STAT 8025

Lecture 2: Spatial Random Processes

Dr. Emily Lei Kang

Division of Statistics & Data Science
Department of Mathematical Sciences
University of Cincinnati

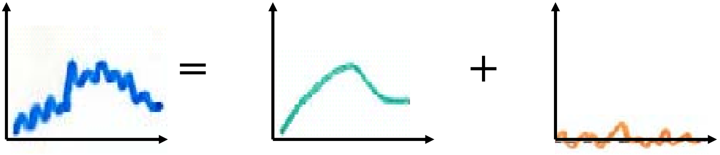
University of Cincinnati

Recap...

- Spatial Statistics: The data's spatial locations play a role in a statistical analysis
- Types of Spatial Data
  - Geostatistical data (also called point-referenced data)
  - Lattice data (also called areal data)
  - Point patterns
- Examples:
  - Annual acid rain deposition in US
  - Population-adjusted mortality rates due to a certain disease in counties of Ohio
  - Locations of residences of individuals with lung cancer within 50 miles of a large incinerator

Structure of Spatial Phenomena

Spatial processes can be decomposed into different components:



- Large-scale variation (also called the first order effect)
- Small-scale variation (also called the second order effect)

Structure of Spatial Phenomena

Large-scale (First Order) Effect

- Variation of the mean value in space, also called *global or large scale trend*
- It measures spatial heterogeneity at the **large scale**
- Modeling usually employs conventional statistical methods (e.g. linear models)
- Covariates are often used to estimate the first order effect, e.g. $\mu(\mathrm{s}) = \mathrm{x}(\mathrm{s})'\boldsymbol{\beta}$

Structure of Spatial Phenomena

Second Order Effect

- Correlation in the deviations of values of the process from the mean
- Results from spatial autocorrelation structure – local or small scale effects; this is usually the core of spatial statistics
- Often deviations from the mean "follow" each other at neighboring sites, resulting in *positive spatial dependence*, but there may also be negative spatial dependence due to competition

Structure of Spatial Phenomena

Some Comments

- The distinction between the first and second order effects is usually a *modeling decision* and does not necessarily represent reality
- Often it depends on the scale and purpose of the study, which effects should be modeled as first order and which as second order
- The combined first and second order effects violate the *iid* assumption made in conventional statistics

Two important types of spatial structure are stationarity and isotropy.

- Stationarity
  - Constant large-scale structure
  - Small-scale structure which depends on the spatial locations only through the their relative positions
- Isotropy
  - Stationary
  - Small-scale structure depends on the spatial locations only through the Euclidean distance between them.
- Formal definitions will be given later

Objectives of Spatial Statistics

- Estimation
  - Estimating treatment effects
  - Estimate the autocorrelation structure, such as spatial structural parameters
- Prediction
  - Prediction of unobserved variables (e.g., spatial prediction/kriging)
- Design Issues
  - Where to take observations or how to arrange treatments in a spatial experiment (briefly discussed later this semester if time permits)

Temporal vs. Spatial Statistics

- Time flows in one direction only, from past to present to future. Not so in space.
- Contrast between time series and geostatistics/lattice data analysis:
  - In time series, observations usually regularly spaced. In geostatistics particularly, but also in lattice data analysis, irregularly spaced data are at least as common as regularly spaced data. So geostatistical and lattice models must be more flexible.
  - In time series models, observations sometimes are assumed to be dependent but identically distributed. In Geostatistics observations are usually assumed to be non-identically distributed (trend).
  - In time series, prediction usually is extrapolating to a future time. In space, interpolation is usually more important – we can "go back" in space.

Traditional Numerical Summaries

- Numerical summaries: Mean, median, mode, standard deviation, range, interquartile range (IQR).
  - They reduce the data to a few numbers, not useful for geostatistical data, because the numerical summaries ignore the location.
- Plots:
  - Histogram: Gives more information about the distribution of the data (if the data are iid), but location is ignored.
  - Maps, 3D scatter plot, contour plots (requiring some smoothing)
  - Plot data versus each marginal coordinate

Explore the large-scale structure

- Median polish of Z

Median Polishing

Assume that either data are gridded or that a "low resolution" map has been produced by overlay a grid on original data and shifting data points to the nearest grid node.

- Let the rows of the grid be indexed as $k = 1, \ldots, p$ and the columns of the grid be indexed as $l = 1, \ldots, q$
- Take the gridded data to be $\{Y_{k,l} : k = 1, \ldots, p; l = 1, \ldots, q\}$. Note that Y's will be either the original data values (if they are on a grid already) or the shifted values
- Consider the following model for $\mu(\cdot)$:

$$\mu(k, l) = a + r_k + c_l; k, l \in \mathbb{G}$$

where $\mathbb{G}$ is our grid, the constants $\{r_k\}$ and $\{c_l\}$ are the row and column effects, respectively.

Median Polishing

1. Compute the median for each row in the grid. Remove the row median from each observation in a given row (i.e. subtract the median value from each observation)
2. Compute the median in each column of the grid. Remove the column median from each observation in a given column
3. Repeat this procedure until another iteration produces virtually no change
4. Final entries in the extra cells are the median polish estimates of row effects $\{r_k\}$, column effects $\{c_l\}$, and an overall effect a

Example

6	3	11
3	2	4
9	0	0

0	−2	4	3
0	0	0	0
8	0	−2	−2
0	−1	1	3

- Reading: Sun, Y. and Genton, M. G. (2012) Functional Median Polish, Journal of Agricultural, Biological, and Environmental Statistics, 17, 354–376.

Explore the small-scale dependence

- h-scatterplots (lag scatterplots)
  - Assume we observe the process Z at equally spaced locations $s_1, \dots, s_n$.
  - Plot $Z(s_i + he)$ versus $Z(s_i)$ for a fixed vector e (defines direction) of unit length, a fixed scalar h, and for all $i = 1, \dots, n$
  - Requires regular spacing between data locations.
  - Positive (negative) correlation in plot indicates positive (negative) spatial dependencc at that lag.
  - Outliers can be detected with this graph.

Variogram Cloud

* detrended data used; that is, the large-scale structure is fit and subtracted

- Historically has been used for spatial data much more than temporal data.
- Plot $(Z(\mathrm{s}_i) - Z(\mathrm{s}_j))^2$ versus $[(\mathrm{s}_i - \mathrm{s}_j)'(\mathrm{s}_i - \mathrm{s}_j)]^{1/2}$ (Euclidean distance) for all possible pairs of observations.
- The variogram cloud implicitly assumes isotropy (does not differentiate any directions).
- The square-root-differences cloud, use $(Z(\mathrm{s}_i) - Z(\mathrm{s}_j))^{1/2}$ can be more resistant to outliers.

Sample semivariogram

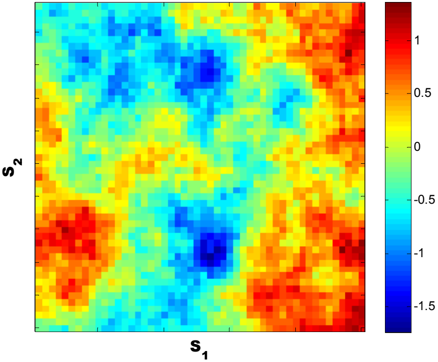
- The traditional sample semivariogram $\hat{\gamma}$ suggested by Matheron (1971) is:

$$\hat{\gamma}(\mathrm{v}) = \frac{1}{2N(\mathrm{v})} \sum_{N(\mathrm{v})} (Z(\mathrm{s}_i) - Z(\mathrm{s}_j))^2$$

  where $N(\mathrm{v})$ are the number of data pairs s_i and s_j separated by v.
- Plot $\hat{\gamma}(\mathrm{v})$ versus different values of v.
- Note that this implicitly assumes stationarity of some kind.
- You can display the variogram along selected directions (e.g., N-S, NW-SE, E-W, and NE-WE) on the same 2-D graph.

Toy Example

Consider a simulated spatial data

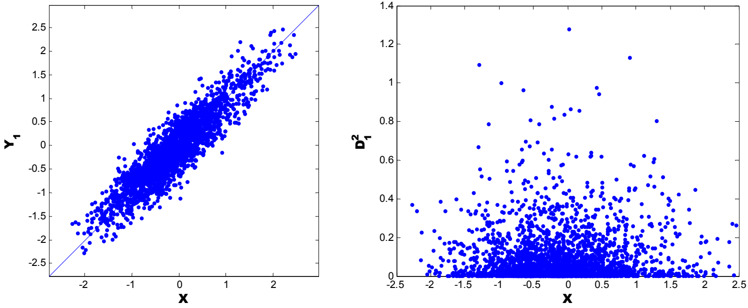


$$Y(\mathrm{s}) = Y((s_1, s_2)), s_1, s_2 = 1, \ldots, 50$$

Toy Example, cont'd

Choose E-W direction and $h = 1$

$$X = Y((s_1, s_2)); Y_1 = Y((s_1 + 1, s_2))$$
$$D_1^2 = (X - Y_1)^2$$

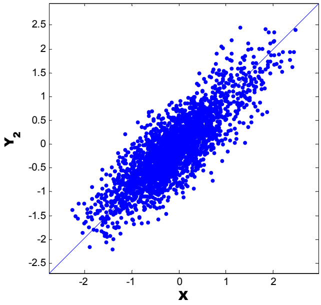


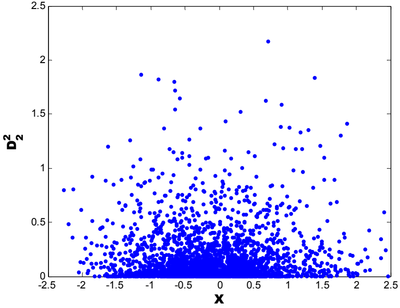
UNIVERSITY OF Cincinnati

Toy Example, cont'd

Choose E-W direction and $h = 2$

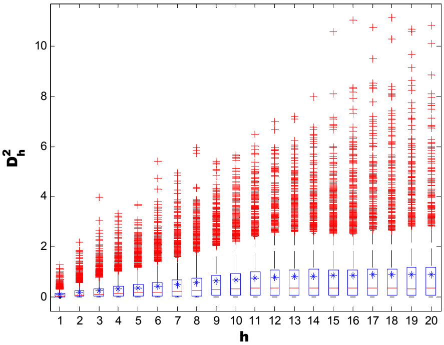
$$X = Y((s_1, s_2)); Y_2 = Y((s_1 + 2, s_2))$$
$$D_2^2 = (X - Y_2)^2$$





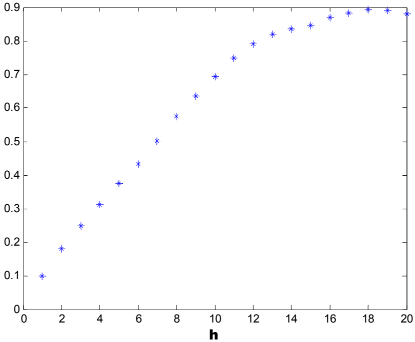
Toy Example, cont'd

- Make a boxplot of $D_1^2, D_2^2, \ldots$ and put them side by side
  - D^2 small: High spatial continuity [smooth, predictable]
  - D^2 large: Low spatial continuity [rough, unpredictable]



* shows ave$\{\cdot\}$ or sample mean

The empirical variogram for the E-W direction is:



Summary

- Structure: large-scale, small-scale
- Exploratory data analysis
- `R` examples

Preview:

- Geostatistical Models