PLSCS/TRES 6200
Spatial Modelling and Analysis

Concepts of Space

D G Rossiter

Cornell University
Section of Soil & Crop Sciences

November 27, 2021

Topic: Types of spaces

- A **vector space** in mathematics is any set of variables that form **metric** axes with a known orientation to a **reference system**.
  - This allows us to compute a **distance** and **direction** between points in that space.
  - If these variables represent **attributes**, we have a **feature** space.
  - If they represent geographic **coördinates**, we have a **metric geographic** space.

- A **non-metric** mathematical space can represent **topology**
  - If these are **geospatial** relations, we have a **topological geographic** space.

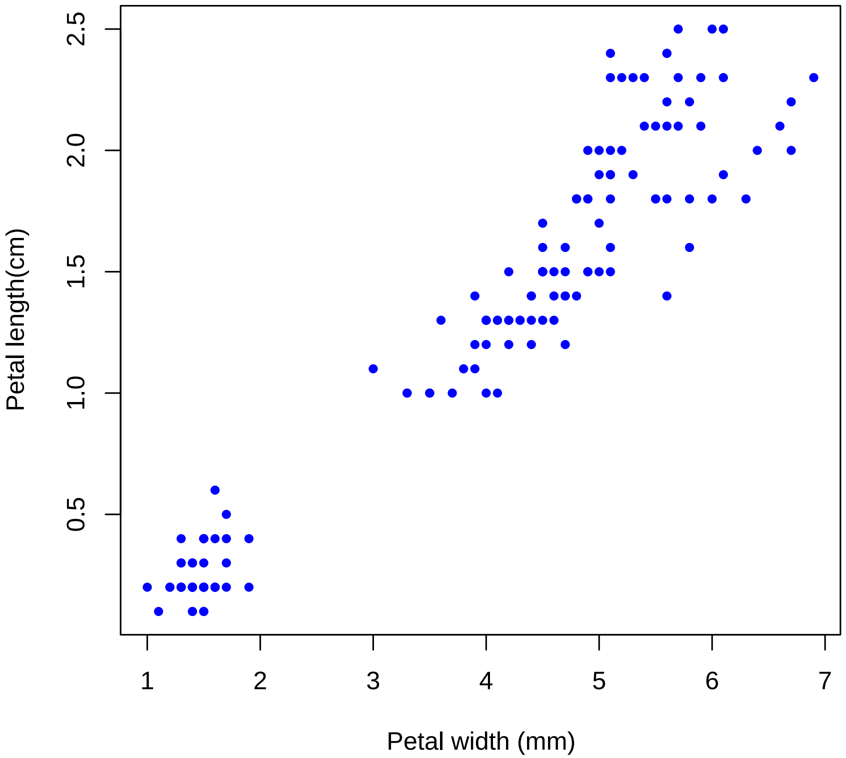
Feature space

This is the vector space formed by any set of variables:

- **Axes** represent each variable;
- **Coördinates** are values of variables, possibly transformed or combined;
- Classified variables should be **ordinal** with some **distance measure**, but sometimes **nominal** variables are considered to be in feature space.
- The observations are related in this ‘space’, e.g. the “distance” between them can be calculated.
- This is the basis of **multivariate analysis**.

Feature space is sometimes referred to as **attribute space**.

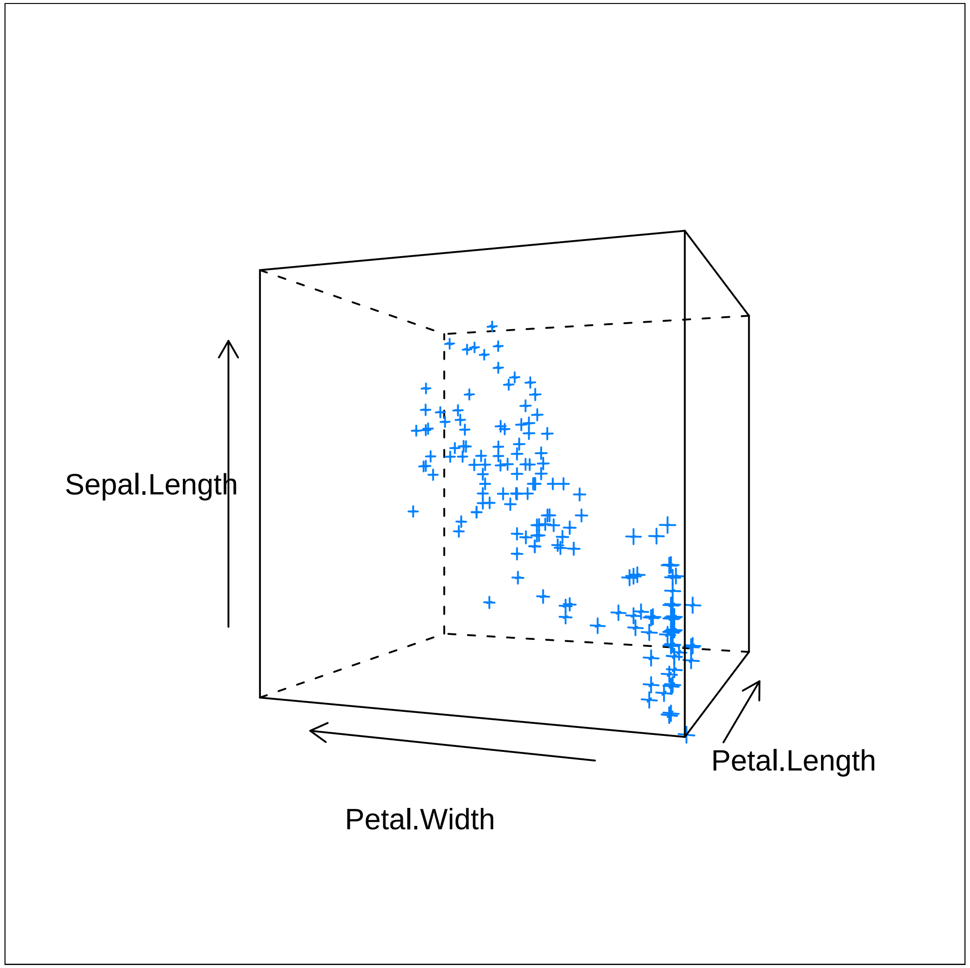
Scatterplot of a 2D feature space



This is a **visualisation** of a 2D feature space using a **scatterplot**; two attributes of individual iris flowers as coördinates.

Source: Anderson, Edgar (1935), *The irises of the Gaspé Peninsula*, Bulletin of the American. Iris Society, 59, 2-5.

Scatterplot of a 3D feature space



Anderson *Iris* data, three attributes as coördinates.

Geographic space

- “Geo” + “graphy” = “Earth” + “mapping”
- Related somehow to the Earth’s surface
- **metric** vs. **topological**

Metric geographic space

- a mathematical space where the axes are **map coördinates** that relate points to some reference location on or in the Earth (or another physical body)

- These coördinates are often in some **geographic coördinate system** that was designed to give each location on (part of) the Earth a unique identification (see lecture on Coördinate Reference Systems)

- However, a **local coördinate system** can be used, as long as there is a clear relation between locations and coördinates;
  - ungeoreferenced aerial or satellite imagery
  - photograph of microscope slide (not the Earth's surface but has metric geometry)

- Key point: can compute **distances**, **angles** of separation, and **areas**.

Metric space

- Coördinates represent "true" distances along their axes

- Axes are 1D **lines** or **arcs** with a **unit of measure** (e.g. metres, kilometres, degrees . . .)

- **One-dimensional**: coördinates are on a line with respect to some origin (0): $(x_1) = x$

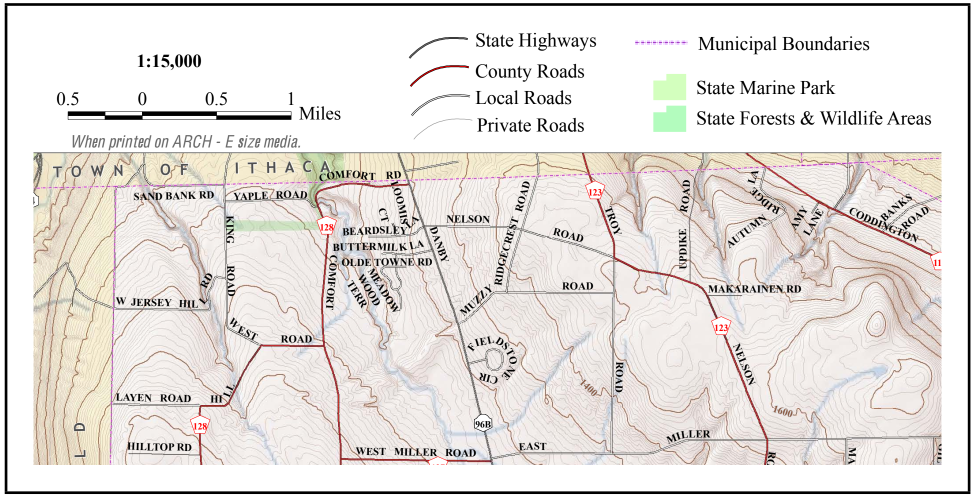
- **Two-dimensional**: coördinates are on a grid with respect to some origin $(0, 0)$: $(x_1, x_2) = (x, y) = (E, N)$
  - **Latitude-longitude** (sometimes called "geographic") coördinates do not have equal distances in the two dimensions; they should be transformed to metric (grid) coördinates for geo-statistical analysis.

- **Three-dimensional**: coördinates are grid and elevation (or depth! a negative elevation) from a reference elevation: $(x_1, x_2, x_3) = (x, y, z) = (E, N, H)$

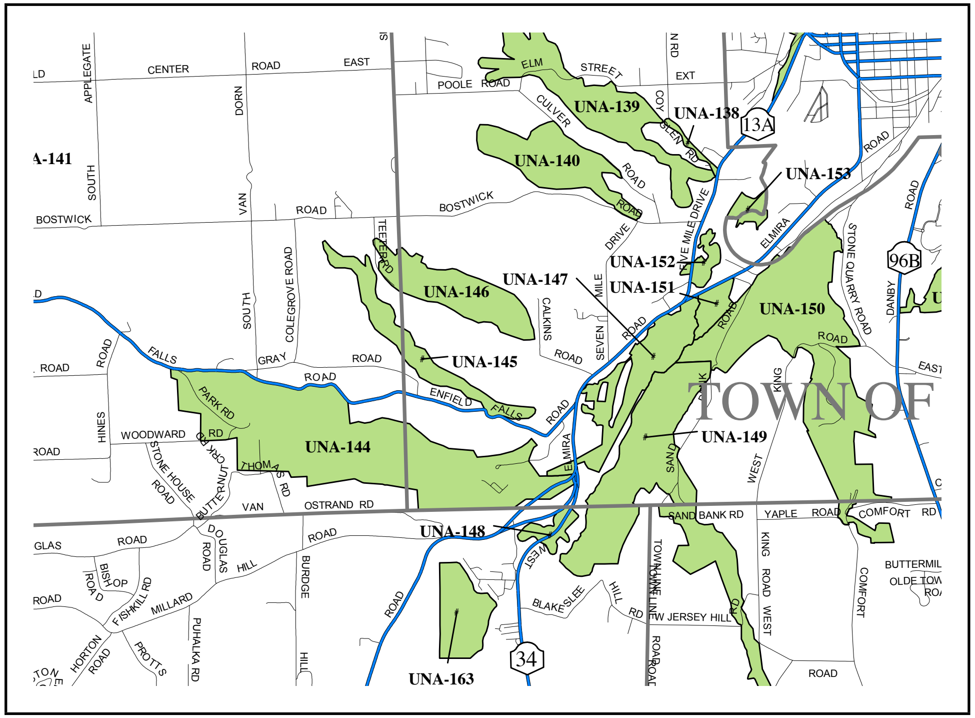
Maps of metric space

- Shows features as **abstract** objects (points, lines, polygons, grids);
- The features (attributes) are **labelled**; these are collected in a **legend**;
  - **continuous** vs. **classified** attributes/features
- A map shows both **metric geographic** and **feature** (attribute) spaces;
- One map can show many **attributes**
  - Special case: contour maps “2.5D”, show the third geographic dimension (height, depth) on a 2D display

Some maps in metric space



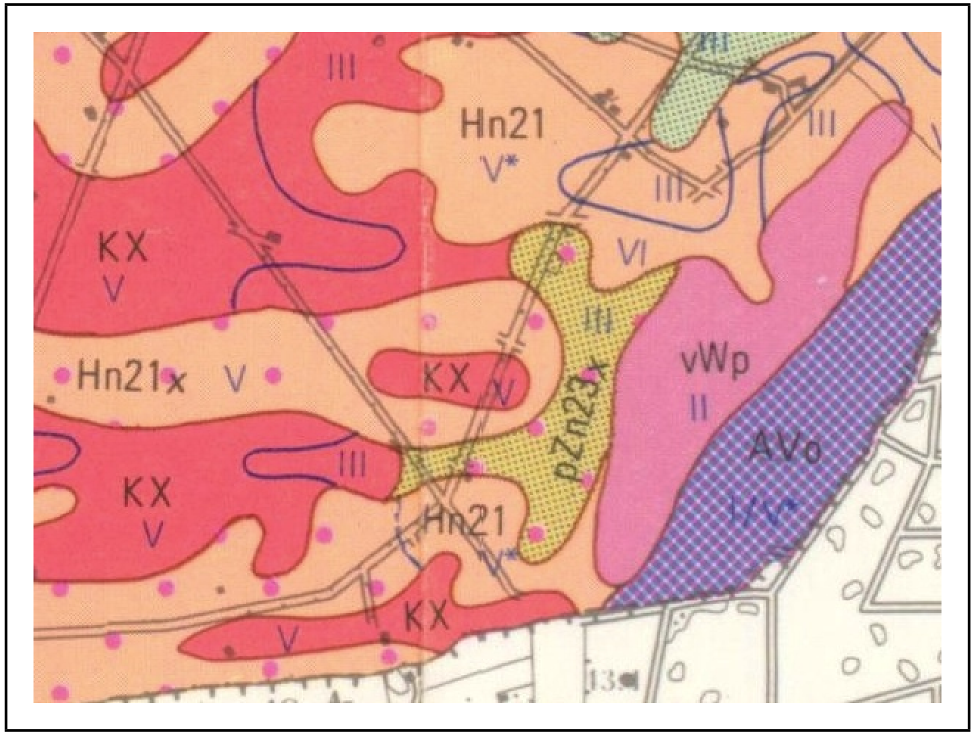
contours; themes: roads (lines); parks & administrative divisions (polygons)



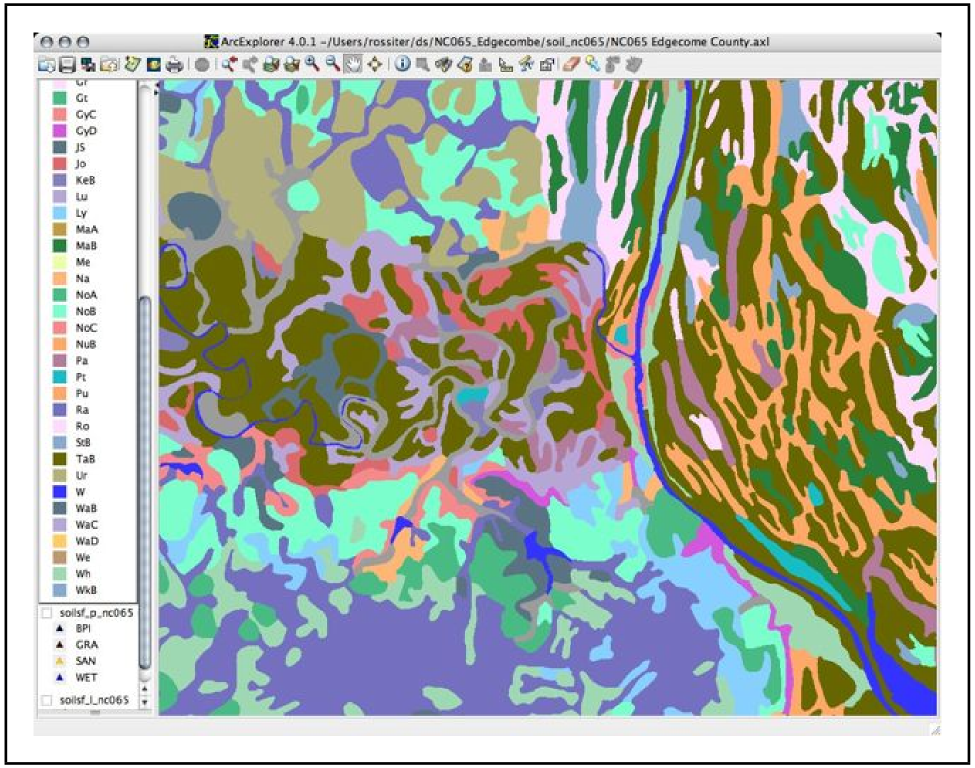
labelled polygons and lines

These are plotted with a metric Coördinate Reference System (CRS), not shown here.

Soil class maps

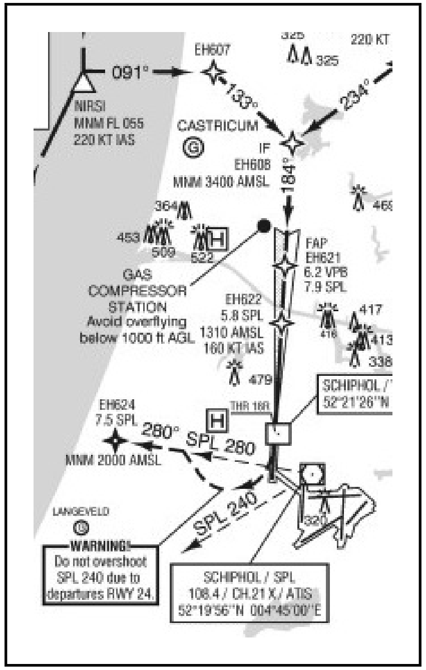


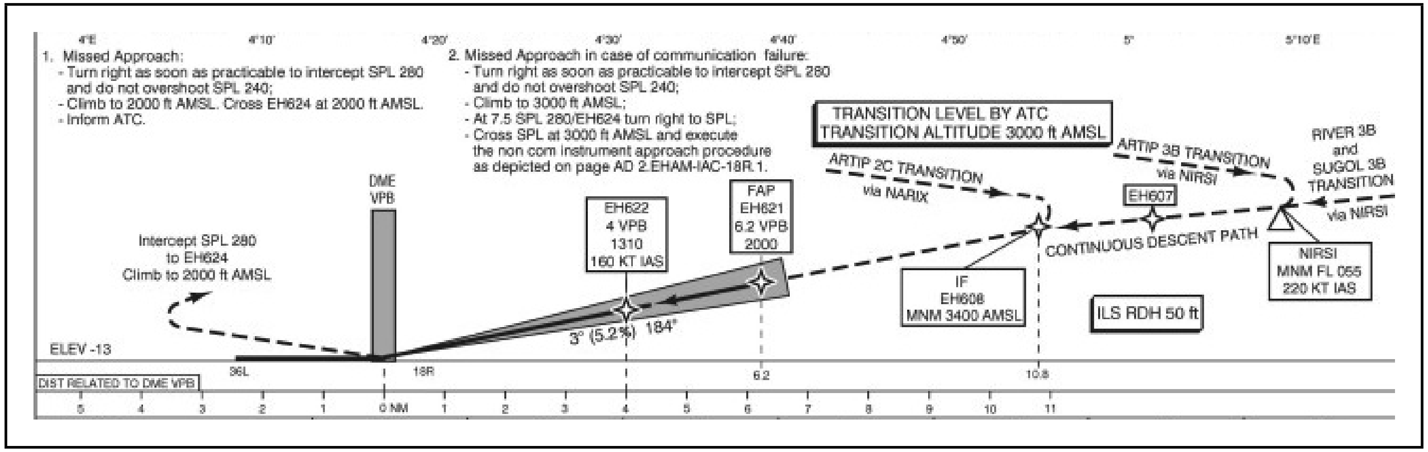
Dutch soil survey 1:50k sheet 34E; themes: soil class; ground-water level class; subsoil condition (“polkadot” overprint)



SSURGO 1:25k, Edgecombe County NC; theme: soil mapping units
n.b. extremely poor, non-connotative, colour scheme

Navigation maps – 1D and 2D



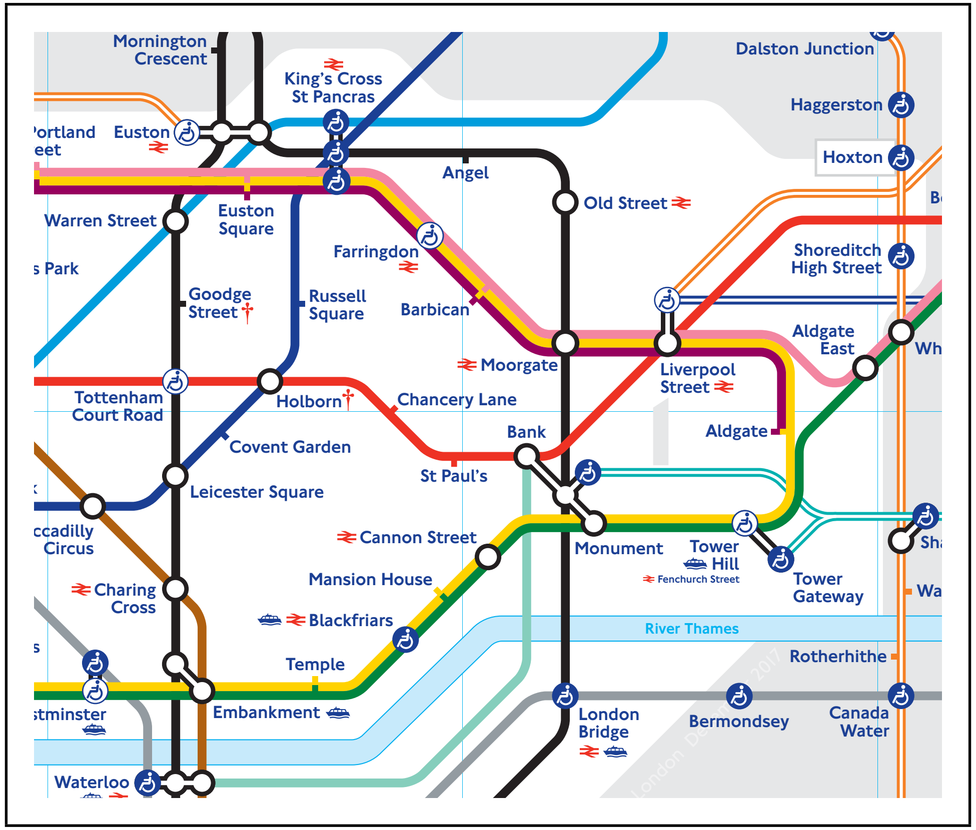


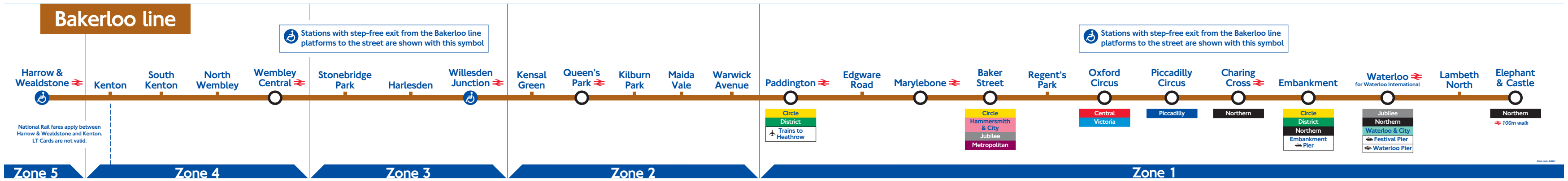
note: different X and Z scales; X scale aligned with runway 18R “Polderbaan”

Topologic geographic space

- Topologic relations are preserved;
  - adjacency, connectivity, containment, intersection . . .
- Distances are not true

True 1D, 2D topology, distorted distances and angles





Note: 3D not correct topology

Typology of spatial data – views of the world

Objects real-world **entities**: can be **discretely** identified “in the field” and located in geographic space

Fields **continuously-varying** properties in space

- Represented in a GIS by some **discretization**, but conceptually continuous
- Measured with some **spatial support** (sample size, instrument field of view, . . .), but conceptually continuous

Networks **interconnected** line and point objects

It's not so simple . . .

- the conceptual definition of the object may be vague (fuzzy **definition**)
- It may be difficult to identify an object in the field (fuzzy **identification**)
- Objects may have **fuzzy boundaries or locations**
- Object concepts may depend on the map **scale**
  - Roads and buildings are conceived of as polygons at large map scale, lines at small map scale
- Continuous variables are measured with some **spatial support** (sample size, instrument field of view, . . .); this is a lower limit of the **resolution** with which we can describe the conceptually-continuous field

Typology of spatial data – conceptual models of spatial objects

Point a single point location, defined by coordinates

Line a set of ordered points, connected by straight line segments

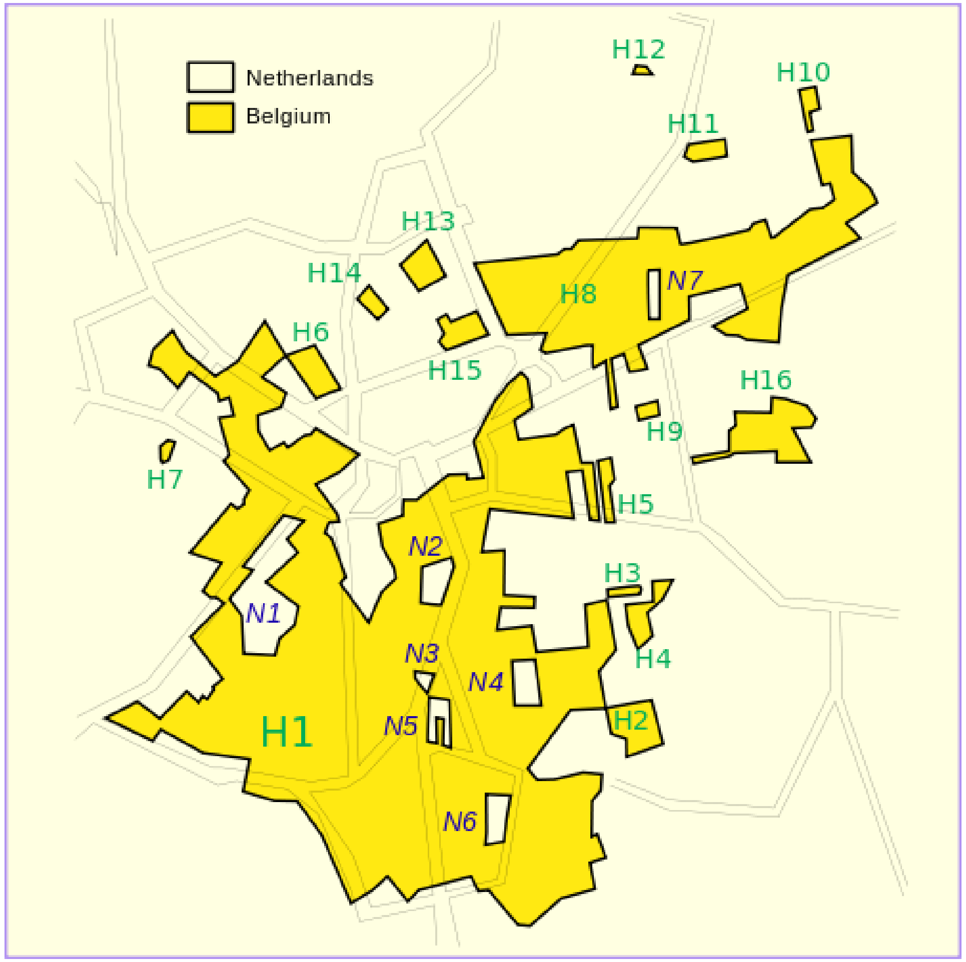
Curves a set of control points (not necessarily on the curve); and a mathematical function of coordinates (e.g., splines)

Polygon an area delineated by one or more lines, possibly containing holes (and holes within holes . . .)

Network a group of curves connected at points (**vertices**); mathematical “graph”

Grid a collection of points or cells, organised in a regular lattice covering an area

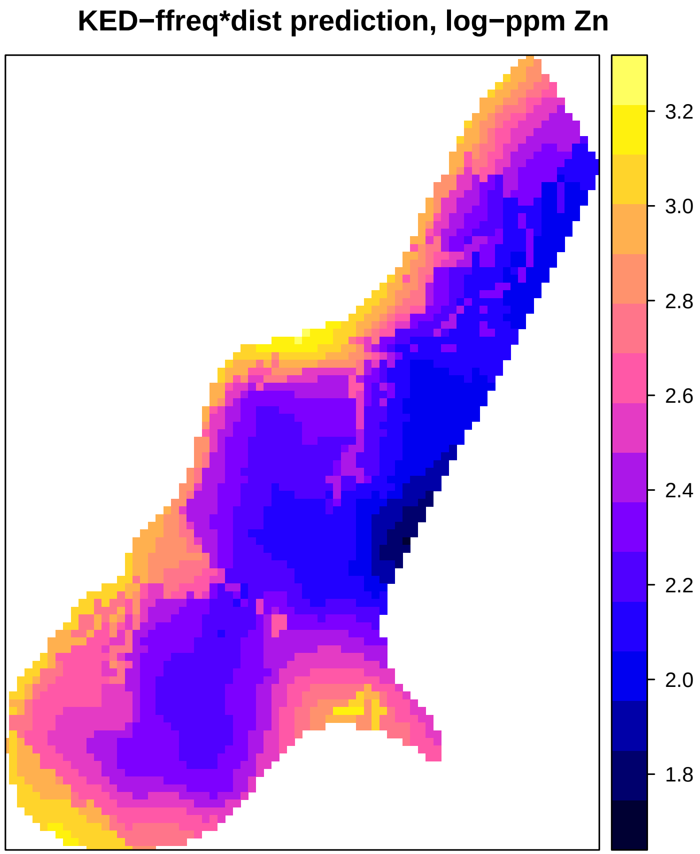
Polygons

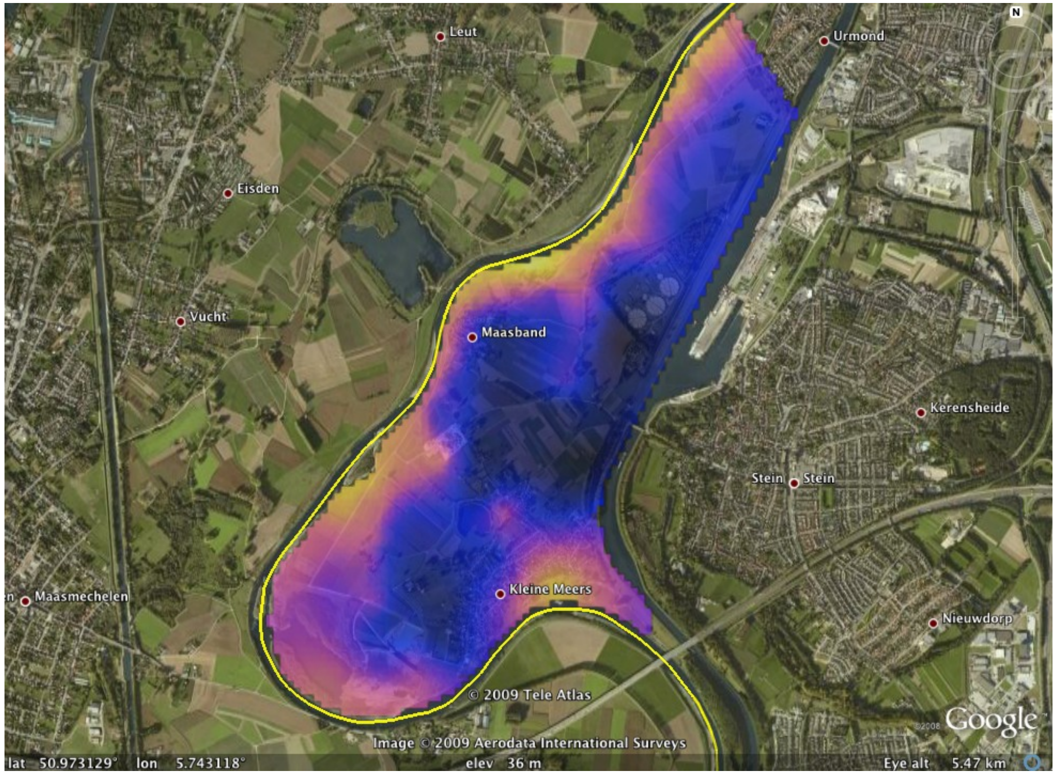


Baarle-**N**assau (NL), Baarle-**H**ertog (*Baerle-Duc*) (B); source Wikimedia commons

“H” polygons are B inside NL (i.e., holes in the surround unlabelled “N”); “N” polygons are N inside B (i.e., holes in a “H” polygon).

Grids





40 x 40 m square cells

conceptually smooth

Cell values could be centre point predictions, block averages, block maxima, block ranges or standard deviations . . .
in any case, one number represents the entire grid cell.

Typology of spatial data – data models

These are how spatial data are represented **inside a GIS** – *not* their conceptual representation

Vector exact mathematical form: 0-dimension = points; 1-dimension = line segments (which can be joined); 2-dimension = areas; 3-dimension = volumes

- Triangulated Irregular Network (TIN) is a vector data model of a 2-D continuous surface conceptual model

Raster a regular **tessellation** (e.g., square or hexagonal grid); fixed resolution

- data values may be grid cell averages, maxima, minima . . . or single values at the centre point

Concepts of scale

1. **cartographic** (map) scale: relation of map distances to ground distances
   - "large" = large area of paper needed to represent a given ground area
2. **geographic** scale: size of area being studied
   - "large"= over a wide ("large") area
3. **process** scale: spatial extent / variability of process operating on landscape
   - e.g., soil erosion: rill, plot, small catchment, river system . . .
4. **measurement** (observation) scale: size ("support") of observations
   - e.g., soil ped, core, profile, pit, trench, . . .
5. **modelling** scale: size of fundamental area at which processes or objects are represented in models ("support")

Blöschel, G., & M. Sivapalan. 1995. *Scale issues in hydrological modelling: a review.* Hydrol. Proc. 9: 251-290. DOI: https://dx.doi.org/10.1002/hyp.3360090305

Temporal scale

The above-mentioned **spatial** scales can also be used to describe **temporal** scales.

(Except for “cartographic”).

Matching scales

Key points:

- **modelling** scale should match **process scale**
- **information** at different measurement scales must be **harmonized**
  - e.g., satellite imagery at different resolutions
  - e.g., demographic information at census ward vs. postal code vs. administrative unit (these also at different levels)
  - up-, down-scaling (see below)

Up- and down-scaling

Upscaling from detailed scale (e.g., lab. experiments) to coarse scale (e.g., ag. field, region, continent . . .)

Downscaling from a coarse scale (e.g., general circulation model of the atmosphere) to a detailed scale (e.g., local weather forecast)

These can be either **spatial** or **temporal** scales.

Often the **inputs** to a model do not have the same scale, so some must be adjusted; and/or the input scale does not match the desired **output** scale.

Issues in spatial scaling

- **Up**scaling: must compress/summarize information
  - e.g., area weighted averaging of properties – but is this meaningful? (e.g., white car in black parking lot → grey pixel)
  - maybe re-run models with coarser scale inputs
  - maybe interpolate including information (somewhat) outside the upscaled resolution

- **Down**scaling: must create new **spatially-explicit** information at a finer scale
  - Just increasing pixel resolution is not creating information!
  - Example: disaggregating a coarse-resolution soil polygon (soil **association** with known landscape relation) to fine-resolution polygons (soil **consociation** = more-or-less homogeneous unit); using expert knowledge + covariates (e.g., terrain classification)

- Bui, E. N., & Moran, C. J. (2001). Disaggregation of polygons of surficial geology and soil maps using spatial modelling and legacy data. Geoderma, 103(1-2), 79-94.https://doi.org/10.1016/S0016-7061(01)00070-2
- Khan, M. R., de Bie, C. A. J. M., van Keulen, H., Smaling, E. M. A., & Real, R. (2010). Disaggregating and mapping crop statistics using hypertemporal remote sensing. International journal of applied earth observation and geoinformation, 12(1), 36-46.https://doi.org/10.1016/j.jag.2009.09.010

The modifiable areal unit problem

Issue: the **same analysis** may give **different results**, i.e., lead to different inferences / conclusions, if the data is aggregated at **different scales**

Example: voting patterns by large → small geographic area (polygon size):

- Vote for B.H. Obama vs. W.M. Romney, US President, 2012:
  1. USA (Obama 51.4%)
  2. NY state (Obama 62.6%)
  3. NY 23rd congressional district (Romney wins, Obama 48.4%)
  4. Tompkins County (Obama 68.2%)
  5. City of Ithaca (Obama 83.3%)
  6. 5th ward (Obama 85.2%)
  7. 5th ward 2nd district (Obama 91.6%),
- Summarize predictor variables at the same scales (party registration, census ethnicity, IRS-reported income . . .)
- **Do you expect the same predictive model?**