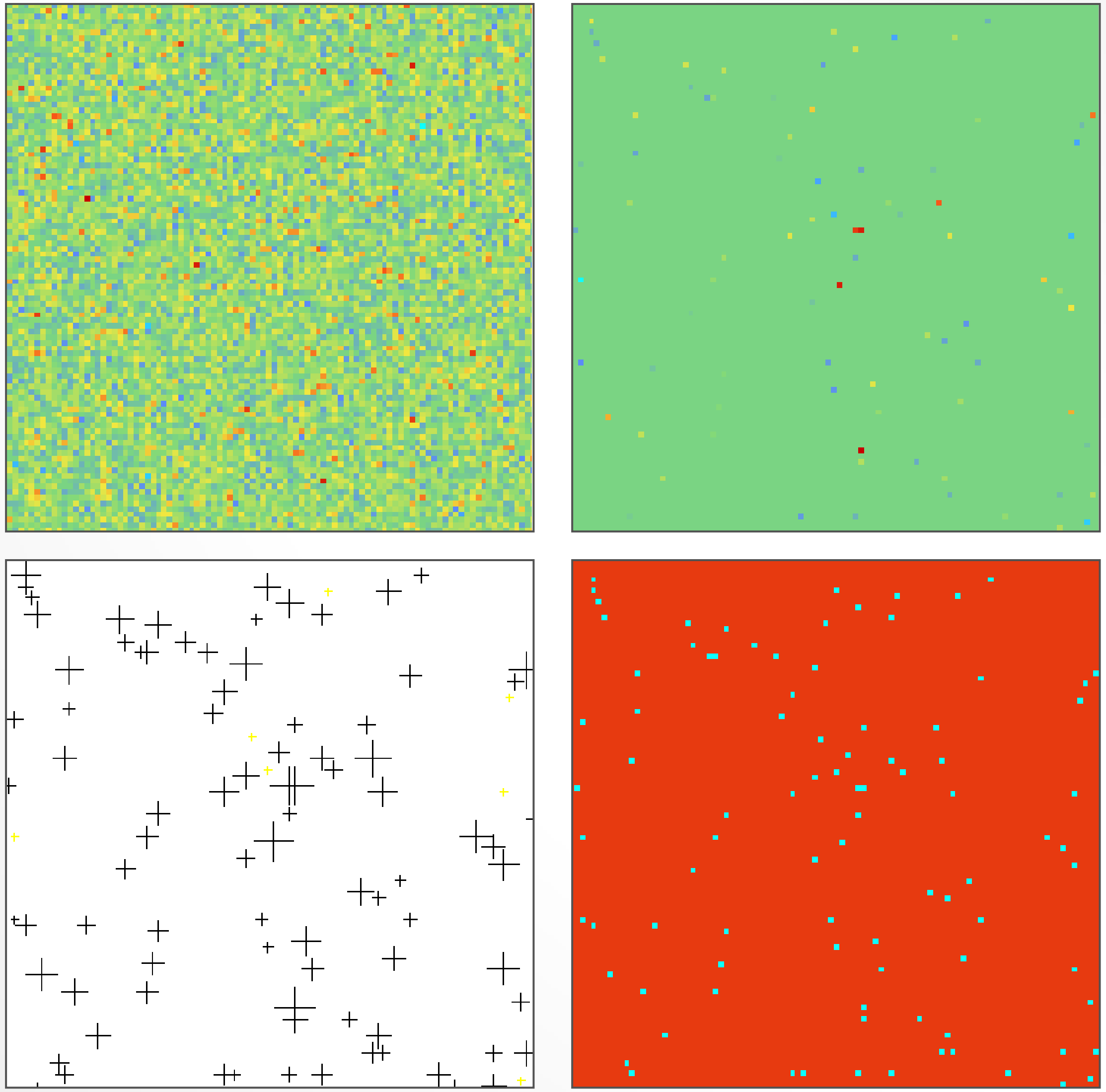


Simple Kriging: smoothing

Nugget Effect

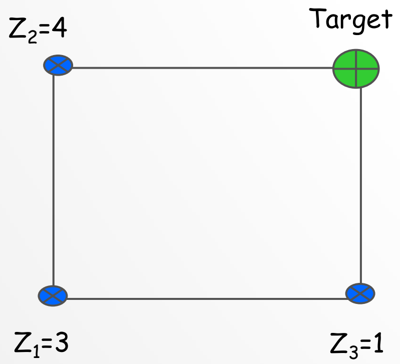


Reality	Kriging
Sampling	St. Dev.



Simple Kriging: Exercise

- Spherical Model with range 1.25m and sill 2
- 3 Data and Target on a square pattern (mesh = 1m)
- Known mean = 2



Simple Kriging: Solution

- Simple Kriging System

$$\begin{bmatrix} C_{11} & C_{12} & C_{13} \\ C_{21} & C_{22} & C_{23} \\ C_{31} & C_{32} & C_{33} \end{bmatrix} \times \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \end{bmatrix} = \begin{bmatrix} C_{10} \\ C_{20} \\ C_{30} \end{bmatrix}$$

$$C(0) = 2$$
$$C(1) = 0.112$$
$$C(\sqrt{2}) = 0$$

$$\begin{bmatrix} 2 & 0.112 & 0.112 \\ 0.112 & 2 & 0 \\ 0.112 & 0 & 2 \end{bmatrix} \times \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \end{bmatrix} = \begin{bmatrix} 0. \\ 0.112 \\ 0.112 \end{bmatrix}$$

- Results

$$\lambda_1 = -0.006$$
$$\lambda_2 = \lambda_3 = 0.056$$
$$\sum \lambda_i = 0.106$$

$$\Rightarrow \begin{cases} Z_0^* = [\Lambda]^t \bullet [Z] + m\left(1 - \sum_i \lambda_i\right) = 2.050 \\ Var(\varepsilon) = C_{00} - [\Lambda]^t \bullet [C_{0i}] = 1.41 \end{cases}$$

Block Kriging

Estimate the average value of Z over a block, starting from sample values:

$$Z_v^* = \lambda_0 + \sum_i \lambda_i Z_i$$

- Simple Block Kriging is (almost) identical to Simple Point Kriging, replacing Point index (◦) by Block index (v):

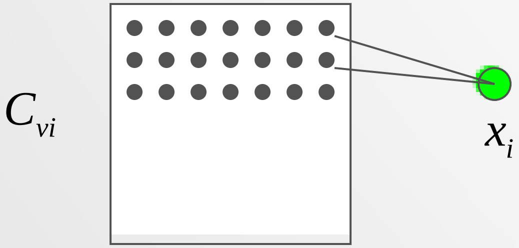
$$\sum_j \lambda_j C_{ij} = C_{0i} \qquad \forall i$$

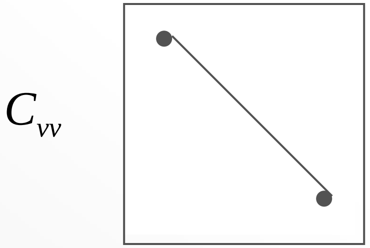
$$\sum_j \lambda_j C_{ij} = C_{vi} \qquad \forall i$$

Point $$\Rightarrow \begin{cases} Z_o^* = \sum_i \lambda_i Z_i + m\left(1 - \sum_i \lambda_i\right) \\ Var\left(\varepsilon_0\right) = C_{00} - \sum_i \lambda_i C_{0i} \end{cases}$$

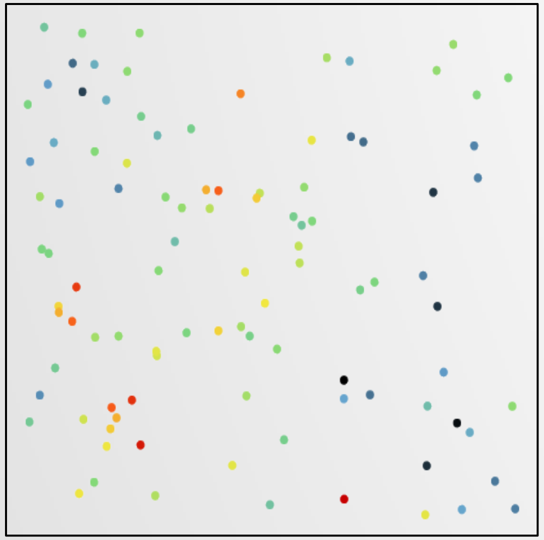
Block $$\Rightarrow \begin{cases} Z_v^* = \sum_i \lambda_i Z_i + m\left(1 - \sum_i \lambda_i\right) \\ Var\left(\varepsilon_v\right) = C_{vv} - \sum_i \lambda_i C_{vi} \end{cases}$$

- Requires calculation of block-data and block-block covariances: **discretization**





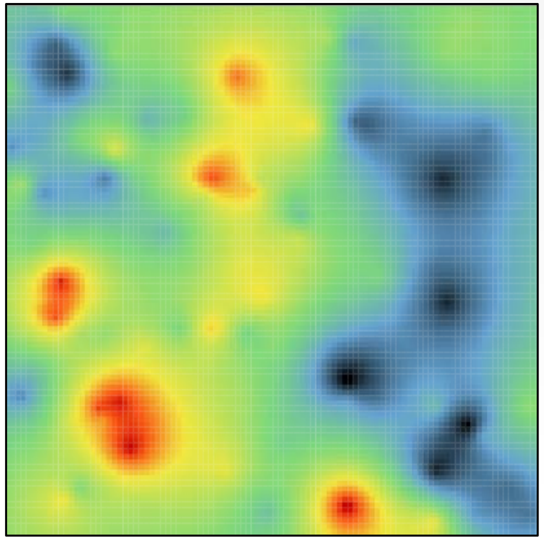
Block Kriging: Example



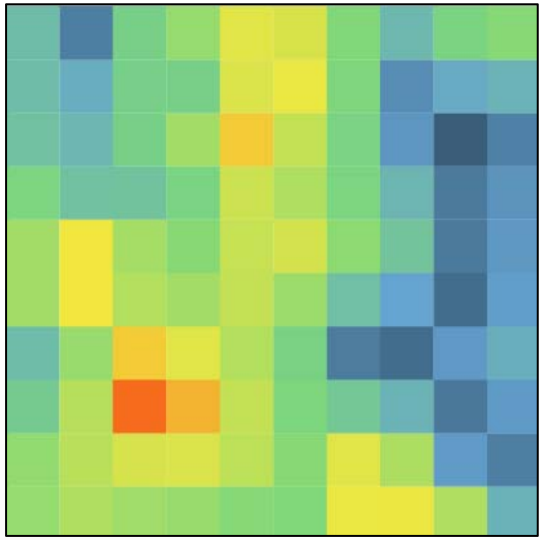
Data

$$Z_v^* = \frac{1}{|V|}\int_V Z^*(x)dx$$

Point Estimation

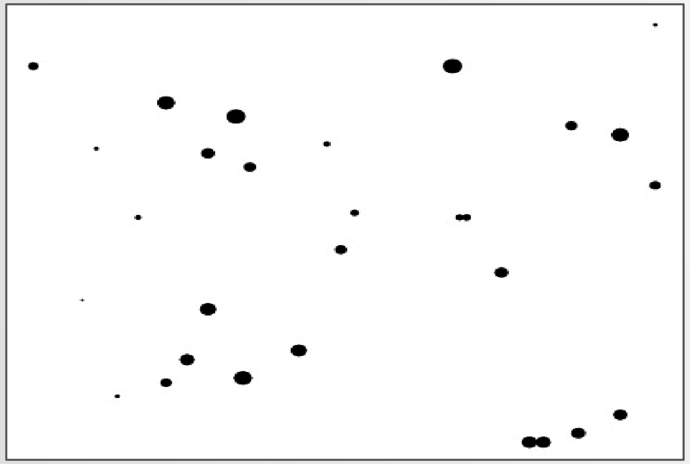


Block Estimation

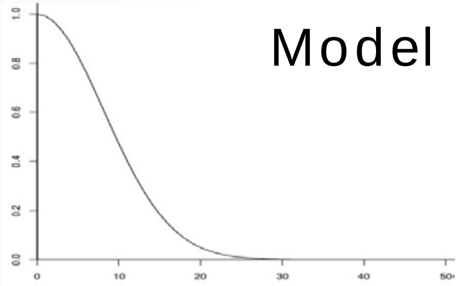


Simple Kriging: Extensions

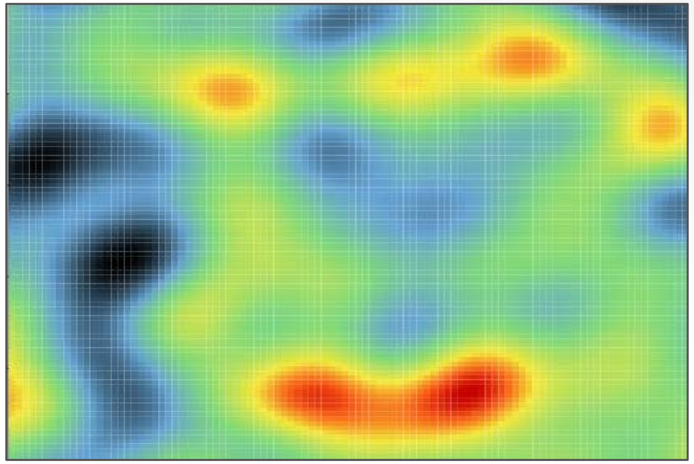
Kriging can be extended to any linear function of the data: e.g. **gradient**



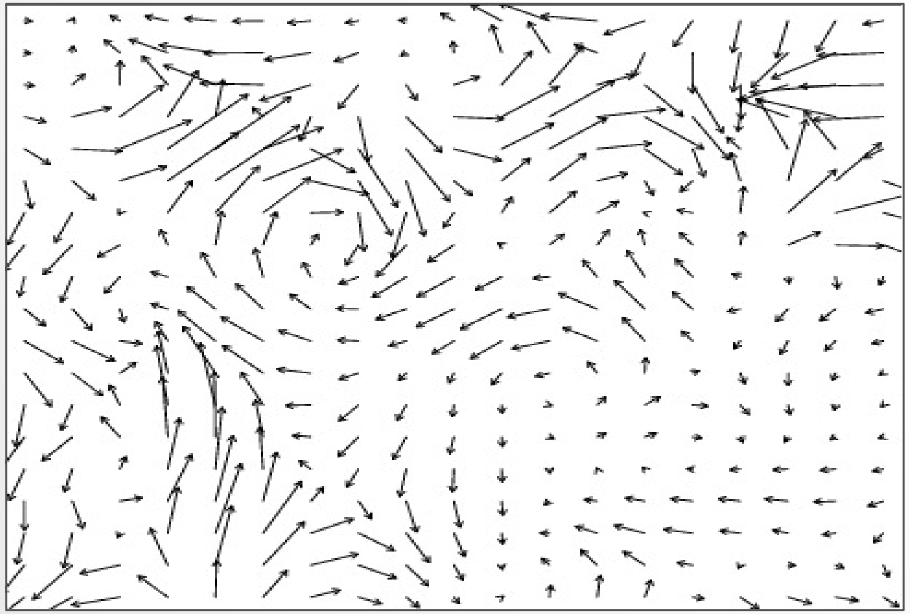
Data



Model

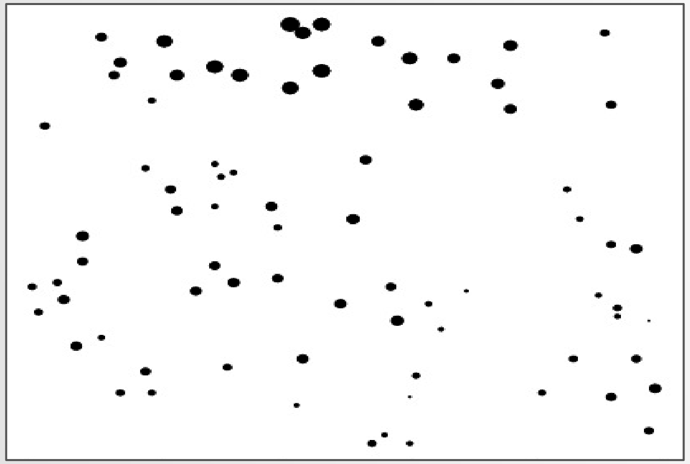


Estimation on a fine grid

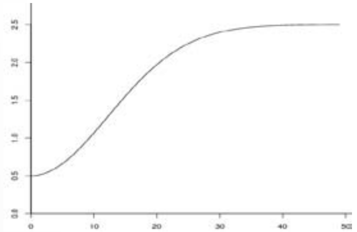


Estimated **Gradients**

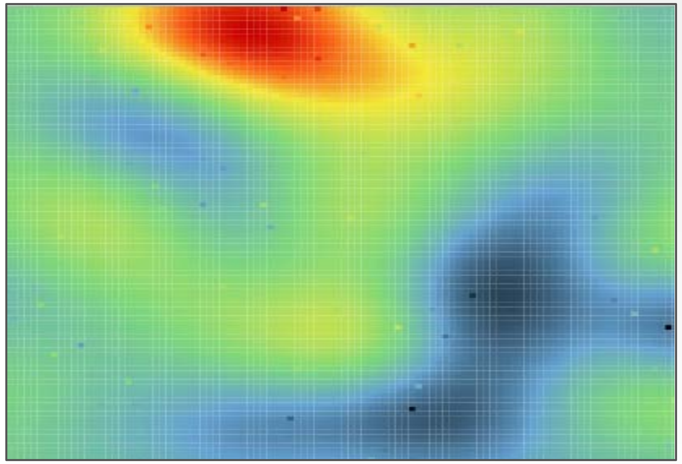
Simple Kriging: Filtering

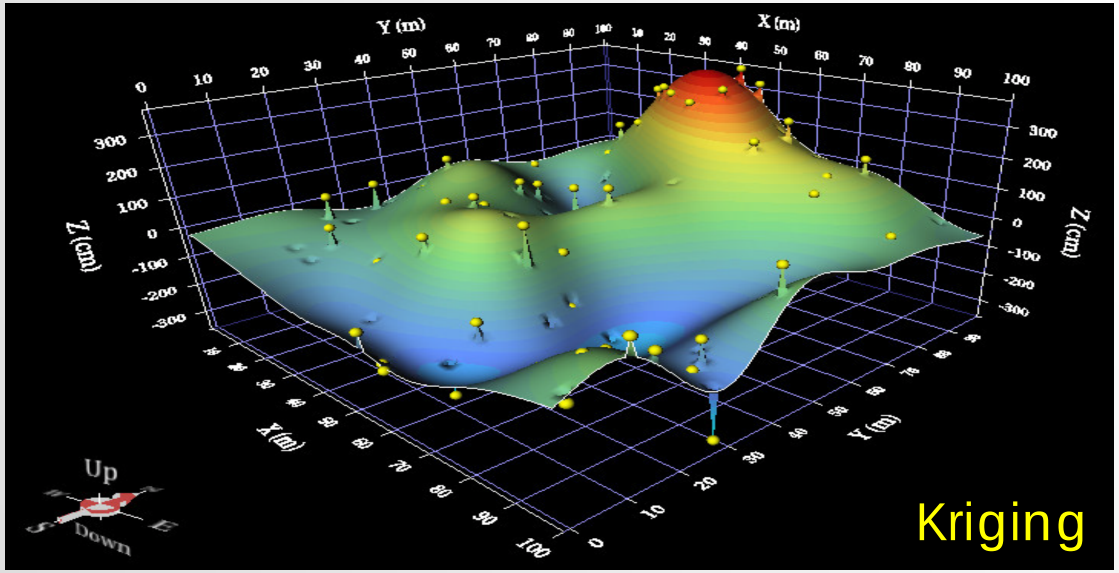


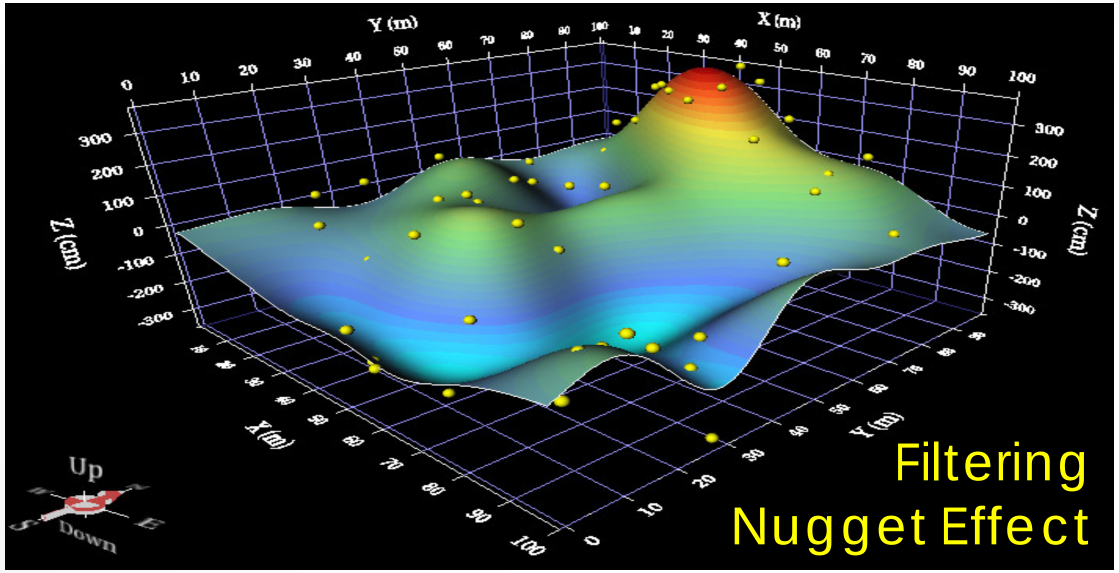
Data



Kriging







Ordinary Kriging

- Intrinsic hypothesis – **Ordinary Kriging** - **Unknown mean**: $E(Z)=m$
- Non bias: $$E(\varepsilon)=E\left(Z_0-\sum_i \lambda_i Z_i-\lambda_0\right)=m-\sum_i \lambda_i \times m-\lambda_0=0 \quad \forall m \Rightarrow \begin{cases} \sum_{\mathrm{i}} \lambda_{\mathrm{i}}=1 \\ \lambda_0=0 \end{cases}$$
- Optimality (under constraints):

$$\Phi=Var\left(Z_0-\sum_i \lambda_i Z_i-\lambda_0\right)+2\mu\left(\sum_i \lambda_i-1\right) \text{ minimum}$$

$$\Rightarrow \begin{cases} \dfrac{\partial \Phi}{\partial \lambda_i}=-\sum_j \lambda_j \gamma_{ij}+\gamma_{0i}+\mu=0 \qquad \forall i \\ \dfrac{\partial \Phi}{\partial \mu}=\sum_i \lambda_i-1=0 \end{cases}$$

- Result $$\begin{cases} Z_0^*=\sum_i \lambda_i Z_i \\ Var(\varepsilon)=\sum_i \lambda_i \gamma_{0i}-\mu \end{cases}$$

Ordinary Kriging

In algebraic terms

- Simple Kriging System:

$$\begin{bmatrix} -\gamma_{11} & \cdots & -\gamma_{1n} & 1 \\ \vdots & \ddots & \vdots & \vdots \\ -\gamma_{n1} & \cdots & -\gamma_{nn} & 1 \\ 1 & \cdots & 1 & 0 \end{bmatrix} \times \begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_n \\ \mu \end{bmatrix} = \begin{bmatrix} -\gamma_{10} \\ \vdots \\ -\gamma_{n0} \\ 1 \end{bmatrix}$$

- Estimation:

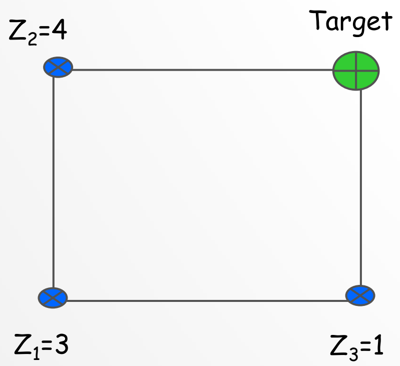
$$Z_0^* = \begin{bmatrix} \lambda_1 & \cdots & \lambda_n & \mu \end{bmatrix} \bullet \begin{bmatrix} Z_1 \\ \vdots \\ Z_n \\ 0 \end{bmatrix}$$

- Variance of Estimation Error:

$$Var(\varepsilon) = -\begin{bmatrix} \lambda_1 & \cdots & \lambda_n & \mu \end{bmatrix} \bullet \begin{bmatrix} -\gamma_{10} \\ \vdots \\ -\gamma_{n0} \\ 1 \end{bmatrix}$$

Ordinary Kriging: Exercise

- Spherical Model with range 1.25m and sill 2
- 3 Data and Target on a square pattern (mesh = 1m)



Ordinary Kriging: Solution

- Simple Kriging System

$$\begin{bmatrix} -\gamma_{11} & -\gamma_{12} & -\gamma_{13} & 1 \\ -\gamma_{21} & -\lambda_{22} & -\gamma_{23} & 1 \\ -\gamma_{31} & -\gamma_{32} & -\gamma_{33} & 1 \\ 1 & 1 & 1 & 0 \end{bmatrix} \times \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \mu \end{bmatrix} = \begin{bmatrix} -\gamma_{10} \\ -\gamma_{20} \\ -\gamma_{30} \\ 1 \end{bmatrix}$$

$$\gamma(0) = 2$$
$$\gamma(1) = 1.888$$
$$\gamma(\sqrt{2}) = 2$$

$$\begin{bmatrix} 0 & -1.888 & -1.888 \\ -1.888 & 0 & -2 \\ -1.888 & -2 & 0 \end{bmatrix} \times \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \end{bmatrix} = \begin{bmatrix} -2 \\ -1.888 \\ -1.888 \end{bmatrix}$$

- Results

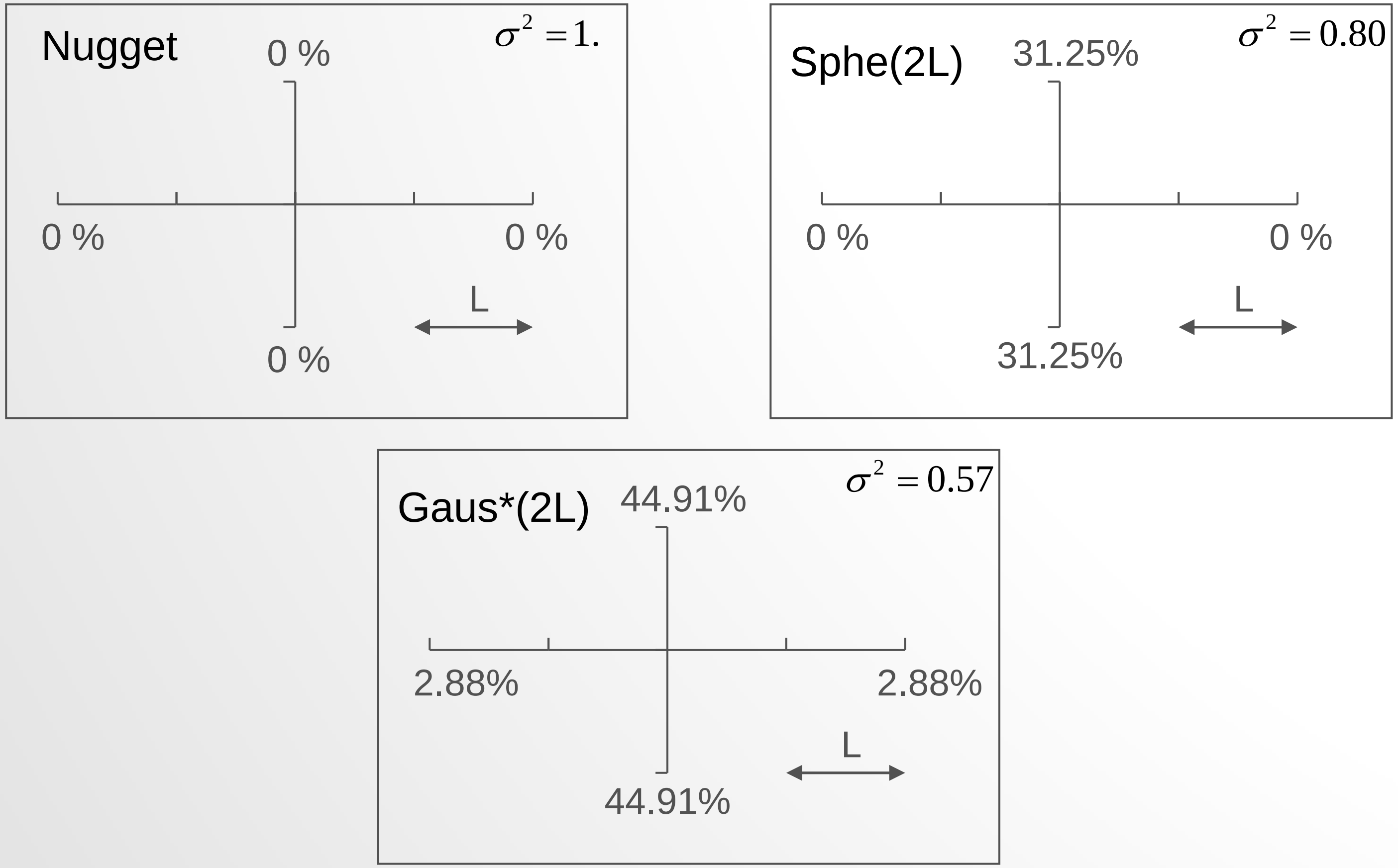
$$\lambda_1 = 0.280$$
$$\lambda_2 = \lambda_3 = 0.360$$
$$\mu = -0.6$$

$$OK \Rightarrow \begin{cases} Z_0^* = \begin{bmatrix} \Lambda \\ \mu \end{bmatrix}^t \bullet \begin{bmatrix} Z \\ 0 \end{bmatrix} = 2.640 \\ Var(\varepsilon) = C_{00} - \begin{bmatrix} \Lambda \\ \mu \end{bmatrix}^t \bullet \begin{bmatrix} C_{0i} \\ 1 \end{bmatrix} = 1.6 \end{cases}$$

$$SK \Rightarrow \begin{cases} Z_0^* = [\Lambda]^t \bullet [Z] + m\left(1 - \sum_i \lambda_i\right) = 2.050 \\ Var(\varepsilon) = C_{00} - [\Lambda]^t \bullet [C_{0i}] = 1.41 \end{cases}$$

Kriging weights

Simple Kriging of the central **Point**



Nugget $\sigma^2 = 1.$

0 %

0 %

0 %

0 %

L

Sphe(2L) $\sigma^2 = 0.80$

31.25%

0 %

0 %

31.25%

L

Gaus*(2L) $\sigma^2 = 0.57$

44.91%

2.88%

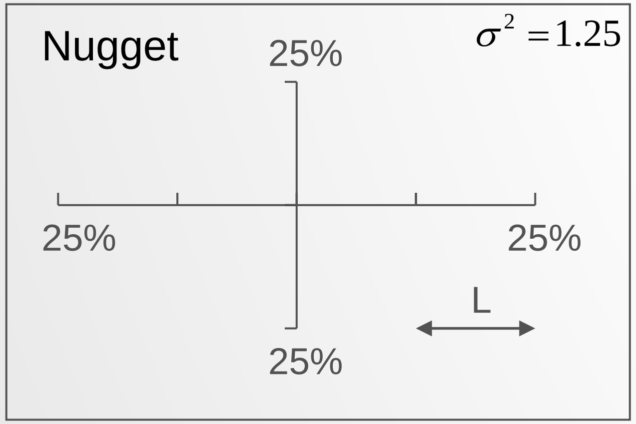
2.88%

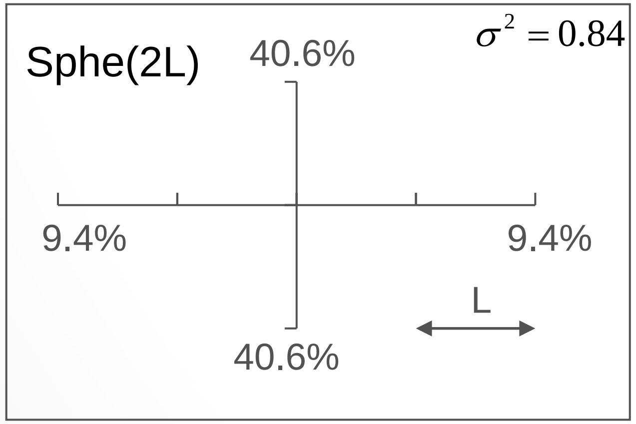
44.91%

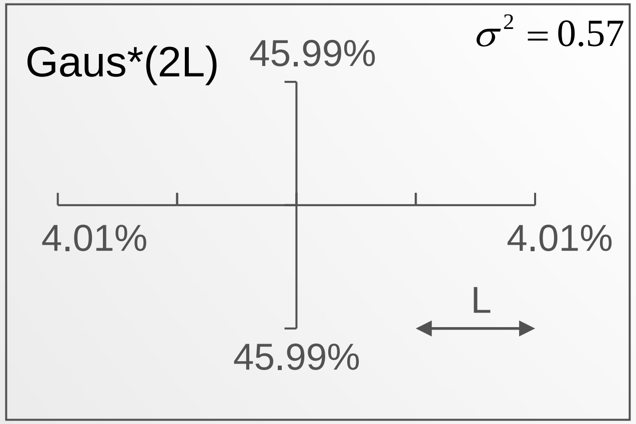
L

Kriging weights

Ordinary Kriging of the central **Point**

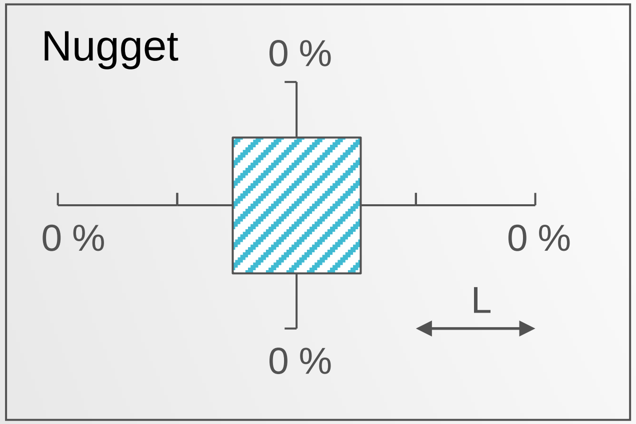


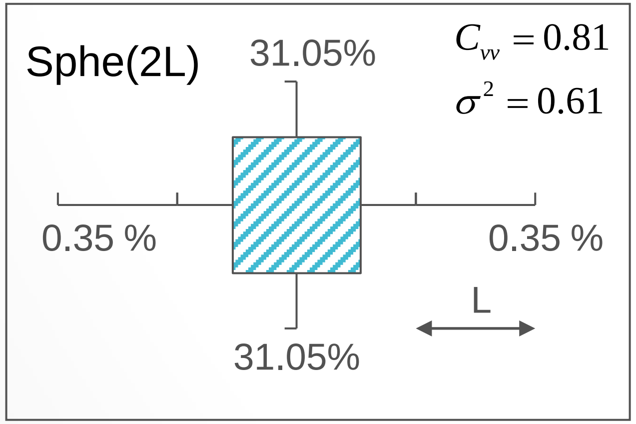


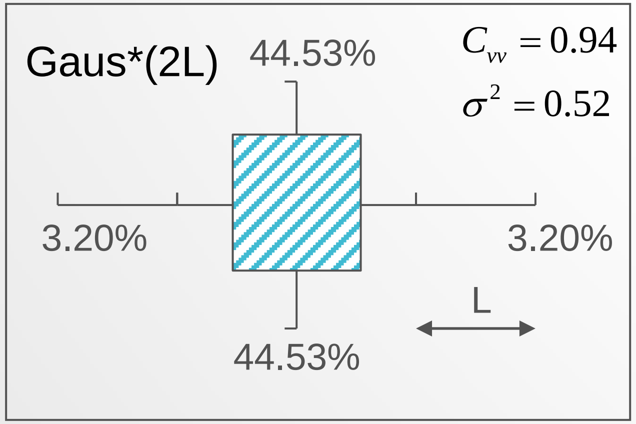


Kriging weights

Simple Kriging of the central **Block**

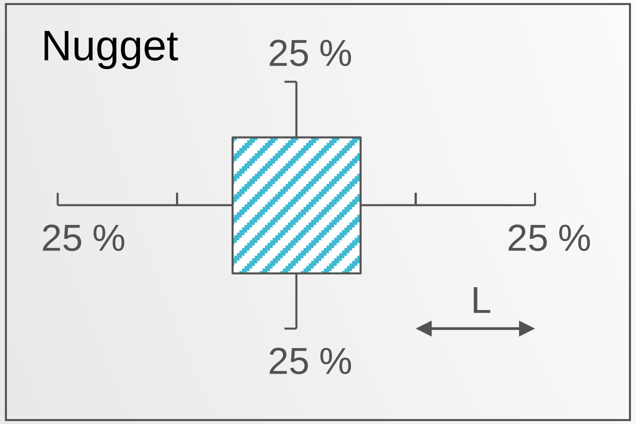






Kriging weights

Ordinary Kriging of the central **Block**



Nugget

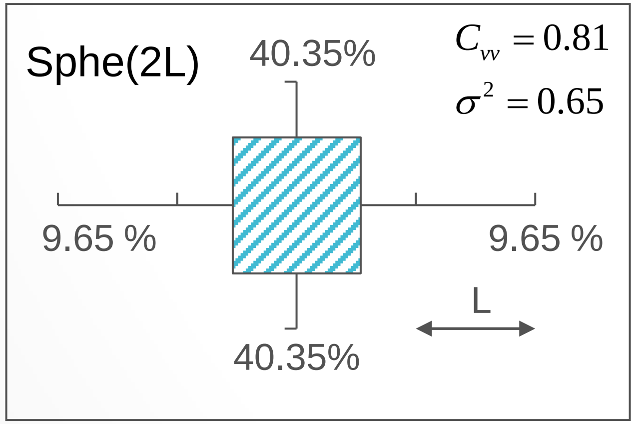
25 %

25 %

25 %

L

25 %



Sphe(2L)

40.35%

$C_{vv} = 0.81$

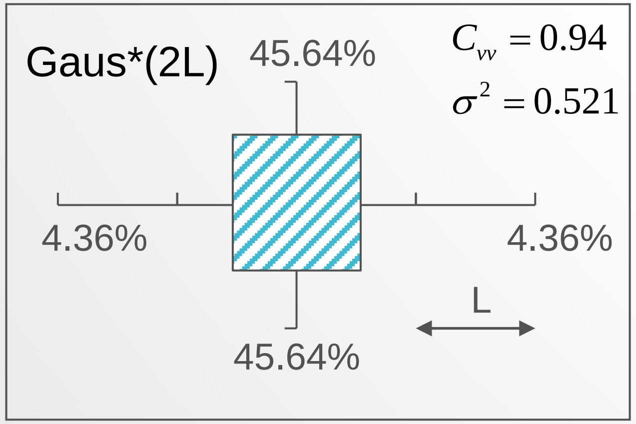
$\sigma^2 = 0.65$

9.65 %

9.65 %

L

40.35%



Gaus*(2L)

45.64%

$C_{vv} = 0.94$

$\sigma^2 = 0.521$

4.36%

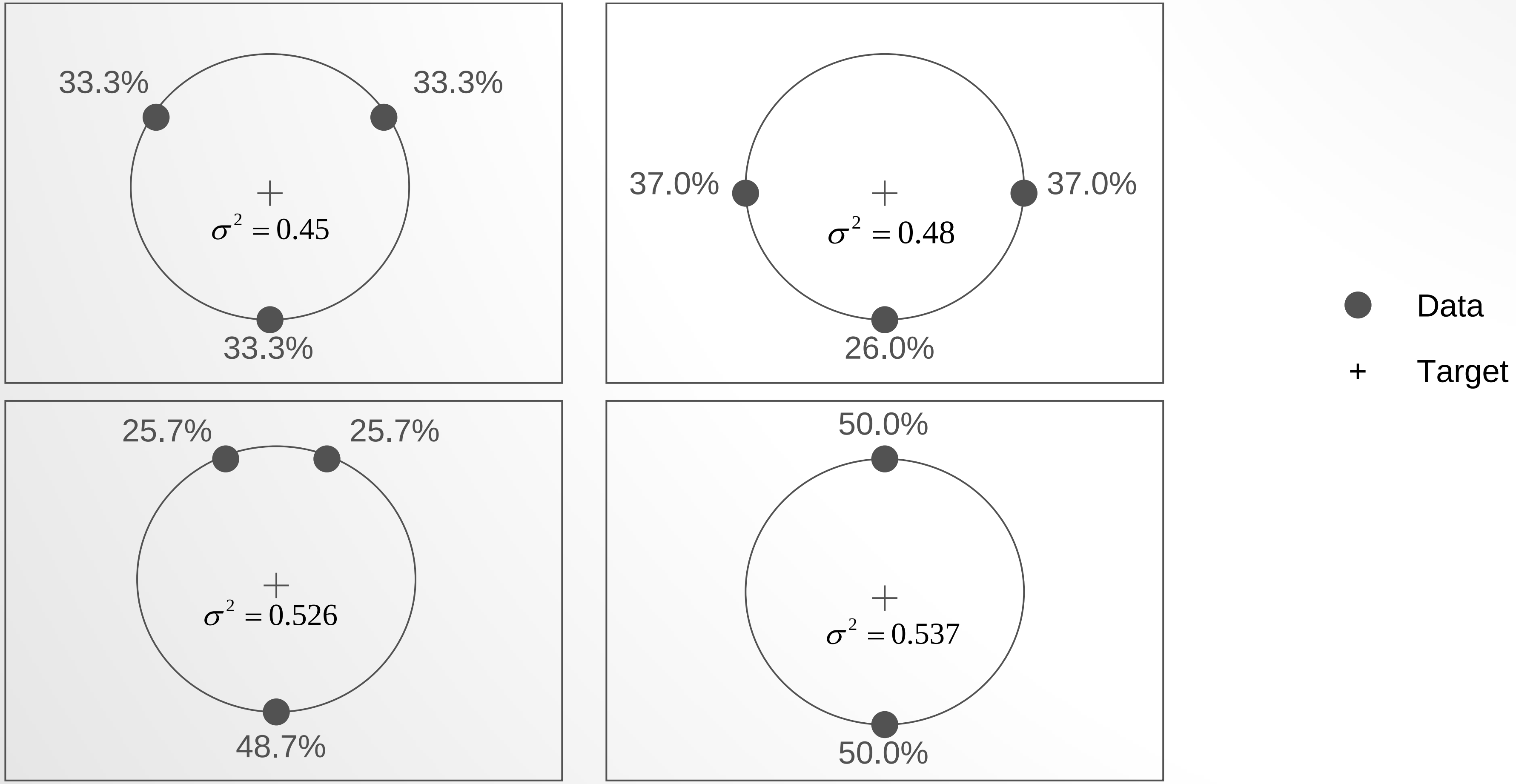
4.36%

L

45.64%

Kriging weights: Declustering

Ordinary Kriging – Model with large range



Kriging Properties

- The Kriging system is **regular** if:
    - the model is authorized
    - there is no redundant information (ex. duplicate points)
    - the solution is unique
- Kriging is an **exact interpolator**: at a sample, the punctual estimate is equal to data value and the estimation variance is zero
- Data values are not used for the determination of the Kriging weights, nor for the variance of estimation error
- Kriging weights are not modified when modifying the sill of the Model
- Variance of estimation error is proportional to the sill of the Model

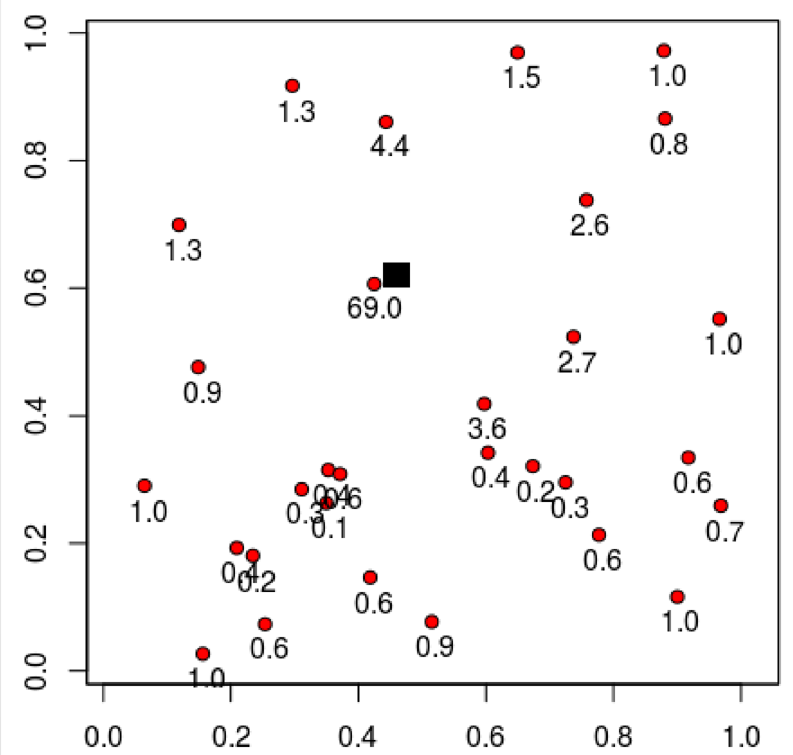
Neighborhood

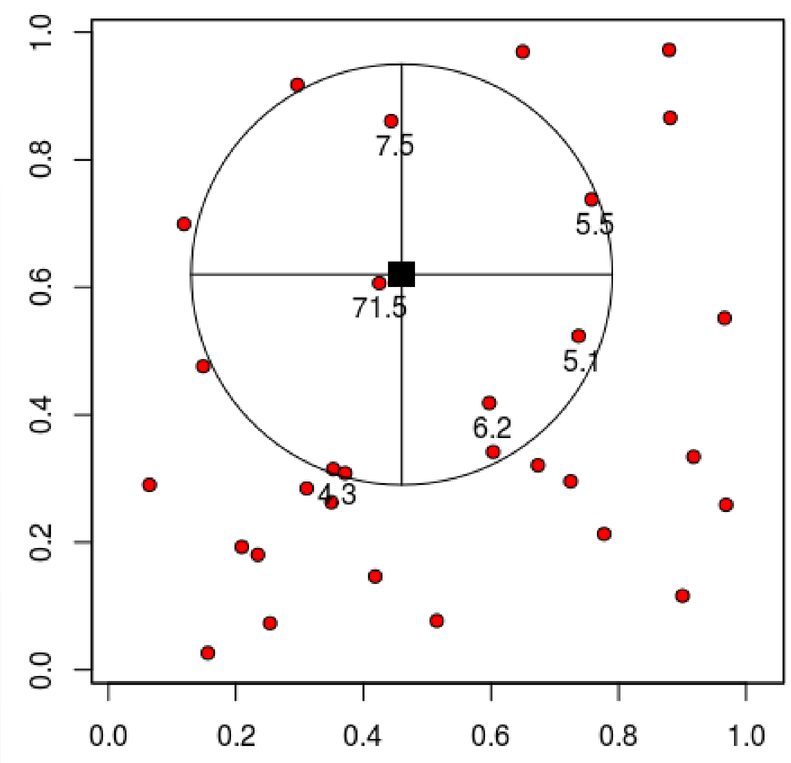
The neighborhood defines the subset of samples used in kriging of a target site:

- **unique**: all samples
- **moving**: only closest samples to the site (used if data set too large ~400)

Display of kriging weights for target site (square)

Unique



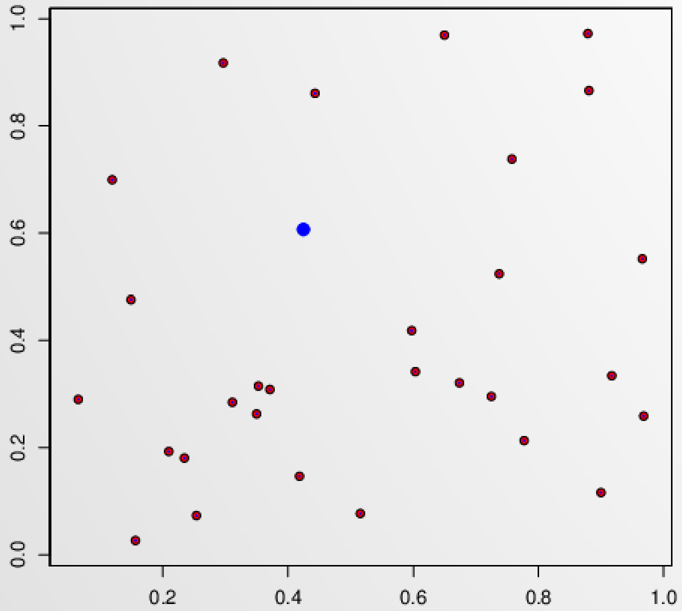


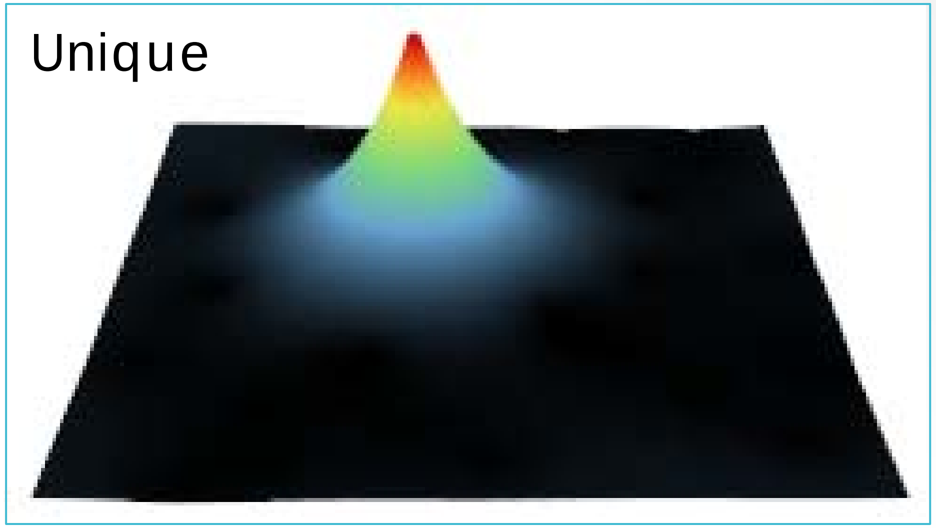
Moving
Radius=0.33
2 pts / quadrant

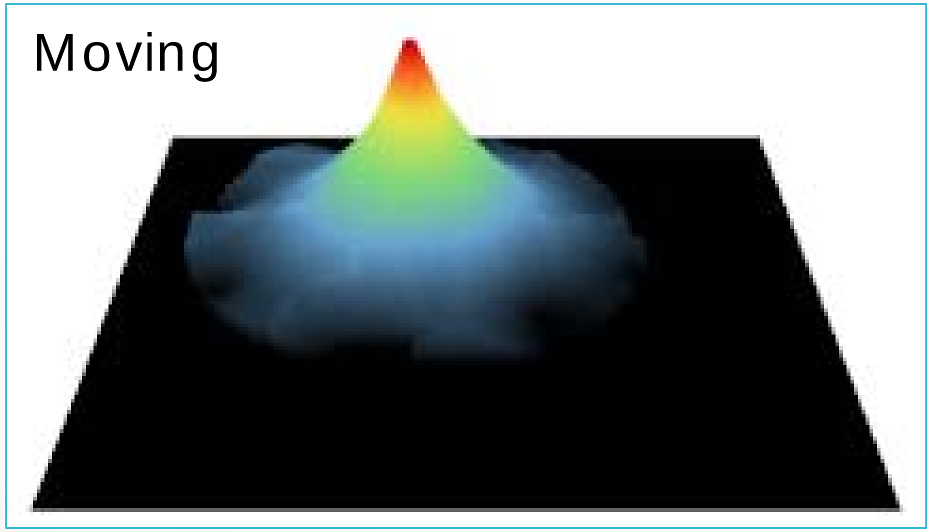
Neighborhood

- Kriging weight relative to **blue** sample when estimating each grid node

- Moving neighborhood:
  radius = 0.33
  maximum= 2 pts per quadrant







Back to Jupyter Notebook!

- We focus on the cell from **Interpolation** section:
  - Kriging Temperature for Year 2008-T2 at 25m Depth
    - Display Estimation Results
    - Display Associated Standard Deviation
    - Change the Neighborhood parameters

Cross-validation

For each datum:

- Suppress the data Z_{i_0}
- Perform Kriging at that data site: $Z_{i_0}^* = \sum_{i \neq i_0} \lambda_i Z_i$

Evaluate:

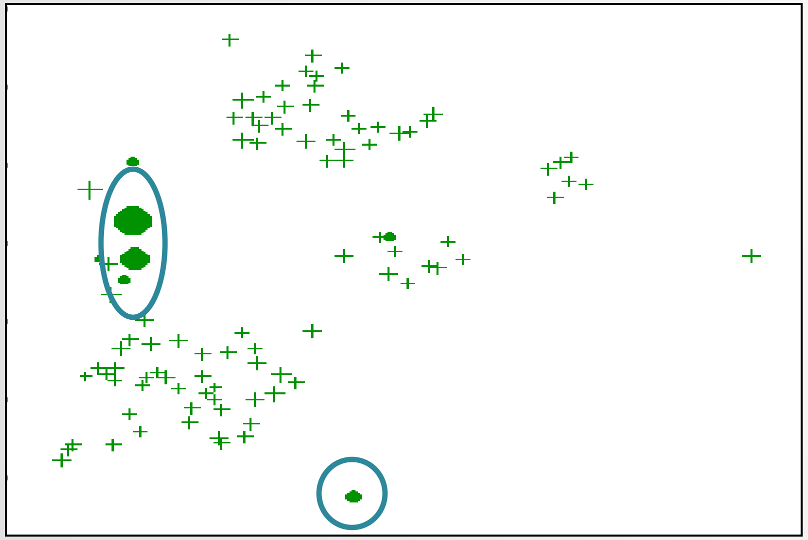
- The error between data value and its estimation $Z_{i_0}^* - Z_{i_0}$
- Normalized by the st. deviation of estimation error $\dfrac{Z_{i_0}^* - Z_{i_0}}{\sqrt{Var(\varepsilon_{i_0})}}$

Calculate statistics (mean and variance) on these errors

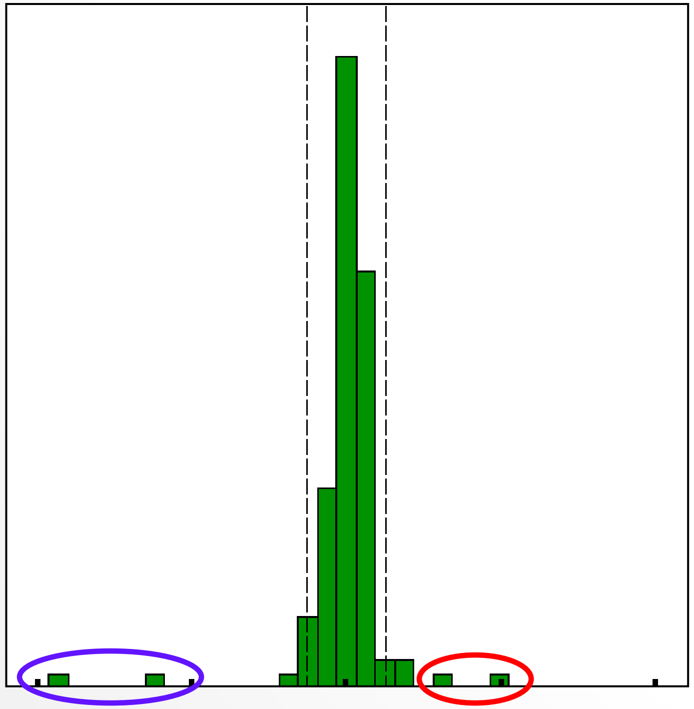
	Mean	Variance
Error	0	$<<D^2$
Normalized error	0	≈ 1

Cross-validation

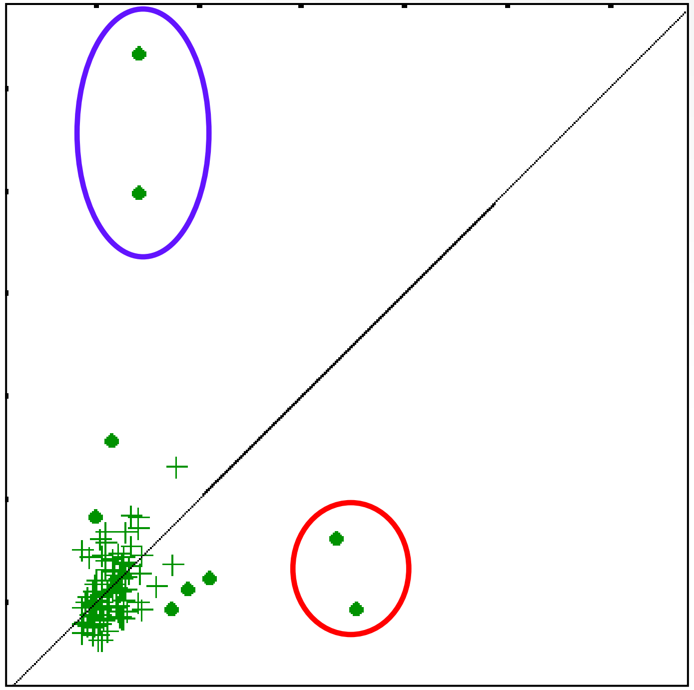
Base Map of normalized errors



Histogram of normalized errors



Correlation Z | Z*



◯ Outliers $=$ ◯ $\dfrac{Z^* - Z}{\sqrt{Var(\varepsilon)}} < -2.5$ $+$ ◯ $\dfrac{Z^* - Z}{\sqrt{Var(\varepsilon)}} > 2.5$

Back to Jupyter Notebook!

- We focus on the cell from **Cross-Validation** section:
  - Kriging Temperature for Year 2008-T2 at 25m Depth
    - Display Cross-Validation Results

End of presentation

Thank you for attention!