L8: Spatial statistics and interpolation

Longley et al., 2005, **Geographic Information Systems and Science**:
- ch. 4: The nature of geographic data
- ch. 14: Query, measurement and tranformation

Sampling of geographic data:

- spatial autocorrelation
- spatial heterogeneity
- sampling

Interpolation

- why interpolation?
- Interpolation methods:
    - Global: classification
    - Local: Thiessen polygons, IDW
    - Geostatistical method: Kriging

Sampling of geographical data

tendency of data objects at some location in space to be related

Spatial autocorrelation (spatial dependency)

1st law of geography (Tobler's law):
Everything is related to everything else, but nearby things are more related than distant things.

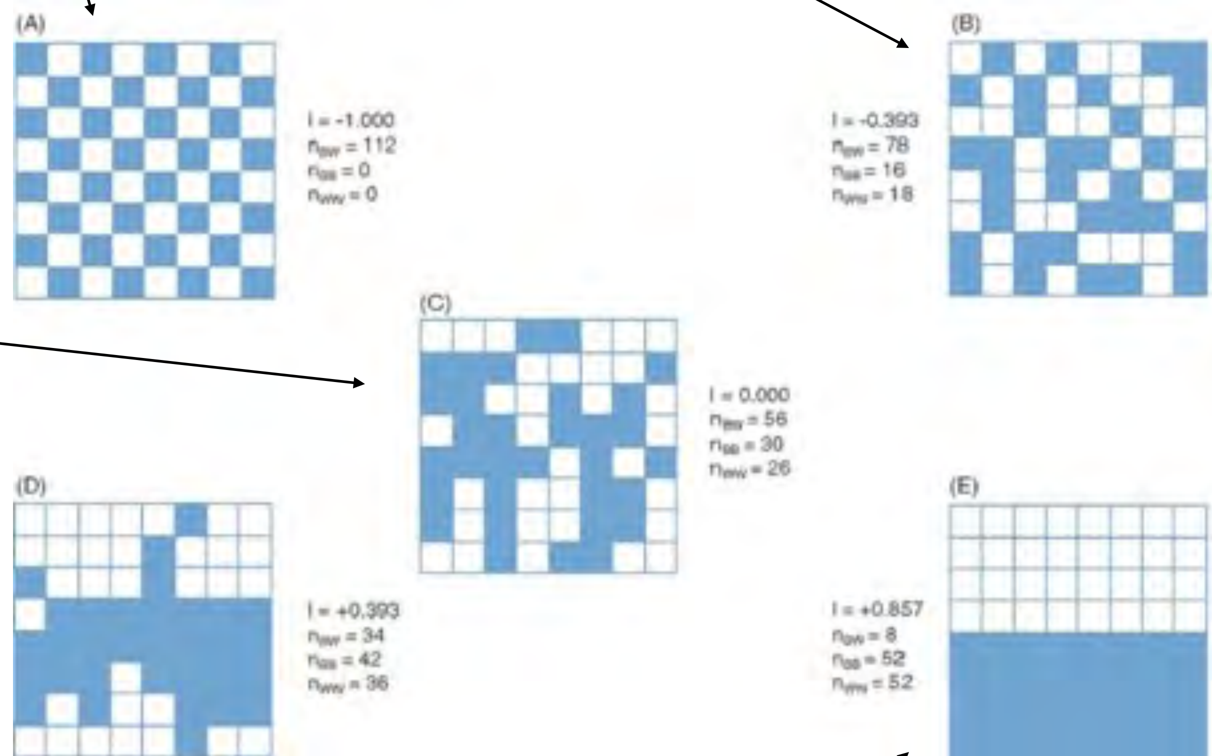
similarities

in position

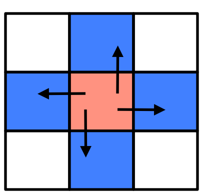
in attributes

Negative spatial autocorrelation: objects that are close togerther in space are more disimilar than objects that are further apart.

Zero spatial autocorrelation: attributes are independent of location.



Neighbourhood: 4 direct neihbouring cells



Positive spatial autocorrelation: objects that are similar in location are also similar in attributes.

Spatial heterogeneity

tendency of geographic places and regions to be different from each other

Differences in

how the landscape looks

how the processes work on the landscape

Example

Sahara desert

Antarctic

differences

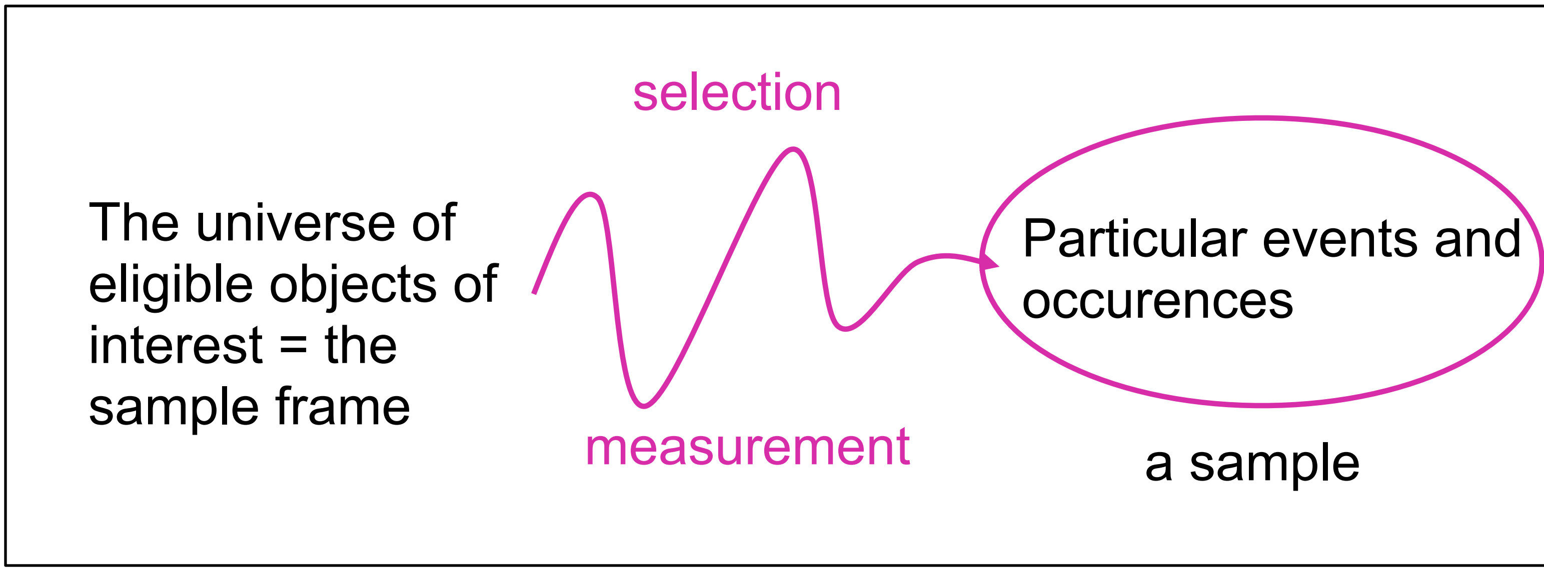
Amazon basin

The Alps

Sampling

It is not possible to use all the objects/events/occurences from the real world in the analysis and representation of geographic phenomena.

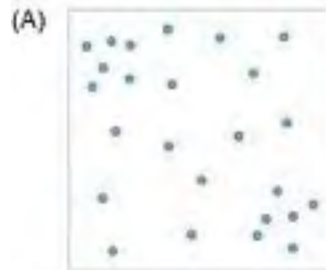
Spatial sampling



Methods for inference allow us to conclude about the characteristics of populations from which the samples were drawn.

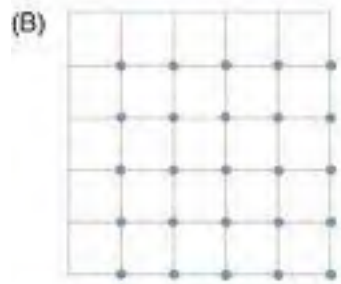
Types of spatial sampling

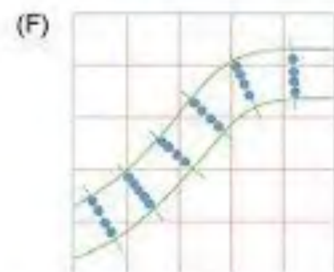
simple random sampling



clustered sampling

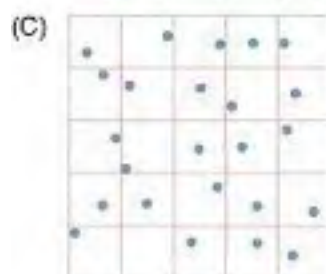
stratified sampling

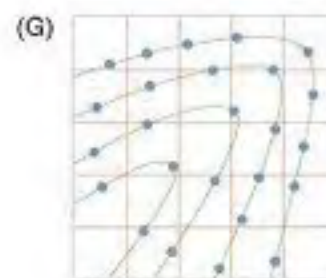




transect sampling

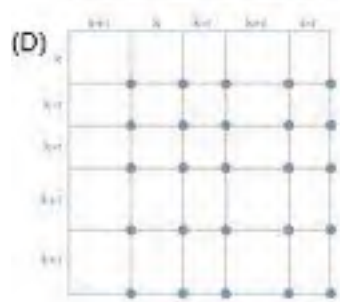
stratified random sampling





contour sampling

stratified sampling with random variation in grid size

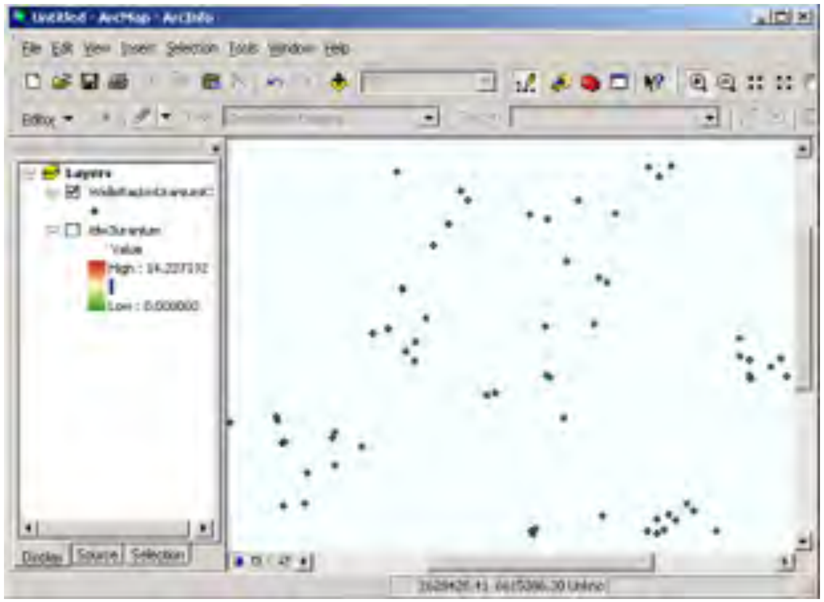


How do we select which locations to take the samples from?

Interpolation

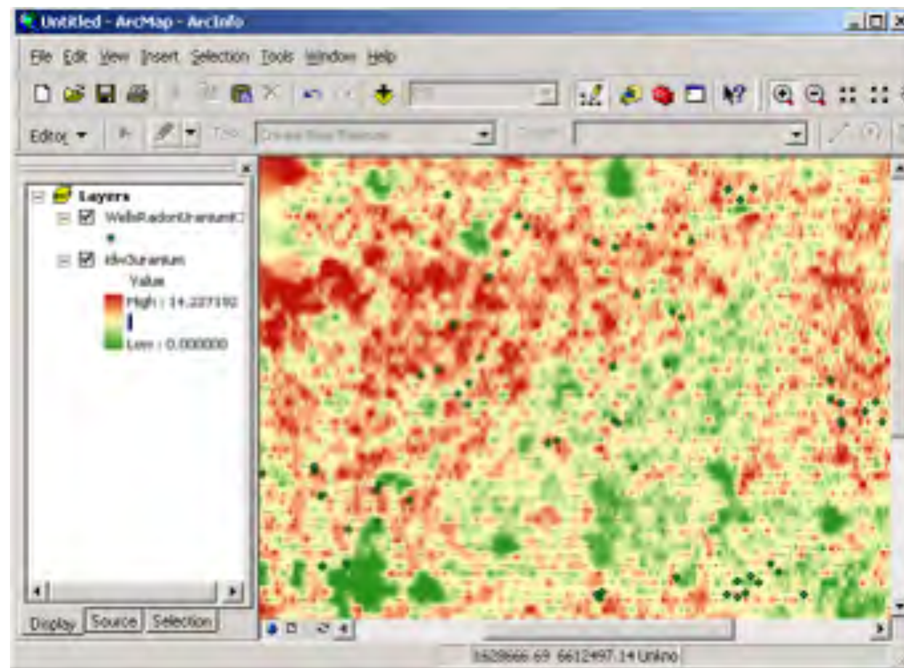
Why interpolation?

How to infer values at unsampled locations?



Values of a field have been measured at a number of sample points

Spatial interpolation

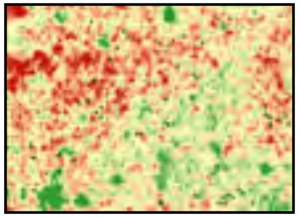


Interpolation is the procedure of predicting the value of attributes at unsampled sites from measurements made at point locations within the same area

(Burrough, 1998)

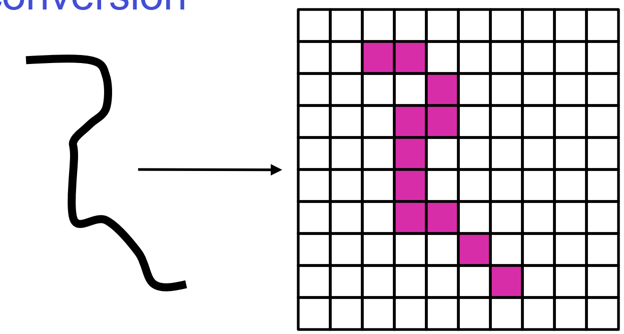
When do we need to interpolate?

converting point data to a continuous field



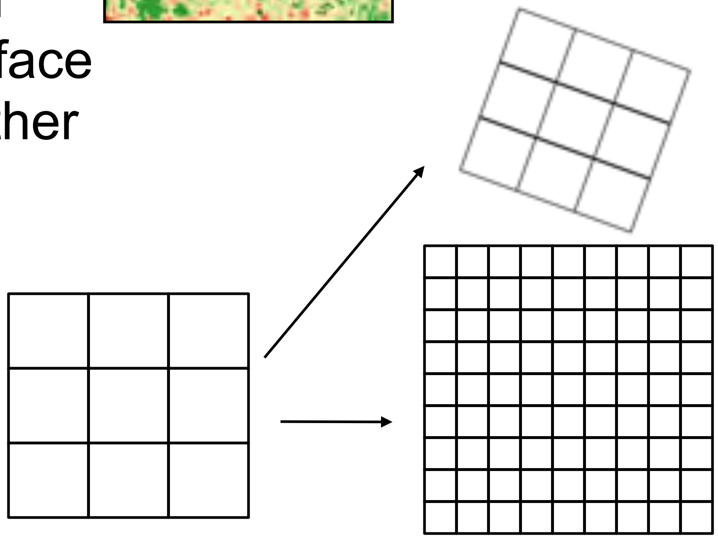
A different data model is needed to represent a continuous surface.

Vector to raster conversion



Resolution or cell orientation of surface is needed in another format.

Conversion of scanned images



Interpolation methods

Global methods

use all available data

Global prediction with classification models

Local methods

use data in the nearest neighbourhood

Inverse-distance weighting (IDW)

Density estimation

Thiessen polygons

L3

Geostatisics

Kriging

Global interpolation methods

Use all available data to predict values for the whole area of interest

Used not for direct interpolation but to examine/remove effects of global variations

Based on standard statistical concepts of mean and variance

Common methods:
prediction by classifcation models, trend surfaces, global regression, etc.

Global prediction using classification models

Areas are divided into regions that can be characterised by the statistical means and variance of attributes measured.

Predictions are based on the mean of all attribute values and the variance in a particular region

Method:
Analysis of variance
(ANOVA)

Assumptions

Most important changes take place at boundaries

Within unit variation is smaller than between units

Typically used in geology: geological maps (bedrock), soil maps

Local interpolation methods

Use data only from the neighbourhood areas of the point of interest to predict the value for the point of interest

Issues

A definition of a search area around the point to be interpolated

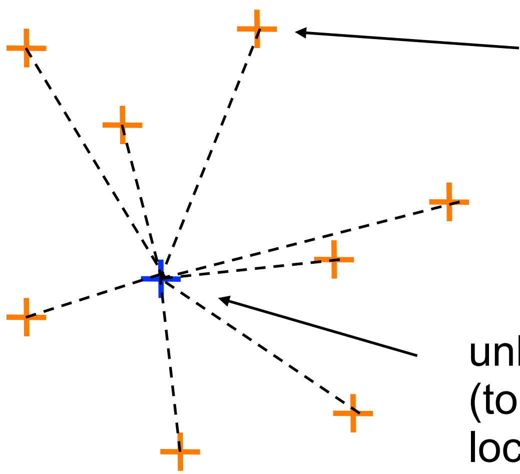
finding the data points within this area

Choosing a mathematical function to represent the variation over this point

Common methods:
IDW, density surfaces, Thiessen polygons, etc.

Inverse-distance weighting (IDW)

The unknown value of a field *z* at a point *x* is estimated by taking a weighted average over the known values:



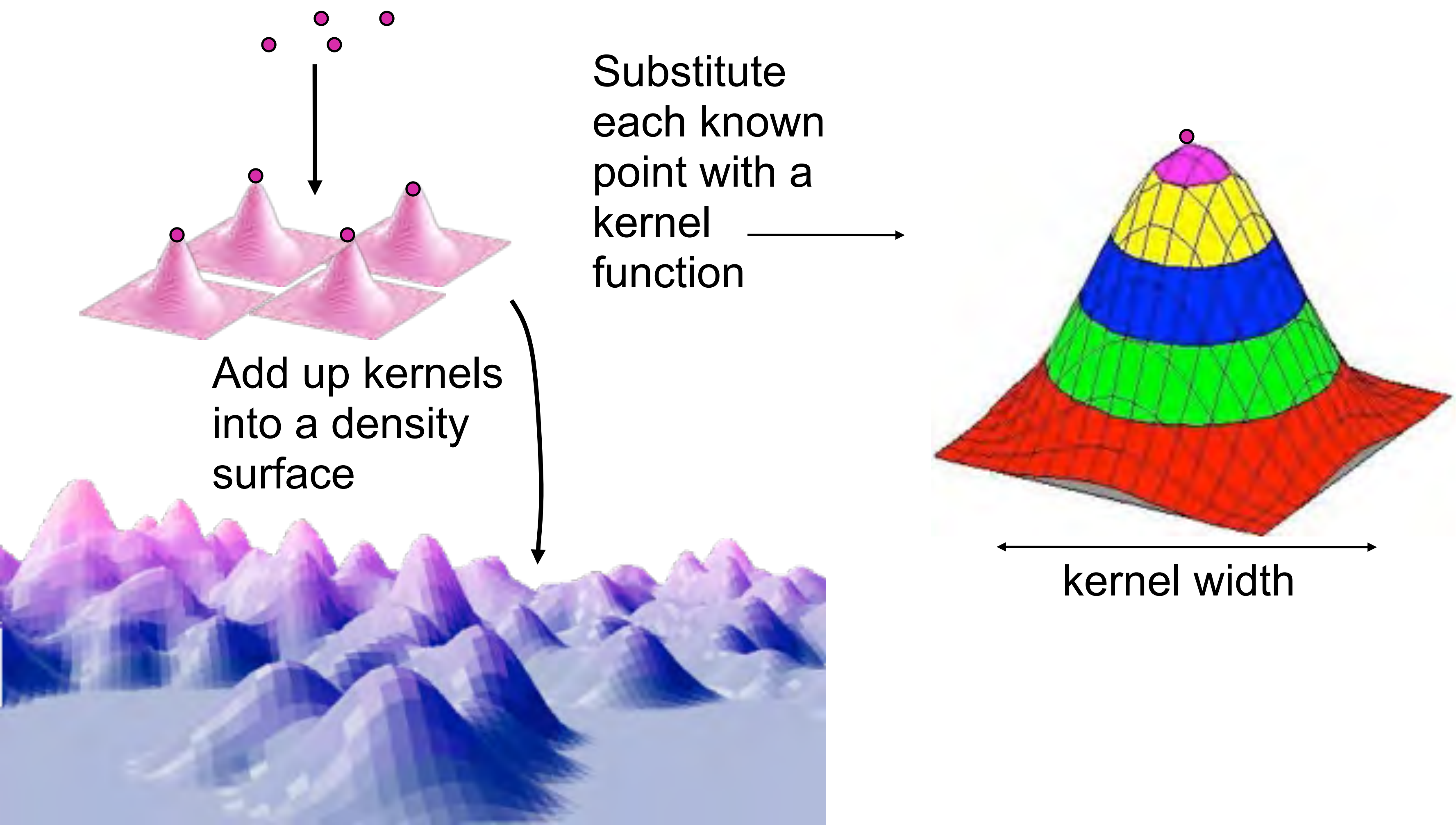
$$z(\mathbf{x}) = \sum_i w_i z_i \Big/ \sum_i w_i$$

$$w_i = 1 \big/ d_i^r$$

Each known value is weighted by its distance from the point x: weights decrease with the r^{th} power of distance (usually r=2).

Density estimation

Density estimation creates a field from discrete point objects: the field's value at any point is an estimate of the density of discrete objects at that point.



Thiessen polygons

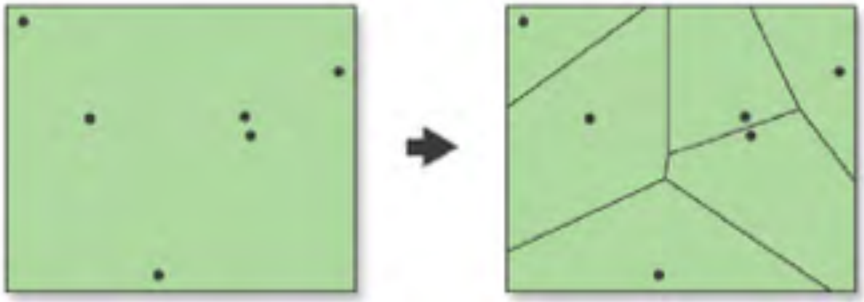
Also known as:
nearest neighbour interpolation

Predictions are provided by the attribute of the nearest sampled point

The form of the surface is determined by distribution of observations. Each point defines a polygon with the following two characteristics:

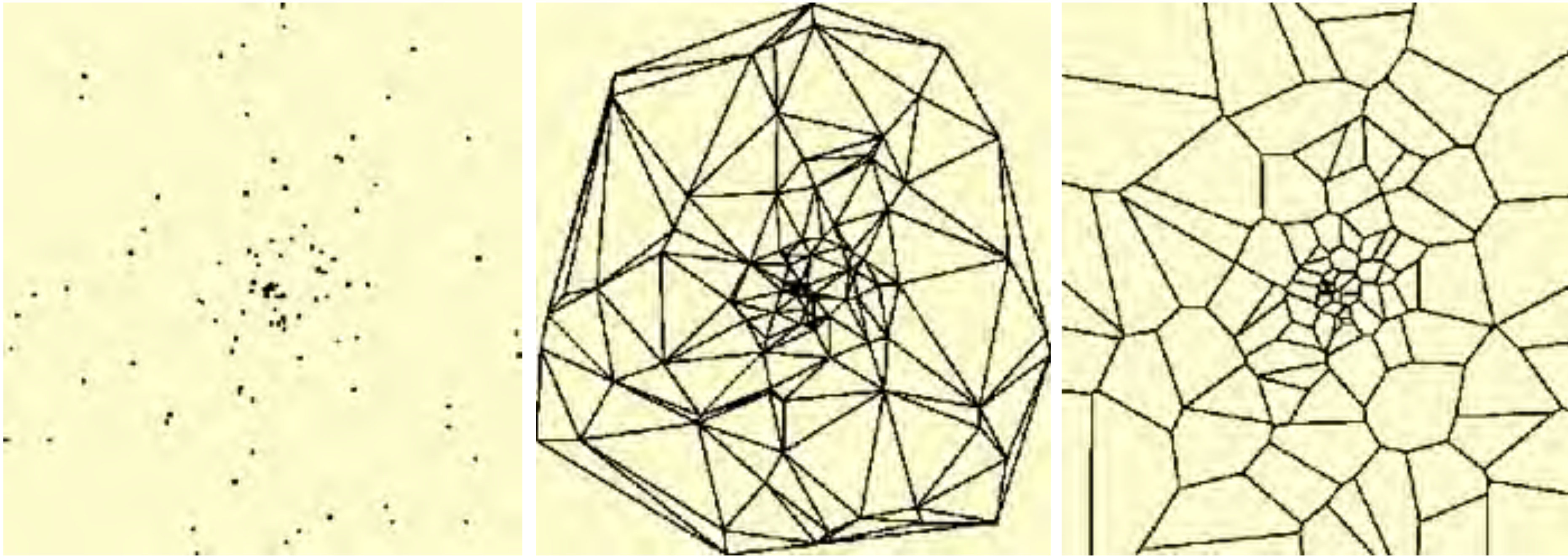
- each polygon contains exactly one input point

- any location within a polygon is closer to its associated point than to any other point.



Thiessen polygons or Voronoi polygons

How Thiessen polygons are calculated:



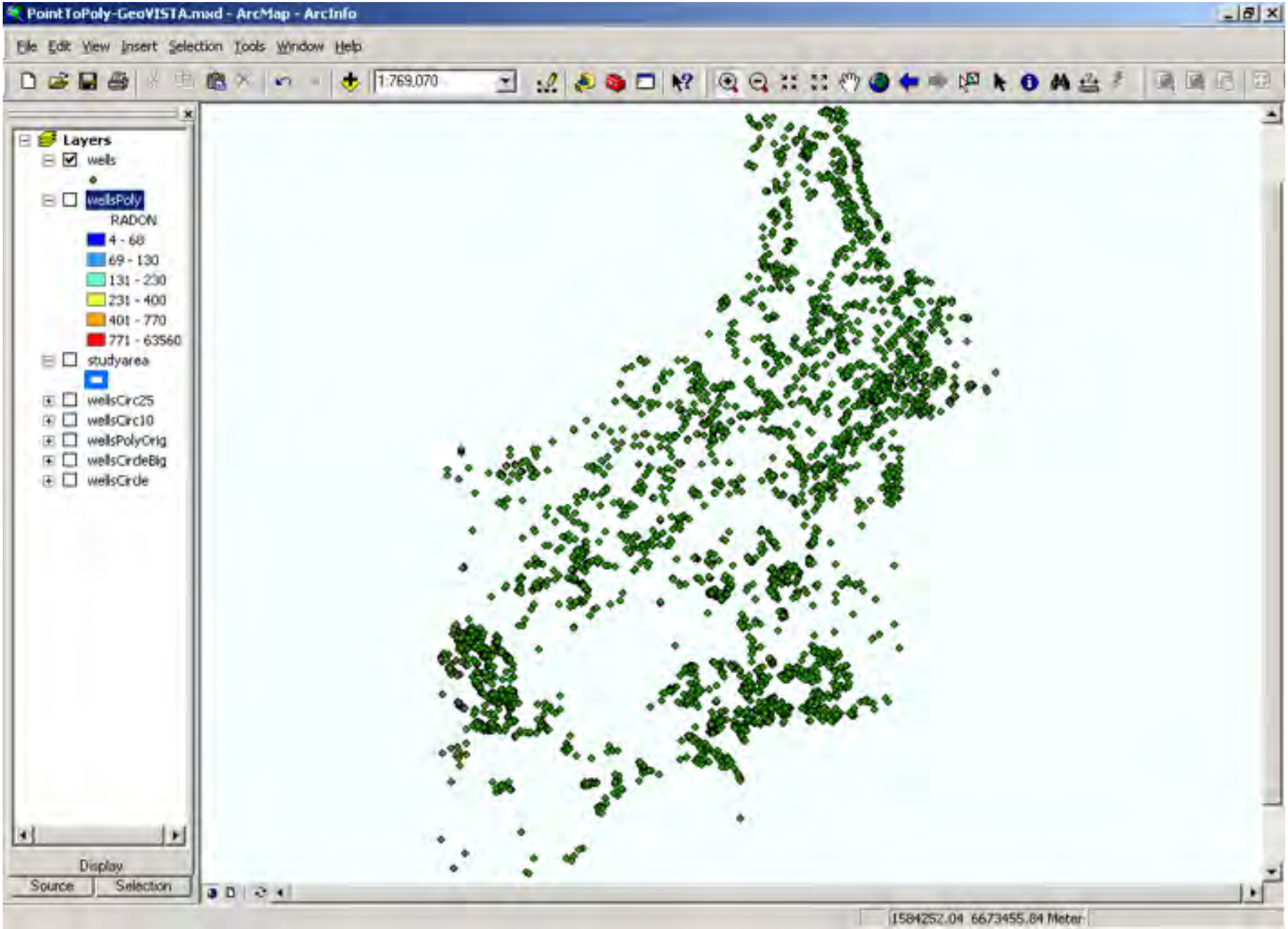
Original data points → Delauney triangulation Thiessen polygons

A triangulation of the vertex set with the property that no vertex in the vertex set falls in the interior of the circumcircle (circle that passes through all three vertices) of any triangle in the triangulation.

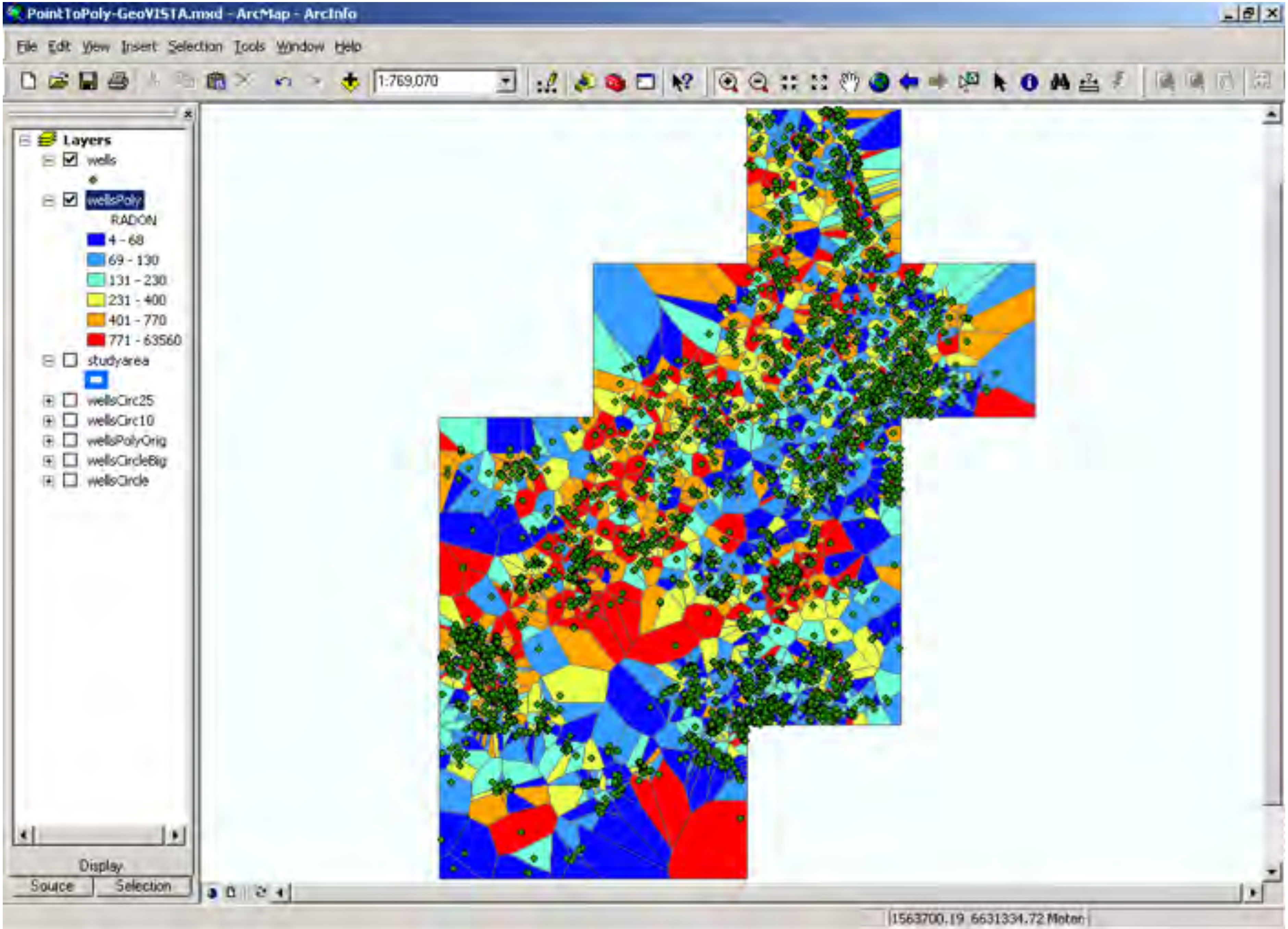
the geometric dual

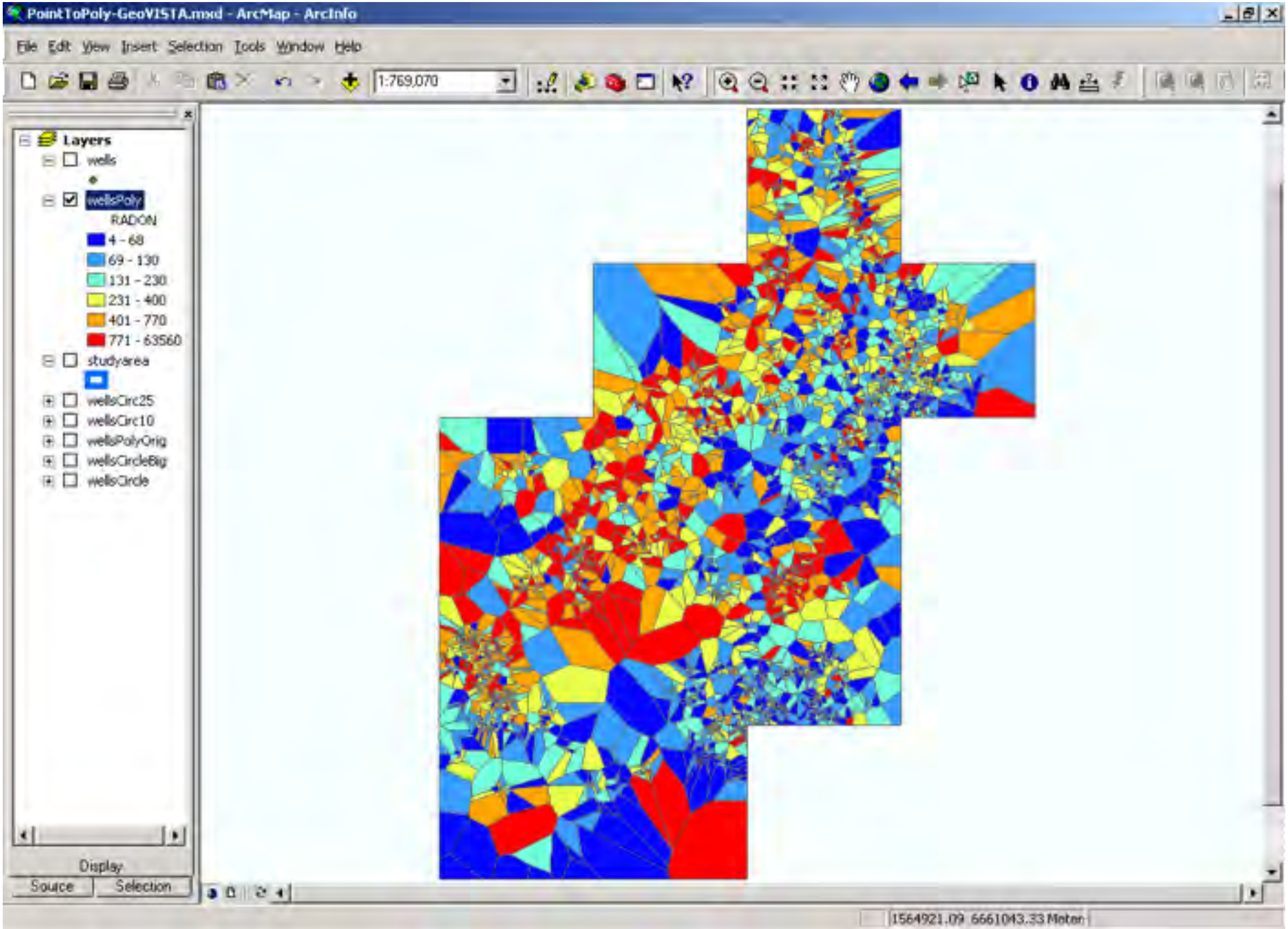
Each polygon is assigned the attribute value of the point that belongs to it.

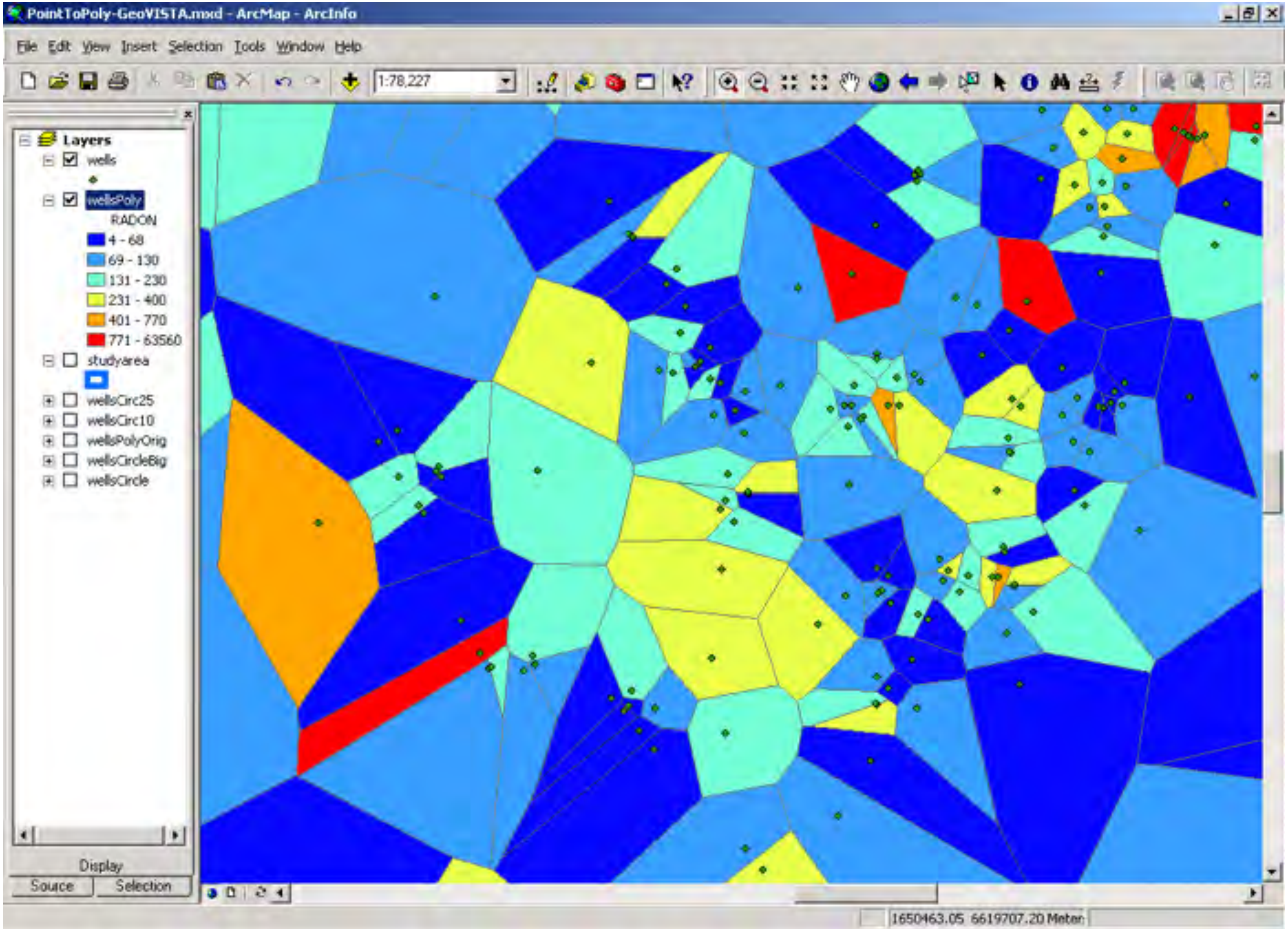
TIN – triangulated irregular network



PointToPoly-GeoVISTA.mxd - ArcMap - ArcInfo
File Edit View Insert Selection Tools Window Help
1:769,070
Layers
wells
wellsPoly
RADON
4 - 68
69 - 130
131 - 230
231 - 400
401 - 770
771 - 63560
studyarea
wellsCirc25
wellsCirc10
wellsPolyOrig
wellsCircleBig
wellsCircle
Display
Source
Selection
1584252.04 6673455.84 Meter


PointToPoly-GeoVISTA.mxd - ArcMap - ArcInfo
File Edit View Insert Selection Tools Window Help
1:769,070
Layers
wells
wellsPoly
RADON
4 - 68
69 - 130
131 - 230
231 - 400
401 - 770
771 - 63560
studyarea
wellsCirc25
wellsCirc10
wellsPolyOrig
wellsCircleBig
wellsCircle
Display
Source
Selection
1563700.19 6631334.72 Meters


PointToPoly-GeoVISTA.mxd - ArcMap - ArcInfo
File Edit View Insert Selection Tools Window Help
1:769,070
Layers
wells
wellsPoly
RADON
4 - 68
69 - 130
131 - 230
231 - 400
401 - 770
771 - 63560
studyarea
wellsCirc25
wellsCirc10
wellsPolyOrig
wellsCircleBig
wellsCircle
Display
Source
Selection
1564921.09 6661043.33 Meters


PointToPoly-GeoVISTA.mxd - ArcMap - ArcInfo
File Edit View Insert Selection Tools Window Help
1:78,227
Layers
wells
wellsPoly
RADON
4 - 68
69 - 130
131 - 230
231 - 400
401 - 770
771 - 63560
studyarea
wellsCirc25
wellsCirc10
wellsPolyOrig
wellsCircleBig
wellsCircle
Display
Source
Selection
1650463.05 6619707.20 Meters

A geostatistical interpolation method: Kriging

Previous interpolation methods

What is the quality of the estimates ?

No detailed/reliable information on how to:
- define the number of points needed to compute the local average
- define the size/shape/orientation of neighbourhood
- Ways to estimate the interpolation weight?
- estimate errors associated with interpolated value

Kriging

A technique of spatial interpolation firmly grounded in geostatistical theory

Developed for use in the mining industry

Underlying principle for kriging: spatial variation of any continuous attribute is too irregular to be modelled by a simple, smooth mathematical function.

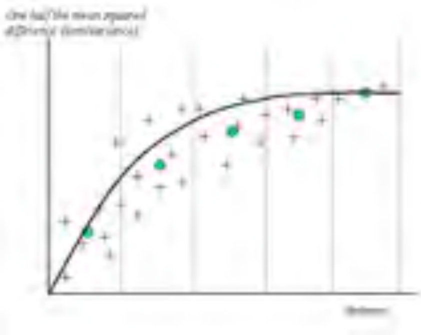
Similar to IDW

Variation is instead described by a stochastic surface, obtained as a weighted combination of neighbouring point values, where weights are derived using **a semivariogram**.

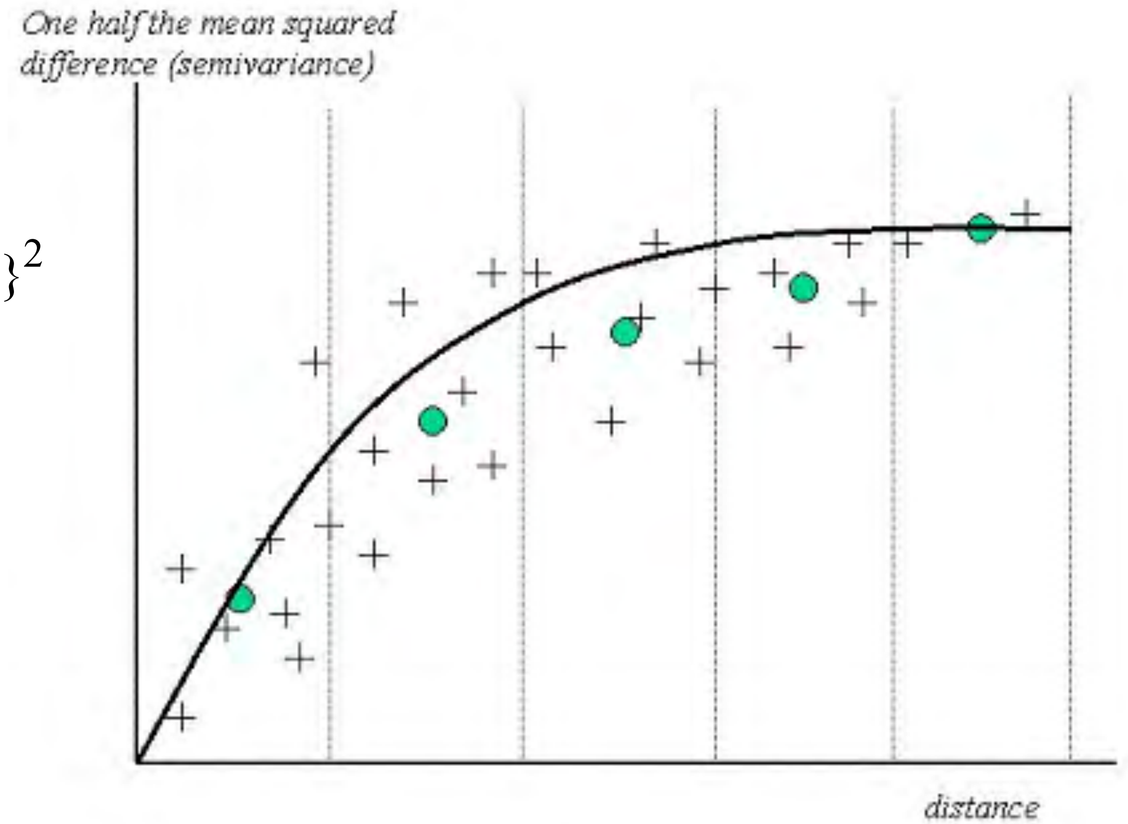
The semivariogram reflects Tobler's Law

differences within a small neighborhood are likely to be small

differences rise with distance



$$\gamma(x) = \frac{1}{2n} \sum_{i=1}^{n} \{z(x_i) - z(x)\}^2$$



A semivariogram. Each cross represents a pair of points. The solid circles are obtained by averaging within the ranges or *bins* of the distance axis. The solid line represents the best fit to these five points, using one of the standard mathematical functions.

The semivariogram

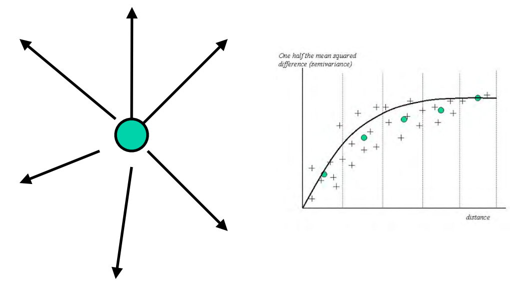
isotropic

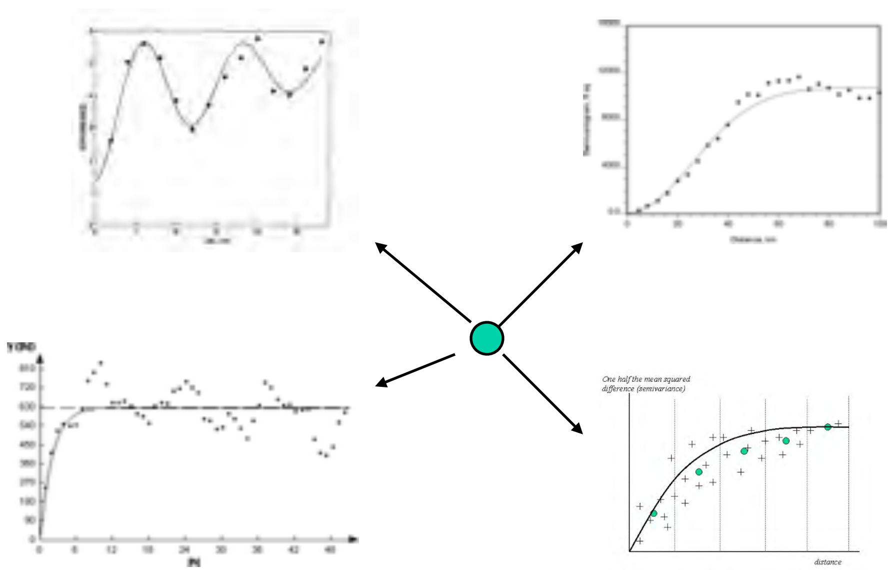
Behaviour of the phenomenon is the same in all directions.

anisotropic

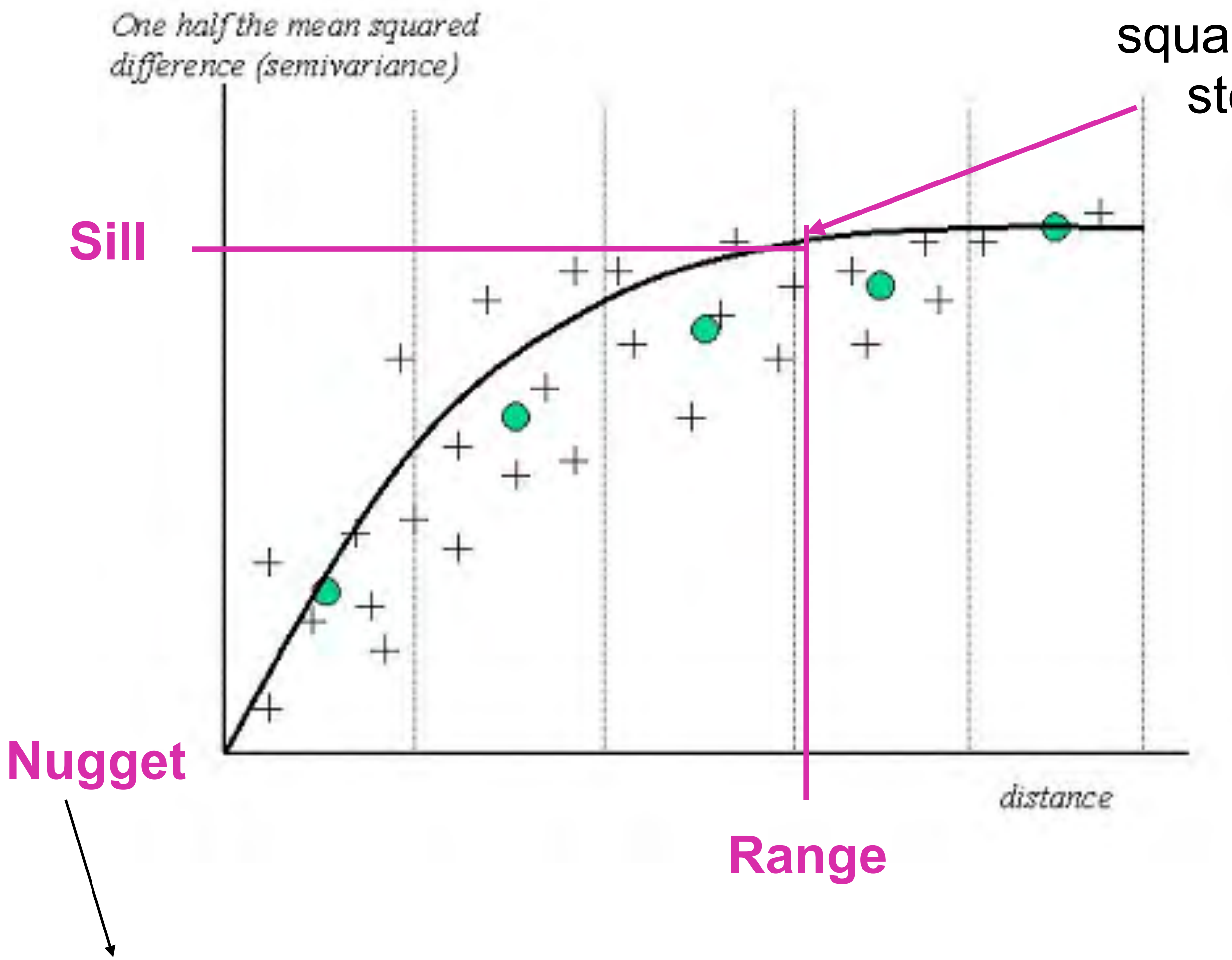
A separate semivariogram is needed for each direction.

Behaviour of the phenomenon is very different in different directions.



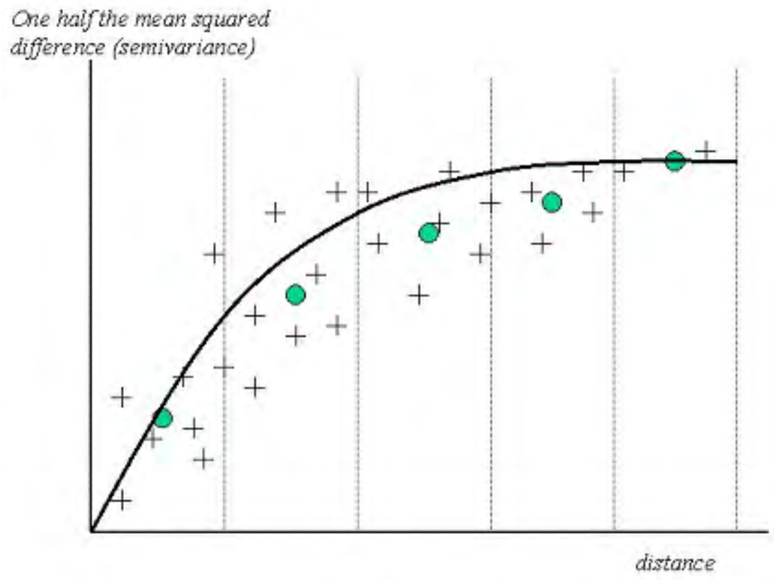


The difference in squared distance increases steeply to a certain point and then no more.



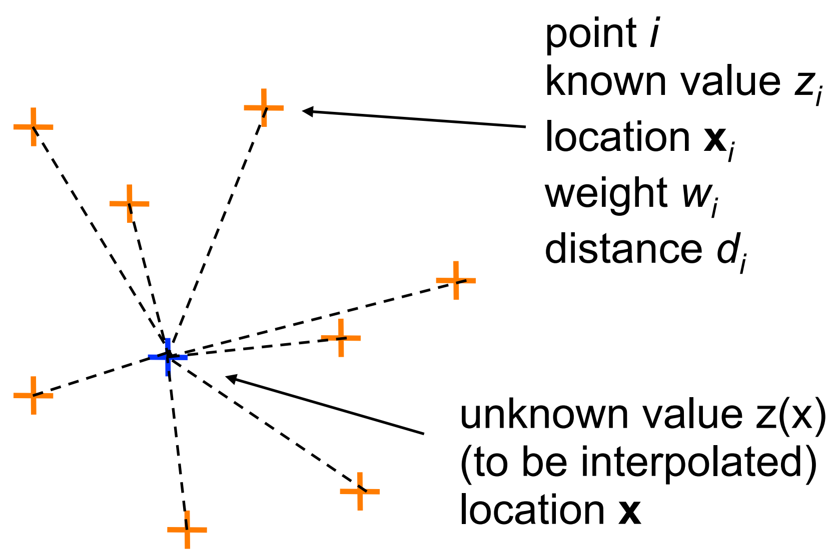
Nugget: the squared difference never falls to zero, not even at zero distance – this is the variation among repeated measurements at the same point.

Once we have the experimental semivariogram (the crosses in this graph), one of the standard mathematical functions is fitted to it (the thick black line in this picture).

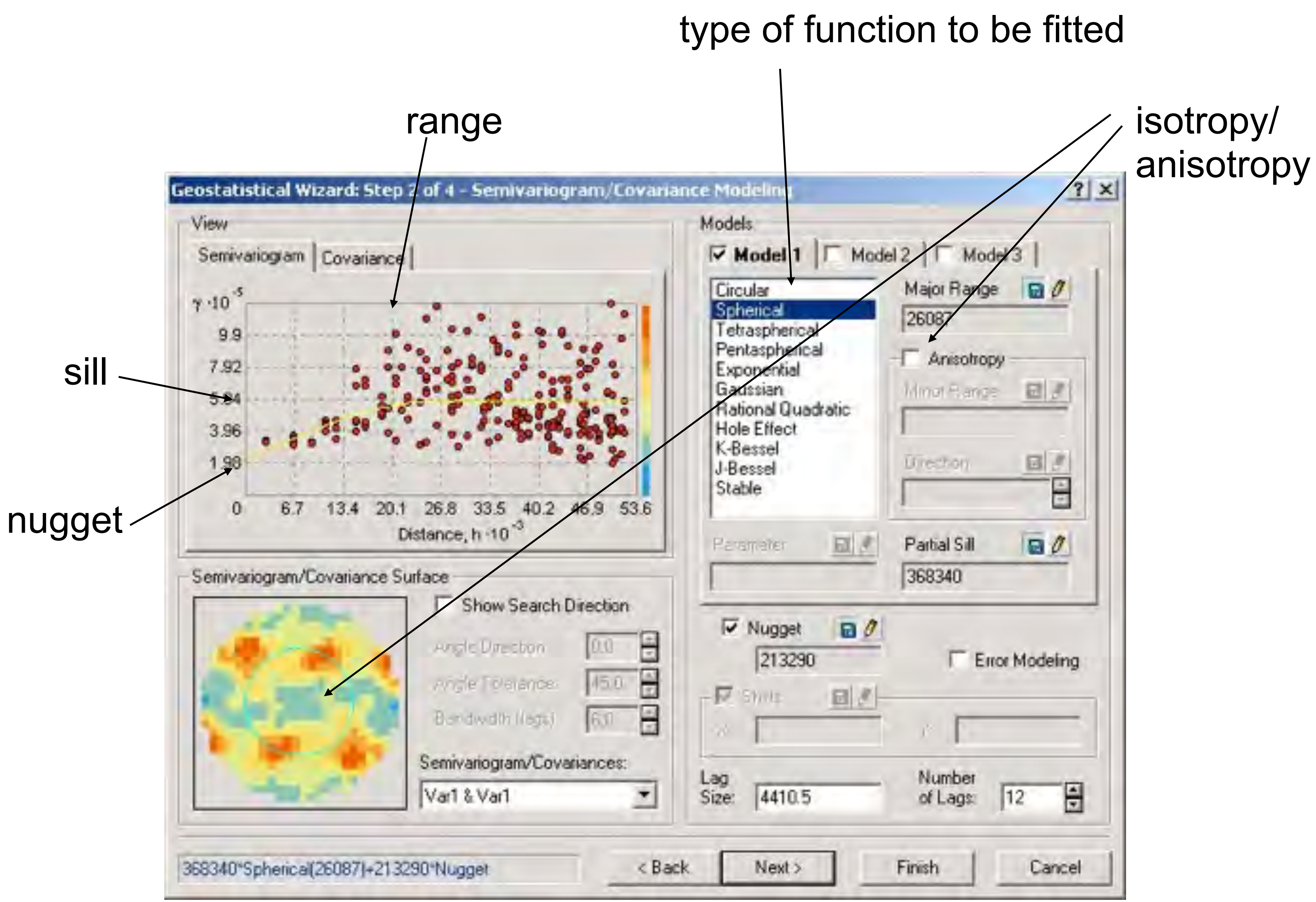


This function is used to calculate the optimal weights w_i for the interpolation, where the unknown value is calculated as a weighted combination of known values (same as with IDW):

$$z(\mathbf{x}) = \sum_i w_i z_i \Big/ \sum_i w_i$$

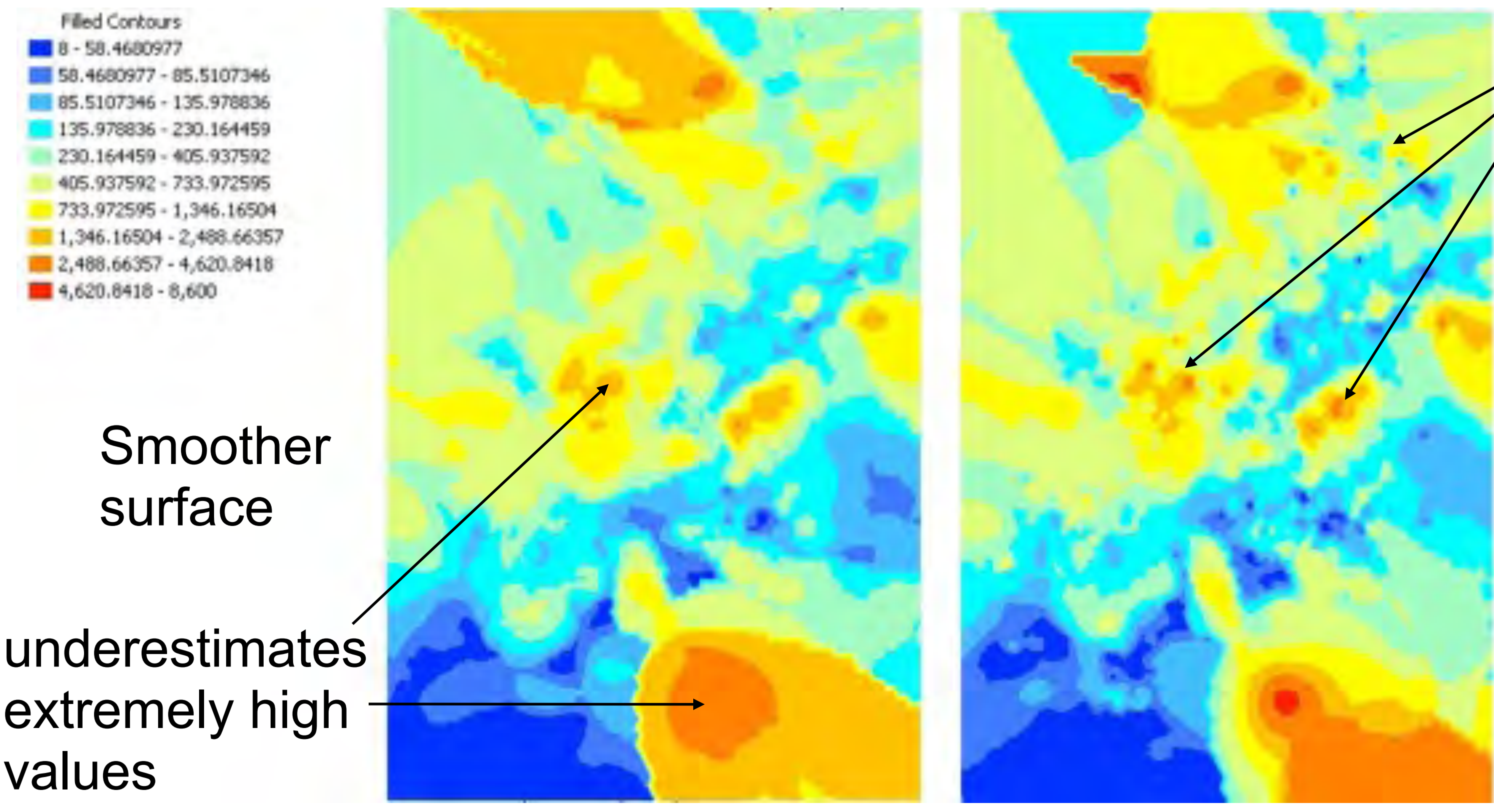


The interpolated surface replicates statistical properties of the semivariogram.


type of function to be fitted
range
isotropy/
anisotropy
sill
nugget
Geostatistical Wizard: Step 2 of 4 - Semivariogram/Covariance Modeling
View
Semivariogram
Covariance
Distance, h ·10^-3
Models
Model 1
Model 2
Model 3
Circular
Spherical
Tetraspherical
Pentaspherical
Exponential
Gaussian
Rational Quadratic
Hole Effect
K-Bessel
J-Bessel
Stable
Major Range
26087
Anisotropy
Minor Range
Direction
Parameter
Partial Sill
368340
Semivariogram/Covariance Surface
Show Search Direction
Angle Direction
Angle Tolerance
Bandwidth (lags)
Semivariogram/Covariances:
Var1 & Var1
Nugget
213290
Error Modeling
Shift
Lag Size:
4410.5
Number of Lags:
12
368340*Spherical(26087)+213290*Nugget
< Back
Next >
Finish
Cancel

Kriging vs. IDW

Steeper surface



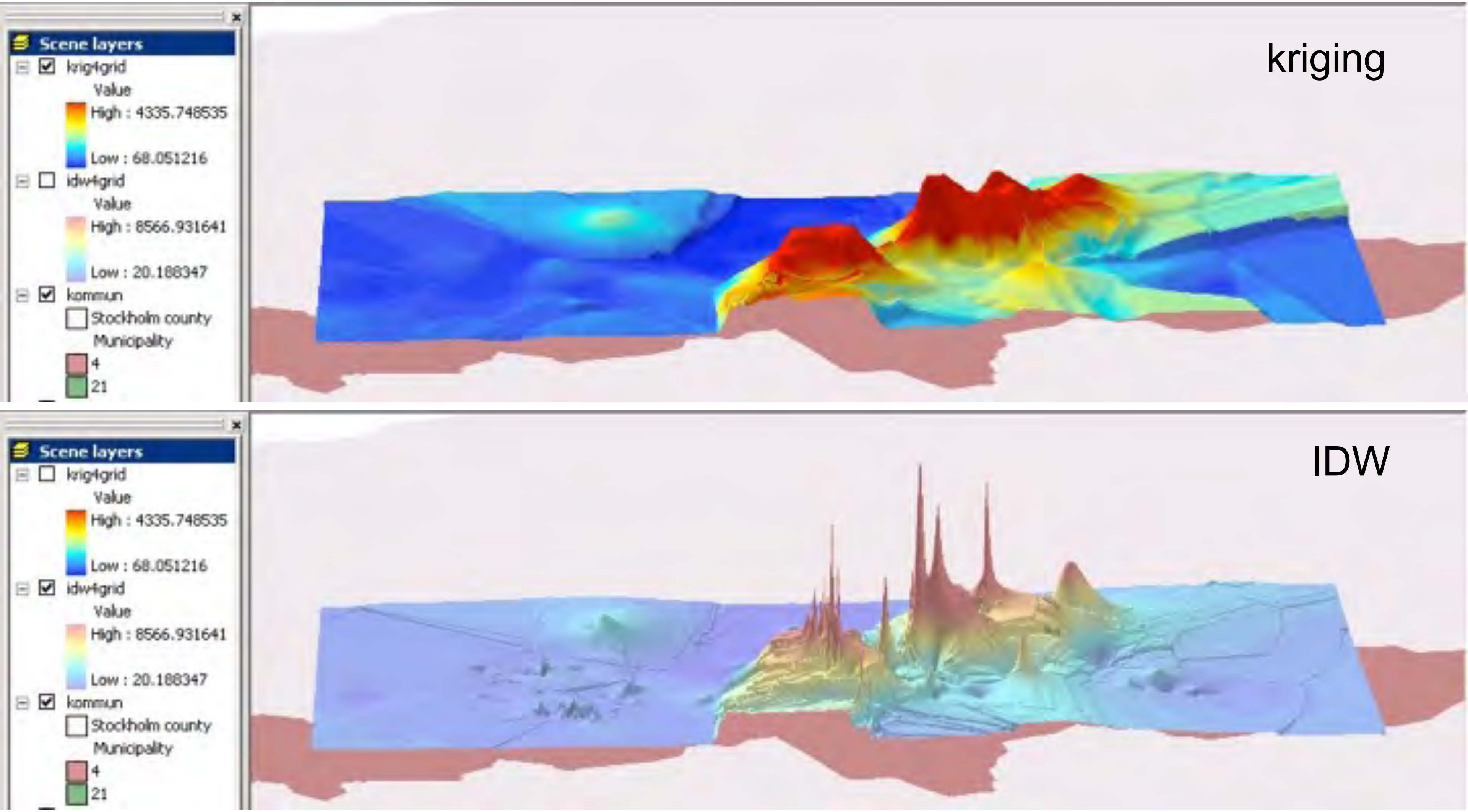
"u" effects around extreme values

Smoother surface

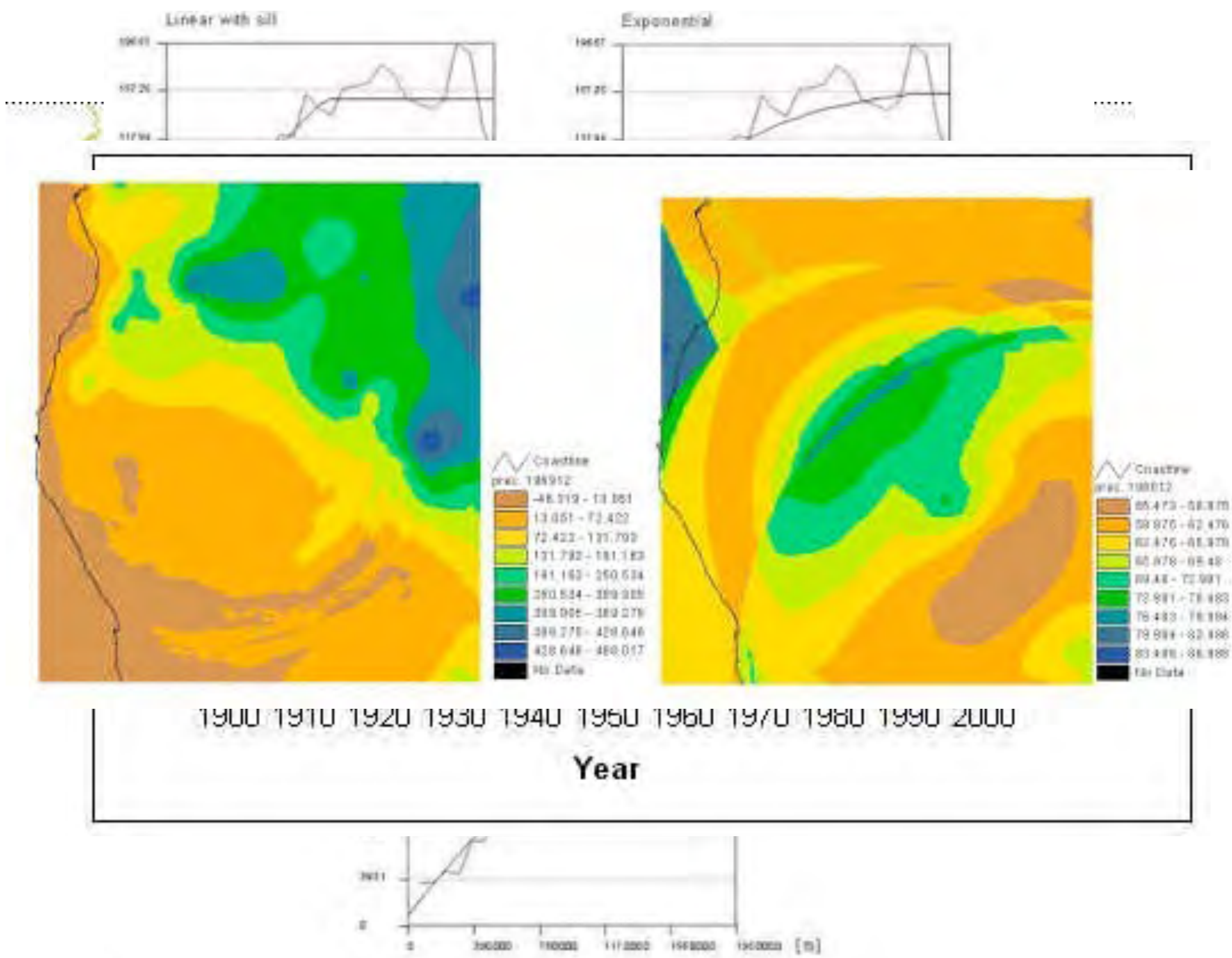
underestimates extremely high values

Comparison of kriging surface with the IDW surface of the same data using the same classification (quantile into 10 classes) and colour scheme for both surfaces.

Kriging vs. IDW

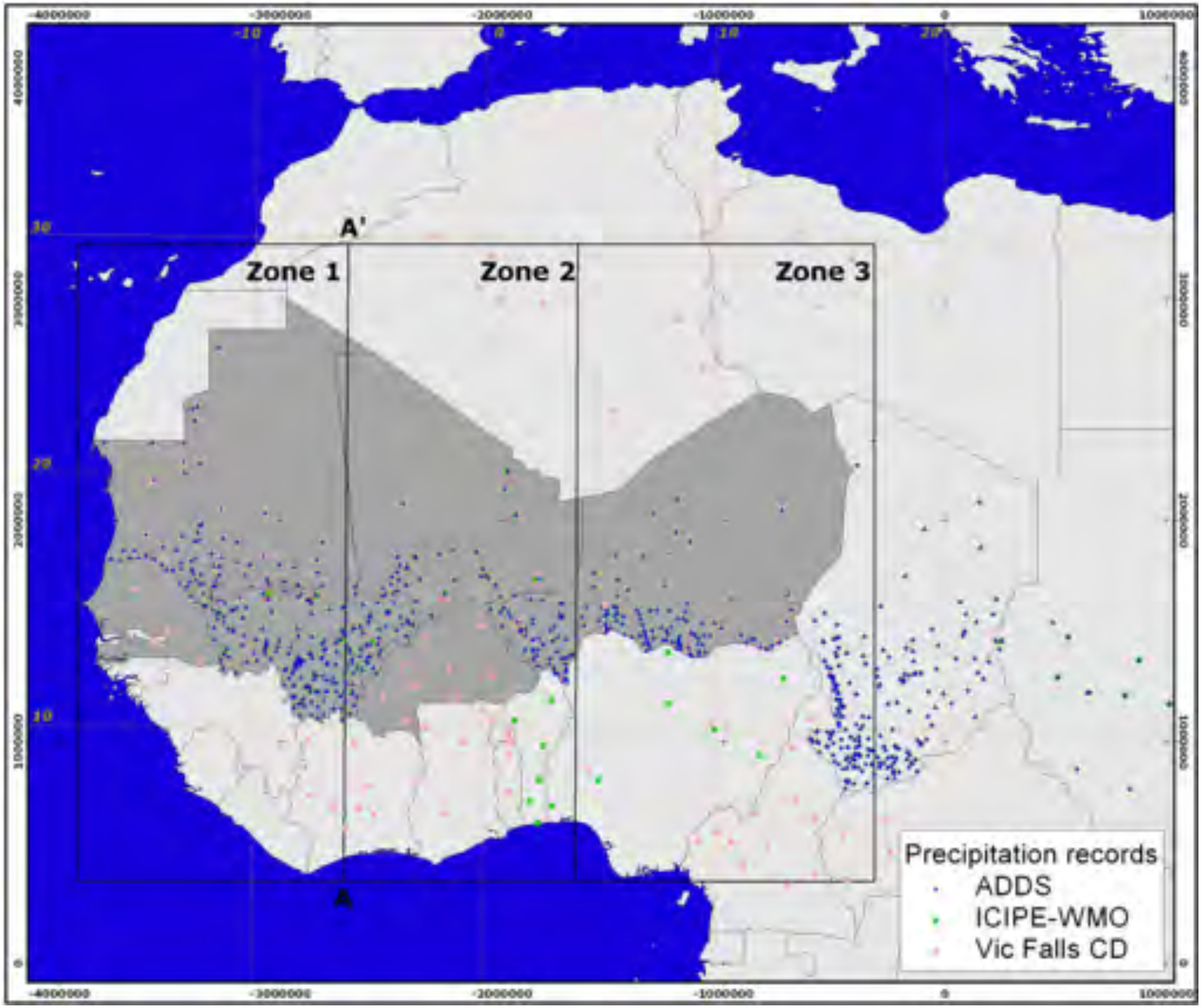


Geostatistisk Interpolering av nederbörden över Okavango



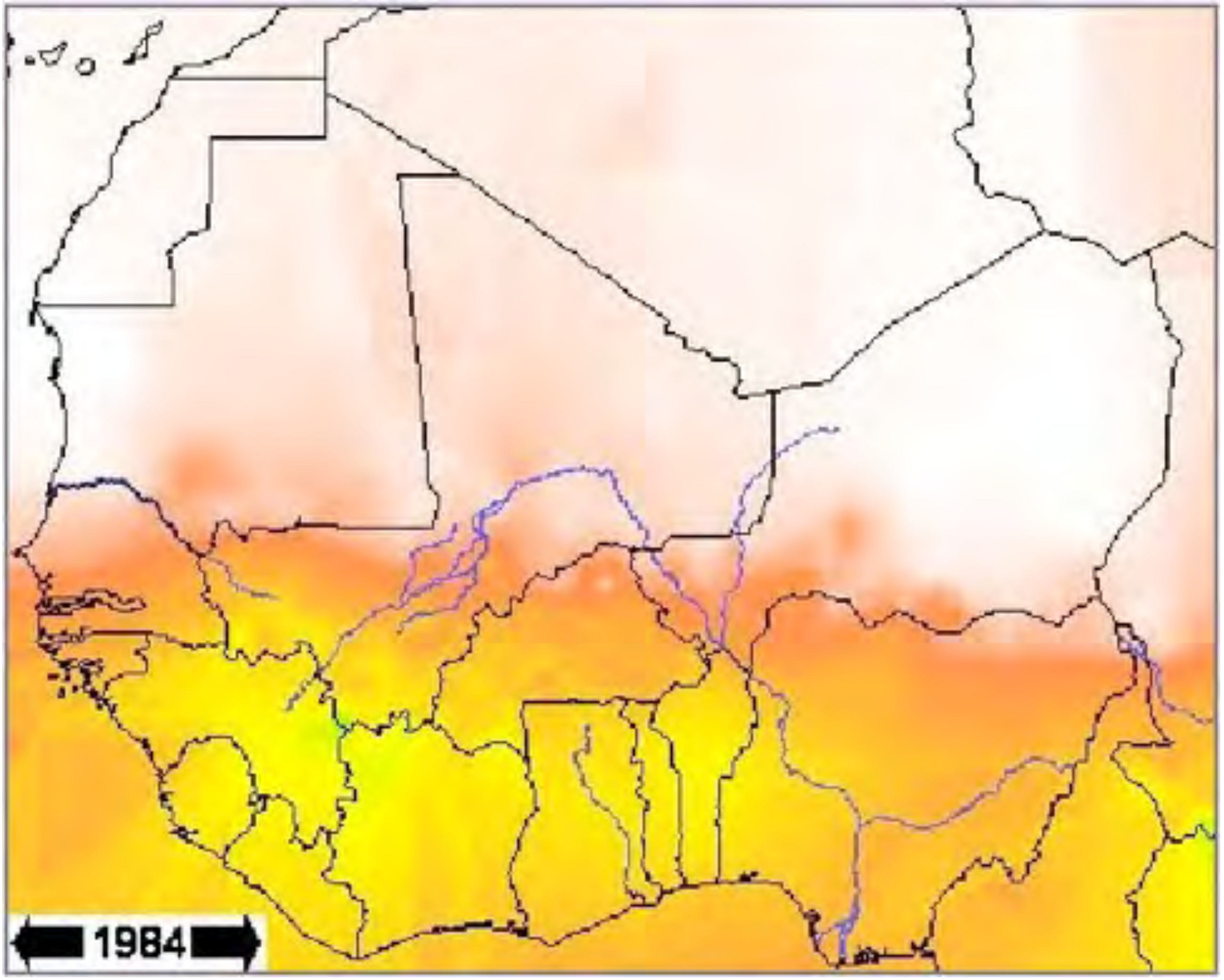
Geostatistisk Interpolering av nederbörden över Sahel

Sahel rainfall stations 1930-1996



Geostatistisk Interpolering av nederbörden över Sahel

Sahel rainfall average 1984



Multi-Criteria Evaluation - MCE

MCE is a method for decision support where a number of different criteria are combined to meet one or several objectives and help to make a decision.

Criterion:
A basis for a decision that can be measured and evaluated

Factor

enhances or detracts from the suitability under consideration

Particular soil types are better for growing wheat than other soil types.

Constraint

limits the alternatives under consideration

A new residential area can not be built inside a national park.

Decision rule – the procedure that combines criteria, often into a single composite index.

Examples

Classification

Classify areas according to how sensitive they are to landslides or erosion

Selection

Choose areas suitable for a particular purpose

Implementation of the decision rule = **Multi-Criteria Evaluation**

MCE in a raster GIS

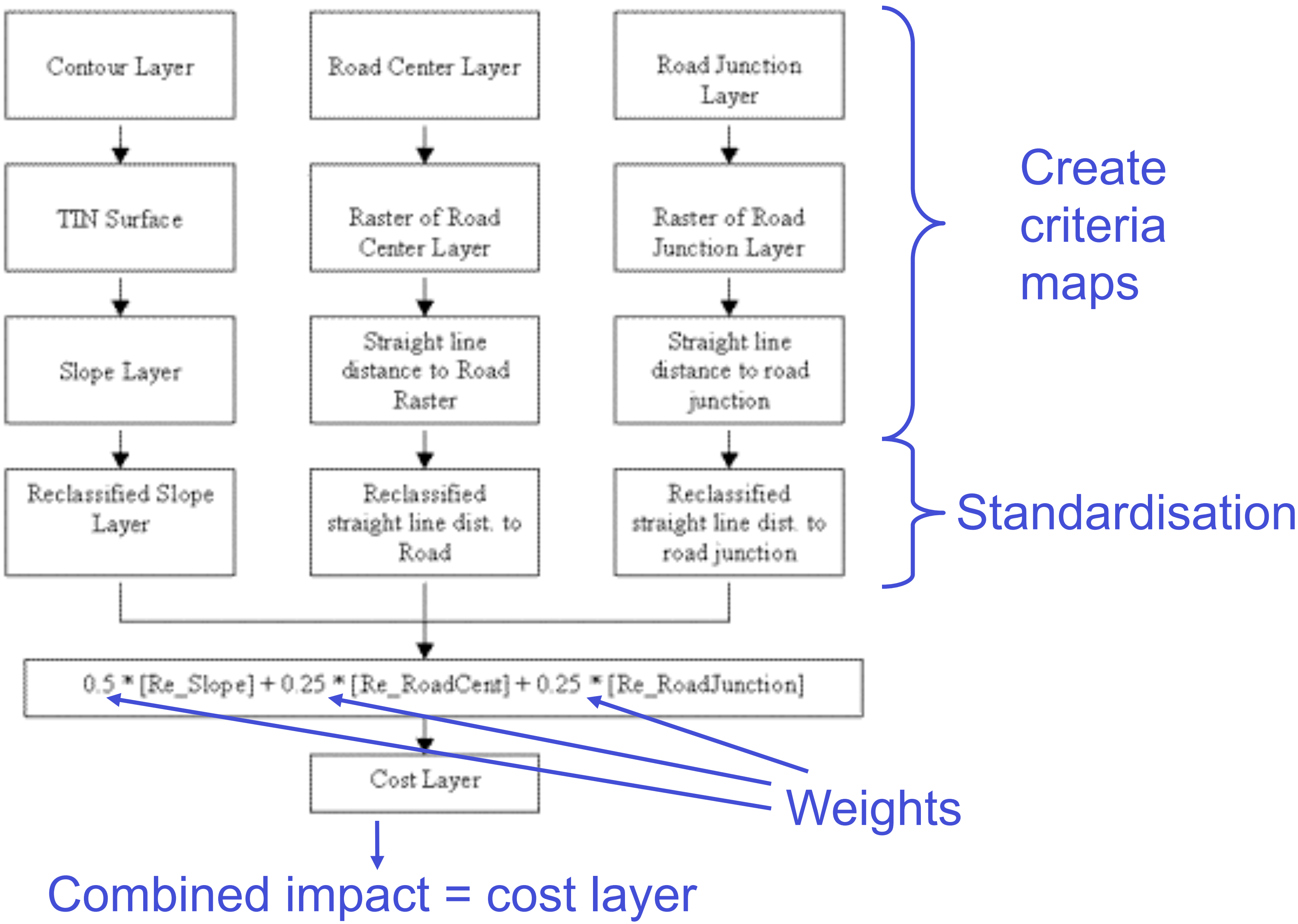
1. Create maps for each criterion.

1	1	0	0	0
1	1	1	0	0
1	1	1	1	0
1	1	1	1	1
1	1	1	1	1

2. Standardise the criteria maps → Same value range for all criteria

3. Assign weights to each criterion

4. Calculate the combined impact of all the criteria by combining all the standardised criteria maps with respective weigths


Contour Layer
Road Center Layer
Road Junction Layer
TIN Surface
Raster of Road Center Layer
Raster of Road Junction Layer
Slope Layer
Straight line distance to Road Raster
Straight line distance to road junction
Reclassified Slope Layer
Reclassified straight line dist. to Road
Reclassified straight line dist. to road junction
0.5 * [Re_Slope] + 0.25 * [Re_RoadCent] + 0.25 * [Re_RoadJunction]
Cost Layer
Create criteria maps
Standardisation
Weights
Combined impact = cost layer

1. Criterion maps

- Proximity maps
  - to roads
  - to town
  - to water
- Limitations
  - not too steep slope
  - not in a national park
- Preferences
  - particular soil types
  - particular landuse

2. Standardisation

Perform scaling so that all factor maps have **the same range**:

Example

Linear scaling:

$x_i = (R_i – R_{min})/(R_{max} – R_{min})*m$ → Puts the values in the [0,m] interval

The desirable feature has to get a high value.

Areas near to roads should get 1, areas far from roads get 0.

3. Assign weights

Many different methods for assigning the weights.

Example: pair-wise comparison of the factors

Each stakeholder produces a comparison matrix for the factors - W_i:

	waterfac	powerfac	roadfac	landfac	Slopefac
waterfac	1				
powerfac	1/5	1			
roadfac	1/3	7	1		
landfac	1/5	5	1/5	1	
slopefac	1/8	1/3	1/7	1/7	1

less important more important

1/9	1/7	1/5	1/3	1	3	5	7	9
extremely	very strongly	strongly	moderately	equally	moderately	strongly	very strongly	extremely

All comparison matrices are combined into one (matrix W):

W_1

W_2

…

W_m

	waterfac	powerfac	roadfac	landfac	Slopefac
waterfac	1				
powerfac	1/5	1			
roadfac	1/3	7	1		
landfac	1/5	5	1/5	1	
slopefac	1/8	1/3	1/7	1/7	1

→

	waterfac	powerfac	roadfac	landfac	Slopefac
waterfac	1				
powerfac	1/5	1			
roadfac	1/3	7	1		
landfac	1/5	5	1/5	1	
slopefac	1/8	1/3	1/7	1/7	1

Weights = eigenvalues of the matrix W

This is a very complicated method for assigning the weights.

But,

weight assignment is **a difficult issue**, as there are usually many stakeholders involved in the process, who usually disagree on how the factors should be combined.

4. Combine criteria

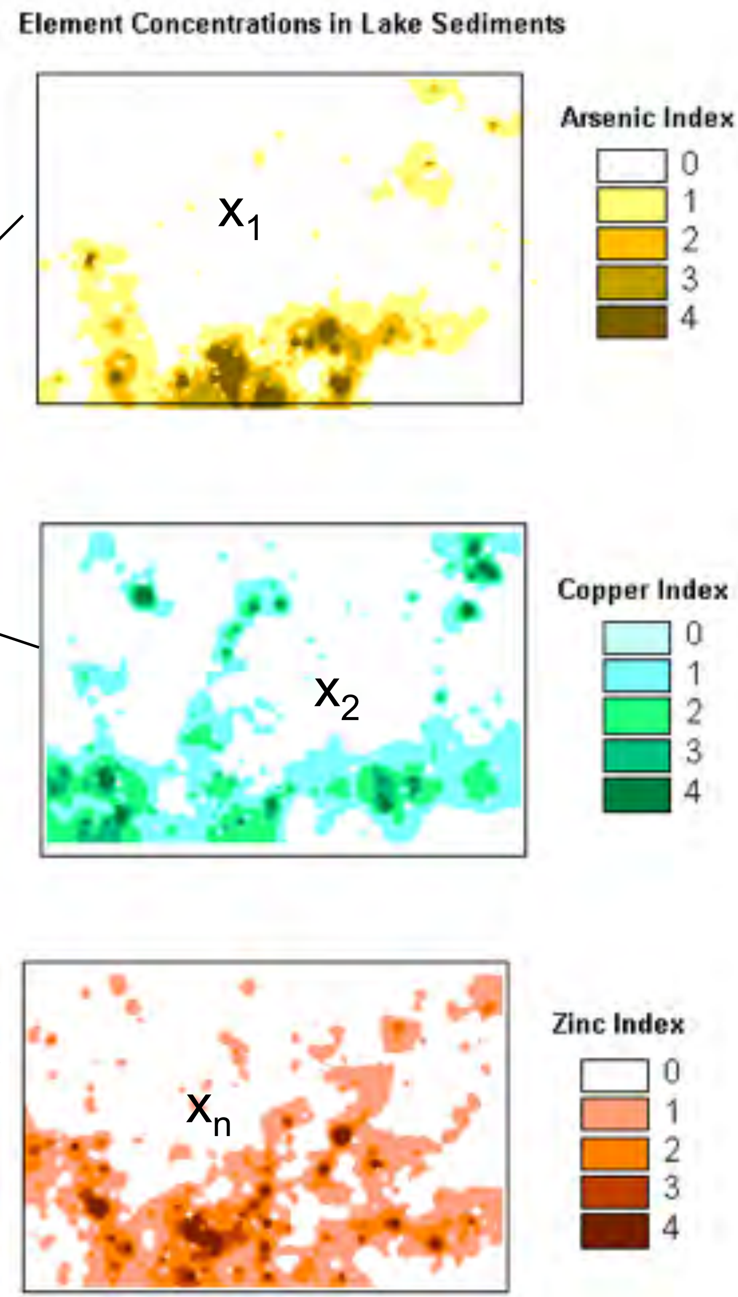
Combined impact of all the criteria:
- a weighted linear combination of standardised factors

$$I = w_1x_1 + w_2x_2 + \ldots + w_nx_n$$

w_1

w_2

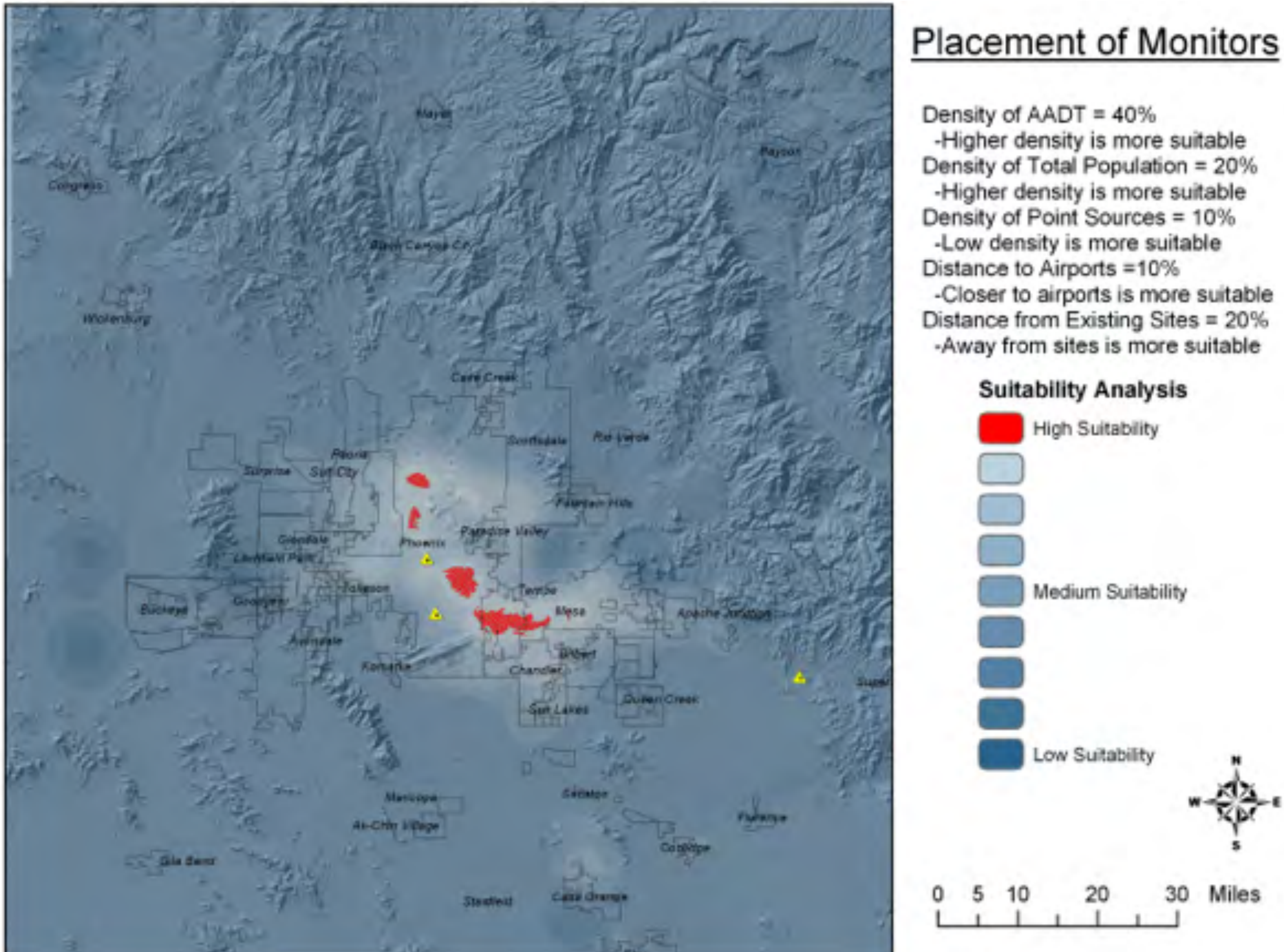
w_n



Calculated by map algebra

Result: **a suitability map**

Suitability for air quality monitors in the Phoenix region, Arizona



Evaluation of the suitability map

Select cells with highest suitability until a certain number is reached.

Cells with a suitability higher than a certain value are classified as suitable for a particular objective

Examples

Suitable for growing a particular kind of crop

Sensitive to erosion

Basis for decision-making

A MCE example

More than 50 meters from water

Areas should be close to water (though not closer than 50 meters)

Find areas suitable for residential development that **meet the following criteria**:

Areas should be close to roads

Areas should be close to services

Not on landuse classes residential, water, industrial, etc.

Multi-objective decisions

What if more than one objective needs to be fulfilled?

2 objectives

- complementary → Can be fulfilled At the same time
- conflicting → Contradict each other

Complementary objectives:
- find areas suitable for both objectives

Create a suitability map for each objective

Combine these in a new MCE procedure

Conflicting objectives:
2 possible solutions

Prioritised solution: put the most important objective first

Conflict resolution: find a compromise between competing objectives.

Is MCE an optimal solution for the decision-making?

How do we choose which criteria are relevant?

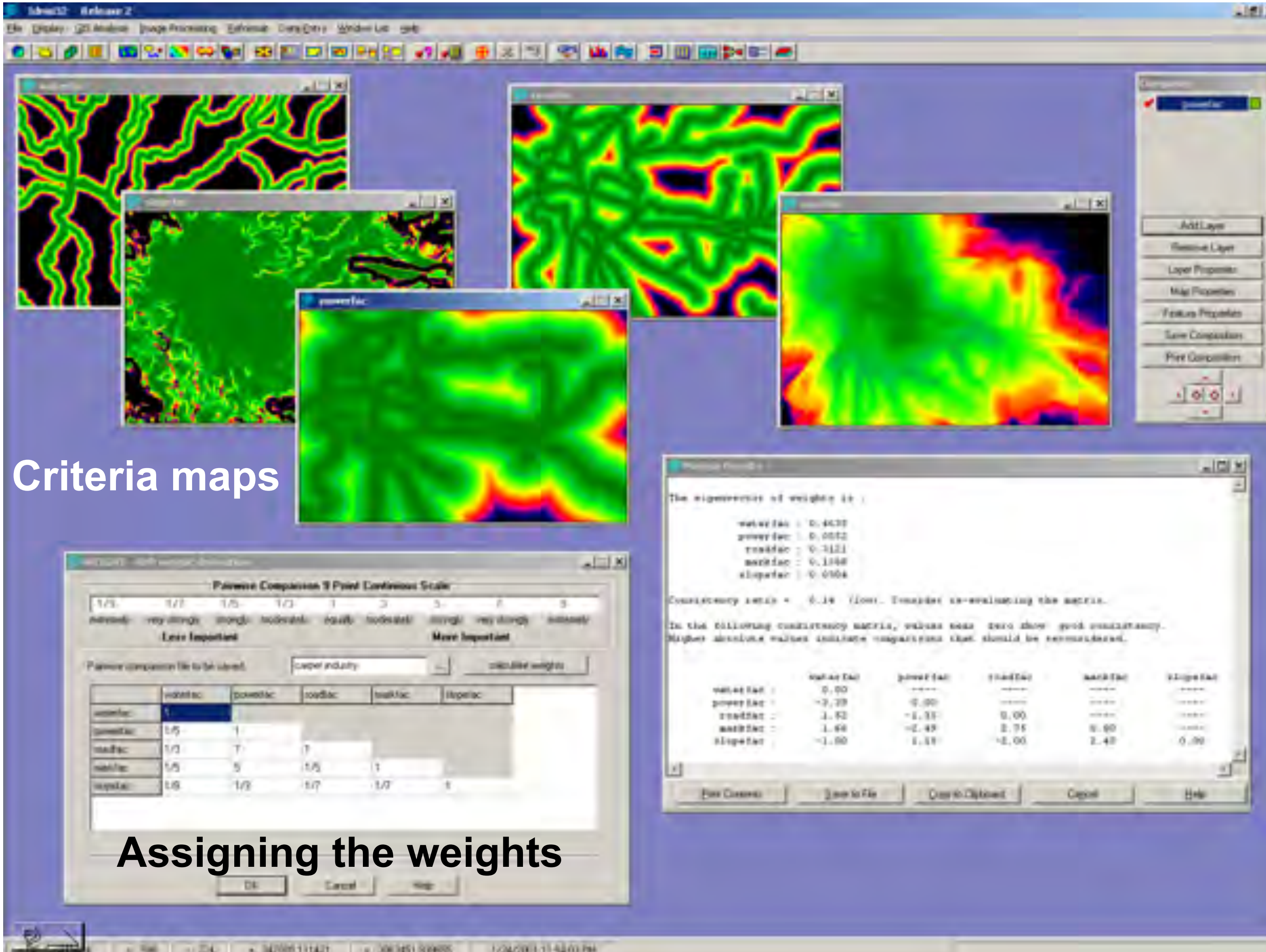
How do we assign the weights?

Geographical data sets often have a high degree of uncertainty.

This uncertainty propagates through the procedure.

The decision-makers need to be aware of this.

MCE in Idrisi



MCE för Cypern

MOLA - Multi Objective Land Allocation