Aalto University
School of Engineering

Geoinformation in Environmental Modelling

Spatial analysis and modelling: grid data

ENY-C2005
Jussi Nikander
1.2.2019
Slides by Paula Ahonen-Rainio and Jaakko Madetoja

Topics today

- About spatial analysis and modelling
  - Various viewpoints for categorizing
- Basics of data modelling
  - Raster topology (back to L-3)
  - Basic concepts of grid data
- Typical tasks with grid data
  - Map overlay – one of the basic methods
  - Watershed, viewshed and cost surface analysis
- "Map algebra" – a set of analysis functions
  - Local, focal, zonal, incremental, and global functions

Examples of potential exam questions relating to this lecture

- Anna esimerkkejä pintojen, kuten korkeusmallin (DEM), spatiaalisesta analyysista.
- Give examples of spatial analysis on surfaces, such as digital elevation model (DEM).

- Selitä mitä tarkoittaa kartta-algebran lokaali/fokaali/zonaali operaatio. Anna esimerkkejä lokaaleista/fokaaleista/zonaaleista operaatioista ja konkreettisista sovelluksista, joissa niitä voidaan käyttää.
- 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.

- Suorita [annettu kartta-algebran operaatio] [annetuille 1…3] rasteritasoille.
- Perform [a given map algebra operation] on the [1...3 given] raster layers.

Spatial analysis

- "A set of methods whose results change when the locations of the objects being analysed change"
  - or when the behaviour (i.e. values in various locations) of a spatially continuous phenomenon changes
  - Note: "location" can be absolute (metric relationships) or relative (topological relationships)
- The process of turning raw data into useful information
  - e.g. for scientific discovery, or effective decision making
  - From hypothesis to confirmation or rejection and then to conclusion (often recursive)
- Collaboration of computer and human user (computational and visual analysis)

Examples of spatial analysis

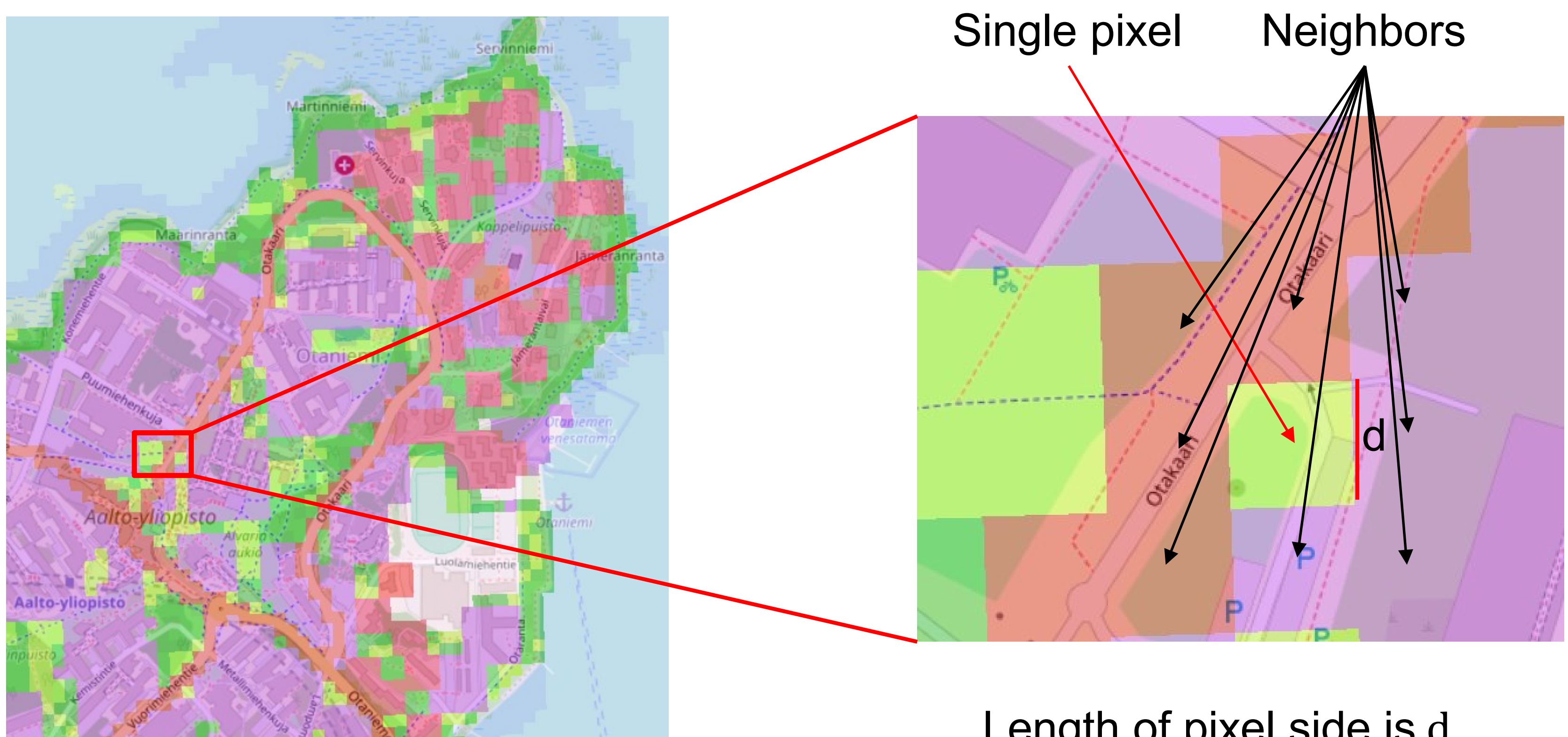
- Analysis of a grid (raster, matrix)
- Analysis of a graph
- Spatio-statistical analysis, basis on statistics

Representation of fields

- Data (incl. field data) can be measured on
  - Numerical scale
    - values measured on ratio, interval, or cyclic scales
    - e.g. height, elevation, temperature, humidity, wind direction
  - Categorical scale
    - values measured on nominal or ordinal scales
    - e.g. soil types, land use, land ownership

- Spatially, a field can be represented e.g. by
  - Irregular or regular point set (possibility to interpolate values between them)
  - Triangulated irregular network, TIN
  - Contours, isolines based on point set (only for visualisation)
  - Polygon network
  - **Grid** (raster)
    - a regular tessellation, i.e. geometry is independent of the variation of the phenomenon

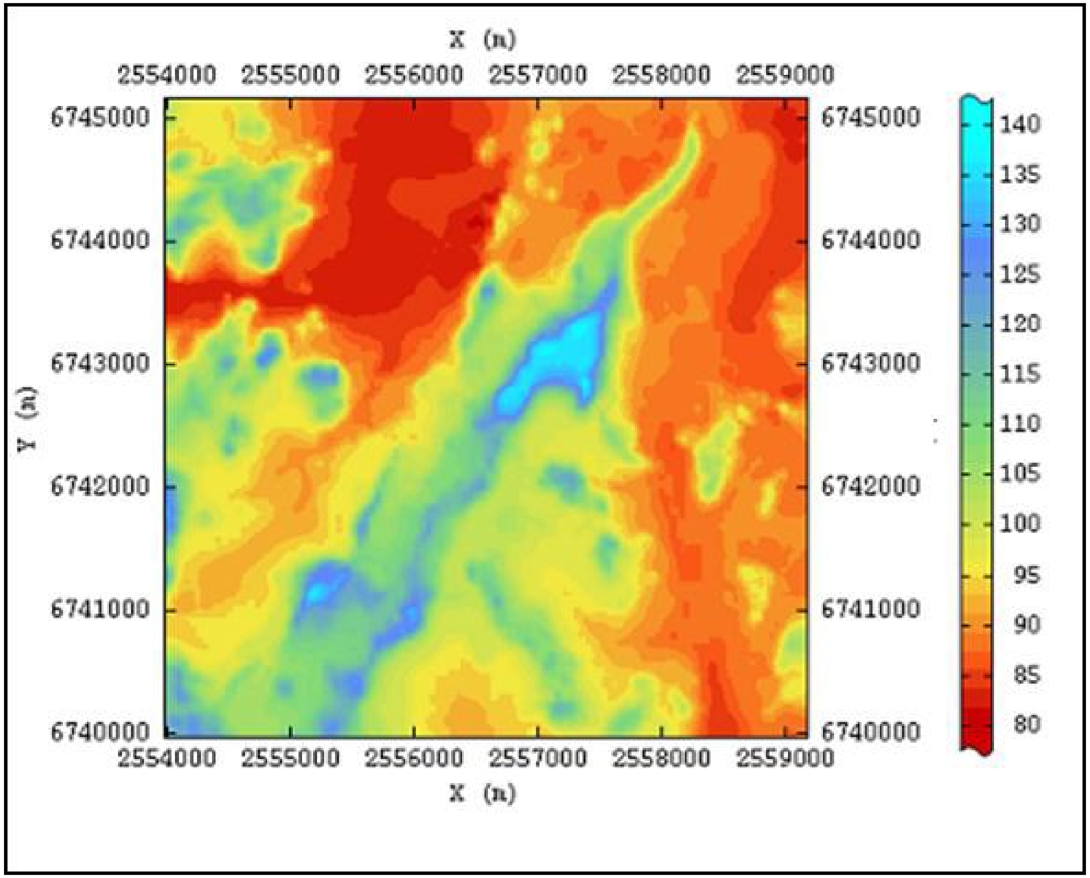
Example: Grid data



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

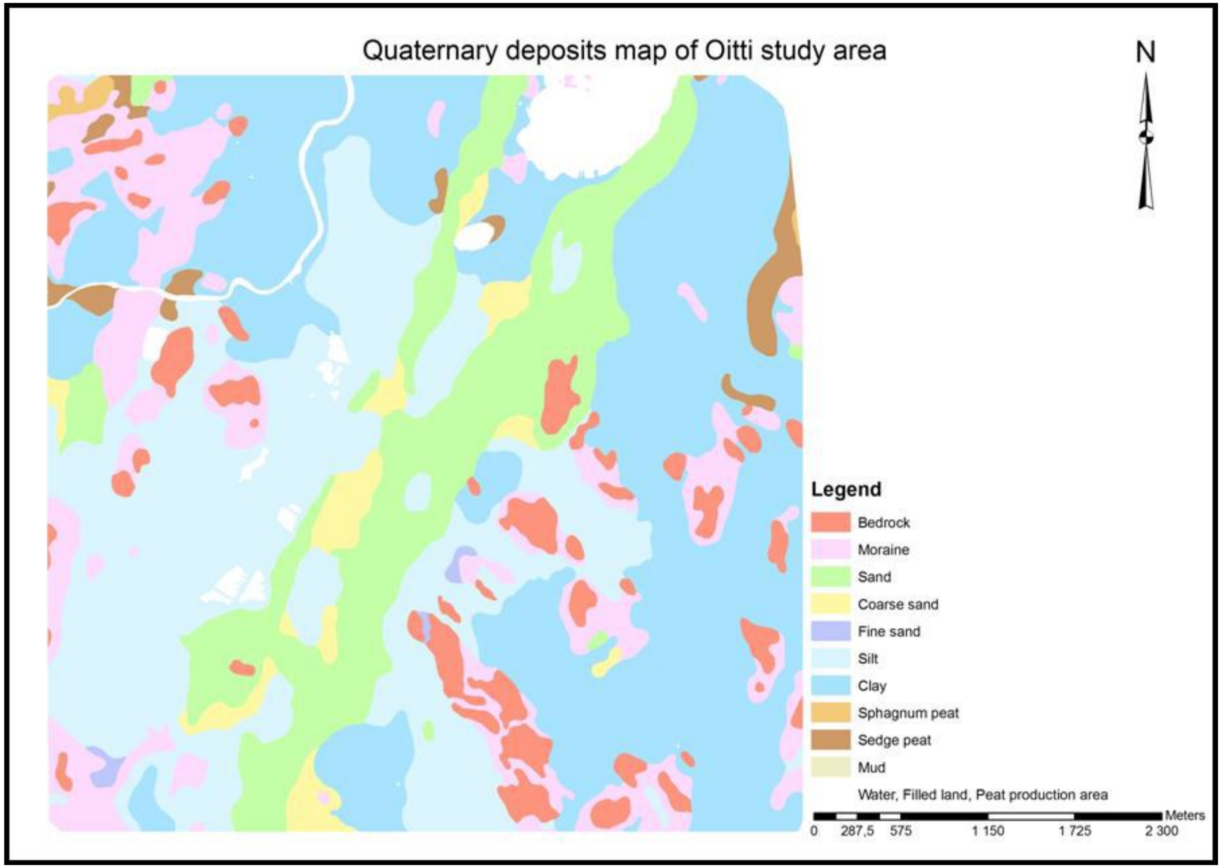
Length of pixel side is d
pixel area is d^2
Here $d = 10m$

Example: Numerical data



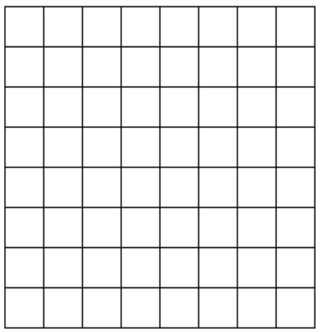
R. Sunila 2009

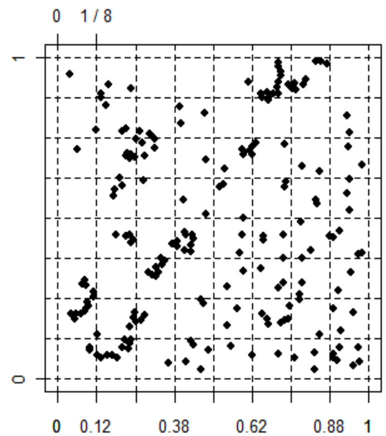
Example: categorical data



R. Sunila 2009

Example:
transformation of a point set to a grid > ratio data



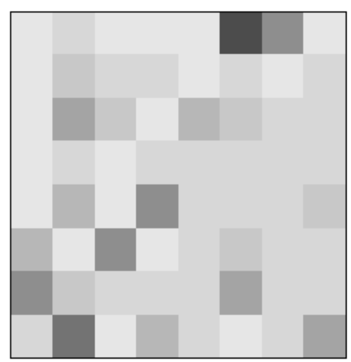


Pistemäiset kohteet voidaan muuntaa pintamaiseksi esitykseksi laskemalla pisteiden määrä annetun hilan ruuduissa

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

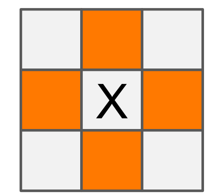
Visual ization



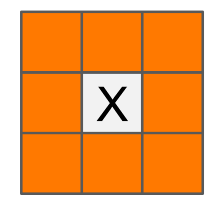
Point objects can be transformed into a field model by counting the sum of points in each cell of the given grid

Grid representation

- 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)

Grid data management

- One layer (taso) presents one theme (or variable)
- Layers are stored separately
- If stored in files, raster layers can be compressed, e.g.,
  - RLE run length encoding
    - The simplest compression method
    - Good result when relatively large homogeneous areas
  - Quadtree (nelipuu)
    - One type of block encoding
    - In addition to compression, also used for indexing
  - These are lossless methods: all information is preserved
    - Spatial data typically should not use lossy compression

Aalto University
School of Engineering

Spatial analysis of grid data

"Raster analysis"

Typical examples of grid analysis:

Overlay analysis

Watershed, Viewshed, Cost surface

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
    - Neighborhood of each pixel (grid cell)
    - Separately defined areas
    - Other means of handling the raster layer(s) (e.g. using a line between two pixels)
- Output of raster analysis is a new raster layer (or layers)

1	1	0
0	1	1
0	0	1

&

0	0	0
1	1	0
1	1	1

=

0	0	0
0	1	0
0	0	1

Foc.avg

3	5	2
7	6	3
8	5	3

=

	4,7	

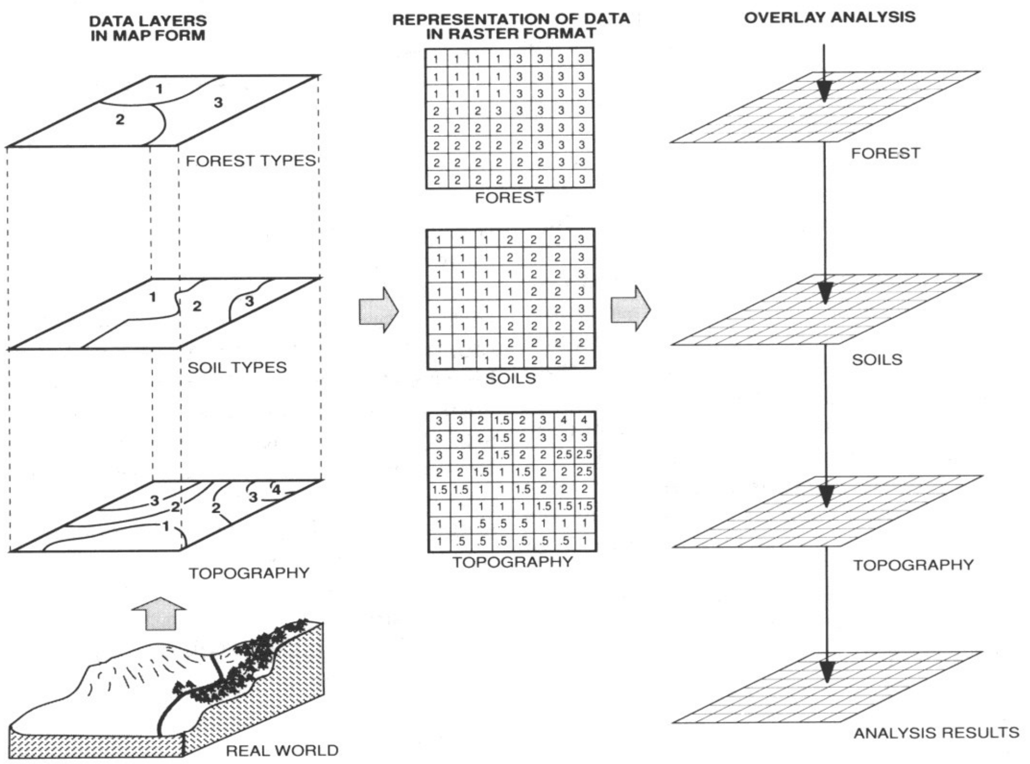
Examples of applications

- 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

Overlay analysis

- History of map overlay
  - (paper) transparencies on top of each other, analysis on the basis of several map layers
- A new data layer as a function of two or more source data layers
- Can be performed to any geometric data type
  - Straightforward computation with grid data (if the resolution and orientation is the same)
  - More complex with vector data (points, lines, polygons), e.g. slivers in polygon overlay
  - Data types must be compatible! -> transformation
- One of the basic operations “at the backstage” in GIS

Raster overlay



Longley et al. 2015

About the overlay analysis

- Basic form based on binary logic $\Rightarrow$ Boolean logic
  - On each layer, possible values for each cell are 1 or 0
  - 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
  - Cf. Assignment A-1

- Potential problems because of inconsistencies
  - Input data may originally be in different scales or in different coordinate systems $\Rightarrow$ 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)

Analysis of surfaces

A typical data set for analysis: Digital elevation model (DEM)

- The same methods are applicable for any kind of continuous surface with interval- or ratio-type variable
- the meaningfulness of the method for the phenomenon in question must be understood

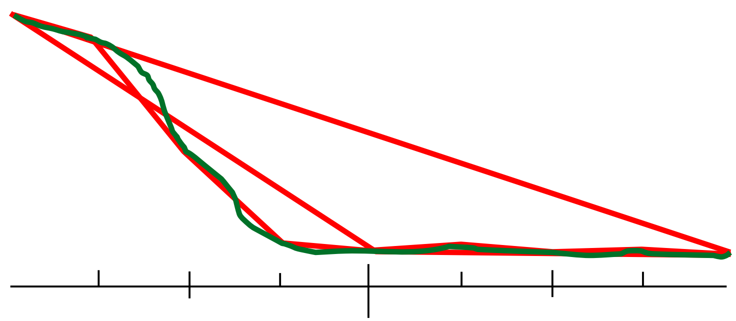
- Key characteristics of DEM: slope, aspect (kaltevuus ja viettosuunta)

Aim of the analysis, e.g.

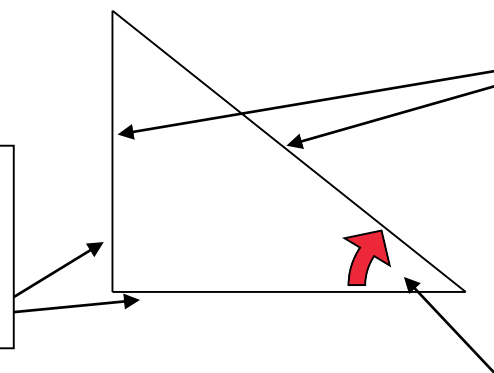
- Catchment and watershed (valuma- ja kokooma-alueet)
- Viewshed, visibility (näkemä)
- [Route over] cost-surface (Reitti kustannuspinnan yli)

Slope (kaltevuus)

- Strongly dependent on scale, various expressions



- There are three alternative definitions of slope



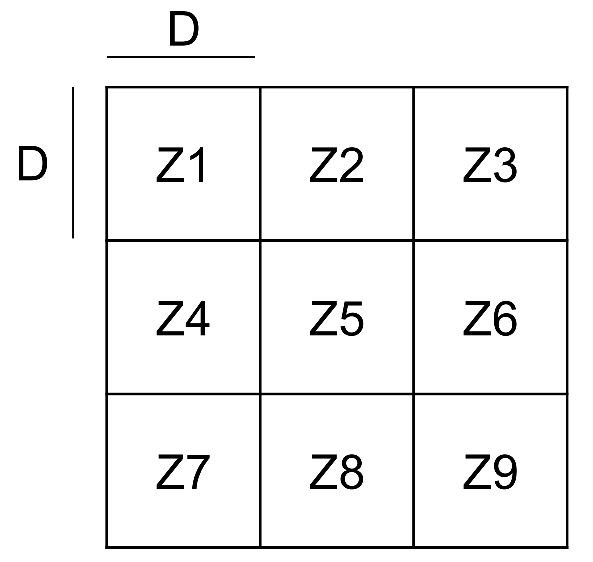
The *ratio* of the change in elevation to the actual distance traveled, range from 0 to 1

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

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)

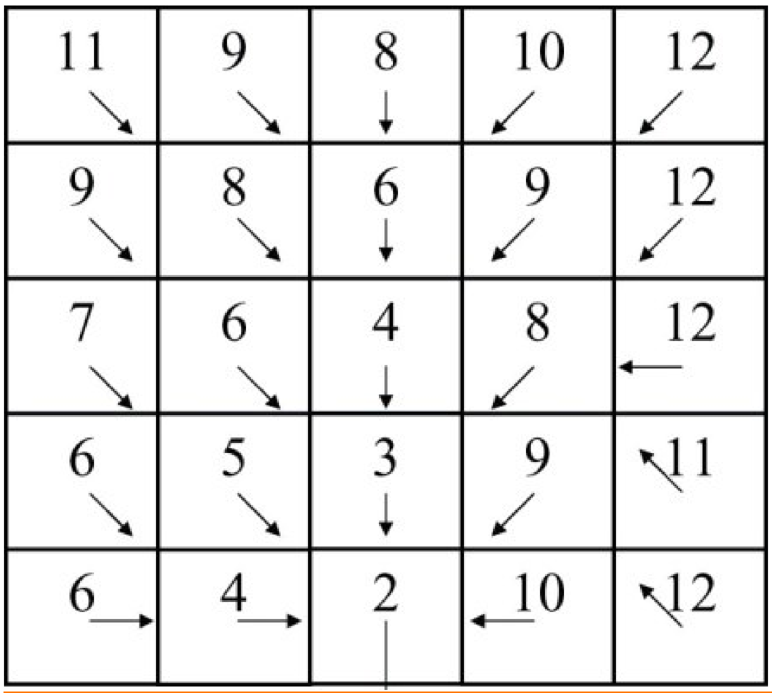


Example: hydrologic analysis of a DEM: watershed

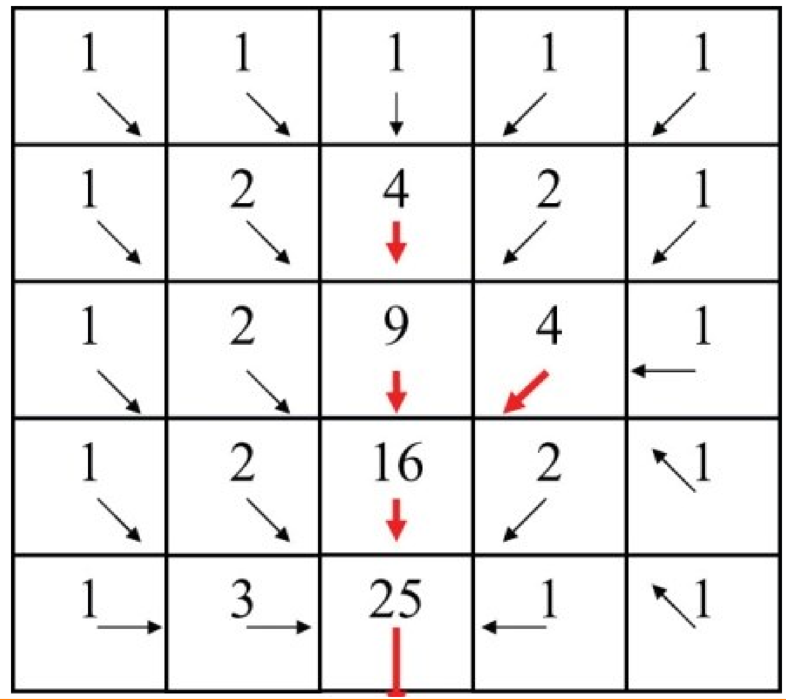
(A) the DEM and inferred flow directions using the rook's case move set

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

A



B



… example: hydrologic analysis

"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. "

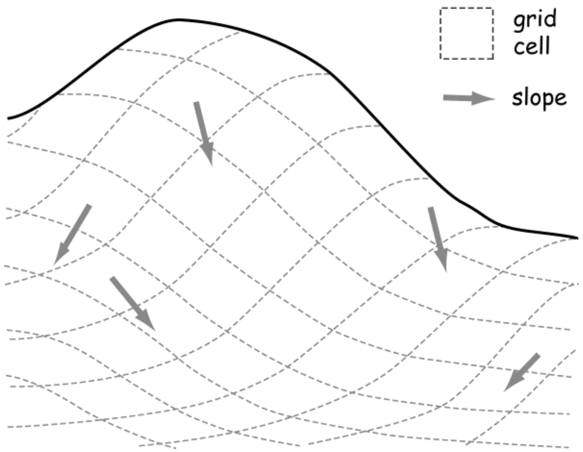
Example: forming a catchment area from DEM (previous slide)

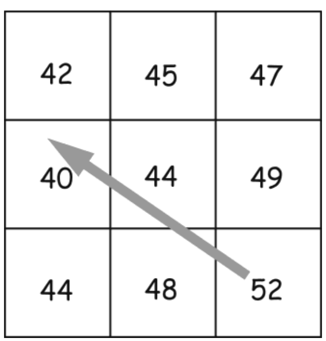
- Direction of flow for each cell: lowest elevation among the eight neighbours
  - If the cell in question is lower than any of its neighbours, water will pool to that cell
- Cumulative amount of water for each cell
- Recognition of watersheds, catchment areas and channels

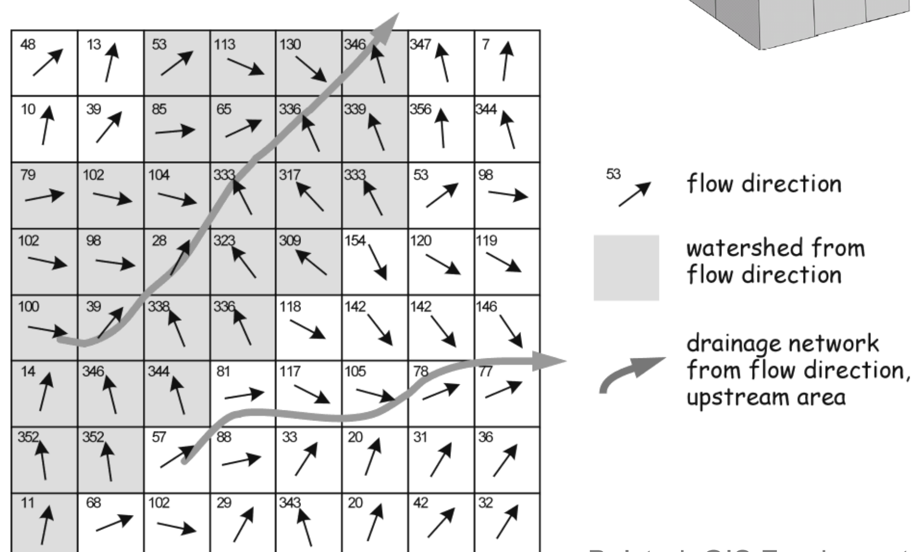
In practice, modelling the behaviour of water is more complex

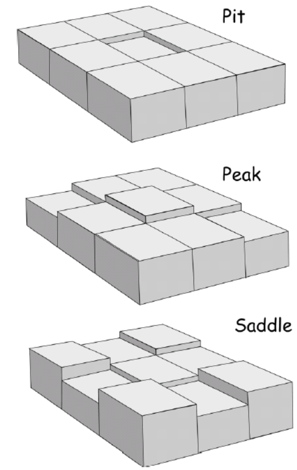
- Some portion of the rainfall is absorbed to the ground
- Soil type has an effect to the flow

Example: Analysing the watershed





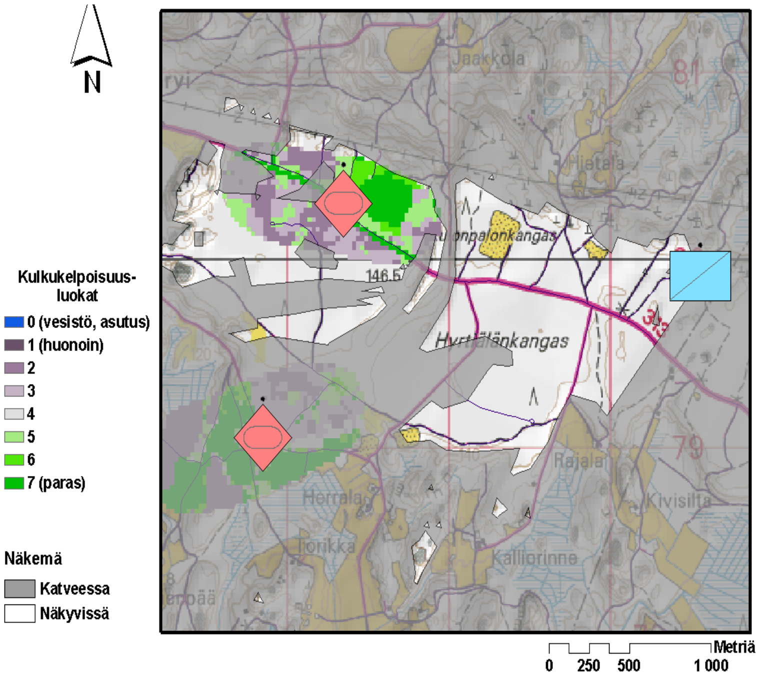




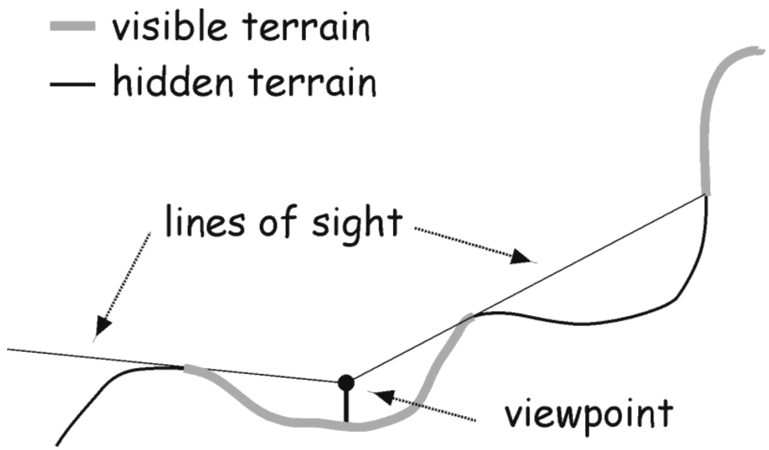
Bolstad: GIS Fundamentals

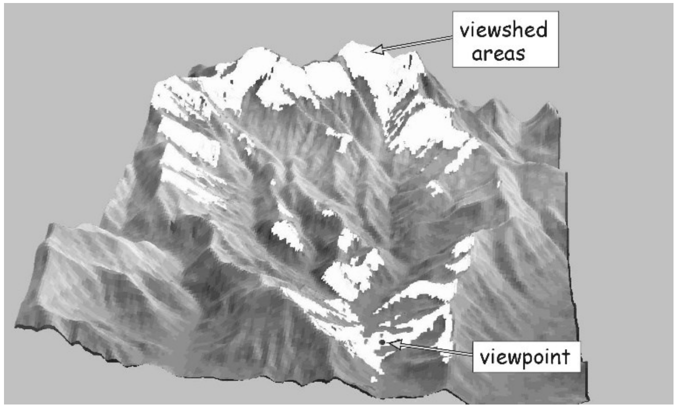
Viewshed

- “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, or even buildings and trees?
- Analysis always for only one vantage point at the time
  - Results can be merged with overlay analysis



Example: Viewshed



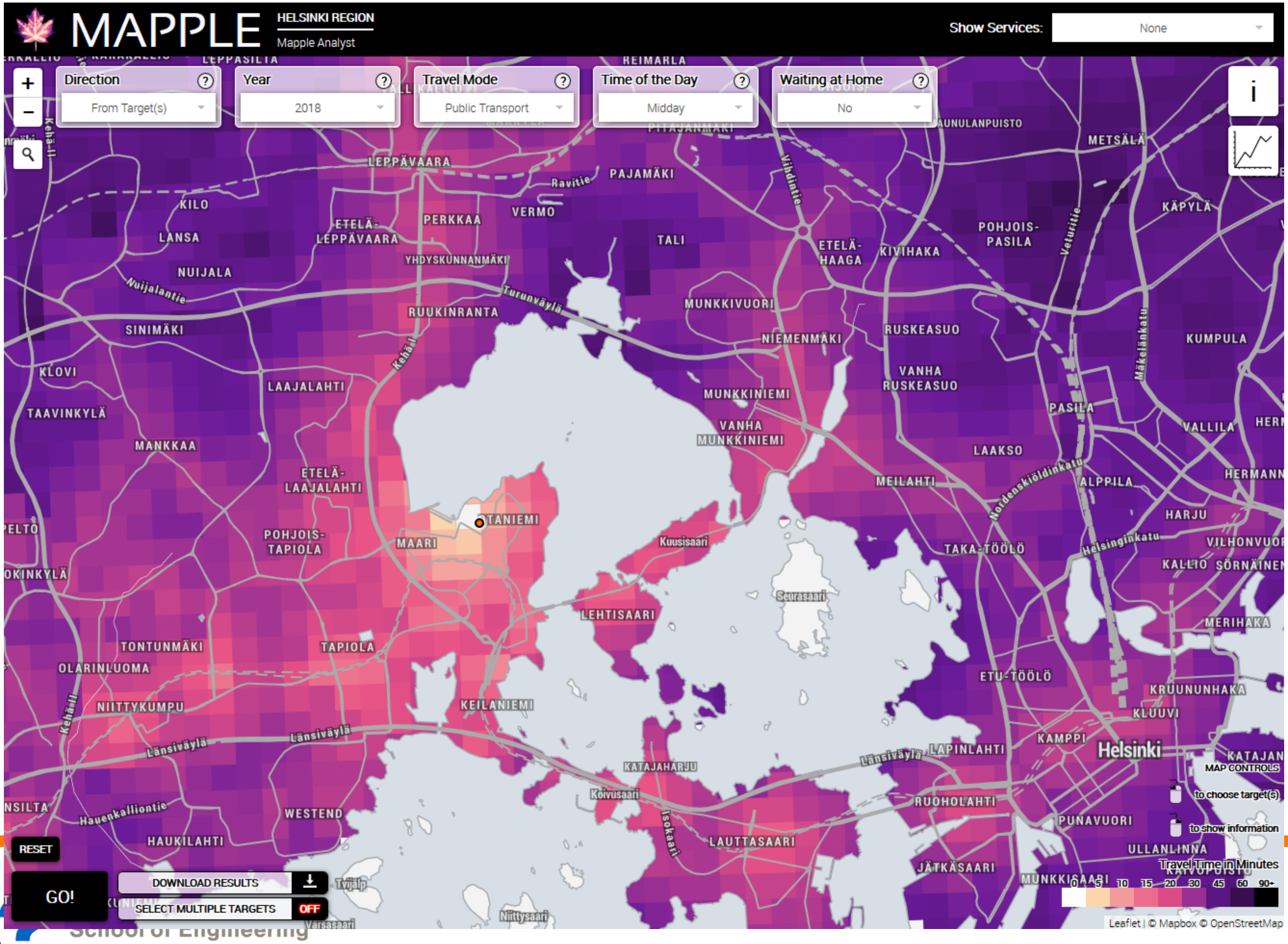


Bolstad: GIS Fundamentals

Cost surface

- Cost surface depicts the cost (e.g. time) required to travel over a pixel
  - Cost depends on the spatial characteristics at each location
    - Density of vegetation
    - Soil type (swamps are difficult to cross)
    - Slope
    - Depth of snow
    - Etc,
- Cost surface can be used to calculate routes
  - Different travel models require different cost surfaces
  - How to travel from starting point to end point
    - Most direct route
    - with minimum cost
- Spreading model from the starting point
  - E.g. travel time maps

Mapple travel time analyst



School of Engineering

Example: Calculating the cost distances (for shortest route travel)

$$\text{distance} = \sqrt{(x^2 + y^2)}$$

$$\text{e.g., } D = \sqrt{(20^2 + 10^2)}$$

$$= 22.4$$

20	10	source cell
22.4	14.1	10
28.3	22.4	20

10 units

cost = distance * fixed cost factor

e.g.,

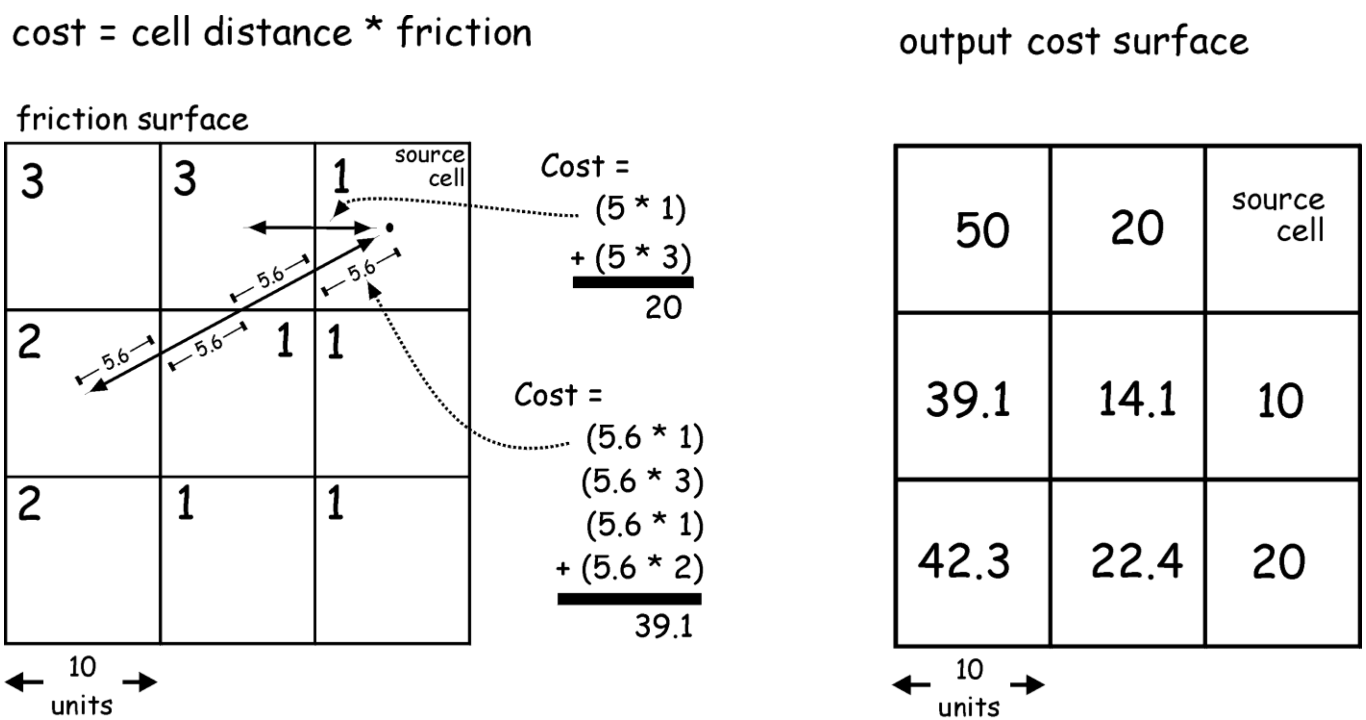
cost = distance * 2

40	20	source cell
44.8	28.2	20
56.6	44.8	40

Continues…

Bolstad: GIS Fundamentals

Example: Cost distances with weighting

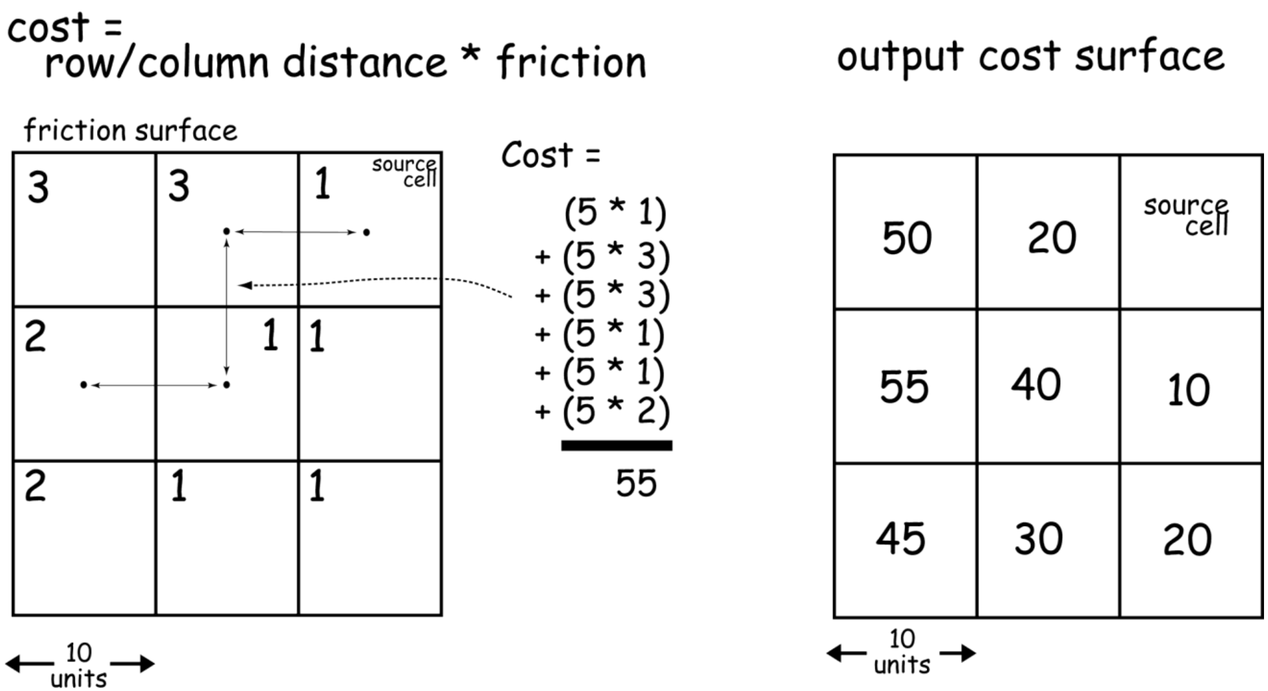


Continues…

Bolstad: GIS Fundamentals

Aalto University
School of Engineering

Example: “Manhattan distances” (diagonal movement forbidden)



Now, what would be the fastest route from source to each other cell using Manhattan distances?

Bolstad: GIS Fundamentals

Aalto University
School of Engineering

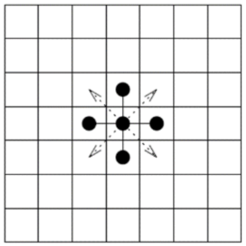
Map Algebra
a formal language for raster analysis

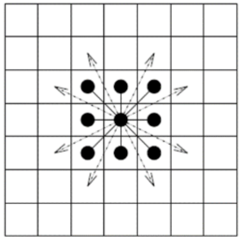
- presented by Dana Tomlin in 1980's
  - revised in:
    Tomlin, D. (2012) *GIS and Cartographic Modeling.*
    ESRI Press.

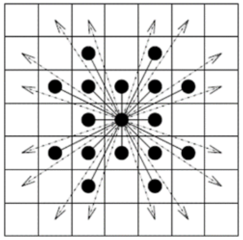
Map algebra (Kartta-algebra)

- A set of primitive operations
  - which allows two or more raster layers (maps) of similar dimensions to produce a new raster layer (map) using arithmetic and logical operations
  - operations on one pixel, neighborhood, zone, entire layer
  - By combining these, one can define a procedure to perform complex tasks
- A formal language, an open standard
- Forms the basis of raster operations in many GISs
  - E.g. ArcGIS Spatial analyst

… map algebra





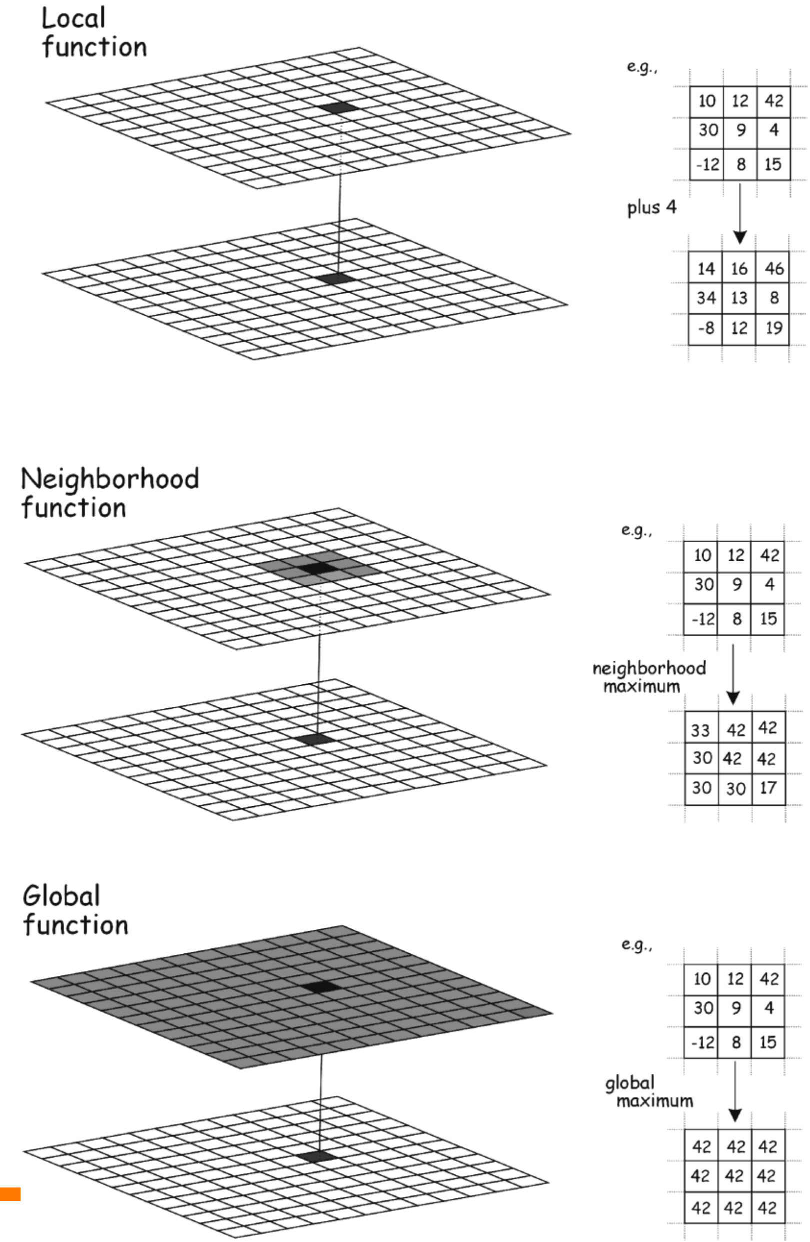


- Key concepts:
- Layers
  - Input-1…input-n
  - NewLayer (= output)
  - Notice the types of data (nominal, ordinal, interval, ratio)
- Location, neighbourhood, zone
  - Local, focal, zonal, incremental, global operations
- Functions
  - Min, max, Boolean, arithmetic, …

Image source: Van Bemmelen et al.: Vector vs. raster-based algorithms for cross country movement planning

…map algebra

- Operations on a raster are of five types
  - 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 within specified zones
  - Incremental operations do on specialized work (e.g. calculate slope, define a network)
  - Global operations compute properties of the entire raster layer



Local functions

- Operations on one pixel at a time from one or more layers, e.g.
  - LocalDifference
  - LocalMaximum
  - LocalRatio
  - LocalProduct

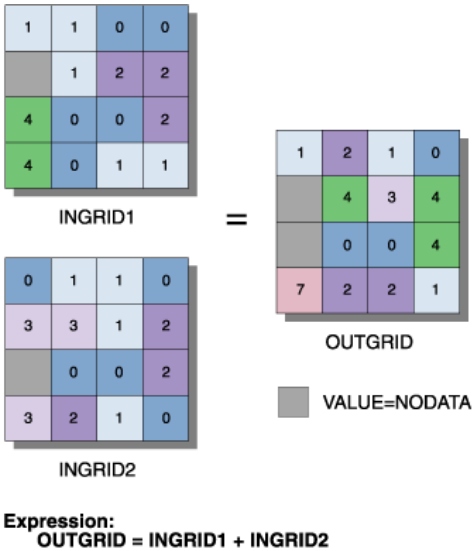
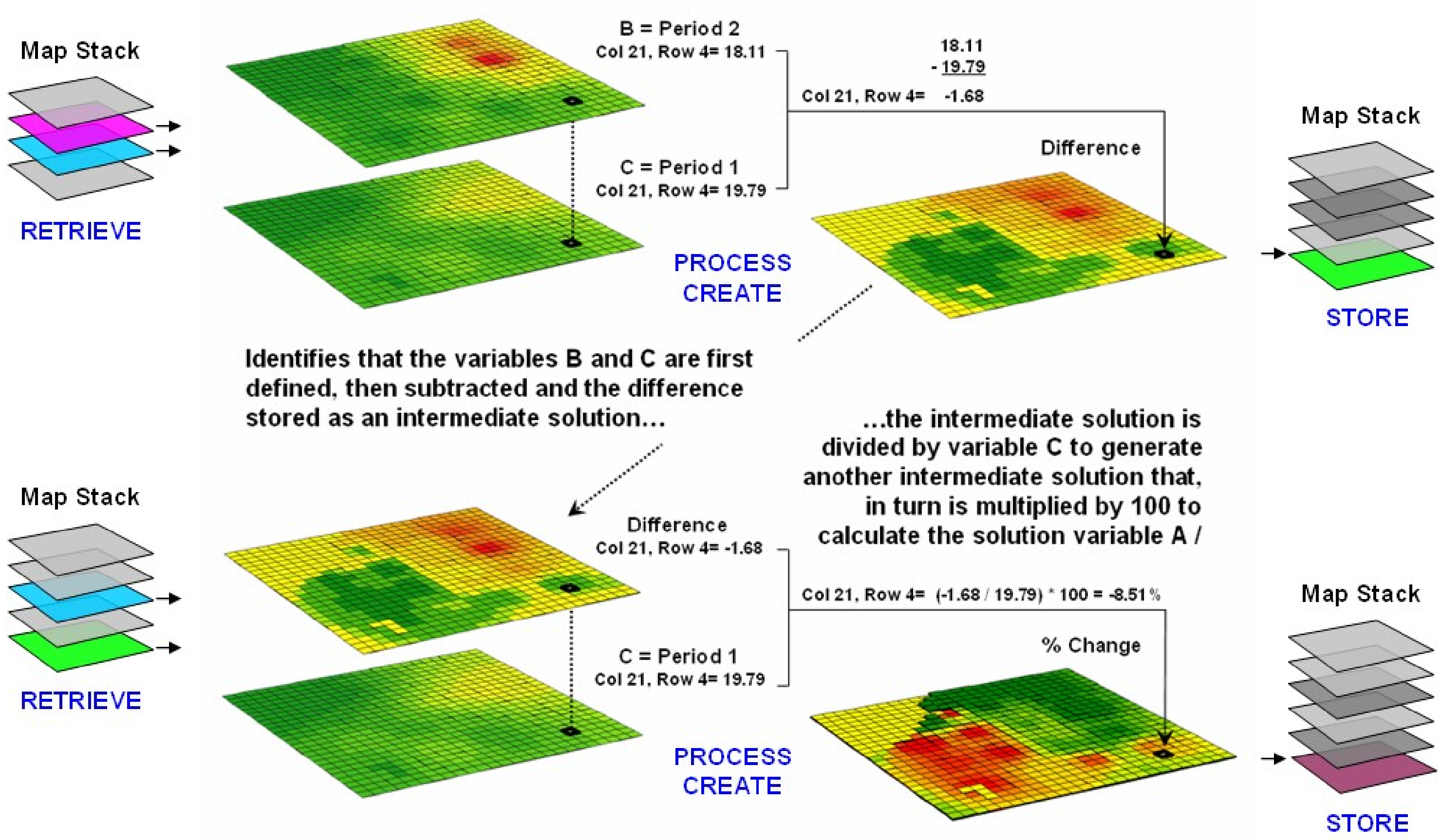


Figure by ESRI

- 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)

Example: Rate of change from period 1 to 2



http://www.innovativegis.com/basis/mapanalysis

Examples of local functions

a)

Input

1	3	1	1
0	N	2	-1
1	2	5	0
0	1	N	N

less than

0	1	0	9
0	5	2	5
0	2	N	2
0	-3	4	8

=

Output

0	0	0	1
0	N	0	1
0	0	N	1
0	0	N	N

b)

1	3	1	1
0	N	2	-1
1	2	5	0
0	1	N	N

equal

0	1	0	9
0	5	2	5
0	2	N	2
0	-3	4	8

=

0	0	0	0
1	N	1	0
0	1	N	0
1	0	N	N

c)

ISNULL

1	3	1	1
0	N	2	-1
1	2	5	0
0	1	N	N

=

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

Notice the NULL values

Bolstad: GIS Fundamentals

Example: Reclassification

a)

Input

1	3	1	1
0	N	2	-1
1	2	5	0
0	1	N	N

Reclass by table

in	out
0	a
1	x
2	b
3	f
4	c
5	s

=

Output

x	f	x	x
a	N	b	N
x	b	s	a
a	x	N	N

b)

1	3	1	1
0	N	2	-1
1	2	5	0
0	1	N	N

Reclass by ranges

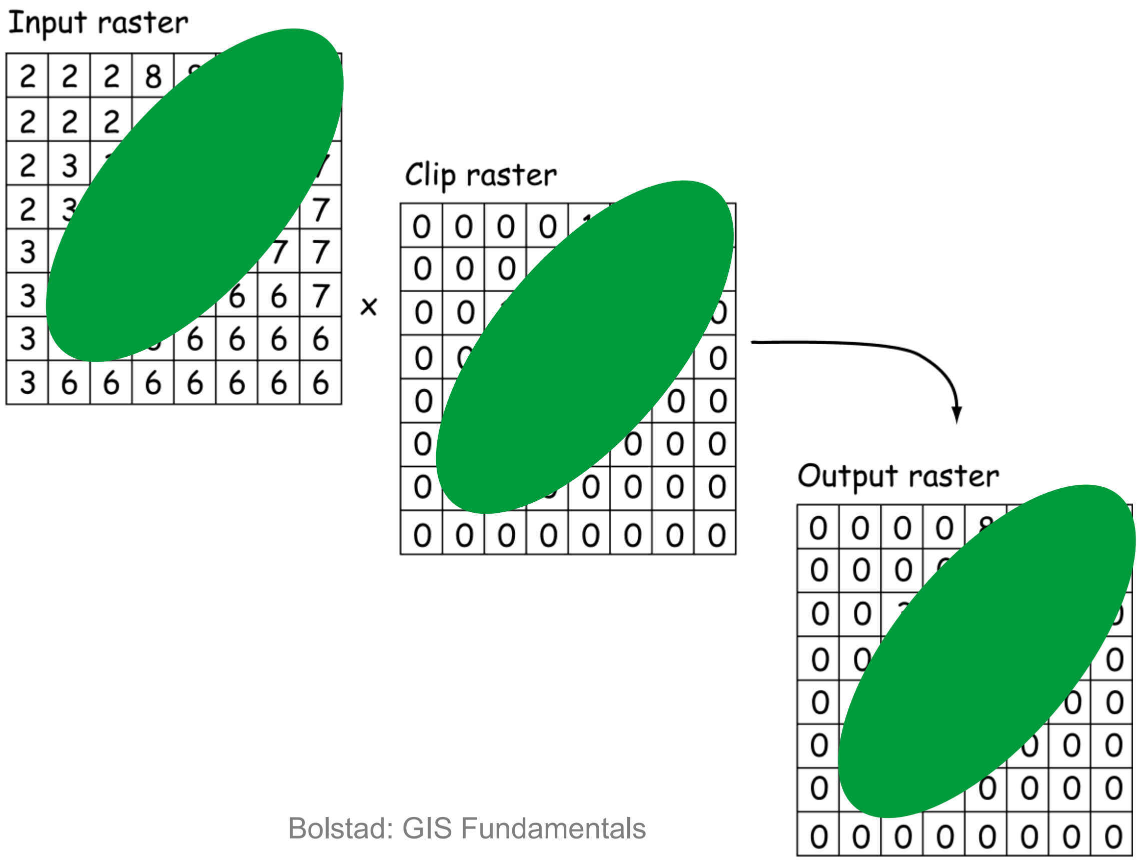
in range	out
0 to 1.5	a
1.5 to 3.5	b
3.5 to 10	c
N	d

=

a	b	a	a
a	d	b	N
a	b	c	a
a	b	d	d

Bolstad: GIS Fundamentals

Example: Clipping



Bolstad: GIS Fundamentals

Focal functions

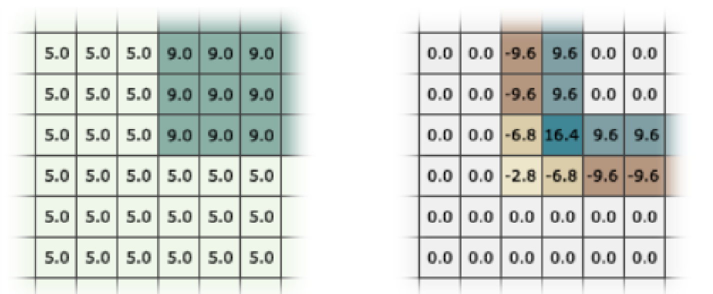
- Operations on a pixel and its 4 (or 8, or 32, or…) neighbouring pixels, e.g.
  - FocalMaximum
  - FocalMean
  - FocalRating
- For example,
  - Smoothing the map layer by FocalAverage; high and low values are smoothed away (continuous values)
  - Generalization by filtering out individual pixels that differ from their surrounding (categorical values)
  - Mobility simulation by producing a cost surface of mobility

Examples

- Focal Sum

- Edge detection
  - This is actually calculated using a custom mask called high-pass filter:

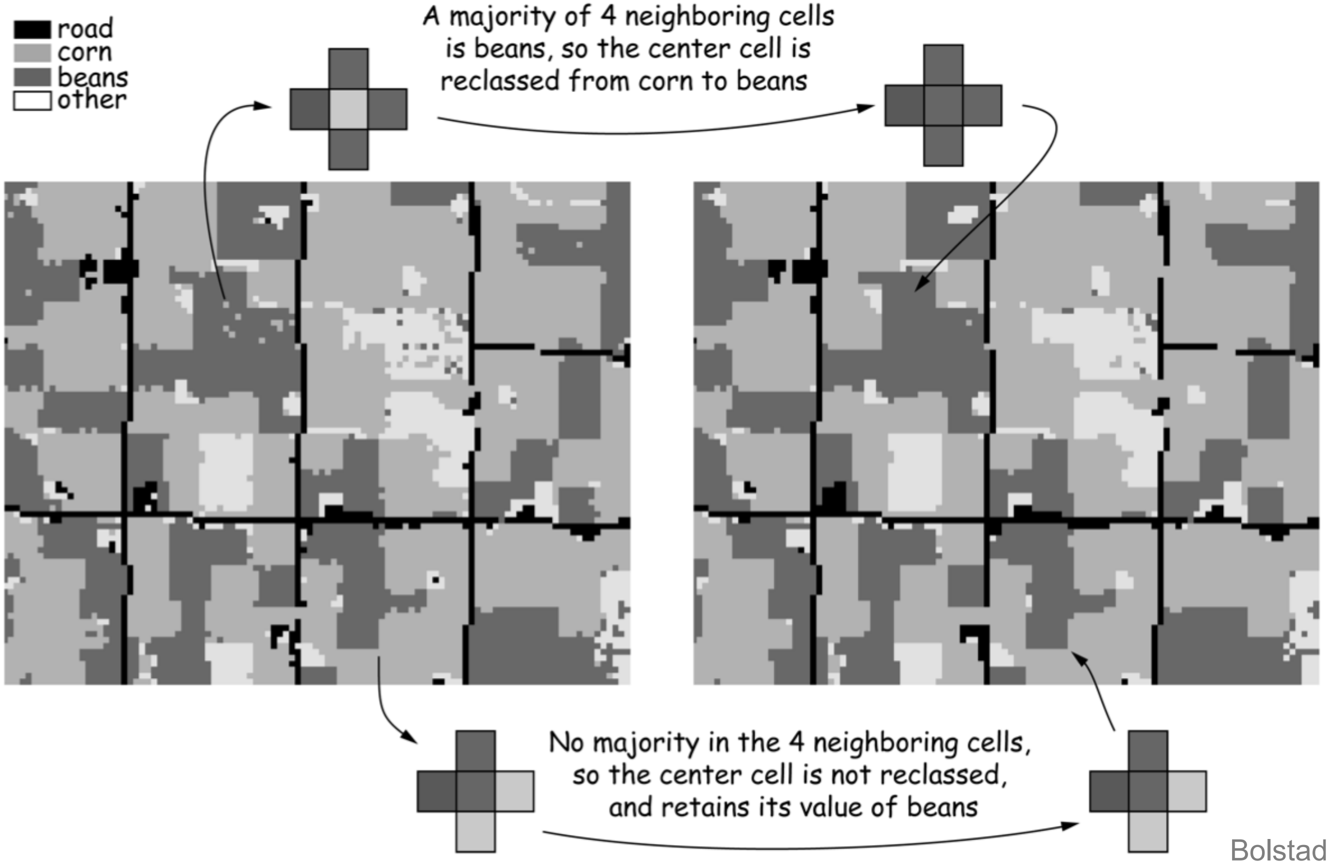
```
-0.7  -1.0  -0.7
-1.0   6.8  -1.0
-0.7  -1.0  -0.7
```



Input with edge where cell values change from 5.0 to 9.0

The HIGH option has detected the edge.
Note that the output values have no relation to the input values.

Example: Filtering with 4-neighbourhood



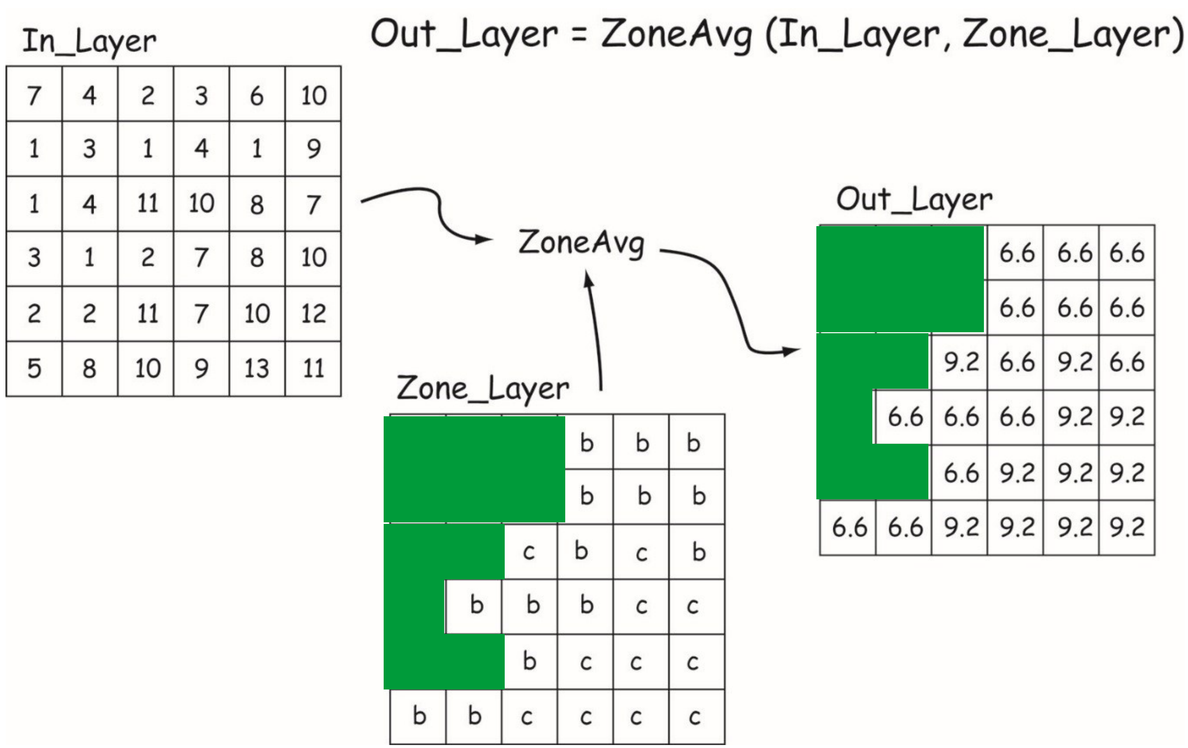
Bolstad: GIS Fundamentals

Zonal functions

- Data is related to zones, each zone is handled separately
- Zones are specified 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

Example: Zonal average



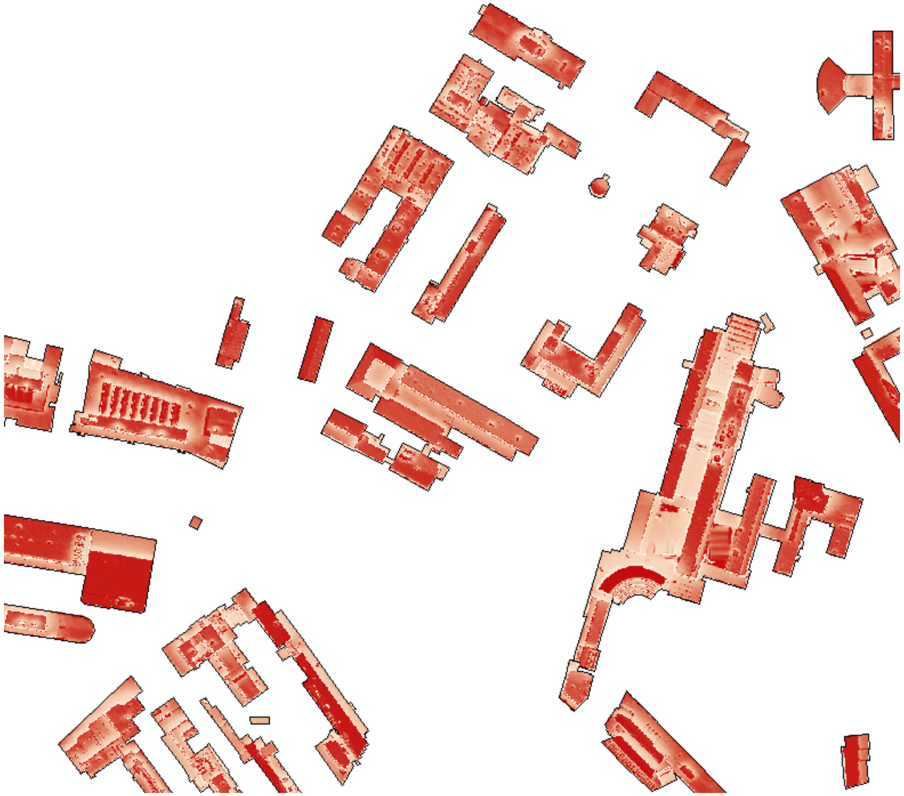
Bolstad: GIS Fundamentals

Example

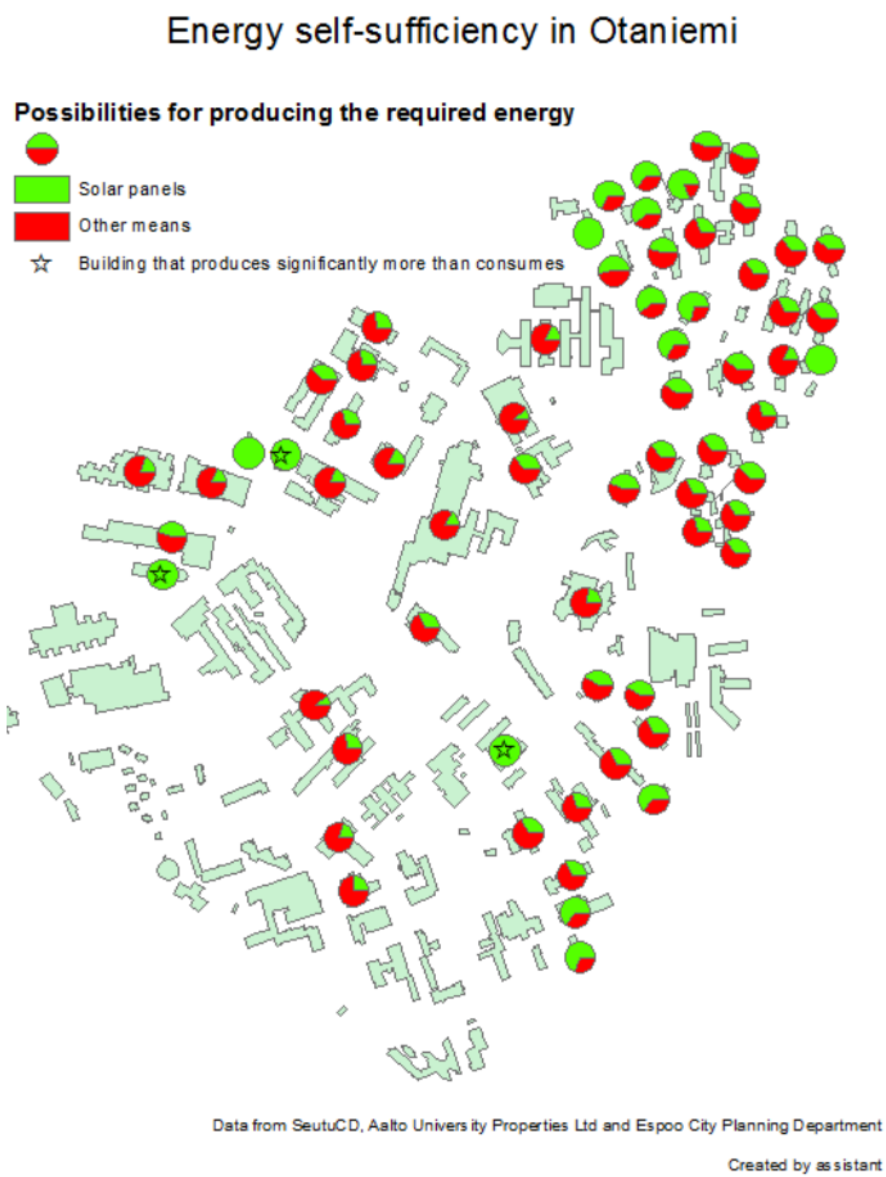
- Energy self-sufficiency in Otaniemi: how big portion of energy consumption can be gathered using solar panels on top of roofs

- Data: Solar energy data, buildings, energy consumption

- Analysis: ZonalSum to get total energy per building



…example

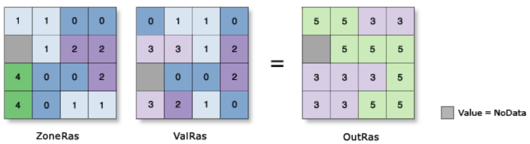

Energy self-sufficiency in Otaniemi
Possibilities for producing the required energy
Solar panels
Other means
Building that produces significantly more than consumes
Data from SeutuCD, Aalto University Properties Ltd and Espoo City Planning Department
Created by assistant

See more details, e.g., in ArcGIS help

ZonalSum

Records, in each output cell, the sum of the values of all cells in the value raster that belong to the same zone as the output cell. Zones are identified by the values of the cells in the input zone raster.

Illustration



Usage tips

Map Algebra

- The zone raster must be an integer value raster.
- The value raster can be integer or floating point.
- The data type of the output raster is floating point. This is because the value for the Sum tends to be quite large, and may not be possible to represent with an integer value. Consider for example a zone that is 2500 rows and columns of cells in size, and the value of each cell is 1000. The sum for that zone would be 2500x2500x1000 = 6.25 billion. If an integer output is required and the range is within ± 2.147 billion, then you can apply the INT function.
- Learn more about how to specify the input raster dataset in the Map Algebra expression of Raster Calculator.

Syntax

See Zonal Statistics

To perform the equivalent to map algebra ZonalSum, set the statistics type to Sum.

Map Algebra syntax

ZonalSum(<zone_grid>, <value_grid>, {DATA | NODATA})

Parameter	Explanation
<zone_grid>	An integer raster that identifies the zone for each cell. The zone raster can be an expression resulting in a raster.
<value_grid>	An integer or floating-point raster that defines the values of the cells to be added together. The sum of the values of the value raster in each zone will be assigned to every cell in that zone on the output raster. The value raster may be an expression resulting in a raster.
{DATA | NODATA}	A keyword defining the manner in which NoData values on the <value_grid> that fall within a zone defined by the <zone_grid> will influence the output results. • DATA — A keyword that specifies that if a NoData value exists in a cell on the value raster within any particular zone defined by <zone_grid>, the cell will be ignored, and only cells on the value raster that have data values within the zone will be summed for the output. • NODATA — A keyword that signifies that if a NoData value exists on the value raster within any particular zone defined by <zone_grid>, there is insufficient information to complete the calculation to determine the sum of the values for all the cells within the zone, and the entire zone will receive the NoData value on the output raster.

Map Algebra example

```
zonalsum(zonegrid, valuegrid)
zonalsum(zonegrid, valuegrid, NODATA)
zonalsum(zonegrid, valuegrid, DATA)
zonalsum(zonegrid, (ingrid1 + ingrid2))
zonalsum((ingrid1 + ingrid2), ingrid3, NODATA)
zonalsum(reclass(ingrid1,tab1), (ingrid2 div ingrid3 * 4))
```

ArcObjects syntax

See Zonal Statistics

Aalto University
School of Engineering

Incremental functions

- Incremental functions do various specialized analysis on a raster layer
  - Less primitive general operations and more specialized operations that cannot be (easily) done directly by using the defined primitives
- For example flow analysis functions can calculate aspect and gradient and flow
  - Flow in this case tells which neighboring pixel the water flows from that particular pixel
- Linkage functions can be used to create explicit networks (e.g. the road network)

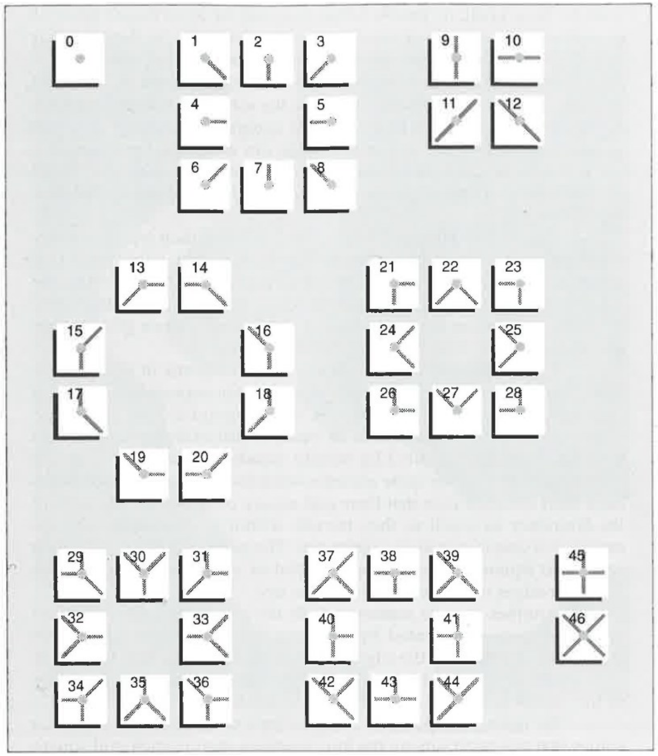
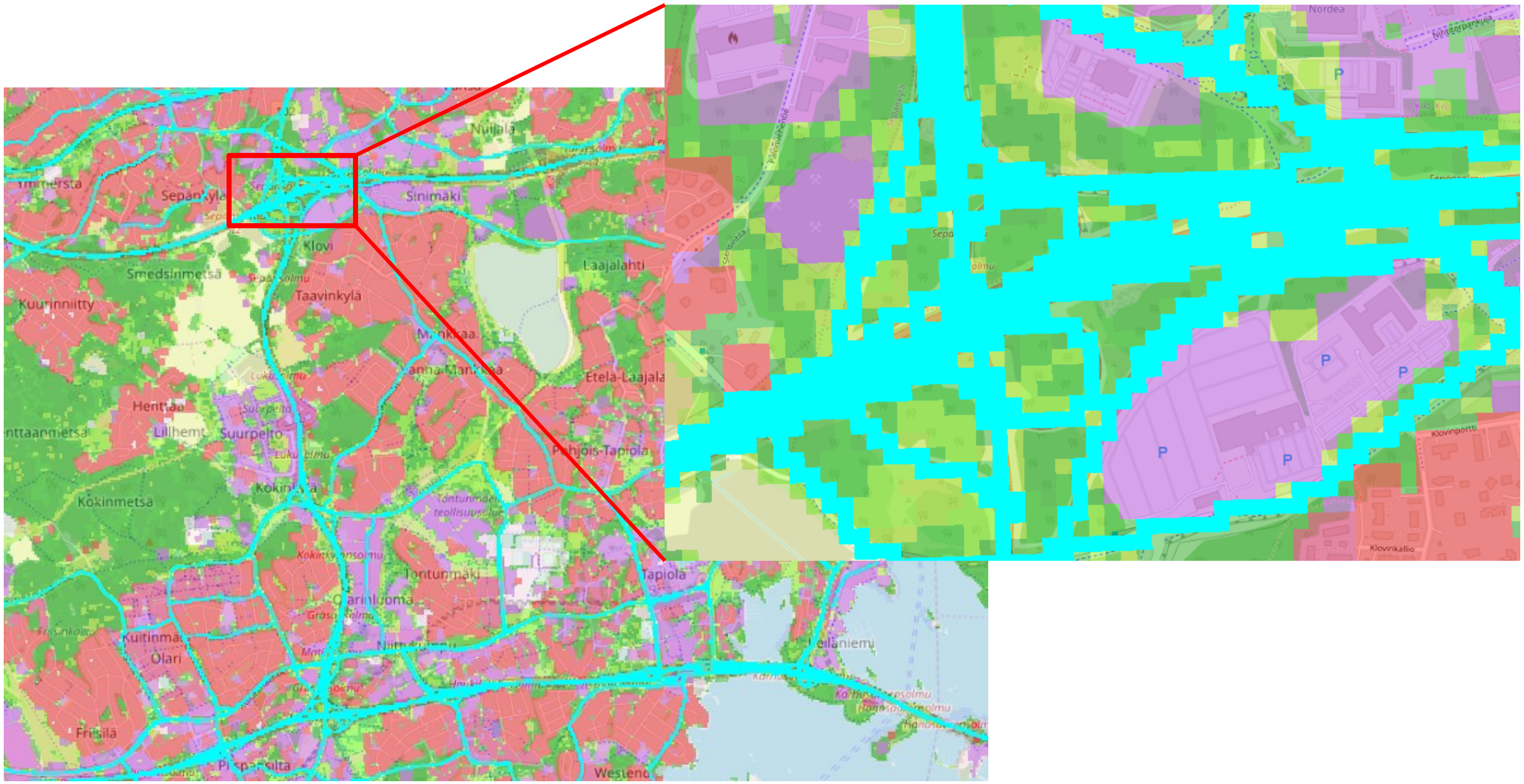


Image source: Tomlin: Geographic Information Systems and Cartographic modeling

Street network finding problems



Aalto University
School of Engineering

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