Aalto University
School of Engineering

Geoinformation in Environmental Modelling

Modelling: from concepts to data models

ENY-C2005
Jussi Nikander
16.1.2019
Slides by Paula Ahonen-Rainio and Jaakko Madetoja

Topics today

- Recap:
  - Two approaches in spatial modelling: **discrete objects and fields**
- Representation of objects and fields in **data models**: vector and raster data
- Spatial distribution: **relationships** between objects

Examples of potential exam questions relating to this lecture

- Paikkatieto ei voi koskaan esittää täydellisesti todellisen maailma kohteita tai ilmiöitä. Miksi ei, vai voiko?
- Geographic data can never fully represent objects or phenomena of real world. Why not, or can it?

- Todellisuutta voidaan mallintaa diskreetteinä objekteina (object-based modelling) tai spatiaalisesti jatkuvina ilmiöinä (field-based modelling). Esittele nämä mallinnustavat ja pohdi niiden soveltuvuutta kaupunkiympäristön, maanpinnan korkeuden ja väestötietojen mallinnukseen.
- We can model environment, at the conceptual level, as discrete objects (object-based modelling) or spatially continuous phenomena (field-based modelling). Explain these modelling approaches and discuss how they can be used for modelling urban environment, elevation of terrain, and demographic (population) information.

- Millä eri tavoilla spatiaalisesti jatkuva ilmiö voidaan esittää vektori-ja rasteritietona?
- In which different ways can a spatially continuous phenomenon be represented in vector and raster data?

Recap:
Geoinformation as a model of environment

Why do we model?

What should the model represent and how?

– complex, continuous, infinite real world
– a model is always selective and simplifying = interpretation
  $\Rightarrow$ level of detail, accuracy/reliability
– Discrete objects or spatially continuous fields?

- Geographic data:
  – location + attributes = static model
  – + time = dynamic model: processes, change

Recap:

Example: Discrete objects

Spatially discrete objects – Building blocks & urban infrastucture

Wikimedia

Recap:

Example: Field

Spatially continuos phenomenon – Depth of snow

yle.fi (Seppo Savolainen)

Object-based models (objekti, kohde)

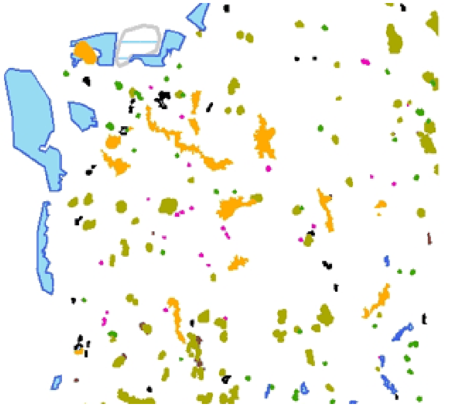
- Spatial representation of geographic reality in the form of geometric objects:

Objects known by their location and extent in an **otherwise empty** 2- or 3-dimensional space

  - Objects need to be **relevant**, **identifiable**, and **describable**
  - *Relevance* means that the object needs to be of interest to us
  - *Identifiable* means that we need to be able to distinguish the object from the rest of the world
    - location and geometry of the objects is important here
  - *Describable* means that the object needs to have characteristics we can represent
    - Attributes are important here, as is location and geometry

Example: Discrete objects

Road center lines
(Digiroad)
lines



Nature reservation areas
(SYKE)
polygons

Swimming places
points

Kuvat: paikkatietoikkuna.fi

Objects…

Geometric primitives in vector geometry:

- 0D: point
  - point object, sampled point (centre of area), reference point (for text labels)
- 1D: line
  - line object, line segment, line string, curve (with interpolation), polyline, closed polyline, ring
- 2D: polygon
  - Area object, interior area, polygon (simple, complex)
  - polygon with one exterior boundary and 0, 1 or more interior boundaries: e.g. boundary of Kauniainen is the interior boundary of Espoo
  - polygon network (with topology)
- 3D: volume

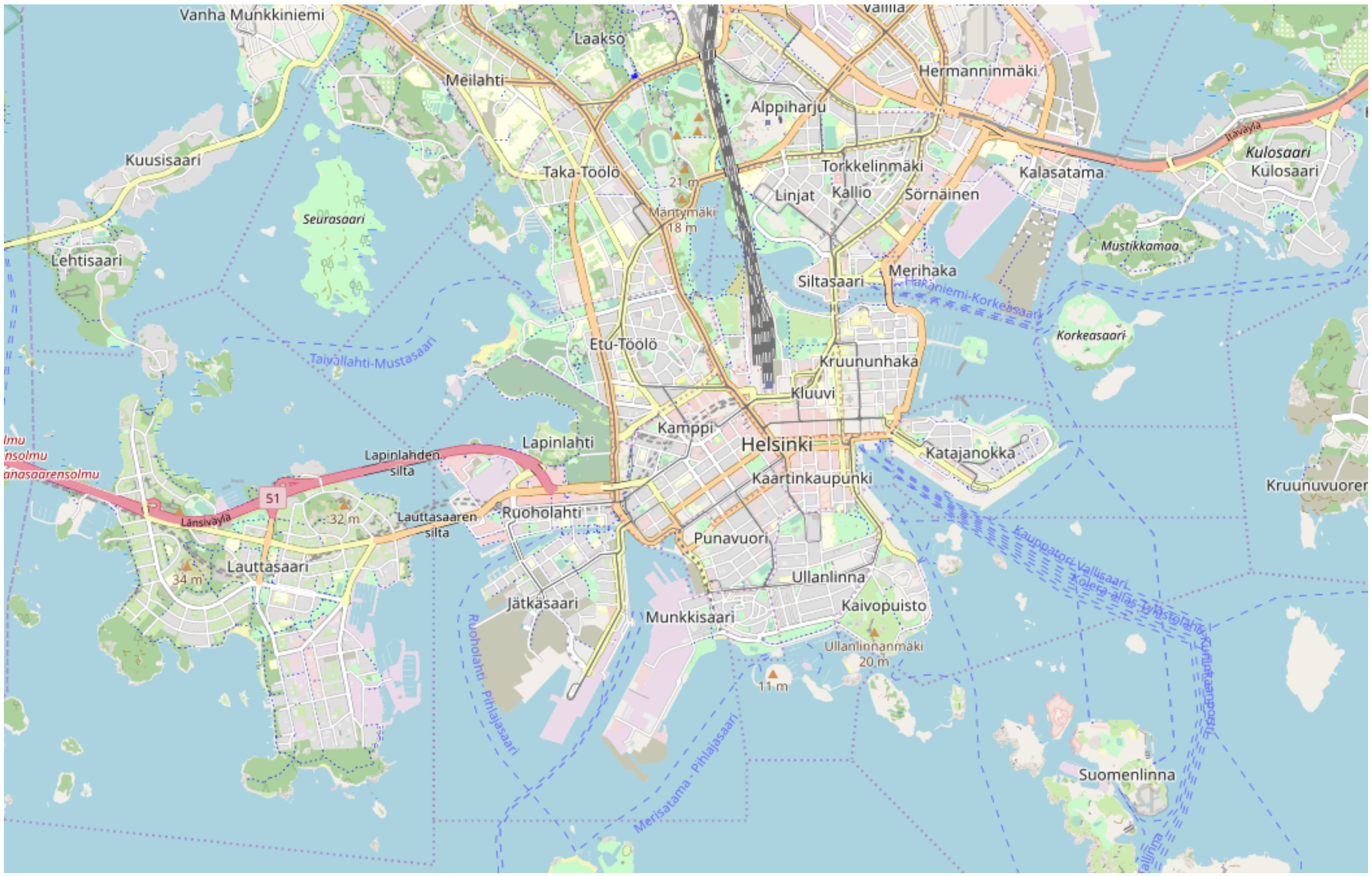
Objects and the vector data format

- Objects are often represented using a **vector** data format

- Geometry
  - Points, lines, polygons
- Attributes
  - ID, unique identifier
  - (spatial and aspatial) properties, *incl. object class/type*
  - *object class: a tree*
  - *Attribute information field: species*
  - *Possible attribute values: spruce, pine, birch, other*

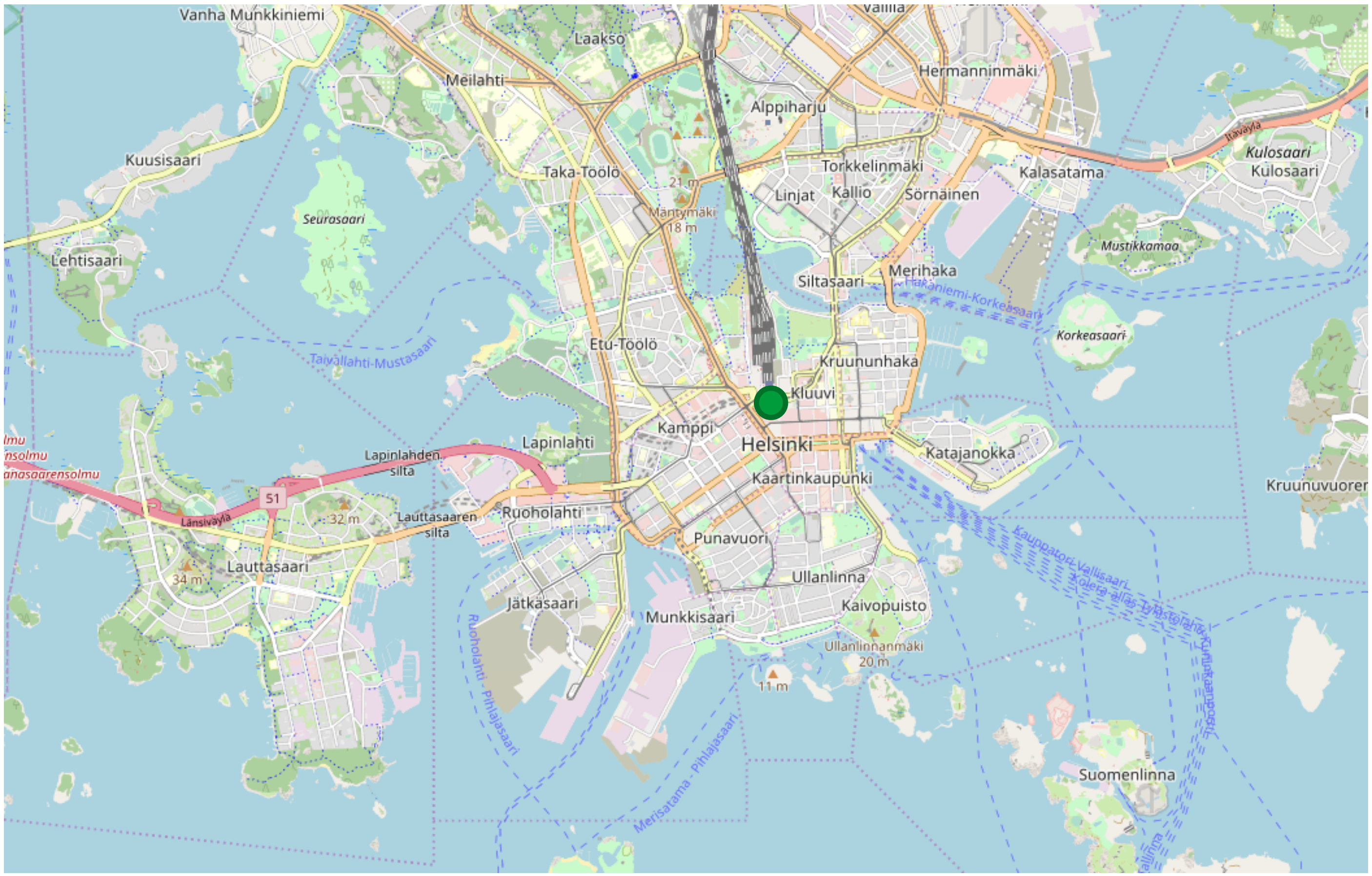
- Relations between objects
  - E.g. topological relations, metric relations (more about these later)

Classroom exercise: place Helsinki city center on the map



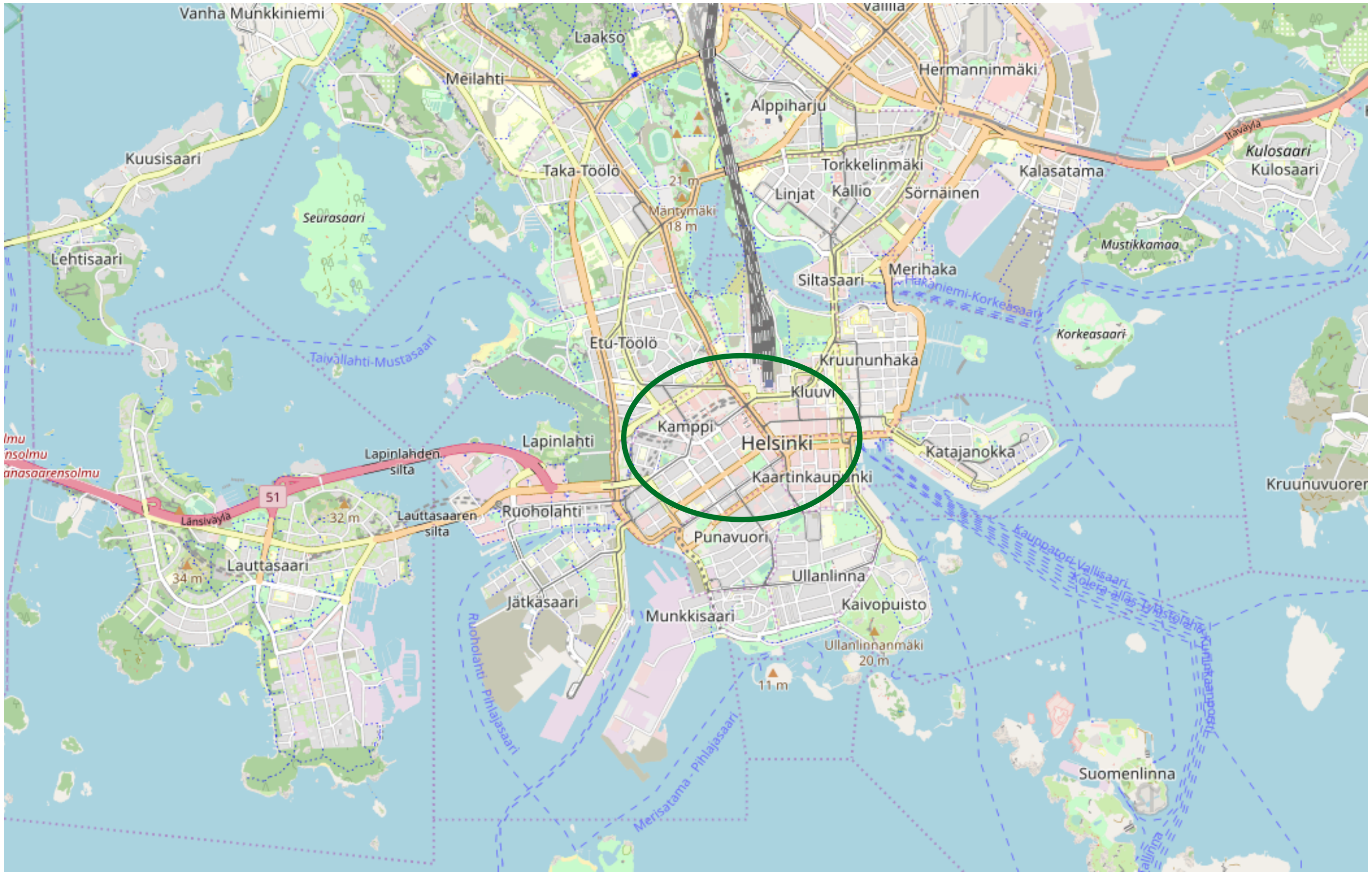
picture: OSM

Classroom exercise: place Helsinki city center on the map



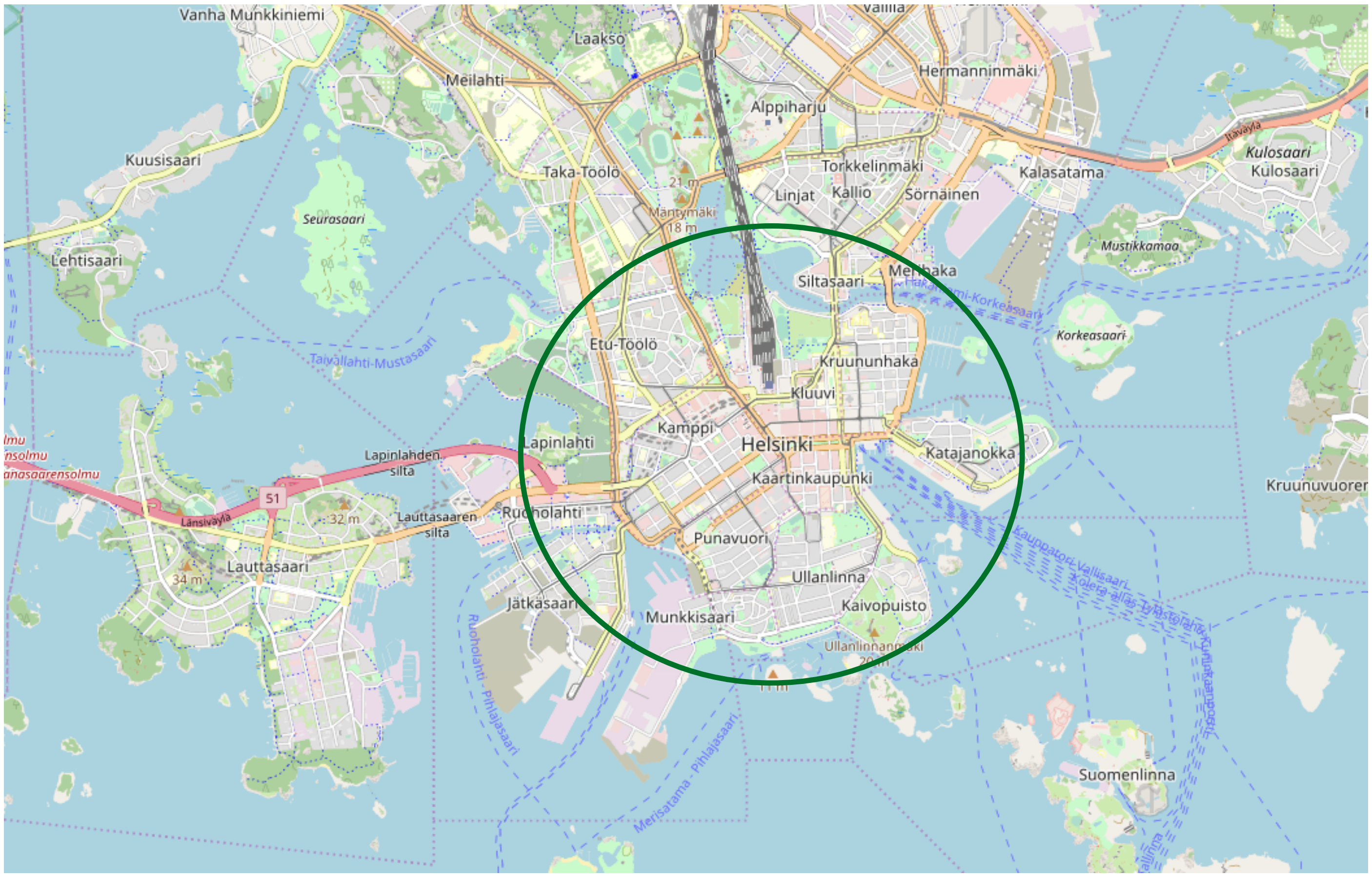
picture: OSM

Classroom exercise: place Helsinki city center on the map



picture: OSM

Classroom exercise: place Helsinki city center on the map



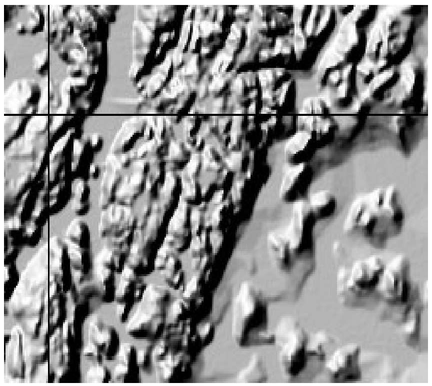
picture: OSM

Field-based models (kenttä, jatkumo)

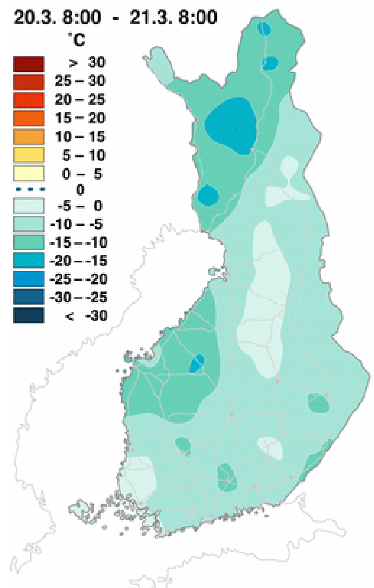
- Spatially continuous phenomenon, i.e., a variable with a value in any location of space (1…3D space)
    - A phenomenon that does not naturally split into homogeneous areas
    - Spatial continuity: every location has a value
      $\Rightarrow$ a function that gets a value in any location x, f(x)

  – Continuous or categorized values
    - Values on a continuous measurement scale (metric values): e.g. relief, precipitation, noice level, population density
    - Categorized values (nominal or ordinal measurement scales): e.g. land cover, soil type

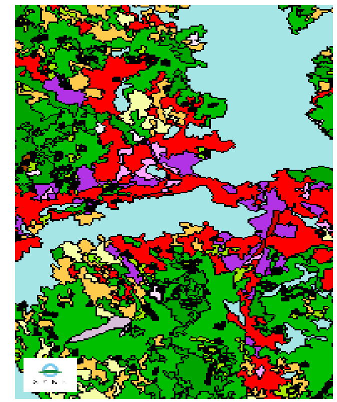
Example: Fields



Elevation
Maanpinnan korkeus
(visualized with shading)



Lowest temperature
Alin lämpötila
(visualized with zone map based on isolines)



Land use and cover
Maankäyttö
ja maanpeite
(visualized with chorochromatic map)

Kuvat: paikkatietoikkuna.fi, ilmatieteenlaitos.fi, syke.fi

Raster data model for fields

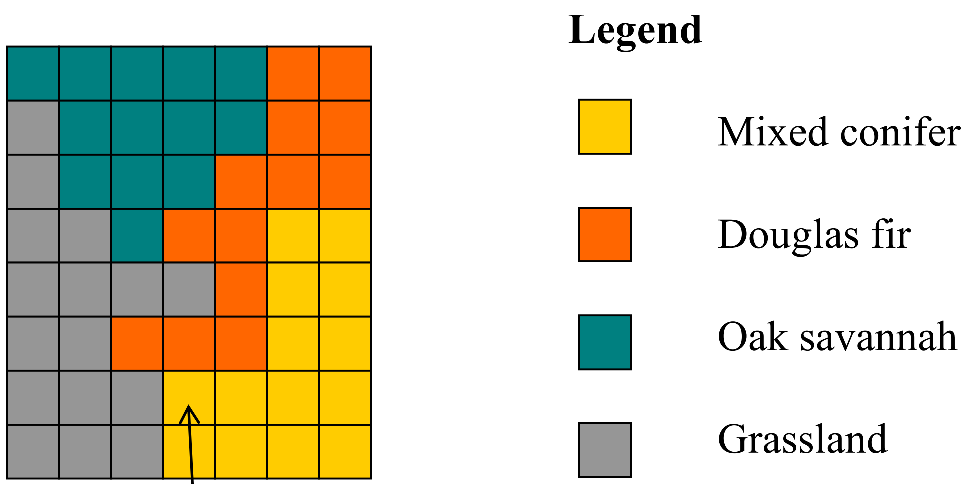
Raster (or a grid)

- A regular tesselation; composed of identical, regular **pixels**
  - the geometry is independent of the variation of the phenomenon;
    e.g. digital elevation model DEM
- The size of a pixel (or a grid-cell) = **resolution**
  - Should reflect the nature of the variation of the phenomenon and the purpose of use of the data

Examples:

- Land cover in raster by satellite imagery
- Demografic statistics (household income, number of children, voting behaviour…)
  - In some cases, the data is collected as individual objects, but then generalized into grid because of the privacy issues)

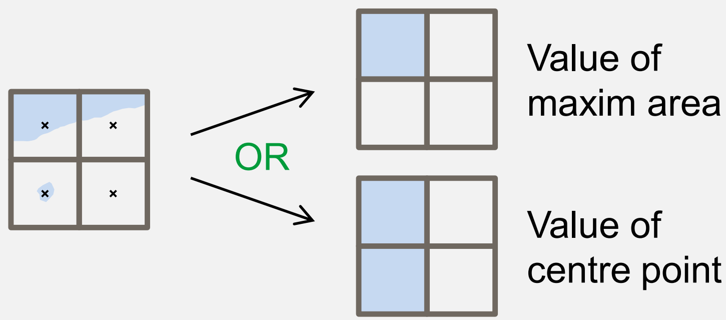
Example: Raster geometry



Longley et al. 2015

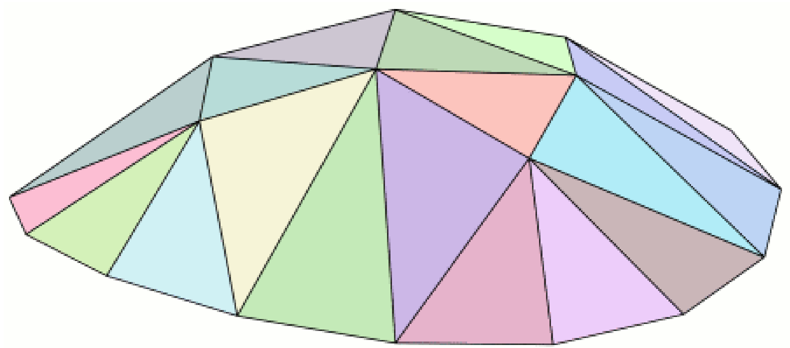
Notice:
Colour visualizes the values to users;
the actual data are either codes indicating the classes (as in this case) or metric values.

The value of a cell (pixel) can be the value of the variable in the centre point or the value dominating in the cell area.



Vector data models for fields

Triangulated Irregular Network, TIN
- Sample points: values of the variable (e.g. elevation of Earth surface) known in these points
- The points are linked so that they form triangles
- Linear interpolation within triangles, i.e. the plane defined by a triangle approximates the "surface" of the variable



Set of points in regular grid or irregularly
- Spatial interpolation among the set of points, with other methods than TIN (in a later lecture)

When the field represents a categorical variable, a "categorical surface"
- Polygon network
  - Adjacent polygons; Irregular areas
    Cf. a grid of geometrically regular grid cells
  - Idea: no variation of the value of the variable within a polygon

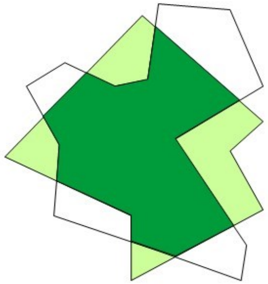
Reading:
O'Sullivan & Unwin (2010): Geographic information analysis pp. 5-9
Longley et al. (2015): Geographic Information Systems and Science Ch. 3.4-3.6, pp. 61-69

Overlay analysis (assignment A-1)

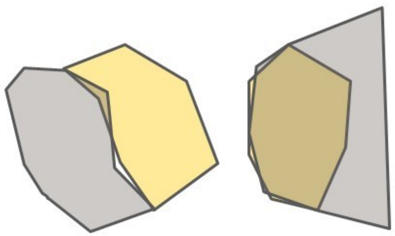
- One of the basic analyses in geoinformatics
- For vector and raster data

Examples from ENG-A1001

Esimerkki: Polygon overlay



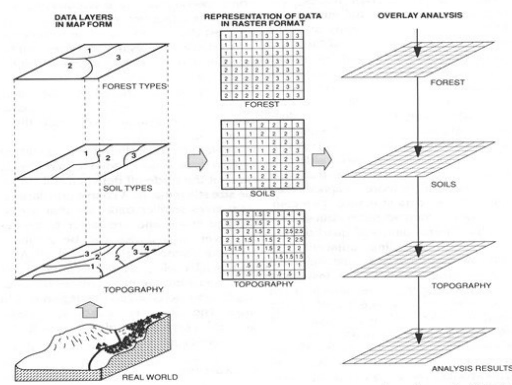
Tässä kahden polygonin leikkaus tuottaa yhdeksän uutta polygonia



Kahden eri prosessissa tallennetun polygonin reunaviivat poikkeavat käytännössä toisistaan sielläkin, missä reunaviiva esittää samaa todellisuuden kohdetta
⇒ leikkauksessa syntyy ohuita "pirstaleita", joilla ei ole reaali-maailmassa vastinetta

Aalto-yliopisto
Insinööritieteiden korkeakoulu

Rasteri-overlay



Longley et al. 2005

Overlay-analyysi on tehokas rasteritiedolle: eri rasteritasojen vastinpikseleitä verrataan ja/tai yhdistellään. Edellyttää kuitenkin, että rasteritasoilla on sama orientointi ja resoluutio.

Aalto-yliopisto
Insinööritieteiden korkeakoulu

Potential complications in representation

Real world entities may…

- be, and most often are, 3 dimensional
- change and move over time
- not have simple geometries, e.g. a transport network
- be fuzzy and/or have indeterminate boundaries
- have fractal characteristics

Their representation may…

- be strongly scale-dependent

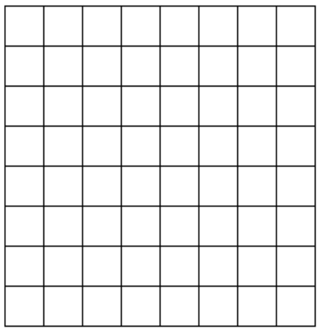
Reading: O'Sullivan & Unwin (2010) pp. 10-18

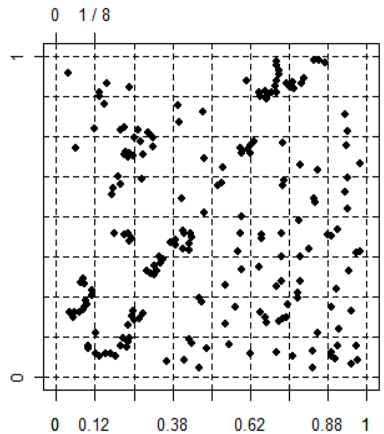
Fractal characteristics example: Level of detail

Wikimedia

almamedia.fi

Transformation of data model



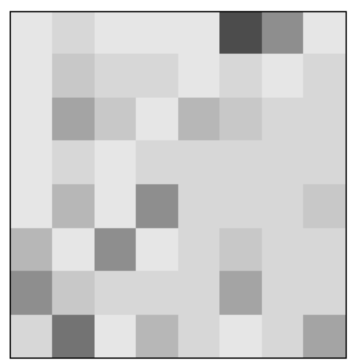


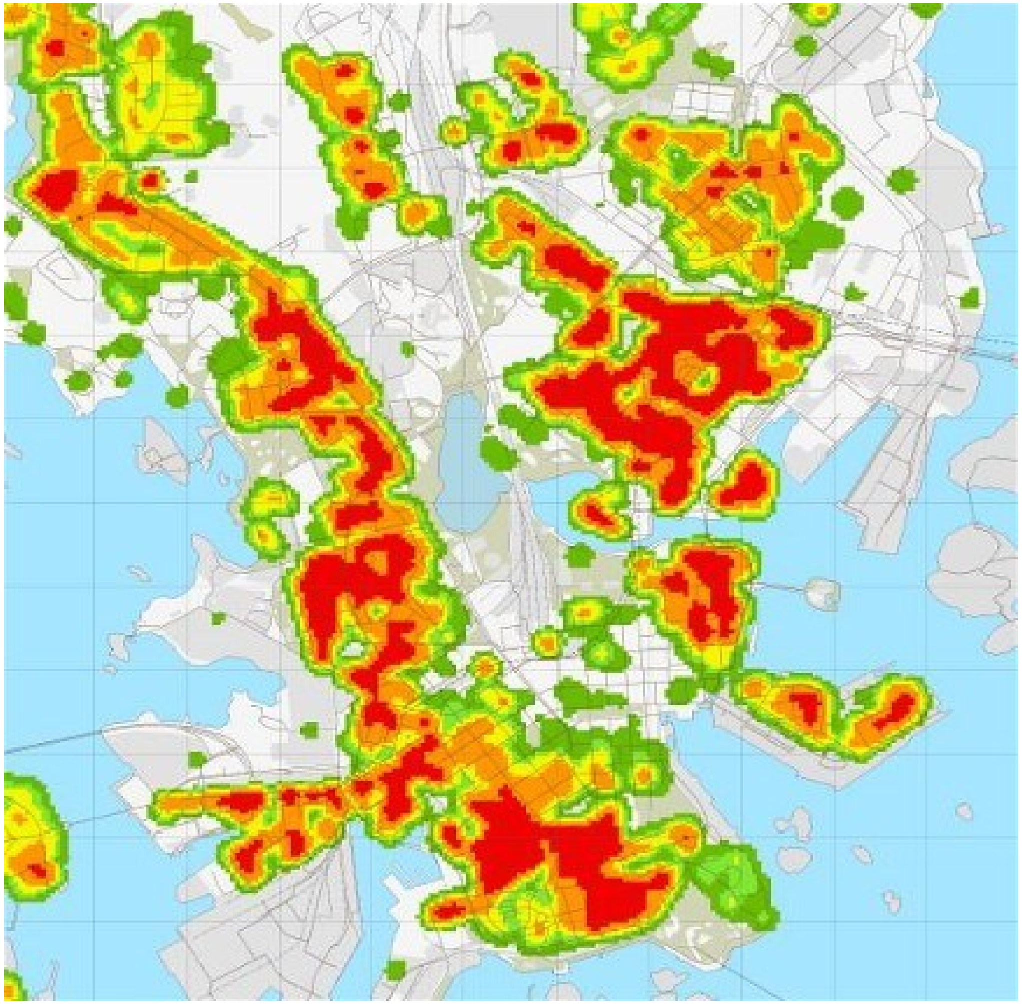
Point objects can be transformed into a field model by counting the sum of points in each cell of the given grid

Transformation

1	3	0	0	1	12	8	0
1	4	3	3	0	2	0	2
1	7	4	1	5	4	2	2
0	3	1	2	2	2	2	3
0	5	1	9	3	3	3	4
5	0	8	0	2	4	3	2
8	4	3	2	2	7	2	3
2	10	1	5	2	1	3	7

Visualization



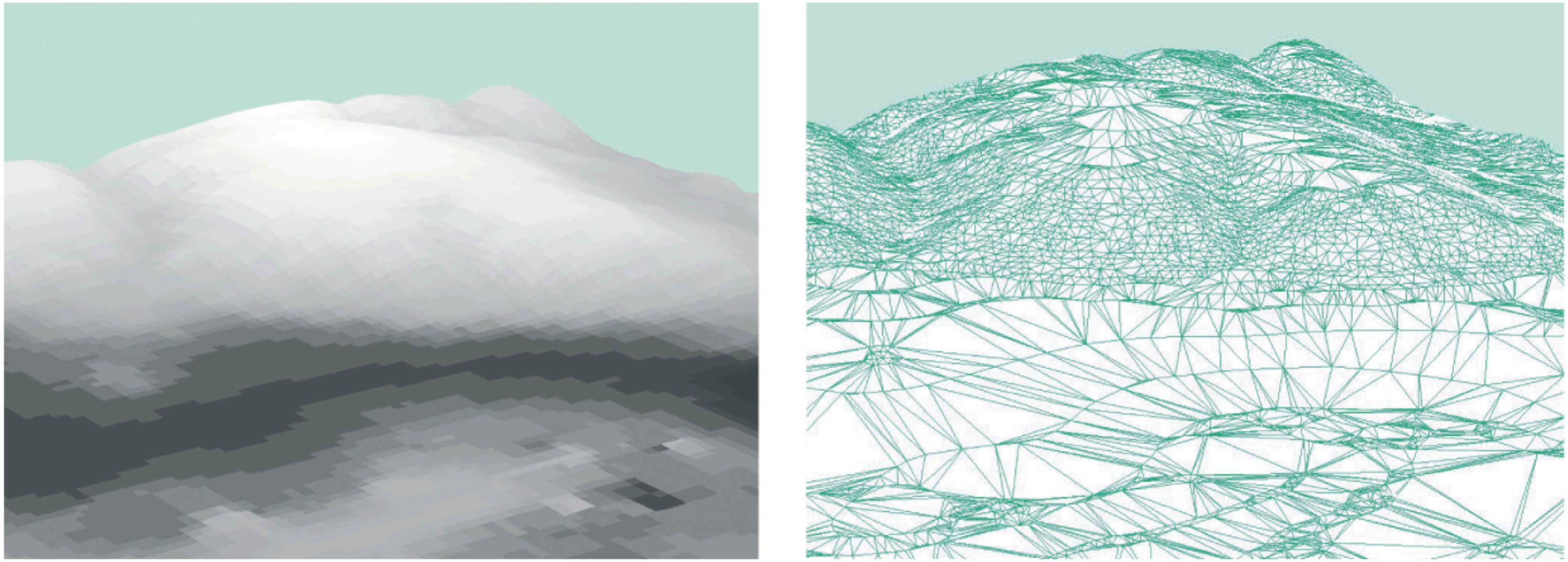


Incident points transformed to and presented as a surface
- Transformation by Kernel density function

(Krisp,2006)

Example: DEM and TIN

DEM: Digital elevation model as a grid



(a) DEM (b) TIN for region of varying complexity

Heywood et al. 2006

Classroom exercise

- How does the model influence the potential accuracy of representation?
  - Think about raster elevation model
    - 25m * 25m raster cells compared to 1m * 1m cells
  - And TIN model
    - Average distance between measured points being 0,5m, 5m, or 50m

Classroom exercise

- How does the model influence the potential accuracy of representation?
  - Think about raster elevation model
    - 25m * 25m raster cells compared to 1m * 1m and 0,1m * 0,1 m cells
    - The smaller the cell, the better the model is able to take into account small landforms
    - Is the value the same everywhere in the cell?
    - If not, how can we estimate values in different locations?
  - And TIN model
    - Average distance between measured points being 0,5m, 5m, or 50m
    - Closer the points, the smaller the individual triangles in the TIN and the better the model is able to take into account small landforms

Spatial distribution of objects

- Distribution of objects in space is in interest of analysis
  - described in terms of spatial relationships between objects
- These relationships can be for example:
  - Distance
    - Euclidian distance, distance in network, travel time, etc.
    - metric relationship
  - Adjacency
    - polygons that share an edge (common boundary)
    - **topological** relationship
    - other definitions are possible, such as, the nearest neighbour or distance-based adjacency
  - Interaction
    - Interaction between two objects is inverse-weighted by their distance; only pairs of adjacent polygons are considered

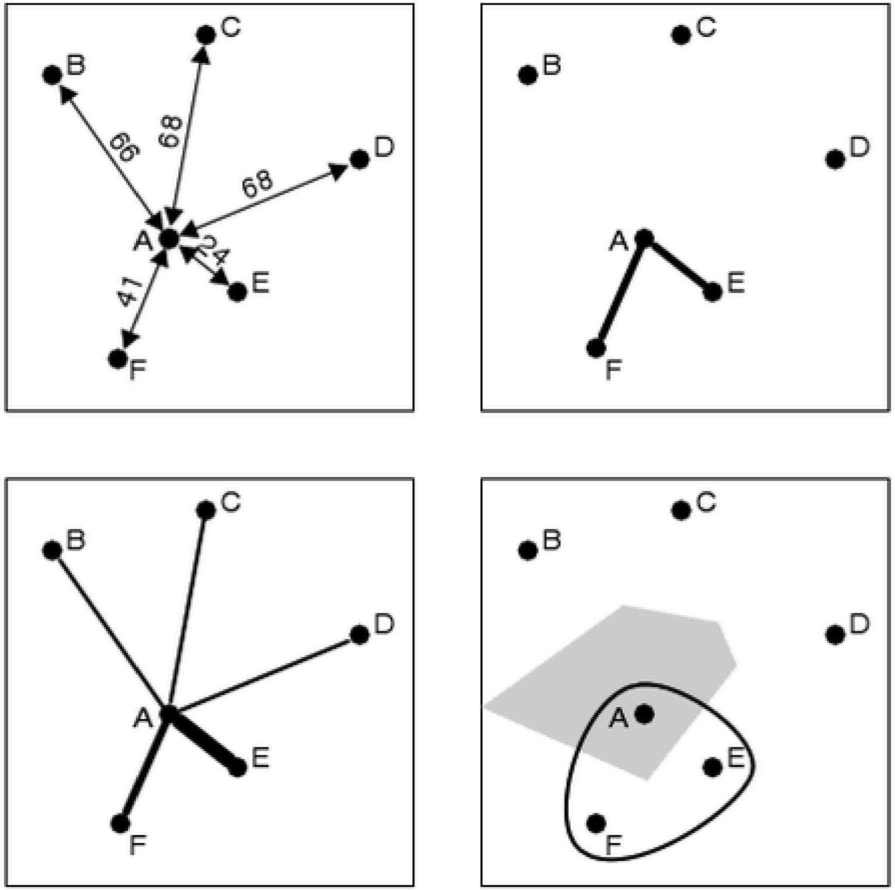


Figure 2.2 A schematic representation of the distance, adjacency, interaction, and neighborhood concepts.

O'Sullivan & Unwin (2010) p. 46

… spatial distribution

- Matrix presentation useful in analysis

Distance matrix D for distances between A, B, C, D, E, and F

$$\mathbf{D} = \begin{bmatrix} 0 & 66 & 68 & 68 & 24 & 41 \\ 66 & 0 & 51 & 110 & 99 & 101 \\ 68 & 51 & 0 & 67 & 91 & 116 \\ 69 & 110 & 67 & 0 & 60 & 108 \\ 24 & 99 & 91 & 60 & 0 & 45 \\ 41 & 101 & 116 & 108 & 45 & 0 \end{bmatrix}$$

Adjacency matrix A for A, B, C, D, E, and F

$$\mathbf{A}_{d \le 50} = \begin{bmatrix} * & 0 & 0 & 0 & 1 & 1 \\ 0 & * & 0 & 0 & 0 & 0 \\ 0 & 0 & * & 0 & 0 & 0 \\ 0 & 0 & 0 & * & 0 & 0 \\ 1 & 0 & 0 & 0 & * & 1 \\ 1 & 0 & 0 & 0 & 1 & * \end{bmatrix}$$

Reading: O'Sullivan & Unwin (2010) pp. 41-50

Proximity polygons

- Proximity polygon on an object is the region of the space that is closer to that object than it is to any other object.
- Defines the neighbourhood of the object
- Thiessen polygons (or Voronoi diagram) as in the figure
- Practical use: closest restaurant, health center, or school
  - Does not consider the road network!
  - Distance can be replaced with travel time (in some cases)

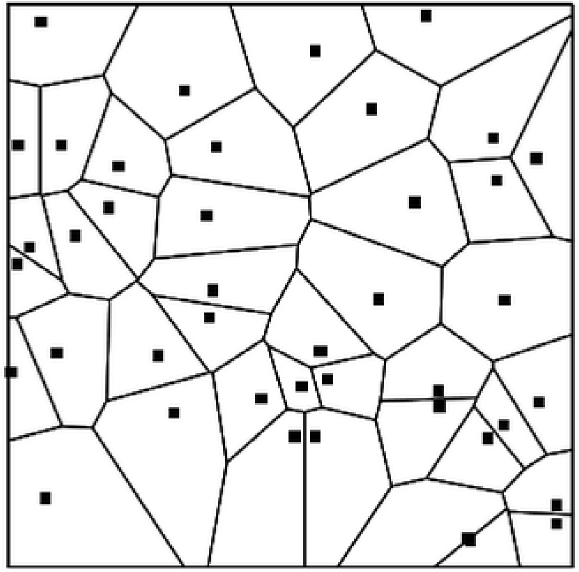


Figure 2.3 The proximity polygons for a set of point events.

O'Sullivan & Unwin (2010) p. 50

About the spatial relations

- Metric relationships
  - Basic relations distance and direction
  - their values change in any distortion of the geometry

- Topological relationships
  - do not change when the geometry is distorted, ”rubber sheet geometry”:
    - ”topological property is a property that remains constant when the spatial framework is stretched or distorted”
  - between points, lines and areas, or grid cells
    - location does not matter, but topology is calculated by using locational coordinates

…topological relationships

- can be identified by computing, but quite often stored in data structures for more efficient processing
  - many data structures based on explicitly stored topology
- are used to validate vector geometry and for operations such as network tracing and tests of polygon adjacency
- Essential topologies:
  adjacency, connectivity, intersection
- Notice the topological primitives:
  - Point > node
  - Line > edge, link; line network, graph
  - Polygon > face; polygon network

Reading: Longley et al. (2015) Ch. 7.2.3 Vector data model

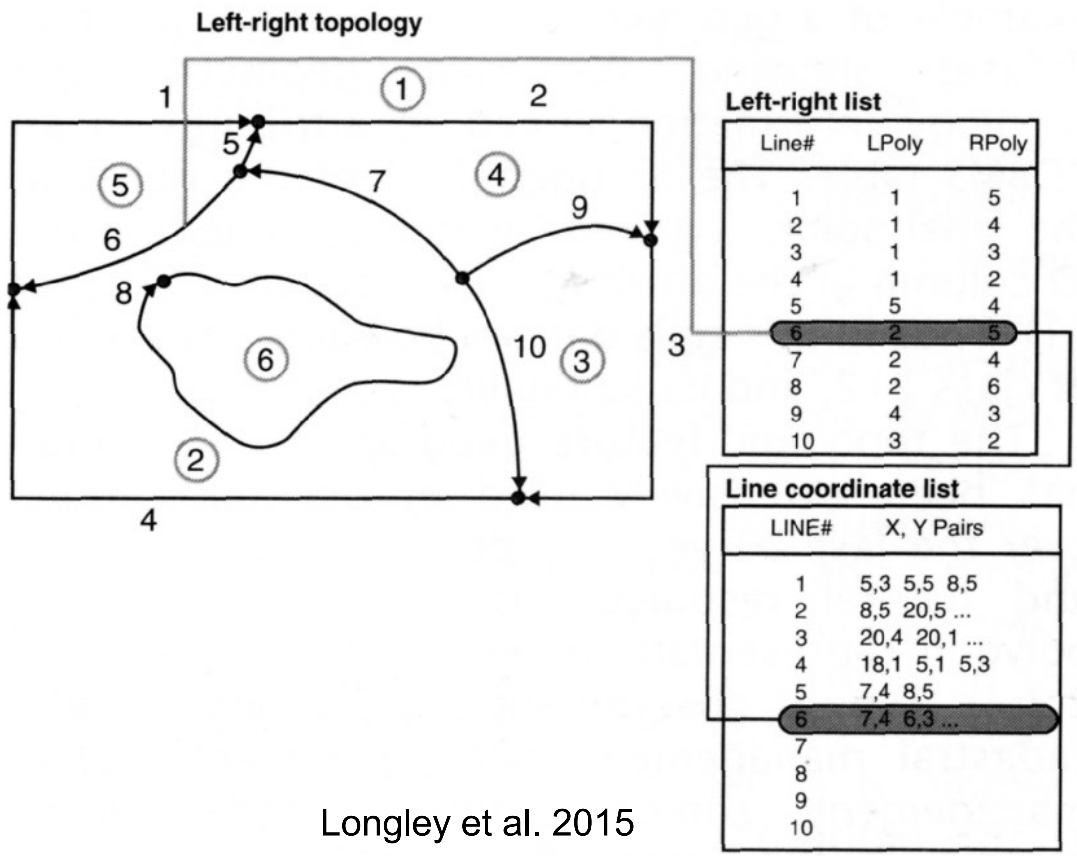
Adjacency

- Between two polygons if they share a boundary line
- between two lines if they share an end point
- can also be defined in higher orders
- Viereisyys

adjacency (viereisyys)

- A line has right and left neighbour area
- A line must have a direction
- Two neighbour areas share an edge
- A bit more complexity leads to topological data structures (not covered here)

Example: real estate property division (kiinteistöjaotus)



Left-right list

Line#	LPoly	RPoly
1	1	5
2	1	4
3	1	3
4	1	2
5	5	4
6	2	5
7	2	4
8	2	6
9	4	3
10	3	2

Line coordinate list

LINE#	X, Y Pairs
1	5,3 5,5 8,5
2	8,5 20,5 ...
3	20,4 20,1 ...
4	18,1 5,1 5,3
5	7,4 8,5
6	7,4 6,3 ...
7	
8	
9	
10	

Longley et al. 2015

Connectivity (or contiguity)

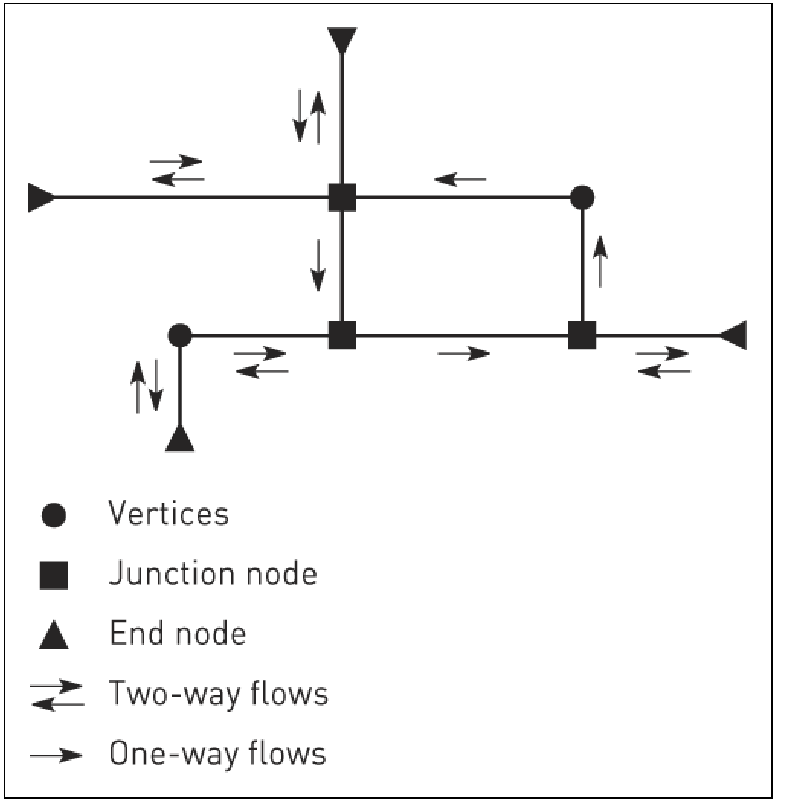
- in a graph two nodes are connected if there is a path between them; otherwise they are disconnected
- Cf. a line network (graph is the concept of graph theory)
- Yhdistävyys, jatkuvuus

connectivity (yhdistävyys, jatkuvuus)

Lines connect to each other in nodes
line = link
point = node
A link can be one-way or two-way
A link can have resistance (or speed)

Example: route optimization

linkki = link
solmu(piste) = node

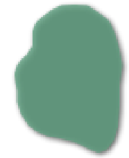


Heywood et al. 2006

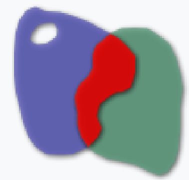
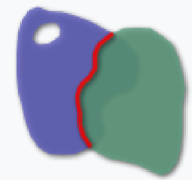
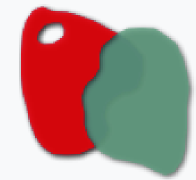
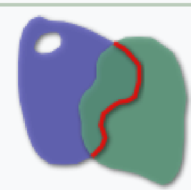
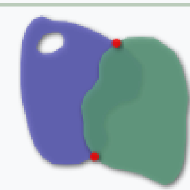
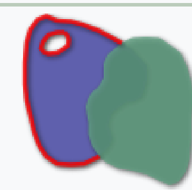
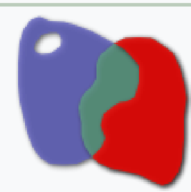
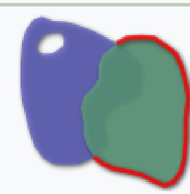
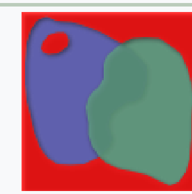
Intersection

- **Intersection** can be defined according to the mathematical topology between any of the elementary spatial objects
  - Intersection is used as the basis of the 9-intersection topology developed by Max Egenhofer
  - The Dimensionally Extended 9-Intersection Model used when topologies are computed ”on-the-fly” (as opposed to being explicitly or implicitly stored in a data structure)
- Leikkaus

b

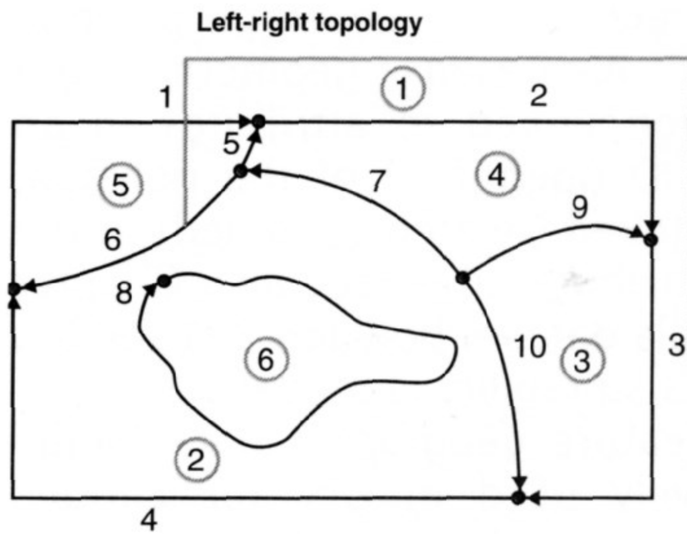


a

	Interior	**Boundary**	**Exterior**
Interior	 $\dim[I(a)\cap I(b)] = 2$	 $\dim[I(a)\cap B(b)] = 1$	 $\dim[I(a)\cap E(b)] = 2$
Boundary	 $\dim[B(a)\cap I(b)] = 1$	 $\dim[B(a)\cap B(b)] = 0$	 $\dim[B(a)\cap E(b)] = 1$
Exterior	 $\dim[E(a)\cap I(b)] = 2$	 $\dim[E(a)\cap B(b)] = 1$	 $\dim[E(a)\cap E(b)] = 2$

Classroom exercise

- Which polygon has the largest number of adjacent polygons?
- How many lines are adjacent to polygon 1?
- Which line is adjacent to smallest number of other lines?
- What about largest number of other lines?



Left-right list

Line#	LPoly	RPoly
1	1	5
2	1	4
3	1	3
4	1	2
5	5	4
6	2	5
7	2	4
8	2	6
9	4	3
10	3	2

Line coordinate list

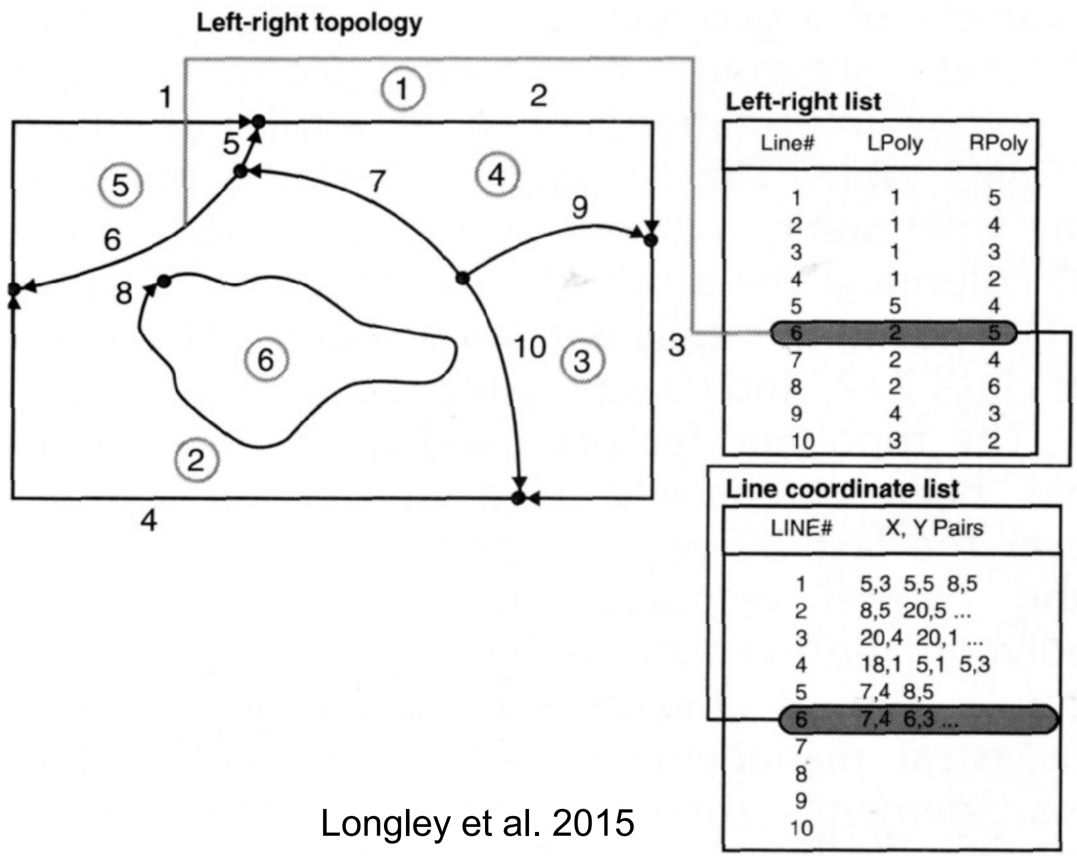
LINE#	X, Y Pairs
1	5,3 5,5 8,5
2	8,5 20,5 ...
3	20,4 20,1 ...
4	18,1 5,1 5,3
5	7,4 8,5
6	7,4 6,3 ...
7	
8	
9	
10	

- Remember elements are adjacent:
  - two polygons if they share a boundary line
  - two lines if they share an end point

Longley et al. 2015

Classroom exercise

- Which polygon has the largest number of adjacent polygons?
  - Polygon number 2
- How many lines are adjacent to polygon 1?
  - 4 lines (1,2,3,4)
- Which line is adjacent to smallest number of other lines?
  - Line 8 (0 lines)
- What about largest number of other lines?
  - All other lines are connected to 4 lines each



Longley et al. 2015

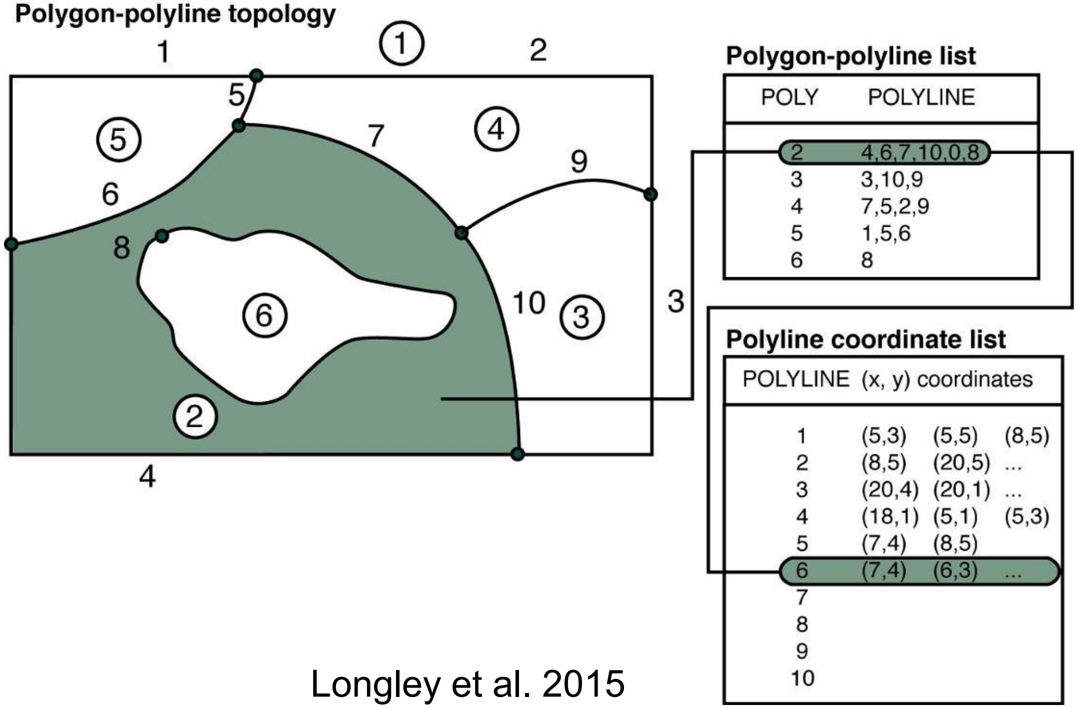
Hierarchical relationship: Composition

- Hierarchy of spatial objects:
  - Dividing area objects into smaller objects; subdivision of areas
  - Splitting longer lines into smaller pieces (line segments)
  - Composing lines by using points
  - Composing polygons by using lines or points

  - Sometimes in literature (e.g. in Longley et al. 2015) **composition** (and decomposition) is introduced as a topological relation

Composition (koostuminen)

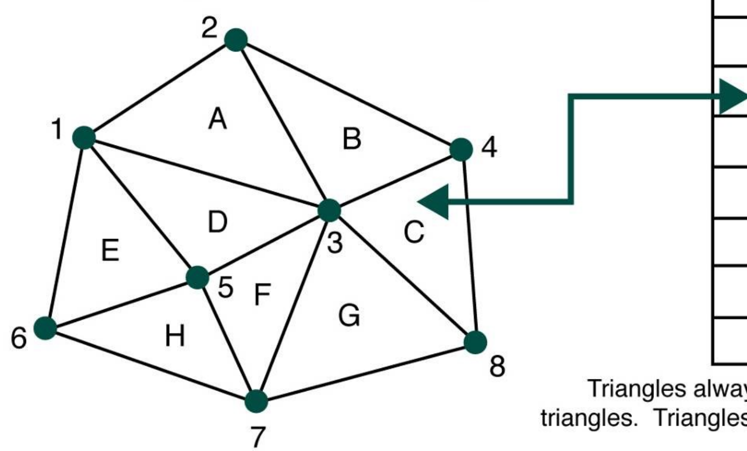
A polygon is composed of lines. A line is composed of points.



Longley et al. 2015

Example: TIN data structure

A TIN is a topologic data structure that manages information about the nodes that comprise each triangle and the neighbours to each triangle



Triangle	Node list	Neighbours
A	1, 2, 3	-, B, D
B	2, 4, 3	-, C, A
C	4, 8, 3	-, G, B
D	1, 3, 5	A, F, E
E	1, 5, 6	D, H, -
F	3, 7, 5	G, H, D
G	3, 8, 7	C, -, F
H	5, 7, 6	F, -, E

Triangles always have three nodes and usually have three neighbouring triangles. Triangles on the periphery of the TIN can have one or two neighbours.

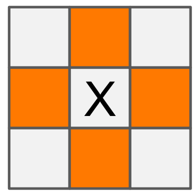
Periphery as a neighbour of a triangle is marked with ”-” in the table.

Raster topology

The primitive of raster data is a pixel;
adjacency is an elementary characteristic of grid/raster structure

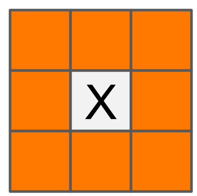
A pixel has

- four neighbours



or

- eight neighbours



Further neighbourhoods can be derived (e.g. 16 or 32) according to what is appropriate in analyses.

Terminology: Elements of geographic data

- Paikkatieto **Geographic data**
  - Sijaintitieto **Spatial data**
    - Vektoritieto **Vector data**
      - Piste **Point**
        Viiva **Line**
        Alue **Polygon**
        Kappale **Volume**
      - Topologia **Topology**
    - Rasteritieto **Raster data**
      - Pikseli, hila-alkio **Pixel, grid cell**
      - Rasteritopologia **Raster topology**
  - Ominaisuustieto **Attribute data**
    - Nominal scale
      Ordinal scale
      Interval scale
      Interval scale
      Cyclic scale

COORDINATES: Reference system

Reading for the lecture

- O'Sullivan & Unwin (2010): Geographic information analysis pp. 5-18, 41-50

- Longley et al. (2015): Geographic Information Systems and Science Ch. 3.4-3.6, 7.23