ENY-C2005 Geoinformation in environmental modeling

Lecture 5: Introduction to raster analysis

Aalto University
School of Engineering

Jussi Nikander

25.1.2023

Based on material originally created by Paula Ahonen-Rainio

Topics today

- **Short introduction to raster data**
- **Spatial analysis with raster data**
  - Map overlay
  - Analysis of surfaces
- **Map algebra**
- **Example problems**
  - Slope and aspect
  - Watershed
  - Cost surface and mobility
  - Viewshed

Examples of potential exam questions

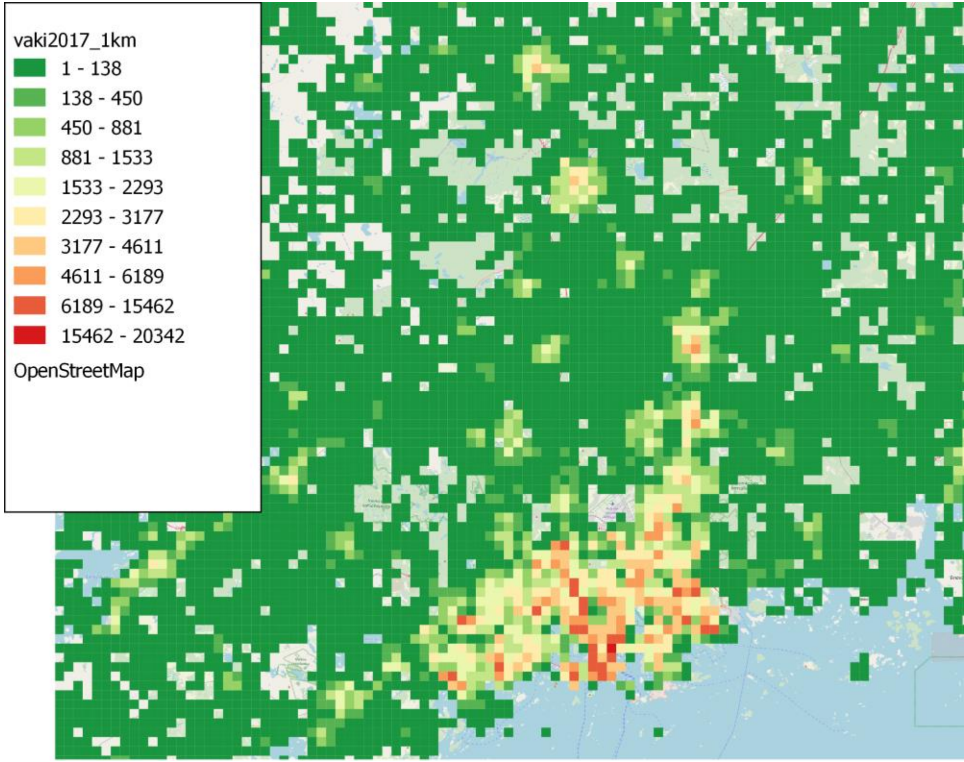
- **Give examples of spatial analysis on surfaces, such as the digital elevation model (DEM).**
- **Explain what a local/focal/zonal operation of map algebra means. Give examples of local/focal/zonal operations and concrete applications where they can be used.**
- **Perform [a given map algebra operation] on the [1...3 given] raster layers**

- **Anna esimerkkejä pintojen, kuten korkeusmallin (DEM), spatiaalisesta analyysista.**
- **Selitä mitä tarkoittaa kartta algebran lokaali fokaali zonaali operaatio. Anna esimerkkejä lokaaleista fokaaleista zonaaleista operaatioista ja konkreettisista sovelluksista , joissa niitä voidaan käyttää**
- **Suorita [annettu kartta algebran operaatio] [annetuille 1…3] rasteritasoille**

Rasters: a short introduction

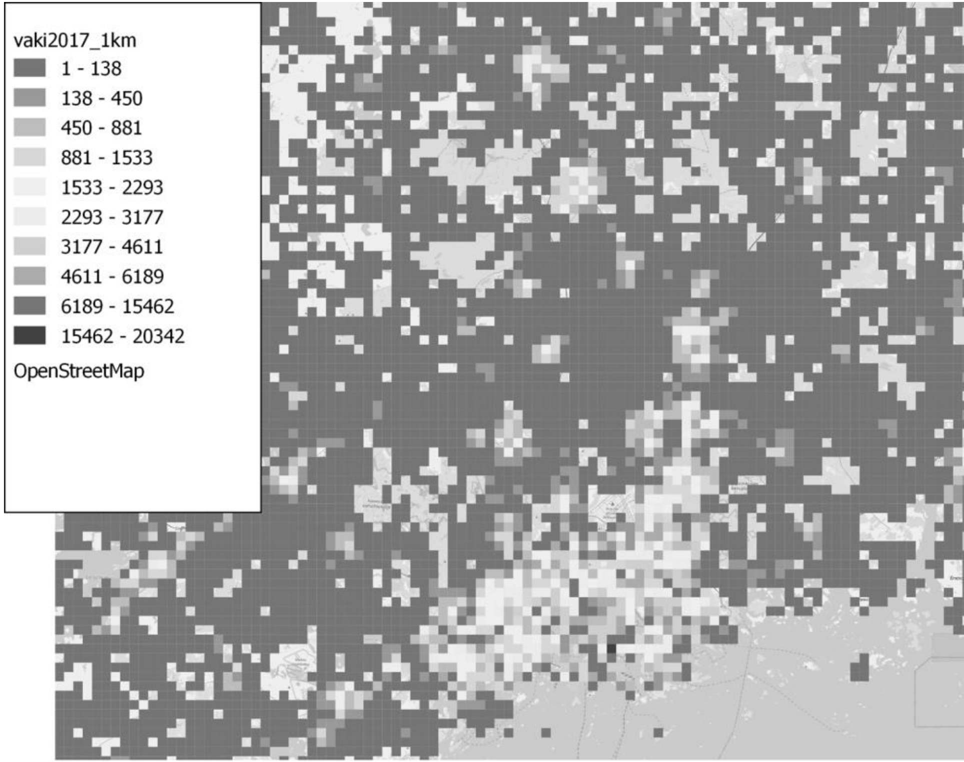
Spatial analysis of raster data

- **Spatial analysis combines location with attribute data in order to reveal patterns**
- **In case of raster analysis the focus is on raster data**
  - Rasters are often used to store field model data
  - Remember that other data models can also be used to represent fields (e.g. TIN)

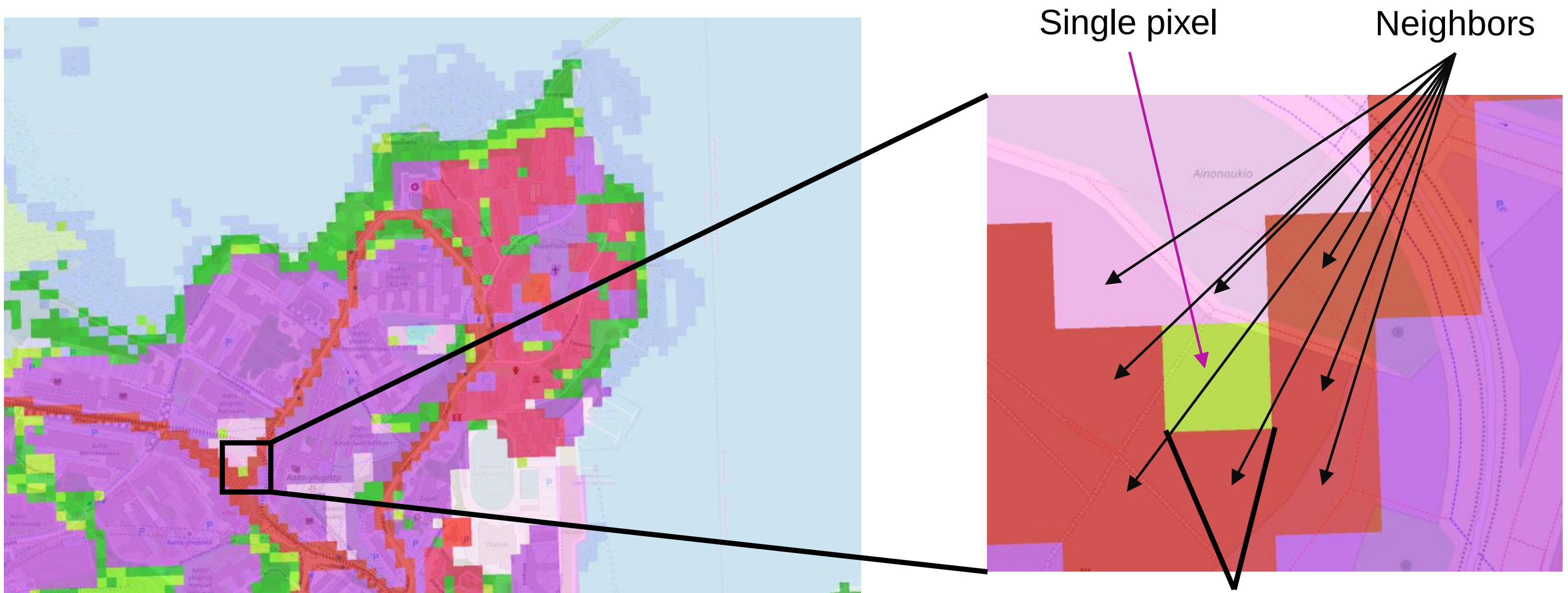


Spatial analysis of raster data

- **Spatial analysis combines location with attribute data in order to reveal patterns**
- **In case of raster analysis the focus is on raster data**
  - Rasters are often used to store field model data
  - Remember that other data models can also be used to represent fields (e.g. TIN)



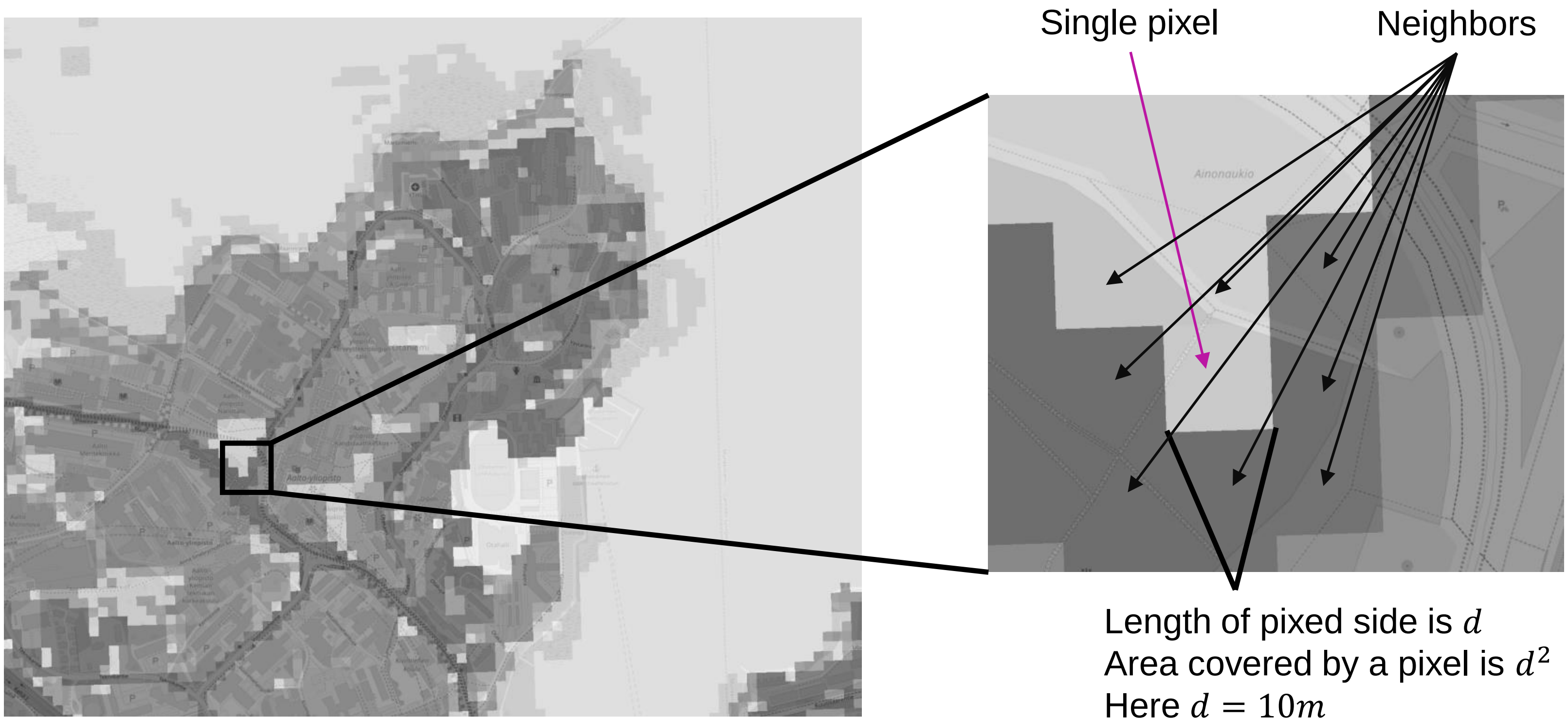
Example: raster (grid) data



Length of pixed side is d
Area covered by a pixel is d^2
Here $d = 10m$
The pixel value is always a number.
Interpretation of the value can be nominal.
Color is for visualization.

Nominal data (CORINE land cover) over basemap (OSM)

Example: raster (grid) data

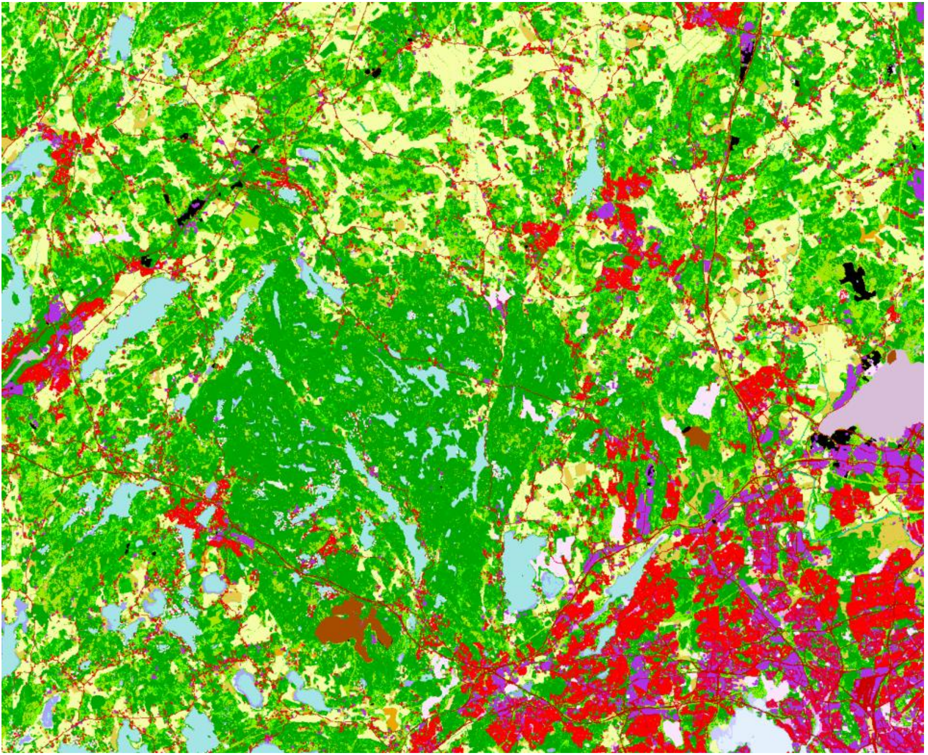


Length of pixed side is d
Area covered by a pixel is d^2
Here $d = 10m$
The pixel value is always a number.
Interpretation of the value can be nominal.
Color is for visualization.

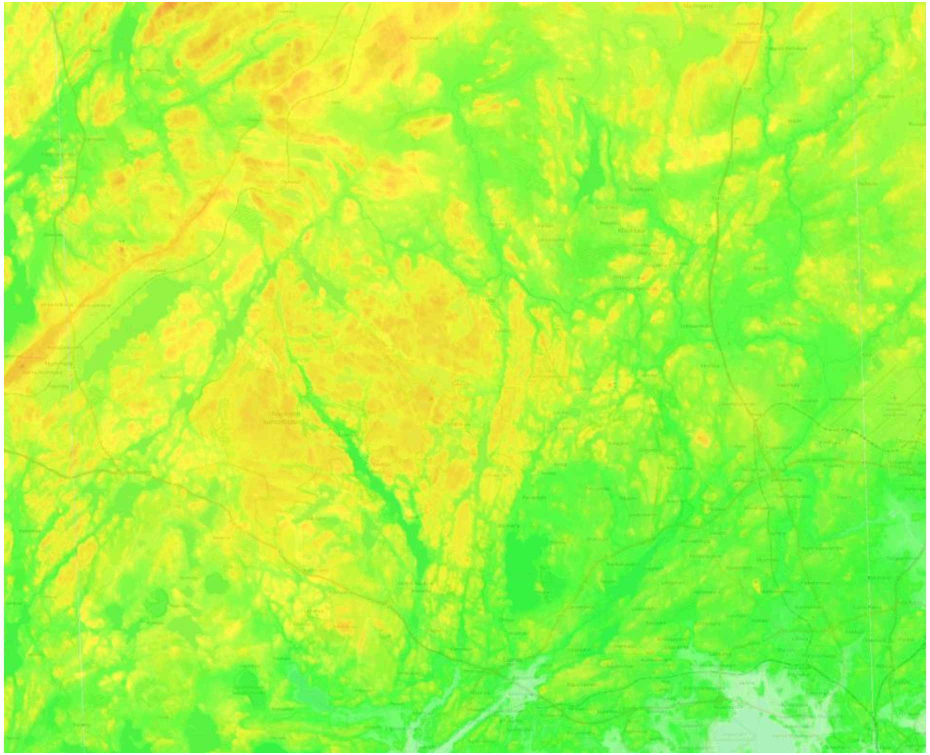
Aalto University
School of Engineering

Nominal data (CORINE land cover) over basemap (OSM)

Categorical and numerical data

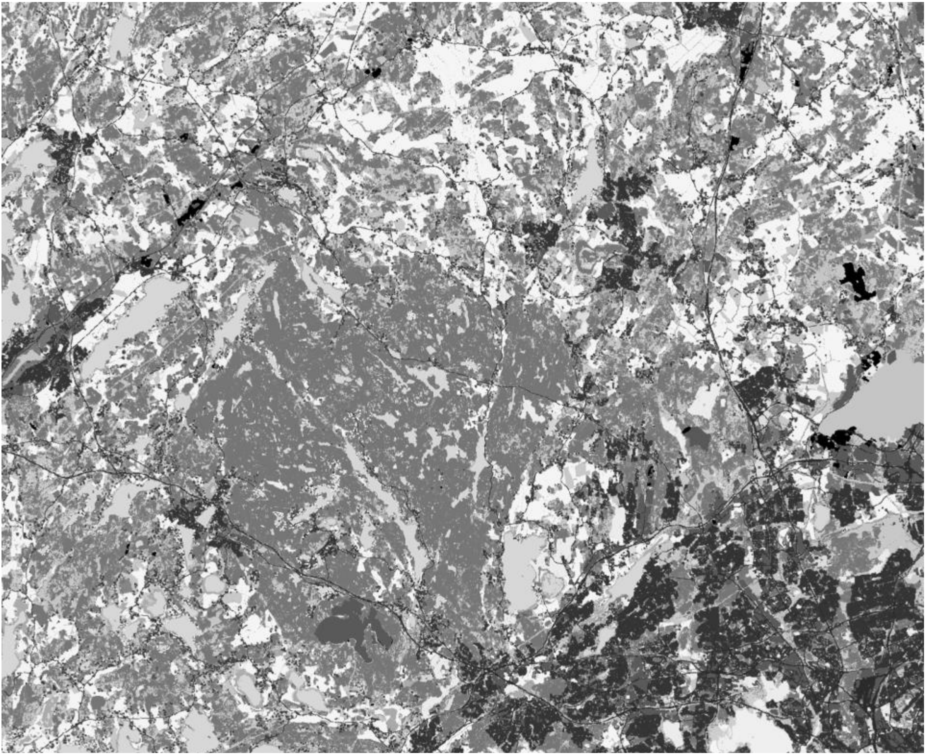


Source: CORINE land cover



Source: MML elevation model

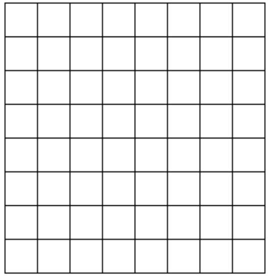
Categorical and numerical data

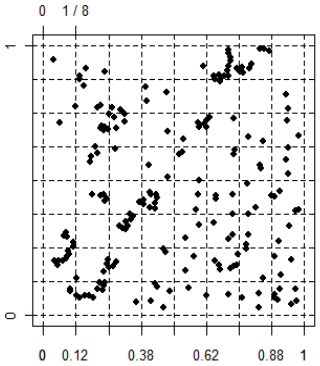


Source: CORINE land cover

Source: MML elevation model

Data relationships: vector and raster

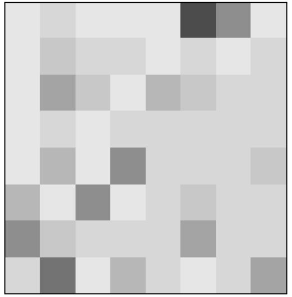




Transformation

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

Visual
ization

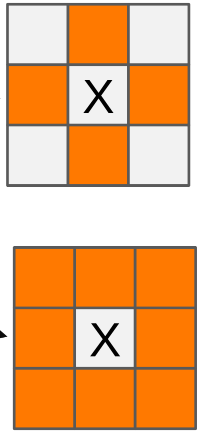


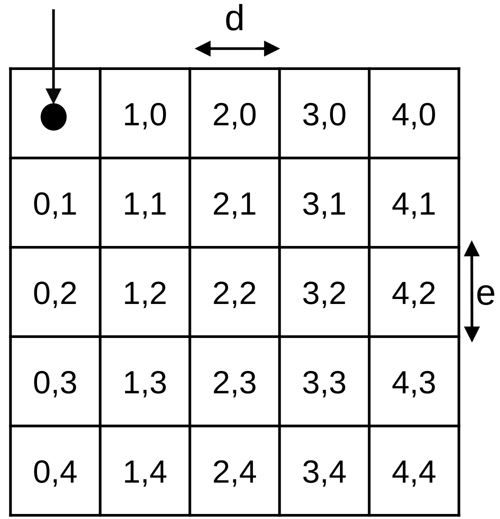
Object data (in this case point objects) can be transformed into a field model (a raster representing number of elements) by counting the sum of points in each cell of the given grid

Rasters, in turn, can be transformed into objects (details depend on the input)

Raster data fundamentals

- **Geometry**
  - Shape of a pixel: usually a square (can be a triangle, a hexagon)
  - Square pixel:
    - 4 neighbours (joint edge)
    - 8 neighbours (joint edge or corner)
- **Implicit topology**
  - Joint edge – adjacency (viereisyys)
  - Joint edge or corner – connectivity (yhdistävyys)
- **Orientation, origin, resolution (size of a pixel/grid cell)**





- **the geographic coordinates for origin are known**
  - Coordinates for other cells are calculated from that using cell position (i,j) and the x- and y-offsets (d and e in the figure): $x_i = x_0 + di$, $y_j = y_0 - ej$

Aalto University
School of Engineering

Raster data management

- **One raster layer contains one theme**
  - e.g. Soil layer, vegetation layer, elevation layer
  - Pixel values according to the theme
- **Layers can be stored in files**
  - E.g. geotiff-images with associated metadata
- **...or in databases**
- **Data can also be compressed**
  - RLE – run-length encoding
  - Quad-tree
  - (other) image compression methods
  - Methods must be lossless, information needs to be preserved

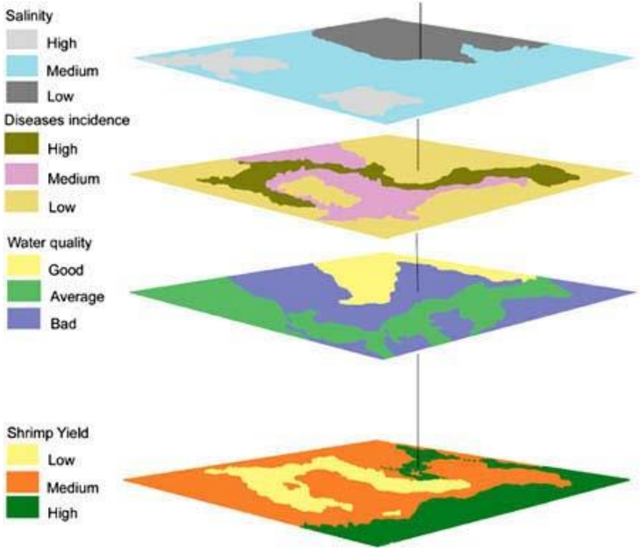


Image source: Geographic information systems in fisheries management and planning. FAO. http://www.fao.org/3/y4816e/y4816e00.htm

Aalto University
School of Engineering

Spatial analysis of raster (grid) data

Basics of raster analysis

- **Raster analysis is based on**
  - Combining data in several raster layers (overlay)
  - Analyzing data in one (or several) raster layers based on
    - Pixels at same coordinates
    - Neighborhood of each pixel (grid cell)
    - Separately defined areas
    - Other means (distance & direction, etc.)
- **Output of raster analysis is a new raster layer (or layers)**

1	1	0
0	1	1
0	0	1

And

0	0	0
1	1	0
1	1	1

=

0	0	0
0	1	0
0	0	1

Focal avg

3	5	2
7	6	3
8	5	3

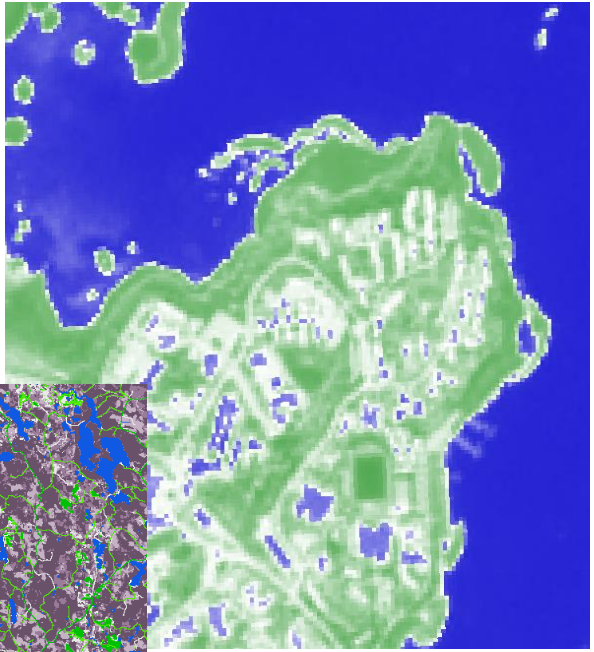
=

	4,7	

Examples of raster analysis applications

- **Environmental analyses with satellite images**
  - E.g., detecting changes in land use, vegetation,...
- **Analyses based on elevation models**
  - Viewshed (visibility, näkemäanalyysi)
    - e.g. locating telecommunication towers, view from a road in landscape planning, military applications
  - Slope (kaltevuus) and aspect (suunta)
    - e.g. watershed (vedenjakaja), catchment areas, estimating growth conditions, risk of avalanche
- **Analyses of demographic (population) and other statistical datasets (in grid form)**
- **Terrain analyses, e.g. for mobility**
  - For crisis management and military applications, prediction of animal movement, forestry

Source: Sentinel 2 NDVI



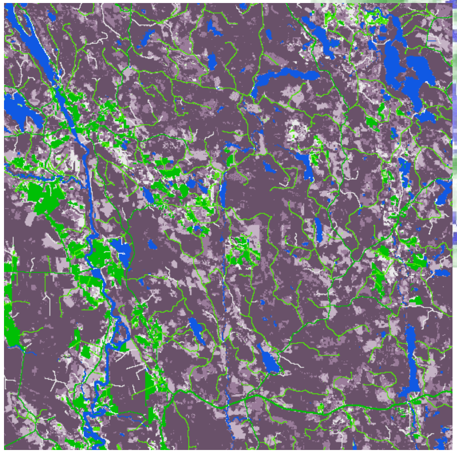
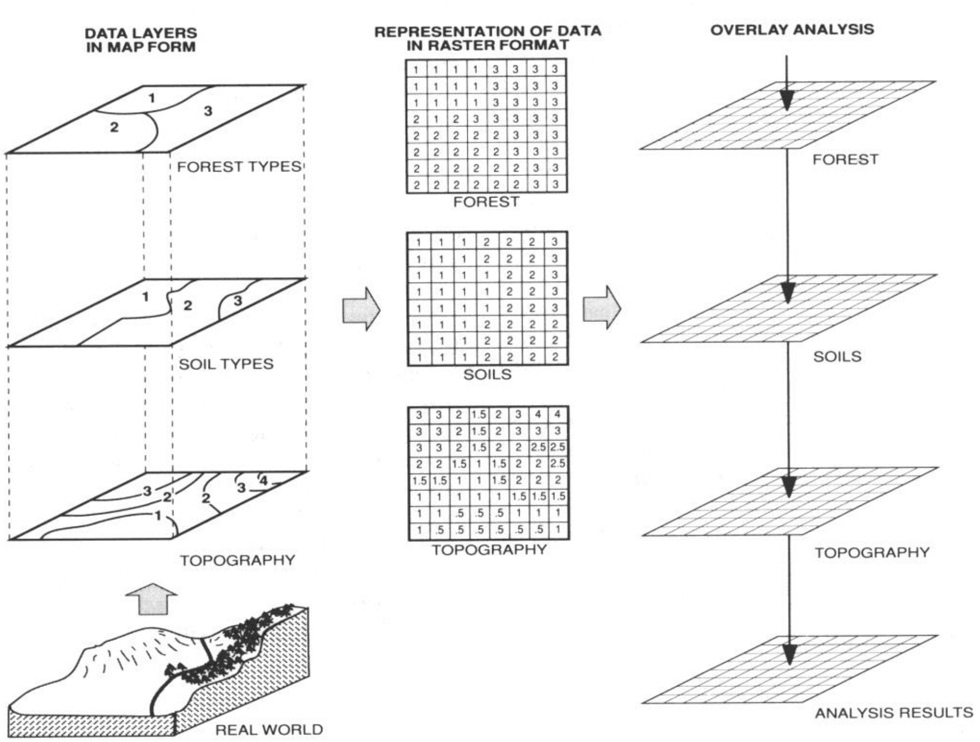


Image source: Jussi Nikander

Raster map overlay

- **The most classic raster data analysis method**
  - Still a backbone of GIS
- **New data layers as a function of two or more source layers**
- **Simple to compute for raster data**
- **Combined Vector and raster map overlay requires data transformation (vector-to-raster)**



Aalto University
School of Engineering

Raster map overlay

- **Basic form based on binary logic ⇒ Boolean logic**
  - On each layer, possible values for each cell are 1 (true) or 0 (false)
  - E.g., suitability of a region to a certain purpose by logical reasoning: different layers have binary values for each location, the result based on the logics on these values
- **Potential problems because of inconsistencies**
  - Input data may originally be in different scales or in different coordinate systems ⇒ resampling
  - Input data may be interpolated (DEMs) or cartographically generalized (e.g. roads on a map are wider than in reality)
  - If the accuracy of the original datasets is not known, results may be of no value (as in any analysis)

T	T	T	T	T
F	T	F	T	T
T	F	T	T	T
T	T	T	T	F
T	T	T	T	F

and

T	T	F	T	T
F	T	T	F	F
F	F	F	F	T
T	F	F	F	T
T	T	F	F	F

=

T	T	F	T	T
F	T	F	F	F
F	F	F	F	T
T	F	F	F	F
T	T	F	F	F

True=1 and False=0

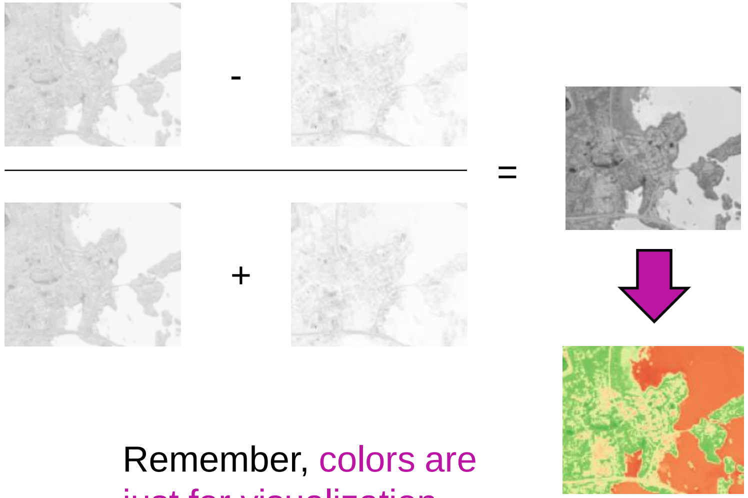
Raster map overlay

- **Map overlay operations can contain a several input layers and their relations to each other can be complex**
- **E.g. calculation of Normalized Difference Vegetation Index (NDVI) takes two input layers**
  - Near infrared band (NIR) and red band
- **Uses the function:**

$$\boldsymbol{NDVI = \frac{NIR - Red}{NIR + Red}}$$

- **This can be also be scaled:**

$$\boldsymbol{NDVI = 100 * \frac{NIR - Red}{NIR + Red}}$$



Remember, colors are just for visualization

Aalto University
School of Engineering

Classroom exercise: raster operations

1	2	2
0	1	2
0	1	1

+

1	0	0
0	1	1
1	2	1

=

1	2	2
0	1	2
0	2	1

*

1	0	0
0	1	1
1	2	1

=

Classroom exercise: raster operations

1	2	2
0	1	2
0	1	1

+

1	0	0
0	1	1
1	2	1

=

2	2	2
0	2	3
1	3	2

1	2	2
0	1	2
0	2	1

*

1	0	0
0	1	1
1	2	1

=

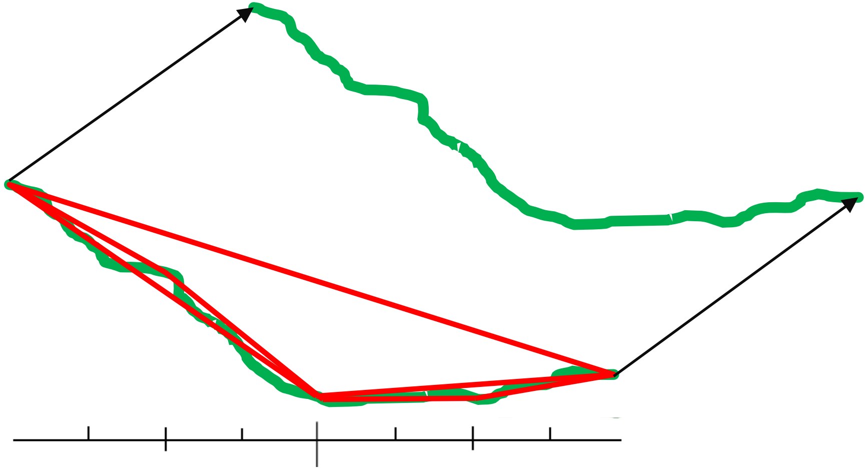
1	0	0
0	1	2
0	4	1

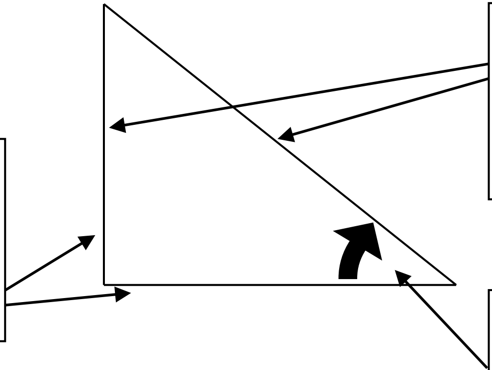
Analysis of raster surfaces

- **Focus is on one raster layer**
  - E.g. a DEM
- **The aim of the analysis is transform one field phenomenon to another**
  - Elevation to slope and aspect
  - Elevation to viewshed
  - Cost surface to movement time
  - Etc.
- **Methods can be applied to any field phenomenon with numerical values**
  - But does it make sense to measure the viewshed of average temperature?
- **The analysis must be meaningful for the phenomenon in question**

Slope analysis

- **Strongly dependent on scale, various expressions**
- **There are three alternative definitions of slope**





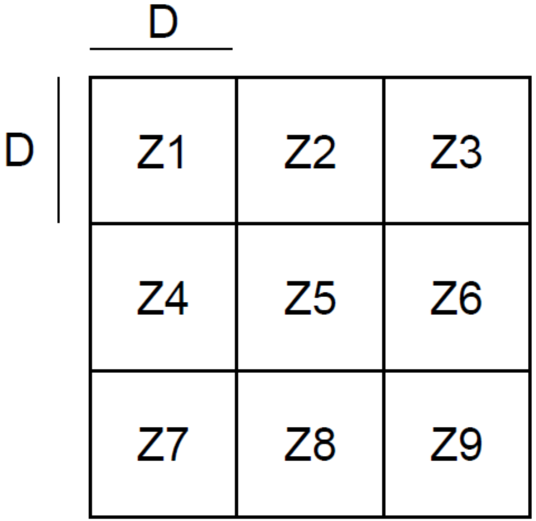
The *ratio* of the change in elevation to the horizontal distance traveled, range from 0 to infinity

The *ratio* of the change in elevation to the actual distance traveled, range from 0 to 1

The *angle* between the surface and the horizontal, range 0 to 90

Slope calculation example

- $b = \frac{Z3+2*Z6+Z9-Z1-2*Z4-Z7}{8D}$
- $c = \frac{Z1+2*Z2+Z3-Z7-2*Z8-Z9}{8D}$
- $tan(slope) = \sqrt{b^2 + c^2}$
- b measures the elevation change in the direction of the x axis
- c measures the elevation change in the direction of the y axis
- The values b and c can also be used to calculate the aspect (direction of the steepest slope)



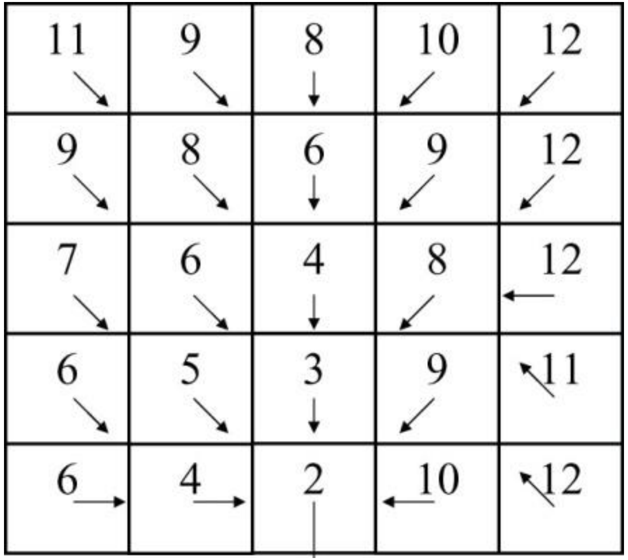
- Note that Z5 is not used in the analysis at all

Surface analysis example: watershed

“A watershed is the area of land where all of the water that is under it or drains off of it goes into the same place. “

- **Example: forming a catchment area from DEM (see also previous slide)**
  - Direction of flow for each cell: lowest elevation among the eight neighbours (aspect of the slope)
  - If the cell in question is lower than any of its neighbours, water will pool to the cell

A) Elevation turned to aspect, aspect used to infer water flow direction (8-neighborhood)

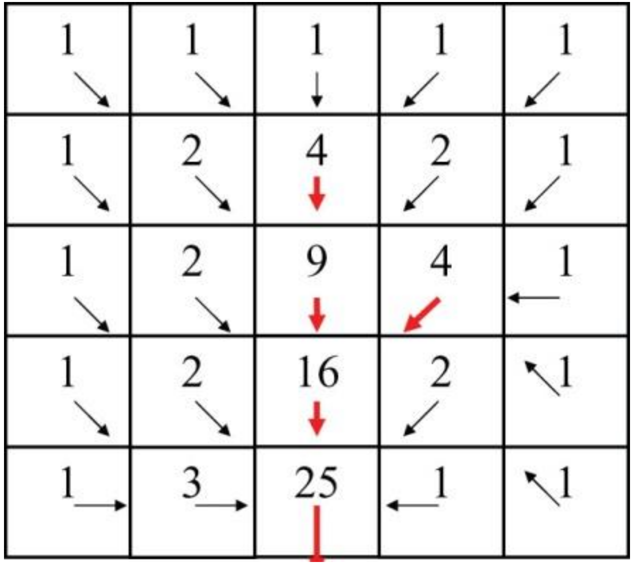


Surface analysis example: watershed

- **Cumulative amount of water for each cell**
- **Recognition of watersheds, catchment areas and channels**

- **In practice, modelling the behaviour of water is more complex**
  - Some portion of the rainfall is absorbed to the ground
  - Soil type has an effect to the flow

B) accumulated flows in each cell and eroded channels based on a threshold flow of 4 units

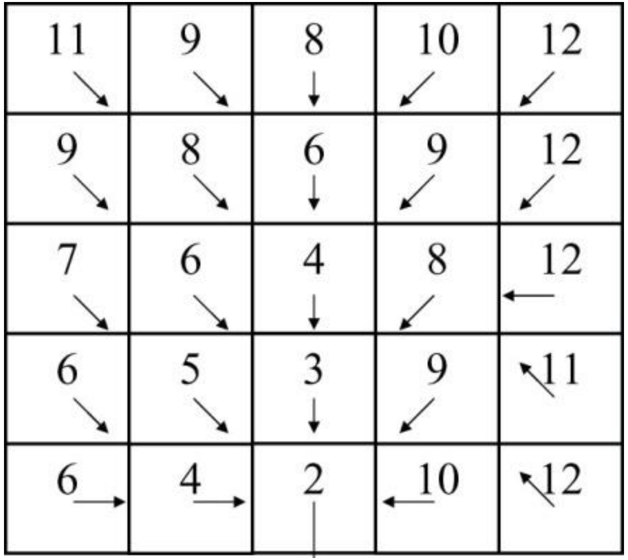


Elevation to aspect to watershed

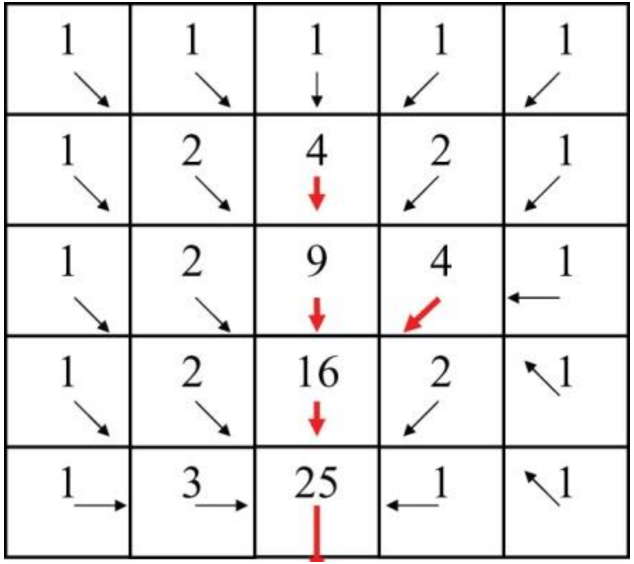
A) Elevation turned to aspect, aspect used to infer water flow direction (8-neighborhood)

B) accumulated flows in each cell and eroded channels based on a threshold flow of 4 units

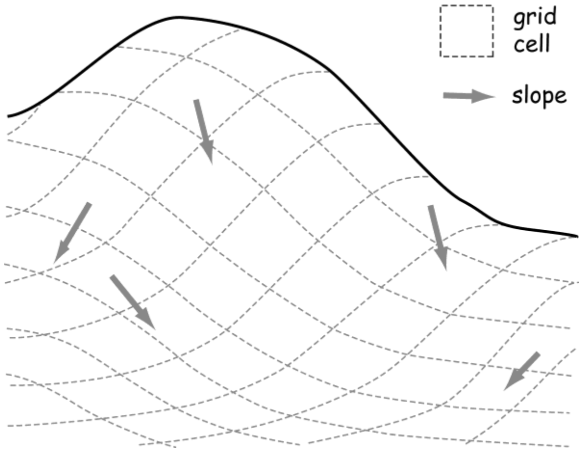
A

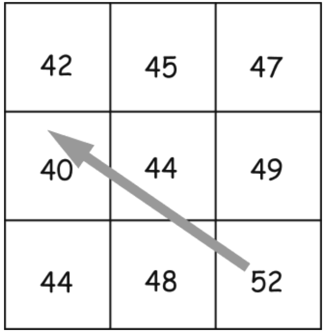


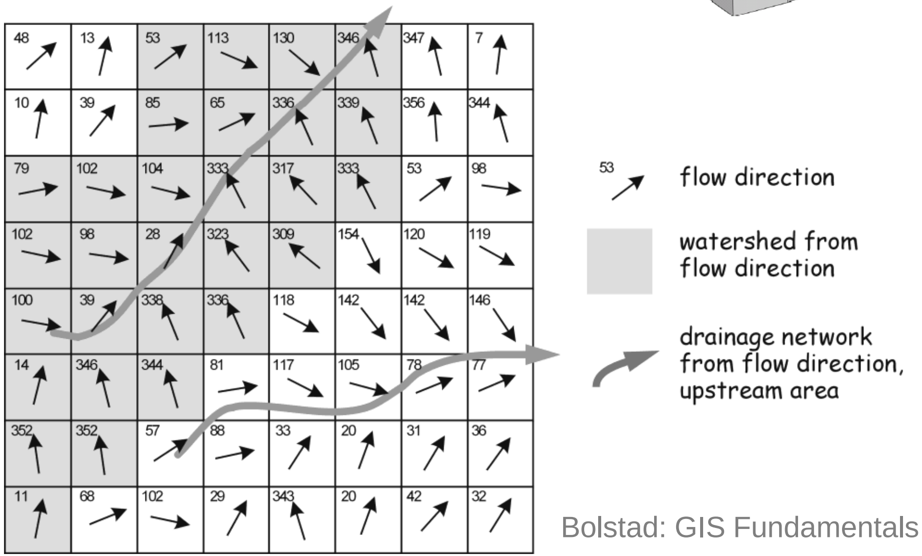
B

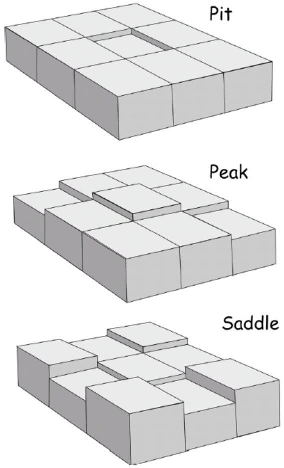


Analyzing the watershed









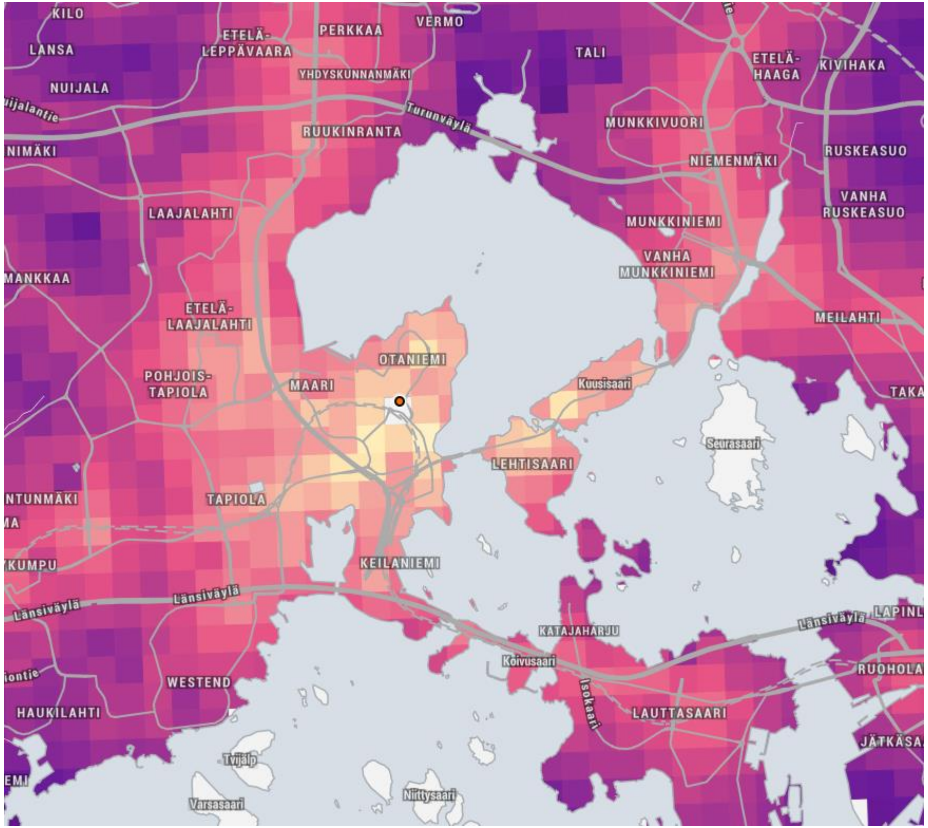
Bolstad: GIS Fundamentals

Cost surface

- **Route on cost surface**
  - Each cell has determined cost, based on for example:
    - the density of vegetation
    - environmental impacts
    - depth of snow
    - Availability of roads
- **Can be to solve how to travel from starting point to end point with minimum cost**

OR

- **To all locations from the starting point**
  - E.g. travel time maps

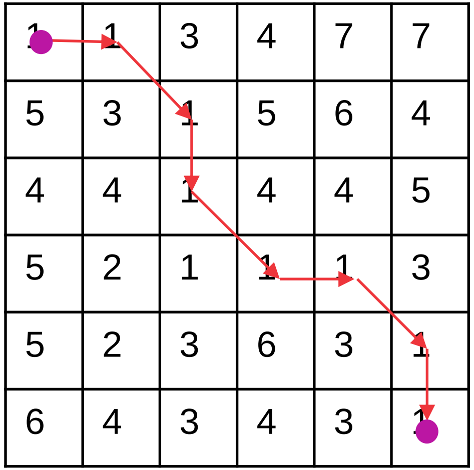


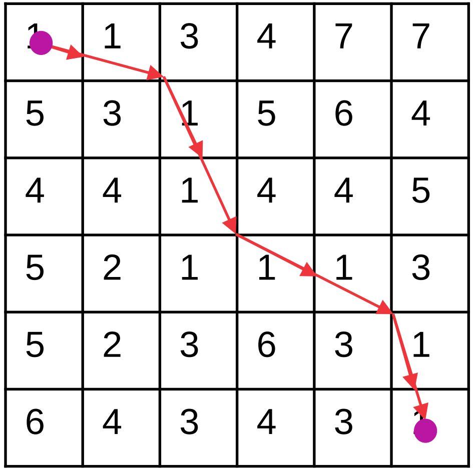
Travel time map

Source: mapple.fi

Cost surfaces

- **The value of each pixel tells how much effort it takes to move through it**
- **The cost of moving from a pixel to its neighbour depends on:**
  - The value of the pixel, its neighbour, or both
  - Whether the neighbour has adjacent edge or adjacent corner with the starting pixel
- **In many models one step is from the *center of one pixel to the center of an adjacent one***





Pixel center to pixel center movement is not very realistic

Cost surfaces with larger neighborhoods

- One solution is to increase the size of pixel neighborhood (number of connections between pixels)

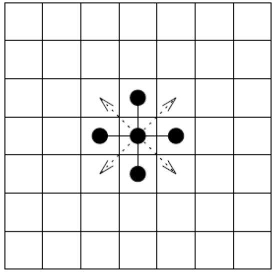


Fig. 4: 8-connected.

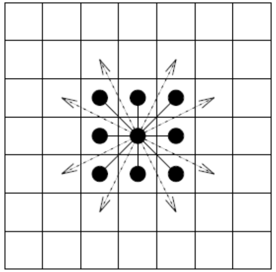


Fig. 5: 16-connected.

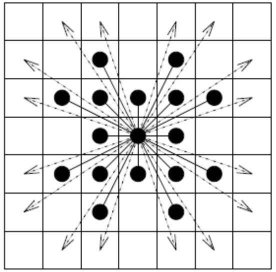


Fig. 6: 32-connected.

- However, more connections increases the effective size of the problem

connect	Maximum Elongation ϵ	Maximum Deviation δ
4	1.41421	0.50000
8	1.08239	0.20710
16	1.02749	0.11803
32	1.01308	0.08114
64	1.00755	0.06155
128	1.00489	0.04951

Fig. 10: Maximum elongation and maximum deviation errors.

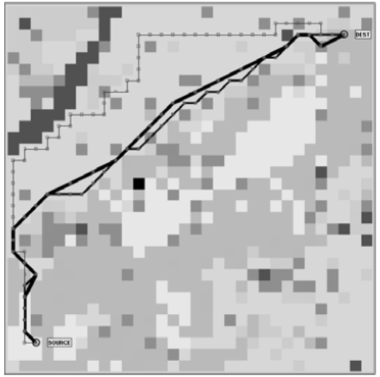


Fig. 11: 4-, 8-, and 16-connected paths.

Aalto University
School of Engineering

Image source: Van Bemmelen et al.

Cost surfaces with movement between locations at pixel borders

- Pixel center to pixel center is conceptually a simple way to model movement
- An alternative is to move from a position at pixel edge to another position at pixel edge
- With limited possible moves from each position, the cost of each move can be calculated easily

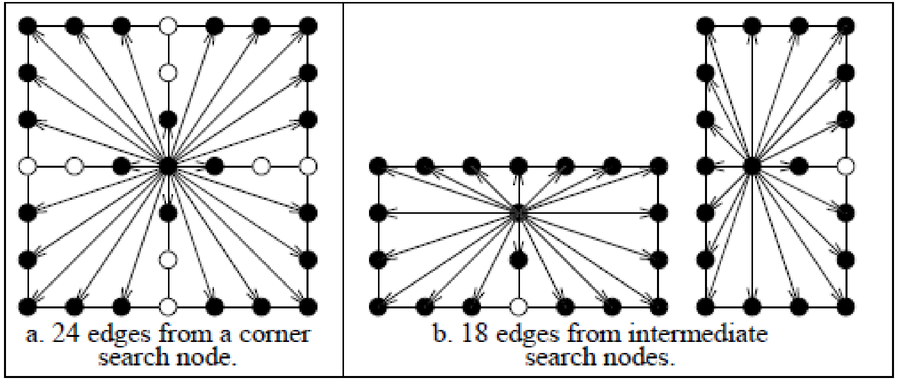


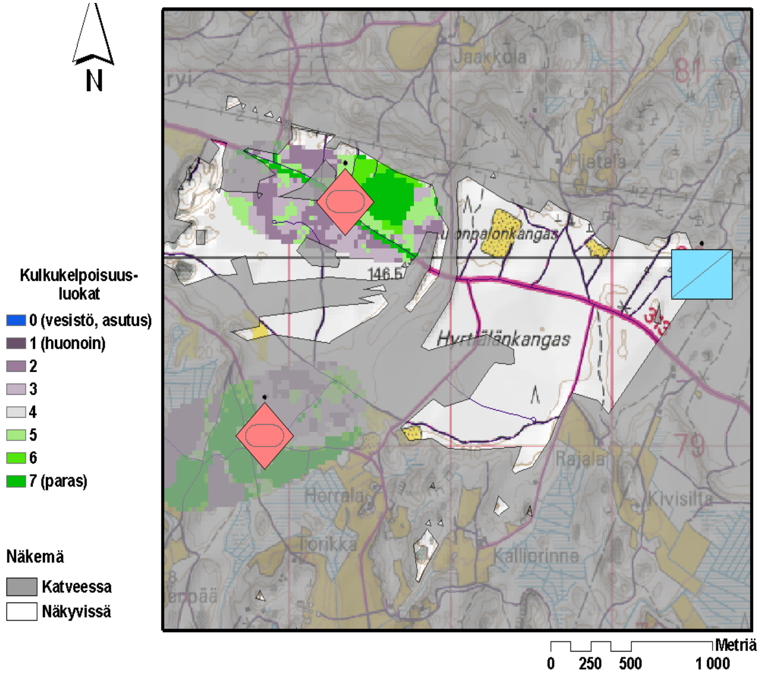
Fig. 13: 2 intermediate nodes.

Aalto University
School of Engineering

Image source: Van Bemmelen et al.

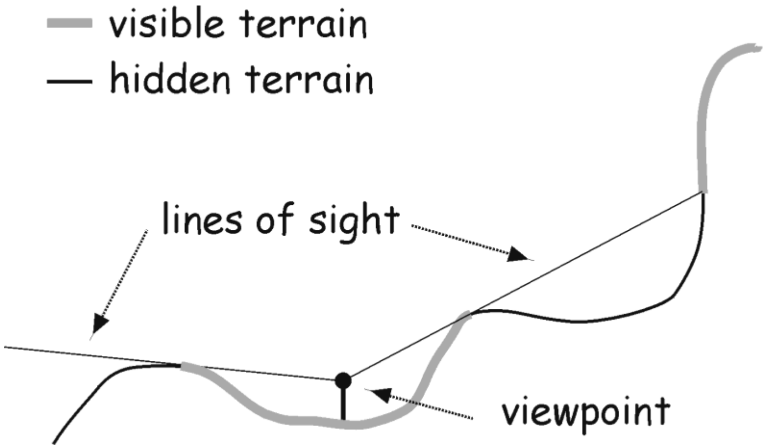
Viewshed

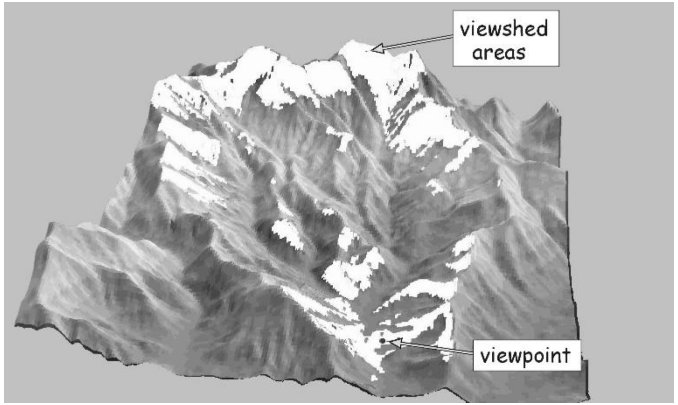
- "A viewshed is an area of land, water, or other environmental element that is visible to the human eye from a fixed vantage point. The term is used widely in such areas as urban planning, archaeology, and military science."
- **Scale-dependent: Elevation model, are buildings and trees included?**
- **Analysis always for only one vantage point at the time**
  - Results can be merged with overlay analysis



Aalto University
School of Engineering

Viewshed examples





Bolstad: GIS Fundamentals

Map algebra: formal language for raster operations

Map algebra basics

- **Created as an algebra for spatial data (stored in raster format)**
- **Data stored in layers**
  - Each layer contains different type of data
- **Map algebra operation takes as input a number of existing raster layers**
  - Mathematical (arithmetic and/or logical) operations done based on the layer data
- **As output it produces a new raster layer**

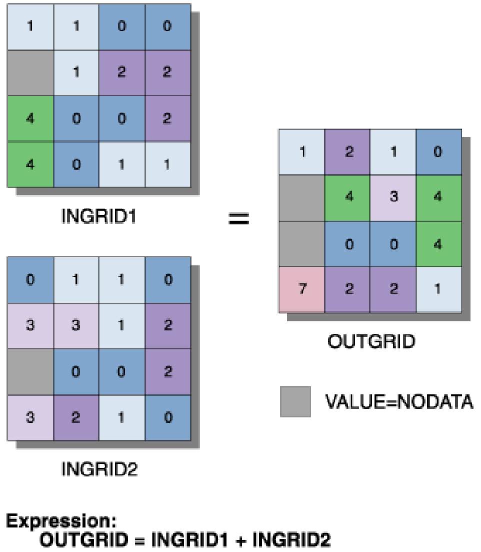
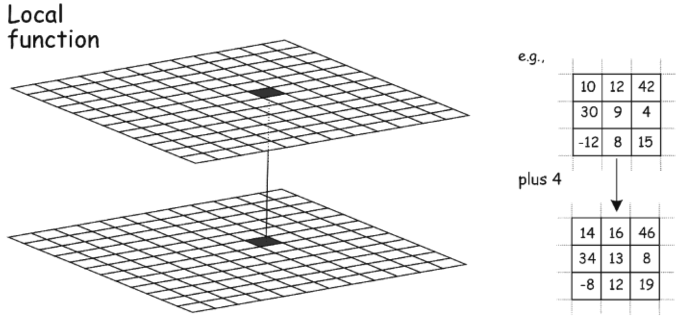


Figure by ESRI

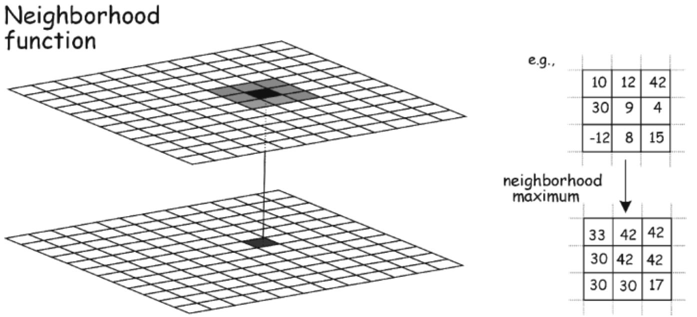
Map algebra basics

- **One or more input layers**
- **One output layer**
- **In map algebra there are four different types of functions**
  - Local, Focal, Zonal, Incremental and Global functions
- **Functions can define e.g. arithmetic or statistical operations**
  - Sum, difference, min, max, etc.
- **Originally proposed by Dr. Dana Tomlin in the 1980s**

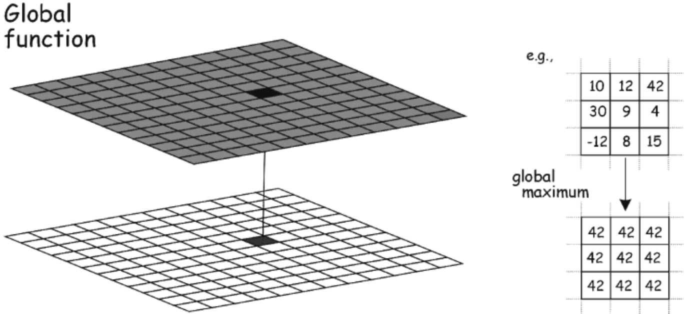
LocalSum



FocalMax



GlobalMax

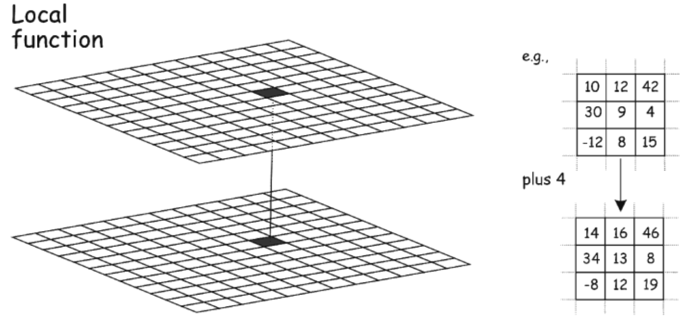


Aalto University
School of Engineering

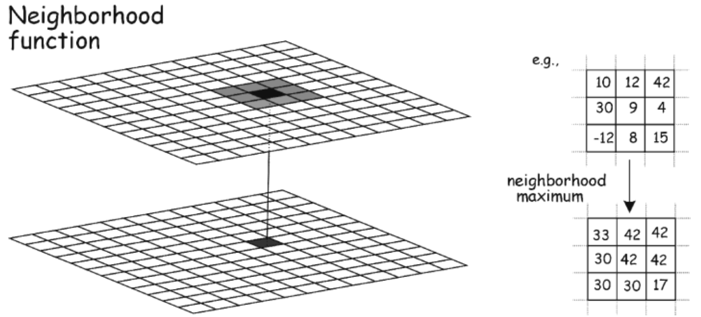
Map Algebra function types

- **Four types of raster operations:**
  - Local operations are determined by the attributes of each cell alone
  - Focal operations are determined by a cell and its neighbors
  - Zonal operations apply to cells with the same value (depicting a zone)
  - Incremental operations increment linear or areal values according to a condition (e.g. computing drainage)
  - (Global operations compute properties for the entire raster layer)
    - Effectively a zonal operation with just 1 zone

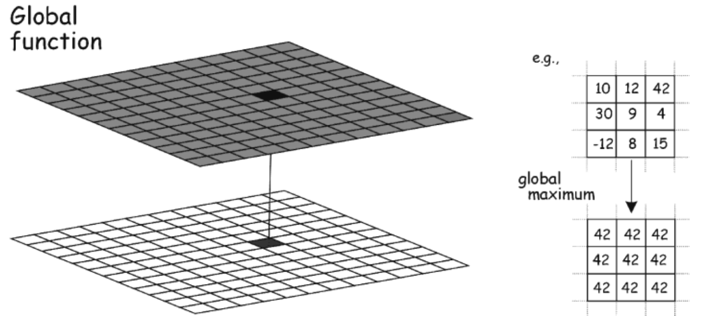
LocalSum



FocalMax



GlobalMax



Aalto University
School of Engineering

Local functions

- **Operations on one pixel at a time from one or more layers, e.g.**
  - LocalDifference
  - LocalMaximum
  - LocalRatio
  - LocalProduct
- **For example,**
  - AverageCost = LocalMean(YourCost and MyCost and HisCost)
  - Comparison of layers: either-or, cf. map overlay
  - From elevation model to contours (in Tomlin's book)

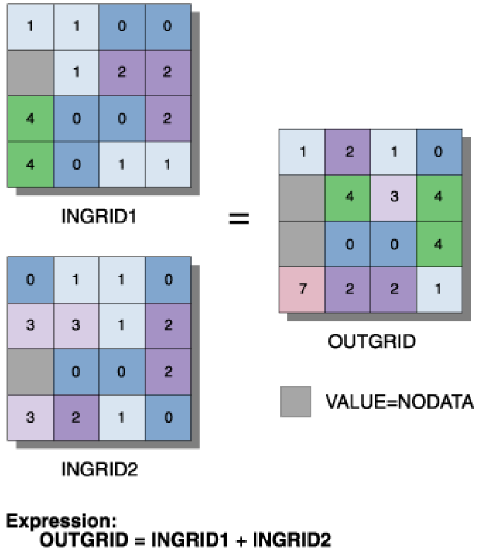
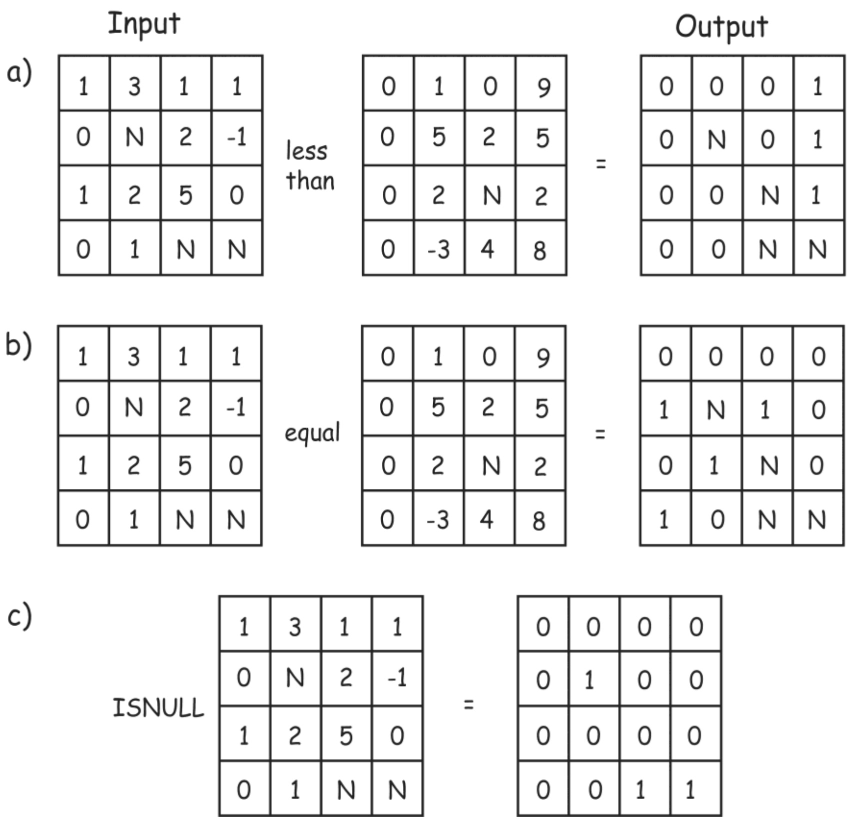
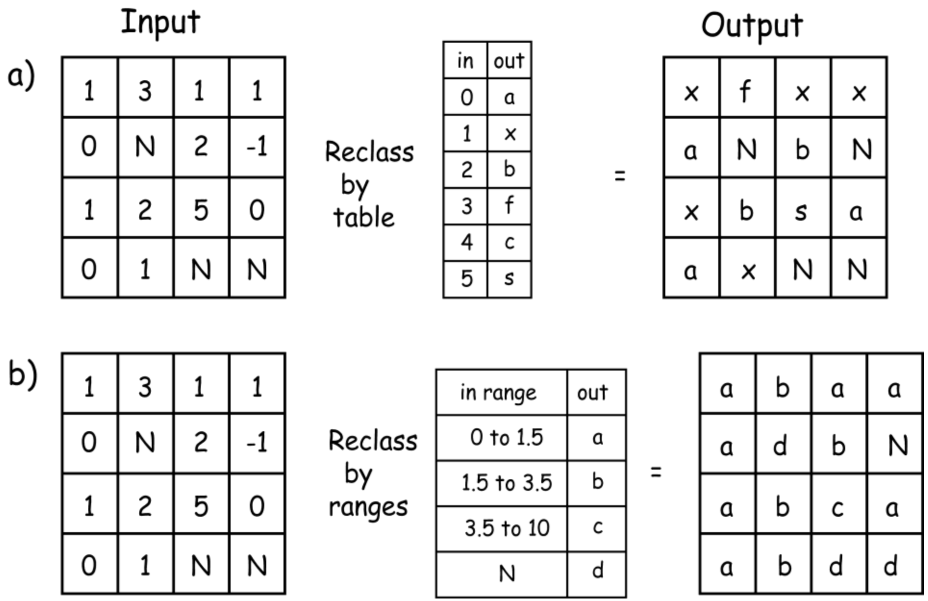


Figure by ESRI

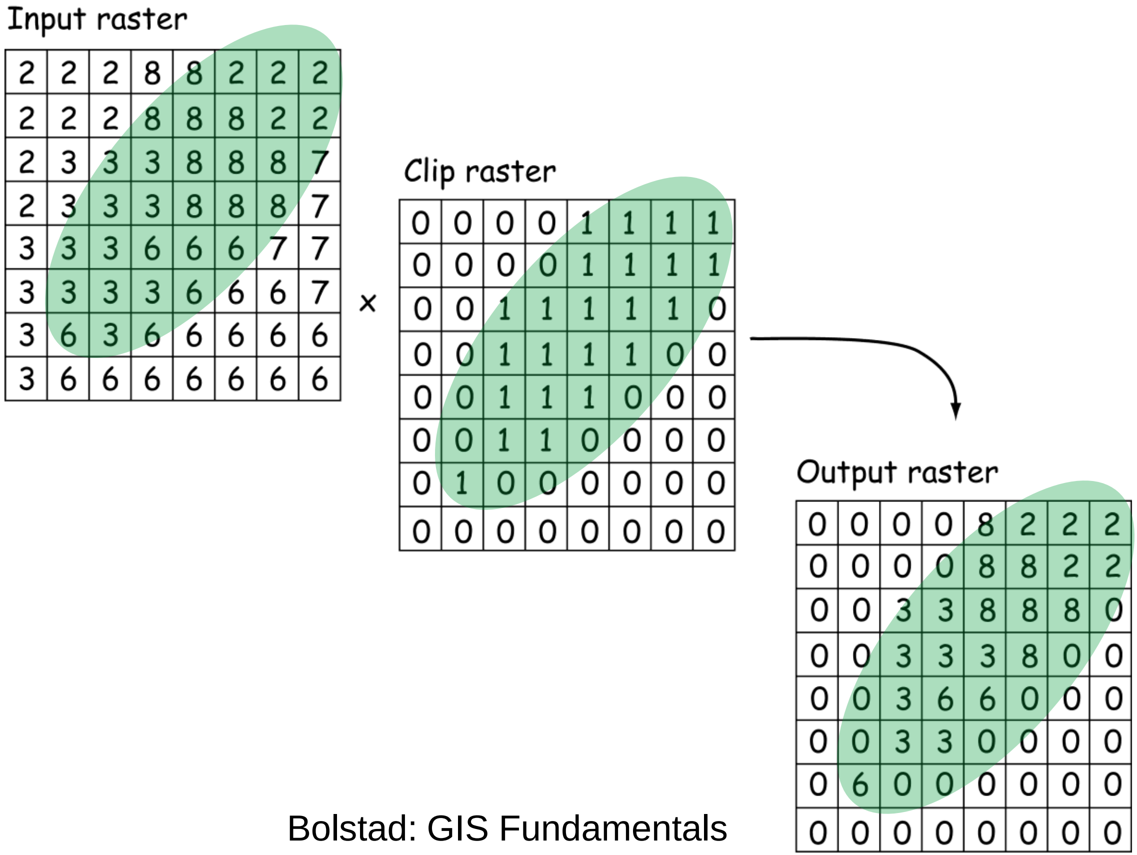
Local function examples





Note the null values
Source: Bolstad: GIS Fundamentals

Map algebra example: clipping



Bolstad: GIS Fundamentals

Map algebra classroom exercise

(

5	3	2
4	4	2
6	3	1

> 3)

AND

(

1	2	3
0	2	3
2	2	2

< 2)

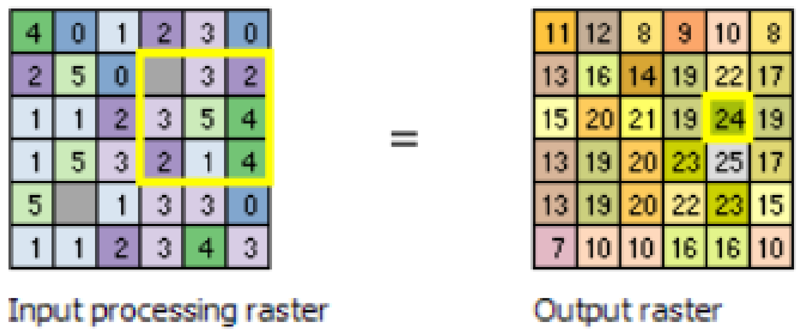
=

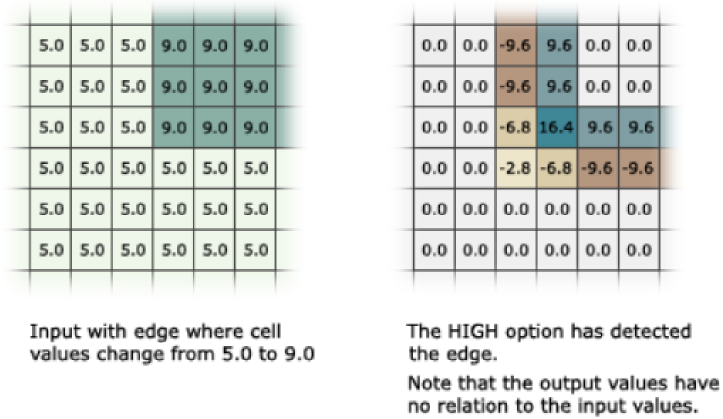
1	0	0
1	0	0
0	0	0

Focal functions

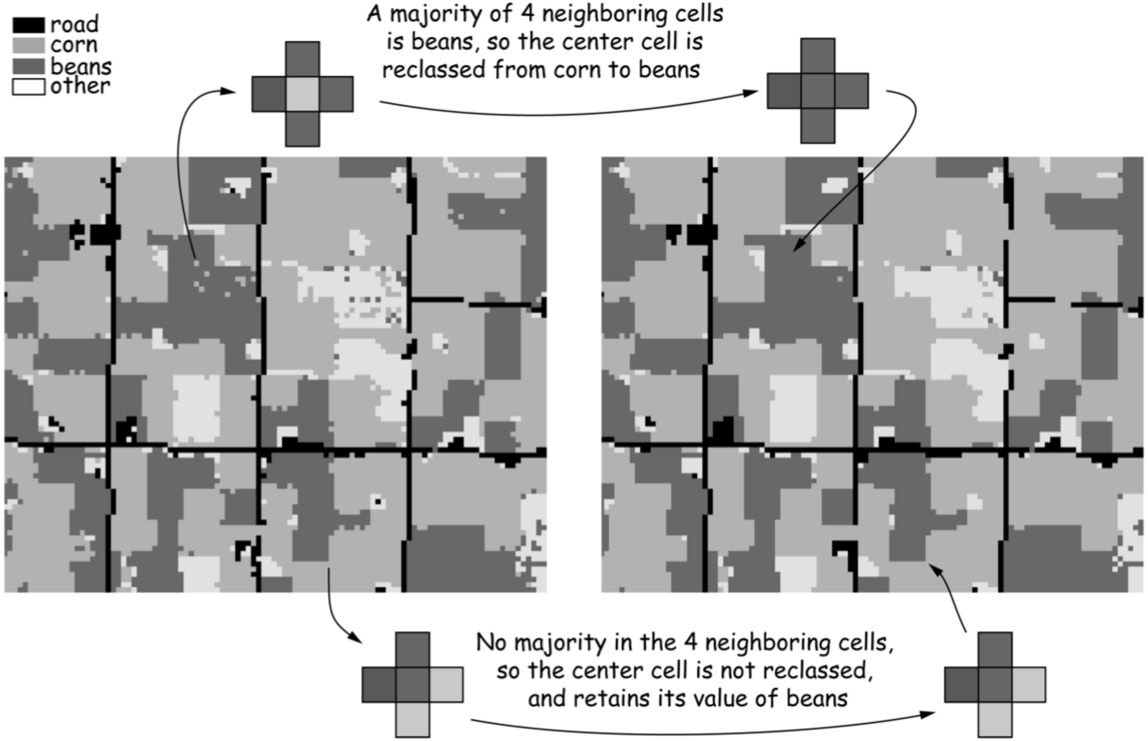
- **Operations on a pixel and its 4 (or 8, or 32, or…) neighbouring pixels, e.g.**
    - FocalMaximum
    - FocalMean
    - FocalRating

- **For example,**
    - Smoothing the map layer by FocalAverage; high and low values are smoothed away (continuous values)
    - Generalization by filtering out individual pixels that differ from their surrounding (categorical values)
    - Edge detection (where values change)





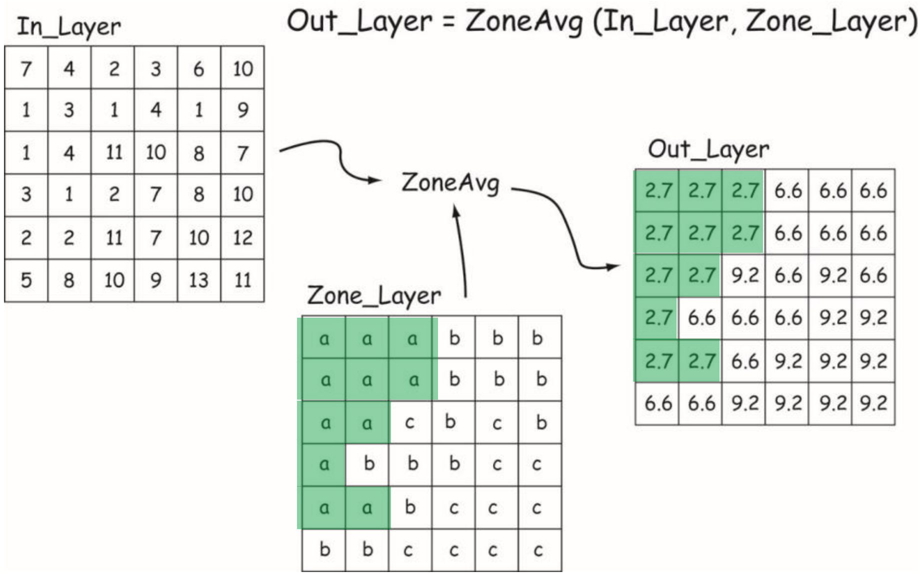
Focal function example: filtering with 4-neighborhood



Bolstad: GIS Fundamentals

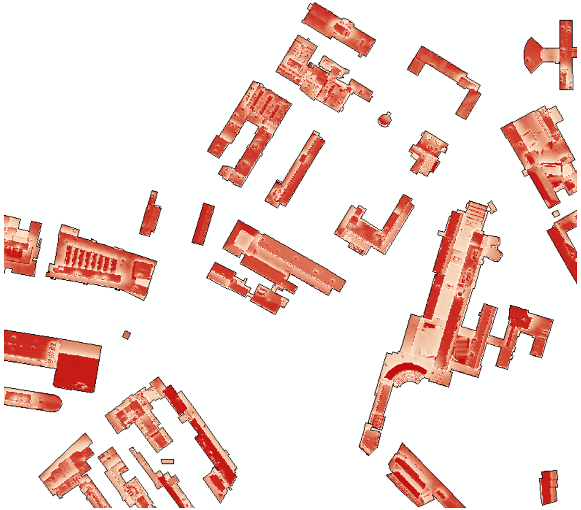
Zonal functions

- **Data is related to zones**
- **Zoning given on one layer, other data in the other layers**
- **For example,**
  - ZonalSum
  - ZonalMaximum
  - ZonalAverage
- **E.g. calculating how many houses in each block**
  - Blocks first formed by “cutting” the area by a road layer (a focal function). Then zonal functions use the blocks:
    - Number of houses -> ZonalSum (value = 1 in each house)
    - Number of residents -> ZonalSum (value = num. of residents in a house)
    - Biggest house of a block -> ZonalMaximum
    - Average size of houses in a block -> Zonal Average

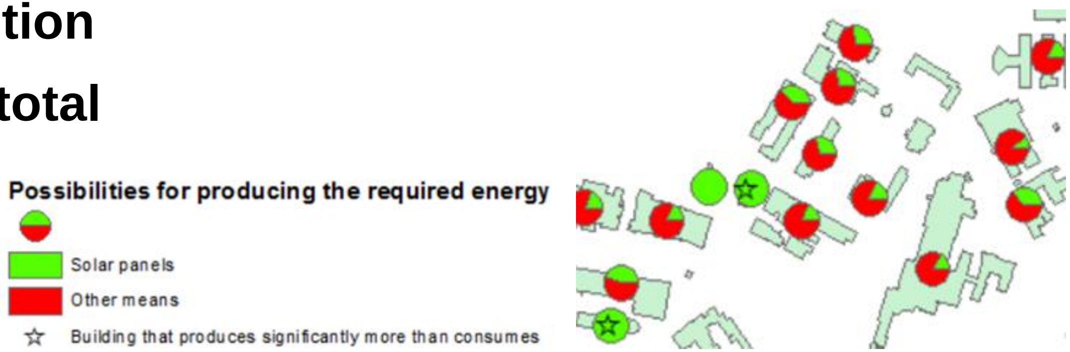


Bolstad: GIS Fundamentals

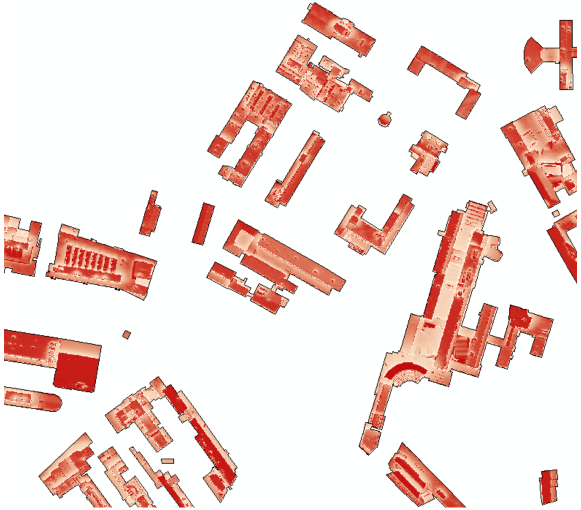
Zonal function example



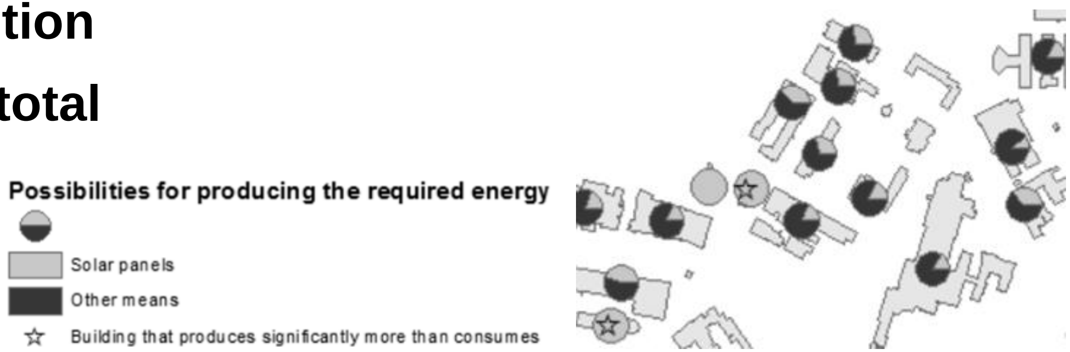
- Energy self-sufficiency in Otaniemi: how big portion of energy consumption can be gathered using solar panels on top of roofs
- Data: Solar energy data, buildings, energy consumption
- Analysis: ZonalSum to get total energy per building



Zonal function example



- Energy self-sufficiency in Otaniemi: how big portion of energy consumption can be gathered using solar panels on top of roofs
- Data: Solar energy data, buildings, energy consumption
- Analysis: ZonalSum to get total energy per building



Reading for the lecture

- **O’Sullivan & Unwin (2010): Chapter 9.5 Map Algebra (pp. 270-273)**
- **http://gisgeography.com/map-algebra-global-zonal-focal-local/**
- **Longley et al. (2015): Chapter 14.3 Analysis of Surfaces**