L3: Basic Analysis Tools

Thomas Gumbricht
thomas@karttur.com
www.karttur.com

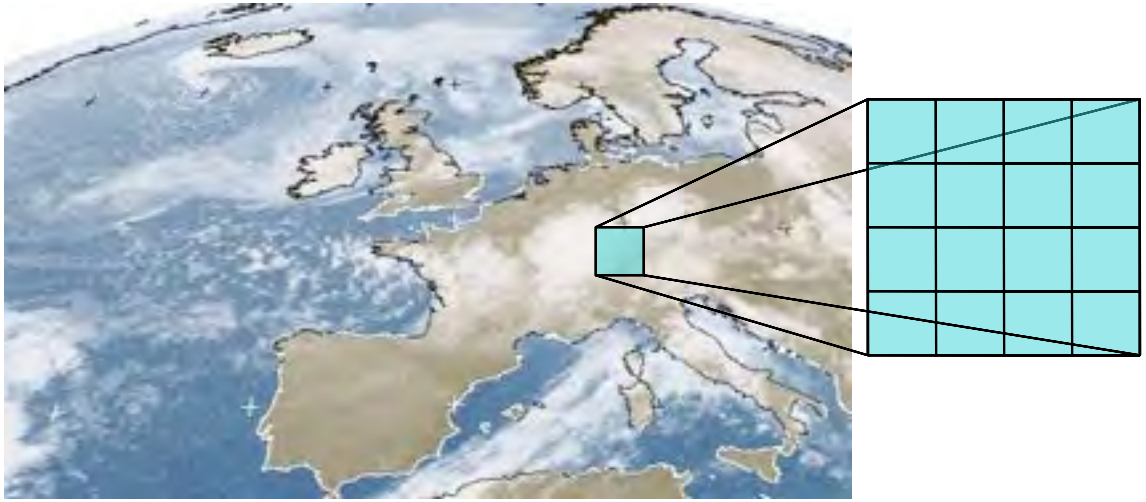
Georeferencing

- how data are linked with their geographic locations

Basic Spatial Analysis methods

- queries in geospatial databases
- transformations
    - buffering
    - point in polygon
    - polygon overlay
    - spatial interpolation
- raster analysis
    - map algebra
    - DEM, slope and aspect

Georeferencing



linking information to the Earth's surface → essential in GIS

Georeferencing must be:

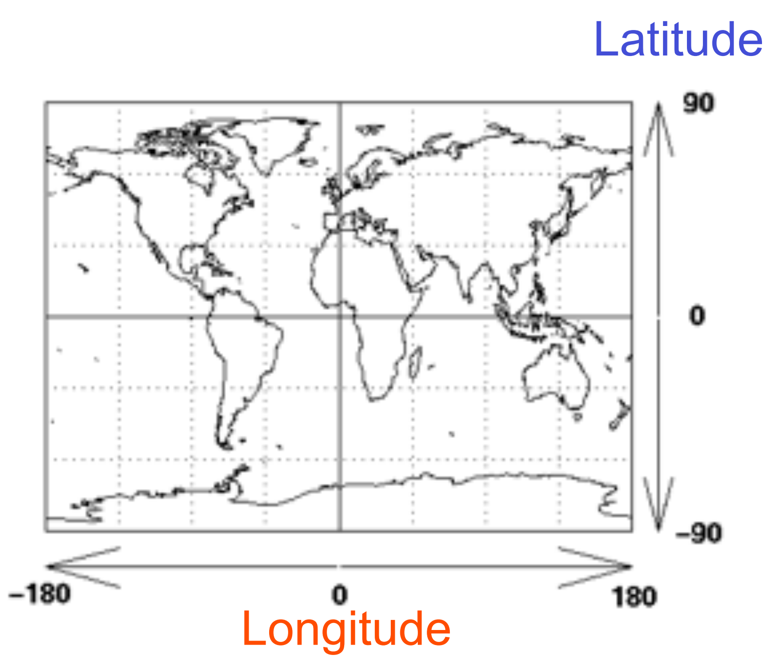
- unique, linking information to exactly one location
- shared, so different users understand the meaning of a georeference
- persistent through time, so today's georeferences are still meaningful tomorrow

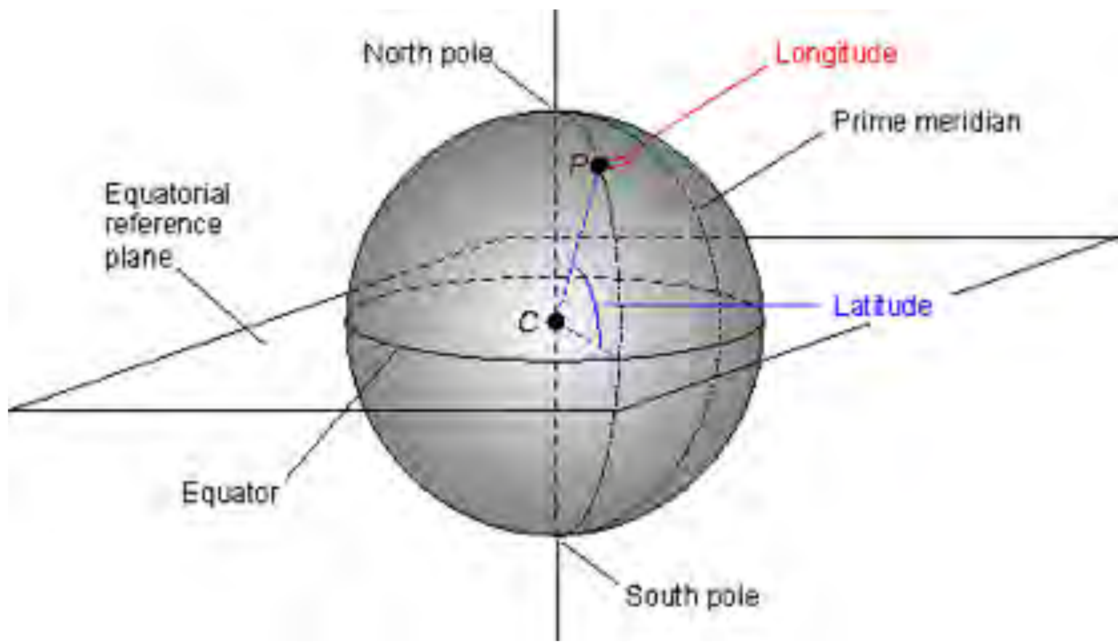
Types of georeferencing

- Some georeferences are **metric**
  - They define location using measures of distance from fixed places
    e.g., distance from the Equator or from the Greenwich Meridian
- Others are **based on ordering**
  - e.g. street addresses in most parts of the world order houses along streets
- Others are only **nominal**
  - Placenames do not involve ordering or measuring

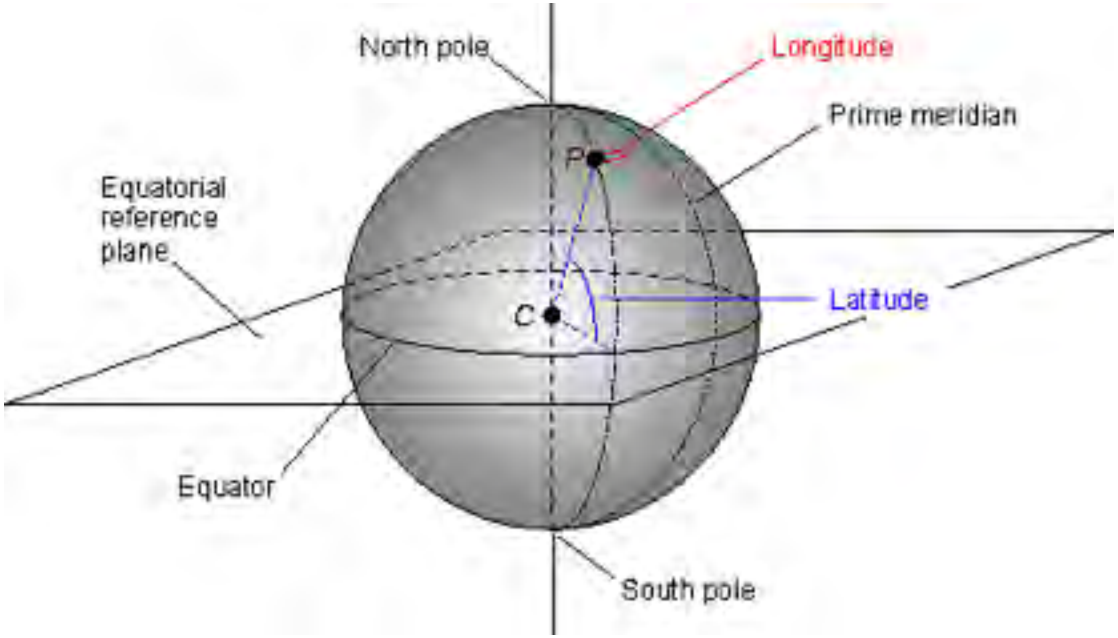
Metric georeferencing by latitude and longitude

- The most comprehensive and powerful method of georeferencing
  - Metric, standard, stable, unique
- Uses a well-defined and fixed reference frame
  - Based on the Earth’s rotation and center of mass, and the Greenwich Meridian





Definition of longitude



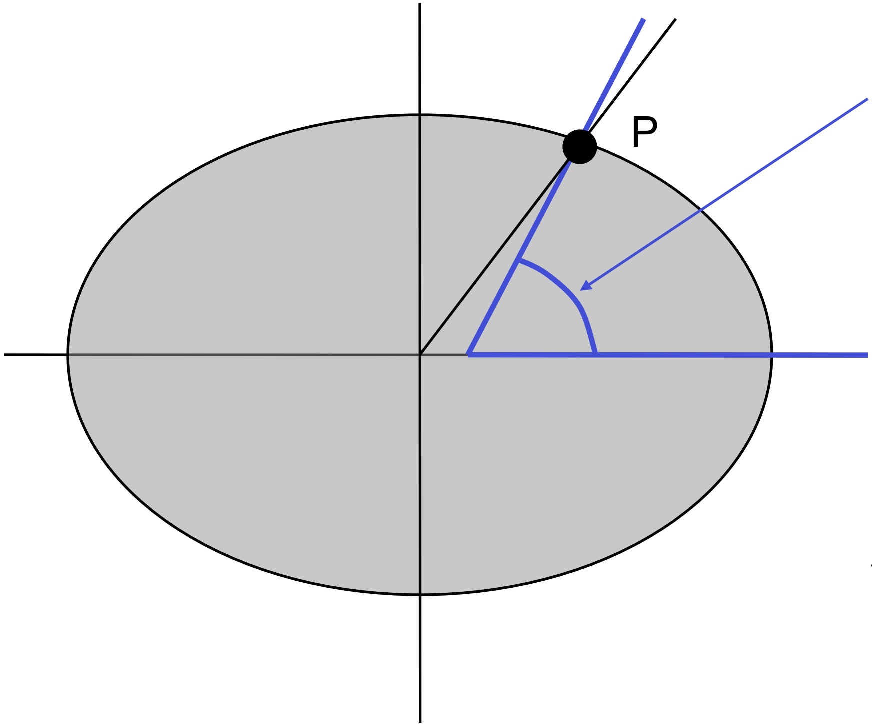
The location of Greenwich defines the Prime Meridian. The **longitude** of the point P is determined by drawing a plane through it and the central axis, and measuring the angle between this plane and the Prime Meridian.

Definition of latitude

- Requires a model of the Earth's shape
- The Earth is somewhat **elliptical**
  - The N-S diameter is roughly 1/300 less than the E-W diameter
  - More accurately modeled as an ellipsoid than a sphere
  - An **ellipsoid** is formed by rotating an ellipse about its shorter axis (the Earth's axis in this case)

National ellipsoids – each country has its own best approximation of the Earth's surface

An international standard ellipsoid (WGS 84)



Latitude (of the point P) is the angle between a line that is perpendicular to the surface of the ellipsoid and the plane of the Equator

WGS 84 ellipsoid

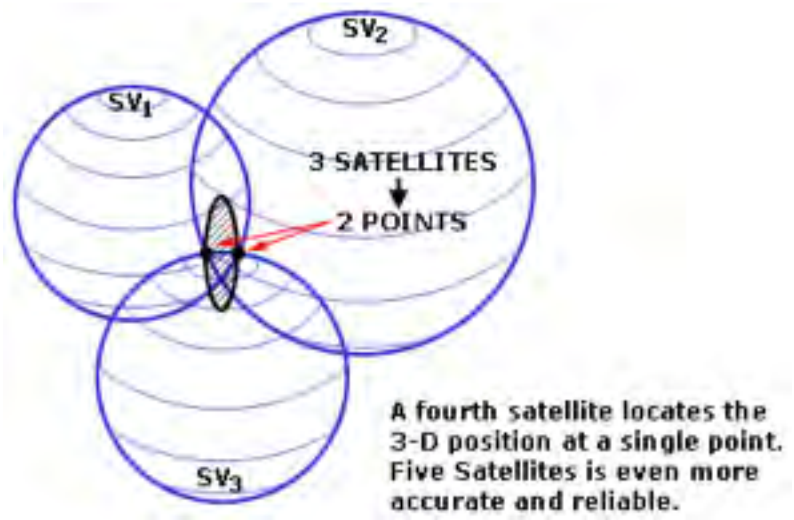
- radius of the Earth at the Equator 6378.137 km
- flattening 1 part in 298.257

Measuring the longitude/latitude:

GNSS (Global Navigation Satellite Systems)

Collections of satellites:

GPS (USA) – Global Positioning System
GLONASS (Russia)
Galileo (EU) – under development



Direct, accurate measurements of latitude and longitude

GPS: 10m accuracy from a simple, cheap unit

- differential GPS capable of sub-meter accuracy
- Sub-centimetre accuracy if observations are averaged over long periods

Field data collection with a GPS and external antenna

Constellation of GPS satellites

Galileo satellite

Georeferencing by place names

the earliest form of georeferencing

the most commonly used in everyday activities

Many names of geographic features are **universally recognised**, others may be understood **only by locals**.

Stockholm, London, New York

Lill-Jansskogen, Uggleviken, Siska

Names work **at many different scales**.

From continents to small villages and neighborhoods

Names may **pass out of use** in time

Where is Östra Aros?

Renamed to Uppsala in 1273, when the seat of the archbishop was moved there from Gamla Uppsala.

Different names in different languages for the same place

Beč (Croatian, Serbian)
Bécs (Hungarian)
Dunaj (Slovenian)
Vena (Russian)
Vídeň (Czech)
Viena (Portuguese, Spanish)
Vienna (Italian, English)
Vienne (French)
Wiedeń (Polish)
Wien (German, Swedish)

Benátky (Czech, Slovak)
Benetke (Slovenian)
Velence (Hungarian)
Venecia (Spanish)
Venedig (Danish, German, Swedish)
Venezia (Italian)
Venise (French)
Wenecja (Polish)

Georeferencing by postal addresses and postal codes

Why are postal addresses and codes useful:

- Every dwelling and house is a potential destination for mail
- Dwellings and houses are usually arrayed along streets, and numbered accordingly
- Streets have names that are unique within local areas
- Local areas have names that are unique within larger regions

↓

If these assumptions are true, then a postal address is a useful georeference

Lakes, mountains, and rivers cannot be located using postal addresses

Urban-style addresses have been extended recently to many rural areas

For natural features

In rural areas

When does georeferencing by postal addresses and codes fail?

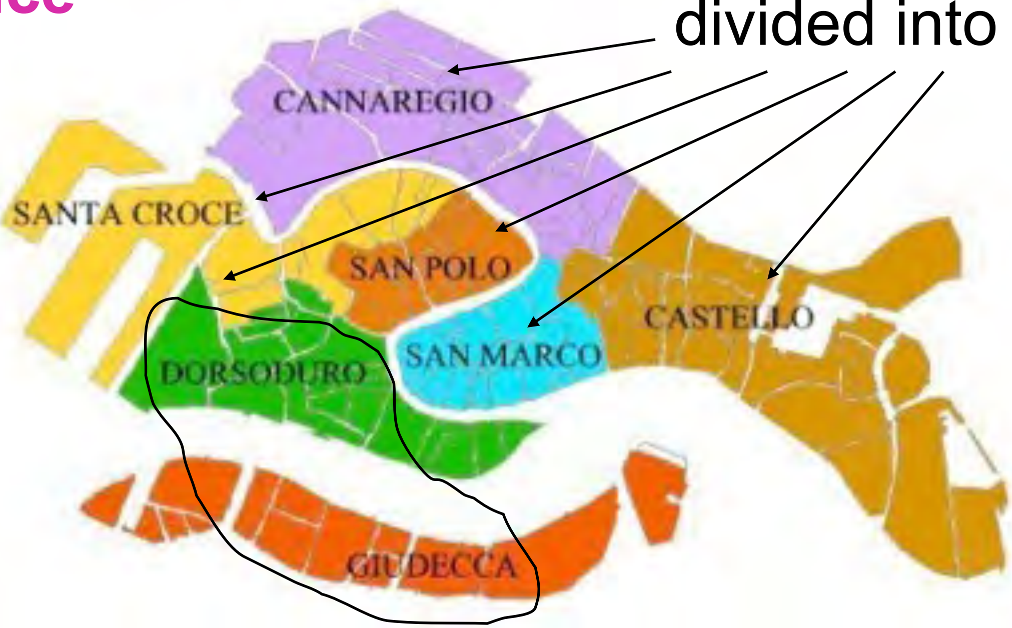
When numbering on streets is not sequential

Japan

Venice

Venice

divided into six districts = **sestieri**



The word is from *sesto*, or sixth; and is used only for Italian towns divided into six districts.

house-numbering **by district**

only six series of numbers for the entire city

Each house has a unique number in the district, from one to several thousand, generally numbered in the order of time when the house was built, which means that the numbers are not distributed in a readily understandable manner.

Basic Spatial Analysis

Spatial Analysis → Turns raw data into useful information

→ Reveals patterns, trends, and anomalies that might otherwise be missed

→ Helps the human intuition in situations where the eye might deceive

A **method of analysis** is **spatial** if the results depend on the **locations** of the objects being analysed:

- the results change if you move the objects (results are not invariant under relocation)

→ necessary attributes → geometric → location

necessary attributes → non-spatial

Spatial analysis using queries

Querying databases

→ Structured Query Language - SQL

SQL is a standard language used to analyse relational databases.

Table = relational database = Object Class

add Geometry

Feature Class = geospatial database

Column = property

Row = object

Attributes of STATES

FID	Shape*	AREA	STATE_NAME	STATE_FIPS
41	Polygon	51715.656	Alabama	01
49	Polygon	576556.687	Alaska	02
35	Polygon	113711.523	Arizona	04
45	Polygon	52912.797	Arkansas	05
23	Polygon	157774.187	California	06
30	Polygon	104099.109	Colorado	08
17	Polygon	4976.434	Connecticut	09
27	Polygon	2054.506	Delaware	10
26	Polygon	66.063	District of Columbia	11
47	Polygon	55815.051	Florida	12
43	Polygon	58629.195	Georgia	13
48	Polygon	6381.435	Hawaii	15
7	Polygon	83340.594	Idaho	16
25	Polygon	56297.953	Illinois	17
28	Polygon	36399.516	Indiana	18
12	Polygon	56257.219	Iowa	19
32	Polygon	82195.437	Kansas	20
31	Polygon	40318.777	Kentucky	21
46	Polygon	45835.898	Louisiana	22
2	Polygon	32161.664	Maine	23
29	Polygon	9739.753	Maryland	24
13	Polygon	8172.482	Massachusetts	25
50	Polygon	57898.367	Michigan	26
9	Polygon	84517.469	Minnesota	27
42	Polygon	47618.723	Mississippi	28
34	Polygon	69831.625	Missouri	29
1	Polygon	147236.031	Montana	30

Record: 0 Show: All Selected Records (0 out of 51 Selected.)

Queries:

- the most basic analysis operations
- used to select information from one or several tables.

How many houses are on this street?

Which parcels are located within 200m from the coastline?

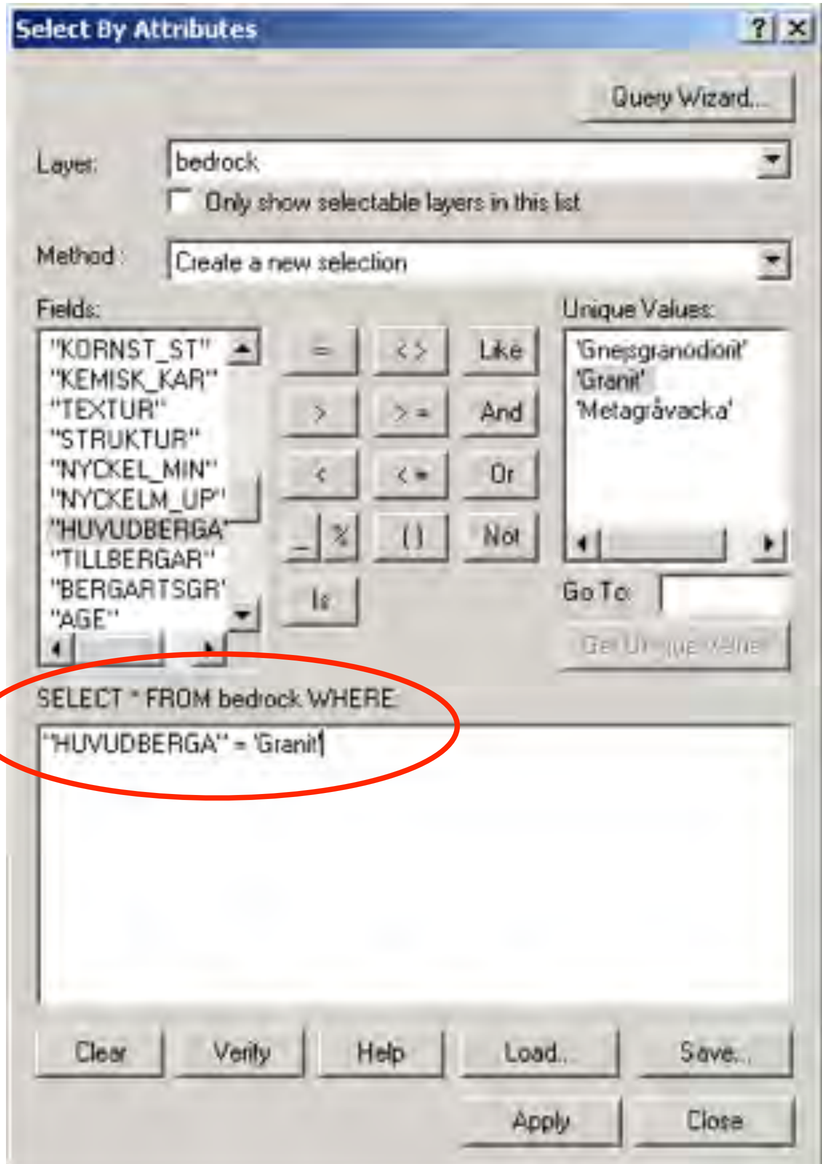
Which is the closest city from Stockholm going north?

SQL examples using a list of student names:

SELECT name FROM list
(selects all names)

SELECT name FROM list WHERE university = "KTH"
(selects names of students studying at KTH)

SELECT name FROM list WHERE grade > 3.0
(selects names of students with a grade greater than 3.0)



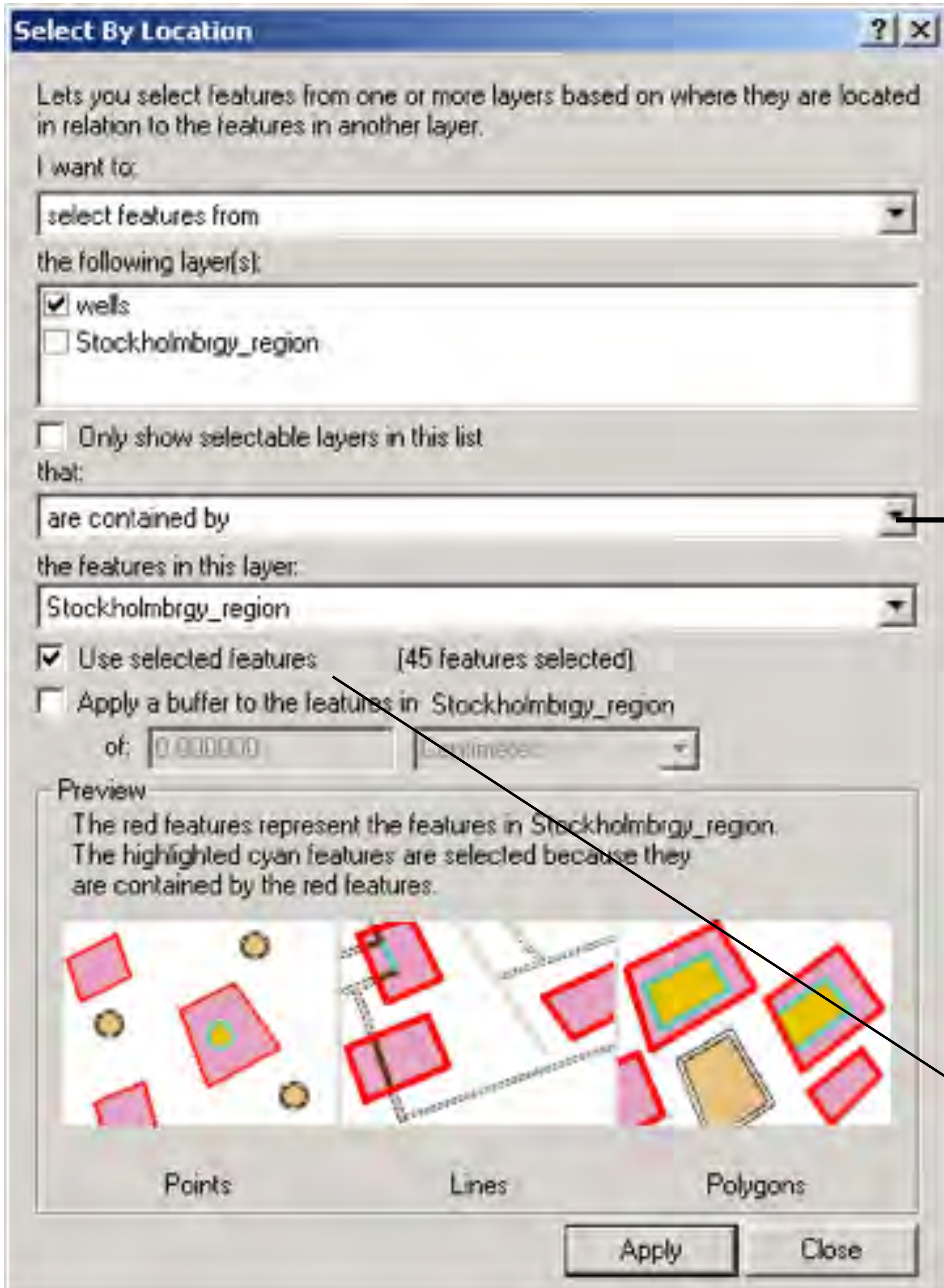
SQL example in ArcGIS: selecting the bedrock polygons according to attribute value

SELECT * FROM bedrock WHERE:
“HUVUDBERGA” = ‘Granit’

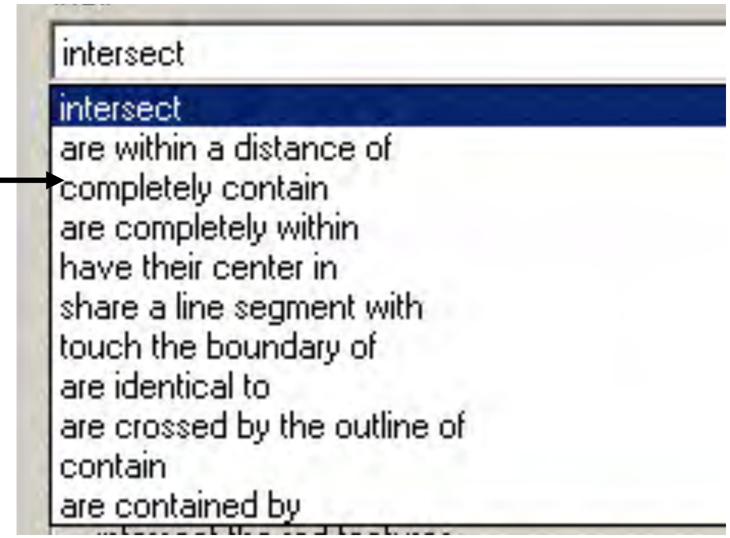
Spatial queries

SQL can be extended with spatial requirements such as:

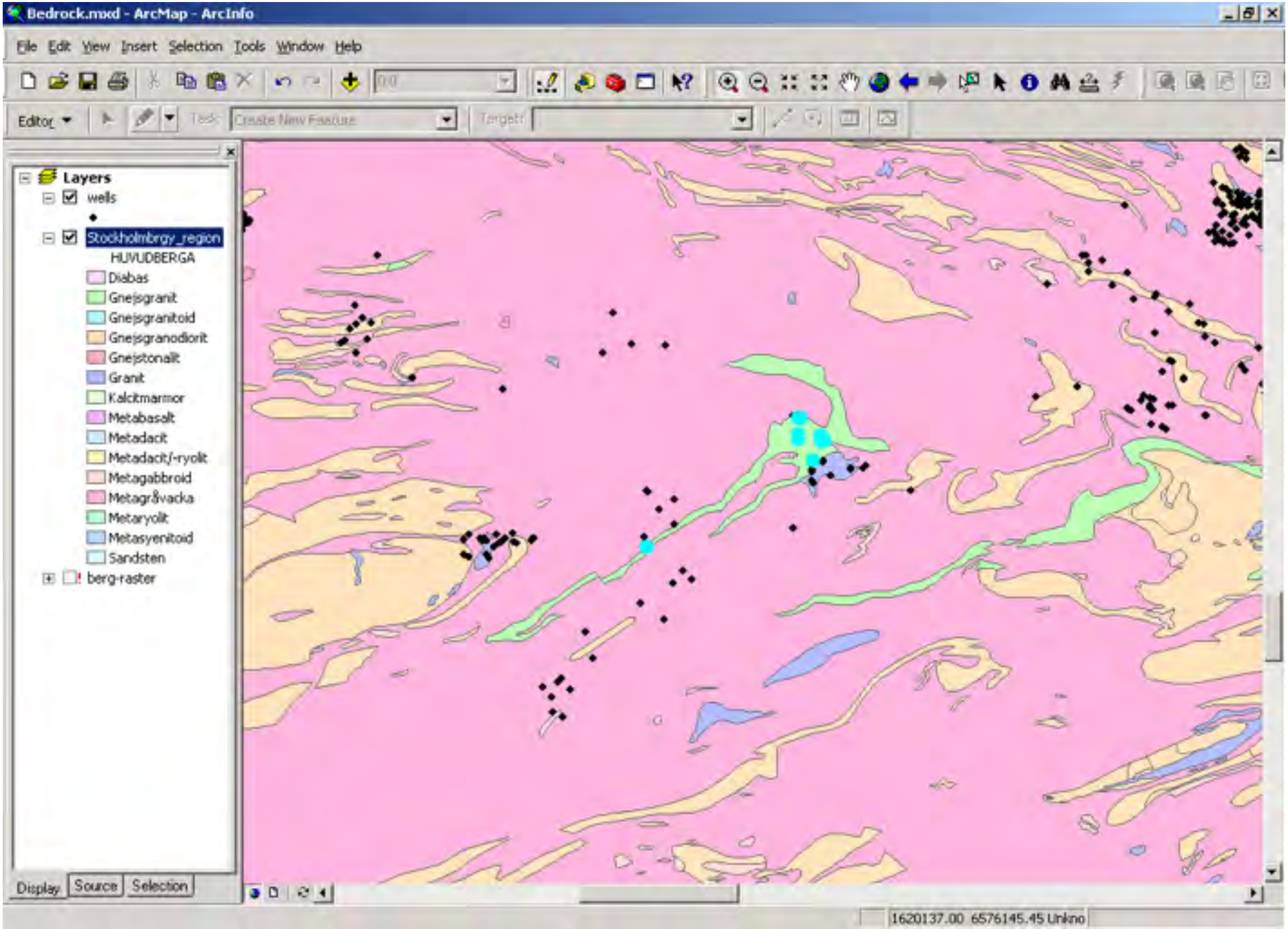
- Intersect
- Contain
- Have their centre in
- Are within a distance of …
- Touch the boundary of …
- etc.



Spatial requirements



SELECT * FROM bedrock WHERE:
“HUVUDBERGA” = ‘Gnejsgranit’


Bedrock.mxd - ArcMap - ArcInfo
File Edit View Insert Selection Tools Window Help
Editor
Task: Create New Feature
Target:
Layers
wells
Stockholmbrgy_region
HUVUDBERGA
Diabas
Gnejsgranit
Gnejsgranitoid
Gnejsgranodiorit
Gnejstonalit
Granit
Kalcitmarmor
Metabasalt
Metadacit
Metadacit/-ryolit
Metagabbroid
Metagråvacka
Metaryolit
Metasyenitoid
Sandsten
berg-raster
Display
Source
Selection
1620137.00 6576145.45 Unkno

Geometriska vektoroperationer

Beräkning av avstånd

Euklidiskt avstånd

$$d(1,2) = \sqrt{(x_1-x_2)^2 + (y_1-y_2)^2}$$

där

$d(1,2)$ är avståndet mellan puntkerna 1 och 2

punkt 1 har koordinaterna (x_1,y_1) och,

punkt 2 har koordinaterna (x_2,y_2).

Geometriska vektoroperationer

Beräkning av avstånd

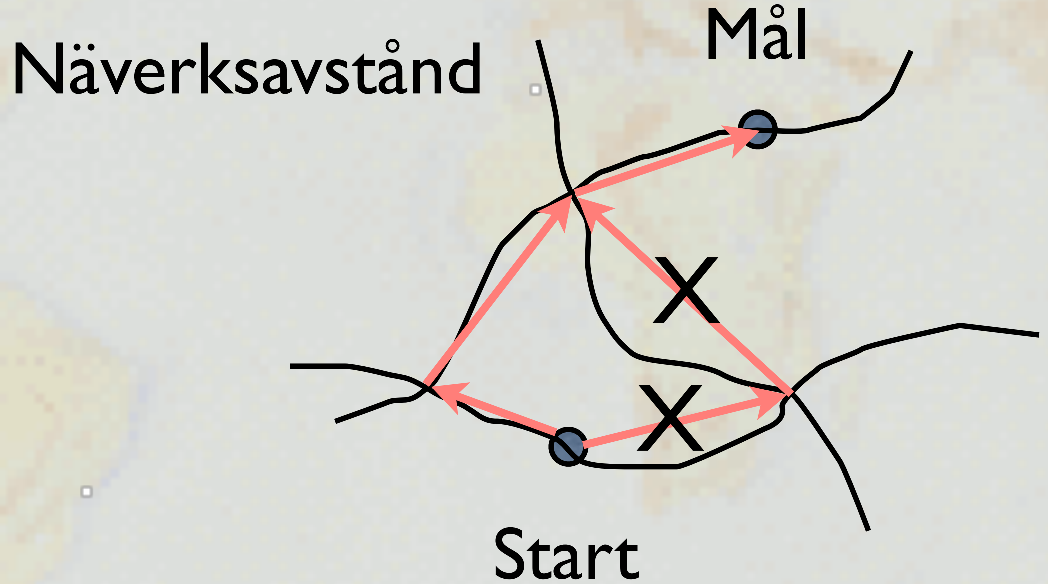
Manhattan avstånd

Euklidiskt avstånd

Manhattan avstånd

Geometriska vektoroperationer

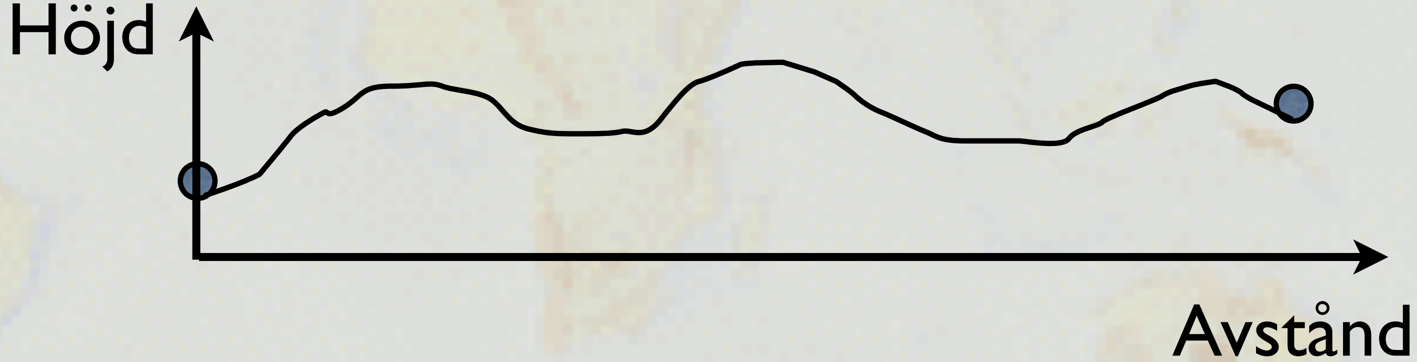
Beräkning av avstånd



Geometriska vektoroperationer

Beräkning av avstånd

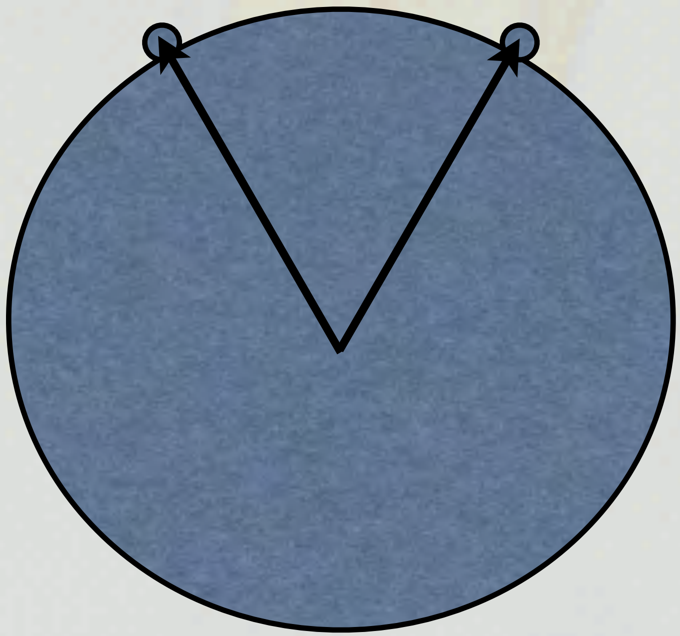
Topografiskt avstånd (över 3D yta)



Geometriska vektoroperationer

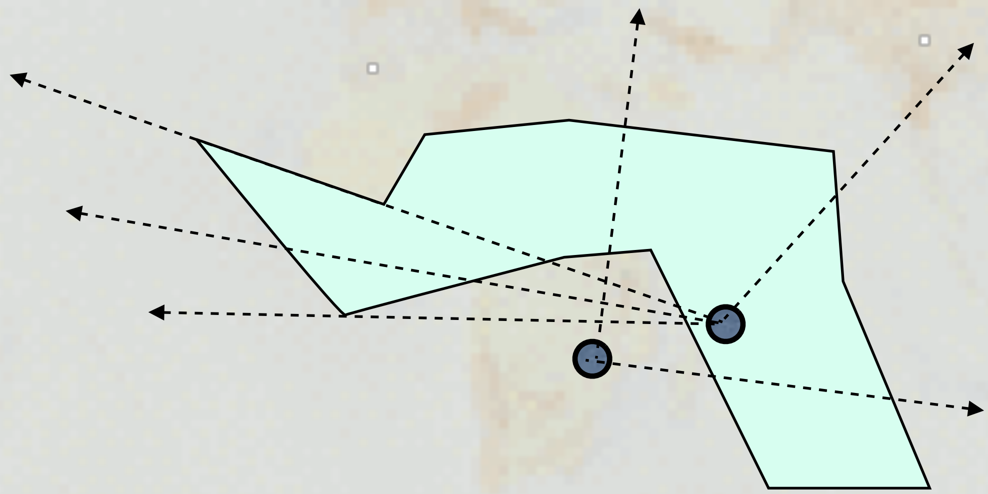
Beräkning av avstånd

Sfärsikt avstånd (med hänsyn till jordyans rundning)



Geometriska vektoroperationer

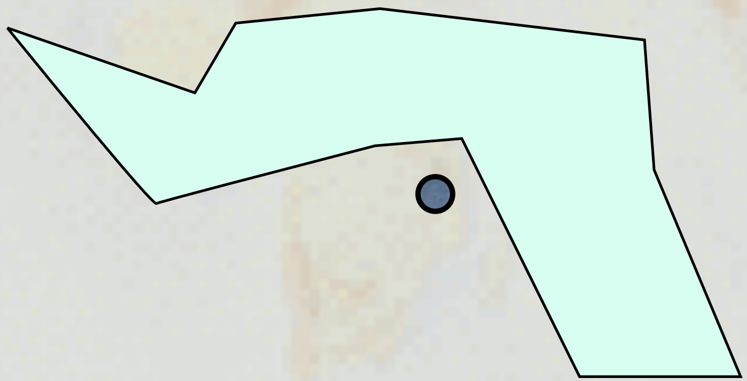
Polygontillhörighet



Om antalet passaer genom polygonens begränsning = ojämt antal, då ligger punkten inuti polygonen

Geometriska vektoroperationer

Beräkning av en polygons tyngpunkt eller centroid



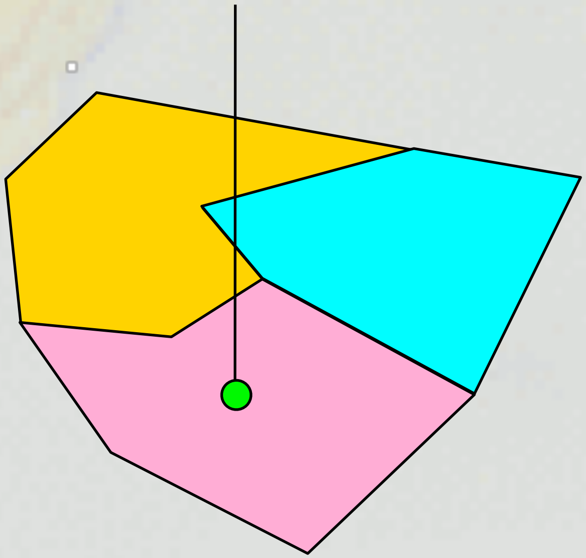
Överlagring i vektordata

Överlagring av punkter på polygoner

Först analyseras polygontillhörighet.

Sedan extraheras valda polygon attribut till punktens attributdata.

Exempel:
- hänföra kriminella aktiviteter till rätt polisdisktrikt
- hänföra röstberättigade till rätt valdistrikt



Överlagring i vektordata

Överlagring av linjer på polygoner

Först klipps linjeobjektet där det delas av polygonskiktet, och nya start- och stoppunkter läggs in. Till skillnad från överlagring av puntker måste en ny linje-vektor skapas.

Sedan extraheras valda polygon attribut (eller linje attribut) till det nya linje-objektets attribut-tabell.

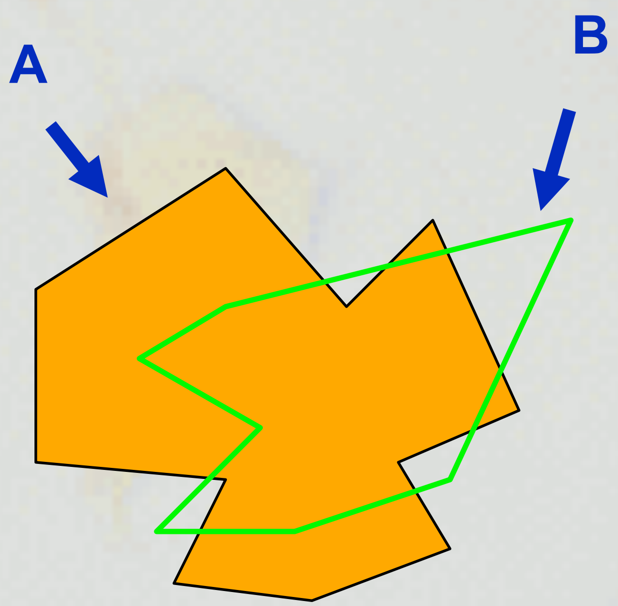
Exempel:

- Vattendragslängder i olika fastigheter
- väglängder i olika län

Överlagring i vektordata

Överlagring av polygoner på polygoner

Överlagring med diskreta objekt hittar inersektioner mellan två polygoner och skapar en ny polygon



I exemplet uppstår 9 new polygons vid intersektionen av polygon A och B.

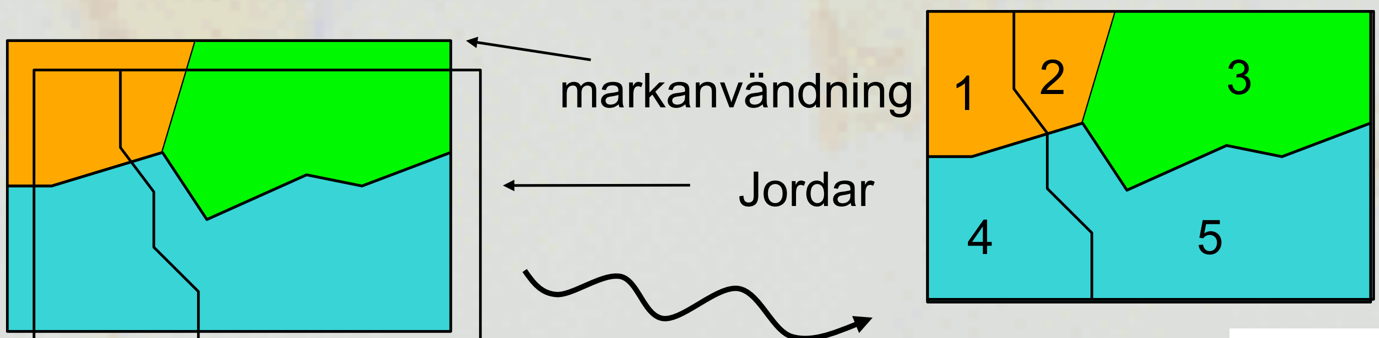
- En bildas gemensamt från A och B.

- Fyra bildas från polygon A men inte Polygon B.

- Fyra bildas från polygon B men inte polygan A.

Överlagring i vektordata

- Två överlappande polygon-lager, som representerar två klassificeringar över samma område
(jordarter och land markägare)
- Överlagringen skapar nya lager från alla kombinationer av intersektioner.
- Varje polygon i det nya lagret har både e jordart och och en markägare (konkatenerade attribut).
- Kan utföras i både raster och vektor



5 polygoner med 2 attribut

Polygon	Soil	Landuse
1	1	1
2	2	1
3	2	2
4	1	3
5	2	3

Transformations

Transformations – simple methods of spatial analysis

change datasets

compare datasets to obtain new datasets

Goal: obtain insight into the data and new knowledge

Spatial interpolation

Buffering

Transformations

Polygon overlay

Point in polygon

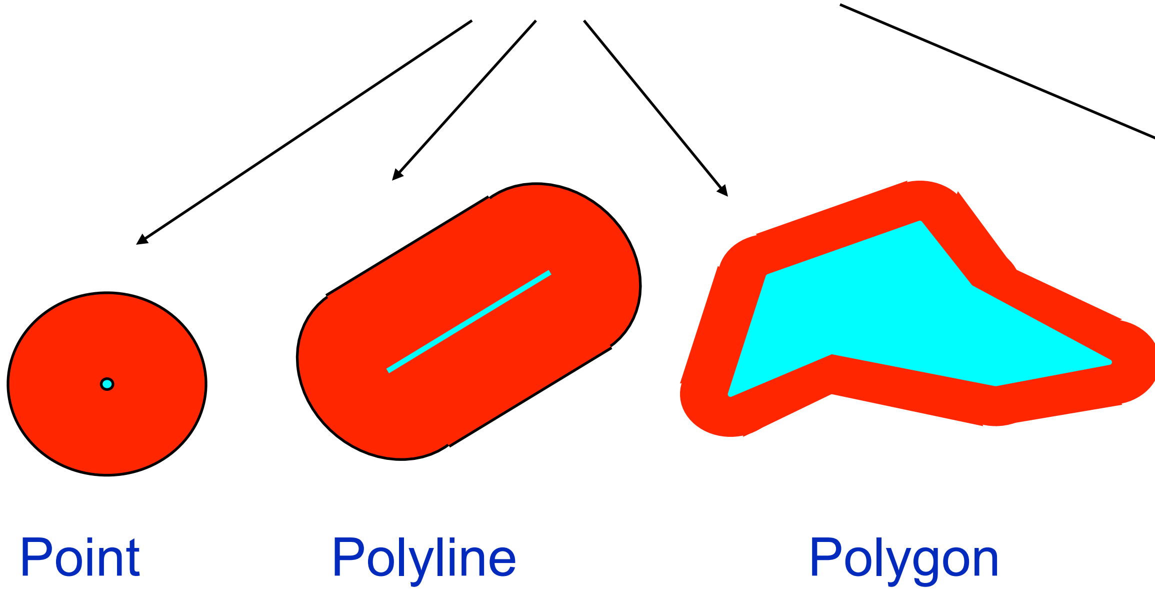
Buffering

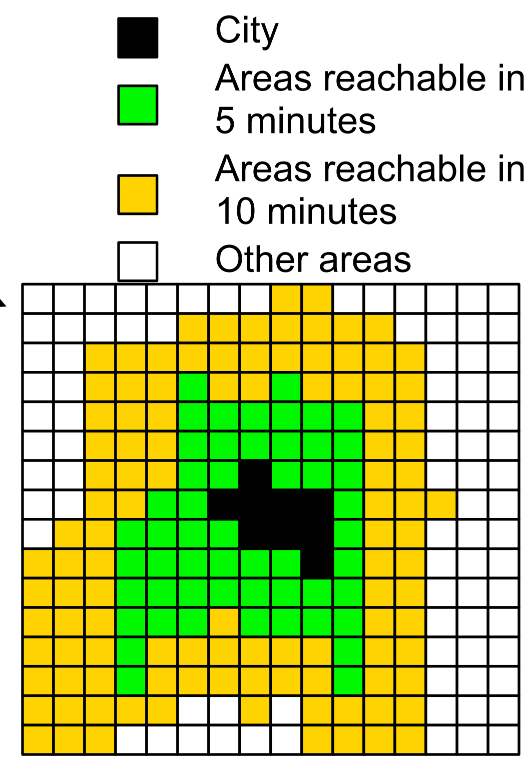
Buffer operation builds a new object consisting of areas within a user-defined distance of an existing object.

Why is this useful?

- to determine areas impacted by a noise pollution
- to determine the areas that would be affected by a flood

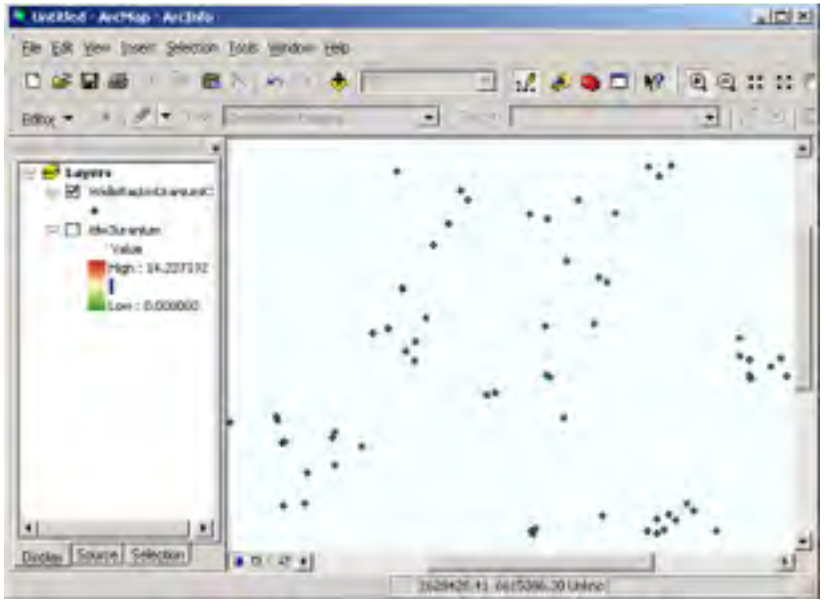
Feasible in either vector or raster mode





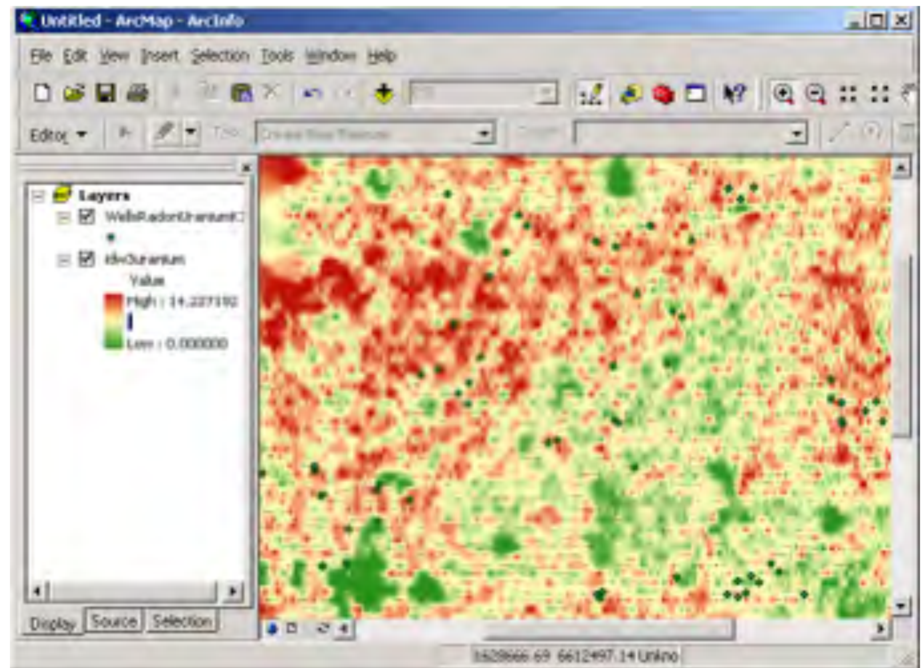
Spatial interpolation

How to estimate the complete field?



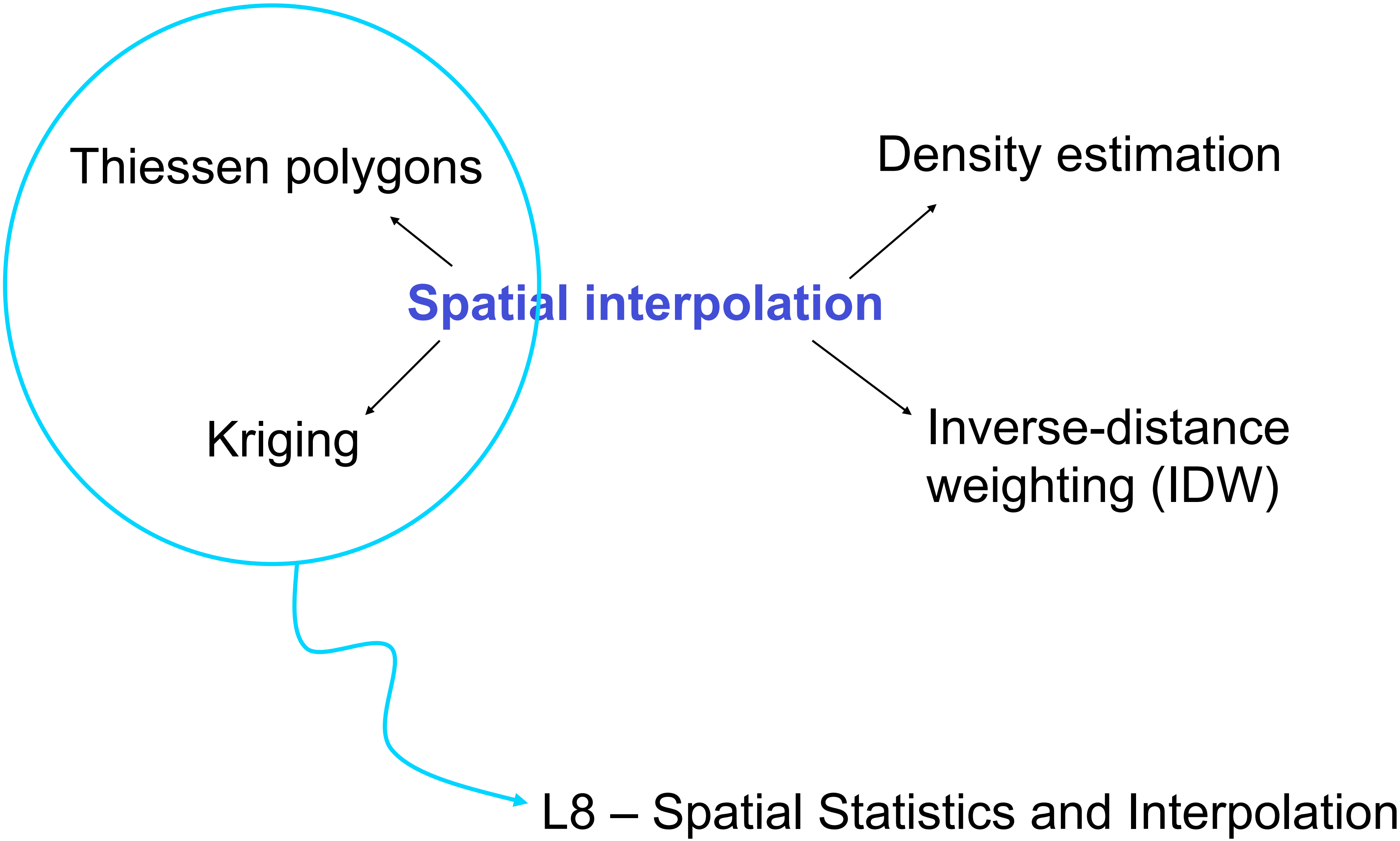
Values of a field have been measured at a number of sample points

Spatial interpolation



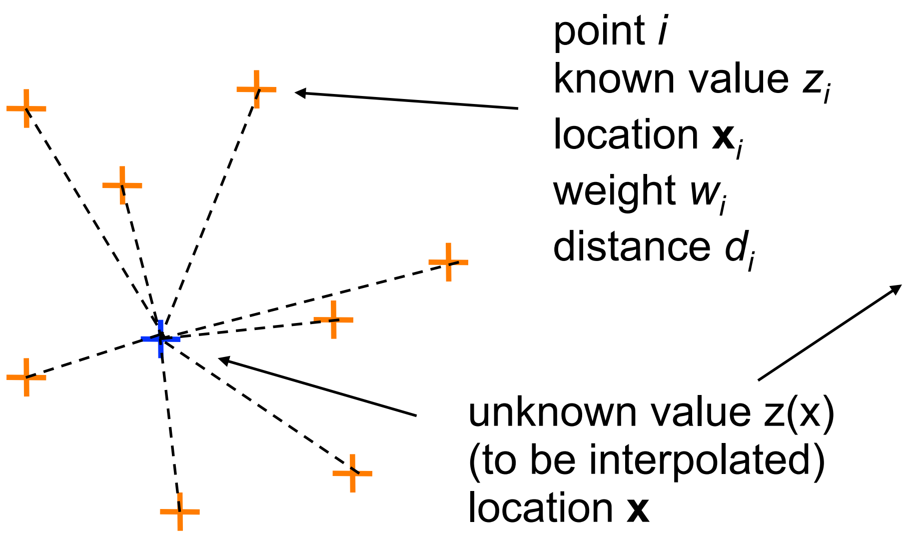
Why do we need this?

- to estimate values at points where the field was not measured
- to create a contour map by drawing isolines between the data points


Thiessen polygons
Density estimation
Spatial interpolation
Kriging
Inverse-distance weighting (IDW)
L8 – Spatial Statistics and Interpolation

Spatial interpolation by Inverse-distance weighting (IDW)

The unknown value of a field *z* at a point *x* is estimated by taking a weighted average over the known values:



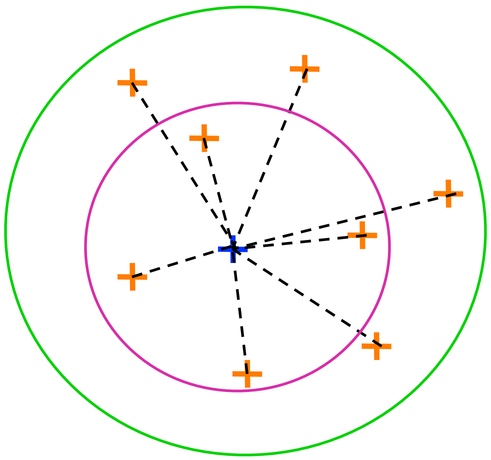
$$z(\mathbf{x}) = \sum_i w_i z_i \Big/ \sum_i w_i$$

$$w_i = 1 \big/ d_i^r$$

Each known value is weighted by its distance from the point x: weights decrease with the r^{th} power of distance (usually r=2).

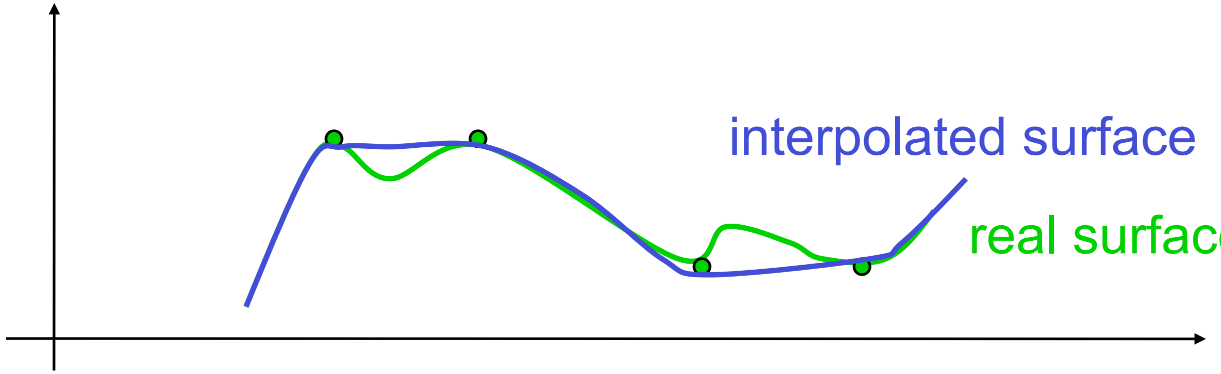
IDW issues

-What is a reasonable size of the search radius?
I.e. how many neighbouring known points should be included in the calculation?



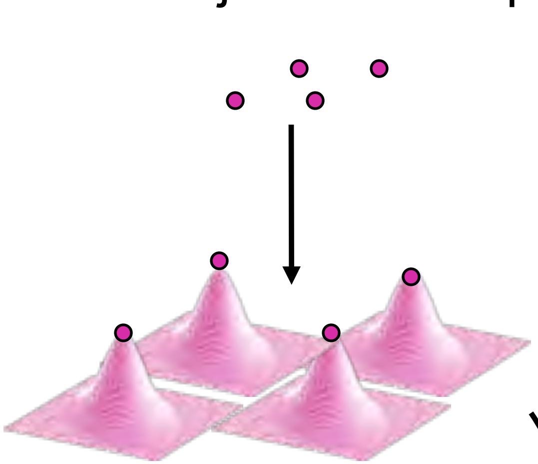
- The range of interpolated values cannot exceed the range of observed values:
sample points should include both min and max value.

-The method misses small peaks and pits:

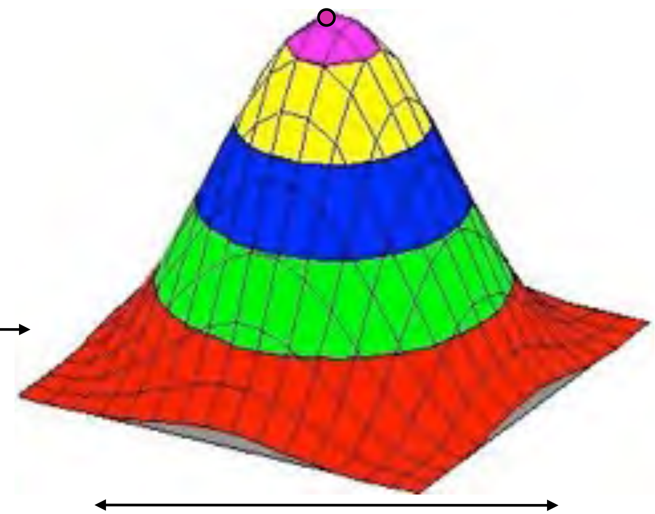


Density estimation

Density estimation creates a field from discrete point objects: the field's value at any point is an estimate of the density of discrete objects at that point.

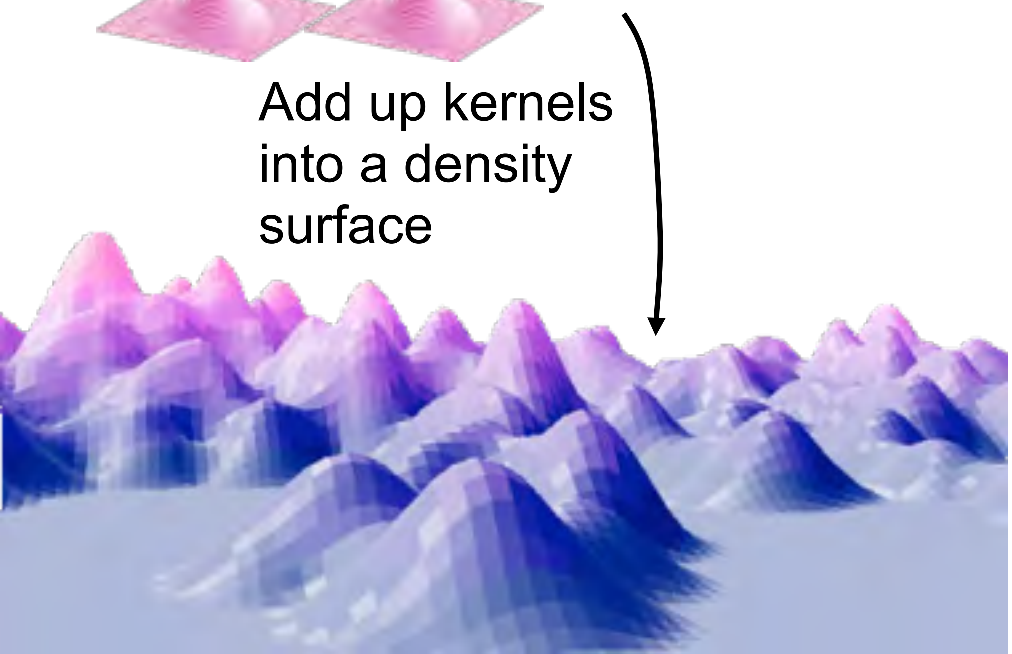


Substitute each known point with a kernel function



kernel width

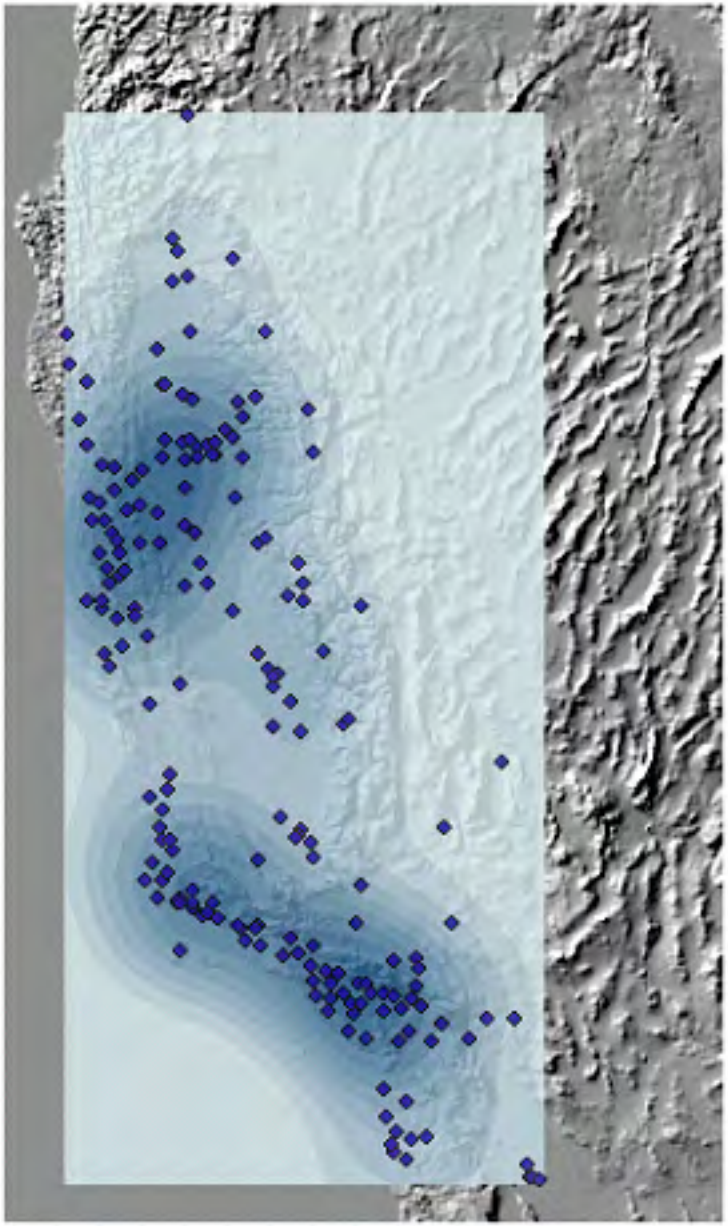
Add up kernels into a density surface



Why can this be used:

- produce a map of population density (a field) from a map of cities with a given population number (discrete objects)

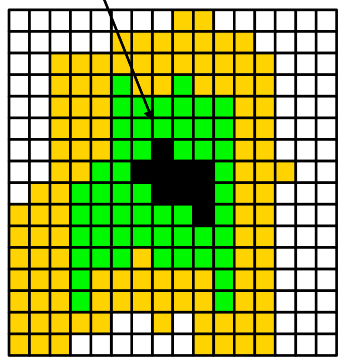
The result of applying a 150km-wide kernel to points distributed over California



When the kernel width is too small (in this case 16km, using only the S California part of the database) the surface is too rugged, and each point generates its own peak.

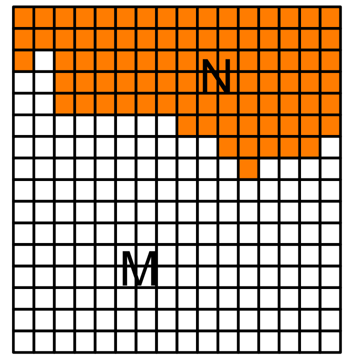
Travel time from the urban area

Cell value:
a_{ij} = travel time to the city



A

A raster map of two municipalities, M and N

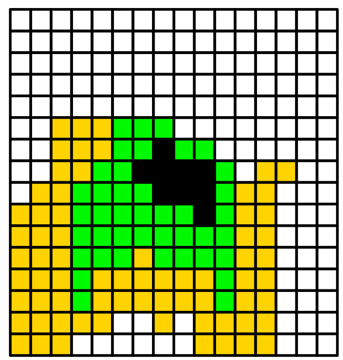


B

Cell value:
b_{ij} = 1 if the cell belongs to municipality M and 0 if it belongs to Municipality N

Cellwise-multiplication

$$C = A*B$$



C

Cell values in the resulting raster:
$c_{ij} = a_{ij} * b_{ij}$

Result: a map showing travel time to the city inside municipality M

Raster Analysis

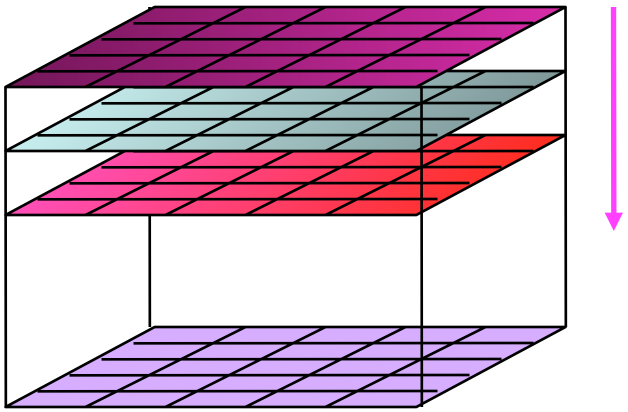
Map algebra

Map algebra involves combining raster maps cell by cell using:

- **boolean operators**

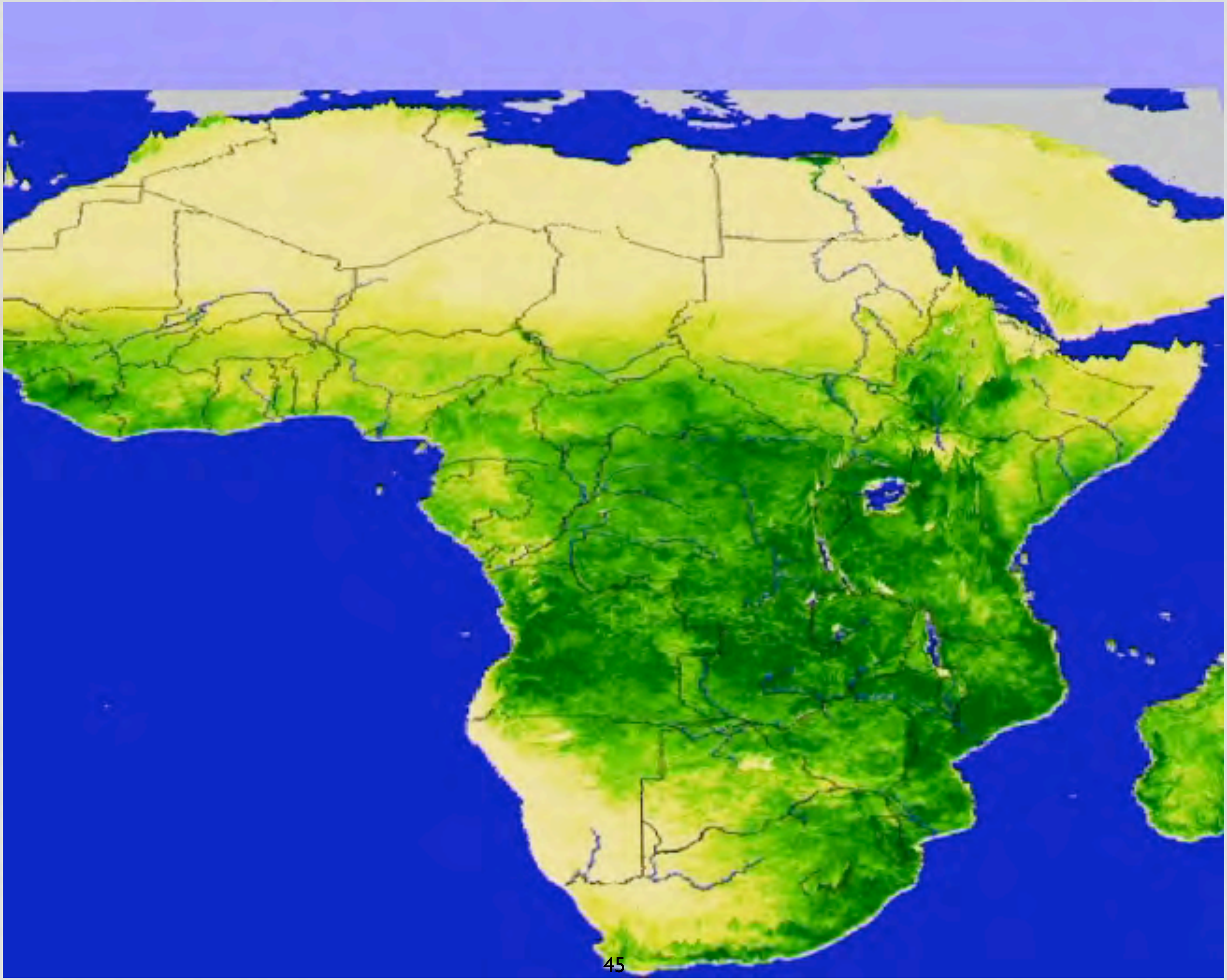
  Where is both A and B
  Where is A or B
  Where is B but not A
  Where is neither A nor B

- **algebraic operations** (+,-, *, /, log, etc)

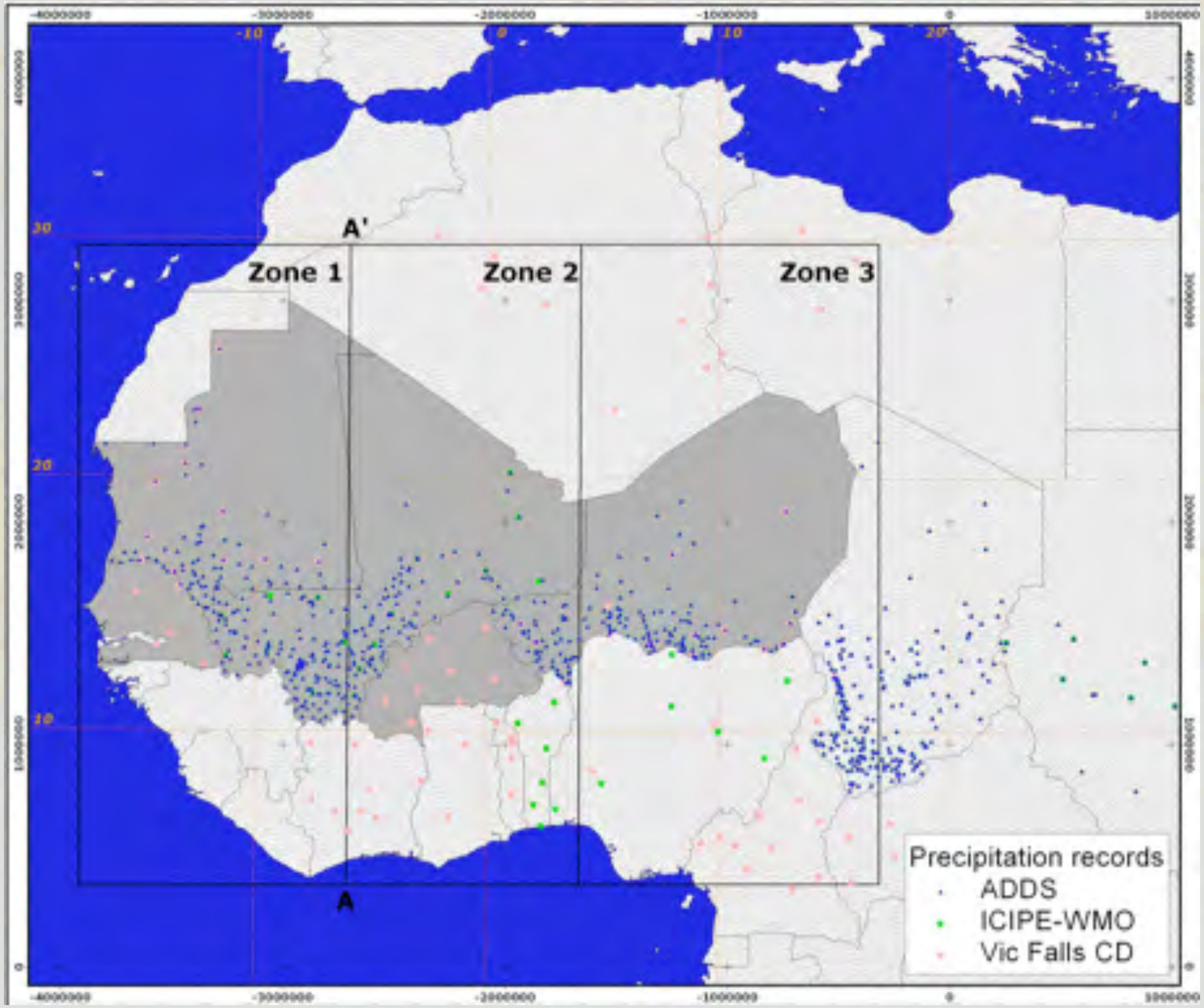


To create the desired map you may need to combine a large number of maps together. Some examples:

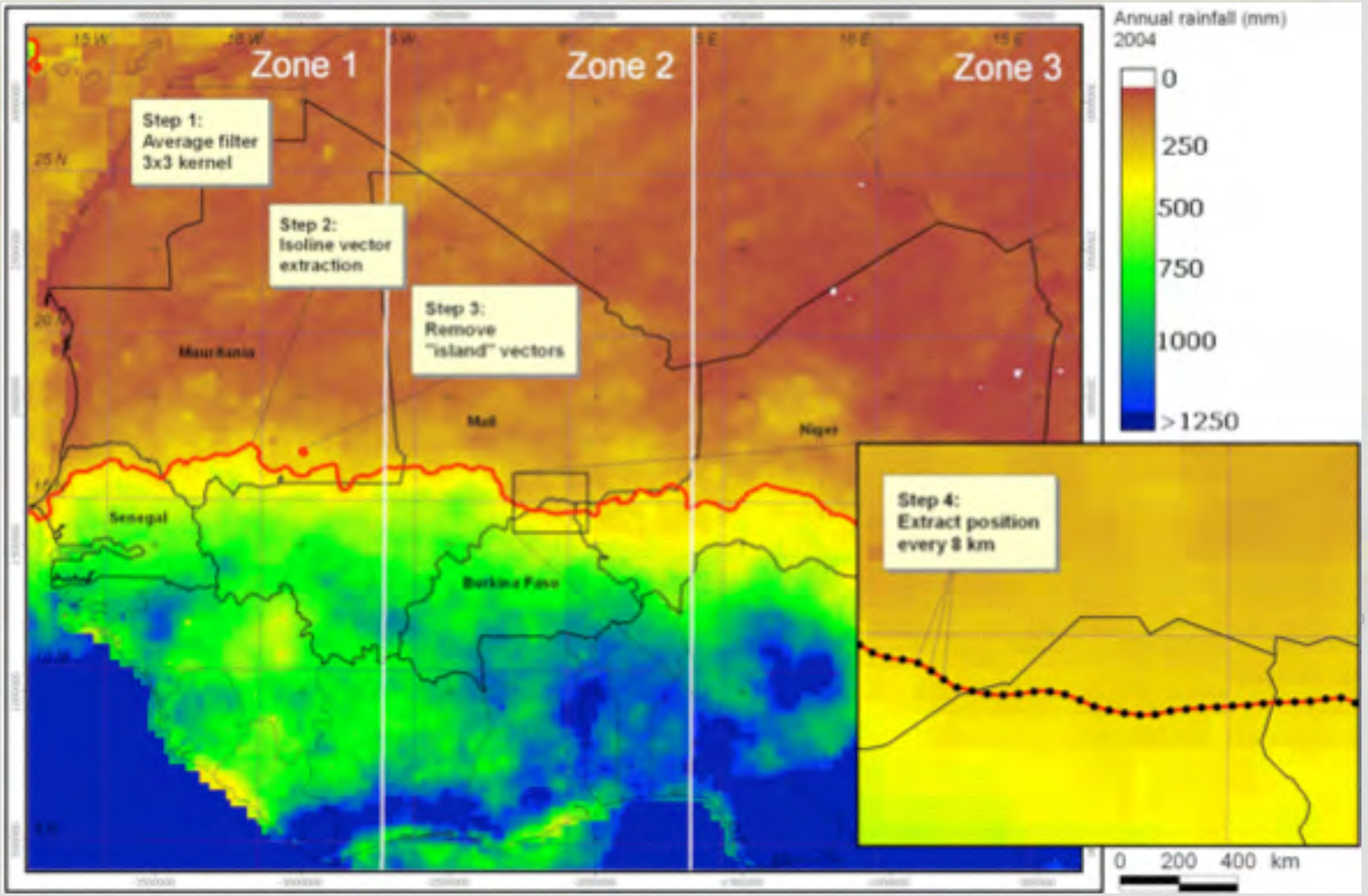
- elevation, slope and aspect
- areas to be excluded or included
- shortest distance to a certain location
- cost of travelling to a certain location
- qualitative data, such as soil, landuse, bedrock
- etc.

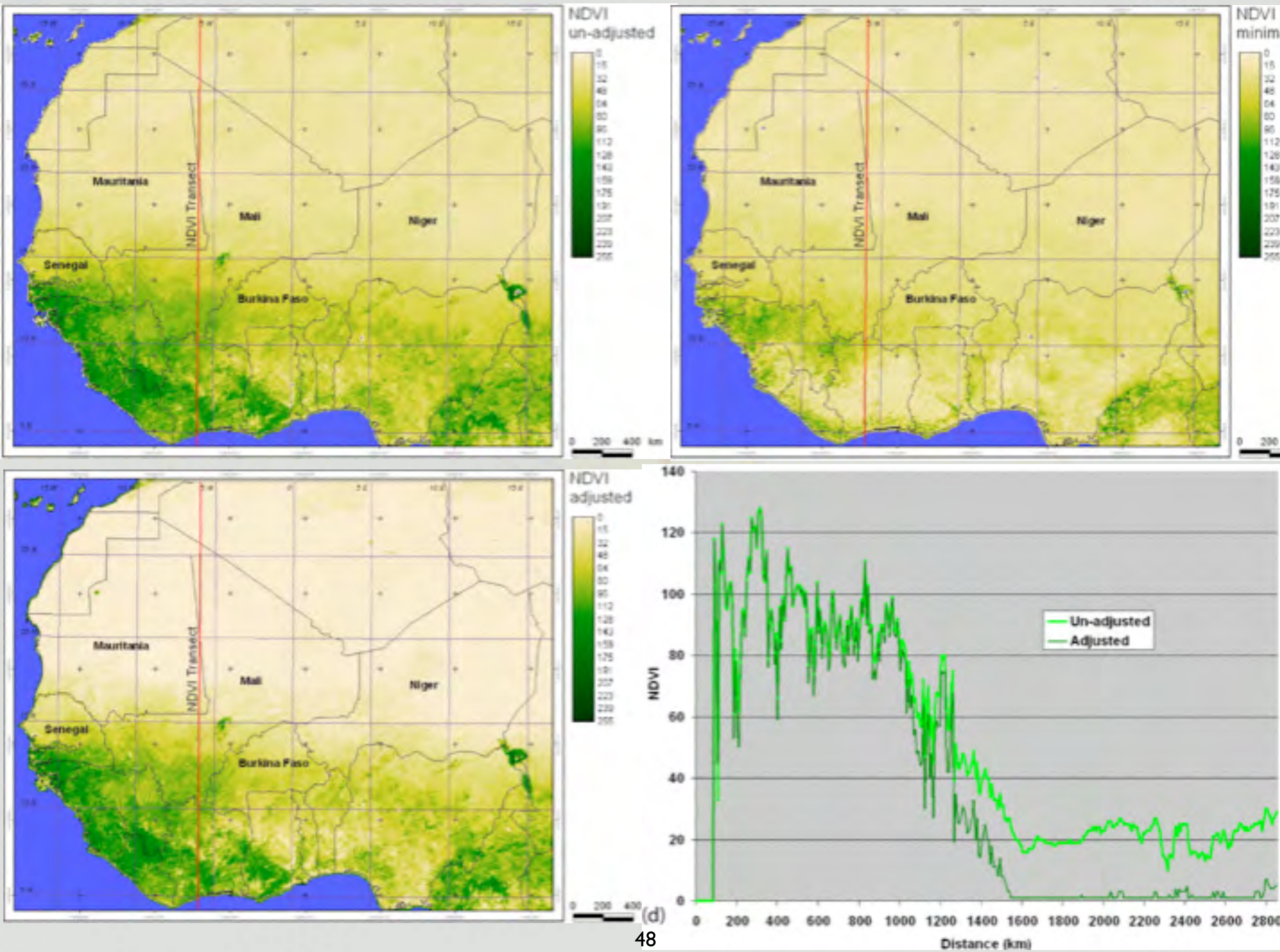


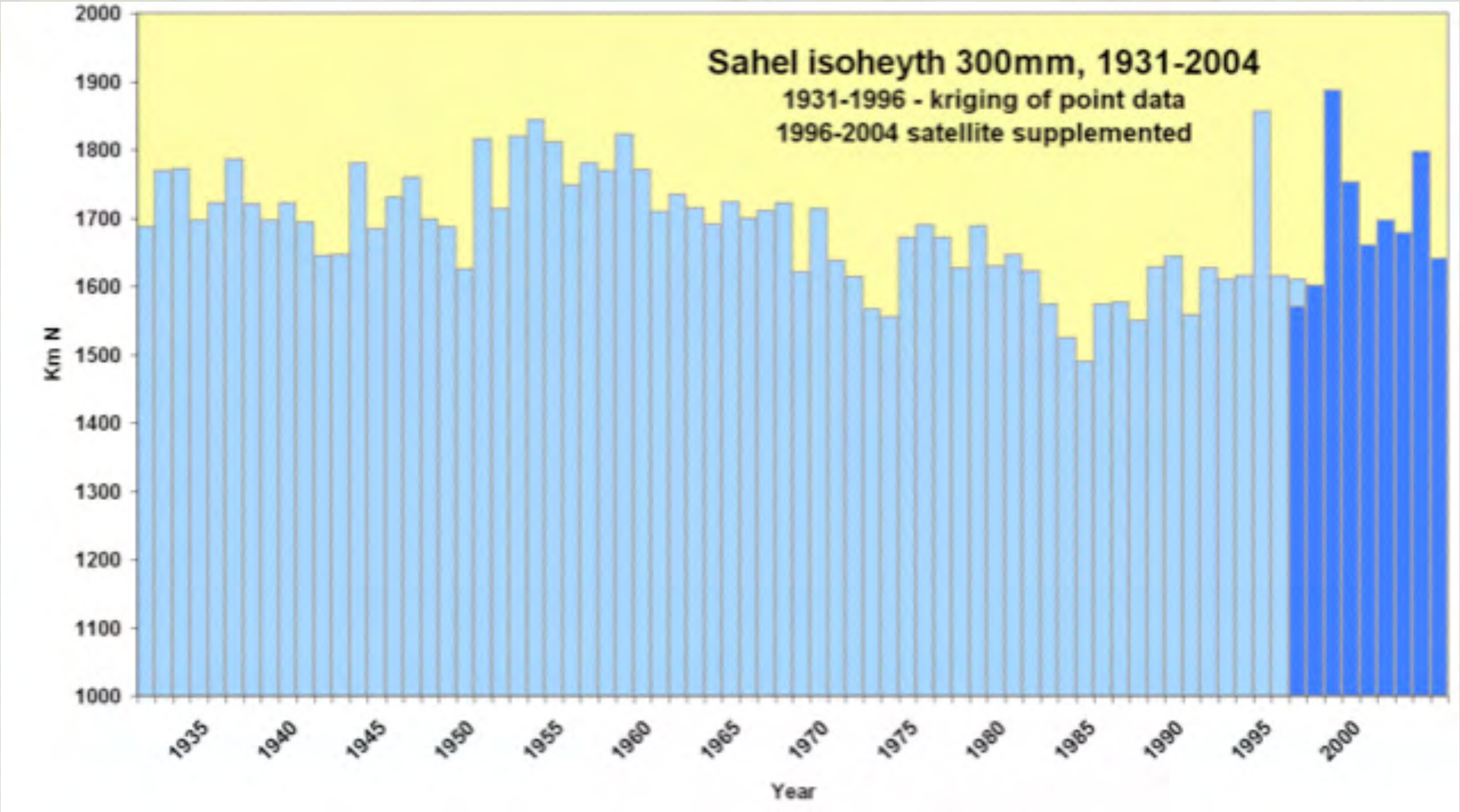
Interpolering av nederbördsdata

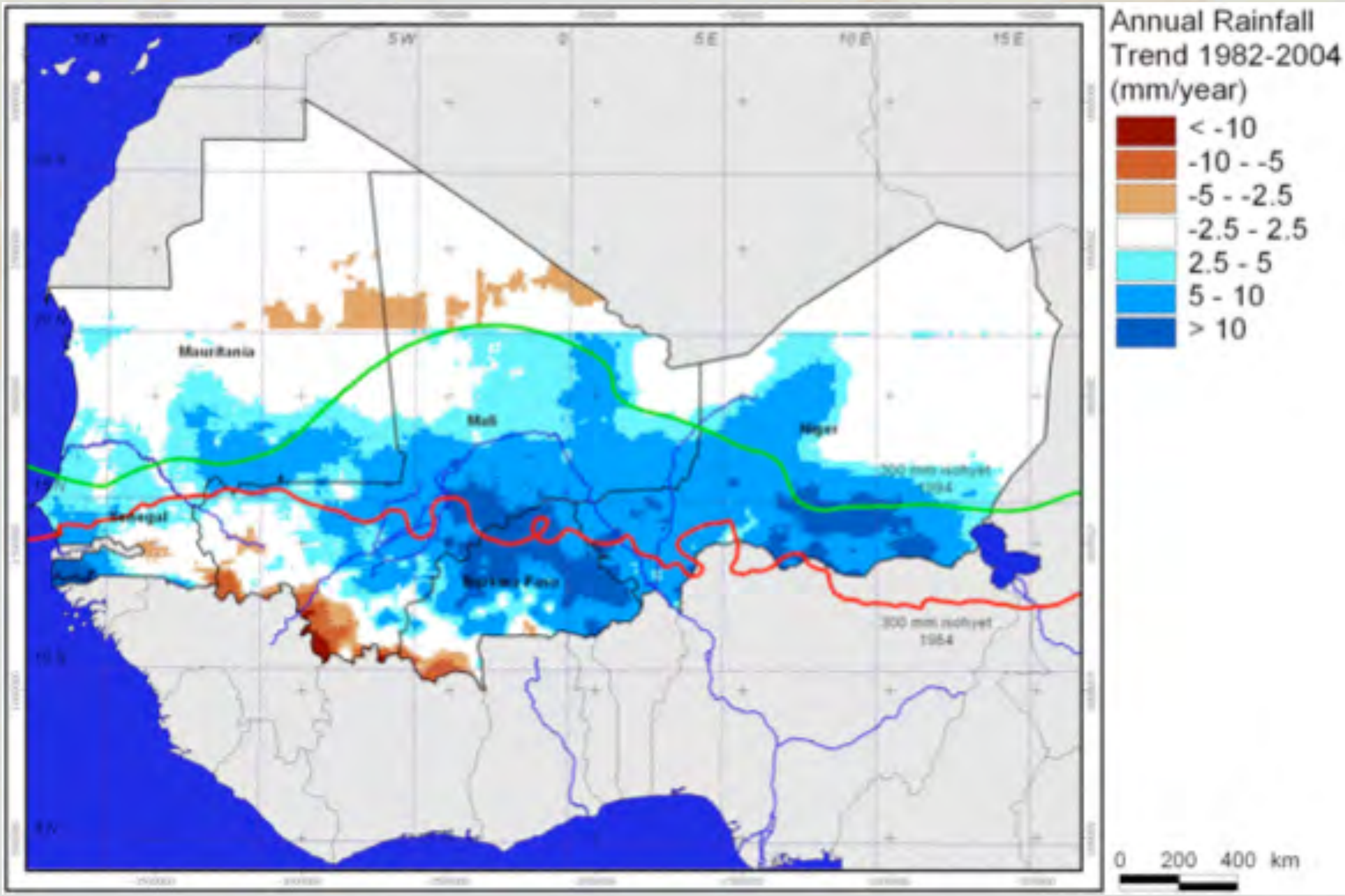


Extrahering av isohyet (isolinje för nederbörd)




NDVI
un-adjusted
Mauritania
Mali
Niger
Senegal
Burkina Faso
NDVI Transect
NDVI
minimu
NDVI
adjusted
0 200 400 km
(d)
Un-adjusted
Adjusted
NDVI
Distance (km)

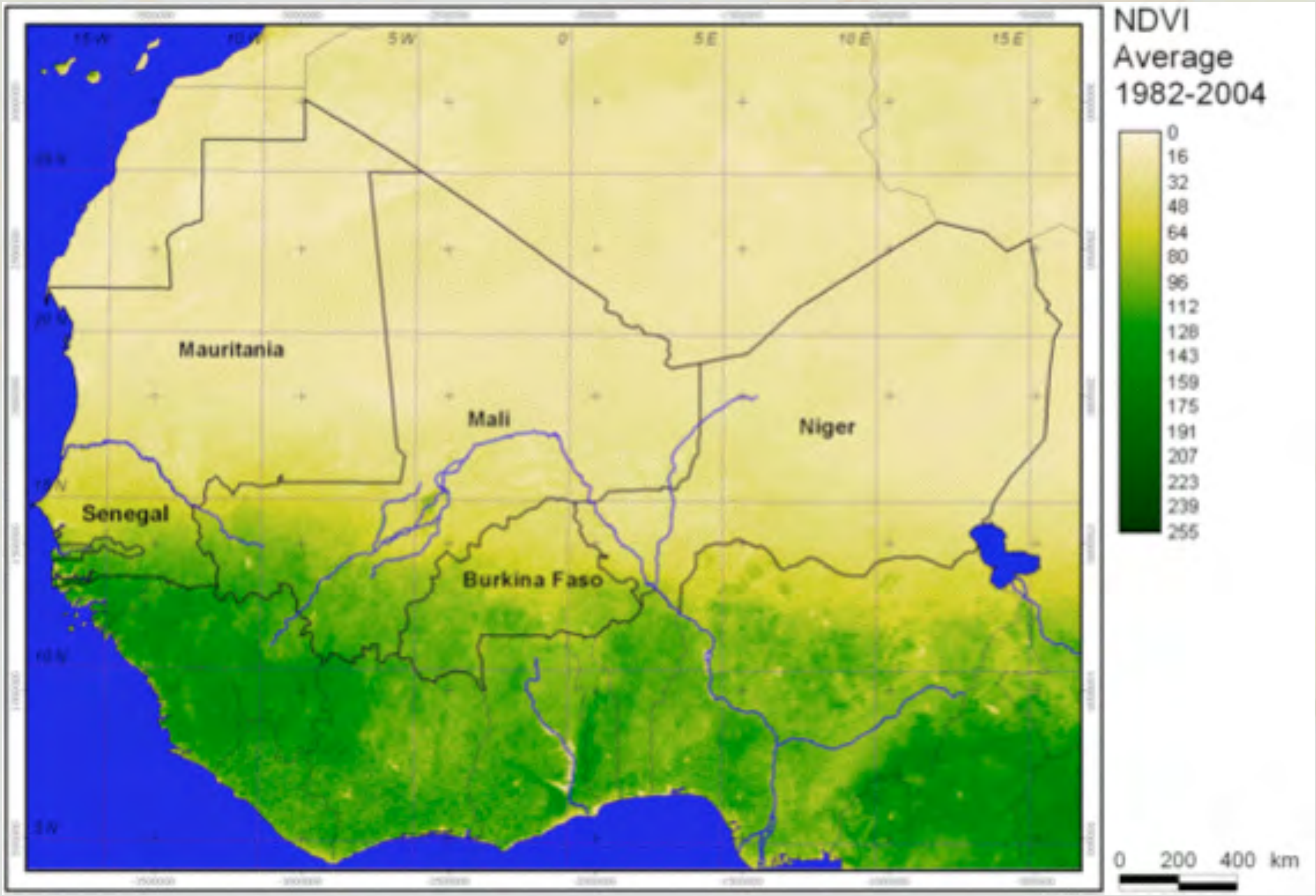

Sahel isoheyth 300mm, 1931-2004
1931-1996 - kriging of point data
1996-2004 satellite supplemented
Km N
2000
1900
1800
1700
1600
1500
1400
1300
1200
1100
1000
1935
1940
1945
1950
1955
1960
1965
1970
1975
1980
1985
1990
1995
2000
Year



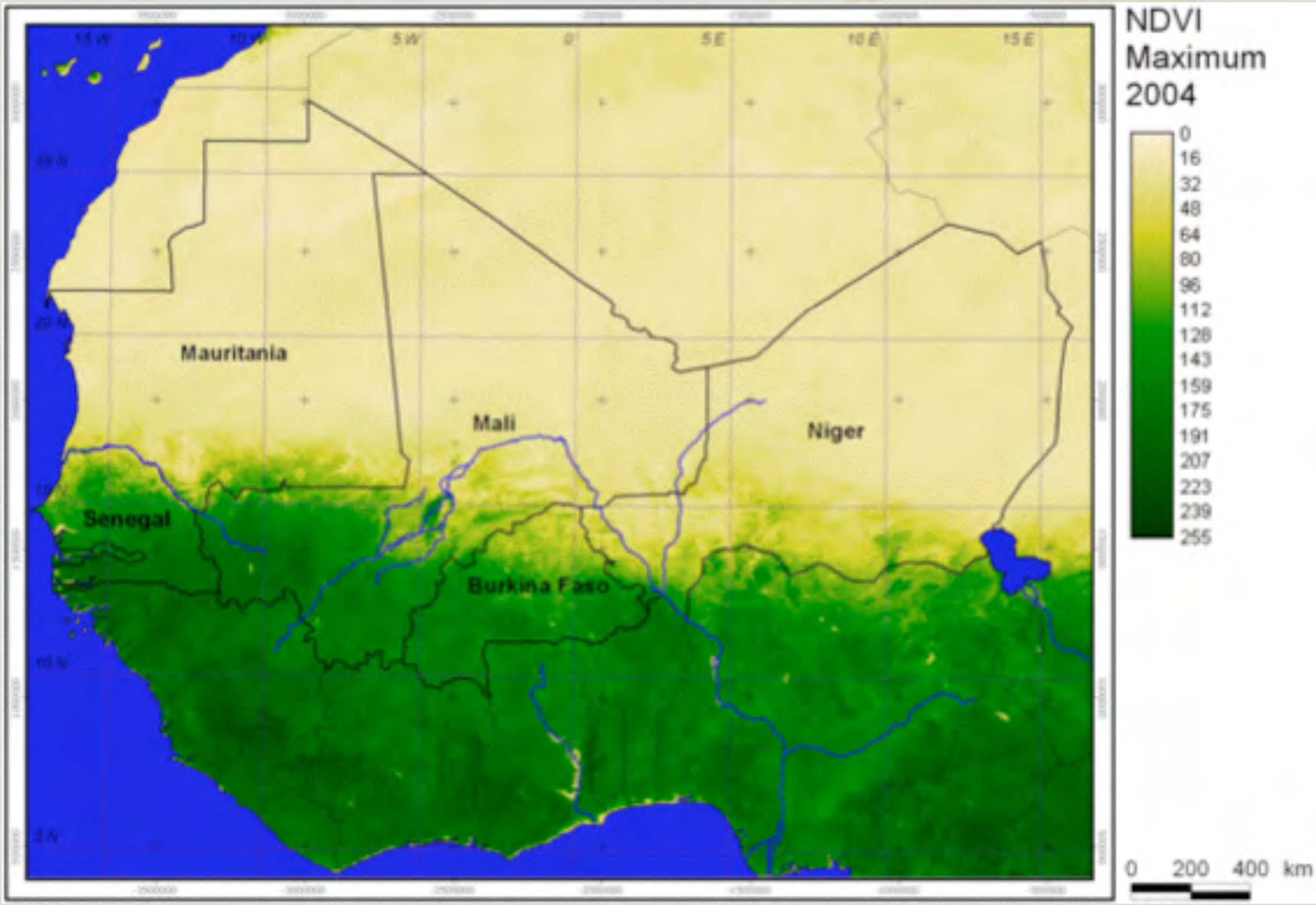
Annual Rainfall Trend 1982-2004 (mm/year)
< -10
-10 - -5
-5 - -2.5
-2.5 - 2.5
2.5 - 5
5 - 10
> 10
Mauritania
Mali
Niger
Senegal
0 200 400 km

$$NDVI = \frac{NIR - red}{NIR + red}$$

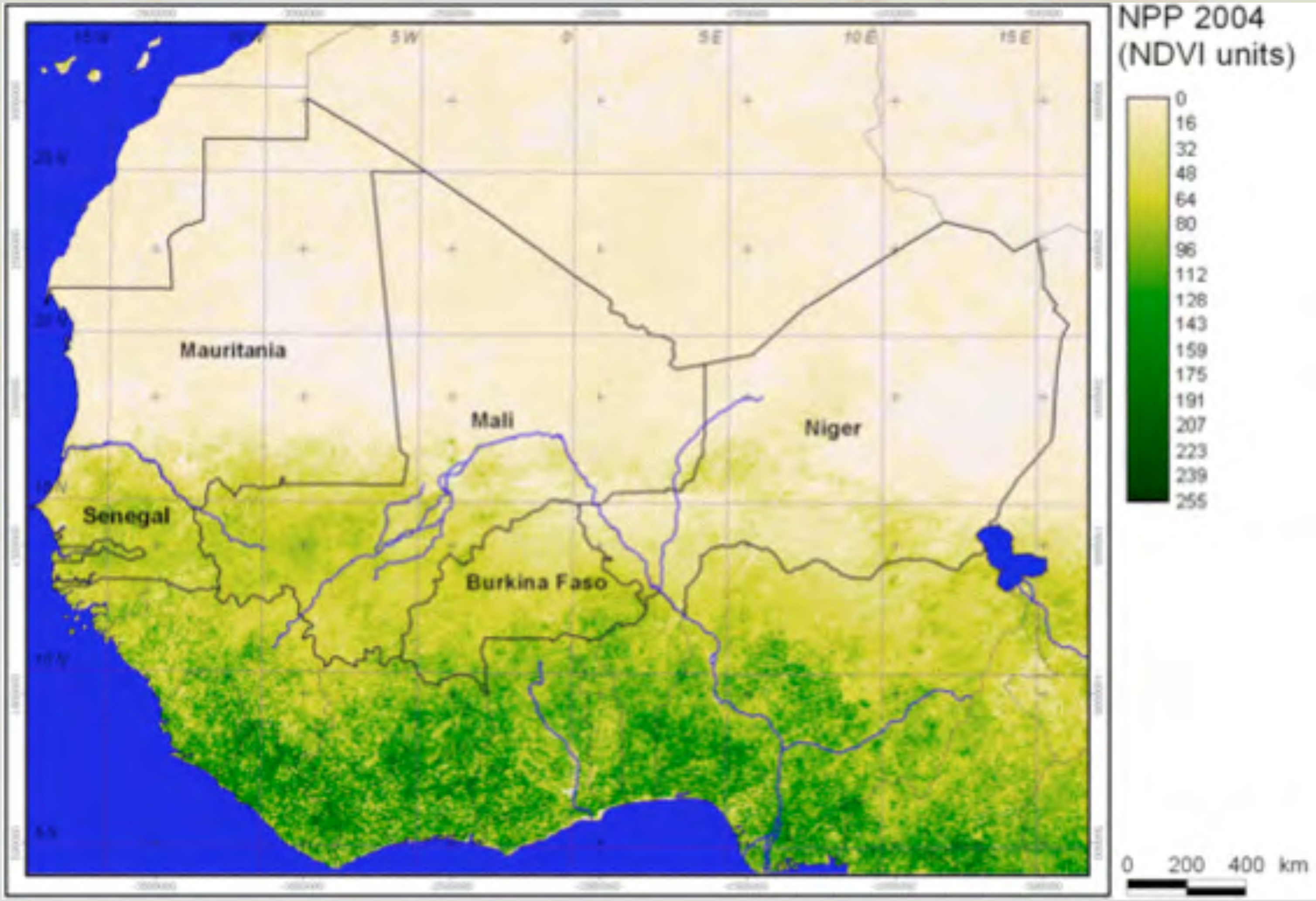
NDVI årligt medelvärde (= integral)



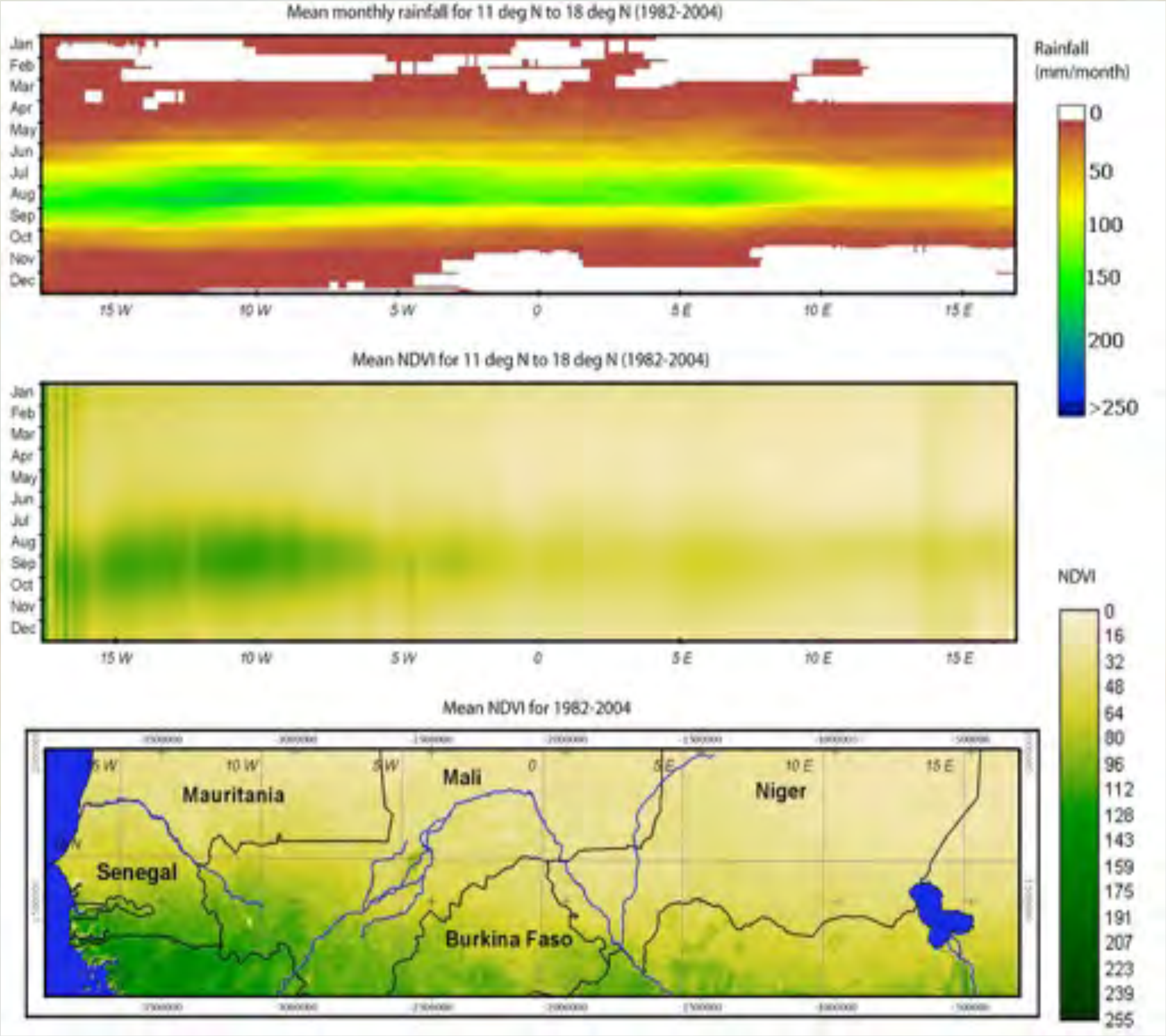
NDVI årligt maxvärde



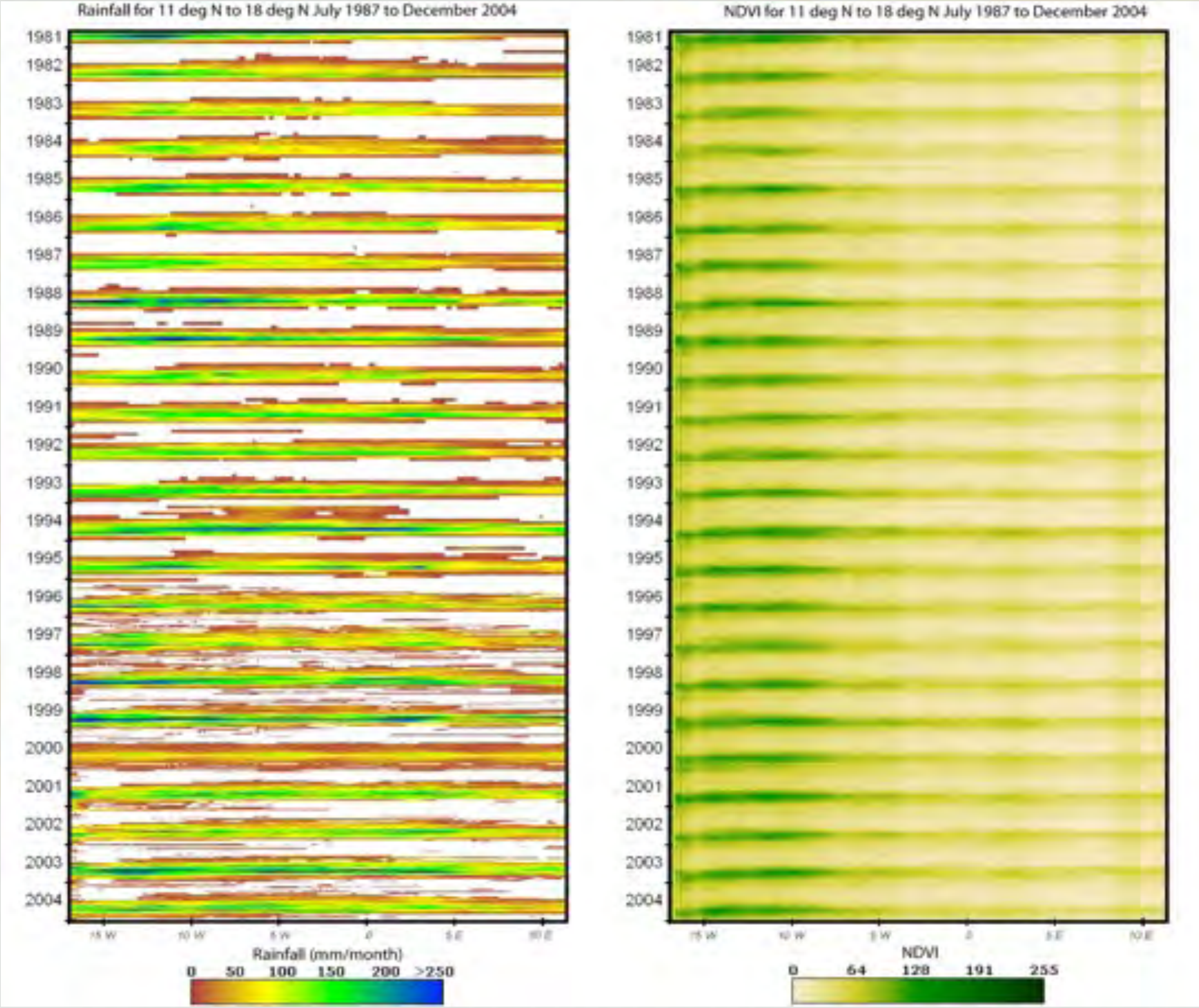
NDVI årlig ökning



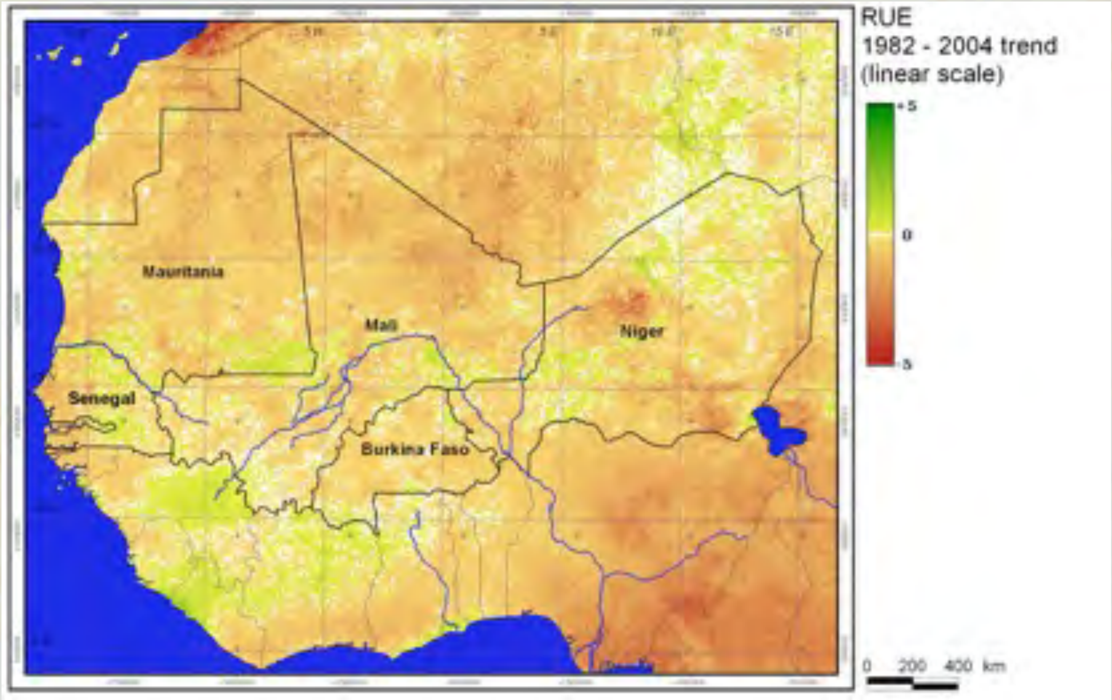
NDVI årlig cykel



NDVI utveckling 1981-2004

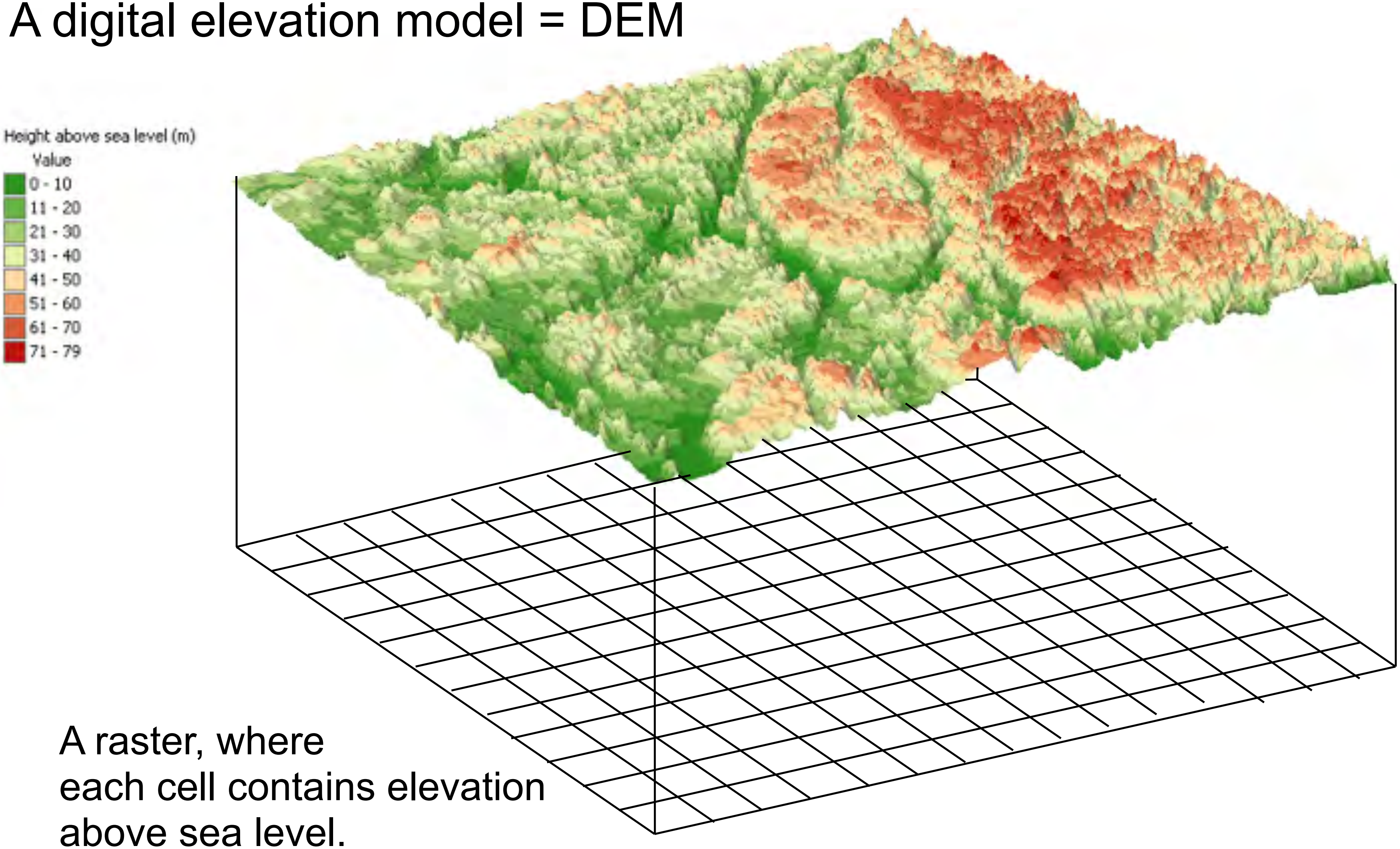


NDVI absoluta och regn-normaliserade trender (OLS)



DEM, slope and aspect

A digital elevation model = DEM



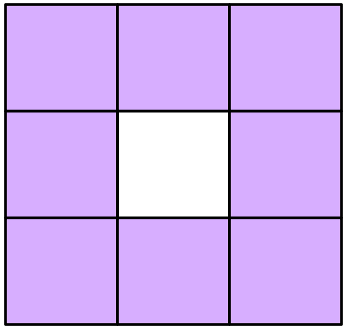
A raster, where
each cell contains elevation
above sea level.

Slope and aspect

Calculated from a digital elevation model

Slope and aspect are calculated at each cell in the grid, by comparing the cell's elevation to that of its neighbors:

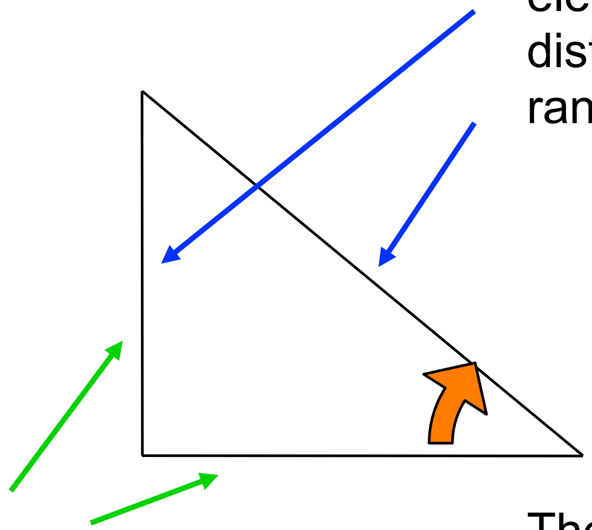
- usually its eight neighbors
- but the exact method varies



In a scientific study, it is important to know exactly what method is used when calculating slope, and exactly how slope is defined.

Alternative **definitions of slope**

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

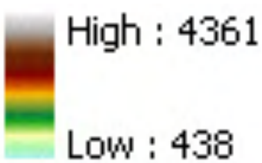


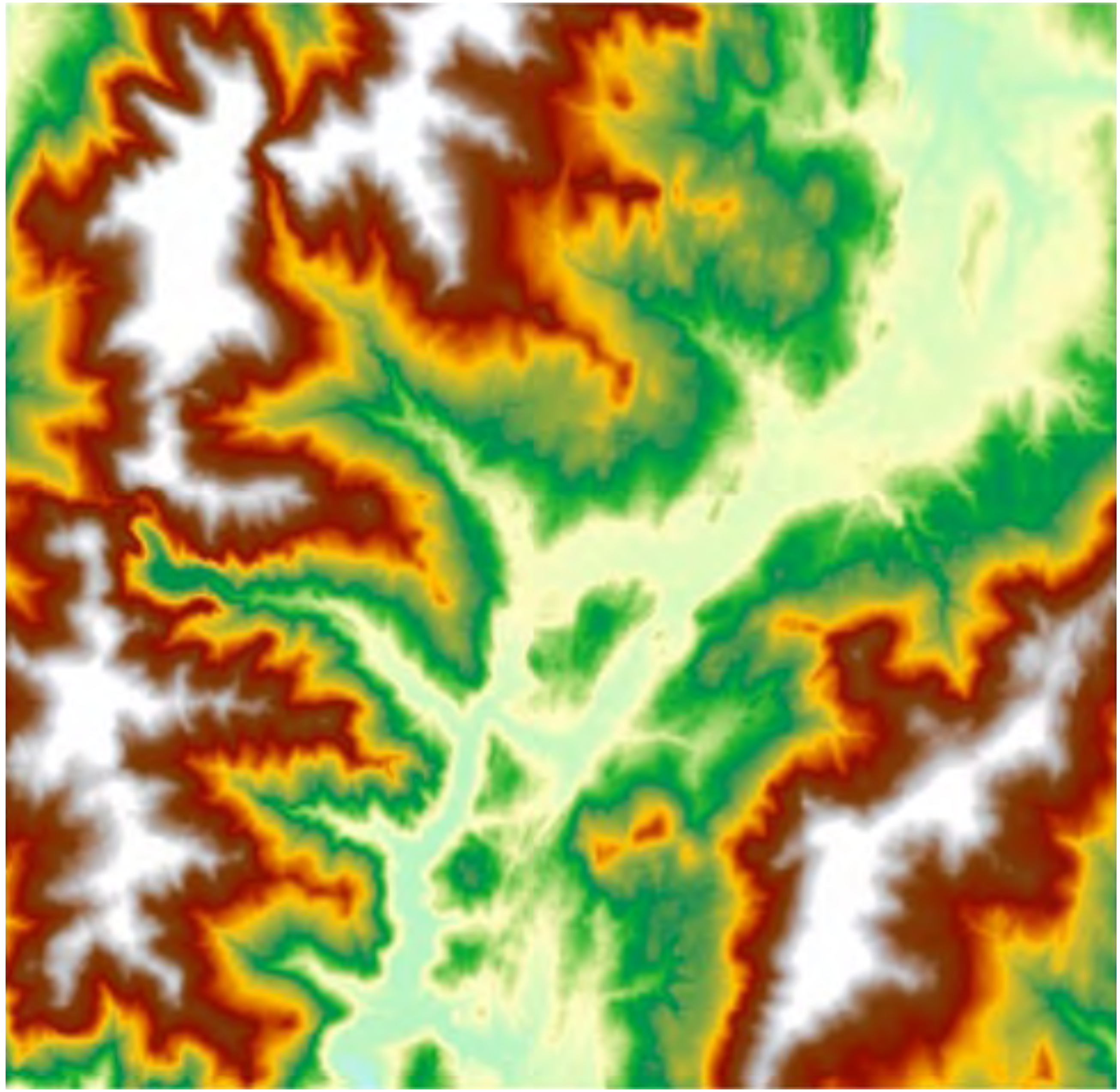
The *ratio* of the change in elevation to the horizontal distance traveled, range 0 to infinity

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

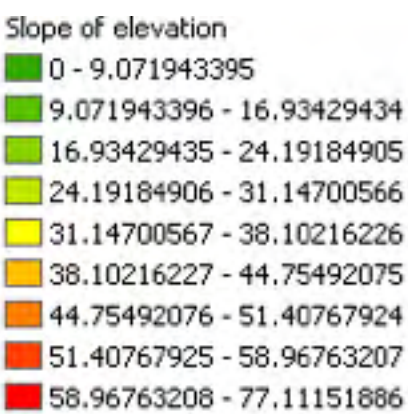
Aspect = direction of slope

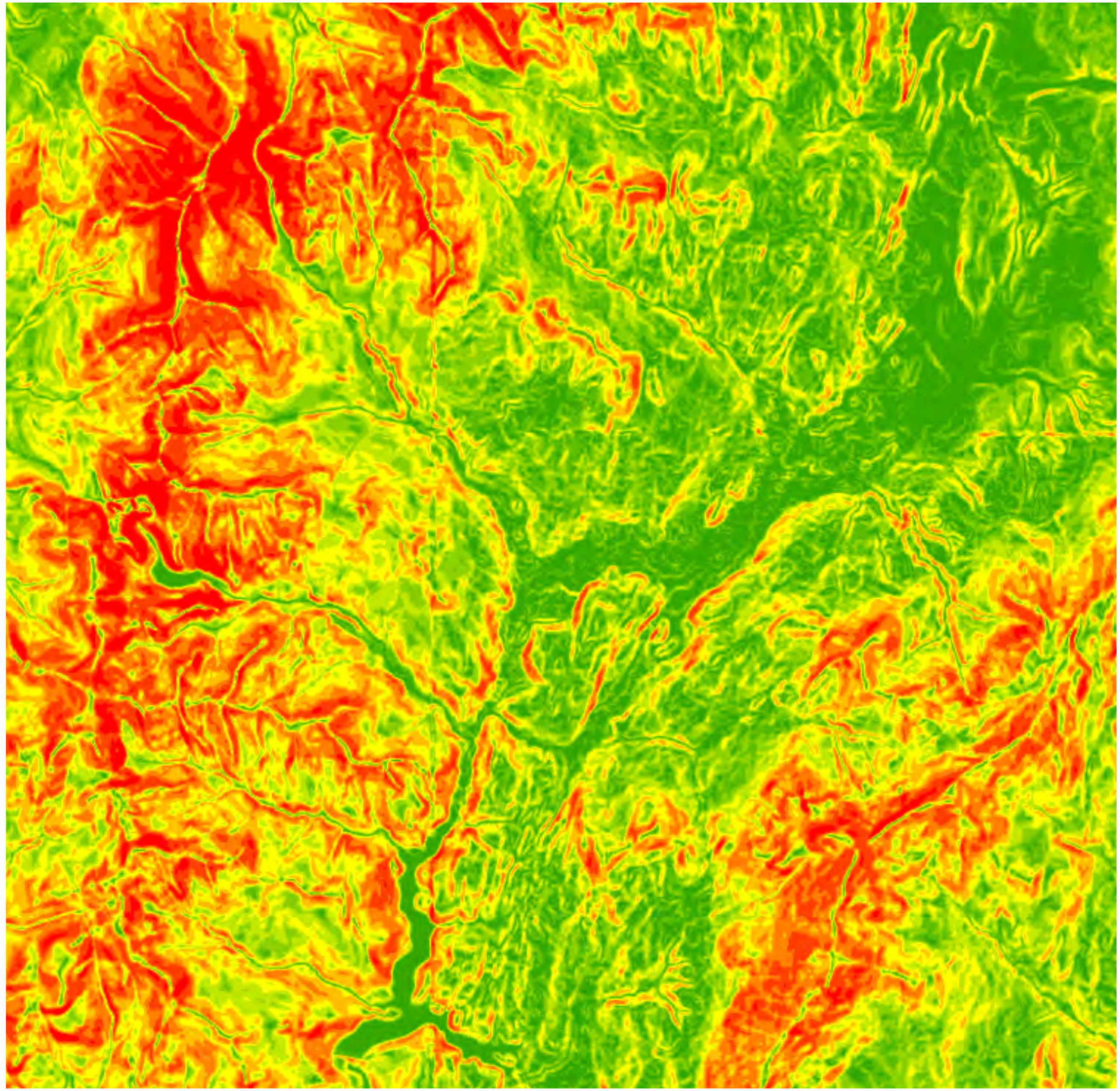
Elevation
(m)





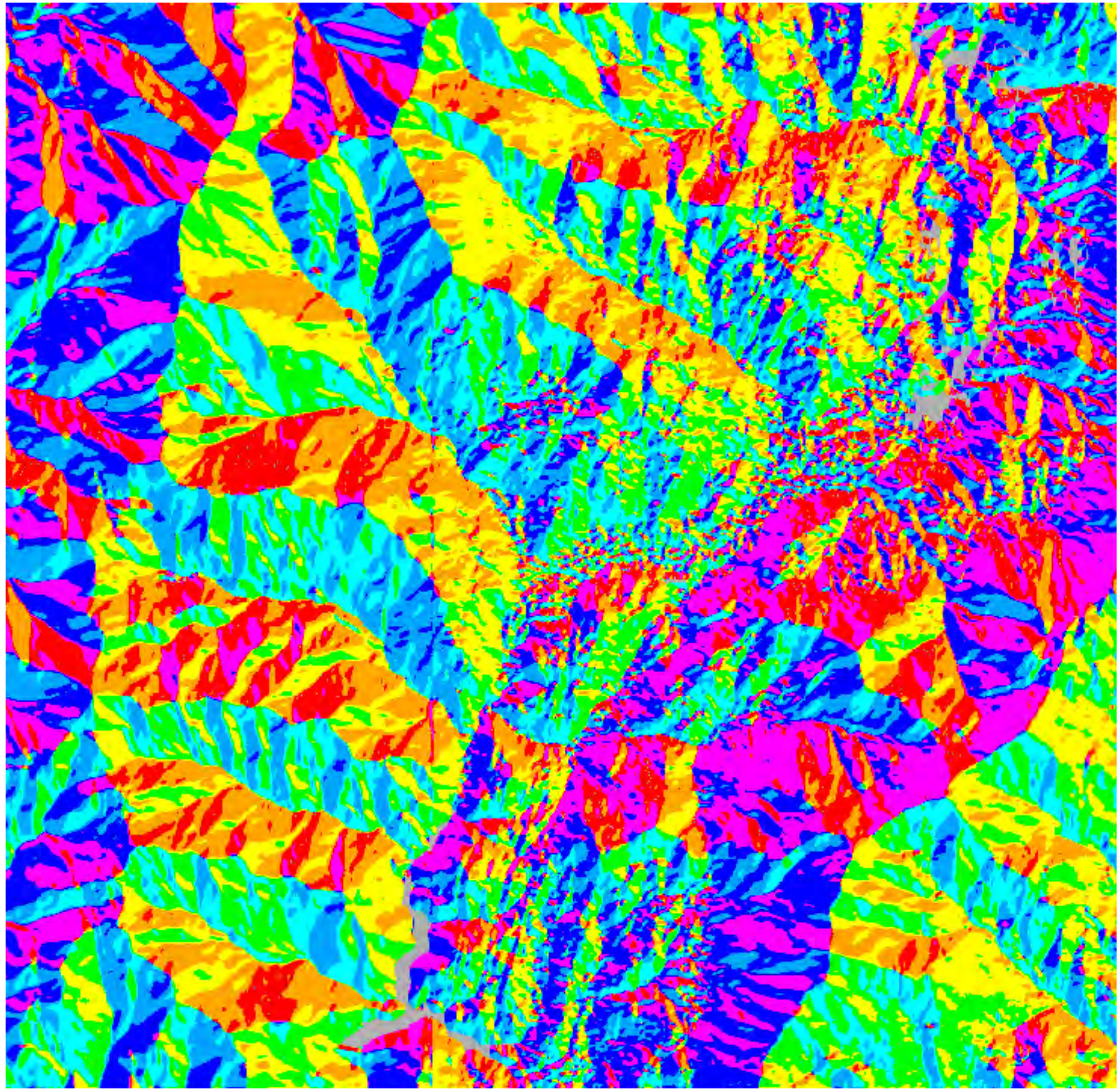
Slope
(degrees)





Aspect
(direction of slope)

Aspect of elevation
- Flat (-1)
- North (0-22.5)
- Northeast (22.5-67.5)
- East (67.5-112.5)
- Southeast (112.5-157.5)
- South (157.5-202.5)
- Southwest (202.5-247.5)
- West(247.5-292.5)
- Northwest (292.5-337.5)
- North (337.5-360)



An example of cartographic modelling with map algebra

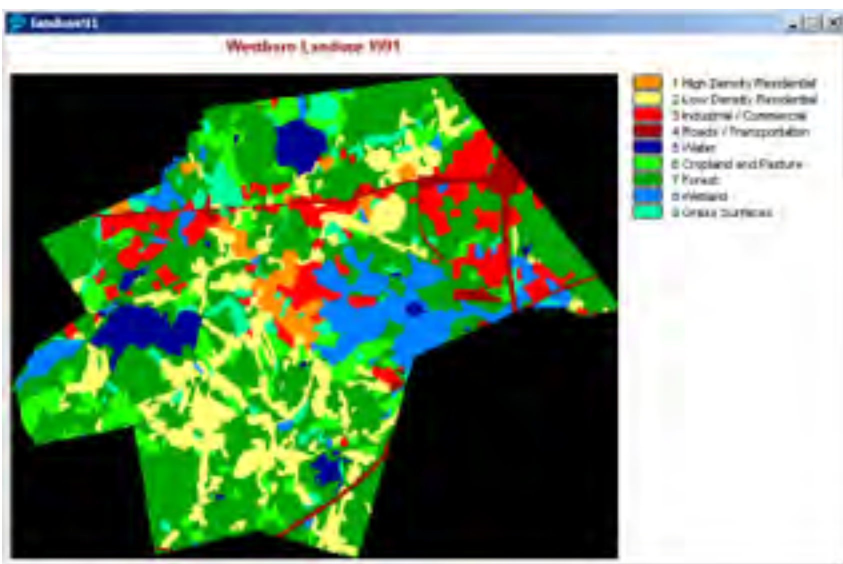
Task:

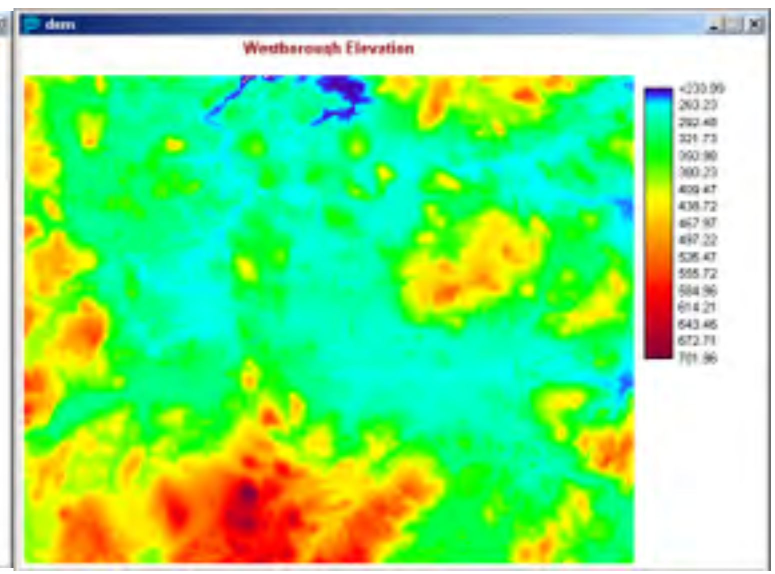
Find areas suitable for a new residential area considering the following aspects:

- the area should be located more than 200m from the water’s edge
- the slope should be less than 5 %
- landuse type should be forest

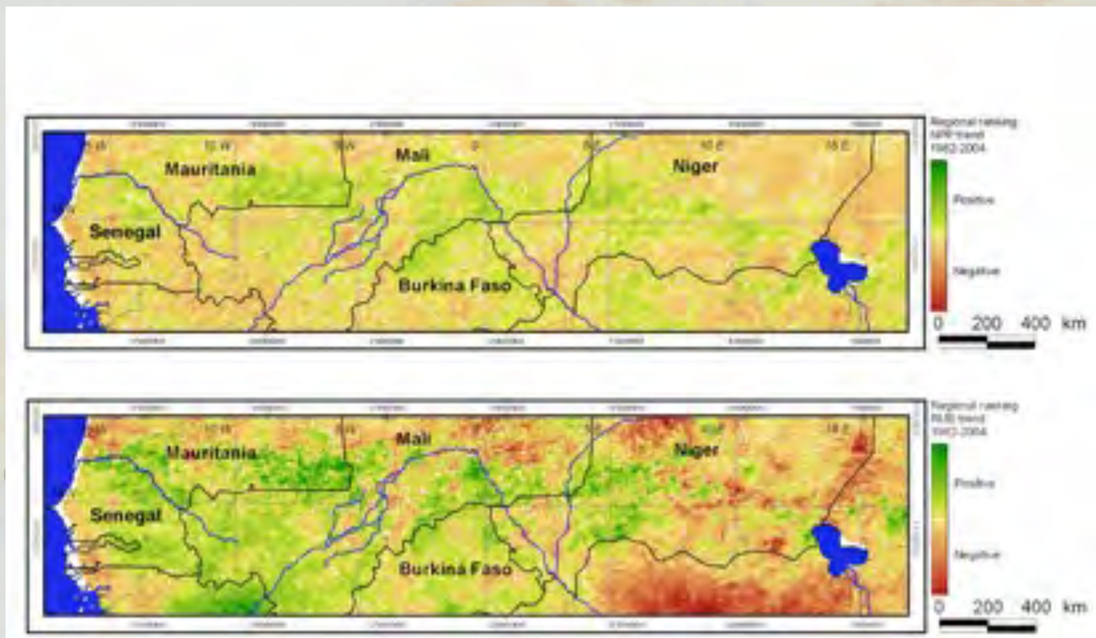
Data:

- landuse map
- Digital Elevation Model, DEM





Data drivna analyser och måldrivna analyser



Ex. Finna de bästa och sämsta platserna = måldrivet