Aalto University
School of Engineering

Geoinformation in Environmental Modelling

Spatial analysis: from points to surfaces
Network analysis

ENY-C2005
Jussi Nikander
6.2.2019
Slides by Paula Ahonen-Rainio and Jaakko Madetoja

Topics today

- From a set of points to a surface
  - Density surface
    - Kernel density
  - Spatial interpolation
    - Thiessen polygons (cf. Voronoi diagram)
    - IDW
    - TIN
    - Kriging – based on spatial statistics
- Network analysis
  - Typical applications
  - Some characteristics of graphs
  - Optimization as an approach to (network) analysis

Examples of potential exam questions relating to this lecture (in English)

- Explain the method of Kernel density estimation. Kernel density estimation and spatial interpolation both create a surface from a point set. However, they are fundamentally different approaches. What is this difference? Give an examples, of what kind of datasets do they fit.

- Explain how to form Thiessen polygons (also called as Voronoi diagram) for a set of points. Draw a set of seven points and the Thiessen polygons for the set.

- What is spatial interpolation? Explain the use of IDW method in spatial interpolation. What kind of limitations relate to this method? [alternatively: interpolation in TIN model]

Same questions in Finnish

- Selitä Kernel-tiheysestimoinnin menetelmä. Kernel-tiheysestimointi ja spatiaalinen interpolointi molemmat tuottavat pinnan pistejoukon perusteella. Ne ovat kuitenkin oleellisesti erilaiset lähestymistavat. Mikä on tämä ero? Anna esimerkit, minkälaiselle datalle nämä lähestymistavat sopivat.

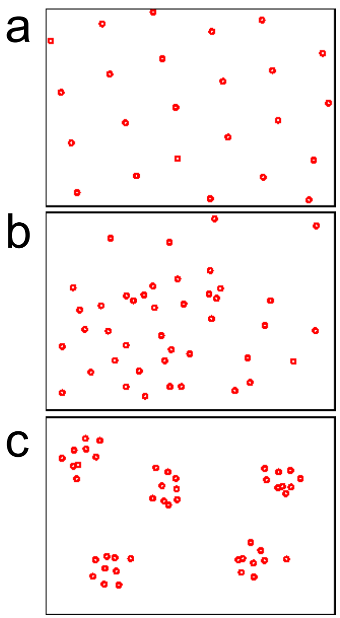
- Selitä kuinka Thiessen-polygonit (eli Voronoi-diagrammi) muodostetaan pistejoukolle. Piirrä seitsemän pisteen joukko ja sille Thiessen-polygonit.

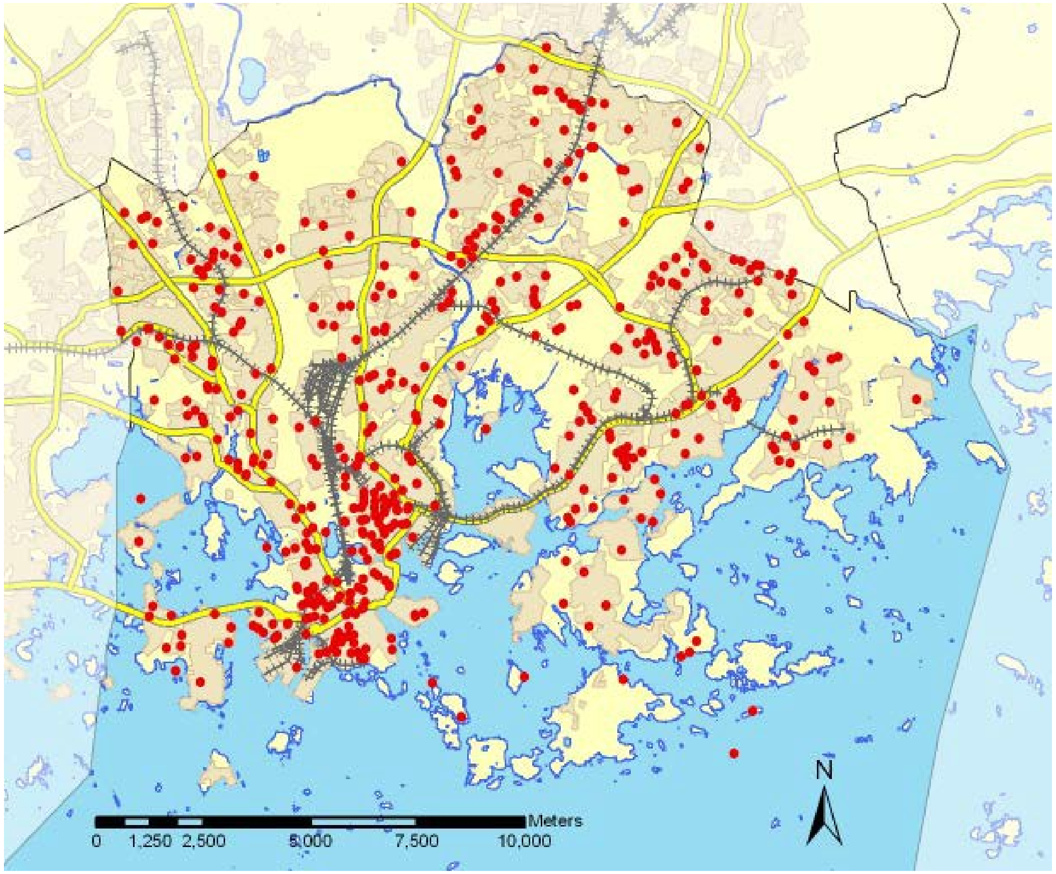
- Mitä on spatiaalinen interpolointi? Selitä miten spatiaalinen interpolointi tehdään IDW-menetelmällä. Mitä rajoituksia menetelmällä on? [tai: interpolointi TIN-mallissa]

From a set of points to a surface

- Density surface
- Interpolating a field on sample points

Case: Incidents of domestic fire in Helsinki

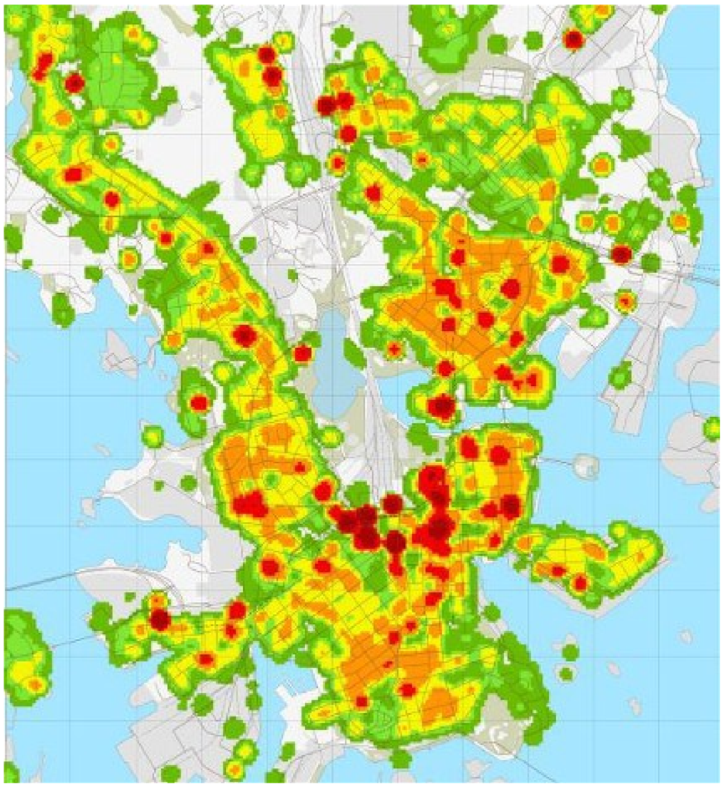


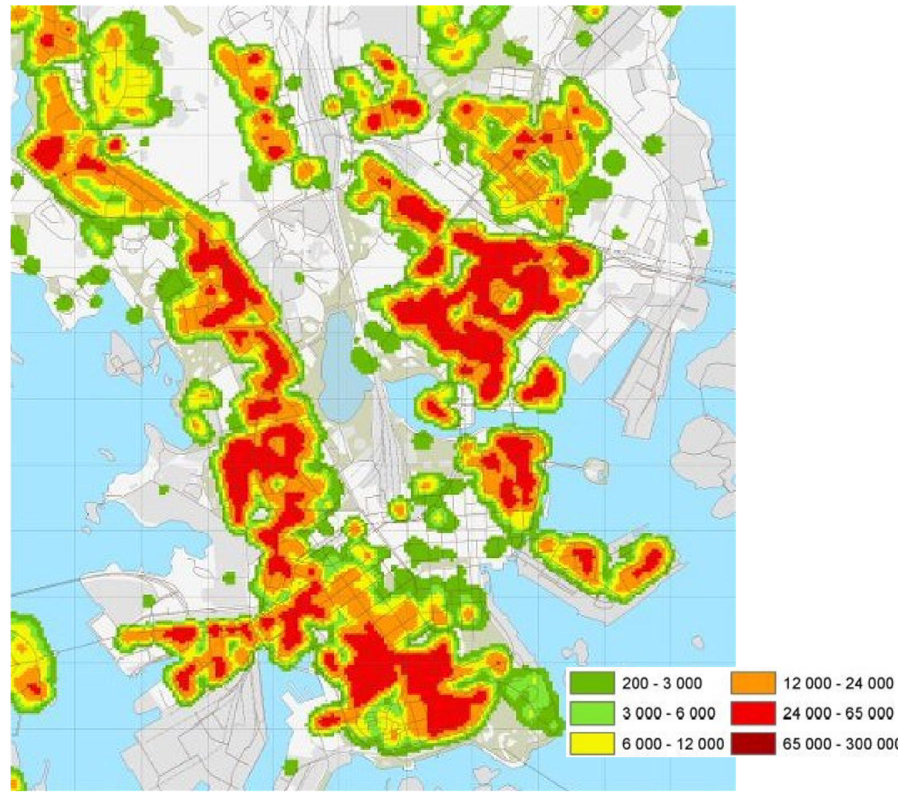


The distribution of the events is not regular (a) nor random (b)

Spatenkova 2009

Case: Incidents of domestic fire in Helsinki





Spatenkova 2009

Kernel density estimation

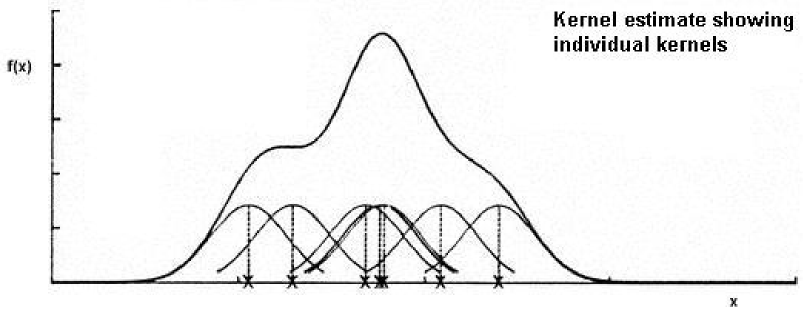
- Transformation from point objects to a density surface to *visualize* a point pattern to detect *hot spots* where the local density is estimated to be high
- to allow comparison of point data with a surface variable
- Density at any location in the study area
  - estimated by counting the number of objects (or events) in a region (=kernel) centered at the location ***p*** where the estimate is made
- Simple kernel: circle centered at the location ***p***
- More sophisticated: a kernel function weighting nearby events
- Complexity: density is scale-dependent! (cf. kernel bandwidth)

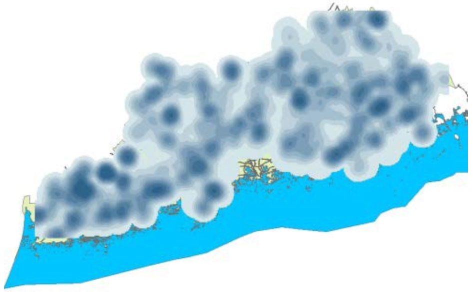
Read more: O'Sullivan & Unwin (2010) pp. 68-71

Longley et al (2015) pp. 310-313

Kernel density surface

- Kernel function (e.g., bell shape) for the local density
  - Each point is replaced by a kernel
  - Density surface is the sum of these





Krisp 2006

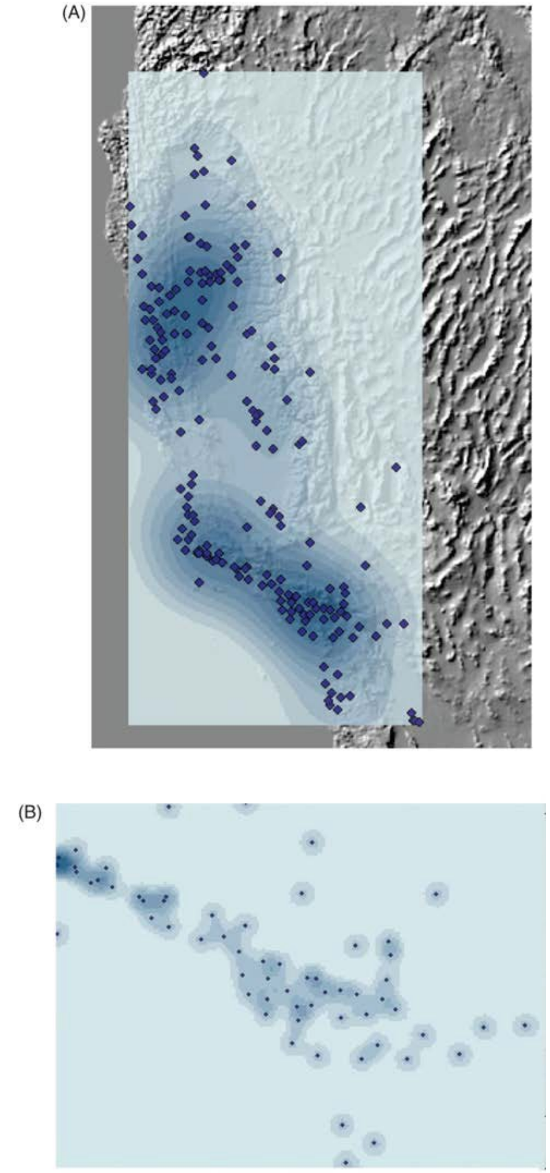
Kernel density surface

- Kernel *bandwidth* affects the resulting density surface
  - Large bandwidth results to smooth variation, densities get close to the global average across the study area
  - Small bandwidth results to surface pattern focusing strongly to individual events and zero densities between remote events

Example: Density estimation using two different distance parameters in the respective kernel functions.

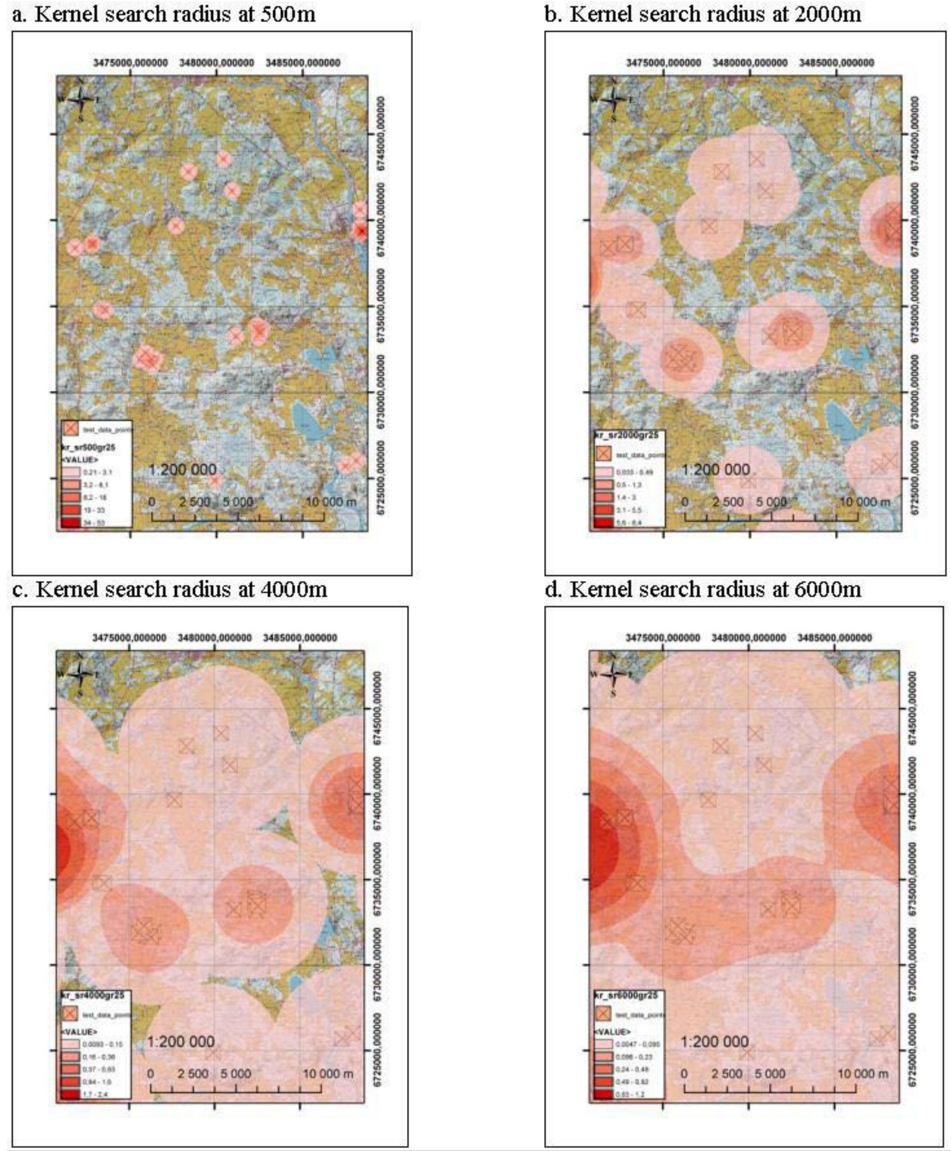
(A) the surface shows the density of ozone-monitoring stations in California, using a kernel radius of 150 km

(B) zoomed to an area of Southern California, a kernel radius of 16 km is too small for this dataset, as it leaves each kernel isolated from its neighbors



Easy-to-use method but understanding the approach is necessary when interpreting the patterns!

Notice the difference in visualization: the original points can be left visible or not (compare to previous slide)



Density surface and spatial interpolation

- Density surface estimates the variation of the density of discrete events (events = point objects)
  - Population of known cities
- Spatial interpolation predicts the values of a spatially continuous variable at unsampled locations from the measurements (known values) made at control points (sample points) in the same area
  - Temperature in measurement points

Spatial interpolation

- Values of a field measured at a number of sample points

  ⇒ need to estimate the continuous field

  i.e. values at points where the field was not measured (points to be estimated)

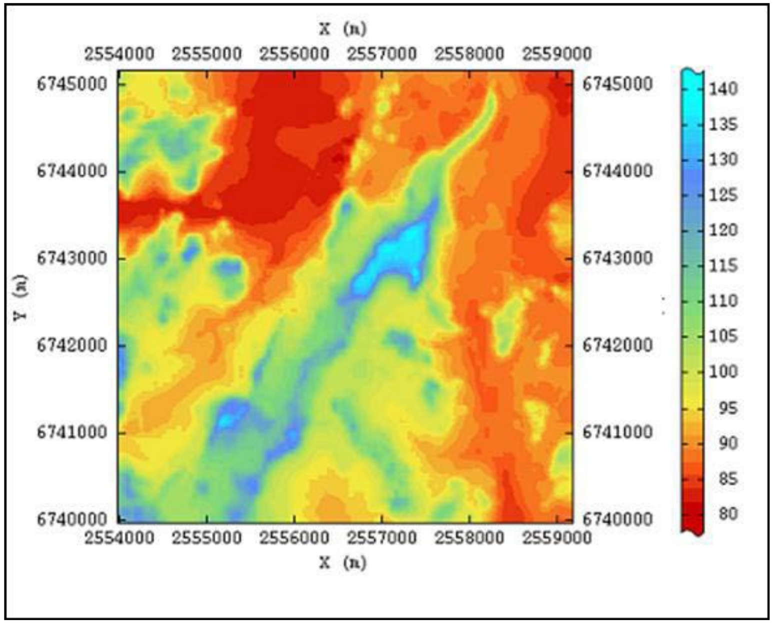
- Approaches to the interpolation :
  - global or local
  - deterministic or stochastic
  - smooth or abrupt changes

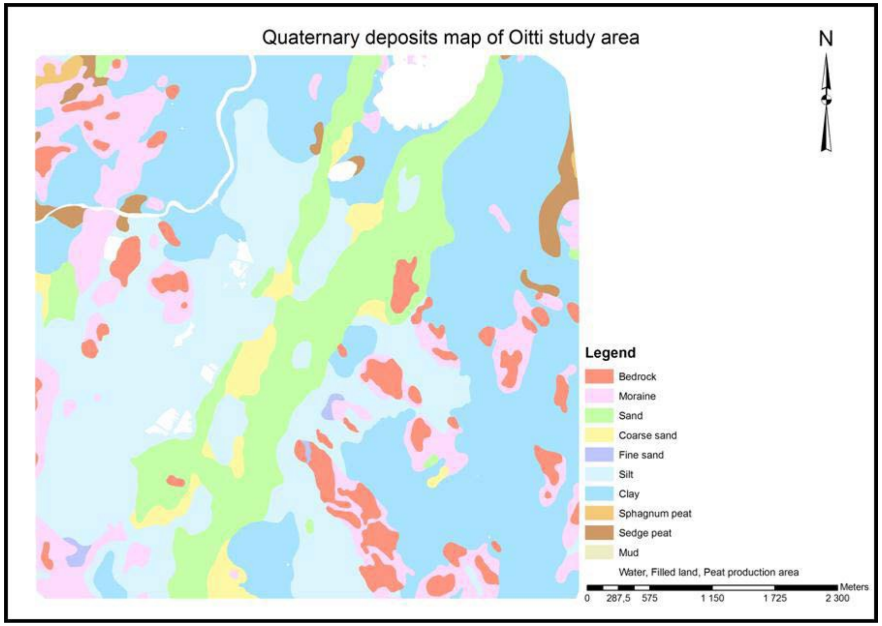
- Which method?
  - Characteristics of the represented phenomenon
  - Consider the uncertainty of the interpolated data!

Spatial autocorrelation and interpolation

- Interpolation is not possible without positive spatial autocorrelation
- Also take into account:
  - Physical obstacles
    - special features, trends
    - anisotropy
  - Continuos or categorized values

Continuous or categorical data

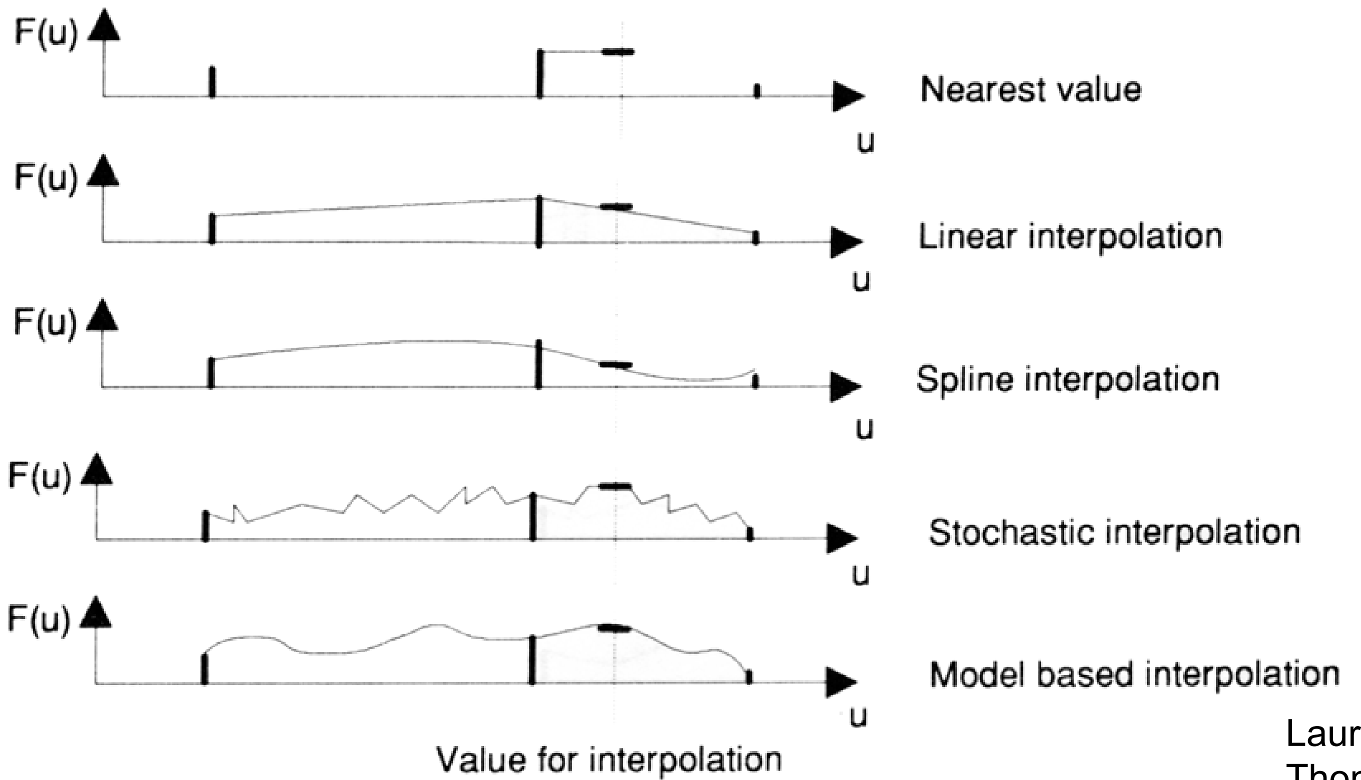




Sunila 2009

Approches in interpolation

The idea of different methods represented in 1-d space:



Laurini & Thompson 1992

Methods for spatial interpolation

- Thiessen polygons: nearest value method
- TIN model: linear interpolation on the planes of triangles
- Local spatial average: fixed distance or fixed number of points; problems with each
- IDW: Inverse distance weighted spatial average
- Kriging: based on study of spatial autocorrelation

Read in O'Sullivan & Unwin (2010) Ch. 9.3 pp. 250-261

Thiessen polygons

- Thiessen polygons: a polygon network
  = proximity polygons (naapuruuspolygonit)
- Constructed by the perpendicular bisectors of lines joining pairs of points
- Formally:

$\{p_1 \ldots p_n\}$ is a finite set of points S on a 2d plane P
$d(x, y) \geq 0$ is a distance function
Thiessen polygon T(p_i) for $p_i \in S$ is defined as the set of locations $p \in P$ where
$d(p, p_i) \leq d(p, p_j) \forall j, i \neq j$

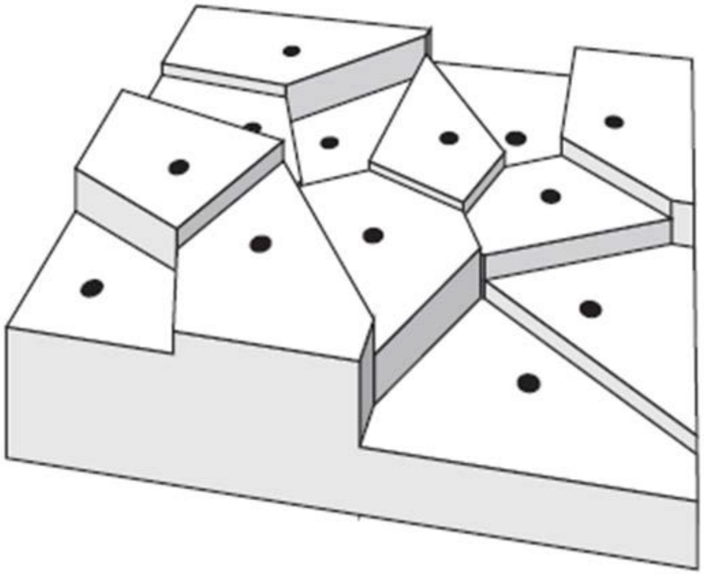
O'Sullivan & Unwin (2010) Fig. 2.4, p. 51

Thiessen polygons

Every unsampled point gets the value of its nearest control point

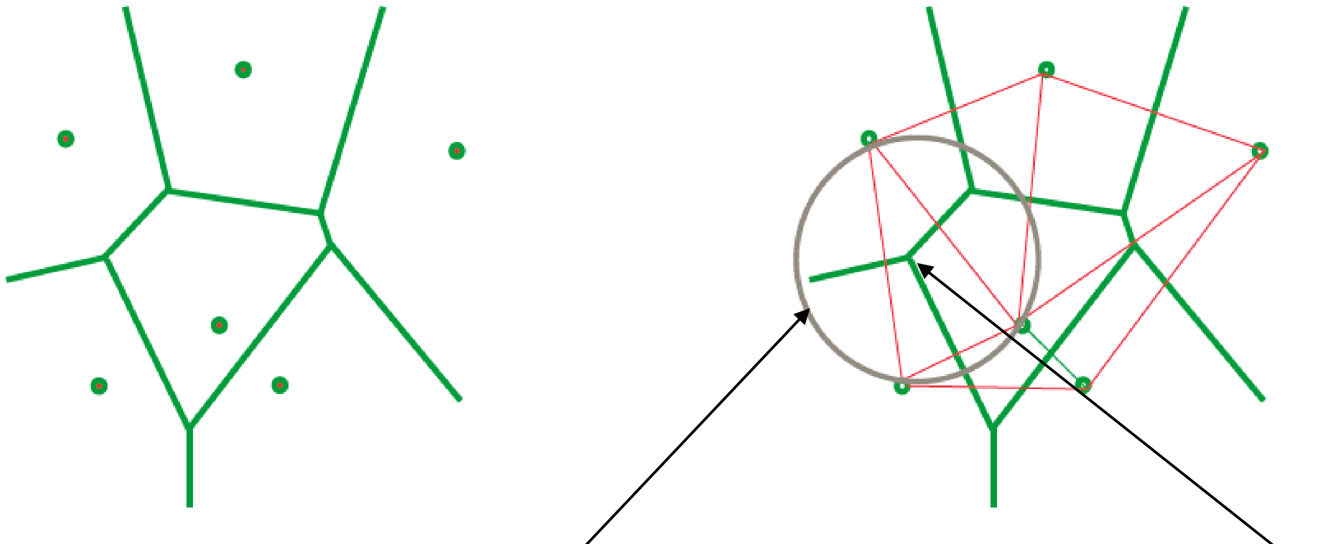
- Proximity polygon of a [control] point: that region of the space which is closer to this point than any other
- Abrupt changes in adjacent polygons
  - May suit well for some phenomena, and not at all for others
  - The approach is quite obvious when visualized; no wrong expectations of the accuracy
- Useful for nominal data
  - But may not be considered *interpolation* in these cases (rather a transformation)

O'sullivan & Unwin (2010) Fig. 9.6 (p. 254)



Notice the connection between Thiessen polygons and triangulation

Thiessen polygons (Voronoi diagram) Delaunay triangulation

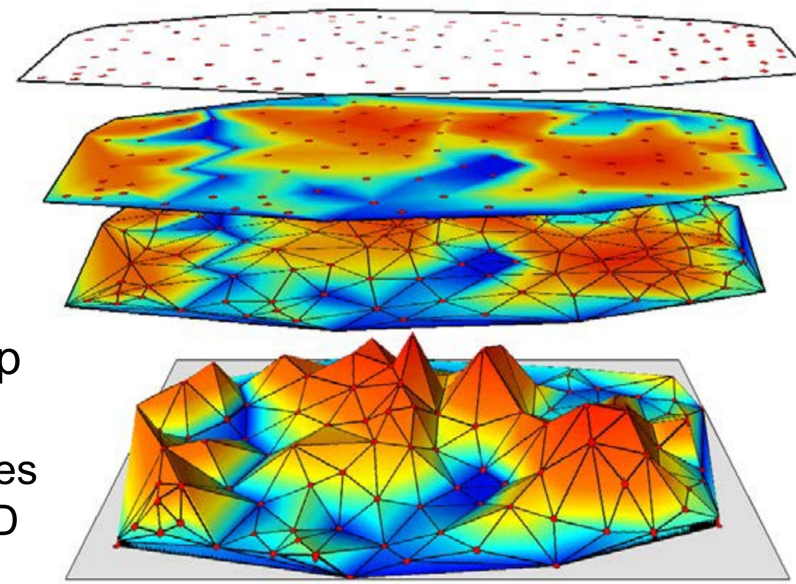


a circle drawn through the three nodes of a triangle will contain no other node

The circle's center point is where three Thiessen polygons meet

TIN represents the surface *per se*

- Sample points are connected to form triangles
- The field inside each triangle is represented as a **plane**
  - The value of the variable (i.e. the field) at any location p inside the triangle can be calculated from the values at the vertices (kärkipisteet) and the distances from the location p to the vertices
    - Notice, not only between two vertices along the edge (as in the case of 1D space) but anywhere on the plane (that is 2D)
  - TIN as such is an interpolation method



www.geosolutions.com

Local spatial average

Estimate = local spatial mean of the values of certain control points

- Use only the sample points within a fixed distance from the point to be estimated
  - There may be locations that are not within the chosen distance from any sample point and therefore remain **without value**
- Use a fixed number of nearest-neighbour sample points
  - Now all locations have nearest neighbours, but the **distances** to them may vary heavily (cf. extent of spatial autocorrelation)
- In both cases, all the sample points in question may locate in the **same direction** from the point to be estimated

Therefore, not recommended

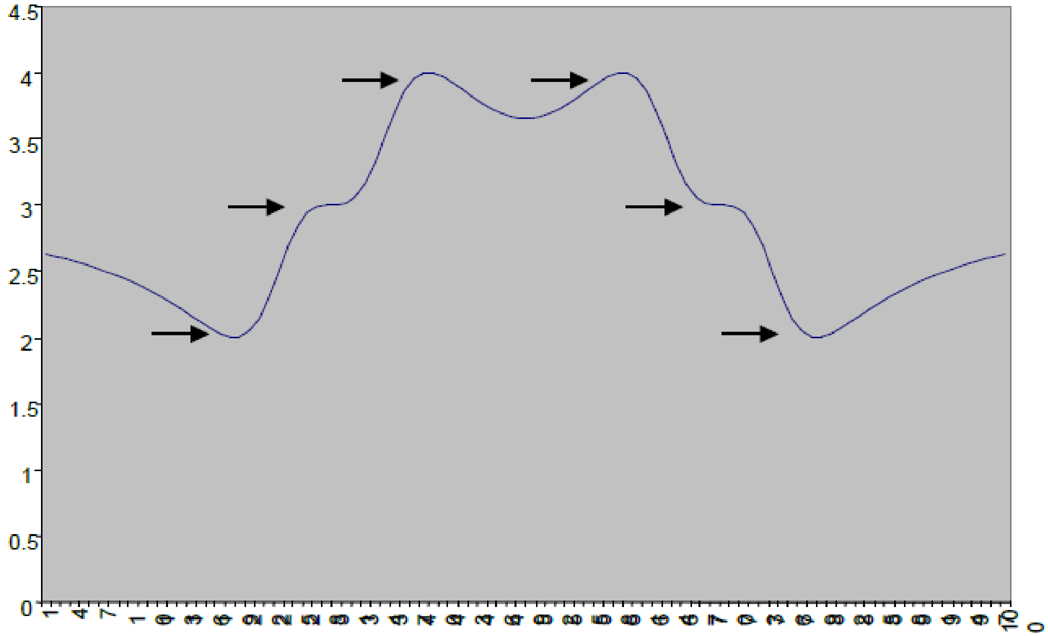
Inverse distance weighted spatial average = IDW

- Nearer locations are considered more prominent

  -> the values of sample points around the point to be interpolated are weighted according to the inverse distance

  - for example, $w \sim 1/d^2$ (decreases the effect of more distant points)
  - weights at any interpolated point shall sum to 1
- Notice: Cannot predict values lower than the minimum or higher than the maximum among the sampled values
  - a characteristic of any averaging technique (undulation problem)

Example of IDW: undulation problem



- If the extreme values (the peaks and valleys) are not included in the sampled data, they cannot be produced by interpolation either
- The sampled values marked by the arrows include the minimum and maximum values, and the interpolated values will be between these
- If there is a peak in the middle of figure, it should have been measured as a sample point

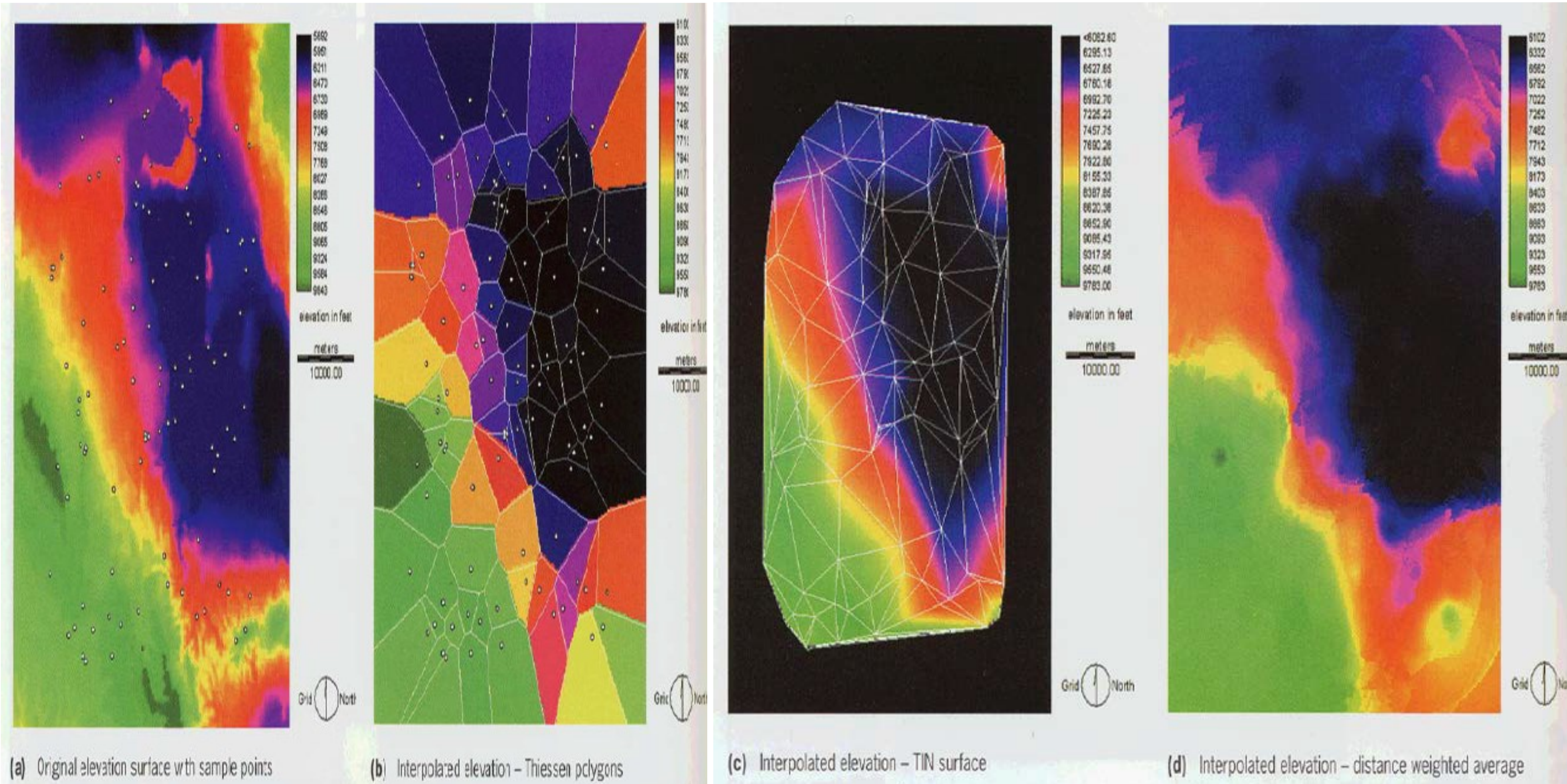
Sample points

The sample points may be

- Surface random
  - Locations are chosen without reference to the shape of the surface being sampled
  - They may, or may not, capture significant features of the surface
- Surface specific
  - Located at places that are important in defining the surface
    - a major change in the shape of the surface, such as peaks, pits, passes, saddle points, along streams, ridges, and other breaks of slope
- Grid sampling

Examples of interpolation results

Heywood et al 2003



(a) Original elevation surface with sample points

(b) Interpolated elevation – Thiessen polygons

(c) Interpolated elevation – TIN surface

(d) Interpolated elevation – distance weighted average

Original surface | Thiessen polygons | TIN surface | IDW

Interpolation of values in a grid

- Values for unknown pixels on the basis of known pixels
- Problem: which pixels are used in interpolation ?
  - Neighbours (4 or 8) – no statistical significance
  - More weight can be given on those pixels that are close to the pixel with unknown value (on the basis of spatial autocorrelation)
- Cf. **focal functions** in map algebra

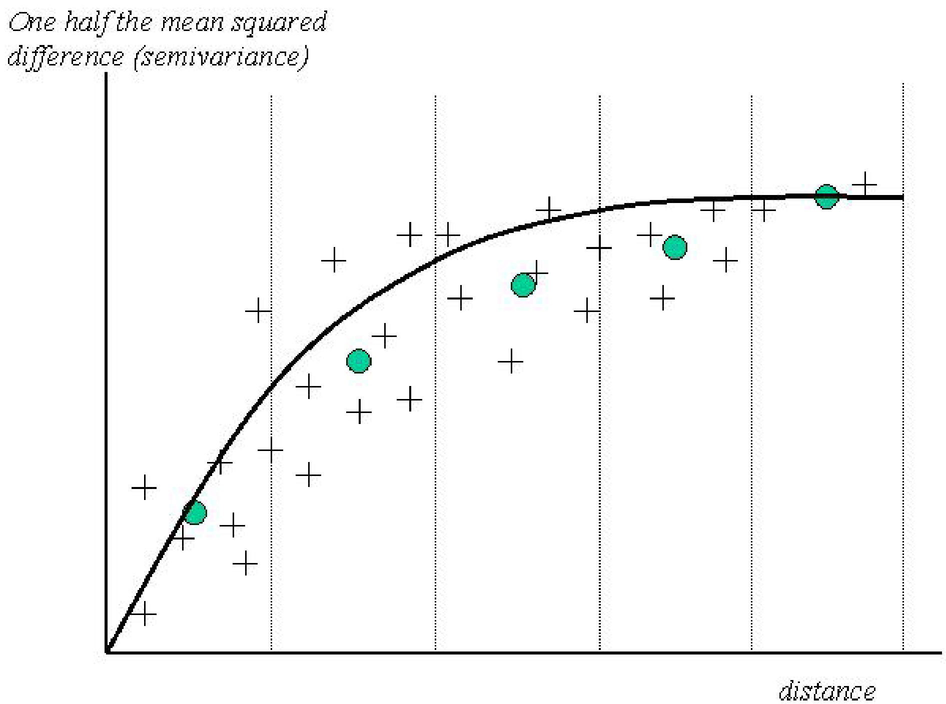
Kriging

- Technique firmly grounded in geostatistical theory;
- The interpolated surface replicates statistical properties of the semivariogram
- *Semivariogram* reflects Tobler's Law:

  " Everything is related to everything else, but near things are more related than distant things"

  => differences within a small neighborhood are likely to be small, differences rise with distance
- Semivariograms can be defined independently for different directions (e.g. each 90º sector)

  => more accurate interpolation
- Kriging needs consideration from the analyst because of its many options



An *isotropic** 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.

*) Does not take into account the direction of distance

Longley et al. (2015):
Section 13.3.6.3
(pp.315-317)
Not part of reading for the exam

Network analysis

- Typical applications
- Some characteristics of graphs
- Optimization as an approach

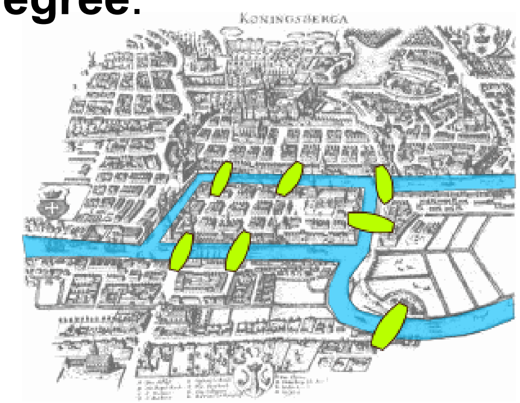
Network analysis applications

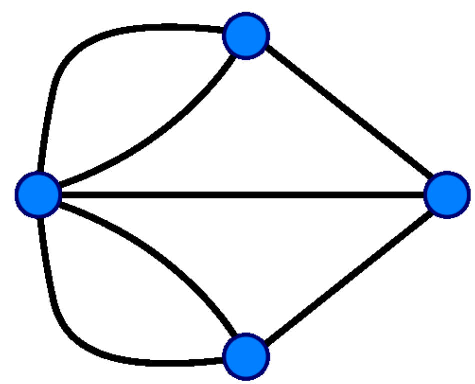
Many typical applications of geoinformatics are based on analysis of network data:

- Route optimization and navigation
  - Shortest route from A to B, or fastest, safest, cheapest… route (weighted links)
  - emergency medical service (ambulance), police, fire service
  - transport of dangerous goods, transport of wood
- Delivery and collection
  - mail delivery, pizza taxi, collection of milk from farms, waste management: collection and recycling (many to many)
- Intelligent traffic
  - E.g. traffic forecasts
- Management of infrastructure
  - Planning, construction, maintenance
  - water supply, sewers, electrical grid, other energy networks (utility management)
- Location allocation
  - Optimal location for a service or business

Case: seven bridges of Köningsberg

- The city of Königsberg in Prussia (now Kaliningrad) included two large islands which were connected to each other, and to the two mainland portions of the city, by seven bridges.
- The problem was to devise a walk through the city that would cross each of those bridges once and only once **and return to the same place**.
- Eulerian cycle
- Euler's argument shows that a necessary condition for the walk of the desired form is that the graph be connected and have exactly **zero nodes of odd degree**.

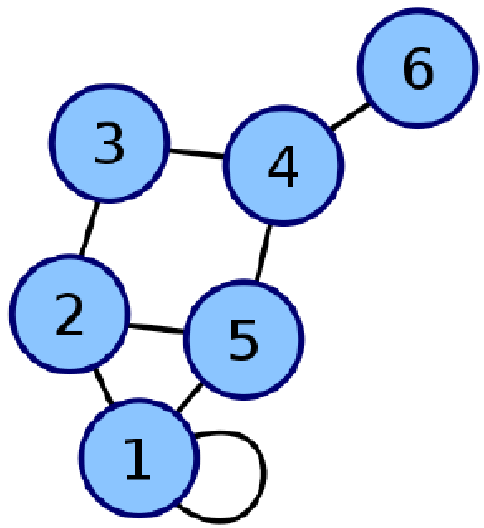




Adjacency matrix of a graph

- Graph can be represented by a linked list or a matrix
- An example of an adjacency matrix of **an undirected graph**

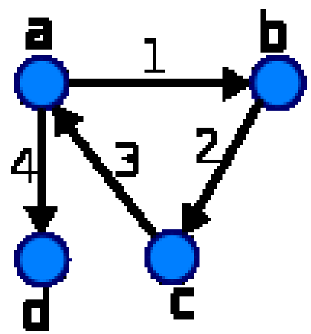
1	1	0	0	1	0
1	0	1	0	1	0
0	1	0	1	0	0
0	0	1	0	1	1
1	1	0	1	0	0
0	0	0	1	0	0



Adjacency matrix of a graph

- Directed graph (digraph, oriented graph)
  - Positive flow to one direction
  - Indegree and outdegree of a vertex
    (lähtöaste, tuloaste)

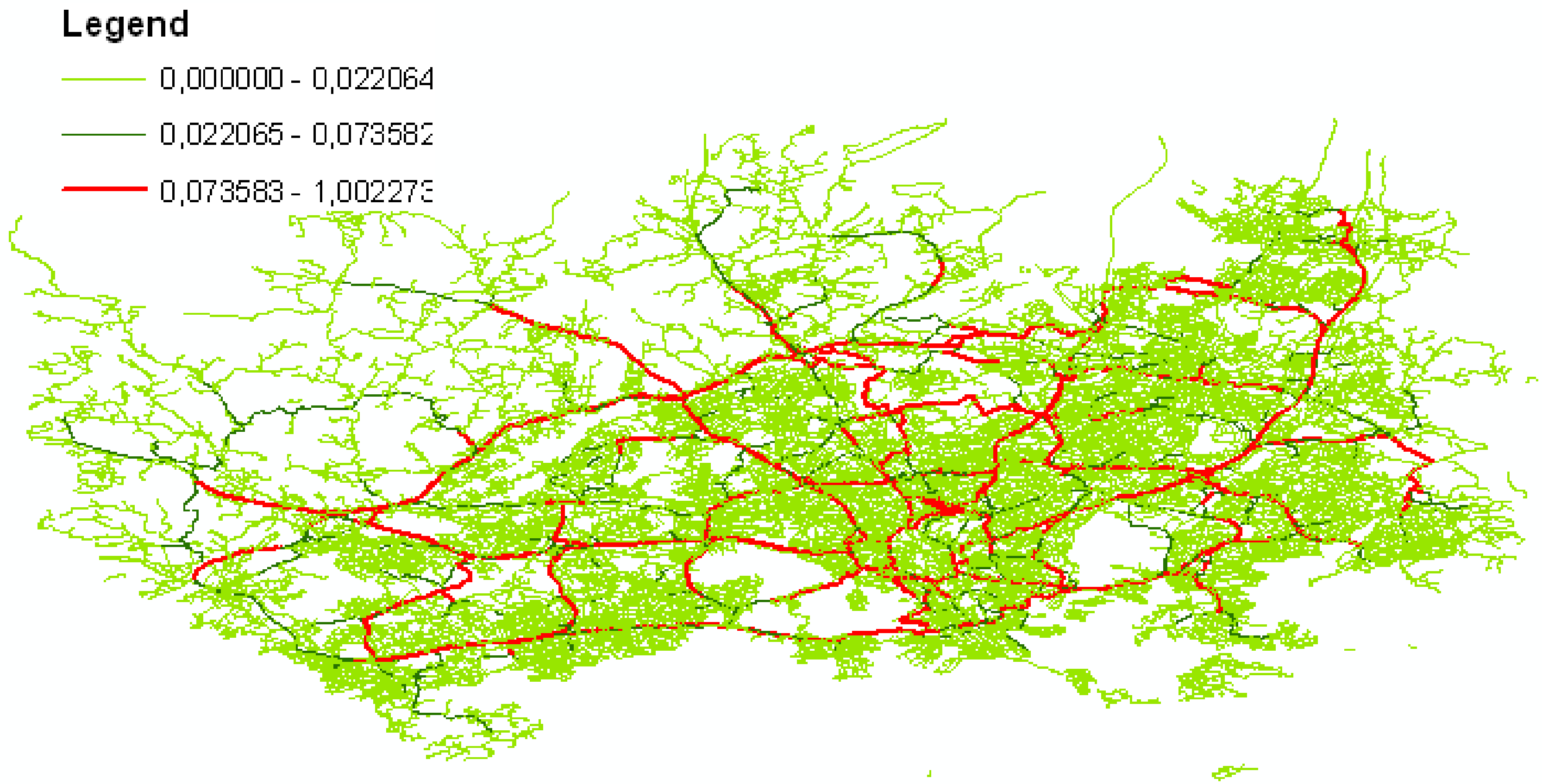
Notice the difference between this matrix and the one in previous slide! This shows the relationship between nodes (a-d) and links (1-4)



	1	2	3	4
a	1	0	-1	1
b	-1	1	0	0
c	0	-1	1	0
d	0	0	0	-1

Example of graph theory in GIS

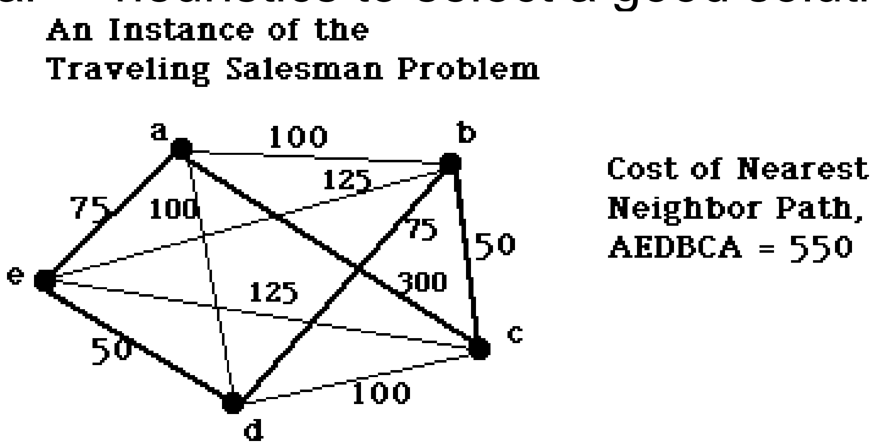
- Analysis of the connectivity or vulnerability of the network
  - Zhang Zhe (Aalto ENG, geoinformatics) studies critical networks
- Case: where are the critical parts of the road network of Helsinki in case of flooding, taking into account the number of population dependent on various parts of the network and possible alternative routings
- E.g. betweenness was a central concept in this study



Normalized betweenness map of Helsinki area Zhang 2016)

Analysis of network by optimization

- Routing in a network
  - Travelling salesman problem: *how to **reach each node once** and return to the starting point with minimal cost?*
  - Hamiltonian cycle
  - (n-1)!/2 possible solutions and no efficient means to find the optimal -> heuristics to select a good solution



Reading for the lecture

- O'Sullivan & Unwin (2010): pp. 68-71, Ch. 9.3 (pp. 250-261)

- Longley et al. (2015): Ch. 13.3.5 (pp. 310-313)