UNIVERSITETET I OSLO

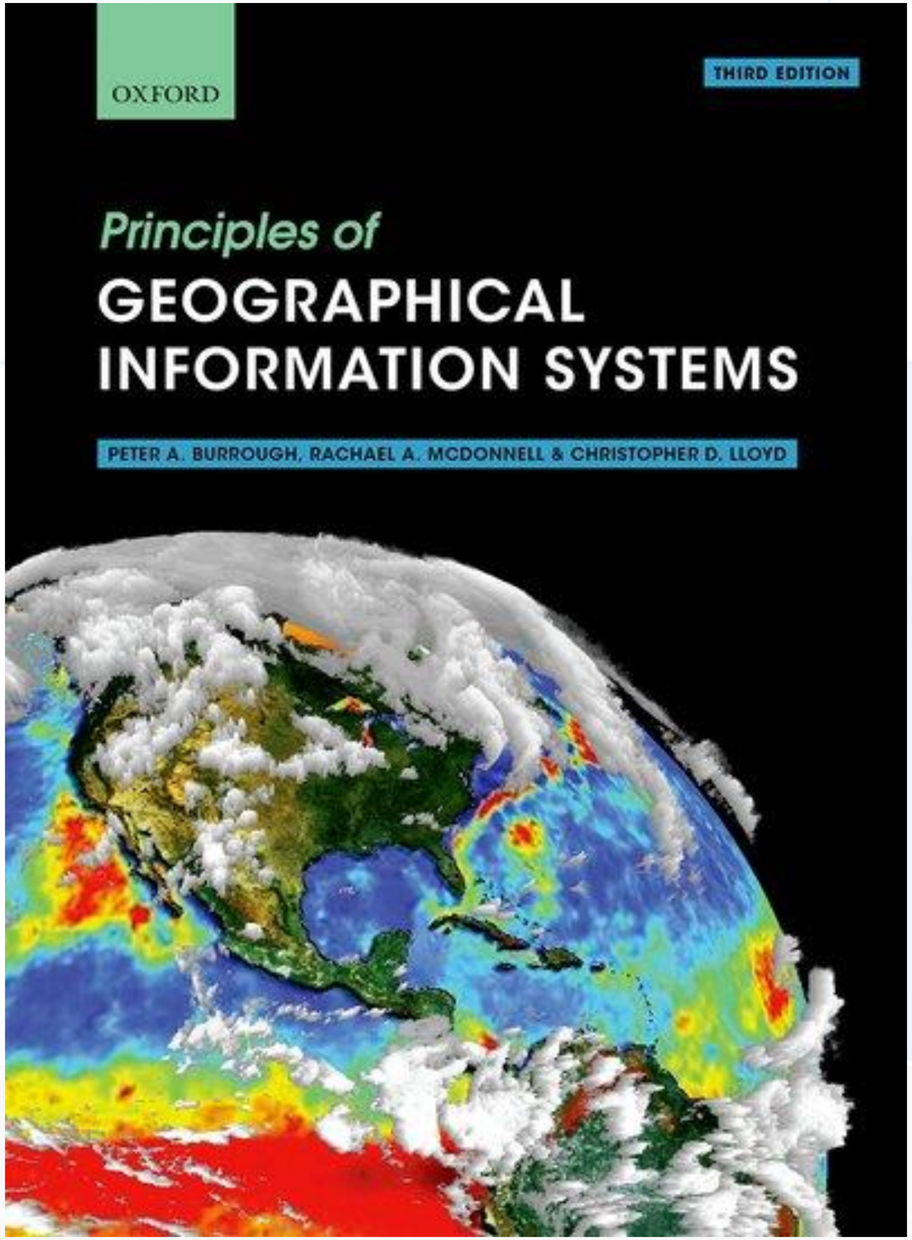
GEO3460 – Geografiske informasjonssystemer (GIS) og geografisk datainnsamling – vår 2025

GEO3460 - Geographical Information Systems (GIS) and Geographical Data Acquisition - spring 2025

Geostatistics – kriging

Luc Girod (luc.girod@geo.uio.no)

Based on the lectures by Dr. Livia Piermattei



- Reference text book:

"*Geographical information systems*"

  - Chapter 9

Learning Objectives

1

What is Geostatistics

- Regionalized variable theory

2

Kriging

- Spatial variation
- Semivariogram
- Basic statistic concept

3

Kriging in ArcGIS

Today's topics

Learning Objectives

1

What is Geostatistics

- Regionalized variable theory

2

Kriging

- Spatial variation
- Semivariogram
- Basic statistic concept

3

- Kriging in ArcGIS
- Summary

Today's topics

Interpolation algorithms

Interpolation methods allow you to construct models of reality (of the phenomenon you are interested in)

A big part of building a **good model** is

your understanding of the phenomenon
how the sample data was obtained
what it represents
what you expect the model to provide

Interpolation algorithms

- Trend surfaces
- Nearest neighbour (Thiessen)
- Inverse distance interpolation / weighted moving average (IDW)
- Spline/local polynomials
- Kriging (analyses of spatial variation)

Interpolation algorithms

- Kriging (analyses of spatial variation) are **geostatistics** interpolating methods to predict values at unmeasured points across the domain.

what was the major disadvantage of the IDW method?

We discussed this earlier this week

Interpolation algorithms

$$\hat{z}(\mathbf{x}) = \frac{\sum_i^n w_i z_i}{\sum_i^n w_i} \qquad \text{where} \qquad w_i = |\mathbf{x} - \mathbf{x}_i|^{-\beta}$$

where:
- $\beta \geq 0$
- β is the inverse distance power (or power coefficient)
- $|x-x_i|$ is the euclidean distance
- n is the number of surrounding points to be included

How interpolate

- **Deterministic methods:**
  - Deterministic models use a mathematical function to predict unknown values and result in hard classification of the value of features.
  - Example: splines, IDW, Natural Neighbour, Trend
- **Statistical (geostatistical) methods:**
  - Statistical techniques are based on statistical models that include autocorrelation: the statistical relationship among the measured points.
  - Observations have a dependence in space (spatial variability)

    → *Regionalized variable theory*
  - Provide some measure of the accuracy of the predictions.
  - Example: kriging

Geostatistics

- **Geostatistics**
  - It is a class of statistics used to analyze and predict the values associated with spatial or spatiotemporal phenomena.
    - It incorporates the spatial (and in some cases temporal) coordinates of the data within the analyses.
  - Geostatistical tools and methods provide interpolated values and measures of uncertainty for those values
  - The measurement of uncertainty provides information on the possible values (outcomes) for each location rather than just one interpolated value

Geostatistics

- **Geostatistics: assumptions**
  - Geostatistical methods were developed for interpreting **data** that **varies continuously** over a predefined, fixed spatial region.
  - The study of Geostatistics assumes that at least some of the spatial variation observed for natural phenomena can be modeled by **random processes** with spatial autocorrelation.
  - It is based on the **theory of regionalized variables**, variable distributed in space (or time).

Geostatistics: Spatial variability

- **Spatial autocorrelation:**
  - presence of systematic spatial variation in a variable
  - positive spatial autocorrelation is the tendency for areas or sites that are close together to have similar values.

> *"Everything in the universe is related to everything else, but closer things are more related."* – Tobler's First Law of Geography

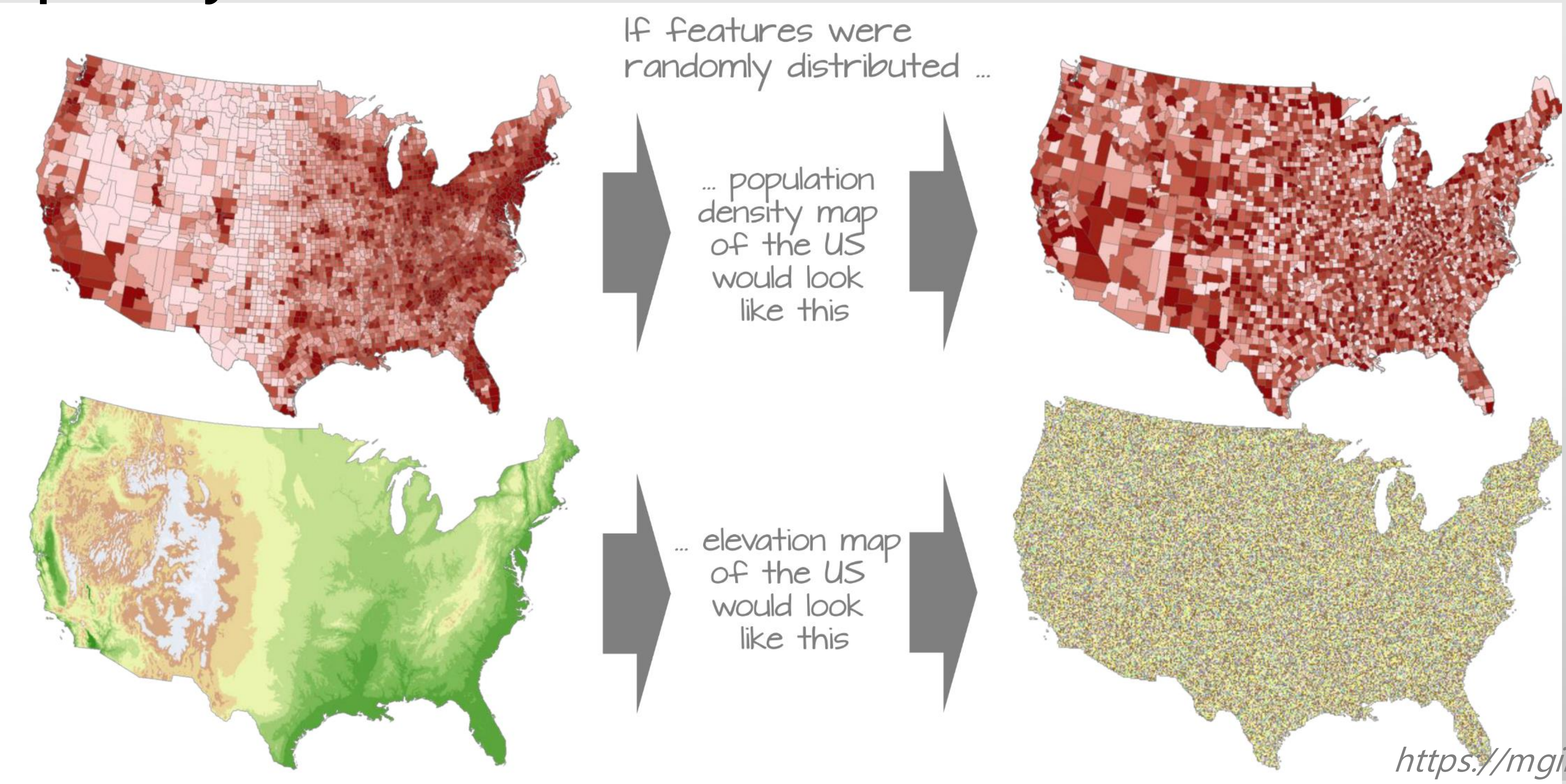
Geostatistics: Spatial variability

- **Examples of spatially dependent variables (regionalized variables)**
  - Rainfall
  - Soil’s hydraulic conductivity
  - Chemical concentration
  - Plant properties
  - Slope
  - Temperature
  - Population characteristics
  - …

Is elevation a spatially autocorrelated variable?

Geostatistics: Spatial variability

- Spatially distributed vs Random distribution



https://mgimond.github.io/

Geostatistics: Spatial variability

spatial autocorrelation
sites which are close together tend to be more similar than those which are further apart

Task: quantitative and objective approach to **quantifying the degree of clustering of similar features and where such clustering occurs**

Geostatistics: Regionalized variable theory

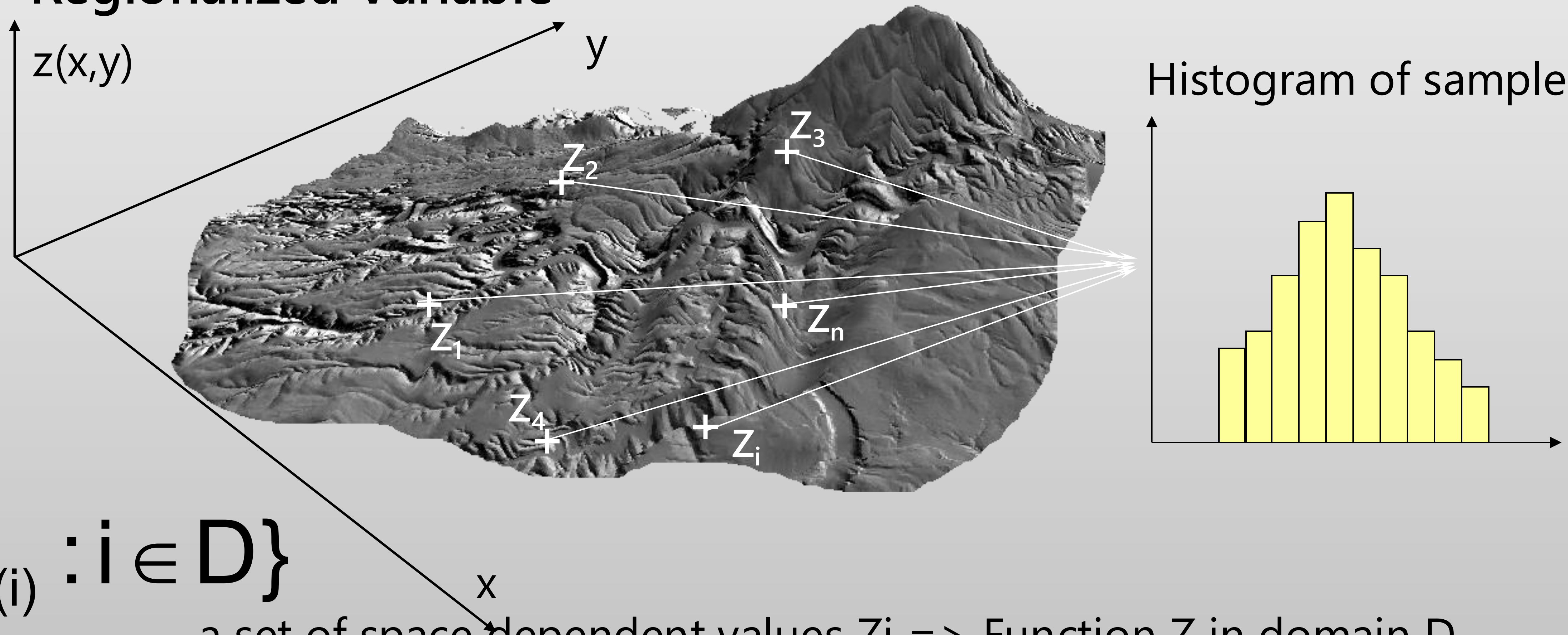
- **Why use regional variables theory**
  - General analysis tool for spatially varying/dependent data
  - A general tool for spatial interpolation
  - A tool for regionalization studies
  - A basis for developing <u>spatial models that consider regional differences</u>

Geostatistics: Regionalized variable theory

- The concept of the theory is that interpolation from points in space should not be based on a smooth continuous object.

Geostatistics: Regionalized variable theory

- Regionalized Variable



$$\{z_{(i)} : i \in D\}$$

a set of space dependent values Zi => Function Z in domain D

Geostatistics: Regionalized variable theory

- **Regionalized Variable**
  - Given a variable Z, measured at a location i, the variability in Z can be broken down into **three** components:

$$z_{(i)} = f_{(i)} + s_{(i)} + \varepsilon$$

Where:

Usually removed by detrending

$f_{(i)} =$ A "structural" coarse scale forcing (e.g. difference in mean levels) or **trend**

What we are interested in

$s_{(i)} =$ Correlated variation, **random local spatial dependency**

$\varepsilon =$ **error variance** (normally distributed) (uncorrelated variation, noise, measurement error)

Geostatistics: Regionalized variable theory

- **Regionalized Variable**
  - Given a variable Z, measured
    be broken down into three c

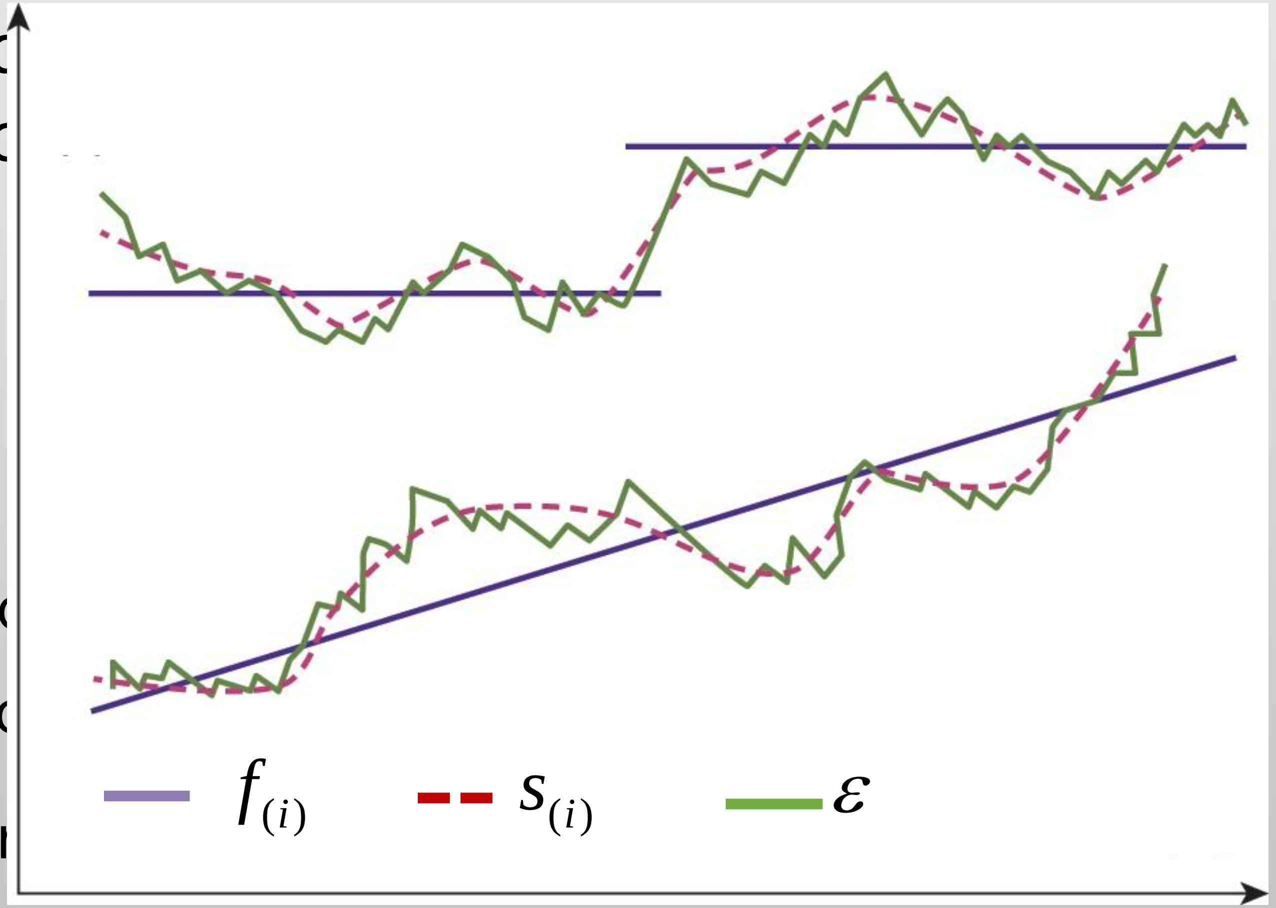
$$z_{(i)} = f_{(i)} + s_{(i)} + \varepsilon$$

Where:

$f_{(i)} =$ A "structural" coarse scale
difference in mean levels) c

$s_{(i)} =$ Correlated variation (rando

$\varepsilon =$ error variance (normally distr



Geost

- Regi
  - Gi be

$Z_{($

Whe

$f_{(i)}$

$s_{(i}$

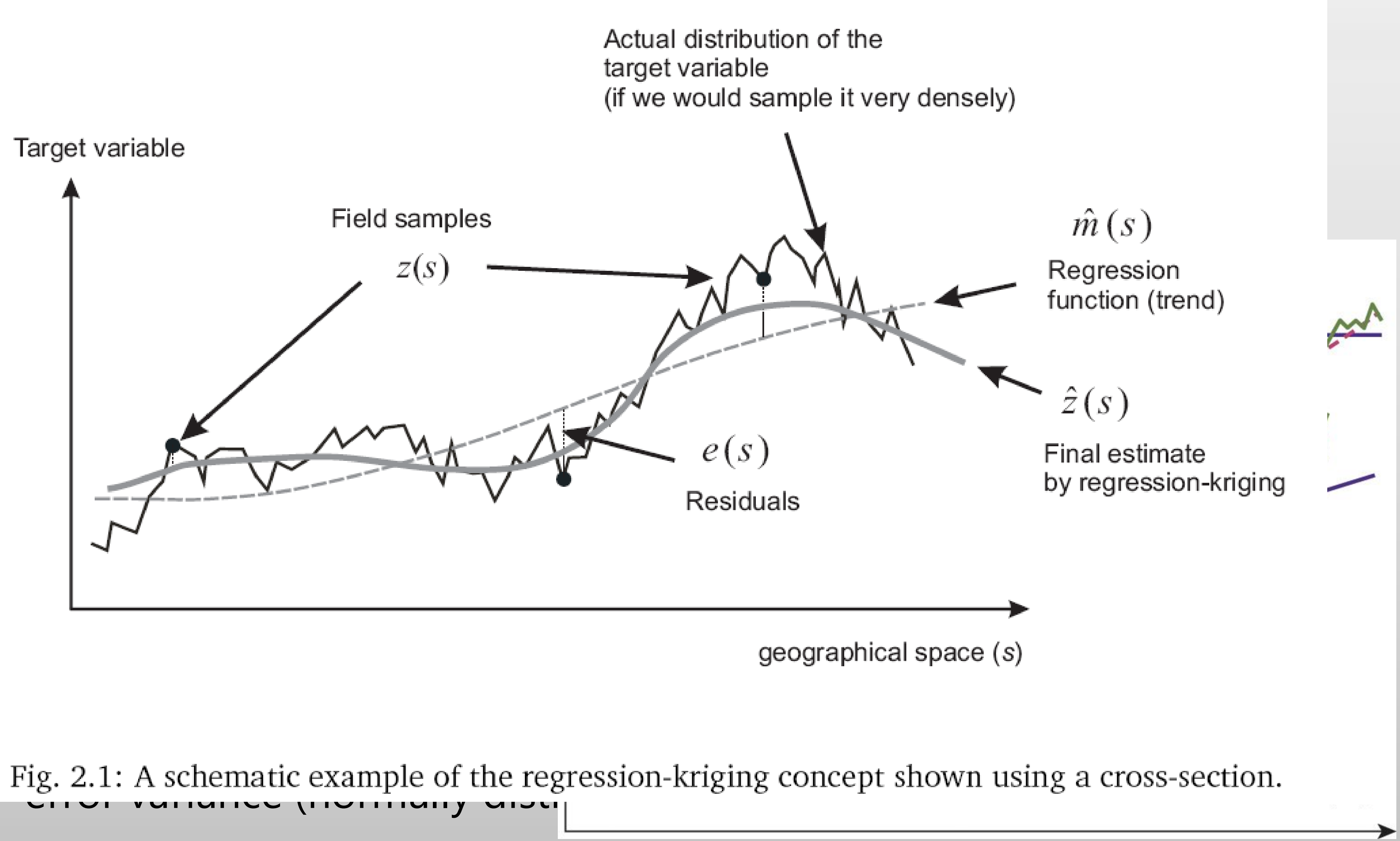


Fig. 2.1: A schematic example of the regression-kriging concept shown using a cross-section.

ε = error variance (normally distr

Geostatistics: Regionalized variable theory

- **Geostatistical methods: divide the spatial variation of a variable into three components**
  - a deterministic model m(x,y)
  - a regionalised statistical (spatially correlated) variation from m(x,y)
    - It is defined by analysing the semivariance, γ
    - **Semivariance** is a measure of the spatial dependence between two observations as a function of the distance between them
  - a random noise (normal error) component

Learning Objectives

1

What is Geostatistics

- Regionalized variable theory

2

Kriging

- Spatial variation
- Semivariogram
- Basic statistic concept

3

- Kriging in ArcGIS
- Summary

Today's topics

Geostatistics: measures of variability

- **Let's review some (not necessarily spatial) statistics**

□ **Arithmetic Mean**

$$\bar{x} = \frac{x_1 + x_2 + x_3 + \cdots + x_n}{n}$$

Example: mean of (7,1,5,2,8) = 7 + 1 + 5 + 2 + 8 = 23/5 = 4.6

□ **Median** (the middle value in a group of numbers)

$$median\ number\ in\ a\ list = \frac{(n+1)}{2}$$

Example: median of (1, 2, 5, 7, 8) = (5+1)/2 = 5 = 3rd Number. *if there are an even number of numbers, take the mean of the middle two

□ **Min/Max** are just what they sound like, the highest and lowest value in a set

□ **Standard Deviation** (a measure of how spread out/scattered the data are from the mean)

$$\sigma = \sqrt{\frac{\sum(x - \bar{x})^2}{N}}$$

Example: standard deviation of (7,1,5,2,8); we know the mean is 4.6

$(7 - 4.6)^2 + (1 - 4.6)^2 + (5 - 4.6)^2 + (2 - 4.6)^2 + (8 - 4.6)^2 = 37.2$

Square root (37.2 /5) = 2.73

Geostatistics: measures of variability

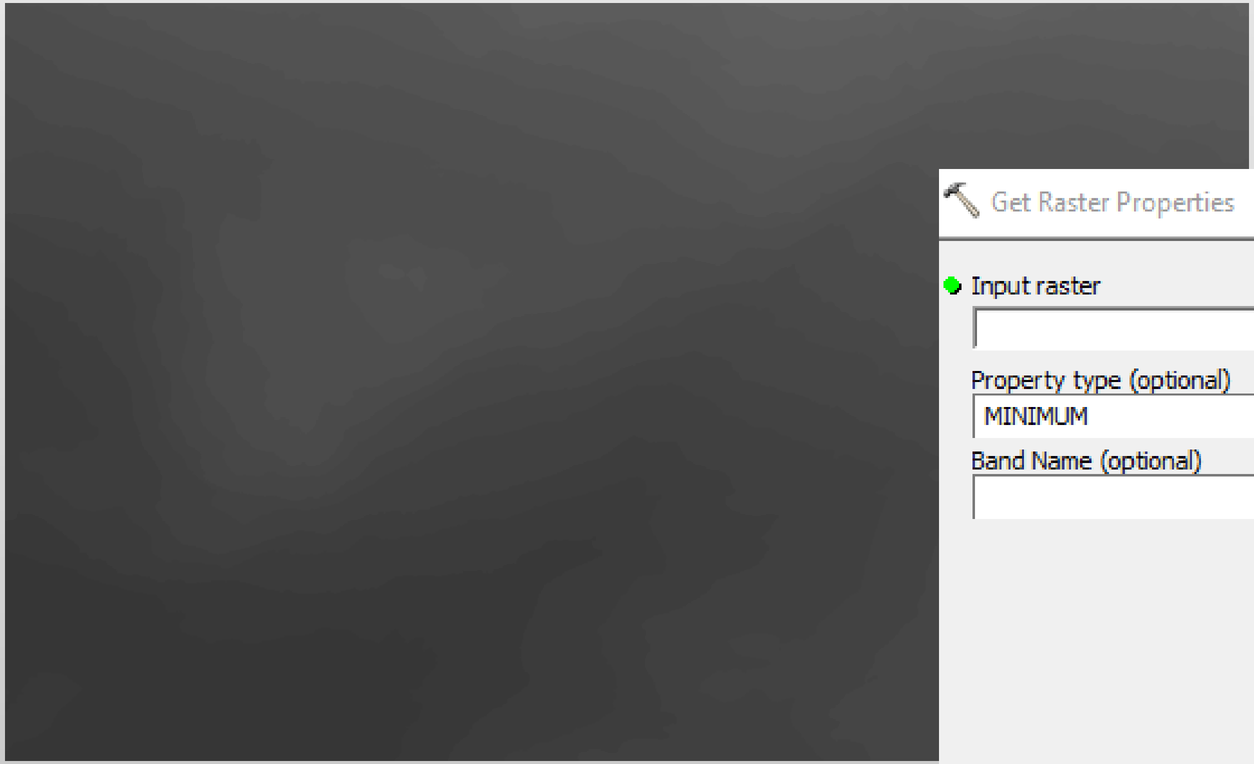
- Let’s review some (not necessarily spatial) statistics

Many of these statistics are readily accessible in ArcGIS

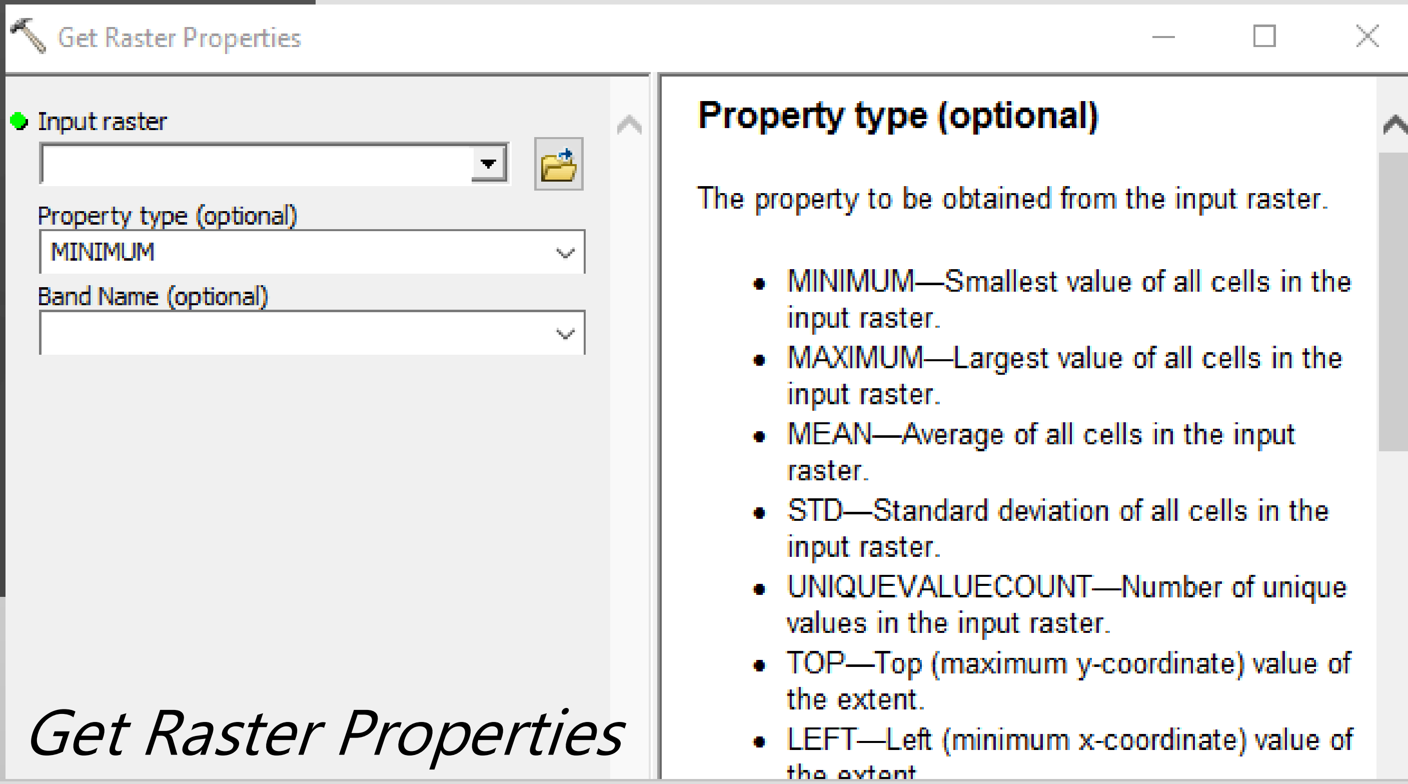
Statistics	
Band_1	
Build Parameters	skipped columns: 1, rows: 1, ignored value(s):
Min	1983,9912109375
Max	2988,2561035156
Mean	2458,4515065964
Std dev.	253,68485774034
Classes	0

Geostatistics: measures of variability

- Let's review some (not necessarily spatial) statistics



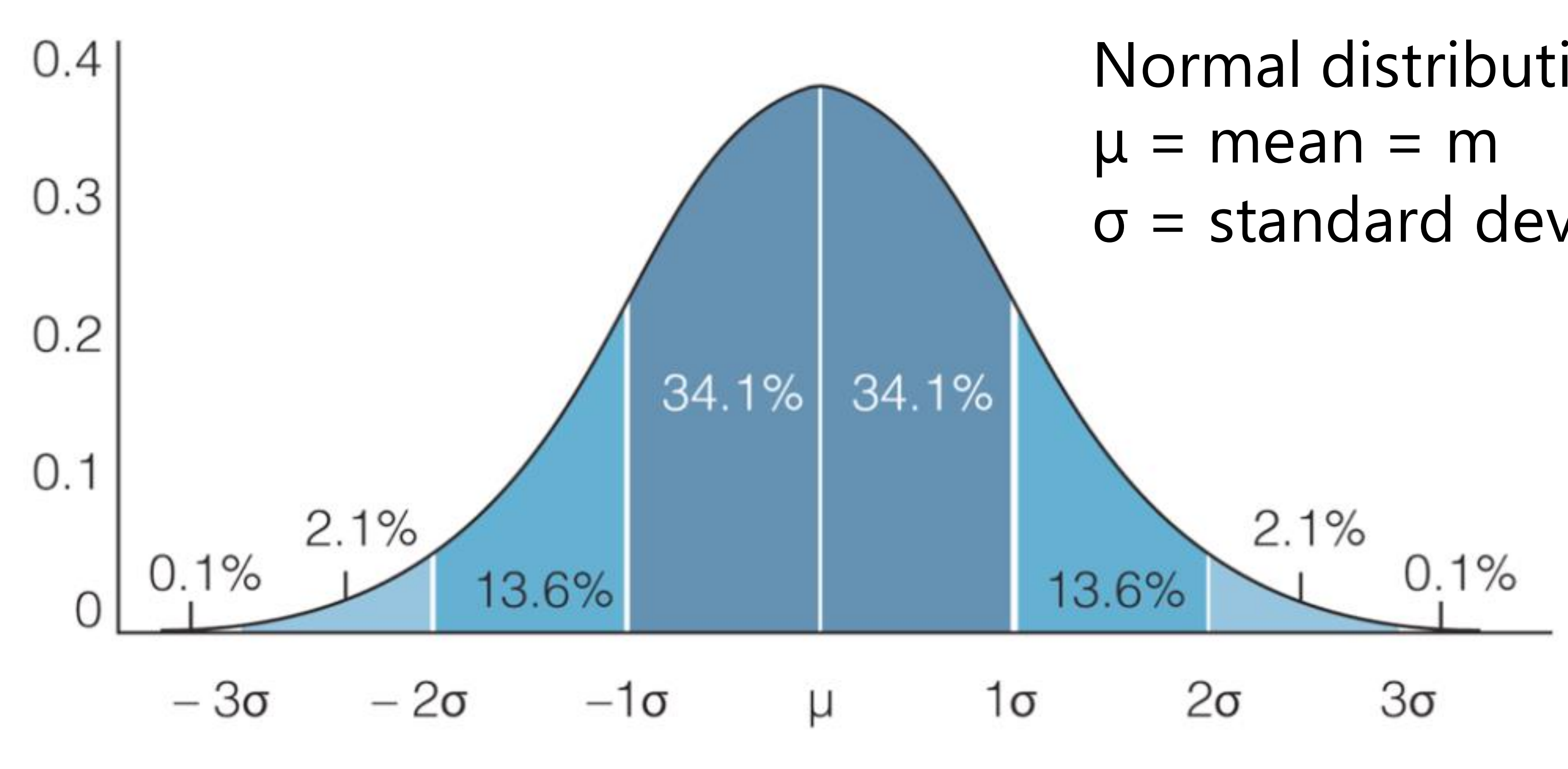
Many of these statistics are readily accessible in ArcGIS



…and if they are not, derive them Tool: *Get Raster Properties*

Geostatistics: measures of variability

- Let’s review some (not necessarily spatial) statistics

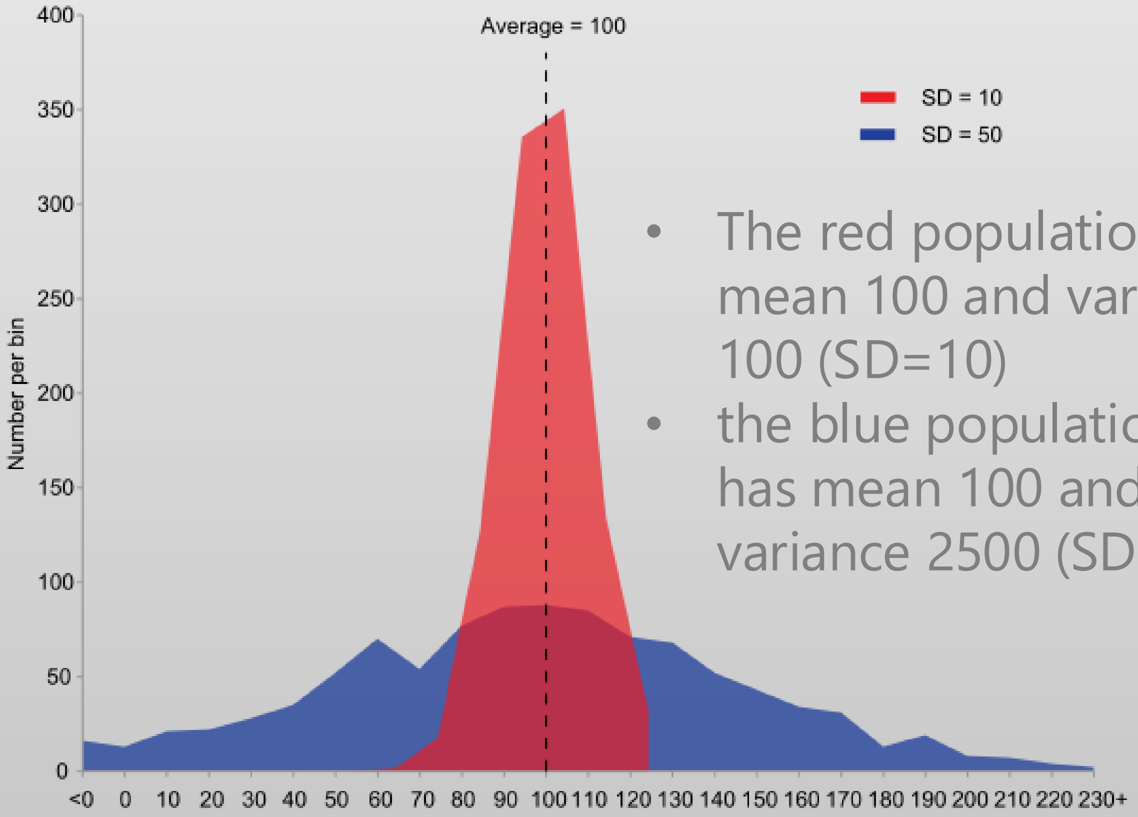


Normal distribution
μ = mean = m
σ = standard deviation = SD

Geostatistics

- Definition of variance
  - <u>Variance</u> is the expectation E[x] of the squared deviation of a random variable x from its mean (m)
  - The variance is the square of the standard deviation

$$Var(x) = \sigma^2 = E[(x - m)^2]$$



- The red population has mean 100 and variance 100 (SD=10)
- the blue population has mean 100 and variance 2500 (SD=50)

Histogram of two populations with the same mean but different variances.

Geostatistics: kriging interpolation

- **Ordinary kriging interpolation (4 steps)**

1. Removing any spatial trend in the data (if present)

2. Computing the experimental variogram, γ, which is a measure of spatial autocorrelation.

3. Defining an experimental variogram model that best characterizes the spatial autocorrelation in the data.
   - Interpolating the surface using the experimental variogram.

4. Adding the kriged interpolated surface to the trend interpolated surface to produce the final output.

Geostatistics: kriging interpolation

- 1. De-trend
  - 1st assumption:
    - Mean and the variation in the entity being studied (all over the domain) is constant across the study area
      → There should be no global trend in the data
  - If the assumption is not met
    - remove the trend from the data before proceeding with the kriging operations

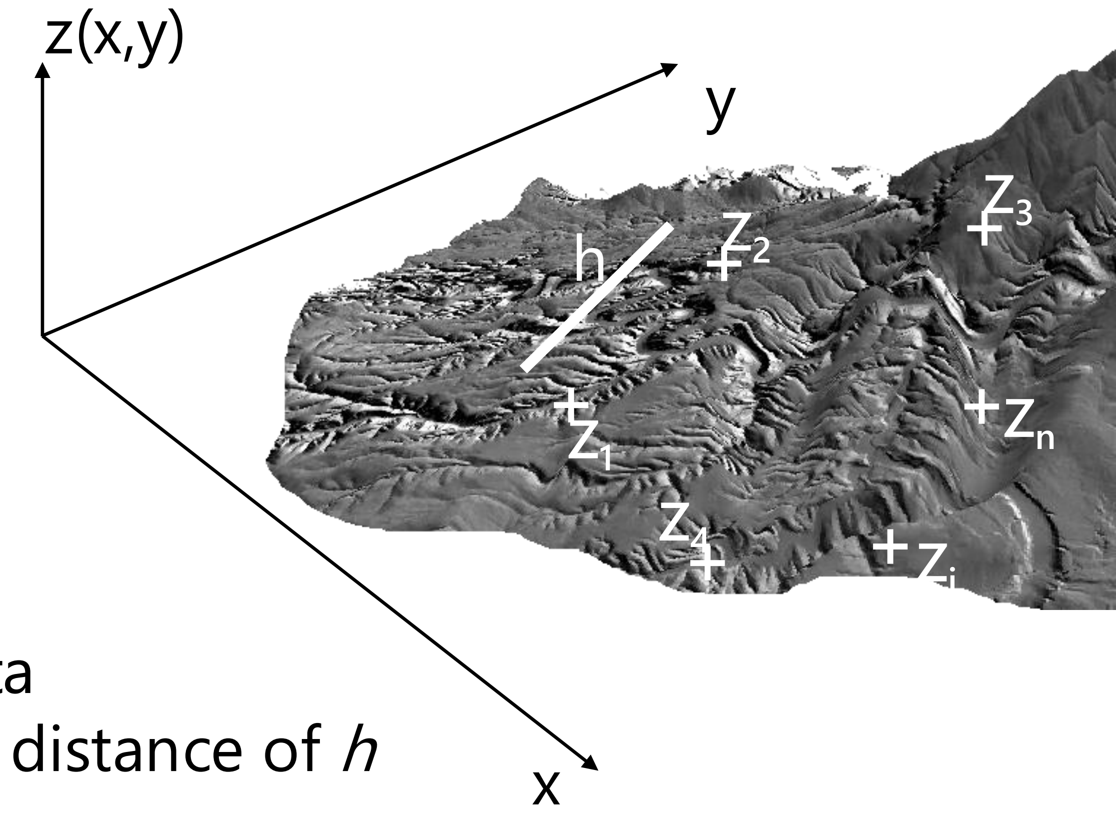
  Note: the modeled trend will be added to the kriged interpolated surface at the end of the workflow.

Geostatistics: kriging interpolation

□ 2. Experimental variogram → semivariance γ

Equation:

$$\gamma(h) = \frac{1}{2n(h)} \sum_{i=1}^{n(h)} [z(x_i + h) - z(x_i)]^2$$



Where

- z = attribute value at a location x,y
- *h* is the distance between ordered data
- *n(h)* is the number of paired data at a distance of *h*

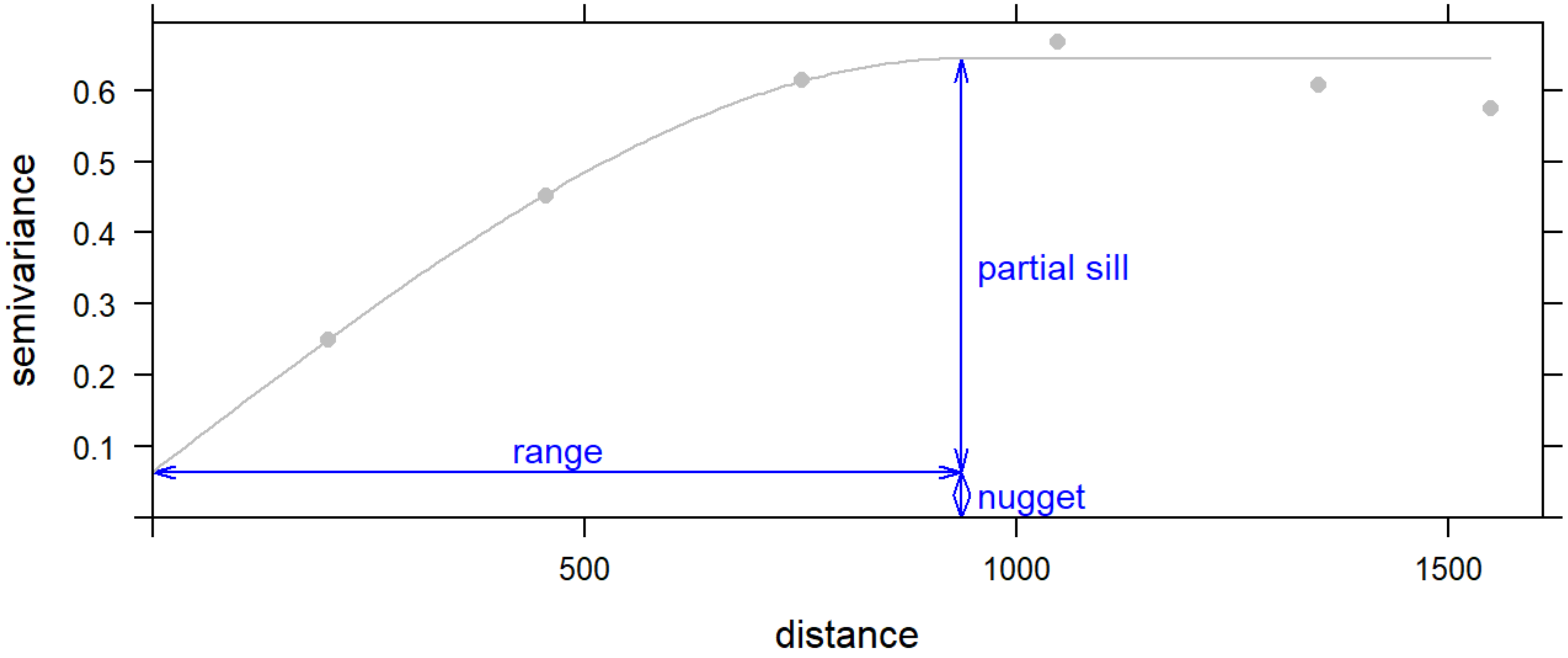
Geostatistics: kriging interpolation

- **2. Experimental variogram → semivariance γ**
  - Semivariance γ is a measure of the spatial dependence between two observations as a function of the distance between them
  - Semivariogram: a plot of semivariances versus distances between ordered data in a graph.
  - 2nd order stationarity:

    The variance of the increment corresponding to two different locations depends only on the vector separating them.

Geostatistics: kriging interpolation

Semivariogram/variogram

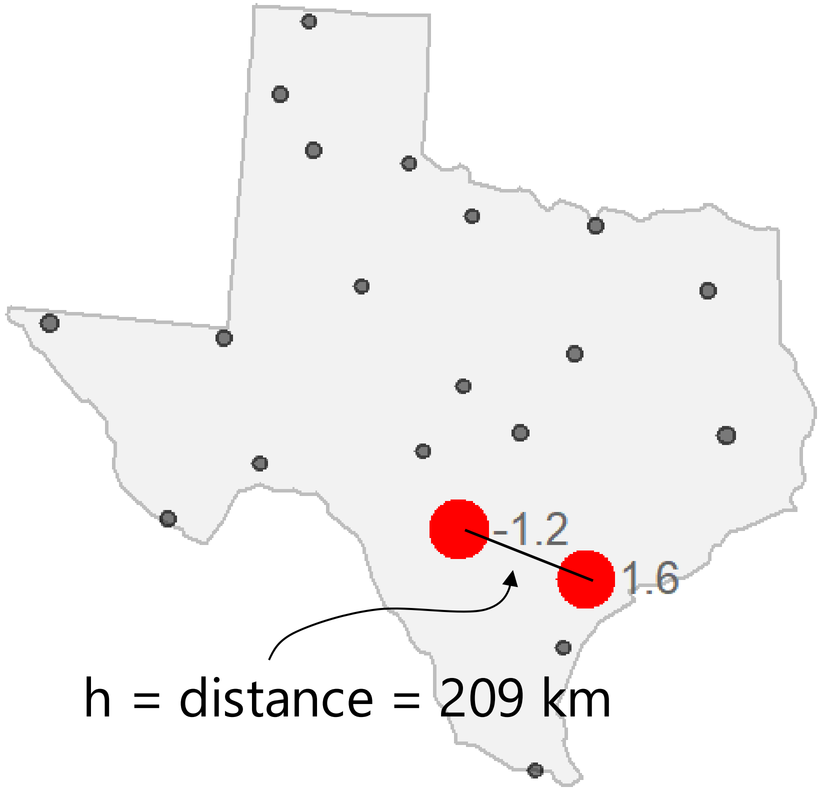


Semivariogram: A plot of semivariances versus distances between ordered data. The variogram is described by **range**, **sill** and **nugget** parameters

- **Partial sill** is the vertical distance between the nugget and the part of the curve that levels off. If the nugget is 0 the partial sill is simply referred to as the sill.
- **Nugget** is the distance between the 0 variance on the y axis. Indicates the random error process
- **Range** is the distance along the x axis where the curve levels off

Geostatistics: kriging interpolation

2. Experimental variogram: example



https://mgimond.github.io/

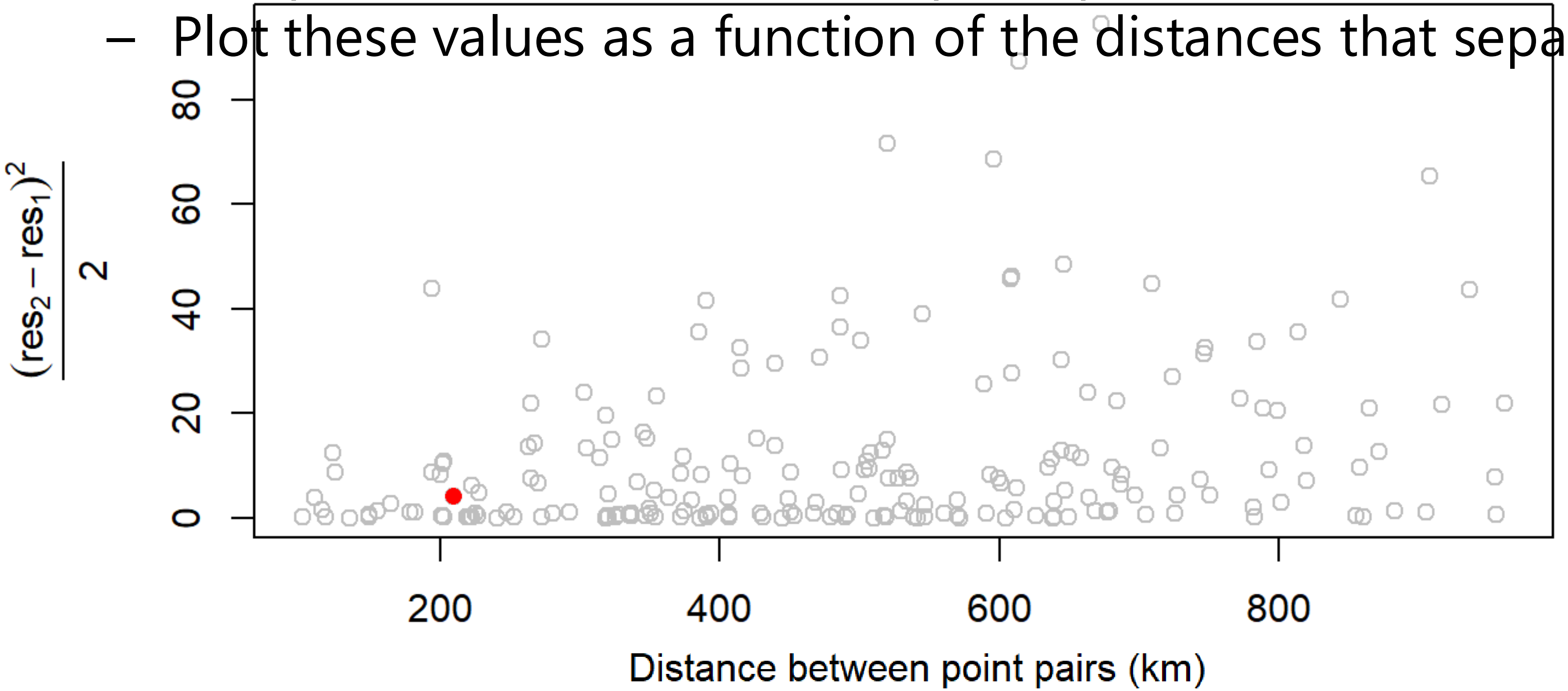
- We are interested in how these attribute values (precipitation residuals) vary as the distance between location point pairs increases
- De-trended the surface → precipitation residuals. De-trended precipitation value is -1.2 and 1.6
- Compute their difference γ

$$\gamma = \frac{(Z_2 - Z_1)^2}{2} = \frac{(-1.2 - (1.6))^2}{2} = 3.92$$

Geostatistics: kriging interpolation

2. Experimental variogram: example

- Compute the difference for all point pairs
- Plot these values as a function of the distances that separate these points



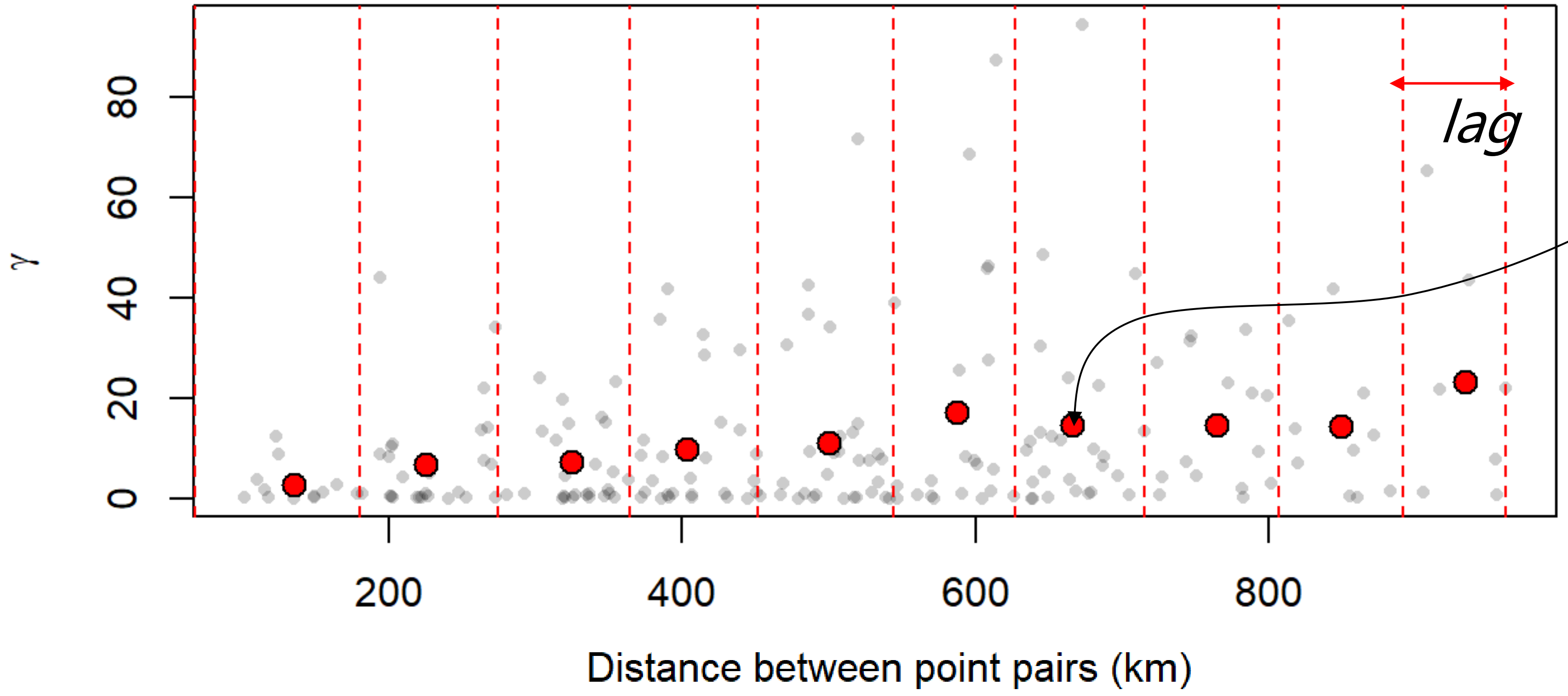
Experimental variogram plot of precipitation residual values

https://mgimond.github.io/

Geostatistics: kriging interpolation

2. Experimental variogram: example

- Sample Experimental Variogram



Average value = sample experimental variogram estimates

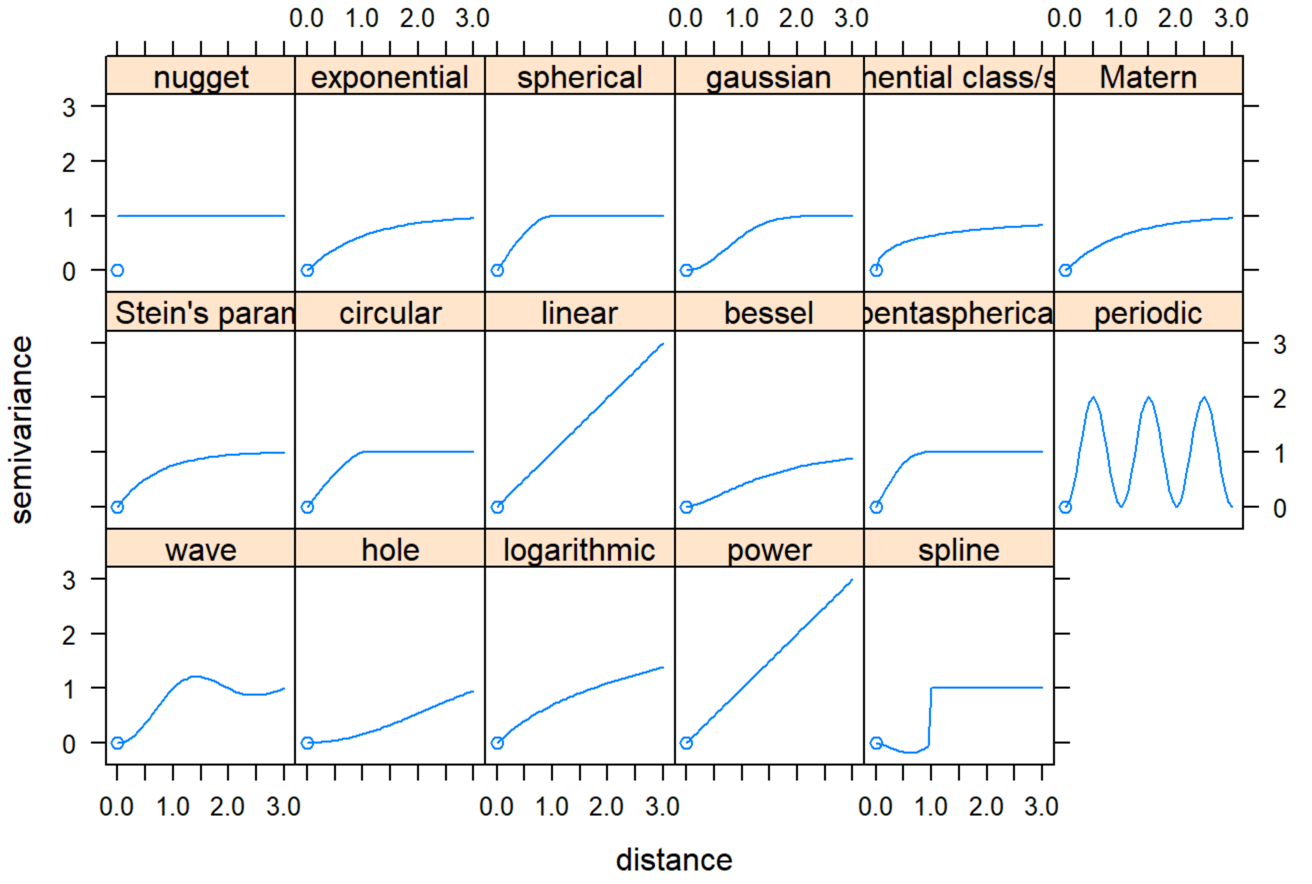
Sample experimental variogram plot of precipitation residual values

https://mgimond.github.io/

Geostatistics: kriging interpolation

3. Experimental variogram model

- Fit a mathematical model to our sample experimental variogram
- Different mathematical models can be used; their availability is software dependent

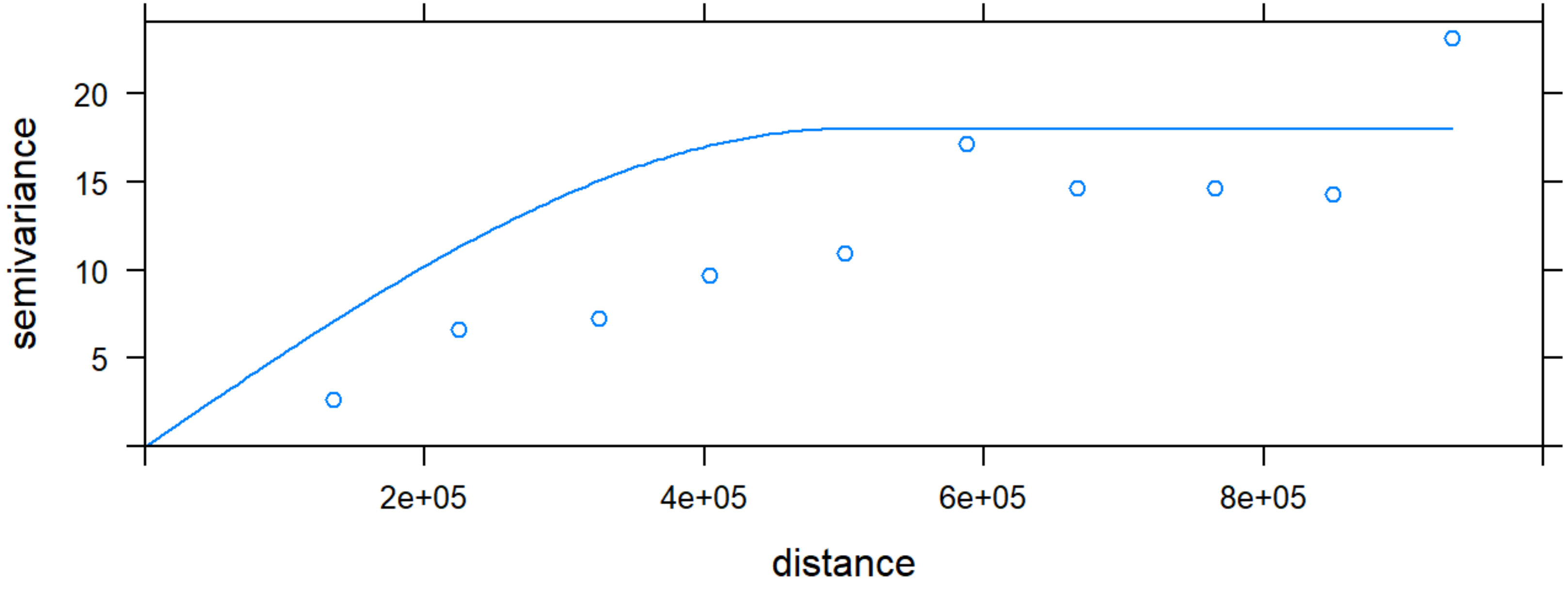


A subset of variogram models available

Geostatistics: kriging interpolation

3. Experimental variogram model: example

– we fit the Spherical function to our sample experimental variogram



A spherical model fit to our residual variogram

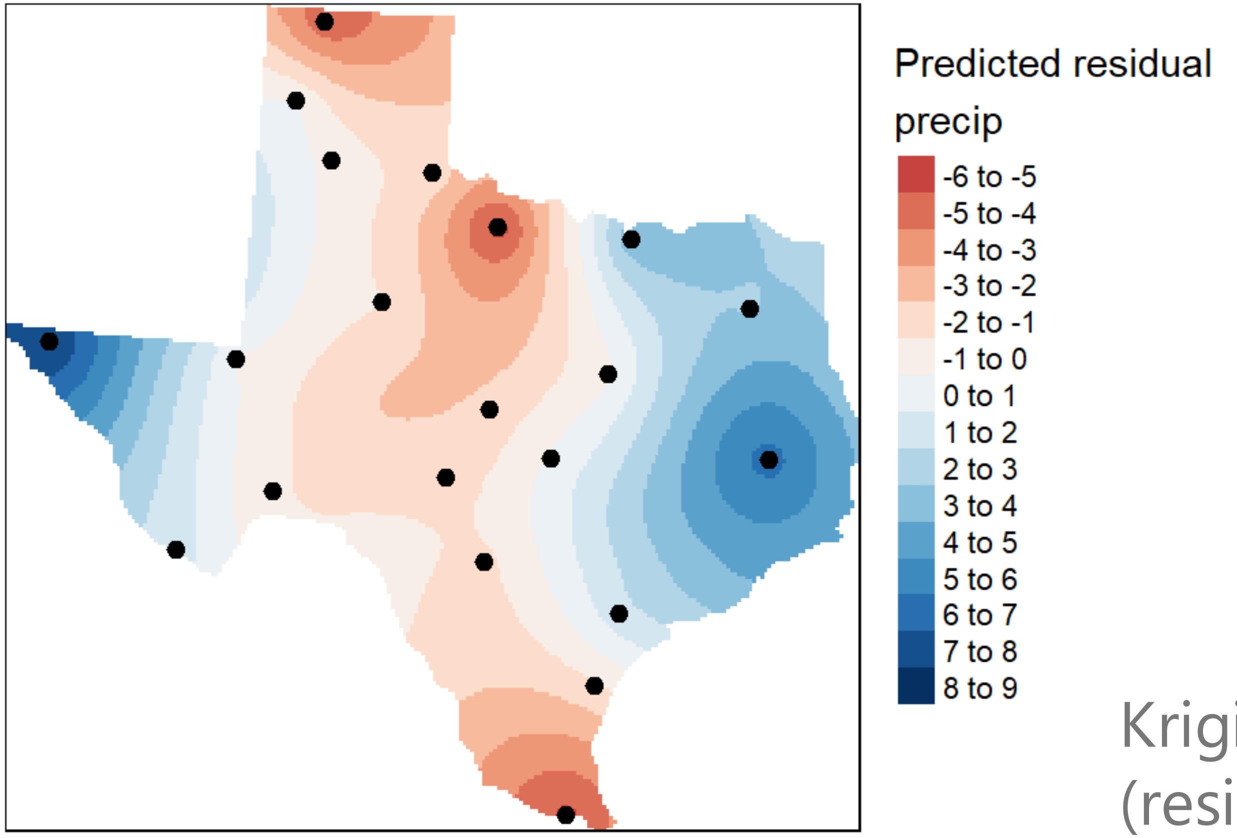
The most popular models are:

- *Spherical*
- *Linear*
- *Gaussian*

https://mgimond.github.io/

Geostatistics: kriging interpolation

4. Kriging Interpolation



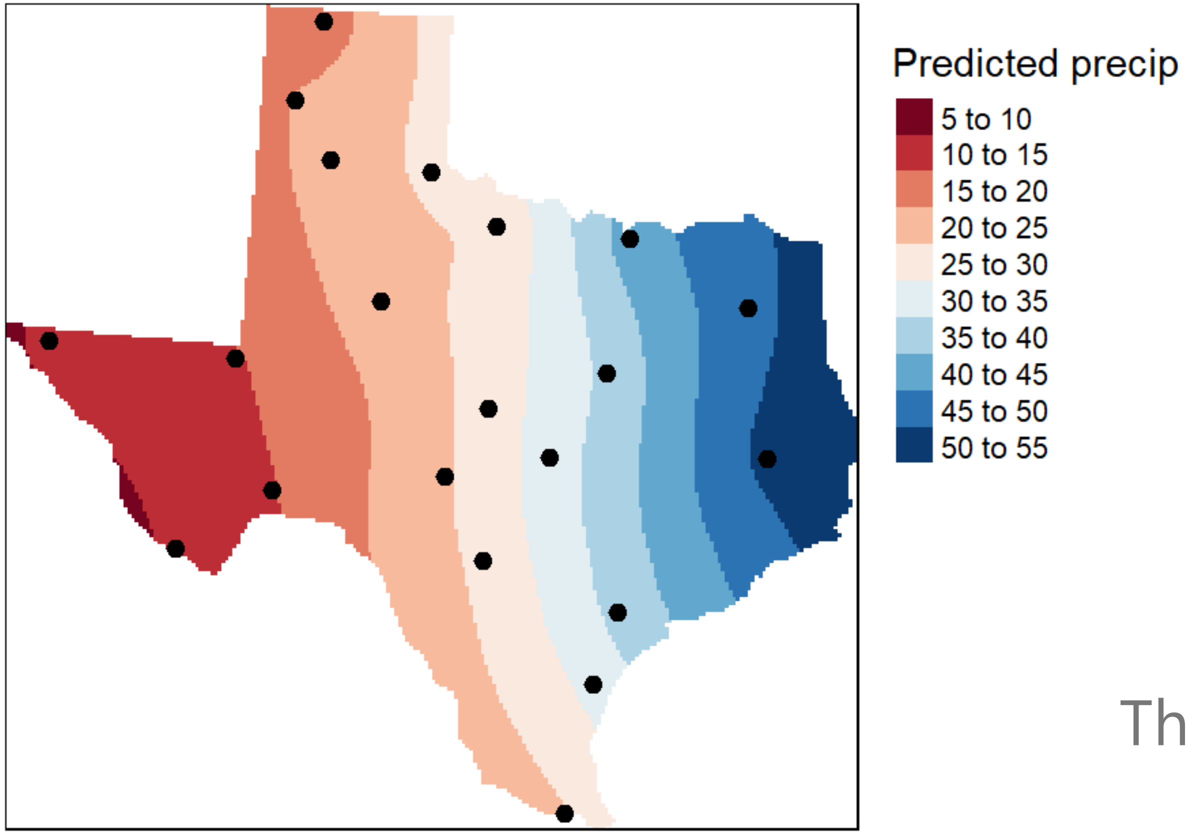
- The variogram model is used by the kriging interpolator to provide localized weighting parameters.

Kriging interpolation of the detrended (residual) precipitation values

https://mgimond.github.io/

Geostatistics: kriging interpolation

4. Kriging Interpolation



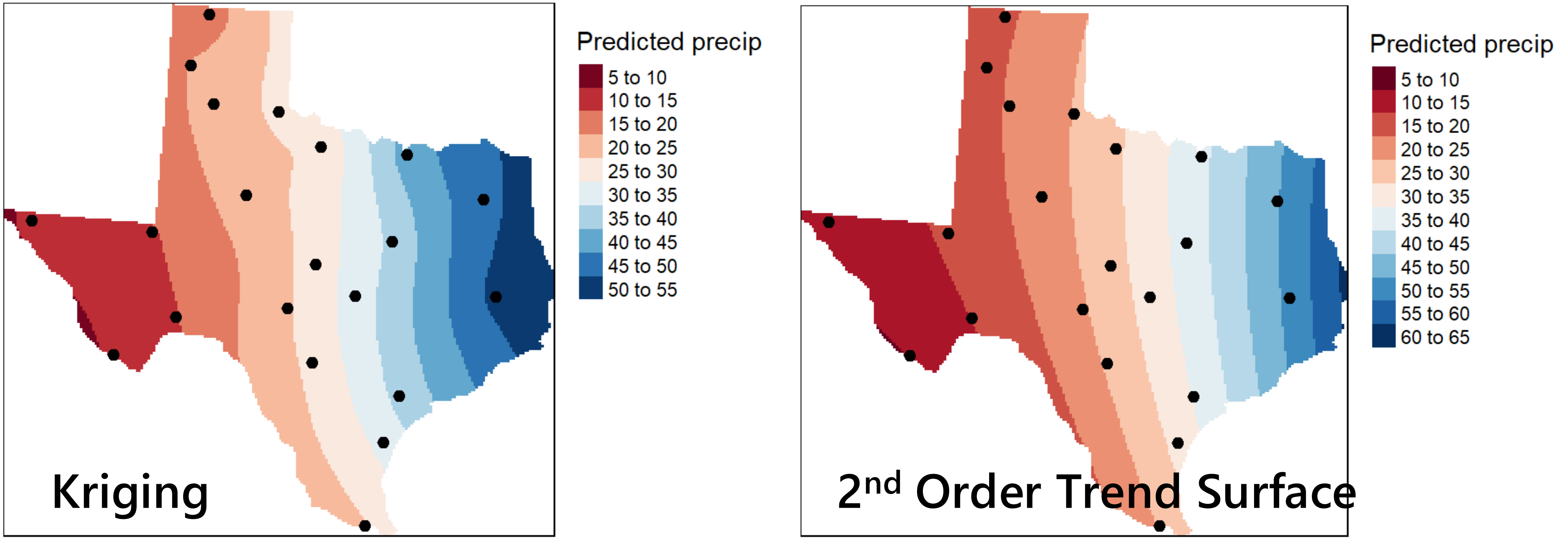
- Combine the interpolated surface (of the residual) with the interpolated surface produced from the trend interpolation

The final kriged surface

https://mgimond.github.io/

Geostatistics: kriging interpolation

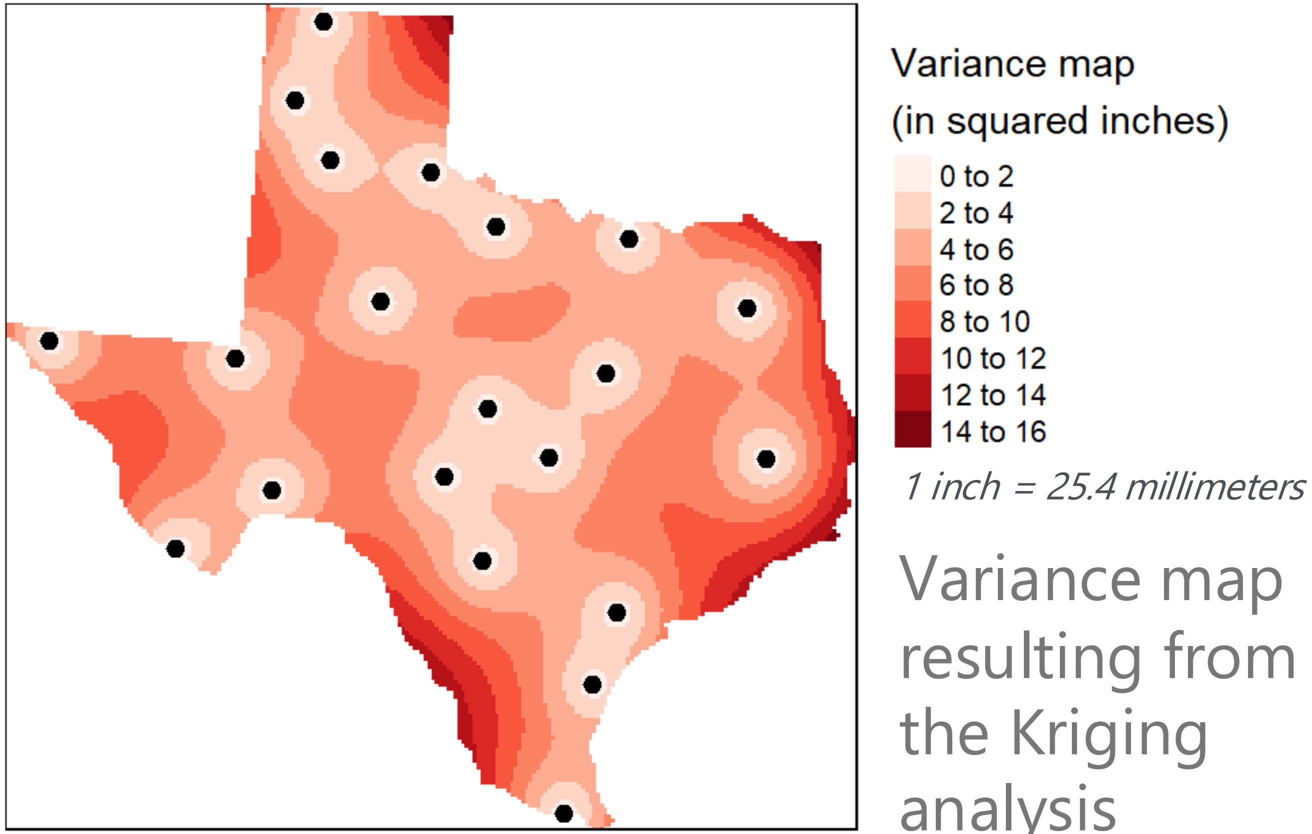
- Comparison Kriging and Trend Surface



https://mgimond.github.io/

Geostatistics: kriging interpolation

Accuracy assessment: variance map



Variance map resulting from the Kriging analysis

- Variance map gives you a measure of uncertainty in the interpolated values.
- The smaller the variance, the better (note that the variance values are in squared units).
- Variance => The average distance of a set of variables from the average value in that set

https://mgimond.github.io/

Learning Objectives

1

What is Geostatistics

- Regionalized variable theory

2

Kriging

- Spatial variation
- Semivariogram
- Basic statistic concept

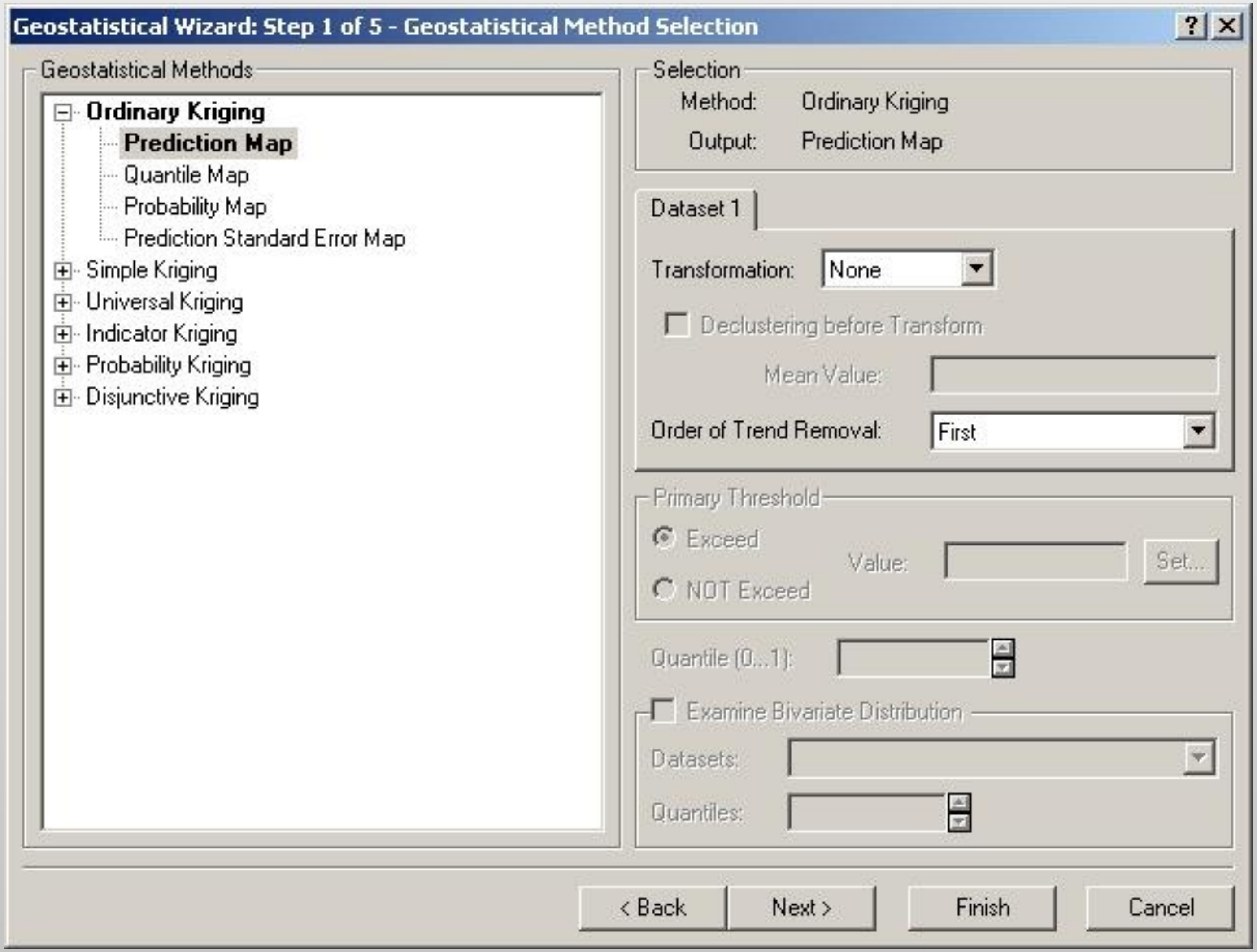
3

- Kriging in ArcGIS
- Summary

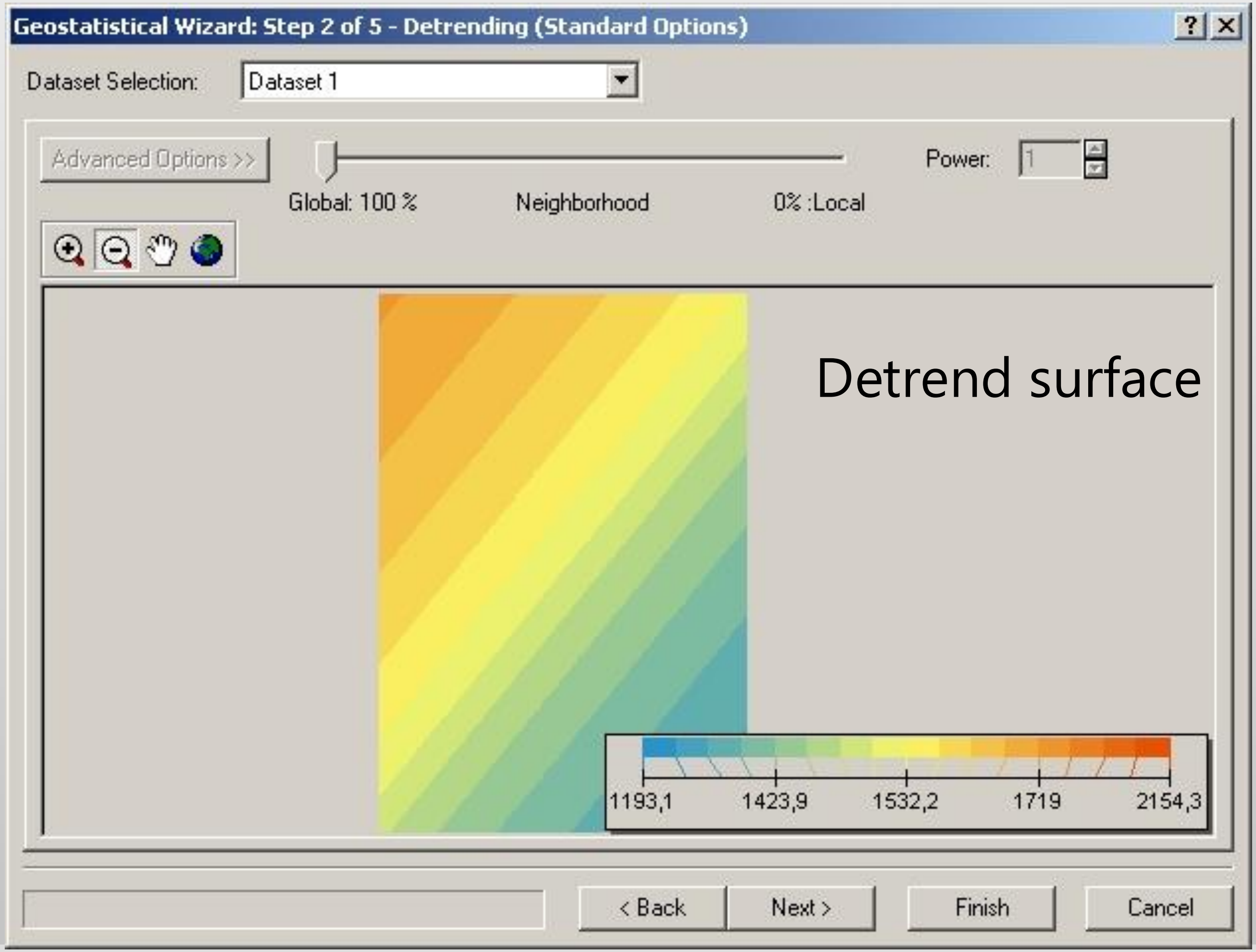
Today's topics

Geostatistics: kriging interpolation in ArcGIS

1. Removing any spatial trend in the data

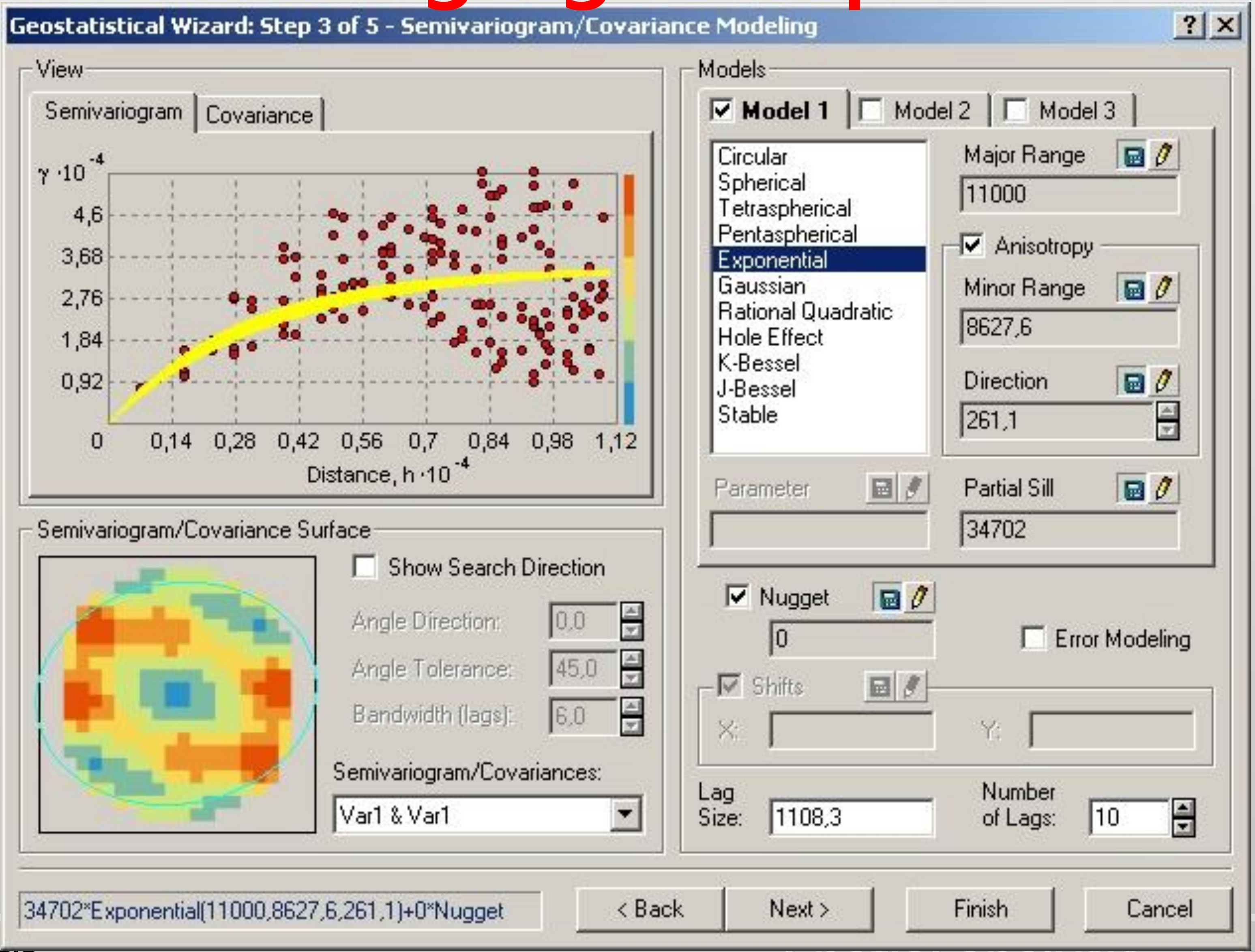


Geostatistics: kriging interpolation in ArcGIS



Geostatistics: kriging interpolation in ArcGIS

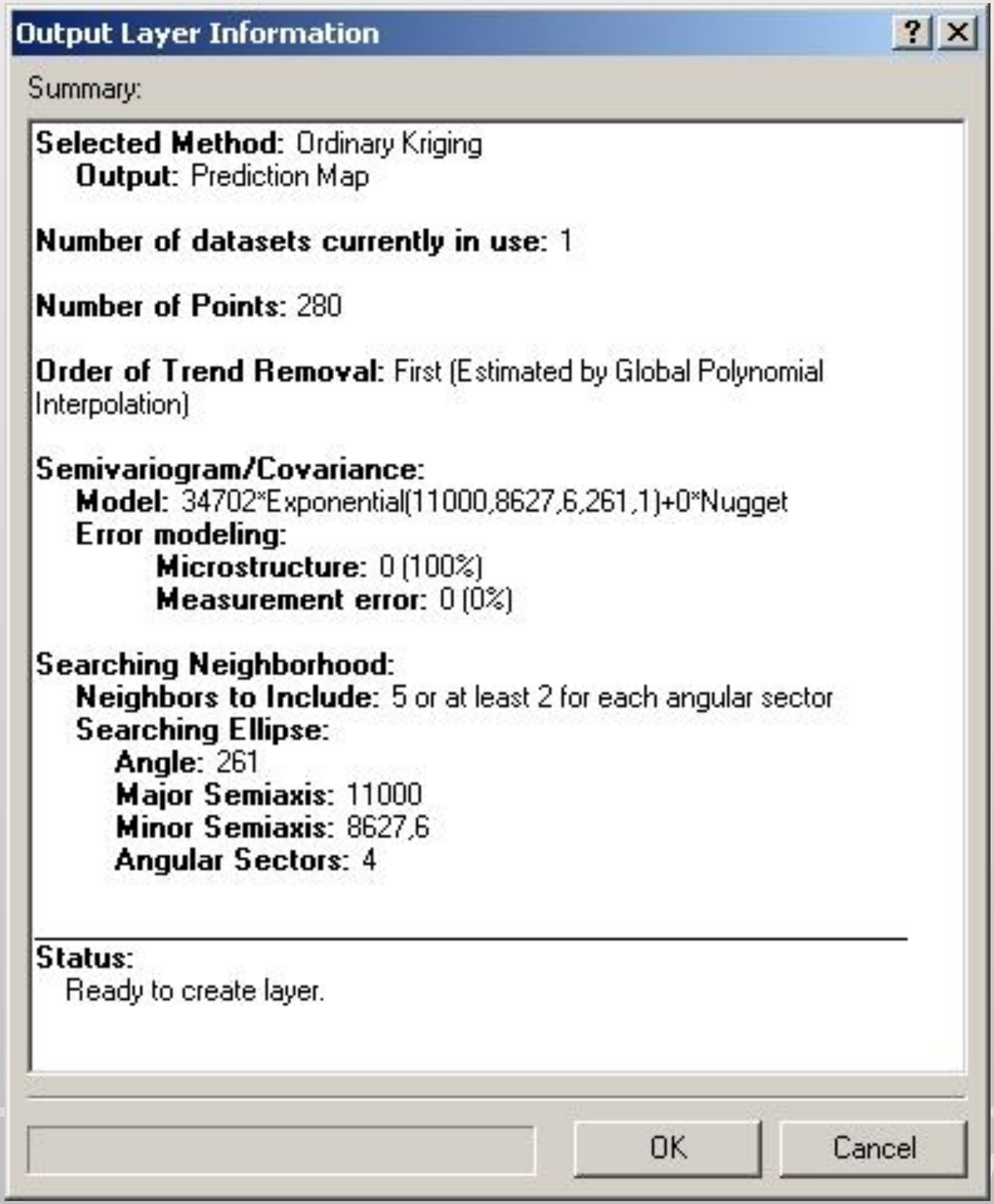
Interpolating the surface using the experimental variogram.



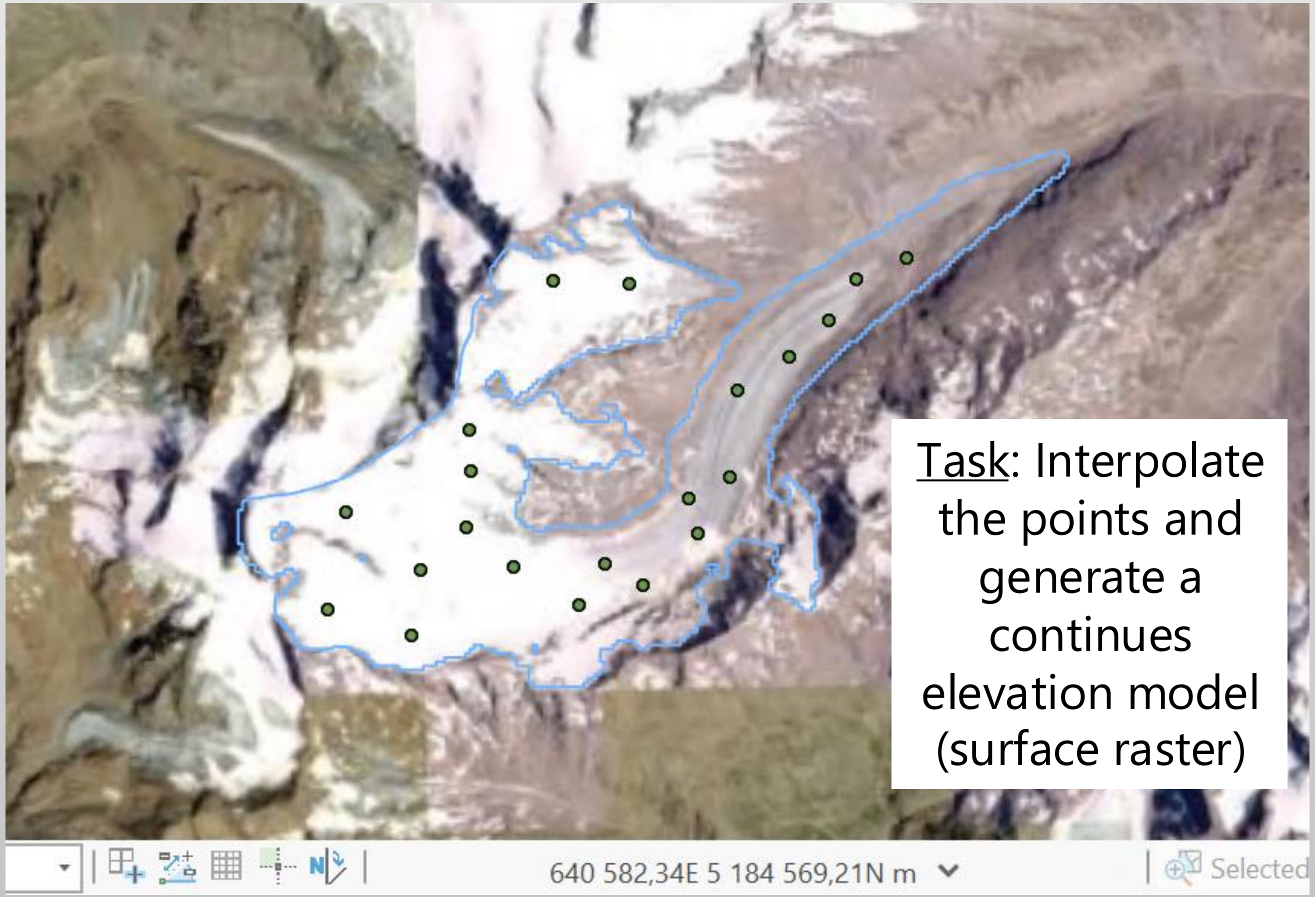
2. Computing the experimental variogram

3. Defining an experimental variogram model that best characterizes the spatial autocorrelation in the data.

Geostatistics: kriging interpolation in ArcGIS



Geostatistics: kriging interpolation in ArcGISPro



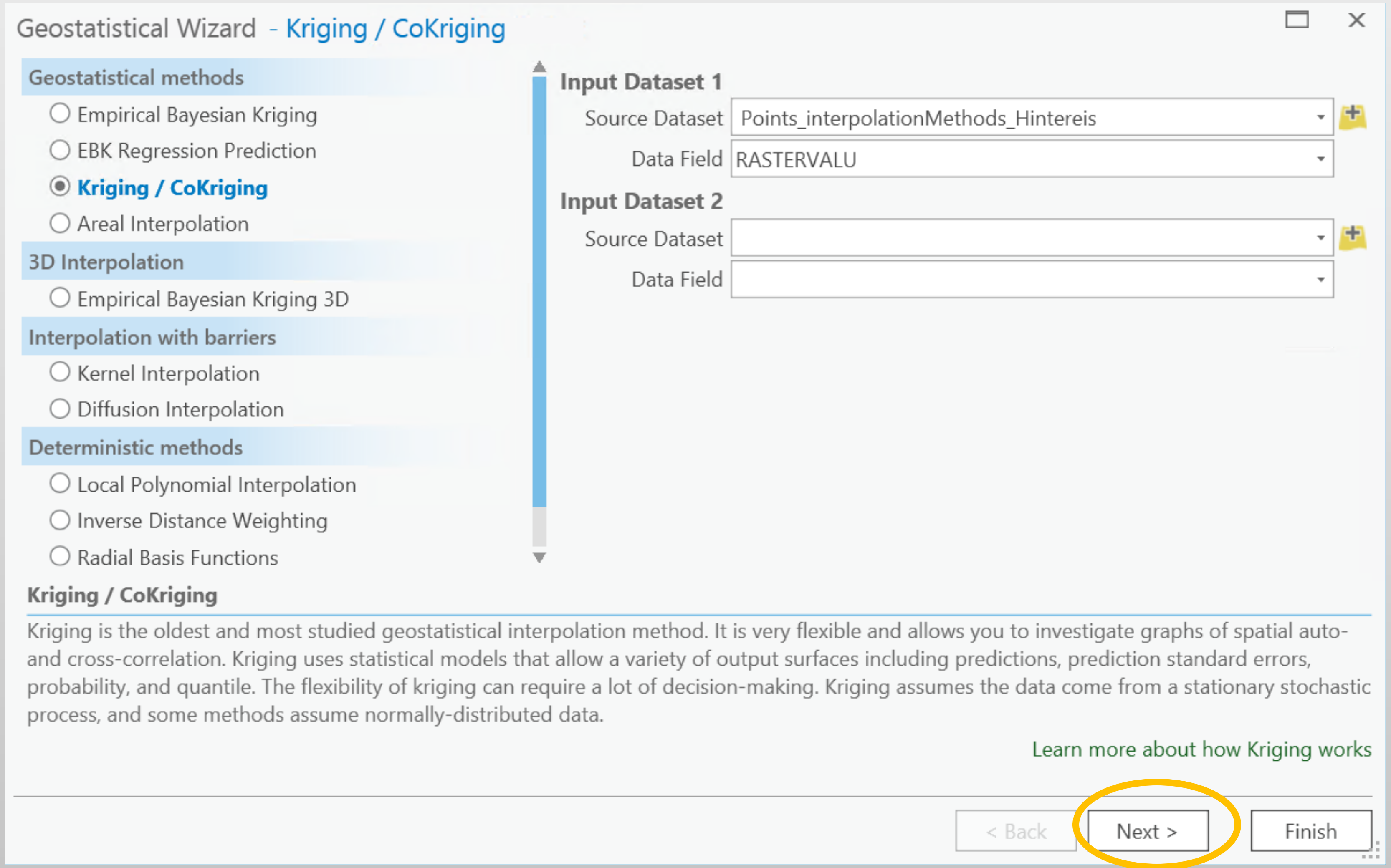
Task: Interpolate the points and generate a continues elevation model (surface raster)

	FID	Shape *	Id	RASTERVALU
1	0	Point	0	2568,93
2	1	Point	0	2612,29
3	2	Point	0	2653,75
4	3	Point	0	2700,09
5	4	Point	0	2736,48
6	5	Point	0	2803,74
7	6	Point	0	2832,32
8	7	Point	0	2849,98
9	8	Point	0	2896,53
10	9	Point	0	2915,8
11	10	Point	0	2946,17
12	11	Point	0	3005,09
13	12	Point	0	3094,7
14	13	Point	0	3224,19
15	14	Point	0	3368,65

0 of 21 selected

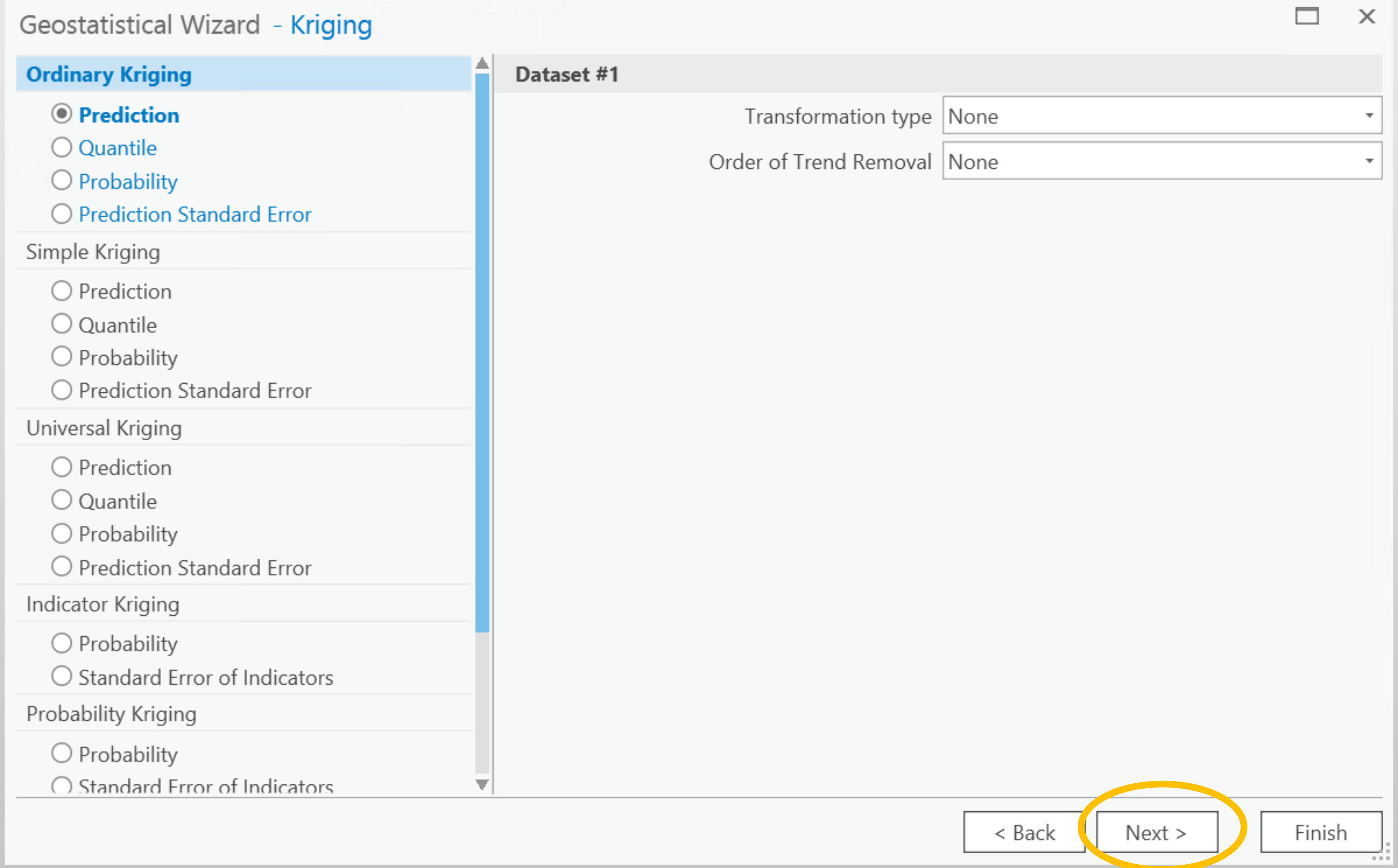
Geostatistics: kriging interpolation in ArcGISPro

Analysis
→ Geostatistical Wizard



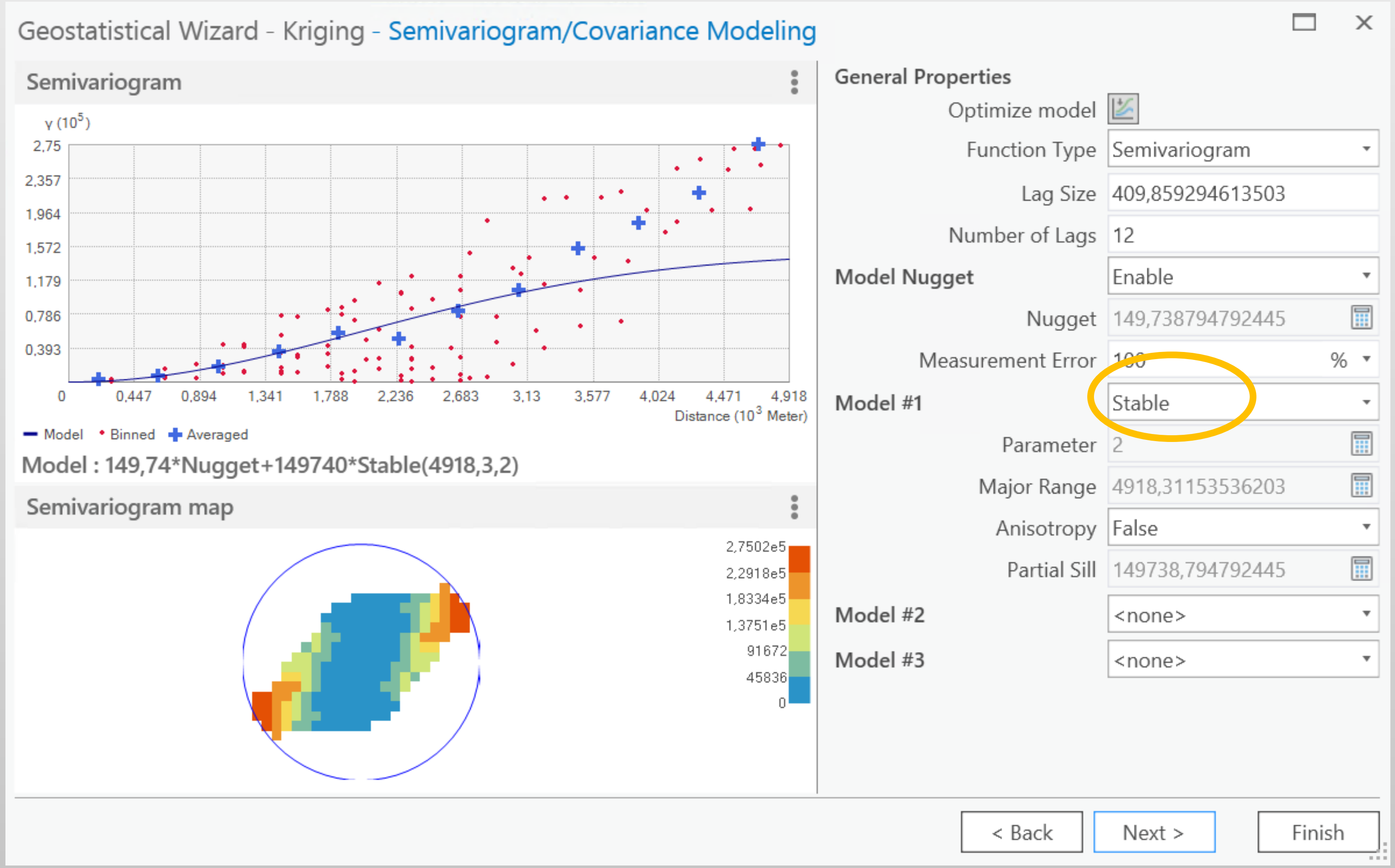
Geostatistics: kriging interpolation in ArcGISPro

Analysis
→ Geostatistical Wizard
→ Ordinary Kriging
- Prediction



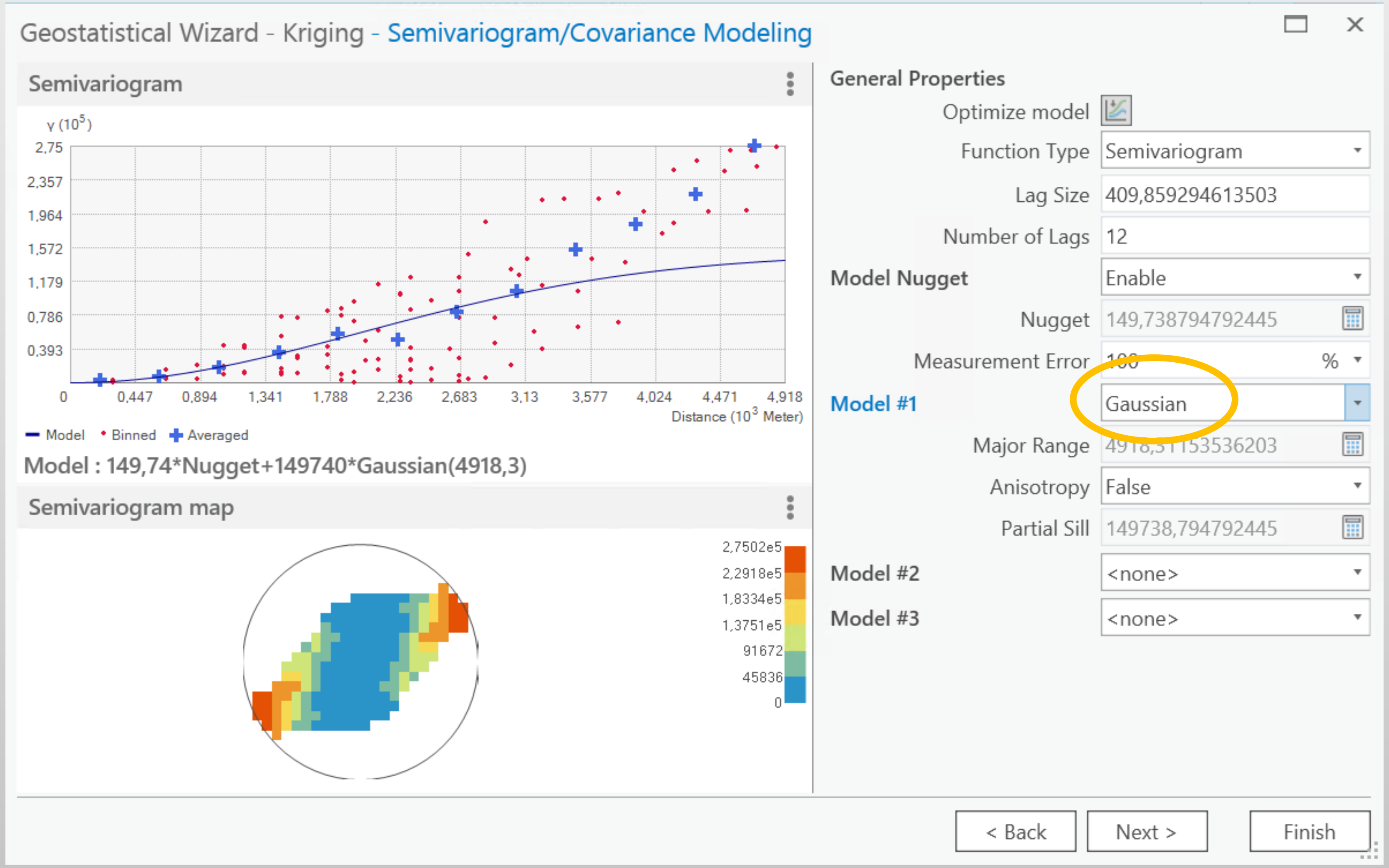
Geostatistics: kriging interpolation in ArcGISPro

Analysis
→ Geostatistical Wizard
→ Ordinary Kriging
- Prediction



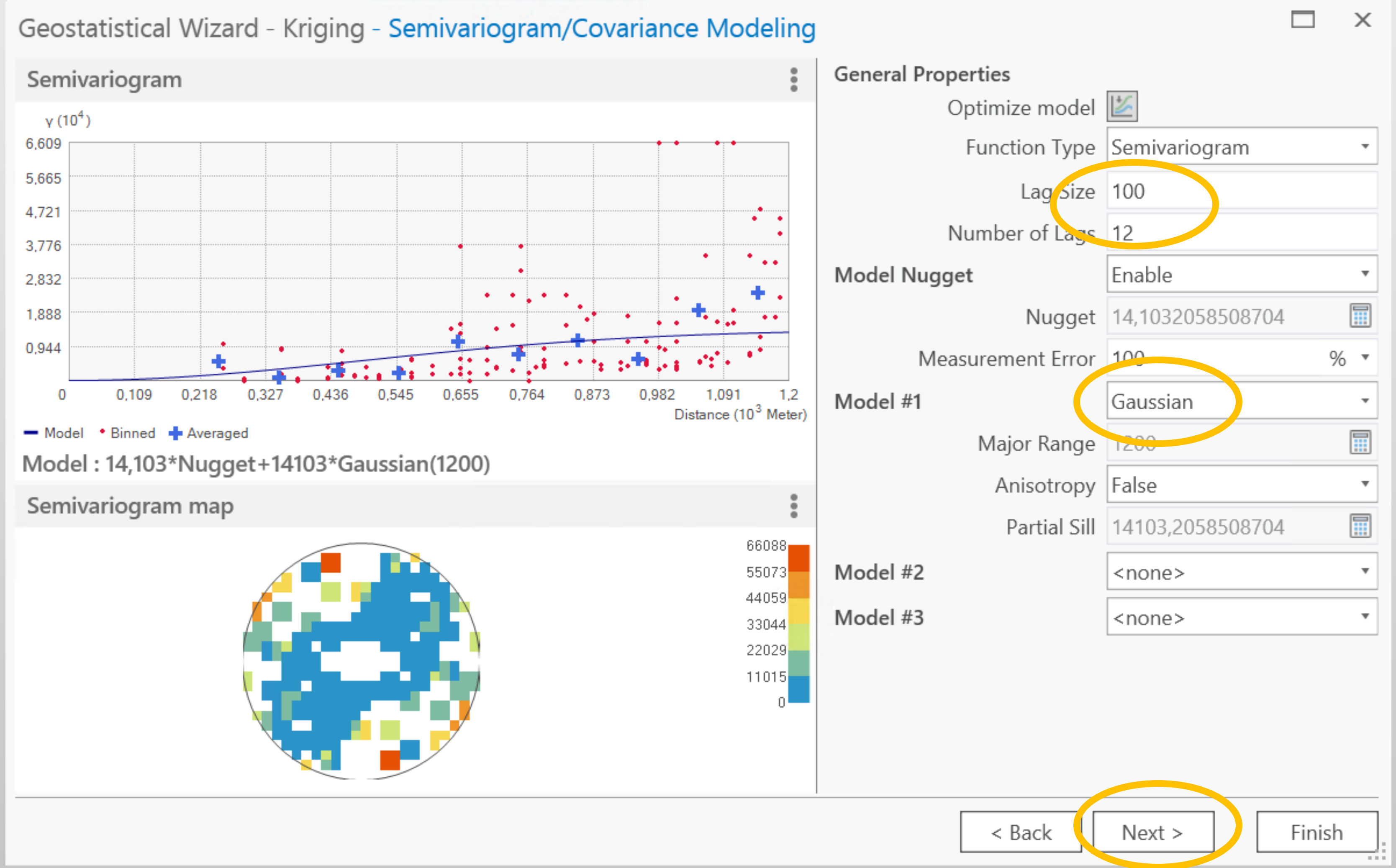
Geostatistics: kriging interpolation in ArcGISPro

Analysis
→ Geostatistical Wizard
→ Ordinary Kriging
- Prediction



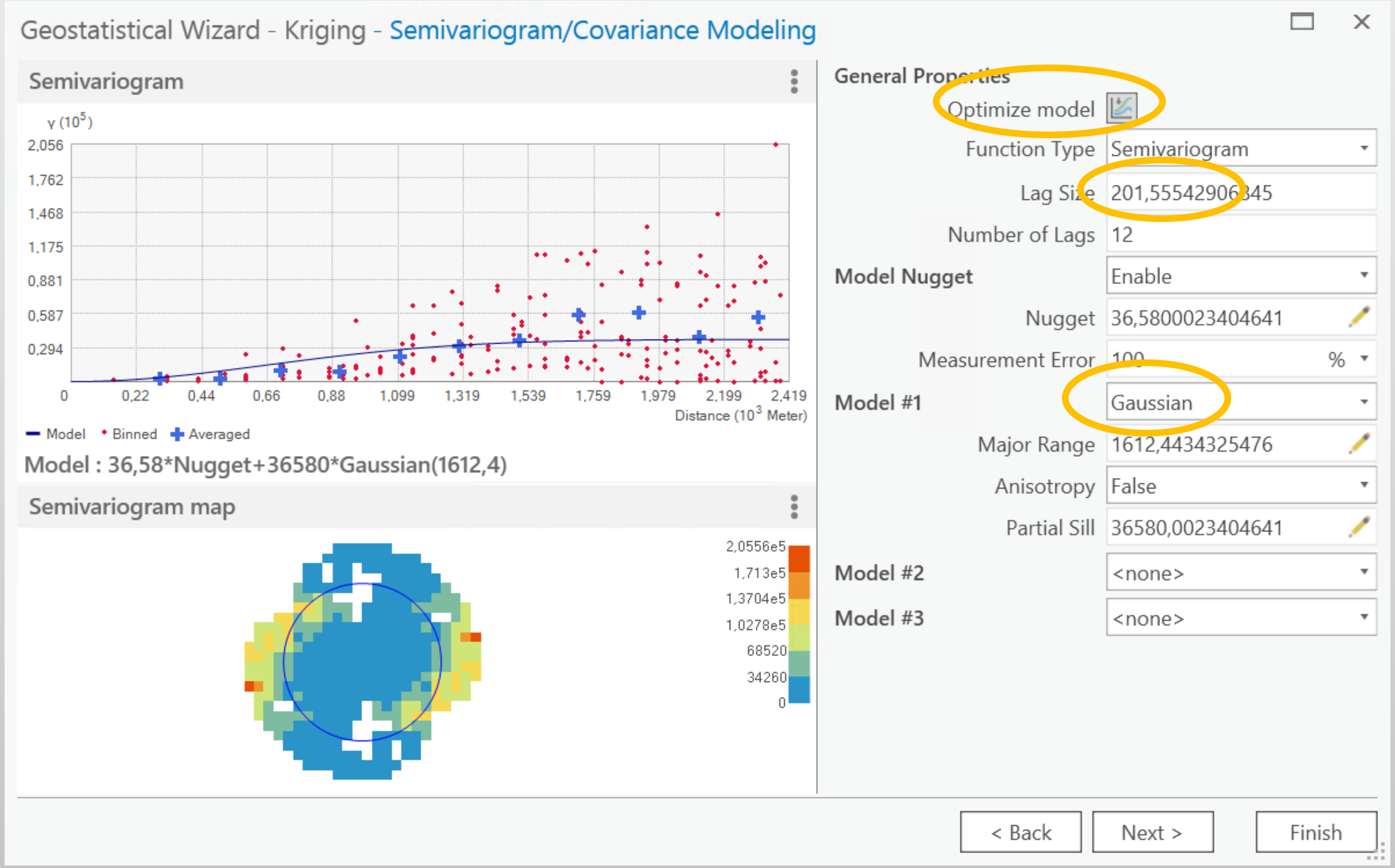
Geostatistics: kriging interpolation in ArcGISPro

Analysis
→ Geostatistical Wizard
→ Ordinary Kriging
- Prediction



Geostatistics: kriging interpolation in ArcGISPro

Analysis
→ Geostatistical Wizard
→ Ordinary Kriging
- Prediction

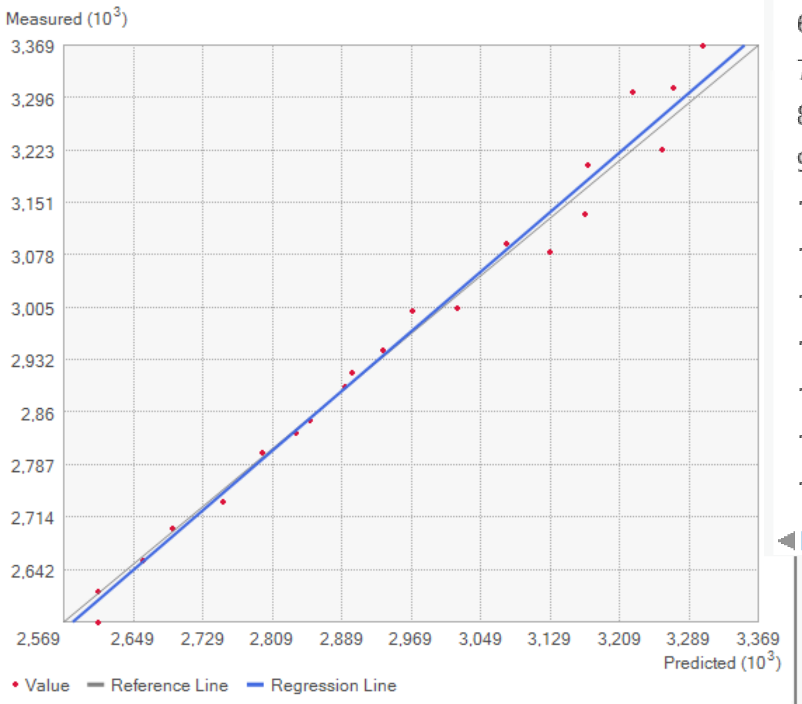


Geostatistics: kriging interpo ArcGISPro

Analysis
→ Geostatistical Wizard
→ Ordinary Kriging
- Prediction

Geostatistical Wizard - Kriging - Cross validation

Normal QQ Plot | Distribution
Predicted | Error | Standardized Error



Regression function: 1,03392748162416 * x + -97,736666776505

Summary | Table

Source ID △	Included	Measured	Predicted
0	Yes	2568,929931640625	2608,380660359934
1	Yes	2612,2900390625	2607,546199618351
2	Yes	2653,75	2659,48993010437
3	Yes	2700,090087890625	2693,003203705832
4	Yes	2736,47998046875	2750,729385042854
5	Yes	2803,739990234375	2797,0456232495
6	Yes	2832,320068359375	2835,11203882072
7	Yes	2849,97998046875	2852,261841743988
8	Yes	2896,530029296875	2891,655770220389
9	Yes	2915,800048828125	2899,81397825937
10	Yes	2946,169921875	2935,29051212356
11	Yes	3005,090087890625	3021,54546769102
12	Yes	3094,699951171875	3077,935401667516
13	Yes	3224,18994140625	3257,05217303706
14	Yes	3368,64990234375	3304,47188769410
15	Yes	3000,6298828125	2969,004309897173
16	Yes	3082,489990234375	3127,23166831878

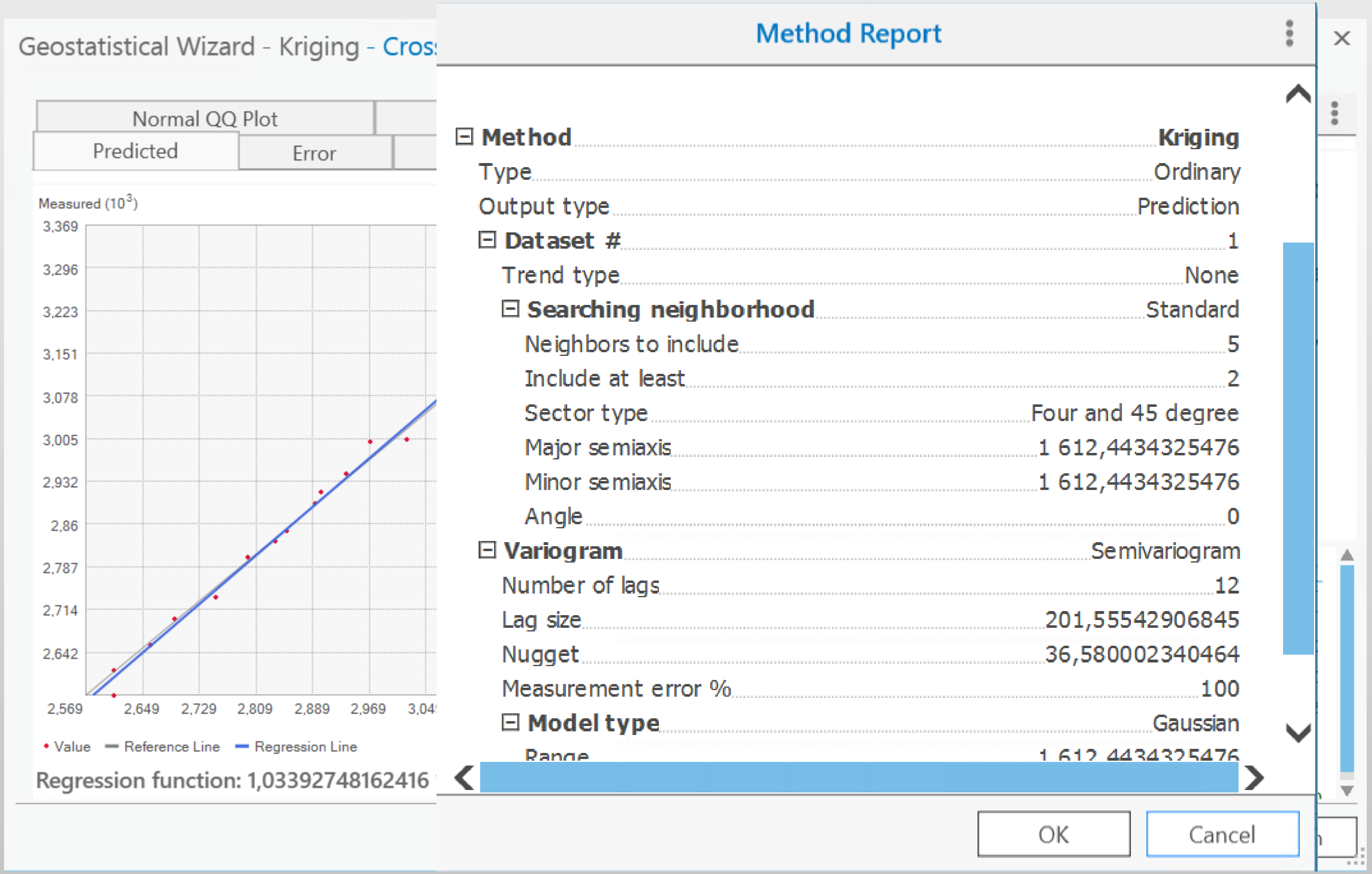
validation works by removing a single point from the dataset and using all remaining points to predict to the location of the point that was removed. The predicted value is then compared to the measured value, and many statistics are generated to determine the accuracy of the prediction.

Learn more about cross validation

< Back | Next > | Finish

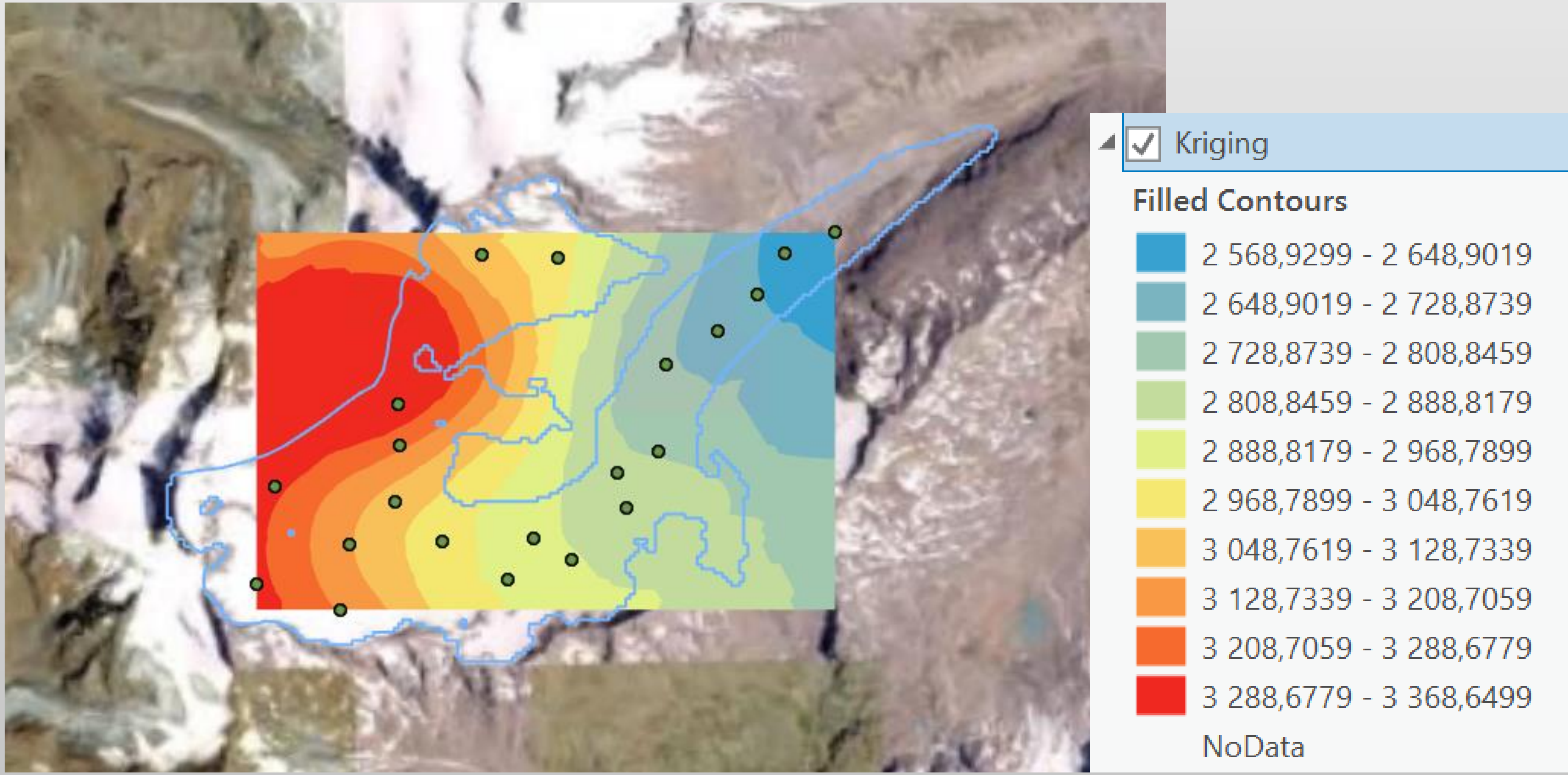
Geostatistics: kriging interpolation in ArcGISPro

Analysis
→ Geostatistical Wizard
→ Ordinary Kriging
- Prediction



Geostatistics: kriging interpolation in ArcGISPro

Analysis
→ Geostatistical Wizard
→ Ordinary Kriging
  - Prediction

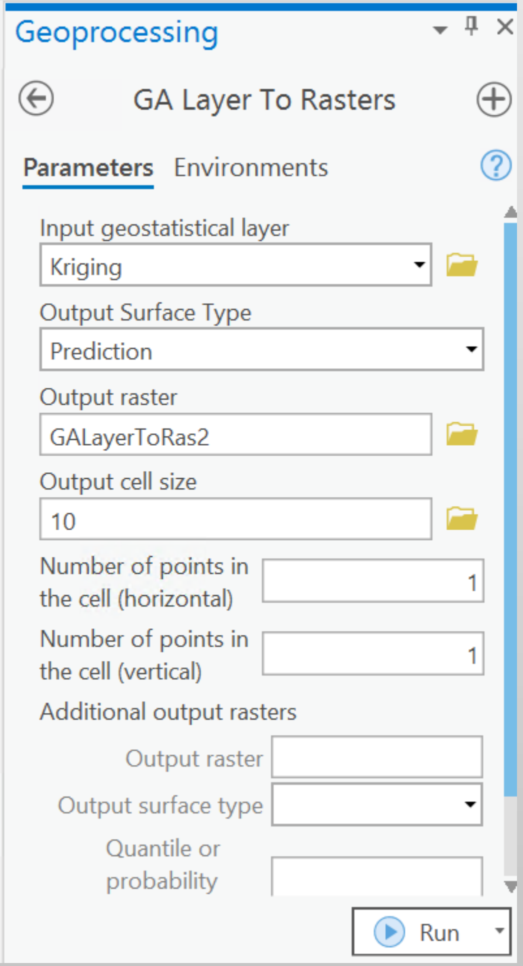


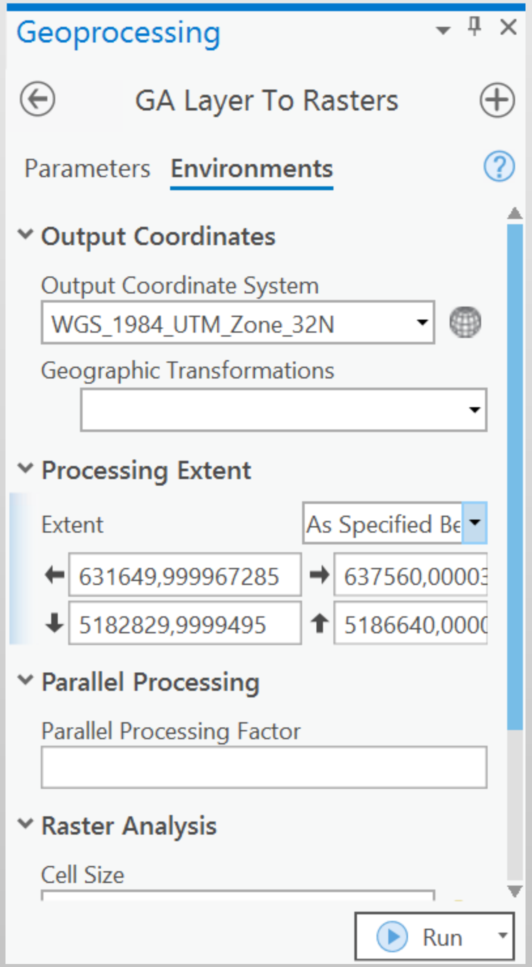
Geostatistics: kriging interpolation in ArcGISPro

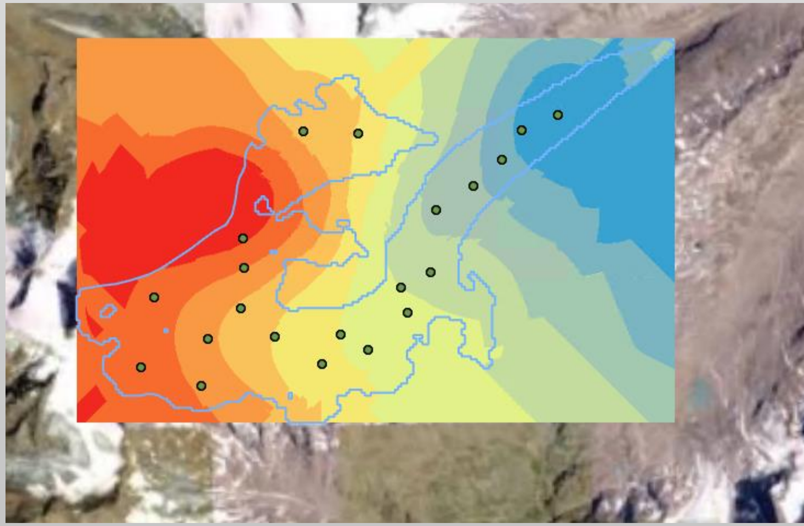
Analysis
→ Geostatistical Wizard
→ Ordinary Kriging
- Prediction

→ Export Layer
- To raster

→ Ordinary Kriging

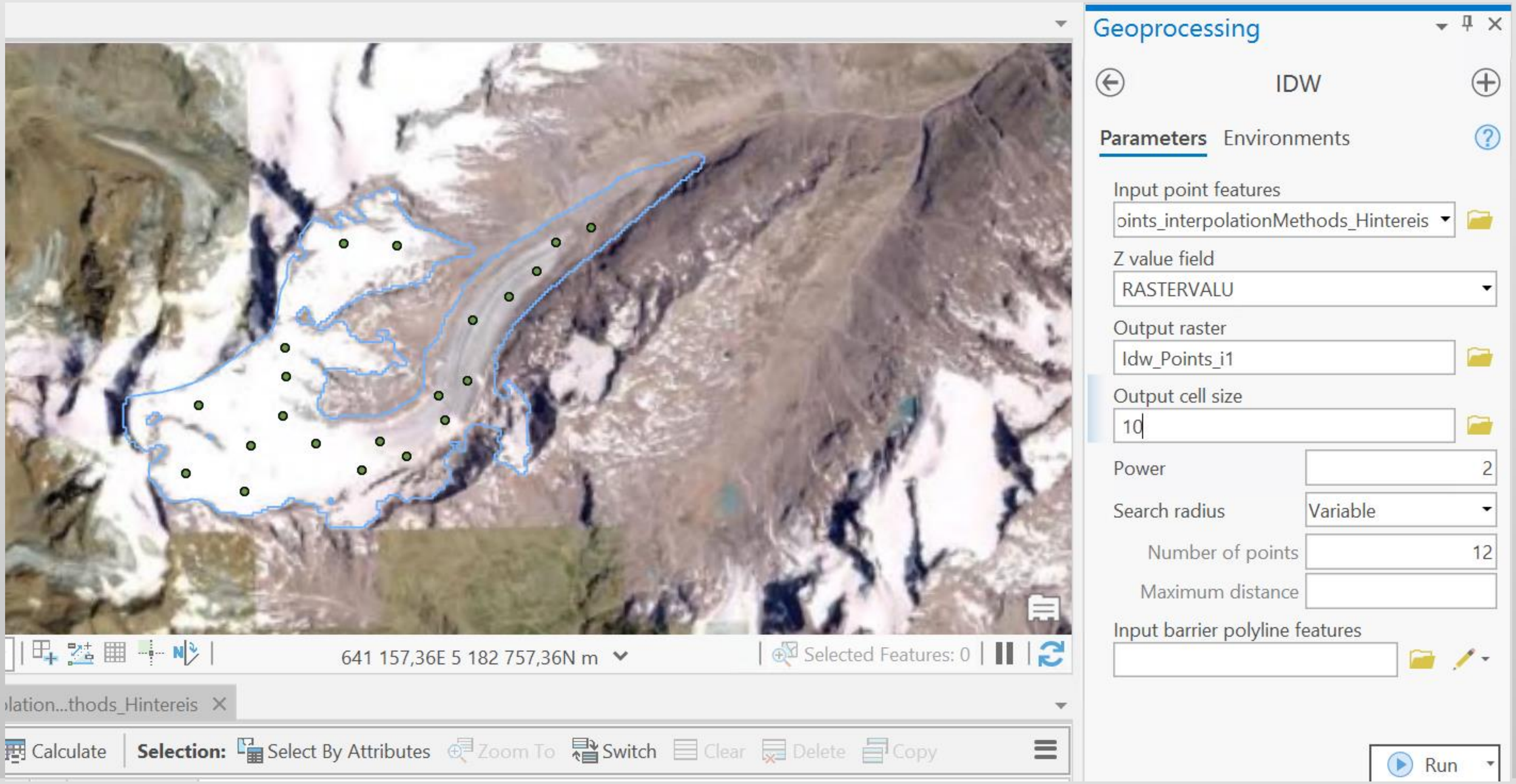






Geostatistics: kriging interpolation in ArcGISPro

Analysis
→ Geoprocessing
→ IDW



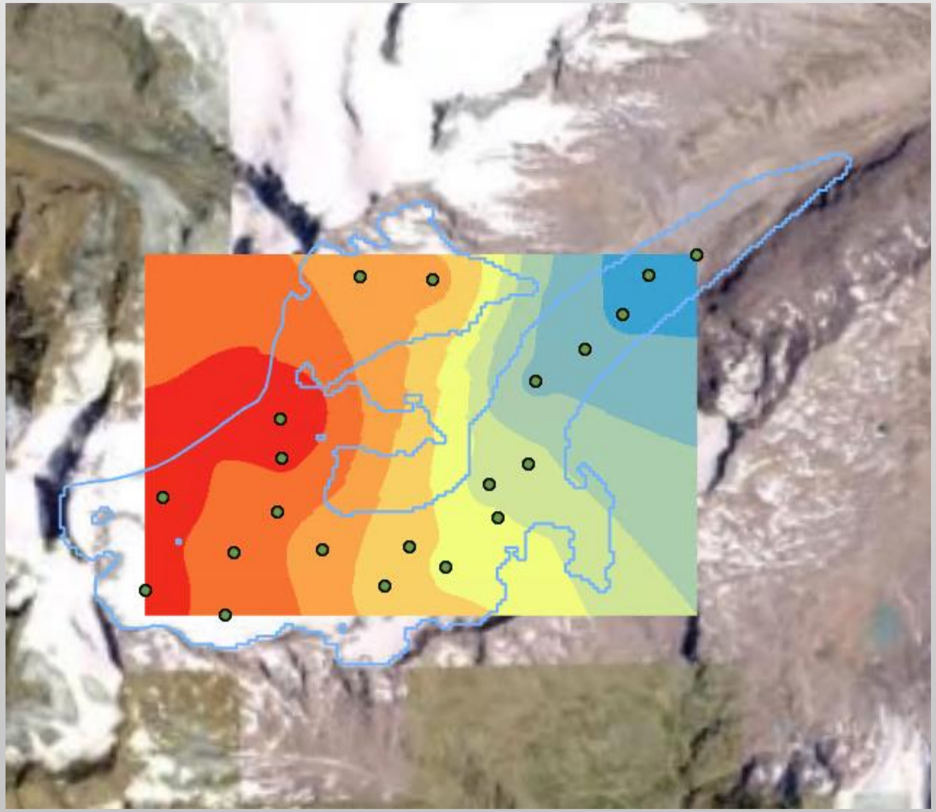
POWER (COEFFIECIENT)= how much weight do you give to the nearest neighbour point close to what we try to interpolate?

Geostatistics: kriging interpolation in ArcGISPro

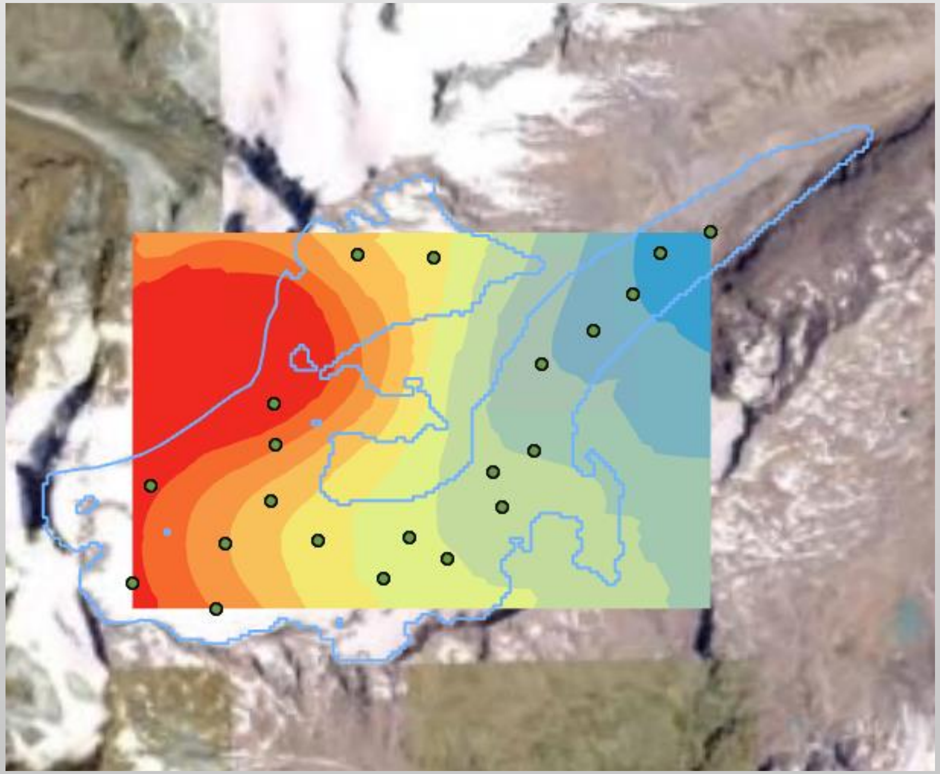
Analysis
→ Geoprocessing
→ IDW

POWER (COEFFICIENT)= how much weight do you give to the nearest neighbour point close to what we try to interpolate?

IDW



Kriging



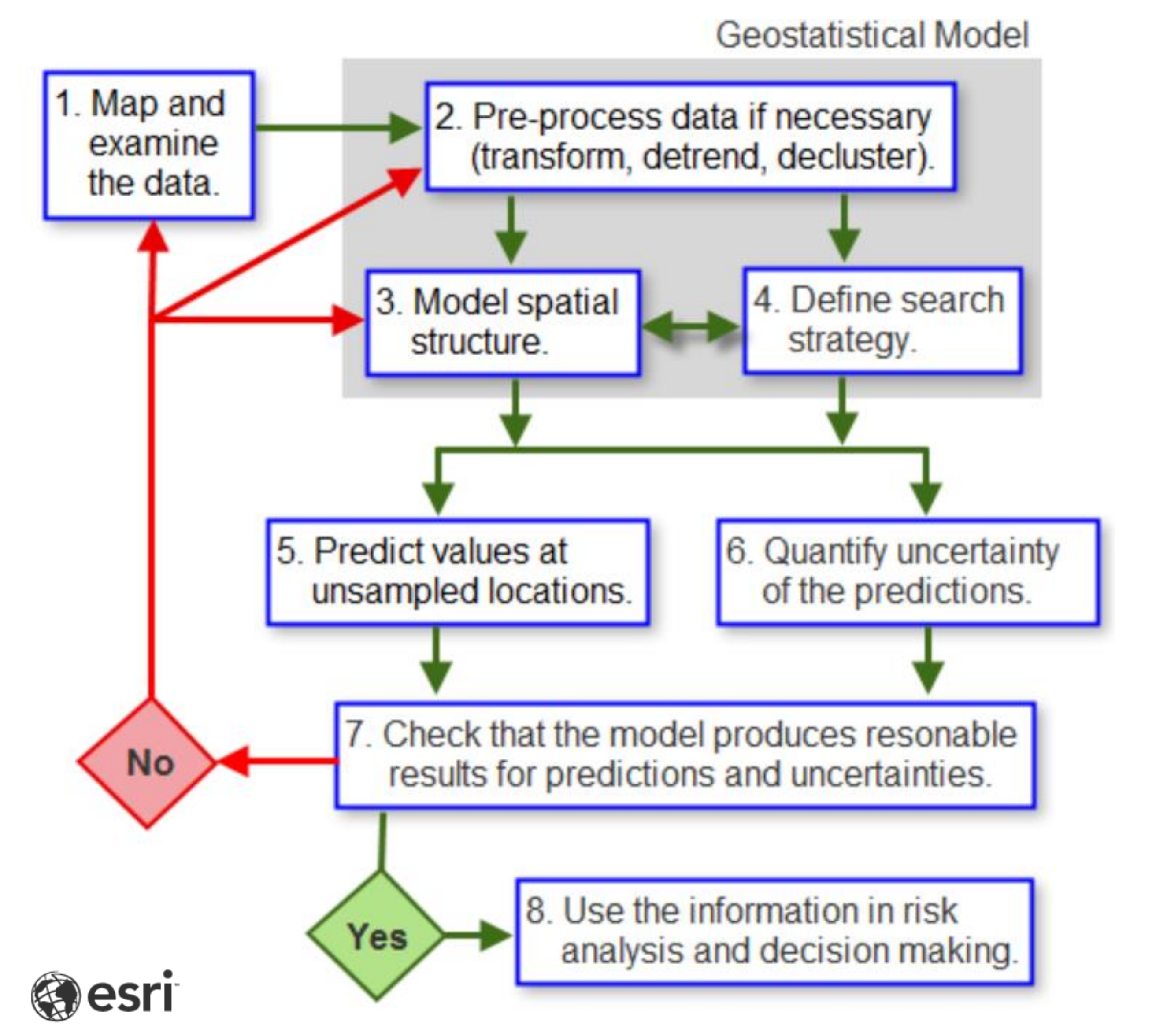
Summary: Regionalised variable

- The nature of the regionalised variable is described by the semivariogram function
- The semivariogram is created from the input data points by plotting the semivariance, y(h), against the distance between the pairs of points (h)
- Semivariogram is used to optimise the interpolation weights and search radius.
- Different models (e.g. spherical, Gaussian, exponential and linear) can be fitted to the semivariogram and that which gives the best fit chosen for the purpose of prediction.

Summary: Spatial variability

- **Geostatistics** tools and methods

  1. Provide interpolated values (predictions for unsampled locations)
  2. Provide measures of uncertainty for those values (prediction)



Thanks!

Feedback questions