GEOG5330: Applied Spatial and Spatiotemporal Analysis

Lecture 1: Overview

Guofeng Cao

www.spatial.ttu.edu

Department of Geosciences
Texas Tech University

`guofeng.cao@ttu.edu`

Fall 2018

Geographic Information Science and Technology (GIST)

- Three core courses
  - GIST 5300: Geographic Information Systems (3)
  - **GIST 5302: Spatial Analysis and Modeling (3)**
  - GIST 5304: Advanced Geographic Information Systems (3)
- Two elective course from
  - **GEOG 5330. Applied Spatial and Spatiotemporal Analysis (3)**
  - GIST 5308. Cartographic Design (3)
  - GIST 5310. GPS Field Mapping (3)
  - GIST 5312. Internet Mapping (3)
  - GEOG 5301. Remote Sensing of the Environment (3)
  - GEOL 5341. Digital Imagery in the Geosciences (3)
  - GEOL 5342. Spatial Data Analysis and Modeling in Geosciences (3)
  - NRM 5404. Aerial Terrain Analysis (4)

Course Description

Content overview

- This course will introduce concepts and commonly used methods in quantitative analysis of (geographic) spatial data
- Contents include:
  - Characteristics of spatial data
  - Representation of spatial data in GIS
  - Commonly used spatial analysis methods
  - Concepts in spatial statistics
- Class webpage: http://www.gis.ttu.edu/gist4302

Course Description

Audience

- This class is intended for students (undergraduate and graduate students) from relevant disciplines (e.g., geography, geology, environmental science and social sciences) who are interested in analysis of spatial data
- Students will be encouraged to engage this course with their thesis/dissertation topics and research interests

Please note:

- The format might be different from other classes you have taken in the GIST program
- General knowledge of statistics or quantitative skills will be very helpful and equations will appear significantly in some lectures
- We have a mix of undergraduate and graduate students with diverse background and expectation

Course Description

Course objectives

- After completing this course, undergraduates students are expected to learn how to:
  - formulate real-world problems in the context of geographic information systems and spatial analysis
  - utilize mainstream software tools (commercial or open-source) to solve spatial problems
  - communicate results of spatial analysis in the forms of writing and presentation
  - have a good understanding of commonly used spatial and spatiotemporal analysis methods
  - apply the introduced methods in the dissertation and thesis research
  - evaluate and assess the results of alternative methods

Course Format

Lectures

- Instructor: Guofeng Cao (guofeng.cao@ttu.edu)
- Science 221
- T: 6:00-8:50pm
- Office hours: Th: 3:00-4:00pm at Holden Hall 211

Lab Assignments

Lab assignments

- ~ 3 hours each week
- Using R

Final project

- The project could be used as a setting for your thesis and dissertation topics, other course topics or research interests
- Group collaboration is encouraged, but for each group, no more than two graduates are allowed
- Start to think of the project ideas early and communicate with the instructor and TA for comments
- Project presentation or poster session: *PechaKucha* style http://en.wikipedia.org/wiki/PechaKucha
- Project report: no more than 8 pages with single space and size 12 font

Grading

Grading policy

- No written exams
- Nine lab assignments: 70%
- Final project: 30% including proposal (5%), class presentation (10%) and project report (15%)
- Class and lab attendance is mandatory

Textbook

- Required
  - Bivand Roger S., Pebesma, Edzer J., and Gmez-Rubio, Virgilio (2008), *Applied Spatial Data Analysis with R*, Springer.
- Optional:
  - Cressie, Noel, and Christopher K. Wikle. Statistics for spatio-temporal data. John Wiley & Sons, 2015.

Topics

- Spatial data representation and manipulation
  - R Basics
  - GIS using R
- Point pattern analysis
  - Species distribution modeling (e.g., MaxEnt)
- Areal data analysis
  - Exploratory analysis for cluster detection
  - Spatial and spatiotemporal regression
  - Change-of-support problem
- Geostatistics
  - Kriging family
  - Bayesian geostatistics
  - Space-time kriging
- Time series analysis
  - Remote sensing imagery
  - Google Earth Engine
- Geospatial uncertainty

Logistics

- R and RStudio
- Github account
- Withdrawing: You are responsible for dropping the class

Introduction to Statistics of Spatiotemporal Data

Components of Spatiotemporal Analysis and Modeling

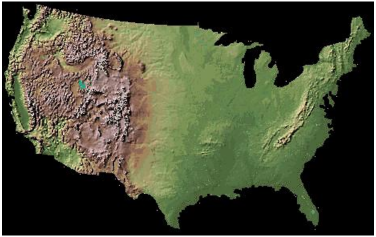
- Data do not equal information
- Components of spatial analysis (geospatial data in particular)
  - Visualization: Showing interesting patterns (mapping, geovisualization)
  - Exploratory spatial data analysis: Finding interesting patterns
  - Spatial modeling, regression: Explaining interesting patterns

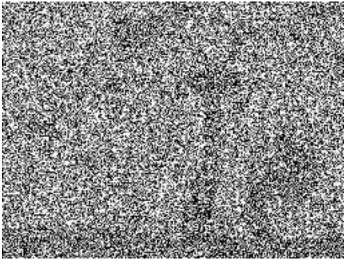
Characteristics of (Geographic) Spatial Data

1. Spatial (and temporal) Context: “Everything is related to everything else, but near things are more related than distant things”
    - Waldo Toblers First Law (TFL) of geography
    - nearby things are more similar than distant things
    - phenomena vary slowly over the Earth’s surface
    - Compare time series

Characteristics of (Geographic) Spatial Data

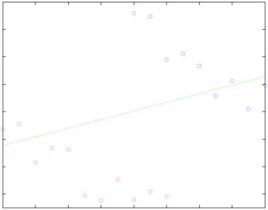
- Implication of Tobler's First Law (TFL)
  - We can do samplings and fill the gap using estimation procedures (e.g. weather stations)
  - Spatial patterns
  - Image a world without TFL:
    - White noise
    - No lines, polygons or geometry (how to draw a polygon on a white noise map?)



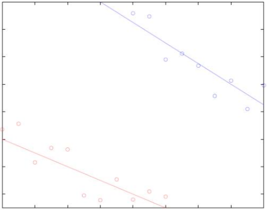


Characteristics of (Geographic) Spatial Data

2. Spatial heterogeneity
   - "Second law of geography" (Goodchild, UCGIS 2003)
   - Earths surface is non-stationary
   - Laws of physical sciences remain constant, virtually everything else changes
     - Elevation,
     - Climate, temperatures
     - Social conditions
   - Implications
     - Global model might be inconsistent with regional models
     - Spatial Simpsons Paradox (a special case of modified areal unit problem, which we will discuss more in the later of this class)



(a) Global Model



(b) Regional Models

Side note: example of Simpson's paradox

- Simpson's paradox usually fools us on tests of performance in real life
- The following is a real life example. Comparison of recovery rates between a new treatment and a traditional treatment for kidney stones.

	New Treatment	Traditional Treatment
Small Stones	93%(81/87)	87%(234/270)
Large Stones	73%(192/263)	69%(55/80)
All	78%(273/350)	83%(289/350)

- Comparison of batting average of two baseball players:

	1996	1997	Combined
Derek Jeter	25.0%(12/48)	31.4%(183/582)	31.0%(195/630)
David Justice	25.3%(104/411)	32.1%(45/140)	27.0%(149/551)

Characteristics of (Geographic) Spatial Data

- In a spatial settings, it is related to modified areal unit problem (MAUP) or omitted variable problem, which will discuss more in the later of this class

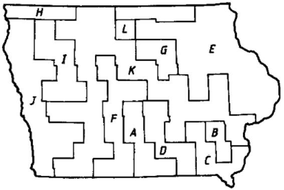


Figure 2a. Zoning system that minimises the regression slope coefficient (-24, r = -.25)

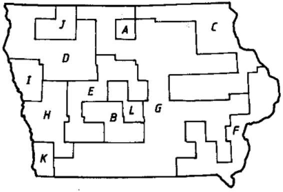


Figure 2b. Zoning system that maximises the regression slope coefficient (12, r = .87)

Figure: Image Courtesy of OpenShaw

Characteristics of (Geographic) Spatial Data

3. Fractal behavior
    - What happens as scale of map changes?
    - Coast of Maine
- Implications
    - Scale is critical for the problem of study
    - Volume of geographic features tends to be underestimated
        - length of lines
        - area of polygons
    - Think of the difference of distances that an ant and elephant needed to travel from where I stand to the center of memorial circle

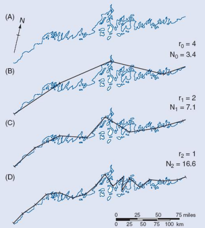


Figure 4.18 The coastline of Maine, at three levels of recursion: (A) the base curve of the coastline; (B) approximation using 100 km steps; (C) 50 km step approximation; and (D) 25 km step approximation

Microscale or Macroscale

Please try to tell whether the following maps are micro- or macro-scale:

- `http://www.smithsonianmag.com/science-nature/macro-or-micro-test-your-sense-of-scale-2208824/?no-ist=`

Characteristics of (Geographic) Spatial Data

Summary: three interrelated characteristics of spatial data

- Spatial context/spatial pattern/spatial structure/spatial dependence/spatial texture..
- Spatial heterogeneity/locality
- Fractal behaviors/scaling effects

Elements of Spatial Data

Elements

- Georeferenced measurements (point or area/region specific samples) Spatial arrangement: regular or irregular (gridded or scattered sampling locations)
- variables/attributes: continuous or discrete (e.g., chemical concentration, soil types, disease occurrences)
- auto- and cross-correlation endemic to spatial data (Toblers first law of Geography)

Types of spatial data

- Point pattern data
- Areal data
- Geostatisticla data
- Spatial interaction or network data

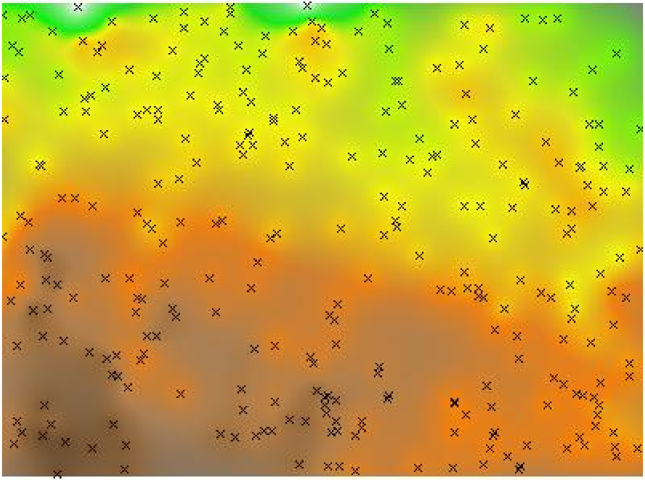
Types of Spatial Data: Geostatistical Data

Geostatistical data

- Attributes vary continuously in space, e.g., temperature, rainfall, elevation
- Measurements of nominal scale (e.g., soil types), or interval/ratio scale (e.g., depth of boreholes)
- Sampling only at fixed set of locations
- Occurs often in physical-related sciences

Types of Spatial Data: Geostatistical Data

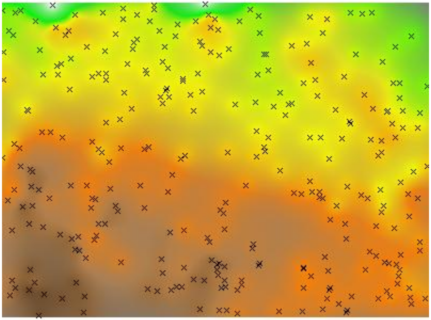
Example: *300 randomly placed points*



Objective

- Mapping spatial variations of regional variables
- Make estimation at unsampled locations

Example: elevation surface generated from 300 points

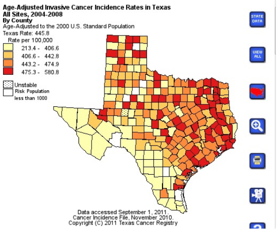


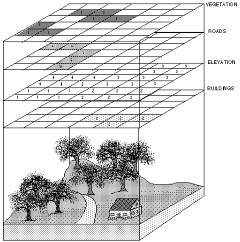
Types of Spatial Data: Areal Data

Areal (lattice) data

- attributes take values only at fixed set of areas or zones, e.g., administrative districts, pixels of satellite images
- Attributes distribute homogeneously within a region
- Lattice or uniform raster data could be taken as a special case of this type of data

Example:



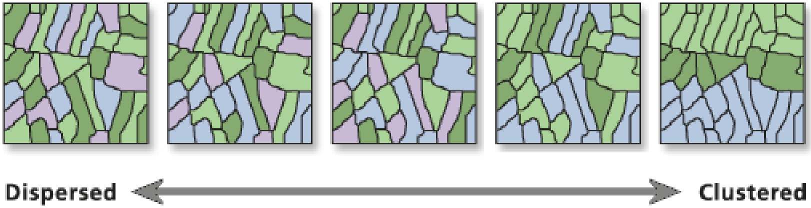


Types of Spatial Data: Areal Data

Objective

- Detect and model spatial patterns or trends in areal values
- Use covariates or relationships with adjacent areal values for inference (e.g., disease rates in light of socioeconomic variables)

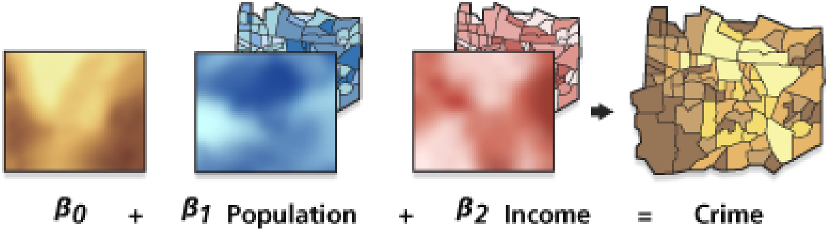
Example:



Types of Spatial Data: Areal Data

Example 2: *find the correlation among maps*

- It is analog to the cases in traditional statistics, but each variable is (multidimensional) 'maps' instead of single 'numbers'



Point pattern data

- series of point locations with recorded events, e.g., locations of trees, epic centers, disease or crime incidents
- attribute values also possible at same locations, e.g., tree diameter, magnitude of earthquakes (marked point pattern)

Example

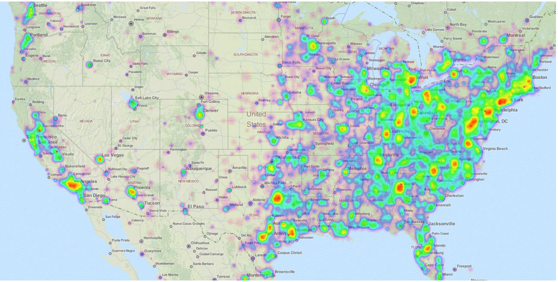
Types of Spatial Data: Point Pattern Data

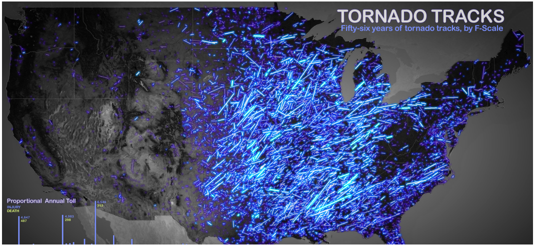
Objective

- detect clustering or regularity, as opposed to complete randomness, of event locations (in space and time)
- If abnormal clustering detected, investigate possible relations with potential factors, e.g., density of disease occurrences with socio-economic status
- Difference with geostatistical point data

Types of Spatial Data: Point Pattern Data

Example:





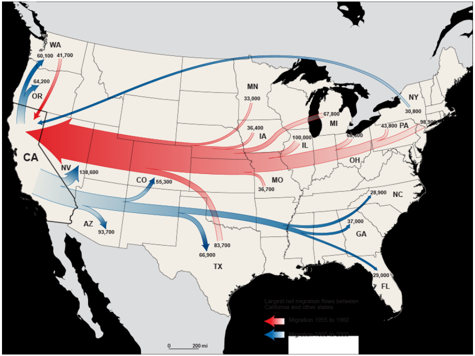
Spatial interaction or network data

- Topological space (not Euclidean space)
- Attributes relate to pairs of points or areas: flows from origins to destinations, e.g., population migrating from CA to TX
- Mostly interested in spatial patterns of aggregate interaction, rather than individuals themselves
- Not a major topic of this class

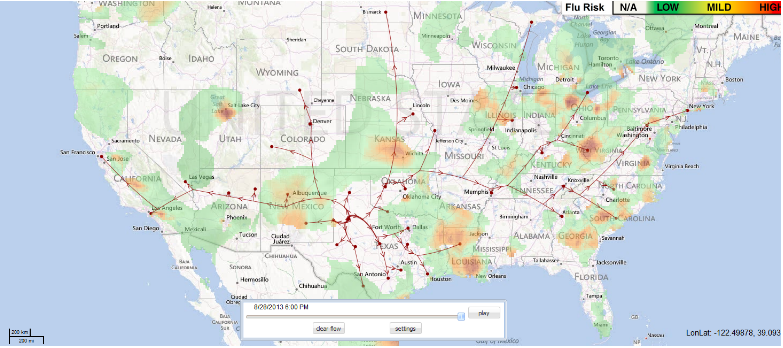
Objective

- Modeling of flow patterns
- Mostly interested in spatial patterns of aggregated interaction, rather than individual behaviors

Example



Example


Flu Risk
N/A
LOW
MILD
HIGH
8/28/2013 6:00 PM
play
clear flow
settings
LonLat: -122.49878, 39.093
200 km
200 mi

Types of Spatial Data: Summary

Summary

- Geostatistical data
- Spatial point pattern
- Areal (lattice) data
- Spatial interaction/network data

Thank you

Questions/comments?