Topic 1: Introduction to Spatial Analysis

This topic gives an overview of main models and applications of the spatial analysis. In particular, we consider

- Some examples of **Spatial Data**.
- **Applications** that use spatial methods.
- **Types** of spatial data.
- **Objectives** of spatial analysis.

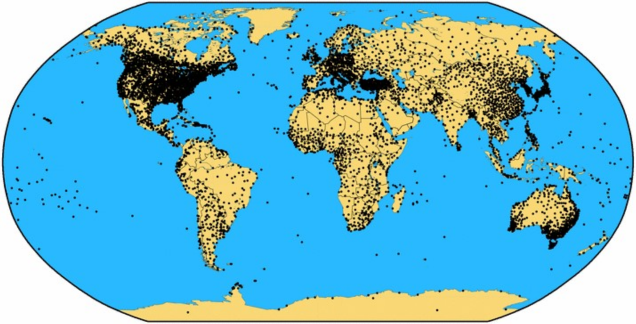
Spatial Data.

Spatial data are used to describe measurements or information that are referenced by specific spatial locations, for example, geographical areas or Earth locations. In numerous classical applications, spatial data represent information collected from the Earth in a corresponding geographic coordinate system.

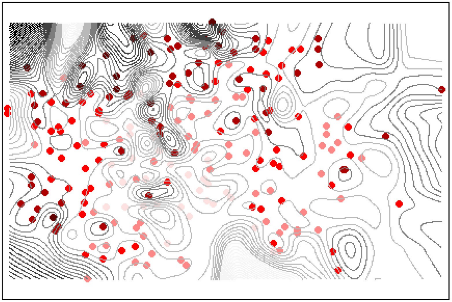
However, these days spatial data are used in many other applications and refer to information collected from any multidimensional spaces or surfaces, for example, brain studies, material surface data, high dimensional physics.

We provide just a few traditional examples of various spatial data:

Global Historical Climatology Network : temperature data for 7280 stations. All have at least 10 years of data.

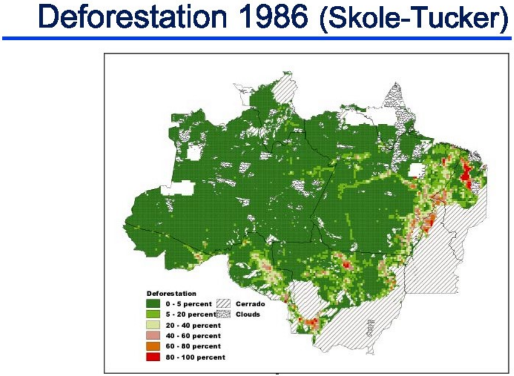


Prices of sold properties and contour lines of areas with the same prices:

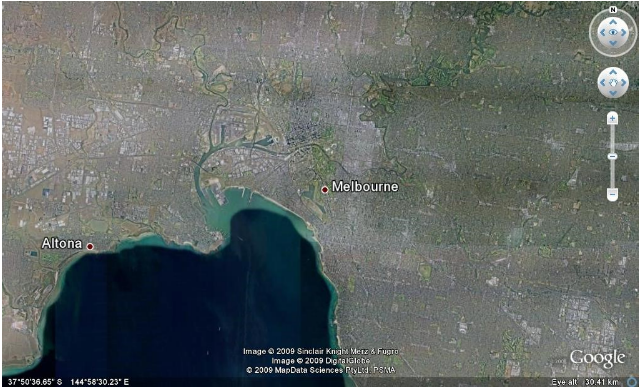


Residential Sales Price, Baltimore MD (1980)
sample points (darker is higher) and contours

Aggregated data (from historical land measurements and modern satellite images) showing deforestation of a specific region.

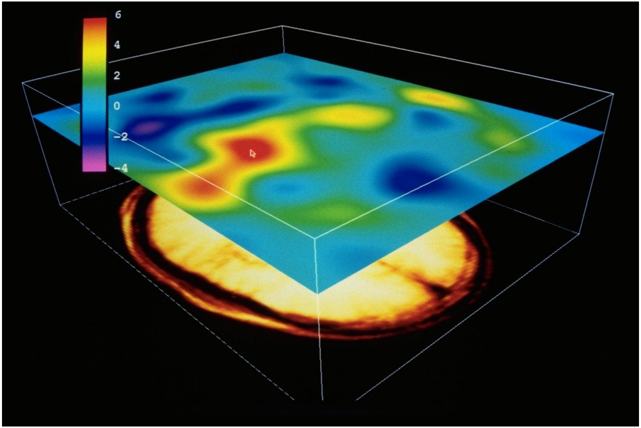


Images, in particular, satellite images showing distribution of objects of interest in a specific region.



GPS data, which often referenced by locations on a set of lines.

Sets of spatial data indexed by time or other (often discrete) indices. For example, a set of brain MRI scans.



Applications that use spatial methods.

Geostatistics

The aim of geostatistics is to study any phenomenon which can be quantified and which develops in a structured way in space (and time). Geostatistics can thus be presented as an extension of statistics or data analysis methods, in that it takes into account the spatial organization of the variables under study.

Geographic information systems

A geographic information system (GIS) integrates hardware, software, and data for capturing, managing, analyzing, and displaying all forms of geographically referenced information.

Environmetrics

Environmental themes include climate change, atmospheric pollution, issues of data collection and interpretation including the design of a monitoring network, and the assessment of probabilities associated with extreme events such as floods and hurricanes.

Encyclopedia of Environmetrics, Wiley, 2002, 2800 pages

Types of spatial data.

Spatial analysis (statistics) is a vast subject, in large part because

- spatial data have many different types:
  - univariate or multivariate,
  - categorical or continuous,
  - real-value (numerical) or not real-valued,
  - observational or experimental;

- the data locations may:
  - be points, regions, line segments, or curves,
  - be regularly or irregularly spaced,
  - regularly or irregularly shaped,
  - belong to a Euclidean or non-Euclidean space.

Distance between two points

Can not be measured in real-life/realistically.

Four important data prototypes:

1. Geostatistical data,
2. Lattice data,
3. Spatial point patterns,
4. Spatial set patterns.

The distinctions between these types are not always clearcut.

Geostatistical Data:

Point observations of a *continuously* varying quantity over a region.

Examples:

- Weekly concentrations of ozone in U.S.
- Annual acid rain deposition in U.S.
- Richness of iron ore within an ore body

One important problem in geostatistics is to predict the ore grade in a mining block from observed samples (Matheron, 1963).

Lattice Data:

Counts or averages of a quantity on subregions that make up a larger region. These data are referred to a lattice.

A lattice of locations often evokes an idea of regularly spaced points. These will be referred to as *regular* lattices, allowing for the possibility of *irregular lattices*, whose relative displacements do not follow a predictable pattern. Statistical models for lattice data need to express the fact that observations nearby (in time or space) tend to be alike.

Examples:

- Presence or absence of a plant species in square quadrats over a study area,
- Number of deaths due to SIDS(Sudden infant death syndrome) in the counties of North Carolina,
- Pixel values from remote sensing (satellites).

Spatial Point Patterns:

Points on a map. Point patterns arise when the important variable to be analyzed is the location of *events*.

The question of interest is whether the pattern is exhibiting complete spatial randomness, clustering, or regularity.

Examples:

- Location of bird's nests in a suitable habitat – evidence of territoriality?
- Location of longleaf pines in a natural forest in North Carolina – evidence of clustering?
- Diameter of the longleaf pines – Do larger trees cluster?

The points may have extra information called marks (for example, the size variable in example 3) attached to them. The mark represents an attribute of the point.

Spatial Set Patterns:

Random geometrical objects. The question of interest is to investigate distributions of these objects on images.

an independent variable that can influence the outcome of a given statistical trial, but which is not of direct interest

Examples:

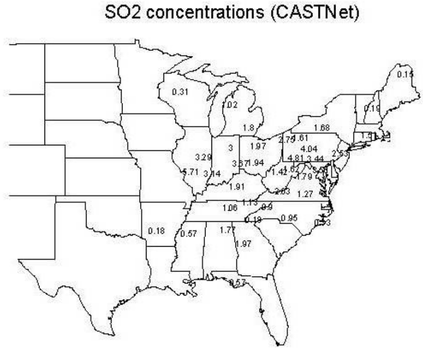
- randomly placed particles,
- images found in microscopy, material science and biology,
- texture analysis of images.

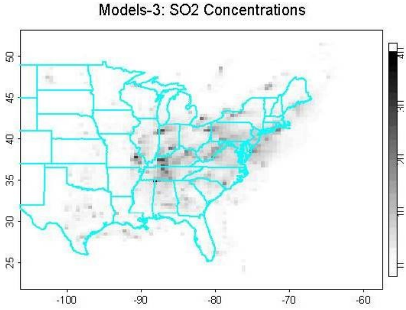
Spatial datasets may also include **covariates,** which are any data that we treat as explanatory, rather than as part of the response. Covariate data may be geostatistical data defined at all spatial locations, e.g. altitude, soil pH, displayed as a pixel image or a contour plot. Covariate data may be another spatial pattern such as point pattern, or a line segment pattern, etc.

Objectives of Spatial Analysis

1. Data preparation: storing, indexing, querying, and manipulating geographic information and spatial data.
2. Visual analysis: output a detailed map, image or movie used to communicate an idea or concept with respect to a region of interest.
3. Statistical estimation: estimating treatment effects in spatial experiments, estimating the autocorrelation structure.
4. Construction of theoretical models.
5. Prediction, interpolation. Prediction of unobserved variables (kriging).
6. Design issues. Where to take observations or how to arrange treatments in a spatial experiment.
7. Simulation: the imitation of some real data or processes.

For example, the first image shows values of SO_2 concentration at locations where these measurements were taken. The second image shows the prediction map obtained by using these measurements.





Temporal vs. Spatial Statistics

- Time flows in one direction only, from past to present to future. Not so in space. It results in much more complicated theoretical models. For example, there are many different definitions of Markov property.
- In time series, observations usually regularly spaced.
- In time series models, observations usually are assumed to be dependent but identically distributed. In spatial statistics observations are assumed to be non-identically distributed (trend).
- In time series, interaction is unidirectional. In space, interaction occurs in all directions.
- In time series, prediction usually is extrapolating to a future time. In space, interpolation is usually more important.

Spatiotemporal Statistics

Spatiotemporal data are observations with identifiable and observed spatial and temporal labels.

Examples:

- earthquakes (locations random in time and space)
- change in locations of trees over time (locations are random in space but nonrandom in time)
- environmental monitoring of water quality (locations are nonrandom in time and space)

Very often space-time data are modeled as:

- a collection of temporally correlated spatial random fields, lattice processes, or spatial point processes, or
- a collection of spatially correlated time series.

Unfortunately such approaches do not work properly in many cases and new models are needed.

Besides a simple regression model for example (more models will be covered in R later this semester).