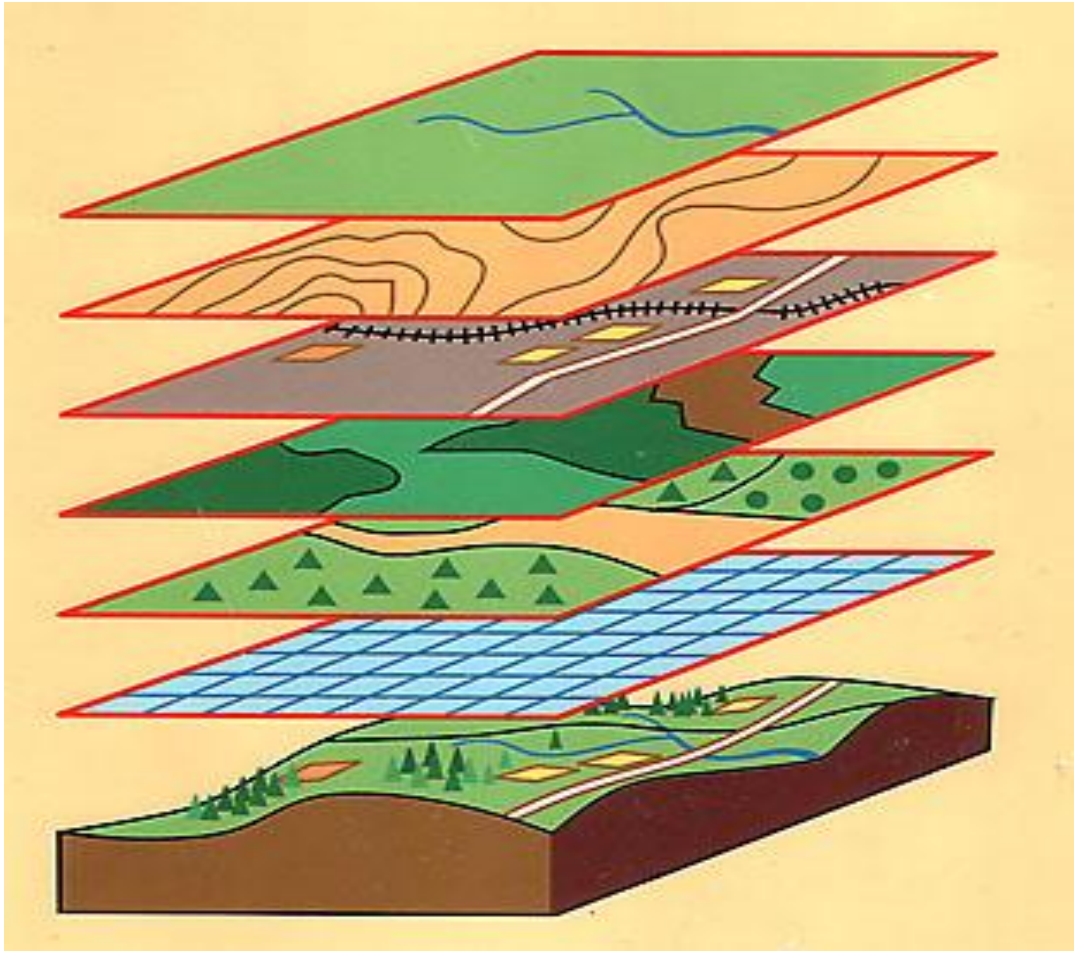
Research Methodology – GIS Basics 1/2



UNIVERSITY OF
EASTERN FINLAND

Content

- Lectures, 4 h
  - Introduction
  - Projections
  - Vector Analysis
  - Raster Analysis
  - GIS systems and software
  - Applications
- Exercises, ArcGIS or QGis, 6 h

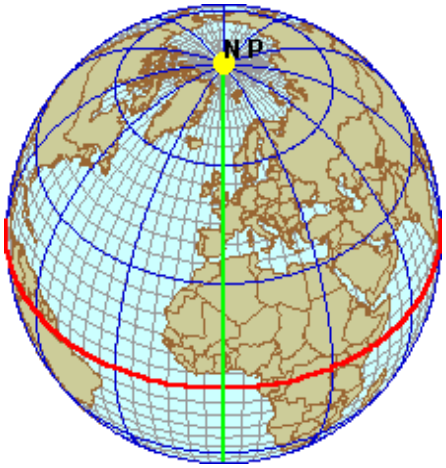
- Basic GIS training available in B. Sc. Level course (enrolment key GIS19): https://moodle.uef.fi/course/view.php?id=6890
- Book (Chapters 6,7,8,9 and 12 enough): http://ezproxy.uef.fi:2048/login?url=http://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&AN=269101

What is GIS? Geographical Information System

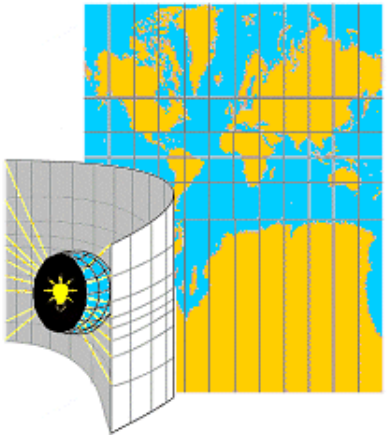
- Computer based program that can create, manage, analyse and visualize different types of geographic data.
- Connects data to a map integrating location with descriptive information.
- Can be used for research, better management and decision making e.g. natural resources, risk analysis (forest fire, floods), new locations for business…
- GIS softwares:
  - ArcGIS (ESRI, license)
  - QGIS (Open source, free)
  - Geomedia (license)
  - MapInfo (license)
  - GRASS GIS, SAGA GIS…

Coordinate Systems Geographic vs. Projected

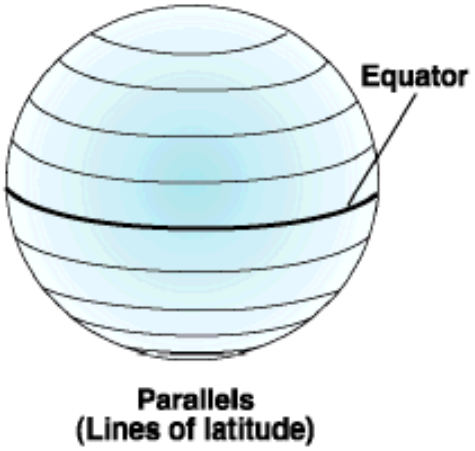
- Geographic Coordinate Systems (GCS)
  - Location measured from curved surface of the earth
  - Measurement units latitude and longitude
    - Degrees-minutes-seconds (DMS)
    - Decimal degrees (DD) or radians (rad)

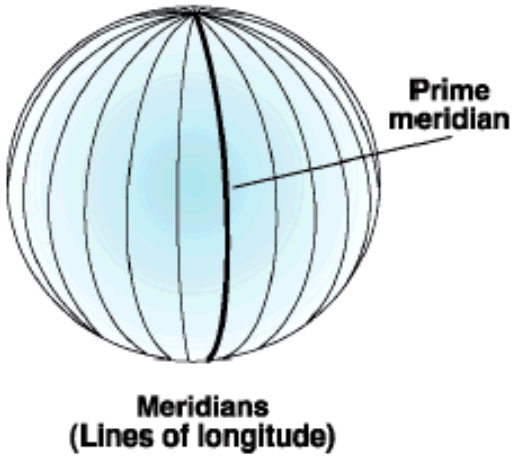


- Projected Coordinate Systems (PCS)
  - Flat surface
  - Units can be in meters, feet, inches
  - Distortions will occur, except for very fine scale maps

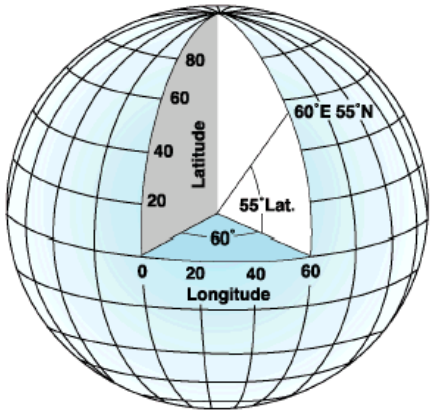


Geographic Coordinate System (GCS)

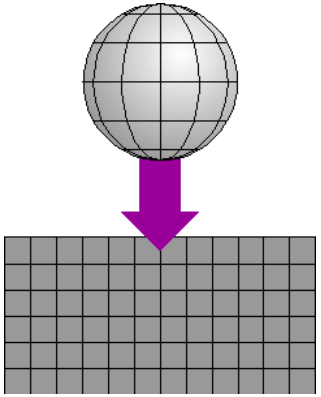




- GCS uses three-dimensional spherical surface to define location on earth.
- Greenwich is the Prime Meridian 0° of longitude
- Parallels – North to South – 0° at the Equator (0°- 90°)
- Meridians – East to West – 0° at the Prime Meridian (0°- 180°)
- Latitude and longitude are angular measurements made from the center of the earth to a point on the surface of the earth
- Example: Joensuu 62.6010° N, 29.7636° E

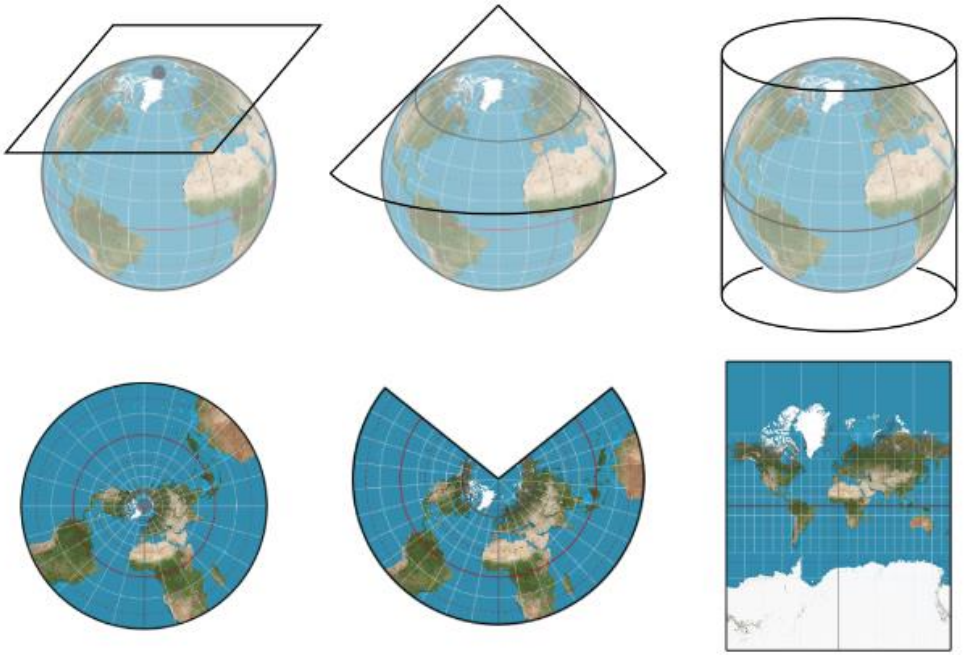


Map Projection



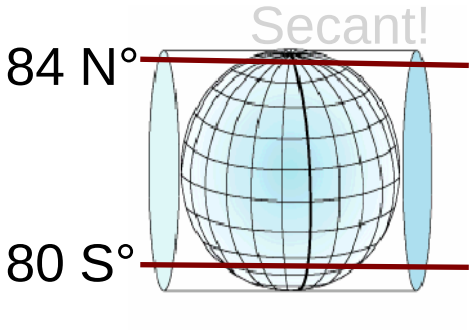
A map projection is a mathematical formula for representing the curved surface of the earth on a flat map.

- Wide variety of projections possible
- Each projection will create a different type of distortion
- Select the best one for your purpose

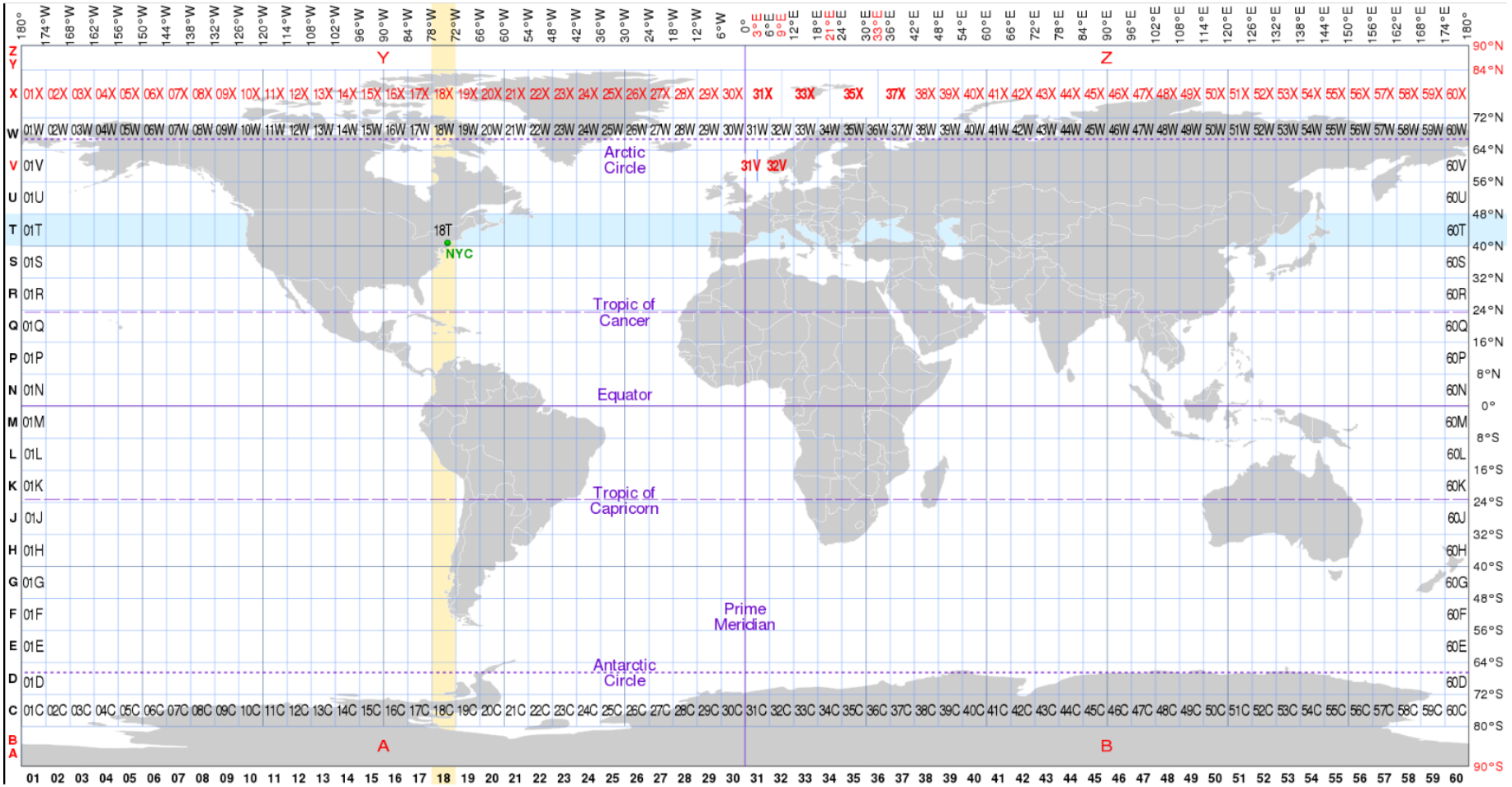


Universal Transverse Mercator (UTM) coordinate system

- One of the most used map projections.
- Divides earth into 60 zones
  - 6^{o} wide of longitude (~800km)
  - Defined by central meridian (example: 120° W)
- Preserves direction and small shapes (conformal projection).
- Extent is from 84°N to 80 °S.
- UTM coordinates are easily recognized by 6 digit for the **x**, and 7 digit for the **y** (most of the time at latitudes of 15° and greater in the Northern Hemisphere)



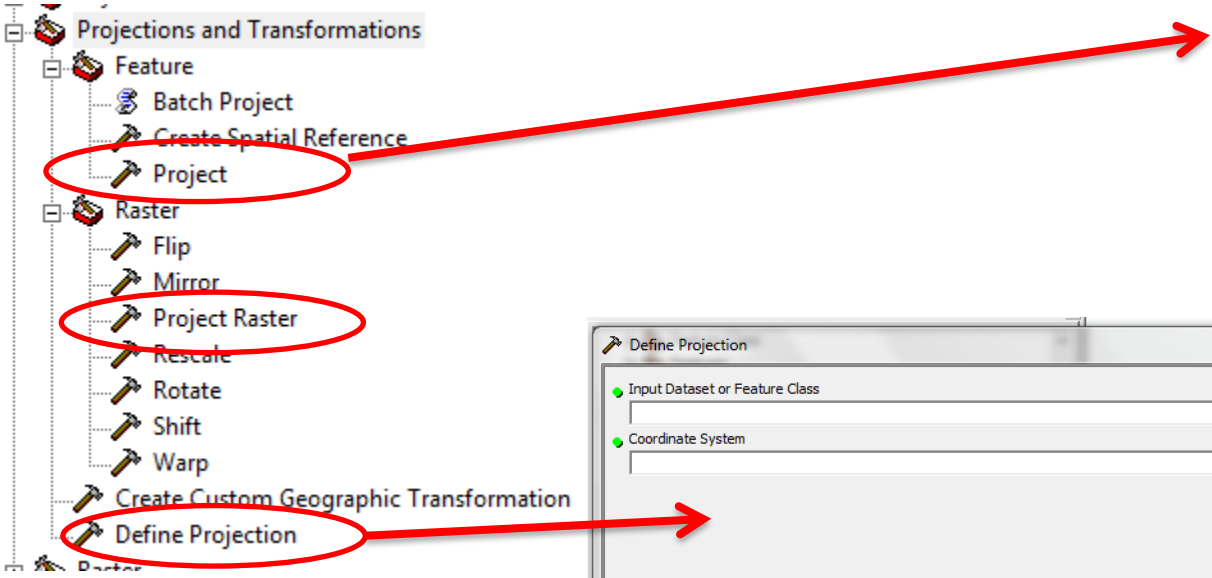
Universal Transverse Mercator (UTM) coordinate system

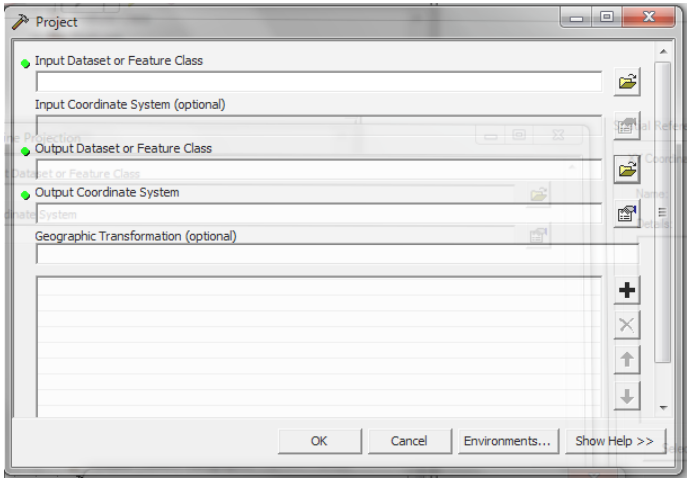


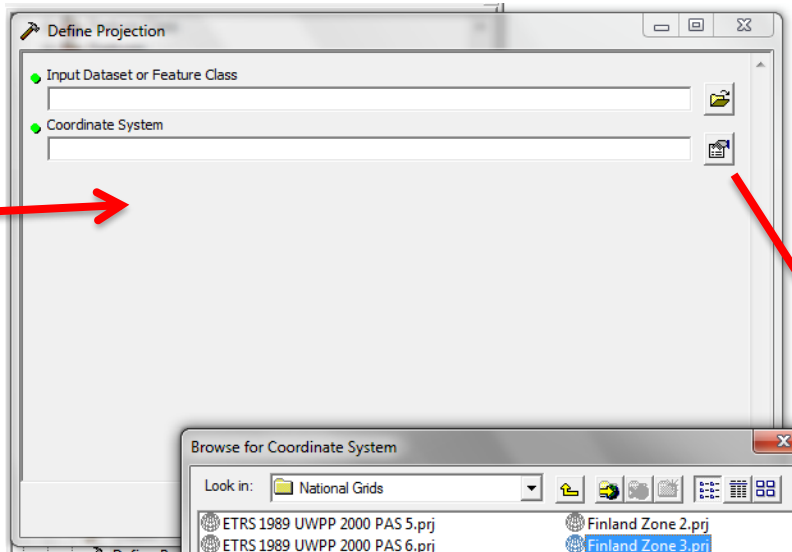
Lambert azimuthal equal-area projection

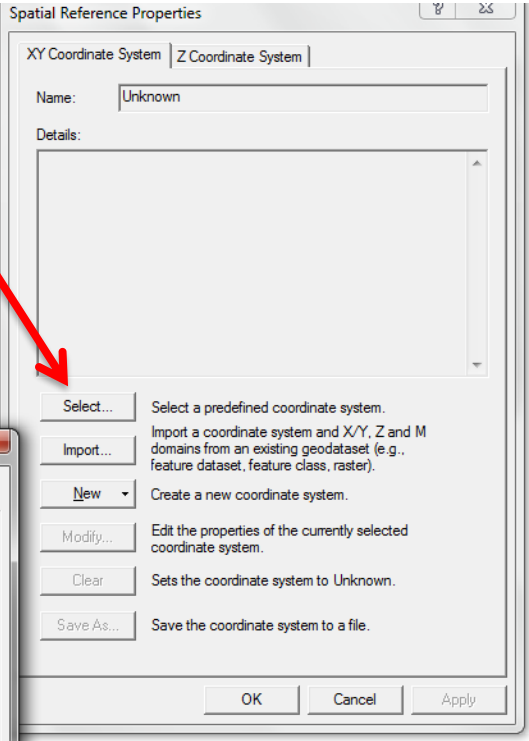
The European Environment Agency recommends to use for European level mapping for statistical analysis and display.

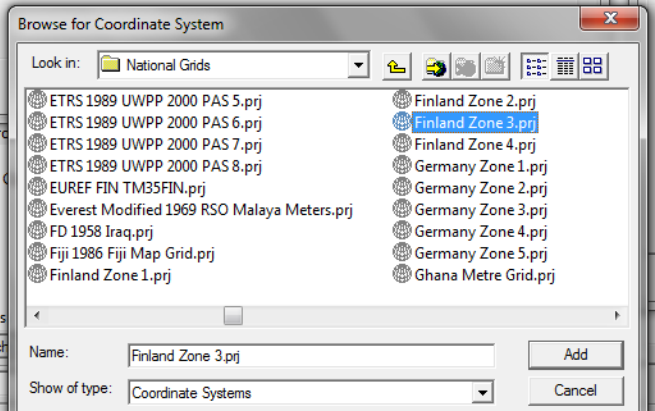
Coordinate systems in ArcGIS









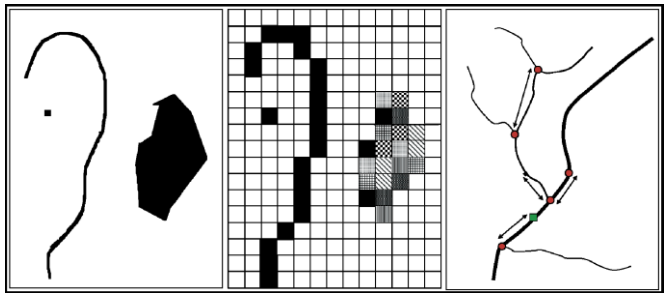


Steps

1) Define projection
2) Project if needed

UNIVERSITY OF EASTERN FINLAND

GIS data representation – vector / raster



- **Vector** – the most common GIS data type consist of points, lines or polygons.
- **Topology** – expresses the spatial relationship of adjacent or connecting features (points, lines and polygons) and how they share the geometry.
- **Raster** – consist of a matrix of cells/pixels, each cell contains a value (e.g. DEM, temperature). Can be aerial fotos, satellite imagery, digital picture, scanned maps

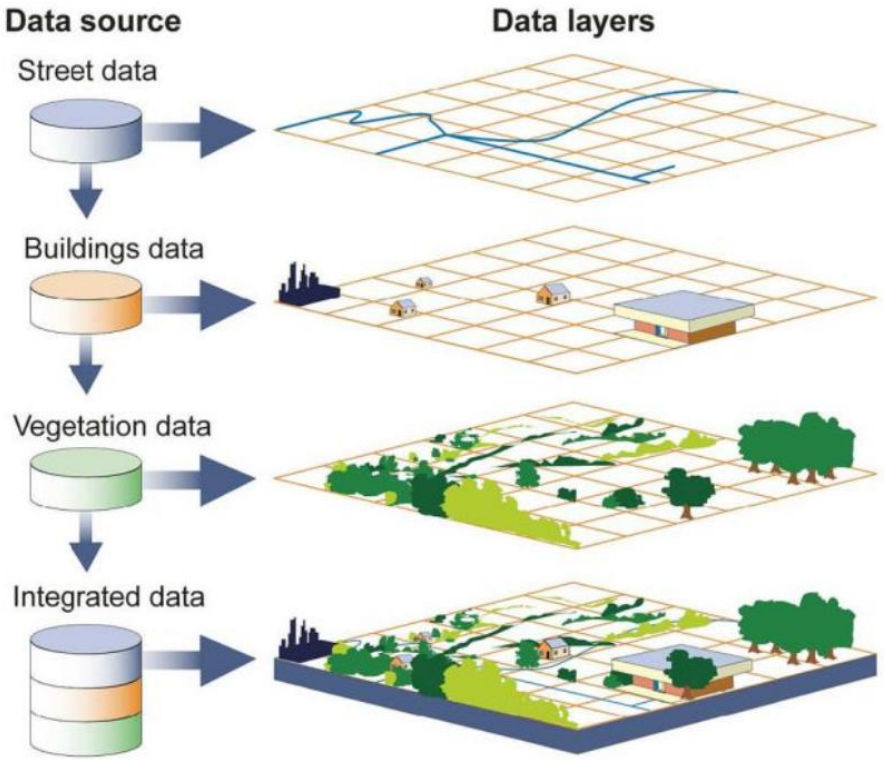


ILLUSTRATION COURTESY U.S. GOVERNMENT ACCOUNTABILITY OFFICE

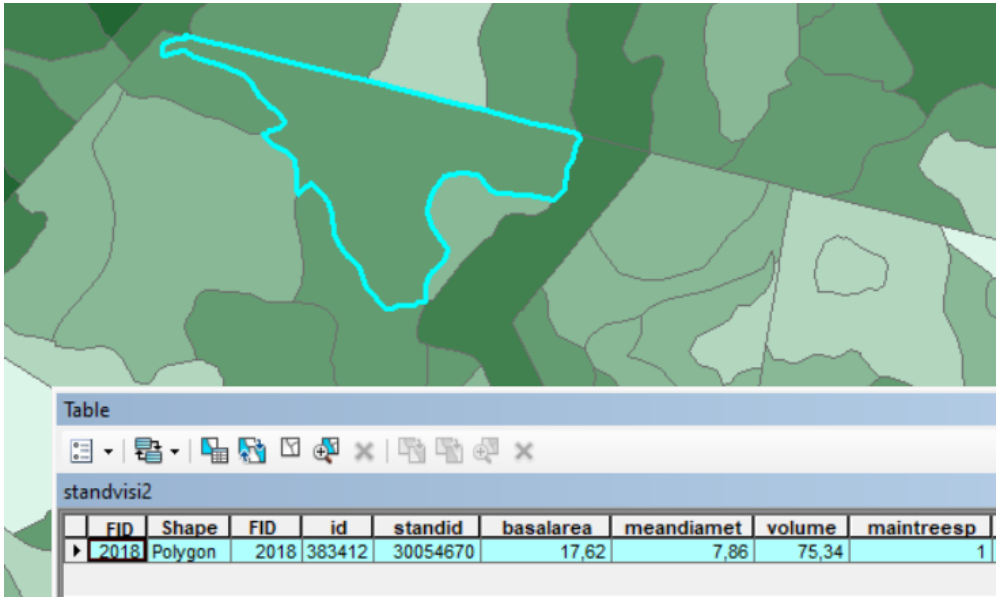
GIS data

Features displayed in GIS have two components:

1) Information describing their spatial location and shape

2) Attribute data describing their non-spatial qualities

Raster representation	*Vector representation*
advantages	
• simple data structure • simple implementation of overlays • efficient for image processing	• efficient representation of topology • adapts well to scale changes • allows representing networks • allows easy association with attribute data
disadvantages	
• less compact data structure • difficulties in representing topology • cell boundaries independent of feature boundaries	• complex data structure • overlay more difficult to implement • inefficient for image processing • more update-intensive



GPS-methods

- Absolute positioning (one receiver)
- Relative positioning (at least 2 receivers)
  - *Differential Global Positioning System (DGPS)* is an enhancement to Global Positioning System that uses a network of fixed, ground-based reference stations to broadcast the difference between the positions indicated by the satellite systems and the known fixed positions. (http://en.wikipedia.org/wiki/Differential_GPS)
    - These stations broadcast the difference between the measured satellite pseudoranges and actual (internally computed) pseudoranges, and receiver stations may correct their pseudoranges by the same amount.
    - The correction signal is typically broadcast over UHF radio modem.
  - *Real Time Kinematic (RTK)* satellite navigation is a technique used in land survey and in hydrographic survey based on the use of carrier phase measurements of the GPS, GLONASS and/or Galileo signals where a single reference station provides the real-time corrections, providing up to centimetre-level accuracy. (http://en.wikipedia.org/wiki/Real_Time_Kinematic)
    - When referring to GPS in particular, the system is also commonly referred to as Carrier-Phase Enhancement, CPGPS. RTK systems use a single base station and a number of mobile units.
    - The base station re-broadcasts the phase of the carrier that it measured, and the mobile units compare their own phase measurements with the ones received from the base station.

GIS Analysis

- You can process ready data or create your own by editing and processing data.
- New information can be produced from spatial databases.
- **Spatial analysis** is a process, where several thematic map layers are combined and overlaid. **Spatial Queries** are done according to location and attributes.
- GIS system is also a modelling tool
- Simulation models are normally utilising spatial and non-spatial data as input

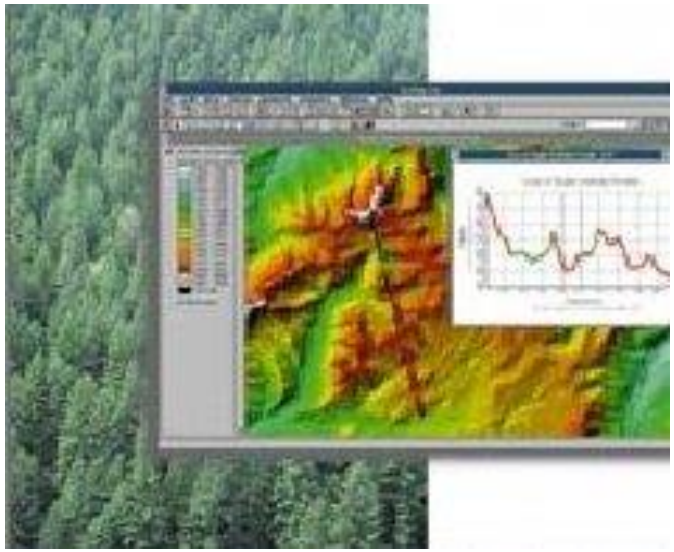
Free GIS data sources:

- ESRI data hub: https://hub.arcgis.com/search
- Natural Earth: https://www.naturalearthdata.com/
- OpenStreetMap: https://gisgeography.com/openstreetmap-download-osm-data/
  - **Natural and land use features** like beaches, forests, and farms.
  - **Places of interest** like restaurants, hotels, police stations, churches, and museums.
  - **Transportation features** like bus stops, traffic signals, and stop signs.
  - **Populated places** like cities, towns, villages, and localities.
- EU open GIS data EUROSTAT: https://ec.europa.eu/eurostat/web/gisco/geodata/reference-data

- Finnish National Land Survey (Maanmittaulaitos): https://tiedostopalvelu.maanmittauslaitos.fi/tp/kartta?lang=en
- Finnish Environment Institute (Ympäristökeskus): https://www.syke.fi/en-US/Open_information/Spatial_datasets/Downloadable_spatial_dataset

Vector GIS tools

- Area, length of line, distance between objects
- Overlay analysis
- Buffer analysis
- Connection to database
- Generalisations or filtering
- Classification
- Network analysis
- Interpolation
- DEM and 3D analysis
- Change



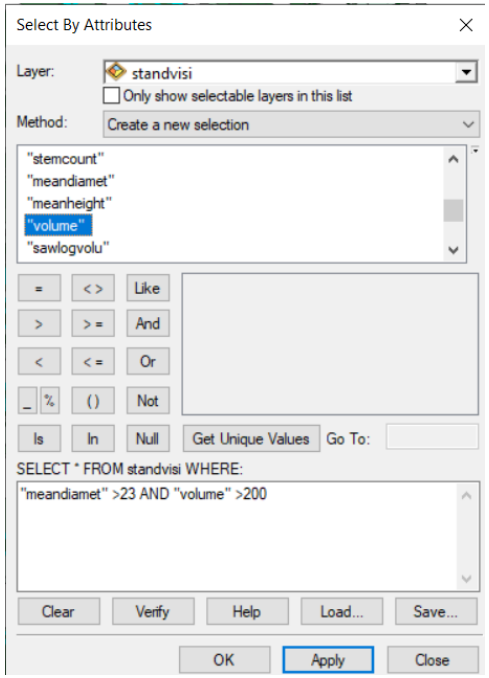
Database queries

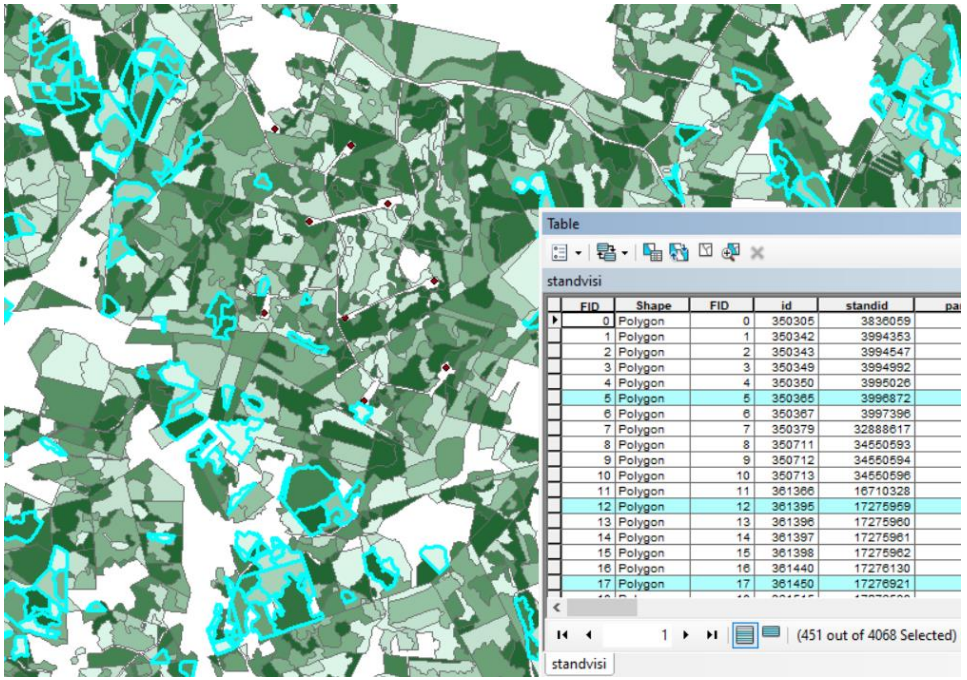
- **Database queries**: a request using the attribute table variables
- **Spatial queries**: uses spatial location e.g. intersection, near, union

- The main difference between attribute data and spatial data is that the **attribute data describes the characteristics of a geographical feature while spatial data describes the absolute and relative location of geographic features**. Attributes are normally stored to separate relational database.
- Communication between application and using query language
- Relational databases utisizes Structured Query Language (SQL)
- First commercial product Oracle (1979)
- Usage of SQL:
  - Interactive use with commands or user-interface window
  - Embedded SQL-commands in software modules

Database queries in GIS interface

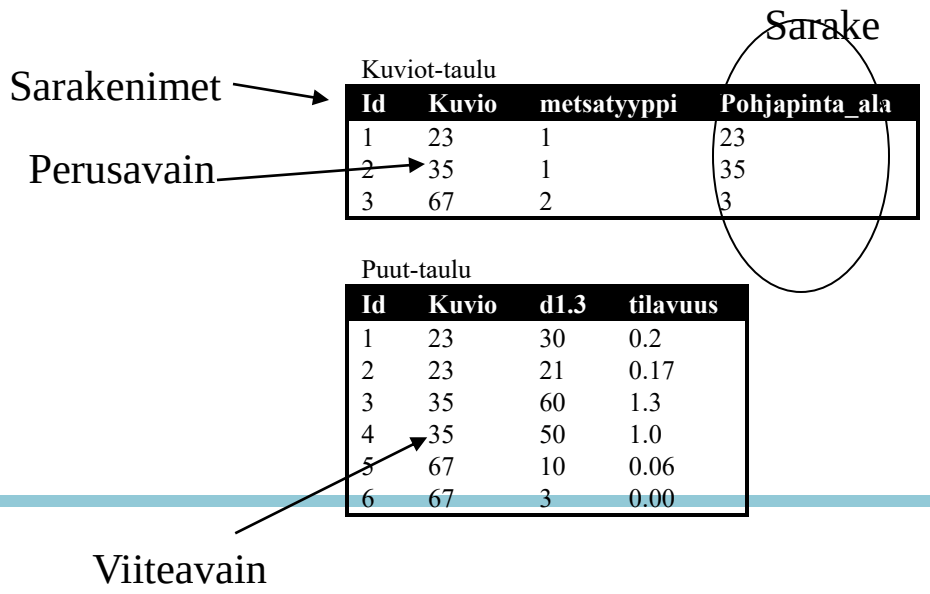
- Search forest stands, which have
  - Mean diameter of trees > 23 cm
  - Volume of stand > 200 m^3/ha





Structure of relational database

- Principles are based of logic and basic mathematics
- Thematic objects organized into **tables**, where we have **records** in **columns** and **fields/attributes** are in columns.
- Tables utilise primary keys in identifation of objects and in connecting/joining tables together. Foreign keys are used in connecting tables.
- Table structure and logic between tables using constrains.



Queries:SELECT

Select	which columns are needed
[distinct]	removal of doubble lines
from	from which table
where	which lines from tables
group by	categorize output
having	specify certain categories
order by	sorting instructions

Sample

Search all people in Lehmo municipality, who has salary > 6500 and worked over 500 hours . Id, name and hours are needed.

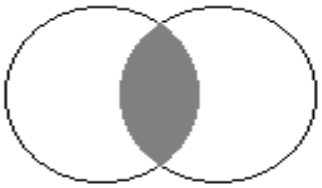
```
SELECT h.tunnus, h.snimi, h.svuosi, SUM(tunnit) AS summa
    FROM Henki h, Prhe p
    WHERE h.tunnus = p.tunnus
    AND (kunta='Lehmo' OR palkka > 6500) AND h.svuosi =
        (SELECT MIN (svuosi) FROM henki)
    GROUP BY h.tunnus, snimi, svuosi
    HAVING SUM(tunnit) > 500
    ORDER BY snimi
```

Vector GIS analysis

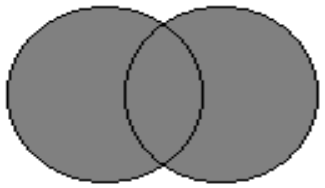
UNIVERSITY OF
EASTERN FINLAND

Overlay analysis

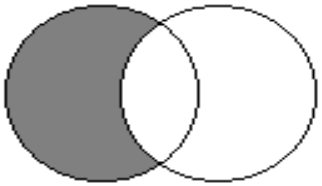
- We can produce from two or more original map layer a new product where needed multiple criteria are met.
- Overlay is utilising logical operators (=, <, >, and, or, nor, not, xor) and arithmetic operators.



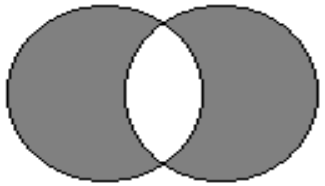
A AND B



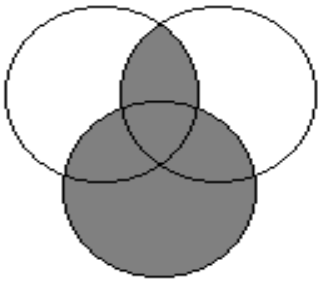
A OR B



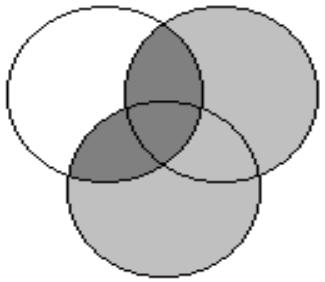
A NOT B



A XOR B



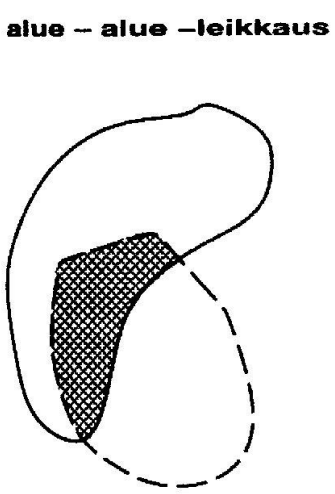
(A AND B) OR C



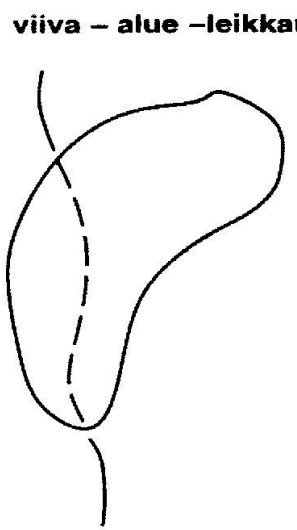
A AND (B OR C)

Overlay analysis

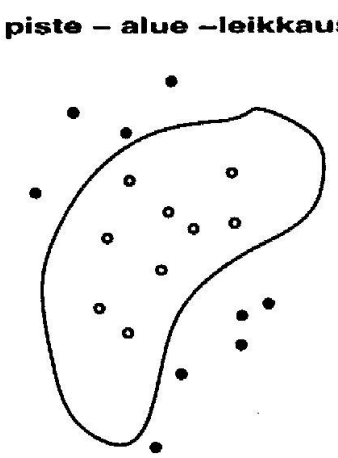
Area-area



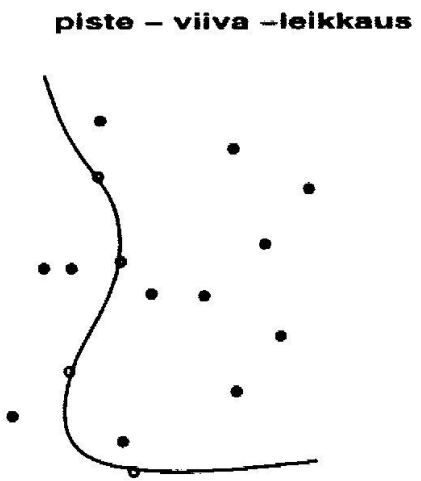
Line - area



Point - area



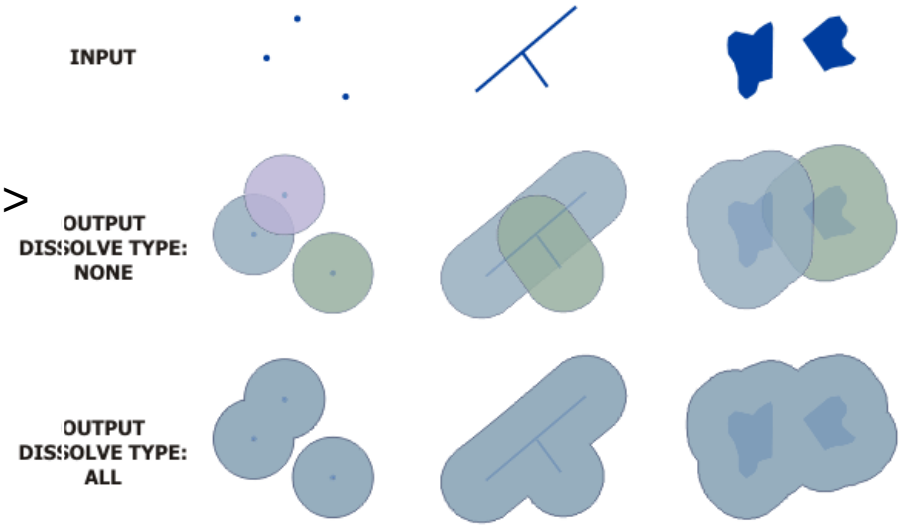
Point - line



Buffering

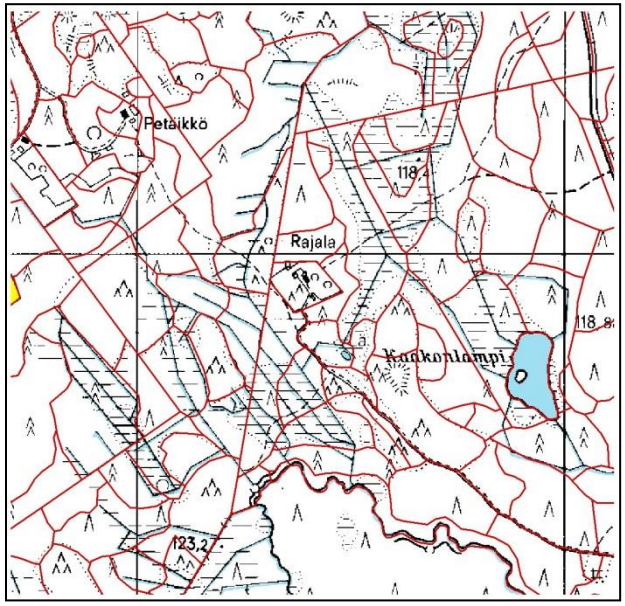
- We can create layers, where specific distance from objects is defined (buffer-function) or search object which are closer than specified distance (near-function)

- Vector buffer is always a polygon
- Raster buffer contain always a distance from object as cell value -> suitable for modelling

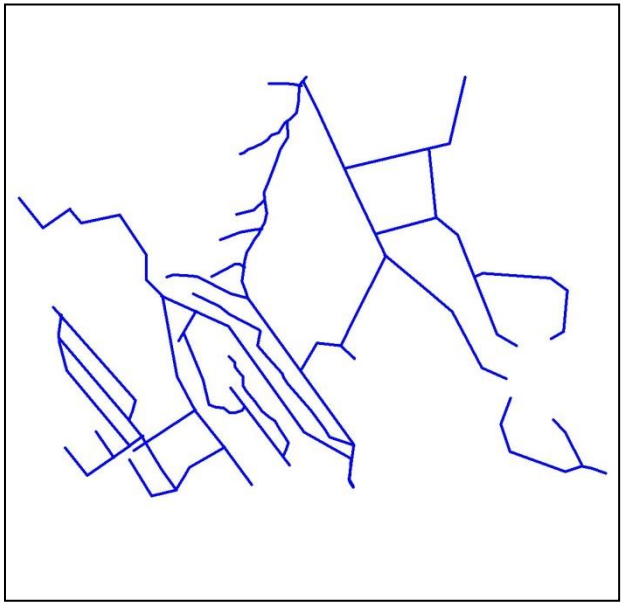


Sample: Distance in spatial growth model, 2D

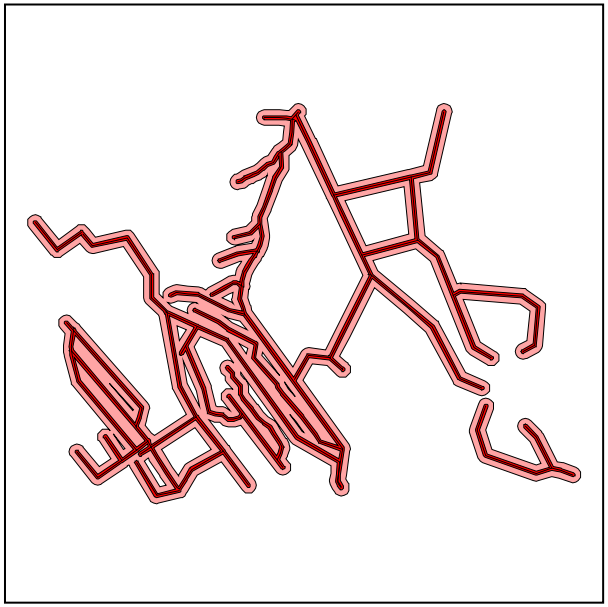
Topographic map



Ditches



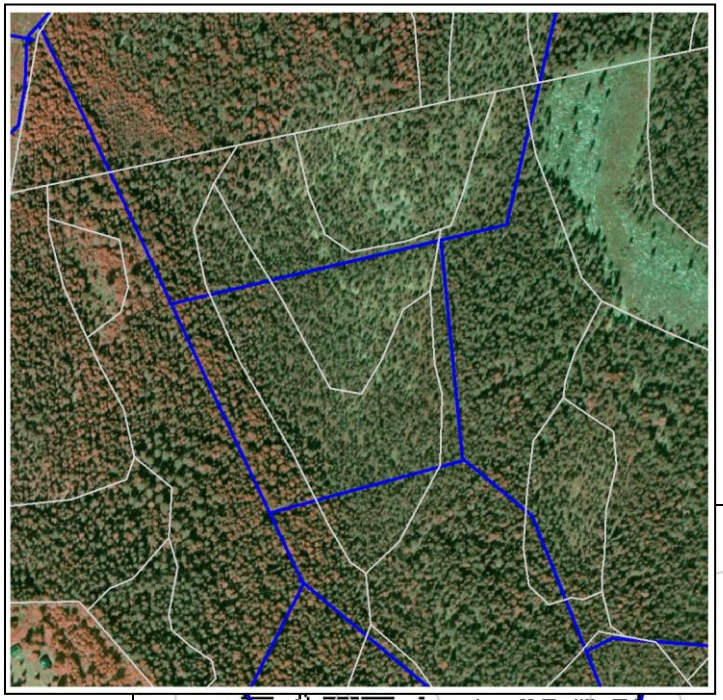
5m and 20 buffers around roads

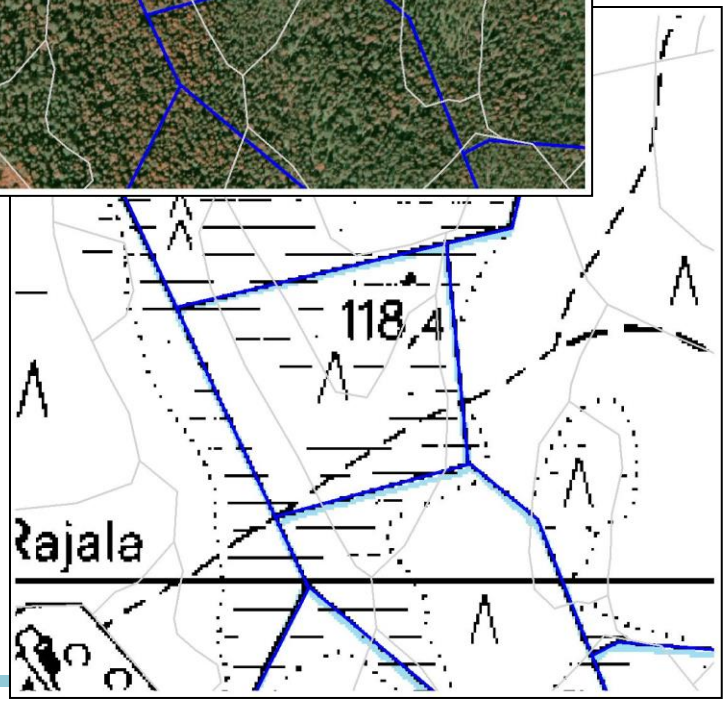


- Distance from roads and effect to growth (Lukkala 1929, Seppälä 1969, Miina 1994
- Lukkala (1929): when distance< 5 m from ditch -> growth=9.6 mm and when distance 6-20 m from ditch -> growth=6.2mm

Sample: Growth model

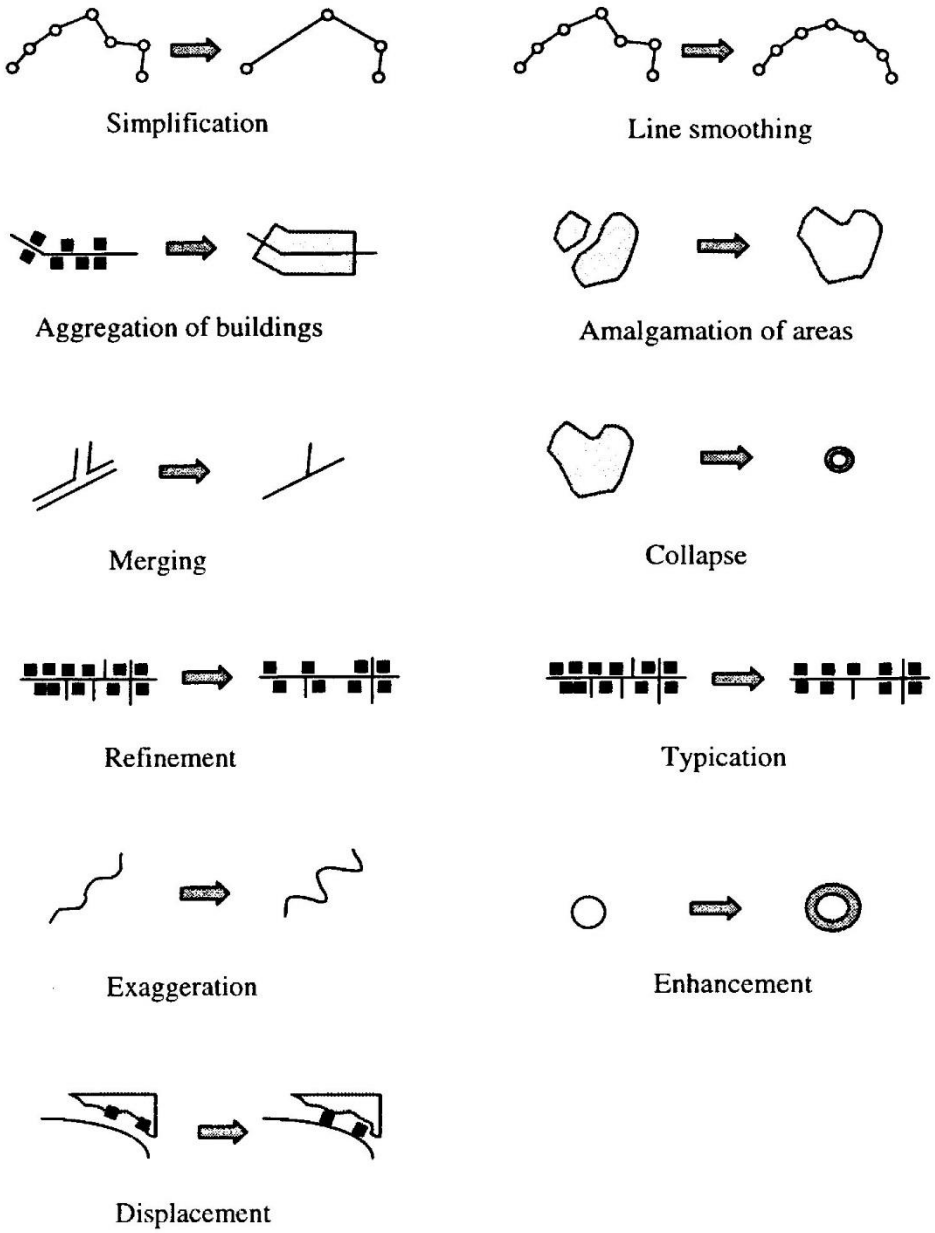
- Distance form ditch can be calculated in GIS system
- Steps:
  1. Extract ditches from topographic database to single layer
  2. Generate buffer
  3. Overlay buffer and stand register
  4. Create different values of growth attributes from buffer zones and other areas
  5. Summarize zonal sum back to stand border level



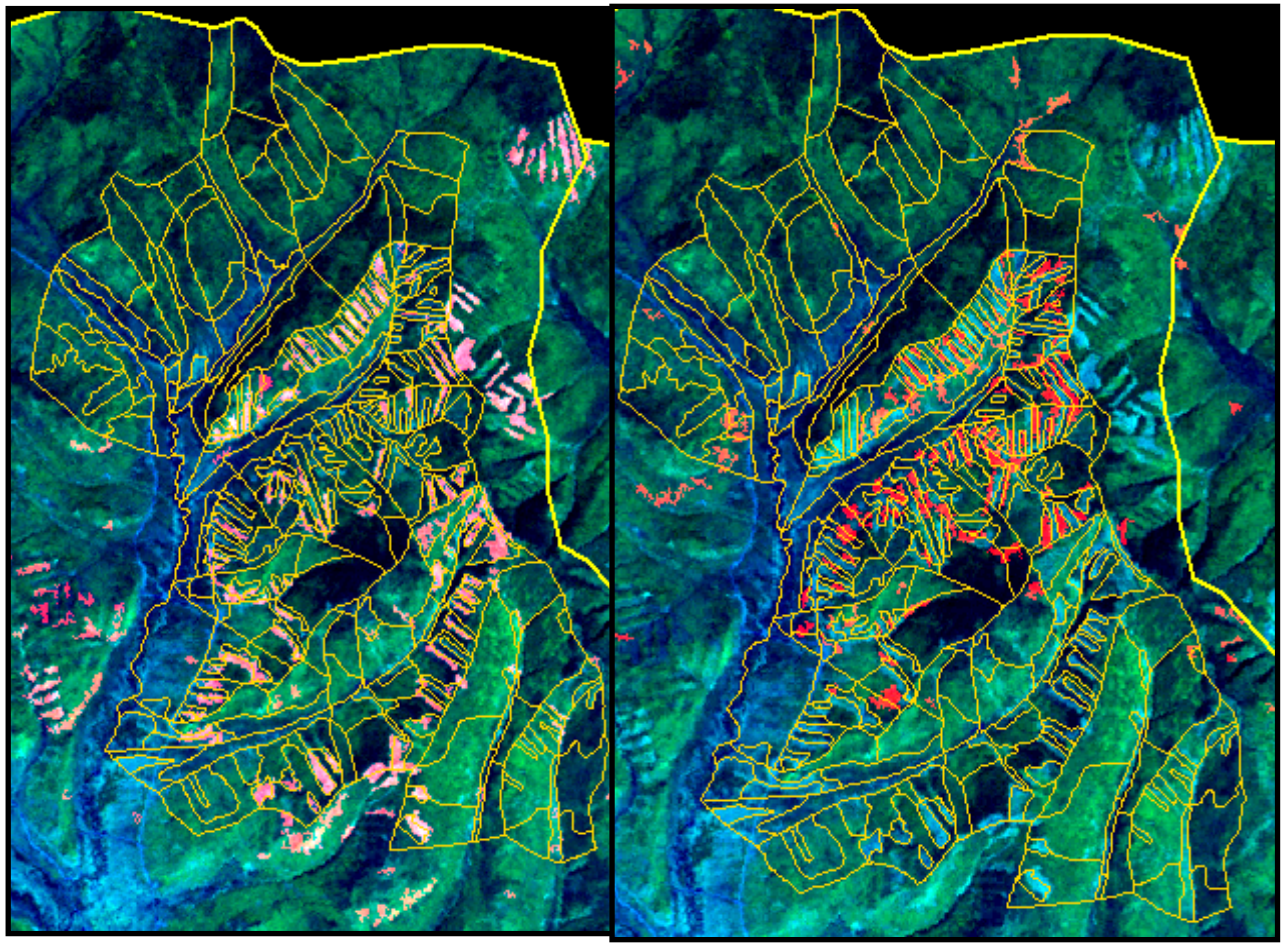


Generalisation

- For visualisation
- Operations in the picture

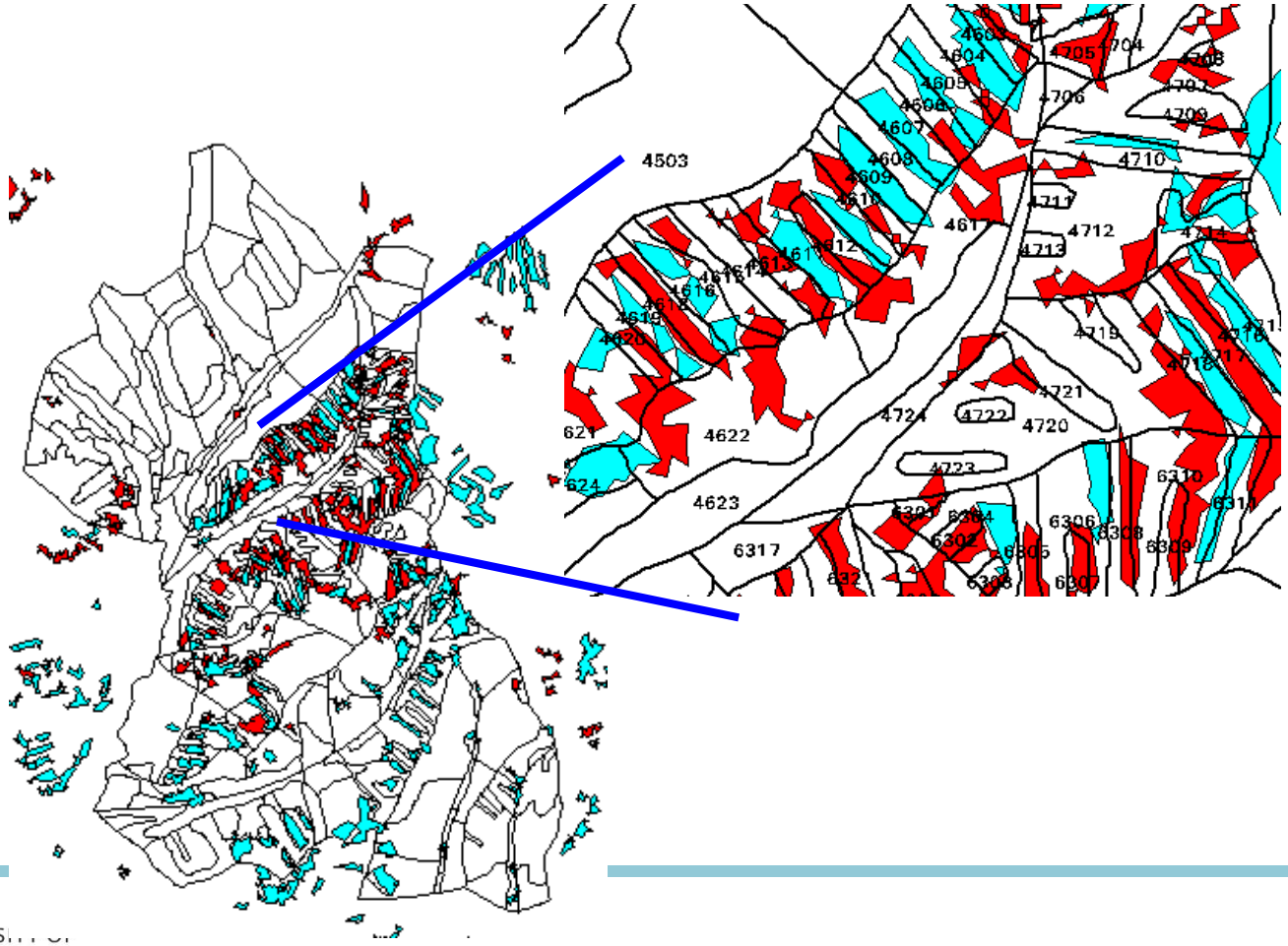


Landsat TM based forest inventory in China



- Feasibility study conserning to forest inventory alternatives in pilot area.
- Landsat TM-imagery were used to demonstrate alternative methods.

Vectorized results of the unsupervised change detection with AutoChange

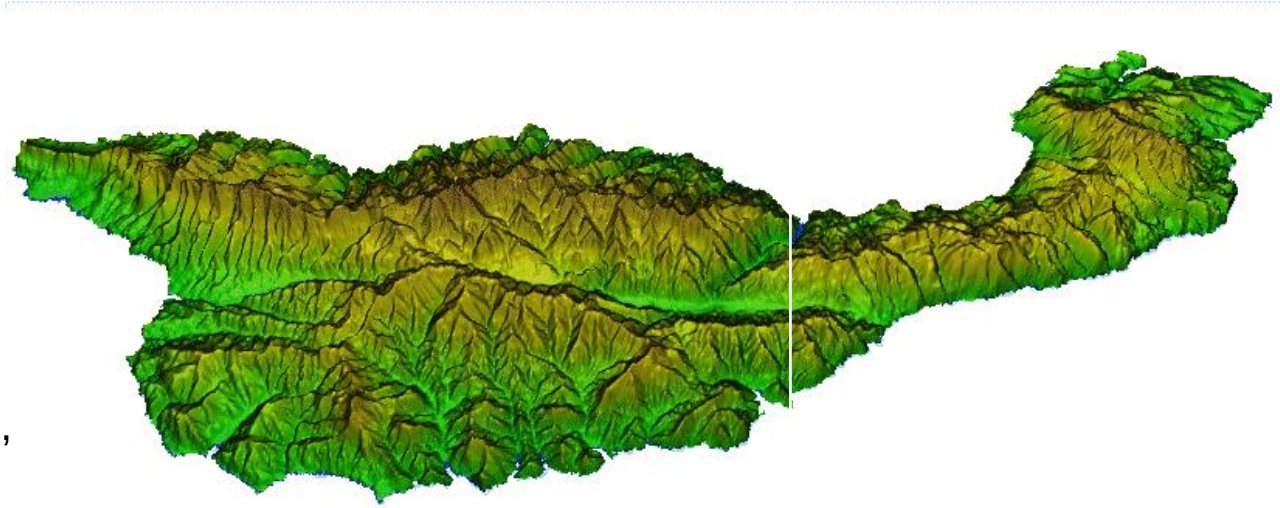


Raster GIS Analysis

Terrain analysis with DEM (digital elevation model)

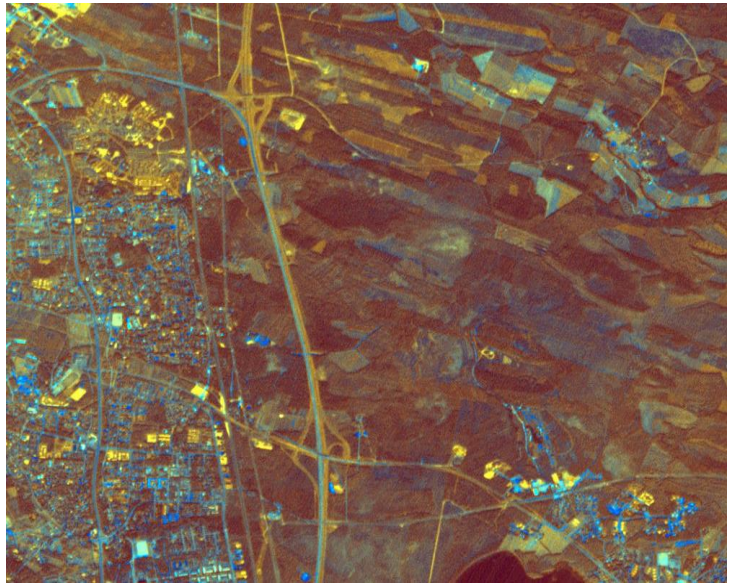
- Slope
- Aspect
- Direction
- Hillshade
- Visibility
- Hydrology Tools:

watershed, flow direction, flow accumulation…



Change analysis

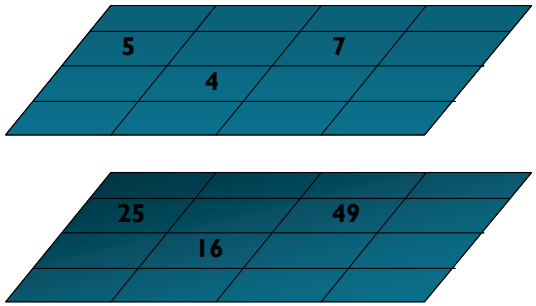
- Time series
  - Snapshots from each timestamp
  - Large changes are visible
  - Classification problems
- Movement map
  - Location is certaing times, eg. Forest fire movement
- Change map
  - Change maps are done
  - Small changes are visible
  - Difficult to understand overal situation



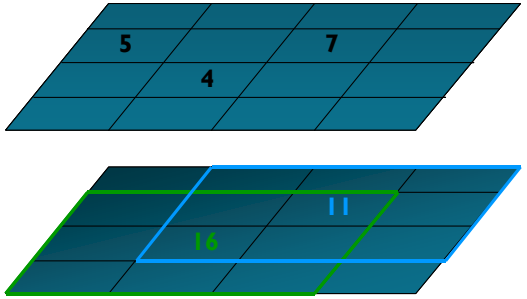
Map algebra and locality of analysis

- Map analysis is utilising map-algebra (Raster Calculator), which is defined using arithmetic operators and spatial functions.
- Four categories of analysis:
  - Local functions
  - Focal functions
  - Zonal functions
  - Global functions

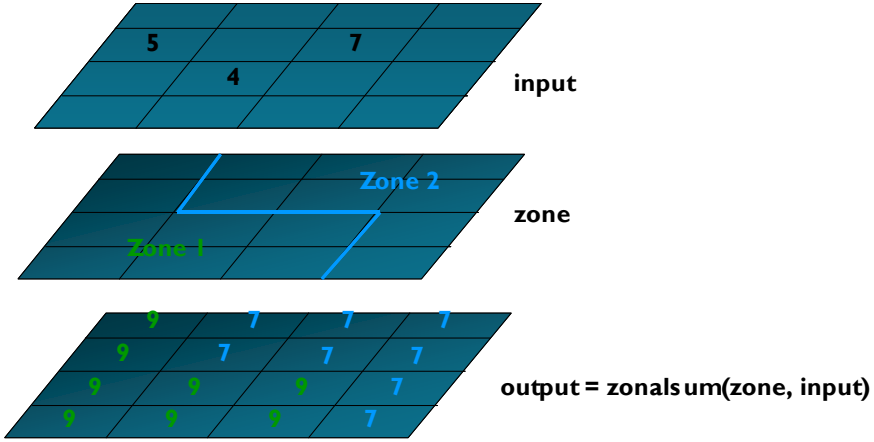
Local



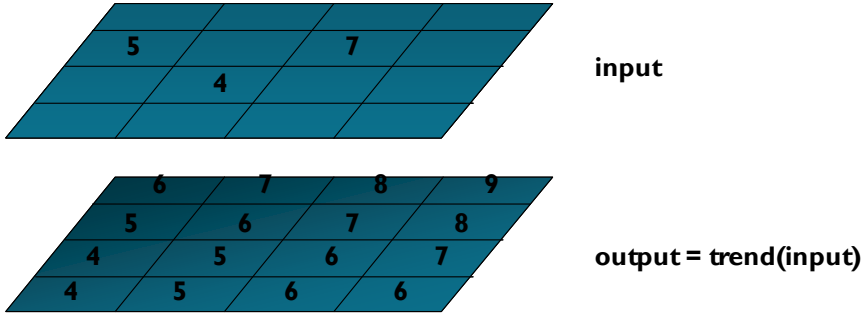
Focal



Zonal



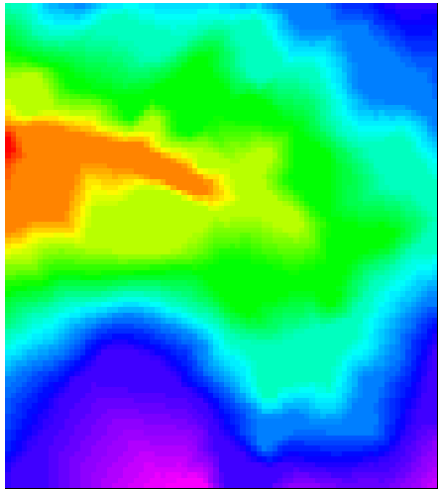
Global



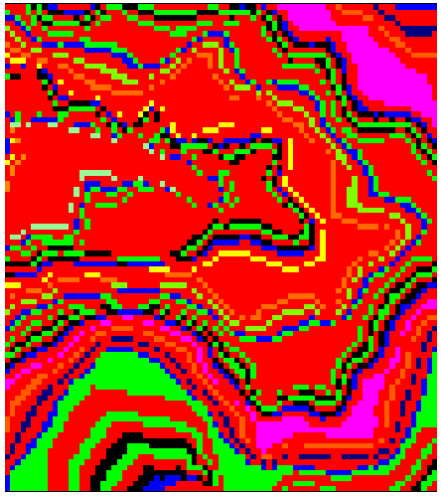
Local functions

- Output value of each cell is a function of the corresponding input value at each location
  - value NOT location determines result
  - e.g. arithmetic operations and reclassification
  - full list of local functions in GRID is enormous
    - Trigonometric, exponential and logarithmic
    - Reclassification and selection
    - Logical expressions in GRID
    - Operands and logical operators
    - Connectors
    - Statistical
    - Other local functions

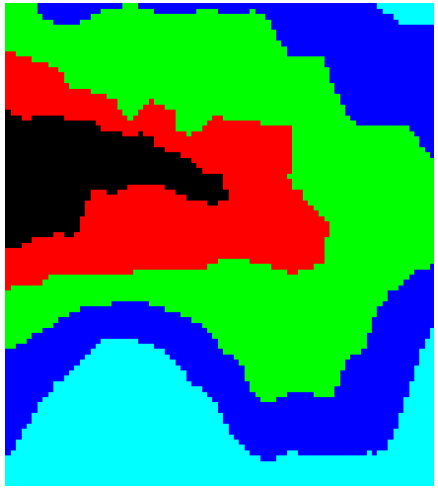
Some examples



input



output = tan(input)



output = reclass(input)

output = log2(input)

input grid

24	50	7	41	32
30	27	8	22	39
14	16	21	16	41
38	6	44	8	7
36	7	32	30	32
38	20	28	28	21
32	2	13	35	49

output grid

32.8	24.3	25.8	24.8	33.5
26.8	21.9	23.1	25.2	31.8
21.8	22.7	18.7	22.9	22.2
19.5	23.8	20.0	25.7	22.3
24.2	27.7	22.6	25.6	21.0
22.5	23.1	21.7	29.8	32.5
23.0	22.2	21.0	29.0	33.3

focal mean

(27 + 8 + 22 + 16 + 21 + 16 + 6 + 44 + 8) / 9 = 18.7

input grid

53	57	33	10	14
78	31	12	22	55
32	9	9	85	26
6	54	33	85	94
75	25	76	49	27
48	16	67	23	89
36	46	82	97	74
45	86	44	42	35

zone grid

1	1	1	2	2
1	1	1	2	2
1	1	1	2	2
5	5	1	2	2
5	5	1	2	2
5	5	10	91	91
5	5	10	91	91
5	5	10	10	10

output (zonal sum) grid

423	423	423	467	467
423	423	423	467	467
423	423	423	467	467
437	437	423	467	467
437	437	423	467	467
437	437	270	283	283
437	437	270	283	283
437	437	270	270	270

zonal sum

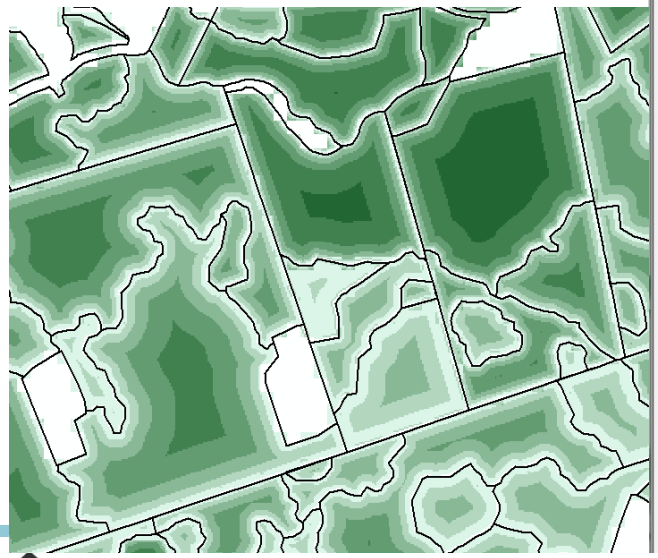
(53 + 57 + 33 + 78 + 31 + 12 + 32 + 9 + 9 + 33 + 76) = 423

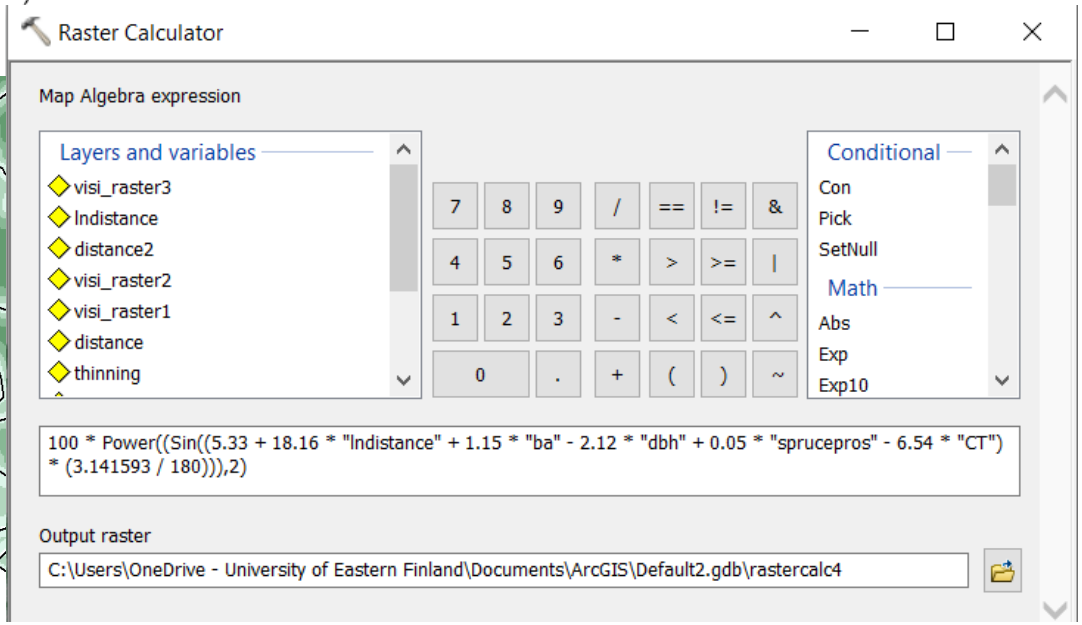
Grid analysis: Map algebra (or “How it works”)

- Map Algebra arithmetic:

Example of calculating the visibility of forest stands

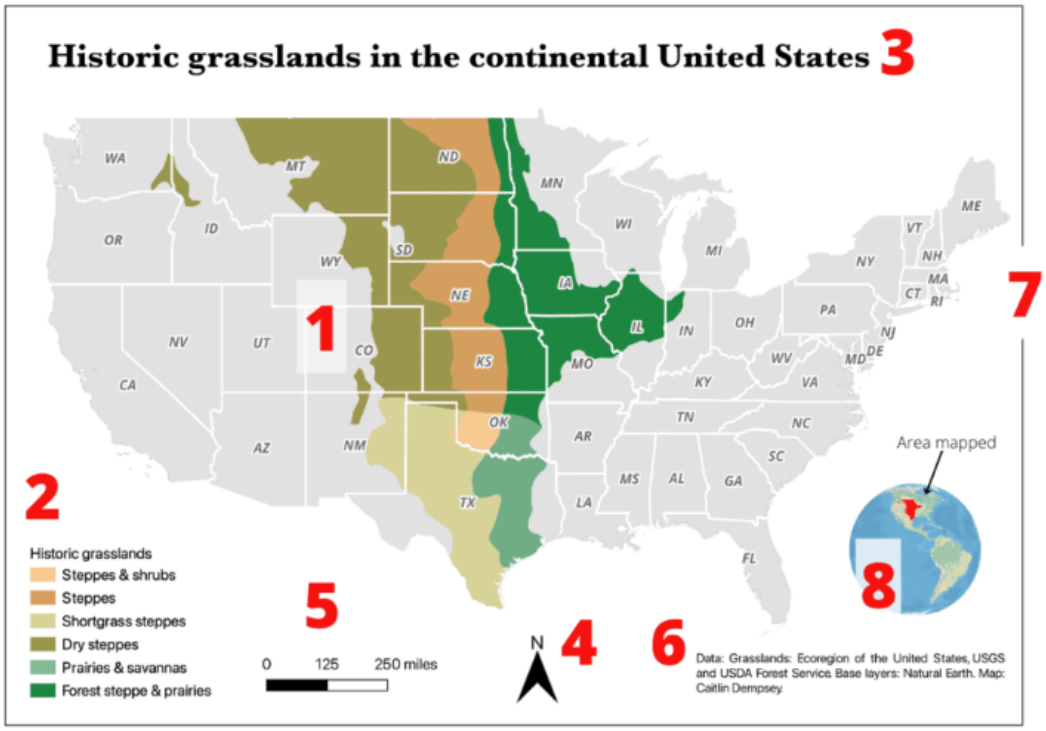
$$Visibility = 100\left(Sin\left(\left(\sum_{i=1}^{n}(k_i M_i) + c\right) * \frac{\pi}{180}\right)\right)^2$$





Map design - Reporting

- Thematic maps and statistics
- Map components need to be planned
  - Objects, symbols
  - Purpose of map
- Objects
  - Lines, areas, legends
- Visualisation
  - Scale, projection, north arrow, general description



A sample map showing the different elements in a map layout. Map: Caitlin Dempsey.

1. Map / data frame
2. Map legend
3. Map title
4. North arrow
5. Map scale bar
6. Metadata / map citation
7. Border
8. Inset / locator map

Map design - Use of colors in maps (Imhof 1962)

1. Clear and strong colors are not good, when they cover large areas.
2. Bright and clear colors near the white areas are also not recommended in large extent
3. Large area and basic colors are best when they quite neutral and grey-smoothed versions. They allow visibility of small bright highlighted objects.
4. Do not use two large areas only. You can easily mix one color to all objects.
5. Color design normally utilise basic colors, smooth colors and harmonise differences.