Spatial Data

SERGIO REY

GPH 483/598
Geographic Information Analysis
School of Geographical Sciences and Urban Planning
Arizona State University
Fall 2010

Outline

1. Spatial Data
   - Types
   - Spatial Effects
2. Spatial Data Analysis
   - Spatial Process
   - Point Pattern Analysis
   - Lattice
   - Geostatistical
   - Network Data
3. Transformations of Spatial Data
   - Area to Point
   - Point to Area
   - Change of Support Problem

Outline

Outline

1. Spatial Data
   - Types
   - Spatial Effects
2. Spatial Data Analysis
   - Spatial Process
   - Point Pattern Analysis
   - Lattice
   - Geostatistical
   - Network Data
3. Transformations of Spatial Data
   - Area to Point
   - Point to Area
   - Change of Support Problem

Outline

Spatial Data is Special

> *Spatial data comes in many varieties and it is not easy to arrive at a system of classification that is simultaneously exclusive, exhaustive, imaginative, and satisfying.*

– G. Upton & B. Fingelton

Types of Spatial Data

Events

addresses of crimes

Continuous surfaces

air quality, rainfall

Discrete spatial objects

county income

What is special about spatial data?

Location, Location, Location

where matters

Dependence is the rule, not the exception

- spatial interaction, contagion
- spatial externalities
- spillovers, copy-catting

Spatial Scale

Inference can change with scale

Nature of Spatial Data

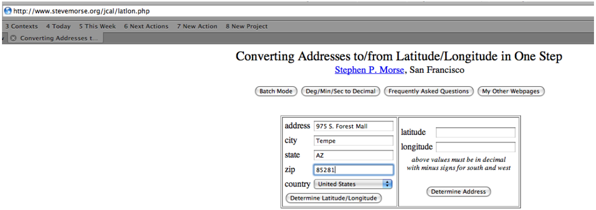
Georeferenced

attribute data together with location

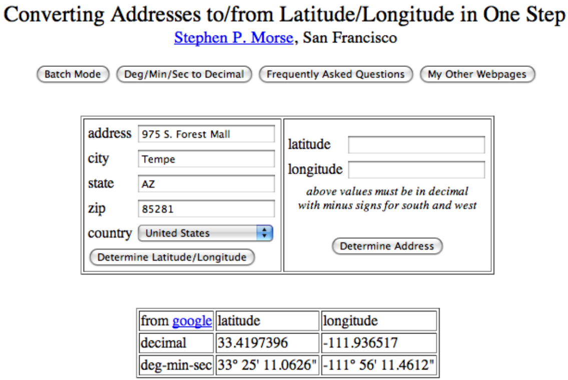
Geocoding

- associate observations with location
- point: latitude-longitude (GPS)
- areal unit: spatial reference

Geocoding on-line



Where is the office?



Converting Addresses to/from Latitude/Longitude in One Step

Stephen P. Morse, San Francisco

Batch Mode | Deg/Min/Sec to Decimal | Frequently Asked Questions | My Other Webpages

address 975 S. Forest Mall

city Tempe

state AZ

zip 85281

country United States

Determine Latitude/Longitude

latitude

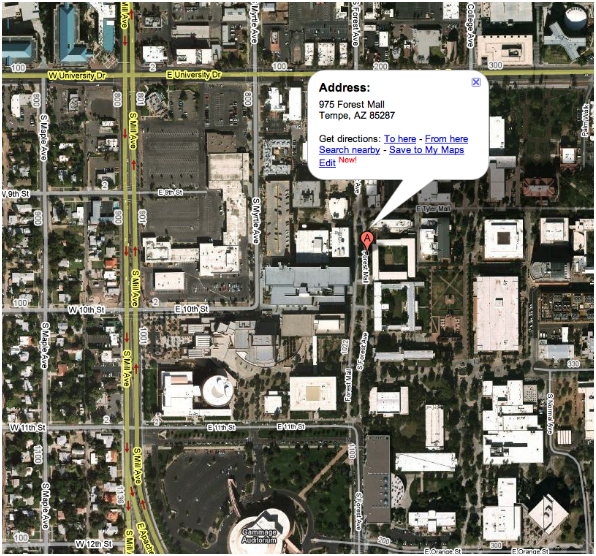
longitude

above values must be in decimal with minus signs for south and west

Determine Address

from google	latitude	longitude
decimal	33.4197396	-111.936517
deg-min-sec	33° 25' 11.0626"	-111° 56' 11.4612"

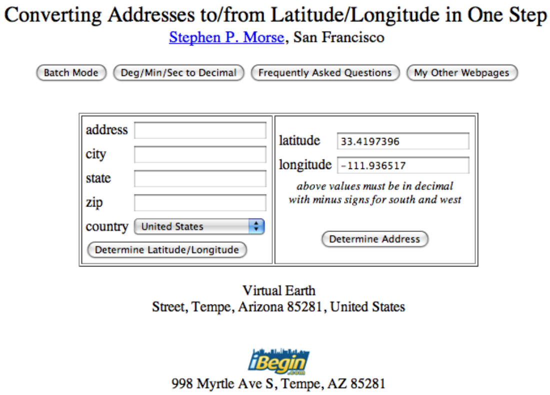
Geocoding: google link



Geocoding


Converting Addresses to/from Latitude/Longitude in One Step
Stephen P. Morse, San Francisco
Batch Mode
Deg/Min/Sec to Decimal
Frequently Asked Questions
My Other Webpages
address
city
state
zip
country
United States
Determine Latitude/Longitude
latitude
33.4197396
longitude
-111.936517
above values must be in decimal
with minus signs for south and west
Determine Address

Geocoding


Converting Addresses to/from Latitude/Longitude in One Step
Stephen P. Morse, San Francisco
Batch Mode
Deg/Min/Sec to Decimal
Frequently Asked Questions
My Other Webpages
address
city
state
zip
country
United States
Determine Latitude/Longitude
latitude
33.4197396
longitude
-111.936517
above values must be in decimal with minus signs for south and west
Determine Address
Virtual Earth
Street, Tempe, Arizona 85281, United States
iBegin
998 Myrtle Ave S, Tempe, AZ 85281

Location

Location as a Given

- in most spatial data analyses no choice in location
- no sampling in the usual sense
- data = attributes augmented with locational information

Outline

1. Spatial Data
   - Types
   - Spatial Effects
2. Spatial Data Analysis
   - Spatial Process
   - Point Pattern Analysis
   - Lattice
   - Geostatistical
   - Network Data
3. Transformations of Spatial Data
   - Area to Point
   - Point to Area
   - Change of Support Problem

Spatial Effects

The Trilogy

- Spatial Dependence
- Spatial Heterogeneity
- Spatial Scale

First Law of Geography

Waldo Tobler

- "everything depends on everything else, but closer things more so"
- Structure of spatial dependence
- Distance Decay
- Closeness = Similarity

Spatial Heterogeneity

Spatial Instability

- Process varies in some way over spatial units
- Multiple forms
  - Discrete = regimes
  - Continuous = expansion method, GWR
- Trade-off
  - spatial homogeneity = stationary process
  - uniqueness = extreme heterogeneity

Spatial Scale

Mismatch

- Spatial scale of the process
- Spatial scale of our measurement

Issues

- points too far apart = miss small distance variation
- area aggregates cannot provide information on individual behavior
- ecological fallacy

Modifiable Areal Unit Problem (MAUP)

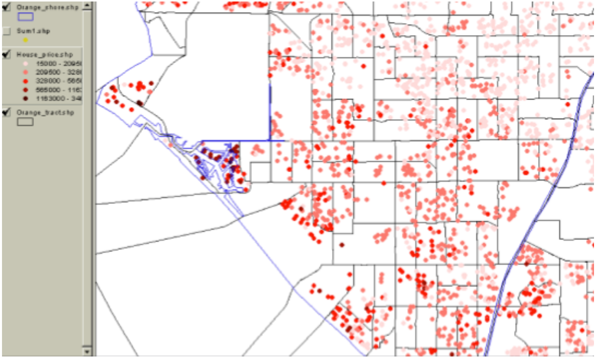
Aggregation Problem

- special case of ecological fallacy
- spatial heterogeneity
- a million spatial autocorrelation coefficients

Zonation Problem

- size
- arrangement

Spatial Heterogeneity: Housing Prices



Spatial Effects

Dependence, Heterogeneity and Scale

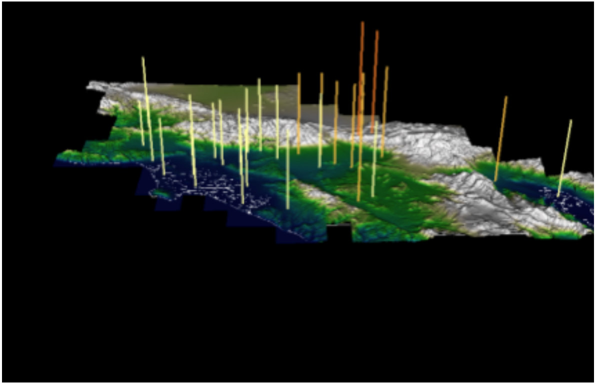
- not necessarily orthogonal
- distinguishing between dependence and heterogeneity can be challenging

Spatial Sampling

Location as an Experimental Design Problem

- Spatial sampling = where to collect the data
  - which villages to survey
  - where to locate air quality monitoring stations

Spatial Sampling



Outline

Spatial Process

Spatial Random Field

- a mathematical construct to capture randomness of values distributed over space
- $\{Z(s) : s \in D\}$
  - $s \in R^d$: location (e.g., lat-lon)
  - $D \in R^d$: index set = possible locations
  - $Z(s)$: random variable at location s

Types of Spatial Analysis

Point Pattern Analysis

spatial distribution of events

Geostatistical Analysis

surface modeling

Lattice Data Analysis

spatial patterns of attributes observed for spatial objects

Outline

Point Pattern Analysis

Data

- mapped pattern = all the events
- not a sample in the usual sense

Spatial Process

- observations as a realization of a random point process
- points occur in space according to a mathematical model

Point Data

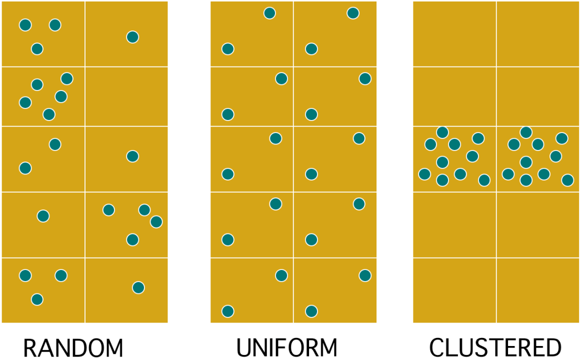
Spatial Domain: D

- Domain is *random*
- *Number* of points is *random*
- *Location* of points is *random*

Focus: Properties of D

- Number of points observed
- Pattern of the point locations

Point Patterns



Point Patterns

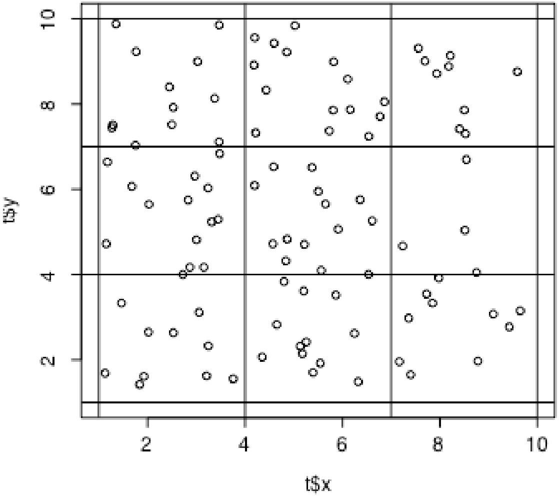
Unmarked Point Pattern

- Only location is recorded
- No other attribute information

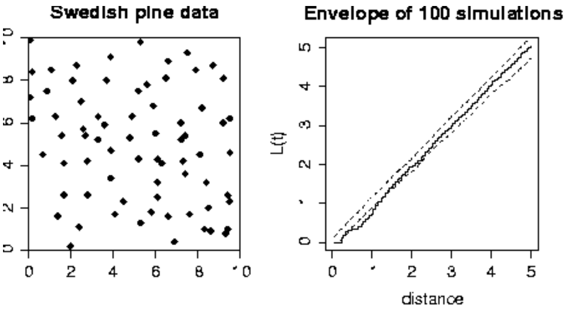
Marked Point Patterns

- Location is recorded
- Stochastic attributes also recorded
- e.g., sales at location, dbh of tree

Point Pattern Analysis: Quadrat Methods



Point Pattern Analysis: Distance Based Methods



Points on Networks

Figure 2: Retail stores assigned to the street network in Shibuya, Tokyo (cells are indicated by different colors)

Burkitt's Lymphoma Data: Time Series of Point Data

Outline

Lattice Data

Spatial Domain: D

- Discrete and fixed
- Locations nonrandom
- Locations countable

Examples of lattice data

- Attributes collected by ZIP code
- census tract
- remotely sensed data reported by pixels

Spatial Domain: D

- Discrete and fixed
- Locations nonrandom
- Locations countable

Examples of lattice data

- Attributes collected by ZIP code
- census tract
- remotely sensed data reported by pixels

Lattice Data

Spatial Domain: D

- Discrete and fixed
- Locations nonrandom
- Locations countable

Examples of lattice data

- Attributes collected by ZIP code
- census tract
- remotely sensed data reported by pixels

Lattice Data

Spatial Domain: D

- Discrete and fixed
- Locations nonrandom
- Locations countable

Examples of lattice data

- Attributes collected by ZIP code
- census tract
- remotely sensed data reported by pixels

Lattice Data

Spatial Domain: D

- Discrete and fixed
- Locations nonrandom
- Locations countable

Examples of lattice data

- Attributes collected by ZIP code
- census tract
- remotely sensed data reported by pixels

Lattice Data

Spatial Domain: D

- Discrete and fixed
- Locations nonrandom
- Locations countable

Examples of lattice data

- Attributes collected by ZIP code
- census tract
- remotely sensed data reported by pixels

Lattice Data

Spatial Domain: D

- Discrete and fixed
- Locations nonrandom
- Locations countable

Examples of lattice data

- Attributes collected by ZIP code
- census tract
- remotely sensed data reported by pixels

Lattice Data

Spatial Domain: D

- Discrete and fixed
- Locations nonrandom
- Locations countable

Examples of lattice data

- Attributes collected by ZIP code
- census tract
- remotely sensed data reported by pixels

Lattice Data: Indexing

Site

- Each location is now an area or *site*
- One observation on Z for each site
- Need a spatial index: $Z(s_i)$

$Z(s_i)$

- s_i is a representative location within the site
- e.g., centroid, largest city
- Allows for measuring distances between sites

Lattice Data: Indexing

Site

- Each location is now an area or *site*
- One observation on Z for each site
- Need a spatial index: $Z(s_i)$

$Z(s_i)$

- s_i is a representative location within the site
- e.g., centroid, largest city
- Allows for measuring distances between sites

Lattice Data: Indexing

Site

- Each location is now an area or *site*
- One observation on Z for each site
- Need a spatial index: $Z(s_i)$

$Z(s_i)$

- s_i is a representative location within the site
- e.g., centroid, largest city
- Allows for measuring distances between sites

Lattice Data: Indexing

Site

- Each location is now an area or *site*
- One observation on Z for each site
- Need a spatial index: $Z(s_i)$

$Z(s_i)$

- s_i is a representative location within the site
- e.g., centroid, largest city
- Allows for measuring distances between sites

Lattice Data: Indexing

Site

- Each location is now an area or *site*
- One observation on Z for each site
- Need a spatial index: $Z(s_i)$

$Z(s_i)$

- s_i is a representative location within the site
- e.g., centroid, largest city
- Allows for measuring distances between sites

Lattice Data: Indexing

Site

- Each location is now an area or *site*
- One observation on Z for each site
- Need a spatial index: $Z(s_i)$

$Z(s_i)$

- s_i is a representative location within the site
- e.g., centroid, largest city
- Allows for measuring distances between sites

Lattice Data: Indexing

Site

- Each location is now an area or *site*
- One observation on Z for each site
- Need a spatial index: $Z(s_i)$

$Z(s_i)$

- s_i is a representative location within the site
- e.g., centroid, largest city
- Allows for measuring distances between sites

Lattice Data: Indexing

Site

- Each location is now an area or *site*
- One observation on Z for each site
- Need a spatial index: $Z(s_i)$

$Z(s_i)$

- s_i is a representative location within the site
- e.g., centroid, largest city
- Allows for measuring distances between sites

Lattice Data: Aggregation and Coverage

Sites are areal units

- Attribute is typically aggregated or averaged
- Aggregated: event counts (number of crimes per tract)
- Averaged: per capita income by state

Coverage

- Lattice data is usually exhaustive in coverage
- e.g., U.S. states, census tracts in San Diego
- Prediction or interpolation not meaningful
- Explaining attribute variation across sites is the focus

Lattice Data: Aggregation and Coverage

Sites are areal units

- Attribute is typically aggregated or averaged
- Aggregated: event counts (number of crimes per tract)
- Averaged: per capita income by state

Coverage

- Lattice data is usually exhaustive in coverage
- e.g., U.S. states, census tracts in San Diego
- Prediction or interpolation not meaningful
- Explaining attribute variation across sites is the focus

Lattice Data: Aggregation and Coverage

Sites are areal units

- Attribute is typically aggregated or averaged
- Aggregated: event counts (number of crimes per tract)
- Averaged: per capita income by state

Coverage

- Lattice data is usually exhaustive in coverage
- e.g., U.S. states, census tracts in San Diego
- Prediction or interpolation not meaningful
- Explaining attribute variation across sites is the focus

Lattice Data: Aggregation and Coverage

Sites are areal units

- Attribute is typically aggregated or averaged
- Aggregated: event counts (number of crimes per tract)
- Averaged: per capita income by state

Coverage

- Lattice data is usually exhaustive in coverage
- e.g., U.S. states, census tracts in San Diego
- Prediction or interpolation not meaningful
- Explaining attribute variation across sites is the focus

Lattice Data: Aggregation and Coverage

Sites are areal units

- Attribute is typically aggregated or averaged
- Aggregated: event counts (number of crimes per tract)
- Averaged: per capita income by state

Coverage

- Lattice data is usually exhaustive in coverage
- e.g., U.S. states, census tracts in San Diego
- Prediction or interpolation not meaningful
- Explaining attribute variation across sites is the focus

Lattice Data: Aggregation and Coverage

Sites are areal units

- Attribute is typically aggregated or averaged
- Aggregated: event counts (number of crimes per tract)
- Averaged: per capita income by state

Coverage

- Lattice data is usually exhaustive in coverage
- e.g., U.S. states, census tracts in San Diego
- Prediction or interpolation not meaningful
- Explaining attribute variation across sites is the focus

Lattice Data: Aggregation and Coverage

Sites are areal units

- Attribute is typically aggregated or averaged
- Aggregated: event counts (number of crimes per tract)
- Averaged: per capita income by state

Coverage

- Lattice data is usually exhaustive in coverage
- e.g., U.S. states, census tracts in San Diego
- Prediction or interpolation not meaningful
- Explaining attribute variation across sites is the focus

Lattice Data: Aggregation and Coverage

Sites are areal units

- Attribute is typically aggregated or averaged
- Aggregated: event counts (number of crimes per tract)
- Averaged: per capita income by state

Coverage

- Lattice data is usually exhaustive in coverage
- e.g., U.S. states, census tracts in San Diego
- Prediction or interpolation not meaningful
- Explaining attribute variation across sites is the focus

Lattice Data: Aggregation and Coverage

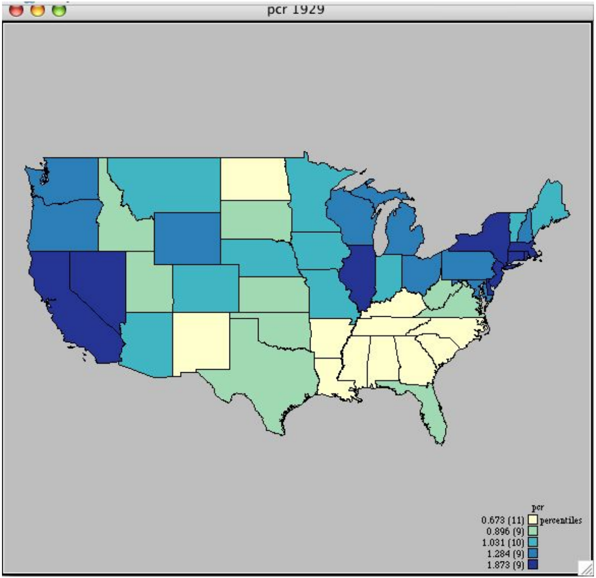
Sites are areal units

- Attribute is typically aggregated or averaged
- Aggregated: event counts (number of crimes per tract)
- Averaged: per capita income by state

Coverage

- Lattice data is usually exhaustive in coverage
- e.g., U.S. states, census tracts in San Diego
- Prediction or interpolation not meaningful
- Explaining attribute variation across sites is the focus

Lattice Data: State Per Capita Incomes



Lattice Data: Spatial Autocorrelation

Outline

Geostatistical Data

Spatial Domain: D

- A continuous and fixed set.
- Meaning $Z(s)$ can be observed everywhere within D.
- Between any two sample locations s_i and s_j you can theoretically place an infinite number of other samples.
- By fixed: the points in D are non-stochastic

Continuous Variation

- Because of the continuity of D
- Geostatistical data is referred to as "spatial data with continuous variation."
- Continuity is associated with D.
- Attribute Z may, or may not, be continuous.

Geostatistical Data

Spatial Domain: D

- A continuous and fixed set.
- Meaning $Z(s)$ can be observed everywhere within D.
- Between any two sample locations s_i and s_j you can theoretically place an infinite number of other samples.
- By fixed: the points in D are non-stochastic

Continuous Variation

- Because of the continuity of D
- Geostatistical data is referred to as "spatial data with continuous variation."
- Continuity is associated with D.
- Attribute Z may, or may not, be continuous.

Geostatistical Data

Spatial Domain: D

- A continuous and fixed set.
- Meaning $Z(s)$ can be observed everywhere within D.
- Between any two sample locations s_i and s_j you can theoretically place an infinite number of other samples.
- By fixed: the points in D are non-stochastic

Continuous Variation

- Because of the continuity of D
- Geostatistical data is referred to as "spatial data with continuous variation."
- Continuity is associated with D.
- Attribute Z may, or may not, be continuous.

Geostatistical Data

Spatial Domain: D

- A continuous and fixed set.
- Meaning $Z(s)$ can be observed everywhere within D.
- Between any two sample locations s_i and s_j you can theoretically place an infinite number of other samples.
- By fixed: the points in D are non-stochastic

Continuous Variation

- Because of the continuity of D
- Geostatistical data is referred to as "spatial data with continuous variation."
- Continuity is associated with D.
- Attribute Z may, or may not, be continuous.

Geostatistical Data

Spatial Domain: D

- A continuous and fixed set.
- Meaning $Z(s)$ can be observed everywhere within D.
- Between any two sample locations s_i and s_j you can theoretically place an infinite number of other samples.
- By fixed: the points in D are non-stochastic

Continuous Variation

- Because of the continuity of D
- Geostatistical data is referred to as "spatial data with continuous variation."
- Continuity is associated with D.
- Attribute Z may, or may not, be continuous.

Geostatistical Data

Spatial Domain: D

- A continuous and fixed set.
- Meaning $Z(s)$ can be observed everywhere within D.
- Between any two sample locations s_i and s_j you can theoretically place an infinite number of other samples.
- By fixed: the points in D are non-stochastic

Continuous Variation

- Because of the continuity of D
- Geostatistical data is referred to as "spatial data with continuous variation."
- Continuity is associated with D.
- Attribute Z may, or may not, be continuous.

Geostatistical Data

Spatial Domain: D

- A continuous and fixed set.
- Meaning $Z(s)$ can be observed everywhere within D.
- Between any two sample locations s_i and s_j you can theoretically place an infinite number of other samples.
- By fixed: the points in D are non-stochastic

Continuous Variation

- Because of the continuity of D
- Geostatistical data is referred to as "spatial data with continuous variation."
- Continuity is associated with D.
- Attribute Z may, or may not, be continuous.

Geostatistical Data

Spatial Domain: D

- A continuous and fixed set.
- Meaning $Z(s)$ can be observed everywhere within D.
- Between any two sample locations s_i and s_j you can theoretically place an infinite number of other samples.
- By fixed: the points in D are non-stochastic

Continuous Variation

- Because of the continuity of D
- Geostatistical data is referred to as "spatial data with continuous variation."
- Continuity is associated with D.
- Attribute Z may, or may not, be continuous.

Geostatistical Data

Spatial Domain: D

- A continuous and fixed set.
- Meaning $Z(s)$ can be observed everywhere within D.
- Between any two sample locations s_i and s_j you can theoretically place an infinite number of other samples.
- By fixed: the points in D are non-stochastic

Continuous Variation

- Because of the continuity of D
- Geostatistical data is referred to as "spatial data with continuous variation."
- Continuity is associated with D.
- Attribute Z may, or may not, be continuous.

Geostatistical Data

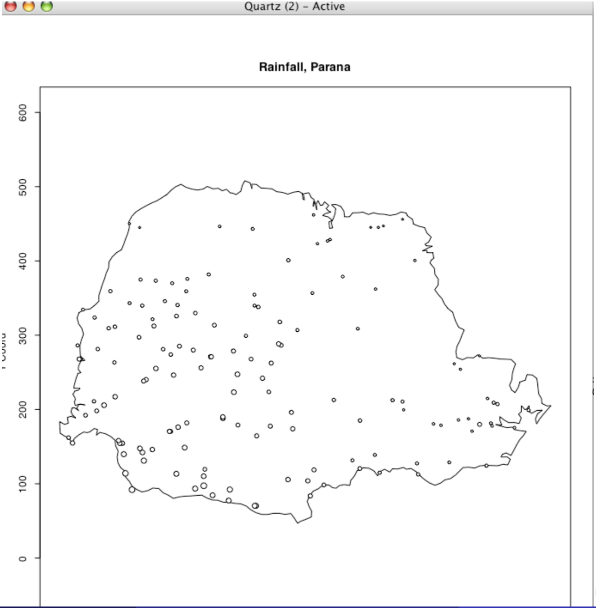
Spatial Domain: D

- A continuous and fixed set.
- Meaning $Z(s)$ can be observed everywhere within D.
- Between any two sample locations s_i and s_j you can theoretically place an infinite number of other samples.
- By fixed: the points in D are non-stochastic

Continuous Variation

- Because of the continuity of D
- Geostatistical data is referred to as “spatial data with continuous variation.”
- Continuity is associated with D.
- Attribute Z may, or may not, be continuous.

Geostatistical Data: Rainfall in Parana State Brazil



Geostatistical Data

Continuous variation

- Potentially measurable anywhere in D
- Impossible to sample D exhaustively

Reconstruction of the surface from observed sites

- Tessellation based methods
- Interpolation
- Kriging

Geostatistical Data

Continuous variation

- Potentially measurable anywhere in D
- Impossible to sample D exhaustively

Reconstruction of the surface from observed sites

- Tessellation based methods
- Interpolation
- Kriging

Geostatistical Data

Continuous variation

- Potentially measurable anywhere in D
- Impossible to sample D exhaustively

Reconstruction of the surface from observed sites

- Tessellation based methods
- Interpolation
- Kriging

Geostatistical Data

Continuous variation

- Potentially measurable anywhere in D
- Impossible to sample D exhaustively

Reconstruction of the surface from observed sites

- Tessellation based methods
- Interpolation
- Kriging

Geostatistical Data

Continuous variation

- Potentially measurable anywhere in D
- Impossible to sample D exhaustively

Reconstruction of the surface from observed sites

- Tessellation based methods
- Interpolation
- Kriging

Geostatistical Data

Continuous variation

- Potentially measurable anywhere in D
- Impossible to sample D exhaustively

Reconstruction of the surface from observed sites

- Tessellation based methods
- Interpolation
- Kriging

Geostatistical Data

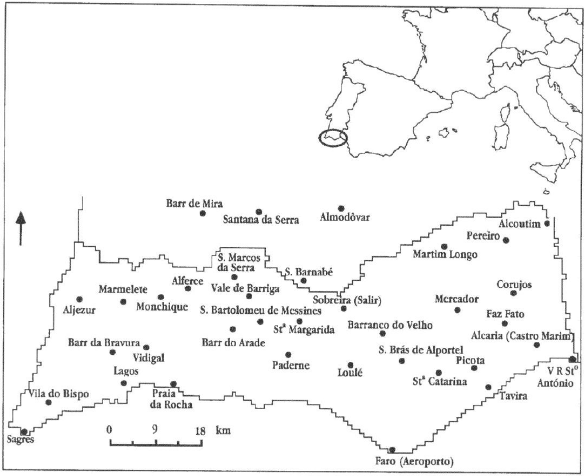
Continuous variation

- Potentially measurable anywhere in D
- Impossible to sample D exhaustively

Reconstruction of the surface from observed sites

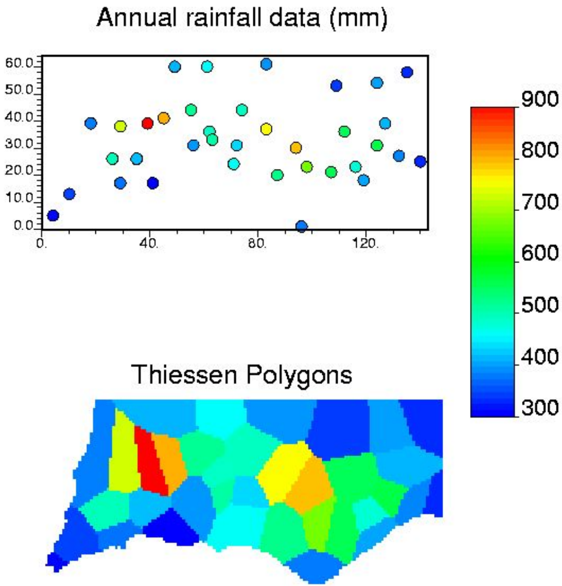
- Tessellation based methods
- Interpolation
- Kriging

Surface Reconstruction: Example[1]

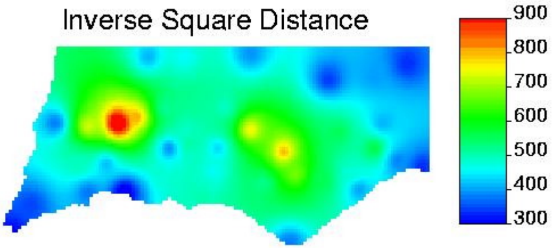


[1] From Goovaerts, P. (1999)“Performance comparison of geostatistical algorithms for incorporating elevation into the mapping of precipitation”. *Geocomputation ’99*.

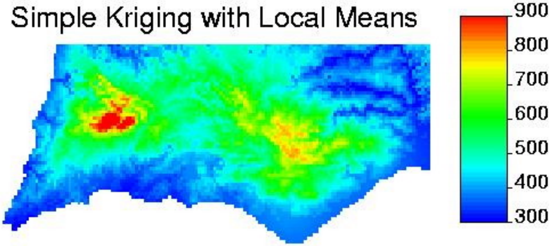
Surface Reconstruction: Tessellation Based Method



Surface Reconstruction: Spatial Interpolation



Surface Reconstruction: Kriging



Outline

1. Spatial Data
   - Types
   - Spatial Effects
2. Spatial Data Analysis
   - Spatial Process
   - Point Pattern Analysis
   - Lattice
   - Geostatistical
   - Network Data
3. Transformations of Spatial Data
   - Area to Point
   - Point to Area
   - Change of Support Problem

Network Data

Networks

- A network is a system of linear features connected at intersections and interchanges.
- These intersections and interchanges are called nodes.
- The linear feature connecting any given pair of nodes is called an arc.
- Formally, a network is defined as a directed graph $G = (N, A)$ consisting of an indexed set of nodes N with $n = |N|$ and a spanning set of directed arcs A with $m = |A|$, where n is the number of nodes and m is the number of arcs.
- Each arc on a network is represented as an ordered pair of nodes, in the form from node i to node j, denoted by (i, j).
- In the GIS literature, a network arc is often called a network link.

Network Data

Networks

- A network is a system of linear features connected at intersections and interchanges.
- These intersections and interchanges are called nodes.
- The linear feature connecting any given pair of nodes is called an arc.
- Formally, a network is defined as a directed graph $G = (N, A)$ consisting of an indexed set of nodes N with $n = |N|$ and a spanning set of directed arcs A with $m = |A|$, where n is the number of nodes and m is the number of arcs.
- Each arc on a network is represented as an ordered pair of nodes, in the form from node i to node j, denoted by (i, j).
- In the GIS literature, a network arc is often called a network link.

Network Data

Networks

- A network is a system of linear features connected at intersections and interchanges.
- These intersections and interchanges are called nodes.
- The linear feature connecting any given pair of nodes is called an arc.
- Formally, a network is defined as a directed graph $G = (N, A)$ consisting of an indexed set of nodes N with $n = |N|$ and a spanning set of directed arcs A with $m = |A|$, where n is the number of nodes and m is the number of arcs.
- Each arc on a network is represented as an ordered pair of nodes, in the form from node i to node j, denoted by (i, j).
- In the GIS literature, a network arc is often called a network link.

Network Data

Networks

- A network is a system of linear features connected at intersections and interchanges.
- These intersections and interchanges are called nodes.
- The linear feature connecting any given pair of nodes is called an arc.
- Formally, a network is defined as a directed graph $G = (N, A)$ consisting of an indexed set of nodes N with $n = |N|$ and a spanning set of directed arcs A with $m = |A|$, where n is the number of nodes and m is the number of arcs.
- Each arc on a network is represented as an ordered pair of nodes, in the form from node i to node j, denoted by (i, j).
- In the GIS literature, a network arc is often called a network link.

Network Data

Networks

- A network is a system of linear features connected at intersections and interchanges.
- These intersections and interchanges are called nodes.
- The linear feature connecting any given pair of nodes is called an arc.
- Formally, a network is defined as a directed graph $G = (N, A)$ consisting of an indexed set of nodes N with $n = |N|$ and a spanning set of directed arcs A with $m = |A|$, where n is the number of nodes and m is the number of arcs.
- Each arc on a network is represented as an ordered pair of nodes, in the form from node i to node j, denoted by (i, j).
- In the GIS literature, a network arc is often called a network link.

Network Data

Networks

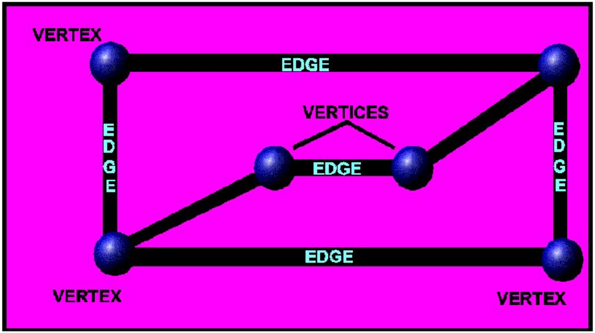
- A network is a system of linear features connected at intersections and interchanges.
- These intersections and interchanges are called nodes.
- The linear feature connecting any given pair of nodes is called an arc.
- Formally, a network is defined as a directed graph $G = (N, A)$ consisting of an indexed set of nodes N with $n = |N|$ and a spanning set of directed arcs A with $m = |A|$, where n is the number of nodes and m is the number of arcs.
- Each arc on a network is represented as an ordered pair of nodes, in the form from node i to node j, denoted by (i, j).
- In the GIS literature, a network arc is often called a network link.

Network Data

Networks

- A network is a system of linear features connected at intersections and interchanges.
- These intersections and interchanges are called nodes.
- The linear feature connecting any given pair of nodes is called an arc.
- Formally, a network is defined as a directed graph $G = (N, A)$ consisting of an indexed set of nodes N with $n = |N|$ and a spanning set of directed arcs A with $m = |A|$, where n is the number of nodes and m is the number of arcs.
- Each arc on a network is represented as an ordered pair of nodes, in the form from node i to node j, denoted by (i, j).
- In the GIS literature, a network arc is often called a network link.

Network Data: Graph Theory



Network Data

In this course

- We will not be analyzing network data per se
- We will be drawing on graph theory to help in ESDA

Uses of graph theory in ESDA

- Properties of adjacency matrices
- Clustering and regionalization algorithms

Network Data

In this course

- We will not be analyzing network data per se
- We will be drawing on graph theory to help in ESDA

Uses of graph theory in ESDA

- Properties of adjacency matrices
- Clustering and regionalization algorithms

Network Data

In this course

- We will not be analyzing network data per se
- We will be drawing on graph theory to help in ESDA

Uses of graph theory in ESDA

- Properties of adjacency matrices
- Clustering and regionalization algorithms

Network Data

In this course

- We will not be analyzing network data per se
- We will be drawing on graph theory to help in ESDA

Uses of graph theory in ESDA

- Properties of adjacency matrices
- Clustering and regionalization algorithms

Network Data

In this course

- We will not be analyzing network data per se
- We will be drawing on graph theory to help in ESDA

Uses of graph theory in ESDA

- Properties of adjacency matrices
- Clustering and regionalization algorithms

Network Data

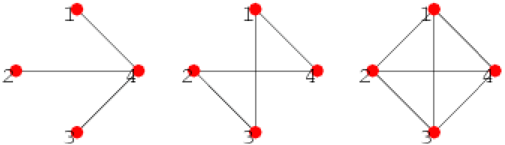
In this course

- We will not be analyzing network data per se
- We will be drawing on graph theory to help in ESDA

Uses of graph theory in ESDA

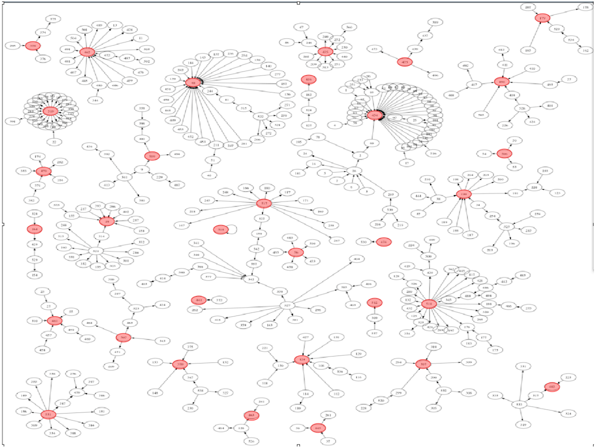
- Properties of adjacency matrices
- Clustering and regionalization algorithms

Network Data: Adjacency Matrices



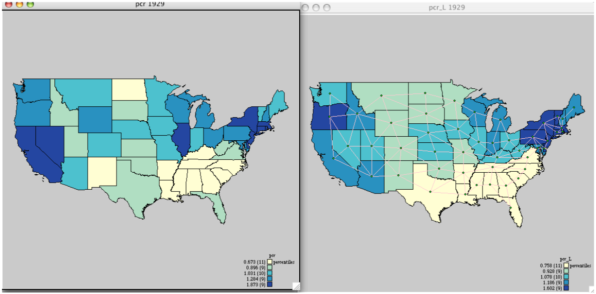
$$\begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix} \quad \begin{pmatrix} 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \end{pmatrix} \quad \begin{pmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix}$$

Network Data: Clustering Visualization



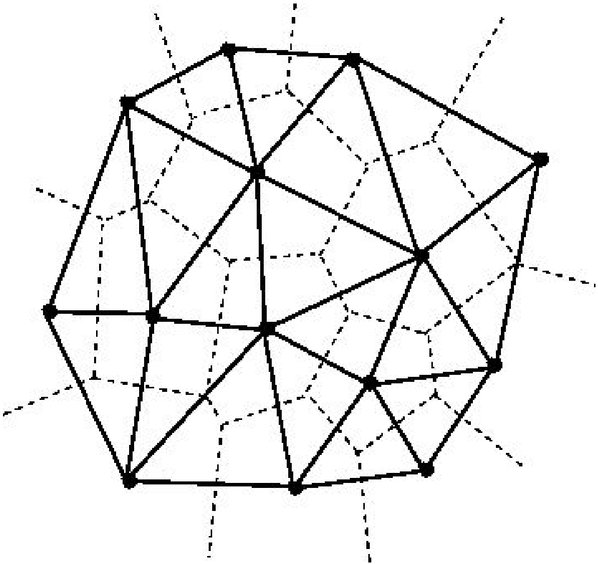
Outline

Area to Point Transformation: Centroids



Outline

Point to Area Transformation: Thiessen Polygons



Outline

Change of Support Problem

Transformation from one spatial framework to another

- Point data to area data
- Regional data: areal interpolation

Scaling

- Upscaling: points to areas or, areas to larger areas
- Downscaling: larger areas to smaller composite areas

Change of Support Problem

Transformation from one spatial framework to another

- Point data to area data
- Regional data: areal interpolation

Scaling

- Upscaling: points to areas or, areas to larger areas
- Downscaling: larger areas to smaller composite areas

Change of Support Problem

Transformation from one spatial framework to another

- Point data to area data
- Regional data: areal interpolation

Scaling

- Upscaling: points to areas or, areas to larger areas
- Downscaling: larger areas to smaller composite areas

Change of Support Problem

Transformation from one spatial framework to another

- Point data to area data
- Regional data: areal interpolation

Scaling

- Upscaling: points to areas or, areas to larger areas
- Downscaling: larger areas to smaller composite areas

Change of Support Problem

Transformation from one spatial framework to another

- Point data to area data
- Regional data: areal interpolation

Scaling

- Upscaling: points to areas or, areas to larger areas
- Downscaling: larger areas to smaller composite areas

Change of Support Problem

Transformation from one spatial framework to another

- Point data to area data
- Regional data: areal interpolation

Scaling

- Upscaling: points to areas or, areas to larger areas
- Downscaling: larger areas to smaller composite areas

Change of Support Problem: Point to Area

Example

- Estimate soil contamination for a census tract given point samples in the tract.
- Sample points $(s(1), \ldots, s(n))$ are the supports
- Prediction is for the area (A), $Y(A)$

Estimation

- Assume a constant mean (contamination), then $E[Y(i)] = \mu \ \forall i$
- Simple estimator: $\hat{Y} = (1/n) \sum_{i=1}^{n} y(i)$
- Problems
  - Ignores spatial correlation in $Y(i)$ over the sites
  - Ignores the difference in variances: $V(Y(A)) < V(Y)$
  - May significantly over or under estimate $Y(A)$
  - Geostatistical methods are better

Change of Support Problem: Point to Area

Example

- Estimate soil contamination for a census tract given point samples in the tract.
- Sample points $(s(1), \ldots, s(n))$ are the supports
- Prediction is for the area (A), $Y(A)$

Estimation

- Assume a constant mean (contamination), then $E[Y(i)] = \mu \; \forall i$
- Simple estimator: $\hat{Y} = (1/n) \sum_{i=1}^{n} y(i)$
- Problems
  - Ignores spatial correlation in $Y(i)$ over the sites
  - Ignores the difference in variances: $V(Y(A)) < V(\hat{Y})$
  - May significantly over or under estimate $Y(A)$
  - Geostatistical methods are better

Change of Support Problem: Point to Area

Example

- Estimate soil contamination for a census tract given point samples in the tract.
- Sample points $(s(1), \ldots, s(n))$ are the supports
- Prediction is for the area (A), $Y(A)$

Estimation

- Assume a constant mean (contamination), then $E[Y(i)] = \mu \; \forall i$
- Simple estimator: $\hat{Y} = (1/n)\sum_{i=1}^{n} y(i)$
- Problems
  - Ignores spatial correlation in $Y(i)$ over the sites
  - Ignores the difference in variances: $V(Y(A)) < V(\hat{Y})$
  - May significantly over or under estimate $Y(A)$
  - Geostatistical methods are better

Change of Support Problem: Point to Area

Example

- Estimate soil contamination for a census tract given point samples in the tract.
- Sample points $(s(1), \ldots, s(n))$ are the supports
- Prediction is for the area (A), $Y(A)$

Estimation

- Assume a constant mean (contamination), then $E[Y(i)] = \mu \; \forall i$
- Simple estimator: $\hat{Y} = (1/n) \sum_{i=1}^{n} y(i)$
- Problems
  - Ignores spatial correlation in $Y(i)$ over the sites
  - Ignores the difference in variances: $V(Y(A)) < V(\hat{Y})$
  - May significantly over or under estimate $Y(A)$
  - Geostatistical methods are better

Change of Support Problem: Point to Area

Example

- Estimate soil contamination for a census tract given point samples in the tract.
- Sample points $(s(1), \ldots, s(n))$ are the supports
- Prediction is for the area (A), $Y(A)$

Estimation

- Assume a constant mean (contamination), then $E[Y(i)] = \mu \; \forall i$
- Simple estimator: $\hat{Y} = (1/n)\sum_{i=1}^{n} y(i)$
- Problems
  - Ignores spatial correlation in $Y(i)$ over the sites
  - Ignores the difference in variances: $V(Y(A)) < V(\hat{Y})$
  - May significantly over or under estimate $Y(A)$
  - Geostatistical methods are better

Change of Support Problem: Point to Area

Example

- Estimate soil contamination for a census tract given point samples in the tract.
- Sample points $(s(1), \ldots, s(n))$ are the supports
- Prediction is for the area (A), $Y(A)$

Estimation

- Assume a constant mean (contamination), then $E[Y(i)] = \mu \ \forall i$
- Simple estimator: $\hat{Y} = (1/n) \sum_{i=1}^{n} y(i)$
- Problems
  - Ignores spatial correlation in $Y(i)$ over the sites
  - Ignores the difference in variances: $V(Y(A)) < V(\hat{Y})$
  - May significantly over or under estimate $Y(A)$
  - Geostatistical methods are better

Change of Support Problem: Point to Area

Example

- Estimate soil contamination for a census tract given point samples in the tract.
- Sample points $(s(1), \ldots, s(n))$ are the supports
- Prediction is for the area (A), $Y(A)$

Estimation

- Assume a constant mean (contamination), then $E[Y(i)] = \mu \; \forall i$
- Simple estimator: $\hat{Y} = (1/n) \sum_{i=1}^{n} y(i)$
- Problems
  - Ignores spatial correlation in $Y(i)$ over the sites
  - Ignores the difference in variances: $V(Y(A)) < V(\hat{Y})$
  - May significantly over or under estimate $Y(A)$
  - Geostatistical methods are better

Change of Support Problem: Point to Area

Example

- Estimate soil contamination for a census tract given point samples in the tract.
- Sample points $(s(1), \ldots, s(n))$ are the supports
- Prediction is for the area (A), $Y(A)$

Estimation

- Assume a constant mean (contamination), then $E[Y(i)] = \mu \; \forall i$
- Simple estimator: $\hat{Y} = (1/n) \sum_{i=1}^{n} y(i)$
- Problems
  - Ignores spatial correlation in $Y(i)$ over the sites
  - Ignores the difference in variances: $V(Y(A)) < V(\hat{Y})$
  - May significantly over or under estimate $Y(A)$
  - Geostatistical methods are better

Change of Support Problem: Point to Area

Example

- Estimate soil contamination for a census tract given point samples in the tract.
- Sample points $(s(1), \ldots, s(n))$ are the supports
- Prediction is for the area (A), $Y(A)$

Estimation

- Assume a constant mean (contamination), then $E[Y(i)] = \mu \ \forall i$
- Simple estimator: $\hat{Y} = (1/n)\sum_{i=1}^{n} y(i)$
- Problems
  - Ignores spatial correlation in $Y(i)$ over the sites
  - Ignores the difference in variances: $V(Y(A)) < V(\hat{Y})$
  - May significantly over or under estimate $Y(A)$
  - Geostatistical methods are better

Change of Support Problem: Point to Area

Example

- Estimate soil contamination for a census tract given point samples in the tract.
- Sample points $(s(1), \ldots, s(n))$ are the supports
- Prediction is for the area (A), $Y(A)$

Estimation

- Assume a constant mean (contamination), then $E[Y(i)] = \mu \ \forall i$
- Simple estimator: $\hat{Y} = (1/n) \sum_{i=1}^{n} y(i)$
- Problems
  - Ignores spatial correlation in $Y(i)$ over the sites
  - Ignores the difference in variances: $V(Y(A)) < V(\hat{Y})$
  - May significantly over or under estimate $Y(A)$
  - Geostatistical methods are better

Change of Support Problem: Point to Area

Example

- Estimate soil contamination for a census tract given point samples in the tract.
- Sample points $(s(1), \ldots, s(n))$ are the supports
- Prediction is for the area (A), $Y(A)$

Estimation

- Assume a constant mean (contamination), then $E[Y(i)] = \mu \; \forall i$
- Simple estimator: $\hat{Y} = (1/n) \sum_{i=1}^{n} y(i)$
- Problems
  - Ignores spatial correlation in $Y(i)$ over the sites
  - Ignores the difference in variances: $V(Y(A)) < V(\hat{Y})$
  - May significantly over or under estimate $Y(A)$
  - Geostatistical methods are better

Change of Support Problem: Point to Area

Example

- Estimate soil contamination for a census tract given point samples in the tract.
- Sample points $(s(1), \ldots, s(n))$ are the supports
- Prediction is for the area (A), $Y(A)$

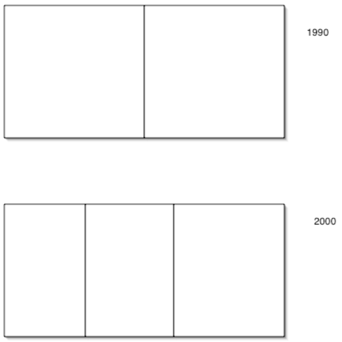
Estimation

- Assume a constant mean (contamination), then $E[Y(i)] = \mu \; \forall i$
- Simple estimator: $\hat{Y} = (1/n)\sum_{i=1}^{n} y(i)$
- Problems
  - Ignores spatial correlation in $Y(i)$ over the sites
  - Ignores the difference in variances: $V(Y(A)) < V(\hat{Y})$
  - May significantly over or under estimate $Y(A)$
  - Geostatistical methods are better

Change of Support Problem: Areal Interpolation

Example

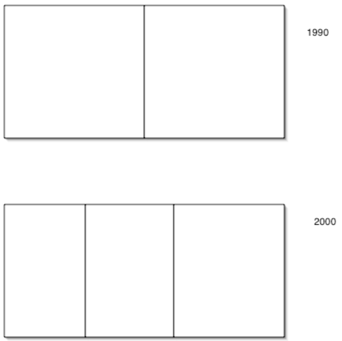
- Census tracts from 2000 and 1990
- Due to population growth, some 1990 tracts may have been split



Change of Support Problem: Areal Interpolation

Example

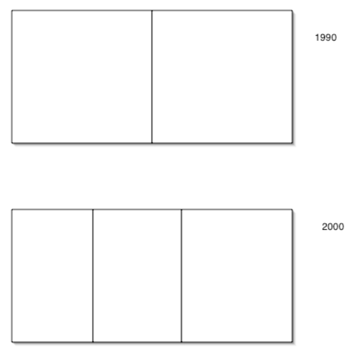
- Census tracts from 2000 and 1990
- Due to population growth, some 1990 tracts may have been split



Change of Support Problem: Areal Interpolation

Example

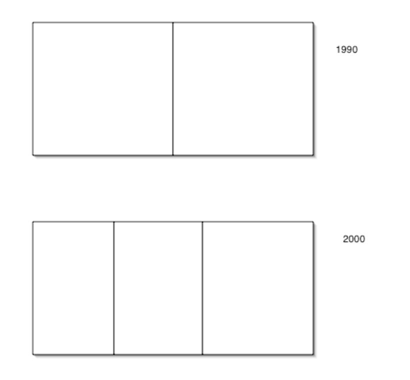
- Census tracts from 2000 and 1990
- Due to population growth, some 1990 tracts may have been split



Change of Support Problem: Areal Interpolation

Example

- Census tracts from 2000 and 1990
- Due to population growth, some 1990 tracts may have been split



Change of Support Problem: Areal Interpolation

Methods

- Cartographic based
- Statistical based

Cartographic

- Polygon overlap
- Adjacency information to smooth

Statistical

- Fit models to surface data on ancillary variables
- Use models to interpolate

Change of Support Problem: Areal Interpolation

Methods

- Cartographic based
- Statistical based

Cartographic

- Polygon overlap
- Adjacency information to smooth

Statistical

- Fit models to surface data on ancillary variables
- Use models to interpolate

Methods

- Cartographic based
- Statistical based

Cartographic

- Polygon overlap
- Adjacency information to smooth

Statistical

- Fit models to surface data on ancillary variables
- Use models to interpolate

Change of Support Problem: Areal Interpolation

Methods

- Cartographic based
- Statistical based

Cartographic

- Polygon overlap
- Adjacency information to smooth

Statistical

- Fit models to surface data on ancillary variables
- Use models to interpolate

Change of Support Problem: Areal Interpolation

Methods

- Cartographic based
- Statistical based

Cartographic

- Polygon overlap
- Adjacency information to smooth

Statistical

- Fit models to surface data on ancillary variables
- Use models to interpolate

Change of Support Problem: Areal Interpolation

Methods

- Cartographic based
- Statistical based

Cartographic

- Polygon overlap
- Adjacency information to smooth

Statistical

- Fit models to surface data on ancillary variables
- Use models to interpolate

Change of Support Problem: Areal Interpolation

Methods

- Cartographic based
- Statistical based

Cartographic

- Polygon overlap
- Adjacency information to smooth

Statistical

- Fit models to surface data on ancillary variables
- Use models to interpolate

Change of Support Problem: Areal Interpolation

Methods

- Cartographic based
- Statistical based

Cartographic

- Polygon overlap
- Adjacency information to smooth

Statistical

- Fit models to surface data on ancillary variables
- Use models to interpolate

Change of Support Problem: Areal Interpolation

Methods

- Cartographic based
- Statistical based

Cartographic

- Polygon overlap
- Adjacency information to smooth

Statistical

- Fit models to surface data on ancillary variables
- Use models to interpolate