

Spatial data analysis

Cristina Catita

cmcatita@fc.ul.pt

Faculty of Science of University of Lisbon

1

Course outline

1. What is Special About Space Data?

2. GIS and Spatial Data Analysis

3. The Process of Spatial Analysis

4. Types of data in Spatial Analysis

5. Events or Point Pattern

6. Areas

7. Continuous Surfaces

Not for today!

2

1. What is Special About Space Data?

Why do we do care about spatial data?

Because...

Space is an essential component of everyday life!



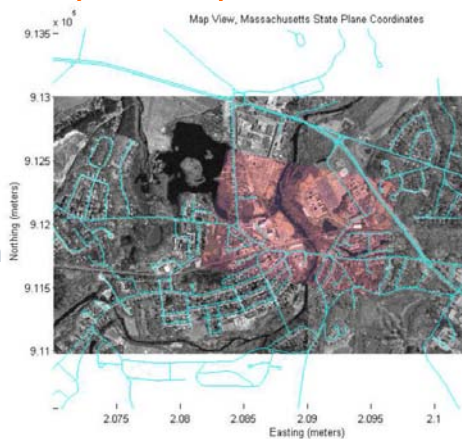
Bing, Google Earth

- We want to know where we are! (in absolute sense)
- We want to know distances between locations! (in relative sense)

1. What is Special About Space Data?

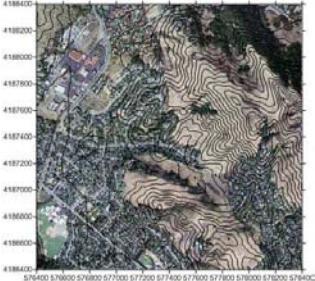
What do we need to know before any kind of analysis?

- 1. A spatial reference system!
- 1. A way to represent geographic entities on digital maps!



1. What is Special About Space Data?

- **Georeferenced data:** location on Earth is represented in a **Coordinate Referenced System**;
- **Many coordinate systems** exist, they evolve over time as accuracy of the Earth measurements improves;



A topographic map showing contour lines and a coordinate grid. The vertical axis (y-axis) ranges from 4188400 to 4188420 in increments of 200. The horizontal axis (x-axis) ranges from 516400 to 516400 in increments of 100. The map shows a complex terrain with various elevations and features.

5

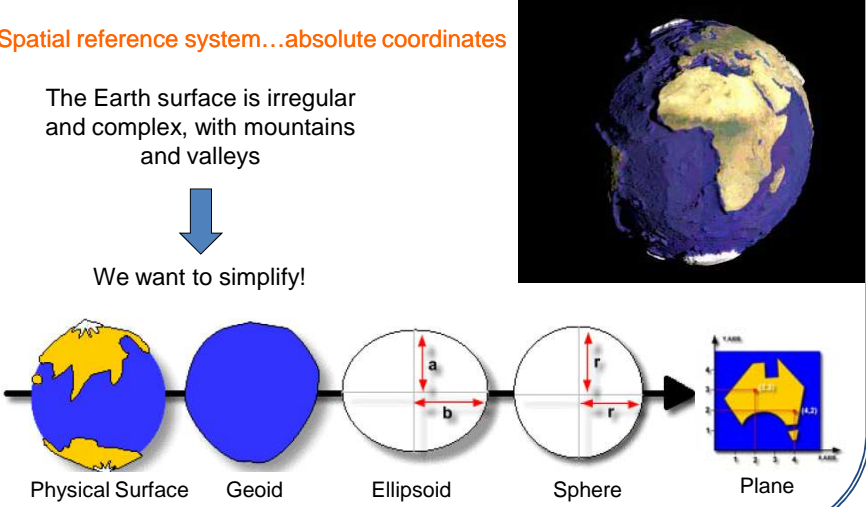
1. What is Special About Space Data?

Spatial reference system...absolute coordinates

The Earth surface is irregular and complex, with mountains and valleys

↓

We want to simplify!



The diagram illustrates the process of simplifying the Earth's surface into a flat plane. It starts with a 'Physical Surface' (a globe with yellow and blue continents and oceans), which is simplified into a 'Geoid' (a blue irregular shape). This is further simplified into an 'Ellipsoid' (a circle with semi-axes 'a' and 'b'), then a 'Sphere' (a circle with radius 'r'), and finally a 'Plane' (a flat map of Australia with a coordinate grid and a point labeled '(4,2)').

6

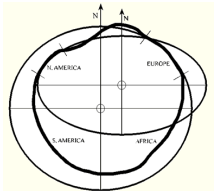
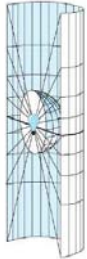
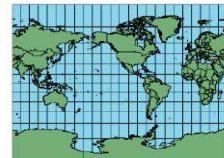
1. What is Special About Space Data?

We want Earth coordinates, but how to get it?

We need a coordinate system, which is based on:

1. a spheroid shape (parameters of the ellipsoid adopted)
2. a datum definition (local ou global datum)
3. and a map projection (Cylindrical, Conical, Azimuthal, etc...)

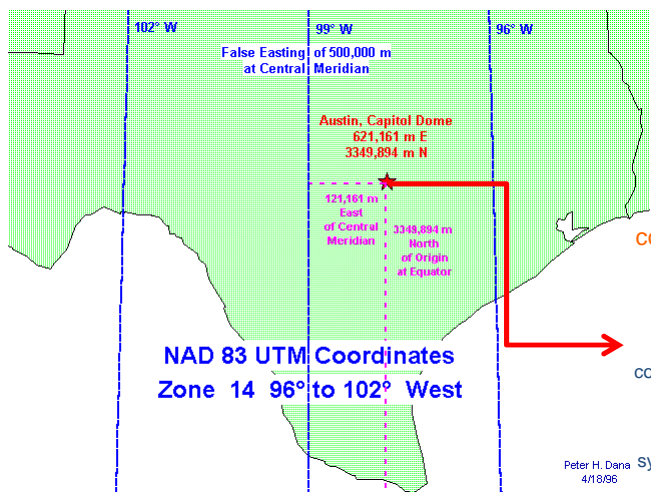
➤ Spheroid + Datum + Map projection


+

=


7

1. What is Special About Space Data?

What kind of coordinates?



What are the coordinates of this point here?

... Geographic coordinates or cartesian coordinates?

What coordinate system are you want to referred it?

Peter H. Dana
4/10/96

8

1. What is Special About Space Data?

2D geographic coordinates (ϕ, λ)

3D geographic coordinates (ϕ, λ, h)

Geocentric coordinates (X, Y, Z)

2D Cartesian coordinates (X, Y)

Images from ITC – Geometric Aspects of Mapping, R. Knipfers
(<http://www.kartografie.nl/geometrics/coordinate%20systems/coordsys.html>)

1. What is Special About Space Data?

Geographic coordinate system

- geoid -> ellipsoid -> (sphere) -> latitude/longitude
- GPS, large regions, data exchange (USGS, Google)
- units are ? degree-minutes-seconds
- requires complex algorithms for distances, areas

From Geospatial Analysis and Modeling MEAS2 – H. Mitasova

1. What is Special About Space Data?

Projected Reference Systems - cartesian coordinate systems based on projections:

- geoid – ellipsoid - developable surface – plane – x,y
- developable surfaces: ?
- type of distortion: ?

Geoid
Simplified representation:
Ellipsoid
Projection on developable
surface
Planar map with
coordinate system

From Geospatial Analysis and Modeling MEA592 – H. Mitsova

11

1. What is Special About Space Data?

How to choose a reference coordinate system?

National/State Coordinate systems defined by:
Reference spheroid/geoid and datum?
Vertical datum ?
Projection ?

GOAL: minimize distortions on maps that were used to measure distances and areas;

12

1. What is Special About Space Data?

How to deal with different coordinate systems?

Data often come in different coordinate systems:
 Different sources of data;
 Different agencies;
 Etc...

Coordinate transformations
TIME CONSUMING (e.g. for raster resampling/reinterpolation to regular grid is required)
LOSS OF PRECISION
GIS: ON-FLY

The diagram illustrates the relationship between different coordinate systems. At the top, two globes represent 'Datum A' and 'Datum B'. Datum A is labeled with coordinates (λ, ϕ) and Datum B with (λ', ϕ') . The transformation between them is labeled 'Transformação de Molodenski (Parâmetros de regressão)'. Below the globes, two coordinate systems are shown: 'Coordenadas Geográficas' (Geographical Coordinates) and 'Coordenadas Cartográficas' (Cartographic Coordinates). The transformation from Geographical to Cartographic is labeled 'Transformação Directa' and from Cartographic to Geographical is 'Equações Inversas'. The transformation from Datum A to Datum B is labeled 'Equações Directas' and from Datum B to Datum A is 'Equações Inversas'.

13

1. What is Special About Space Data?

Another way to reference geographic entities...
 relative to...

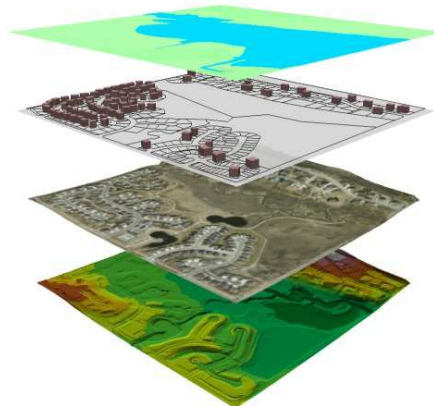
- System of determining the position of an event (e.g. pavement condition) relative to a route system (e.g. streets) by specifying a start position, direction and distance.

The diagram shows a route system consisting of three streets: 1st Ave, Main St, and 2nd Ave. A red line segment on Main St is labeled 'Poor Pavement Condition'. A green dot on Main St is labeled 'Start of Main St'. A pink arrow labeled 'direction' points along Main St towards 2nd Ave. A blue line segment on Main St is labeled '650 feet from 1st Ave'.

14

1. What is Special About Space Data?

Now you know how to get absolute coordinates on the Earth surface!



NEXT STEP:

We will need a model of the world to represent it computationally?

MAYBE WE NEED A:

GIS
(Geographic Information System)

Image Source: <http://www.freewebs.com/waywardgeographics/gis1.bmp>

15

2. GIS and Spatial Data Analysis

"a GIS is a system of hardware, software and procedures to facilitate the management, manipulation, analysis, modelling, representation and display of georeferenced data to solve complex problems regarding planning and management of resources."

(NCGIA - National Centre of Geographic Information and Analysis, 1990)




Image source: http://www.bisag.gujarat.gov.in/is/gis_overview.jpg

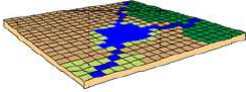
16

2. GIS and Spatial Data Analysis


- Mapped, georeferenced data are transformed into discrete GIS representations
- Two different types of objects/phenomena
 - **continuous fields**: $w=f(x,y)$, $w=f(x,y,z)$
 each point in space is assigned a distinct value, change between two neighboring points is relatively small: elevation, precipitation
 represented as **raster layers**, but meshes, TIN, isolines or points are also used.
 - **discrete objects/features**: lines, points or areas with attributes
 represented as geometry(shape) with attribute table or object based (geodatabase); raster representation is also used : roads, streams, census blocks, land use, schools



Real World



Raster Representation

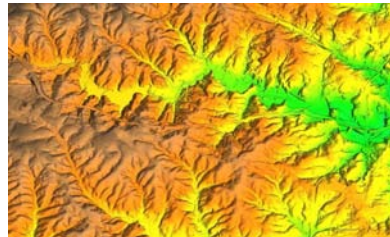
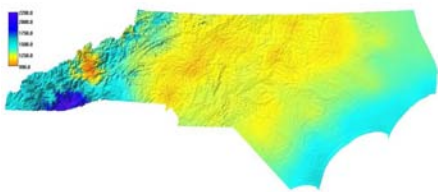


Vector Representation

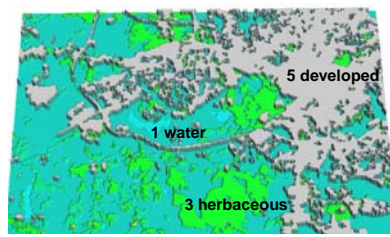

From Geospatial Analysis and Modeling MEA592 – H. Mitsova

2. GIS and Spatial Data Analysis

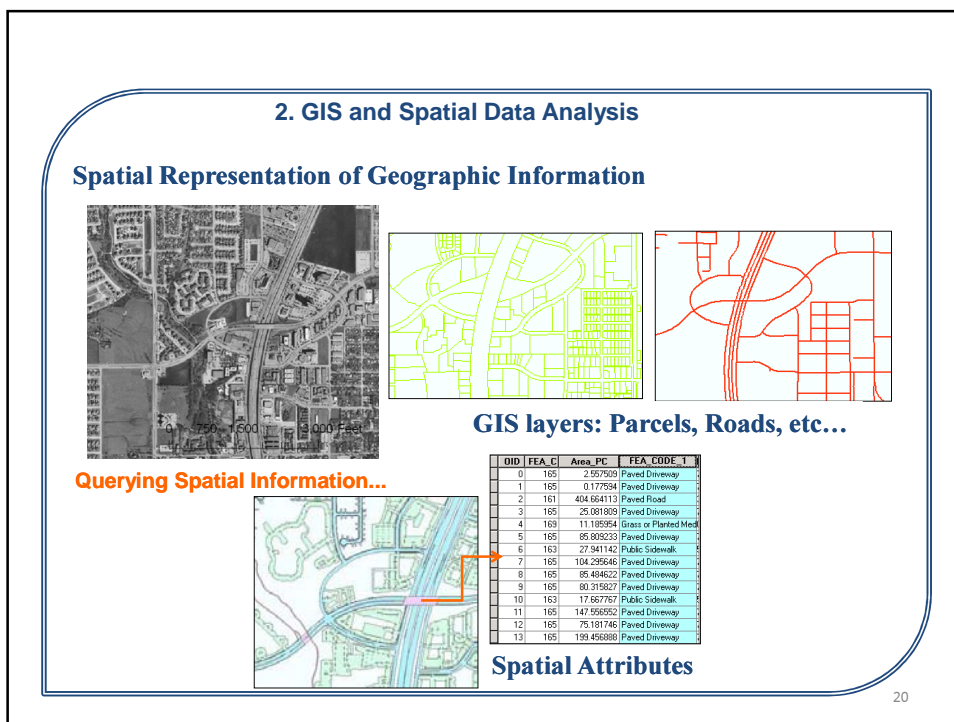
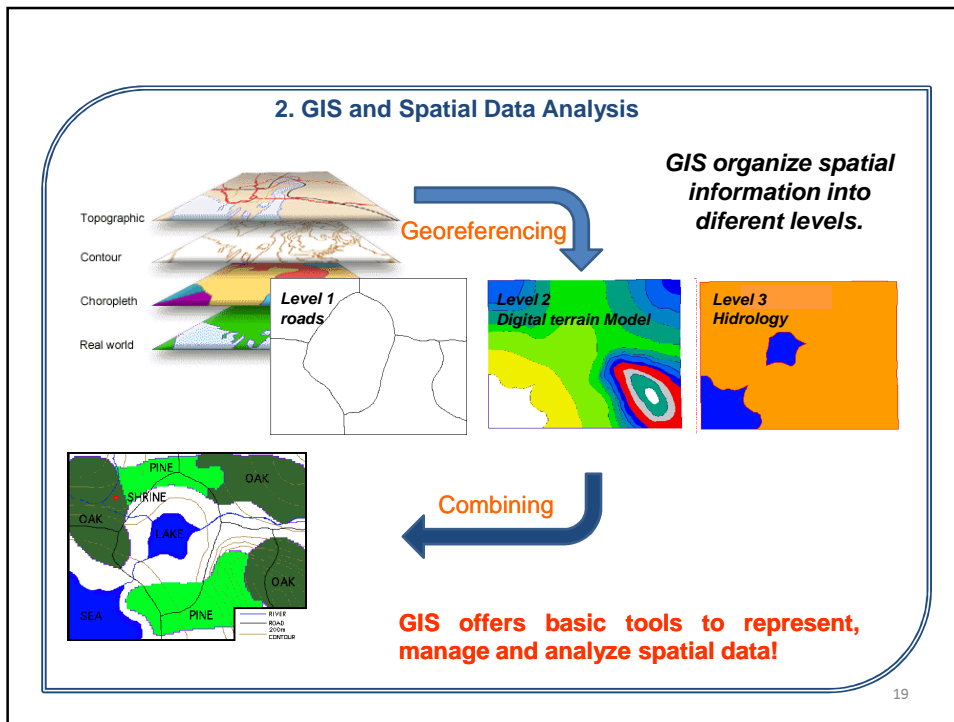
continuous: elevation, precipitation

discrete: land use, roads

From Geospatial Analysis and Modeling MEA592 – H. Mitsova



2. GIS and Spatial Data Analysis

Categories of Spatial Analysis

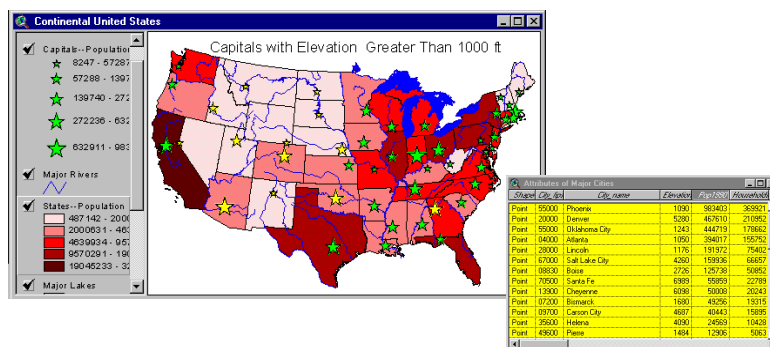
1. **Spatial data manipulation**
 - ❖ The basic functions of any GIS
2. **Spatial data analysis**
 - ❖ Descriptive and exploratory functions using maps
3. **Spatial statistical analysis**
 - ❖ Uses statistical methods to determine if spatial data are “typical” or “unexpected” relative to a statistical model.
4. **Spatial modeling**
 - ❖ Constructing models to predict spatial outcomes.

21

2. GIS and Spatial Data Analysis

1. Spatial Data Manipulation

Classification and retrieval



22

2. GIS and Spatial Data Analysis

1. Spatial Data Manipulation

Classification and retrieval

Map Query 2

Layers: [Map Query 1], [Map Query 1_Count], [Dem2], [Dem2_Count], [Aspect of Dem2], [Slope2]

Sample Values: 0, 3.608, 7.016, 10.523, 14.031

Expression: `[(Slope2) > 10]`

Update Values

Evaluate

View1

Slope2

- 0 - 7.795
- 7.795 - 15.59
- 15.59 - 23.385
- 23.385 - 31.18
- 31.18 - 38.975
- 38.975 - 46.771
- 46.771 - 54.566
- 54.566 - 62.361
- 62.361 - 70.156
- No Data

Aspect2

7

View2

Slope > 10

- False (0)
- True (1)
- No Data

23

2. GIS and Spatial Data Analysis

1. Spatial Data Manipulation

Reclassify values

Reclassify Values

Classification Field: S_value

Classify Unique Lookup

Old Value	New Value
Cenozoic	0
Lower Mesozoic	0
Lower Paleozoic	0
Madison Limestone	1
Precambrian	0
Upper Mesozoic	0
Upper Paleozoic	0

Load Save OK Cancel

View1

Reclass of Geolgrid

- 0
- 1
- No Data

Geolgrid

- Cenozoic
- Lower Mesozoic
- Lower Paleozoic
- Madison Limestone
- Precambrian
- Upper Mesozoic
- Upper Paleozoic
- No Data

View2

Reclass of Geolgrid

- 0
- 1
- No Data

NbrMean of Dem2

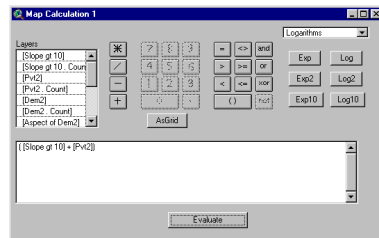
- 872 - 992
- 993 - 1112
- 1113 - 1233
- 1234 - 1353
- 1354 - 1474
- 1475 - 1594
- 1595 - 1715
- 1716 - 1835

24

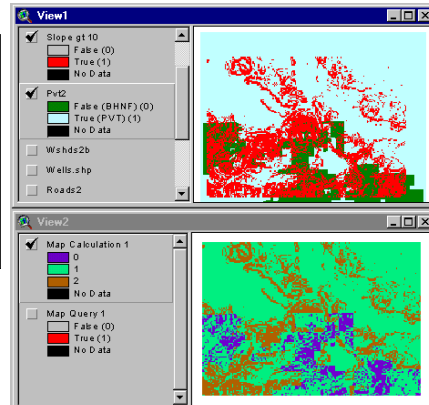
2. GIS and Spatial Data Analysis

1. Spatial Data Manipulation

Map Calculations



e.g. find combinations of public/private land and high/low slope



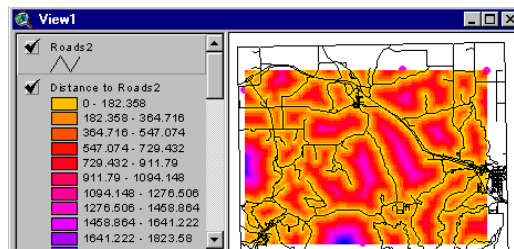
25

2. GIS and Spatial Data Analysis

1. Spatial Data Manipulation

Proximity Functions

- ⌘ Proximity functions measure distances between features
- ⌘ Distance functions
 - ☑ Calculate a surface representing the distance to the nearest target (raster)
 - ☑ Example: Determine the distance from archeological sites to the nearest stream



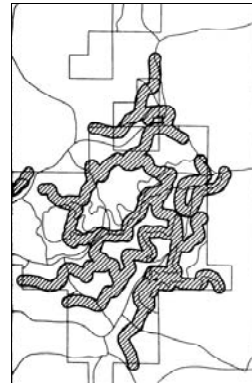
26

2. GIS and Spatial Data Analysis

1. Spatial Data Manipulation

Buffering

- ⌘ Constructs polygons containing all areas within a certain distance of the target
- ⌘ Targets may be points, lines, or polygons
- ⌘ Accomplished in raster using expand or distance functions
 - ☑ Example: How much timber is lost if one does not harvest < 100m from streams?
- ⌘ Often combined with overlay to tabulate areas inside the buffers



27

2. GIS and Spatial Data Analysis

1. Spatial Data Manipulation

Analyze Spatial Information...

OBJECTID*	TAXID	LANDKEY*	ADDITIO
45528	R247410003501	257700003501	ROYAL ADDITION 2LT 5 BLK A(LC)
4947	R274100030301	057200030301	TREYBURN ESTATES
49783	R274100031301	057200031301	TREYBURN ESTATES
49844	R274100031101	057200031101	TREYBURN ESTATES
50309	R274100034101	057200034101	TREYBURN ESTATES
51407	R191500100201	241600100201	CENTRAL CENTER LT 2
51844	R239500000601	046100000601	TREYBURN ADDITION
51846	R239500000501	046100000501	TREYBURN ADDITION
51950	R032000508101	019500508101	GREENHOLLOW ADDITION
51951	R032000509101	019500509101	GREENHOLLOW ADDITION
51960	R032000507001	019500507001	GREENHOLLOW ADDITION
52091	R239500000601	046100000601	TREYBURN ADDITION
52297	R032000508001	019500508001	GREENHOLLOW ADDITION
52440	R239500001401	046100001401	TREYBURN ADDITION
52471	R239500001501	046100001501	TREYBURN ADDITION
52780	R139600100101	227900100101	REPUBLICAN PLAND MOTORBA
53343	R239500000801	046100000801	TREYBURN ADDITION
53993	R239500001701	046100001701	TREYBURN ADDITION
53964	R032000506201	019500506201	GREENHOLLOW ADDITION
53964	R621200000401	621200000401	
54453	R139600100101	228200100101	PLAZA PARK 2
54753	R196200002001	243000002001	CENTRAL CENTER RETAIL LT 2 BL
55417	R240300100101	256100100101	WHITSON ADDN
55479	R193000100201	241900100201	BRIMAC OFFICE PARK LT 2R BLK 1
55734	R021000400101	002200400101	HAGGARD, W. O. ADDITION
55717	R021000400401	002200400401	HAGGARD, W. O. ADDITION
56018	R021000300901	002200300901	HAGGARD, W. O. ADDITION




Which parcels are within 5 km distance from x crossroad?

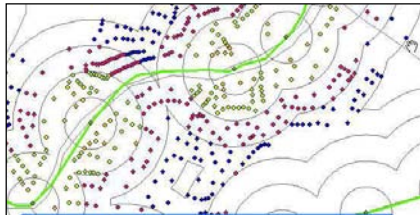
28

2. GIS and Spatial Data Analysis

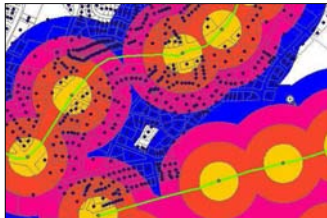
1. Spatial Data Manipulation



1. INPUT
Layer A: Parcels
Layer B: Stop BUS



2. BUFFER ZONE around Stop BUS



3. SPATIAL JOIN between Buffer Zone and Parcels Centroids

SUB_PARCEL	PARCEL_ID	sq_ft_val	X	Y	FromBufDist
000	2205990000	101	227481.3	895218.3	225
000	2205999000	106.07	227496.3	895224.2	225
1000	2206000000	314.812	227262.8	896126.2	225

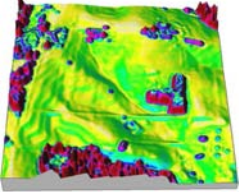
29

2. GIS and Spatial Data Analysis

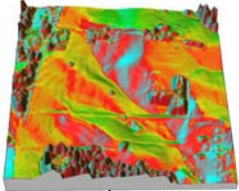
1. Spatial Data Manipulation

Topographic Analysis

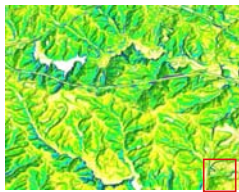
Source: *Modeling with Open Source GIS, H. Mitasova and M. Neteler, OSGEO*



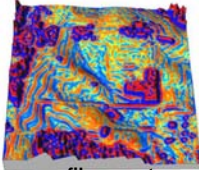
slope



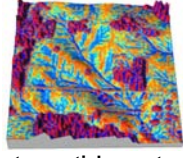
aspect



slope at 10m res., bare ground



profile curvature



tangential curvature

Topographic analysis from different DEMs and DSMs:

- 1m res. lidar DSM
- 10m res. lidar DEM
- 30m SRTM
- 500m DEM

30

2. GIS and Spatial Data Analysis

1. Spatial Data Manipulation

Viewshed analysis

*Source: Geospatial Analysis and Modeling
MEA592 – Helena Mitasova*

90 deg - horizontal
0 deg - under viewers feet
180 deg - above viewers head

visibility analysis for finding a location for a webcam to support real-time monitoring of a watershed

- height of webcam location above the surface 1.75 and 6m

31

2. GIS and Spatial Data Analysis

1. Spatial Data Manipulation

Network analysis

Source: Modeling with Open Source GIS, H. Mitasova and M. Neteler, OSGEO

- Finding best location for new fire station
- Point locations are constrained by network
- Response times also determined by network

Find shortest path between two hospitals based on cost (1/speed) and distance

32

2. GIS and Spatial Data Analysis

1. Spatial Data Manipulation

Network analysis

Modeled Response Time
Threshold Polygons
(Scenario 0)

Department of Geography
and Urban Studies, Temple
University

Modeled Response Time
Scenario 7 (Variation 2)

FIGURE 3 Modeled response times for current geographical arrangement of fire stations.

dded demands to obtain the specific intersection. Finally, we allocated each node to fire station and determined that node intersection from

33

2. GIS and Spatial Data Analysis

1. Spatial Data Manipulation

Cost surface analysis: e.g. accident

Source: Modeling with Open Source GIS, H. Mitasova and M. Neteler, OSGEO

Speed map

● accidentent ◆ firestations

- convert street speed limit data into time (use 5km/hr for off-street areas) ,
- compute cost surface to reach site of an accident from any point in the area,
- find closest fire station, compute shortest path from it

Can be solved both with raster tools and vector network analysis tools

34

3. The Process of Spatial Analysis

2. Spatial Data analysis

Anselin and Getis (1992) divided into 4 components the analysis functions of GIS:

- ❑ **Selection** (sampling of data from the database)
- ❑ **Manipulation** (partitioning, aggregation, overlay, buffering, and interpolation)
- ❑ **Exploration**
- ❑ **Validation**

} Typical features of GIS

} Data Analysis Module

35

3. The Process of Spatial Analysis

2. Spatial Data analysis

What's about....

SPATIAL INTERACTION?

SPATIAL STRUCTURE?

SPATIAL PROCESSES?

Internal Ward Migration

Total Migrants

- 115 - 473
- 163 - 300
- 45 - 149

Gross Inter-Ward Flow

Sum of Two-Way Flow

- 121 - 225
- 101 - 150
- 50 - 100

Migration Density Level

High

Low

Spatial Interaction - Geovisualisation for Glasgow Migration of Populations

Image source: <http://sites.google.com/site/flowmappingexamples/>

The emphasis of **Spatial Analysis** is to measure properties and relationships, taking into account the spatial location of the phenomenon under study.

36

3. The Process of Spatial Analysis

2. Spatial Data analysis

Objectives of spatial analysis...

- **Detect spatial patterns** that cannot be detected by visual analysis.
- **Measure properties and relationships**, taking into account the spatial location of the phenomenon under study.
- **Confirm whether a spatial pattern** found in visual analysis is significant (nonrandom).

37

3. The Process of Spatial Analysis

2. Spatial Data analysis

Components of spatial analysis (Anselin, 2002):

- **Visualization**
Detecting interesting patterns on maps
- **Exploratory Spatial Data Analysis (ESDA)**
Finding interesting patterns;
Quantifying, with global and local indices, spatial associations
- **Spatial Modeling**
Explaining interesting patterns

38

4. Types of data in Spatial Analysis

2. Spatial Data analysis

Types of Spatial Analysis

Methods used in Spatial Analysis can be divided according to the three main categories of data to be analyzed:

- ❖ Events or Points Patterns
- ❖ Areas
- ❖ Continuous data or surfaces

39

4. Types of data in Spatial Analysis

2. Spatial Data analysis

Data types in spatial analysis...

Methods used in Spatial Analysis can be divided according to the 3 main categories of data to be analyzed:

Events or point patterns – phenomena expressed through occurrences identified as points in space, denominated point processes.

Some examples are: *crime spots, disease occurrences, and the localization of vegetal species.*

Areas with Counts and Aggregated Rates – means data associated to population surveys, like census and health statistics, and that are originally referred to individuals situated in specific points in space. For confidentiality reasons these data are aggregated in analysis units, usually delimited by closed polygons (*census tracts, postal addressing zones, municipalities*).

Continuous surfaces – estimated from a set of field samples that can be regularly or irregularly distributed.

40


5. Events or Point Pattern

2. Spatial Data analysis

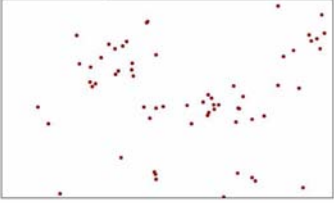
Point Pattern

❖ The basic interest of spatial point pattern analysis will be to detect whether it is distributed at **random** or represents a **clustered** or **regular** pattern.

points in geographic context



pattern of points in uniform plane



Spatial Objects

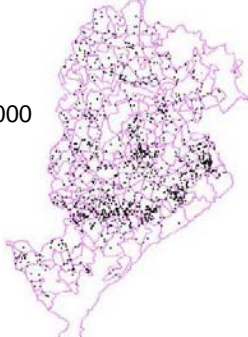
41

5. Events or Point Pattern

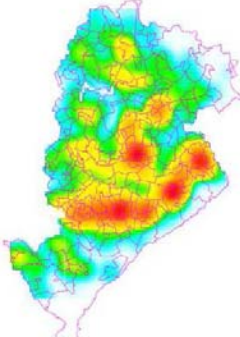
2. Spatial Data analysis

Point Pattern examples... crime data in Belo Horizonte (Brazil)

Taxi robbery 1995-2000



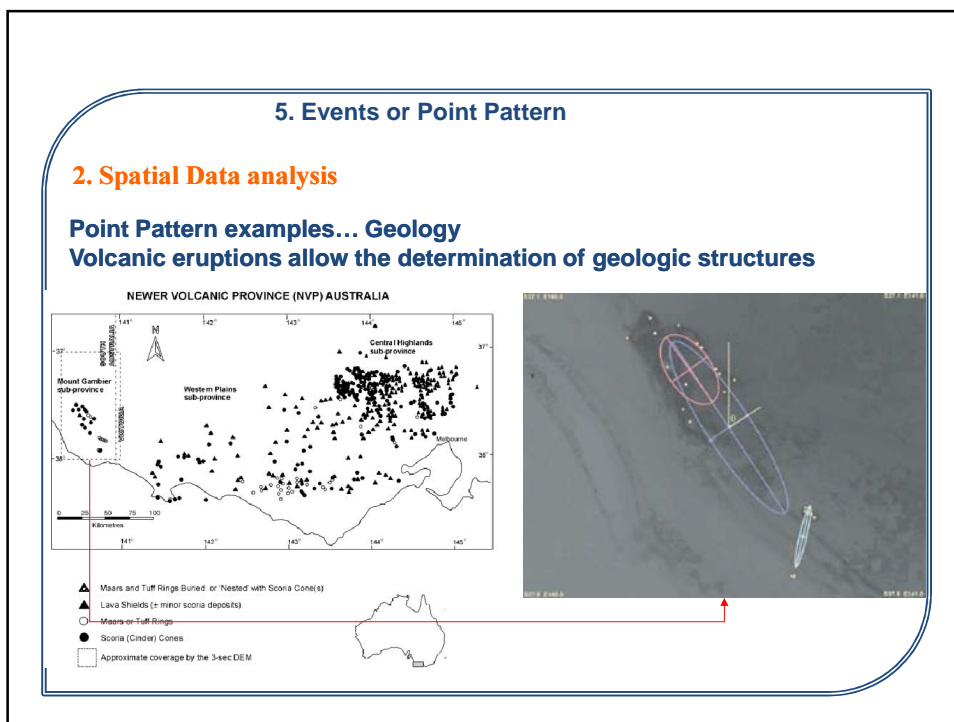
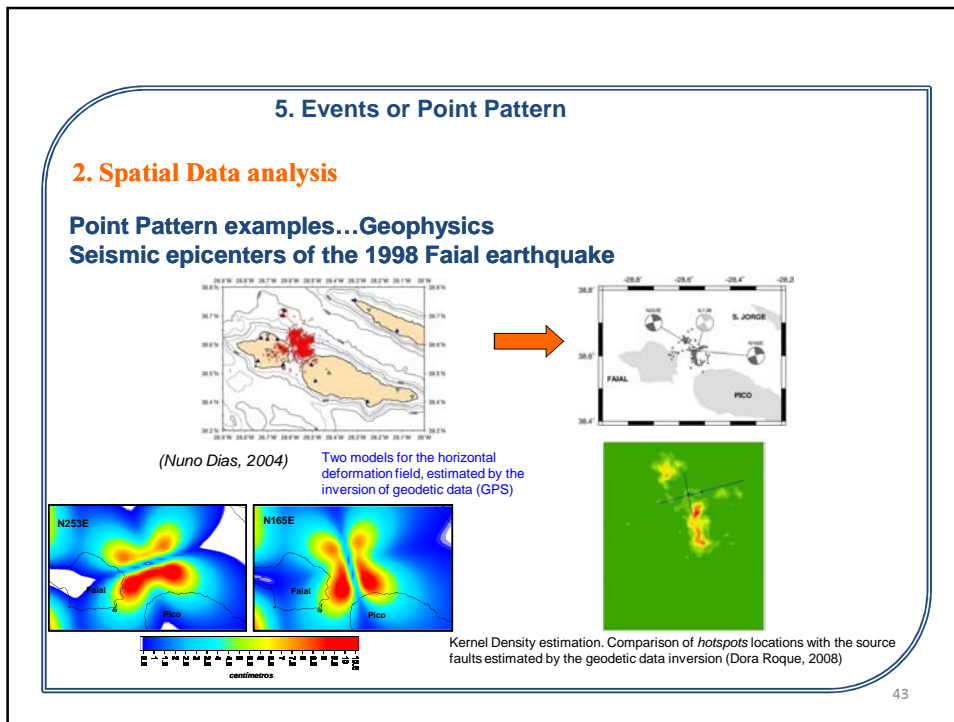
Event locations



Density estimation events/unit area

Source: Renato Assunção (UFMG/Brasil)

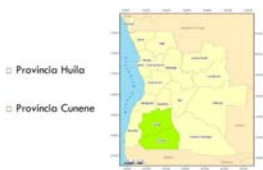
42




5. Events or Point Pattern

2. Spatial Data analysis


Point Pattern examples... Demography



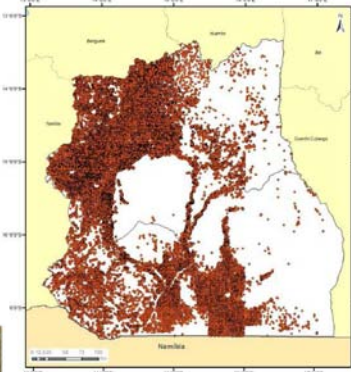
□ Provincia Hwange
□ Provincia Cunene



Vista aérea de um Quimbo



Exemplo, etnia Herero, tribo Himba



Distribuição dos Quimbos

From (Santos, O., 2008)

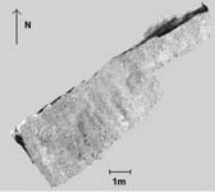
45

5. Events or Point Pattern

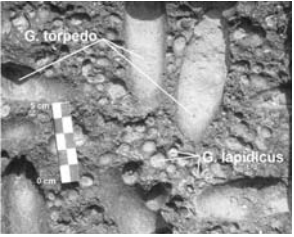
2. Spatial Data analysis

(Trace fossils for paleogeographic studies, Redweik et al., 2010 submitted to AGILE 2010)


Point Pattern examples... Paleontology




Ortofoto

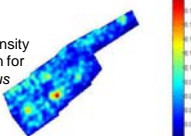


G. torpedo
G. lapidicus





Maps of complete sets of extracted *G.lapidicus*



Kernel density estimation for *G.lapidicus*

5. Events or Point Pattern

2. Spatial Data analysis

Point Pattern examples... Public Health
 People with lung and larynx cancer, in Lancashire (England), between 1974-1983 (Gatrell and Lovett (1978))

Population Backcloth
 large incinerator, active between 1972-1980, is located just upwind of this cluster of cases.

- Lungue cancer
- Larynx cancer

47

5. Events or Point Pattern

2. Spatial Data analysis

- Visualization
 - **Descriptors** of point patterns: *fast perception of the point distribution, (i.e. variability and spatial orientation);*
 - Frequency
 - Intensity
 - Geometric Center
 - Spatial dispersion
- ESDA
 - **Detectors** of point pattern: *Methods to derive summary statistics or plots of the observed distribution to investigate specific hypotheses;*
 - Examining first order effects (large scale trends)
 - Examining second order effects (local effects)

48

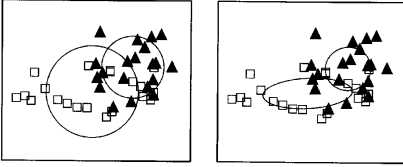
5. Events or Point Pattern

2. Spatial Data analysis

Visualization

Centrography

- **Central Tendency**
 - mean center = center of gravity
average of x and y coordinates
 - median center = location with shortest total distance to all points
distance is non-linear, iterative algorithm
- **Second Moment**
 - standard deviational ellipse
variation around mean center accounting for angle

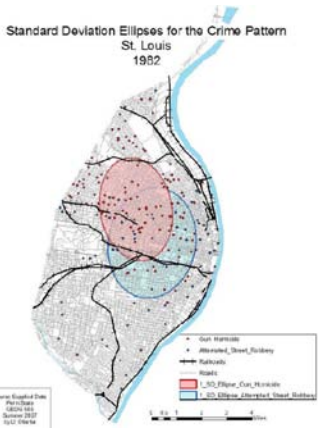


5. Events or Point Pattern

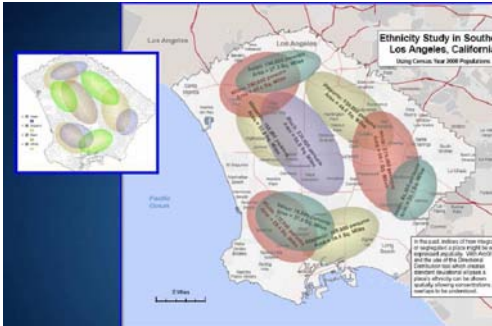
2. Spatial Data analysis

Visualization

Standard Deviation Ellipses for the Crime Pattern
St. Louis
1982



Ethnicity Study in Southern
Los Angeles, California
(Using Census Year 2000 Population)



50

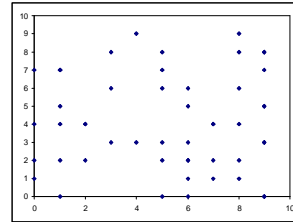
5. Events or Point Pattern

2. Spatial Data analysis

ESDA

Independent random process (IRP)
Complete spatial randomness (CSR)

1. **Equal probability:** any point has equal probability of being in any position or, equivalently, each small sub-area of the map has an equal chance of receiving a point.
2. **Independence:** the positioning of any point is independent of the positioning of any other point.



$$P(k, n, x) = \binom{n}{k} \left(\frac{1}{x}\right)^k \left(\frac{x-1}{x}\right)^{n-k}$$

51

5. Events or Point Pattern

2. Spatial Data analysis

ESDA

IRP/CSR is not realistic

- The **independent random process** is mathematically elegant and forms a useful starting point for spatial analysis, but its use is often exceedingly naive and **unrealistic**.
- If real-world spatial patterns were indeed generated by unconstrained randomness, geography would have little meaning or interest, and most GIS operations would be pointless.

52

5. Events or Point Pattern

2. Spatial Data analysis

ESDA

1. First-order effect

- The assumption of **Equal probability** cannot be satisfied
- ❖ The locations of disease cases tends to cluster in more densely populated areas
- ❖ Plants are always clustered in the areas with favored soils.

2. Second-order effect

- The assumption of **Independence** cannot be satisfied
- ❖ New developed residential areas tend to near to existing residential areas
- ❖ Stores of McDonald tend to be far away from each other.

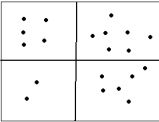
53

5. Events or Point Pattern

2. Spatial Data analysis

ESDA – 1ST ORDER EFFECTS

Density-based point pattern measures - **Quadrat Count Methods**



$$\text{Mean} = \frac{\text{No. of pts. in the region}}{\text{No. of quadrants}} = \frac{20}{4} = 5 \quad \mathbf{n = n. quadrats}$$


$$\text{Variance} = \frac{\sum x_i^2 - \frac{(\sum x_i)^2}{n}}{n-1} = \frac{2^2 + 5^2 + 6^2 + 7^2 - \frac{(20)^2}{4}}{4-1} = 4.5$$

x_i – frequency in each quadrat

Variance to Mean Ratio


➔ $VTMR = \frac{\text{Variance}}{\text{Mean}} = \frac{4.5}{5} = 0.9$

cluster



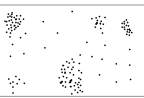
VTMR > 1

dispersed



VTMR < 1

random



VTMR = 1

5. Events or Point Pattern

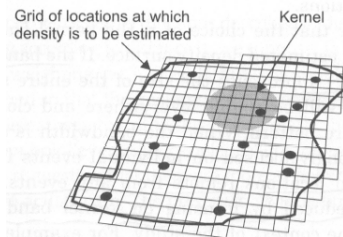
2. Spatial Data analysis

ESDA – 1ST ORDER EFFECTS

Density-based point pattern measures – Kernel Density

The pattern has a density at any location in the study region, not just locations where there is an event

This density is estimated by counting the number of events in a region, or kernel, centered at the location where the estimate is to be made.



$$\hat{\lambda}_p = \frac{no.[S \in C(p,r)]}{\pi r^2}$$

$C(p,r)$ is a circle of radius r centered at the location of interest p

55

5. Events or Point Pattern

2. Spatial Data analysis

ESDA – 1ST ORDER EFFECTS

Density-based point pattern measures – Kernel Density

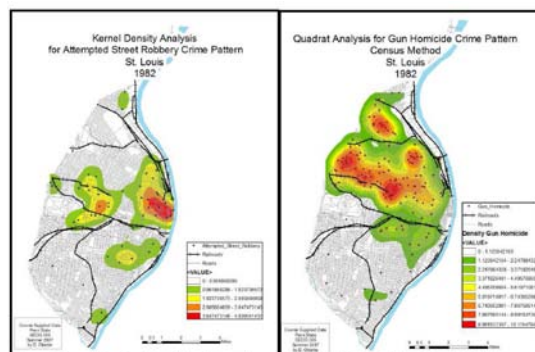


Figure 4. Kernel Density Analysis for Attempted Street Robbery

Figure 5. Kernel Density Analysis for Gun Homicide

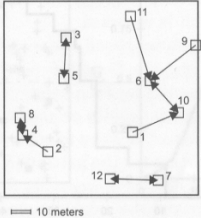
56

5. Events or Point Pattern

2. Spatial Data analysis

ESDA – 2nd ORDER EFFECTS
Distance-based point pattern measures

If clustered, \bar{d}_{\min} has a higher or lower value?



Event	x	y	Nearest neighbor	d_{\min}
1	66.22	32.54	10	25.59
2	22.52	22.39	4	15.64
3	31.01	81.21	5	21.14
4	9.47	31.02	8	9.00
5	30.78	60.10	3	21.14
6	75.21	58.93	10	21.94
7	79.26	7.68	12	24.81
8	8.23	39.93	4	9.00
9	98.73	42.53	6	21.94
10	89.78	42.53	6	21.94
11	65.19	92.08	6	34.63
12	54.46	8.48	7	24.81

$\bar{d}_{\min} = 21.62$

57

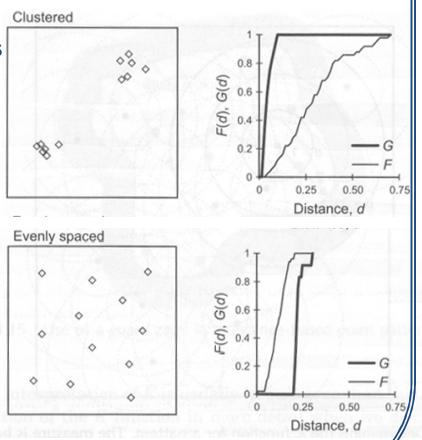
5. Events or Point Pattern

2. Spatial Data analysis

ESDA – 2nd ORDER EFFECTS
Distance-based point pattern measures

$$G(d) = \frac{no.[d_{\min}(s_i) < d]}{n}$$

$$F(d) = \frac{no.[d_{\min}(p_i, S) < d]}{m}$$

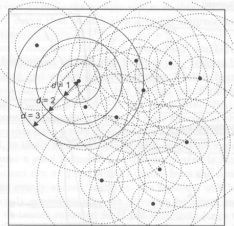


58

5. Events or Point Pattern

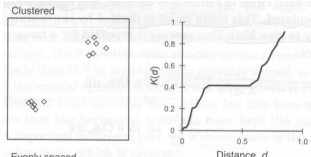
2. Spatial Data analysis

ESDA – 2nd ORDER EFFECTS
Distance-based point pattern measures
K-Ripley (Ripley, 1976) – uses distances between all points

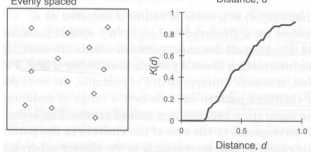


$\lambda = \frac{n}{a}$ is the study area event density

Clustered



Evenly spaced



$$K(d) = \frac{\sum_{i=1}^n no.[S \in C(s_i, d)]}{n\lambda}$$

$$= \frac{a}{n} \cdot \frac{1}{n} \sum_{i=1}^n no.[S \in C(s_i, d)]$$

59

5. Events or Point Pattern

3. Spatial Statistical analysis

Validation

Clusters were detect... But How cluster they are?

We need to give a statistical significance to our exploratory results!

But how?

Processes

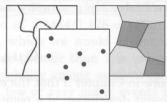
MATHEMATICAL DESCRIPTION
or
COMPUTER SIMULATION

↓

EXPECTED VALUES
and/or
DISTRIBUTIONS

MODELS or THEORIES or HYPOTHESES

Patterns



e.g.
Quadrat counts
Density surfaces
Nearest neighbor
G, F, and K

↓

MEASURES of PATTERN

or

STATISTICS
or
HYPOTHESIS TEST

OBSERVATIONAL DATA

← STATISTICAL TEST or HYPOTHESIS TEST →

Is the observed pattern a likely realization of the hypothesized process?

60

5. Events or Point Pattern

3. Spatial Statistical analysis

Complete spatial randomness, CSR

1. **Equal probability:** any point has equal probability of being in any position or, equivalently, each small sub-area of the map has an equal chance of receiving a point.
2. **Independence:** the positioning of any point is independent of the positioning of any other point.

61

5. Events or Point Pattern

3. Spatial Statistical analysis

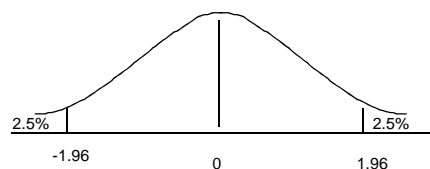
Statistical Hypothesis Testing: Classical Approach

Statistical hypothesis testing usually involves 2 values;

- A measure(s) or index(s) derived from samples (e.g. the Nearest Neighbor Index)
- A *test statistic*, derived from the measure or index, whose probability distribution is known when repeated samples are made.
 - this is used to test the *statistical significance* of the measure/index

We proceed from the null hypothesis (H_0) that, in the population, there is “no difference” between the two sample statistics, or from spatial randomness*

- If the test statistic we obtain is very unlikely to have occurred (less than 5% chance) if the null hypothesis was true, the null hypothesis is rejected



If the test statistic is beyond +/- 1.96 (assuming a Normal distribution), we reject the null hypothesis (of no difference) and assume a statistically significant difference at at least the 0.05 significance level.

62

5. Events or Point Pattern

3. Spatial Statistical analysis

Statistical Hypothesis Testing: Simulation Approach

- Because of the complexity inherent in spatial processes, it is sometime difficult to derive a legitimate test statistic whose probability distribution is known
- An alternative approach is to use the computer to simulate multiple random spatial patterns (or samples) the spatial statistic (e.g. NNI or LISA) is calculated for each, and then displayed as a frequency distribution.

Empirical frequency distribution from 499 random patterns ("samples")

63

5. Events or Point Pattern

3. Spatial Statistical analysis

Under CSR hypothesis, the distribution function $G(w)$ is a Poisson process (Bailey e Gatrell, 1995):

$$G(w) = 1 - e^{-\lambda\pi w^2} \quad w \geq 0$$

And the simulation for $G(w)$, under CSR, is given by:

$$\bar{G}(w) = \frac{\sum_{i=1}^m \hat{G}_i(w)}{m}$$

Where : $\hat{G}_i(w)$ $i=1, 2, \dots, m$ são funções de distribuição empíricas, estimadas a partir de m simulações independentes dos n eventos, na hipótese CSR (n eventos independentes e uniformemente distribuídos).

$$U(w) = \max \{ \hat{G}_i(w) \}, i=1, \dots, m$$

$$L(w) = \min \{ \hat{G}_i(w) \}, i=1, \dots, m$$

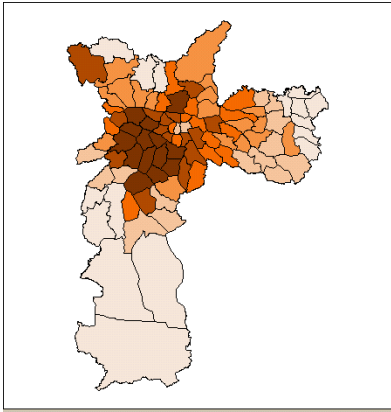
— Envelope Inferior
— Estimado
— Envelope Superior

6. Areas

2. Spatial Data analysis

Areas

❖ The basic interest of spatial analysis of areas is to measure spatial autocorrelation

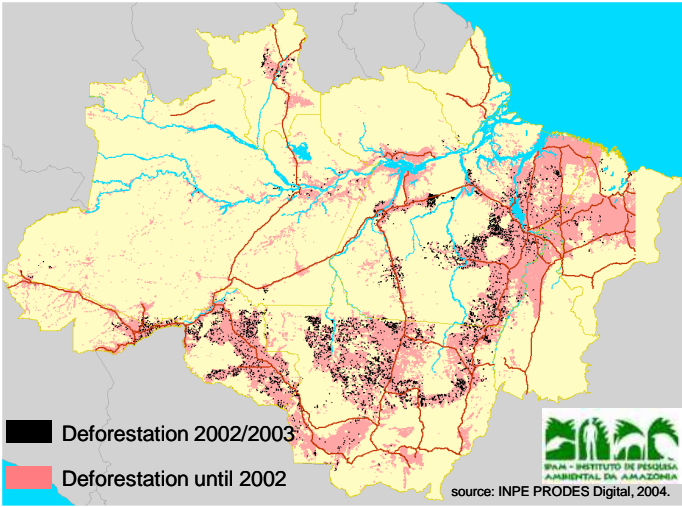


Is this data spatially autocorrelated?

65

Let's look to some examples....

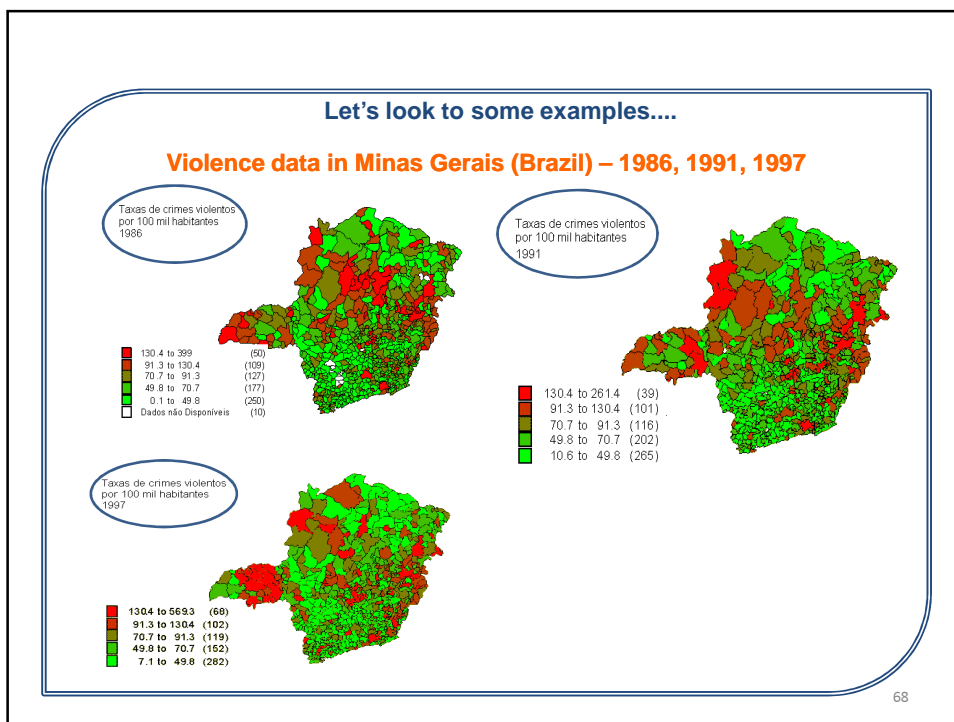
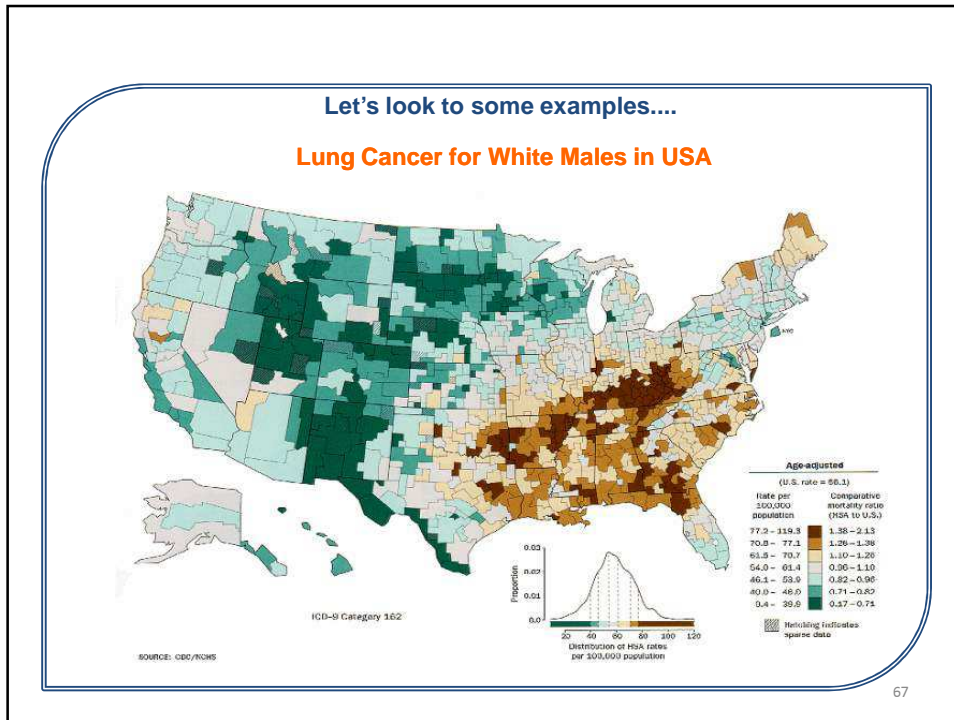
Amazon Deforestation 2003



■ Deforestation 2002/2003
■ Deforestation until 2002

source: INPE PRODES Digital, 2004.

66



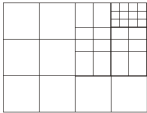
6. Areas

2. Spatial Data analysis


MAUP – Modifiable Area Unit Problem

Scale effect: is the tendency, within a system of modifiable areal units, for different statistical results to be obtained from the same set of data when the information is grouped at different levels of spatial resolution (e.g., enumeration areas, census tracts, cities, regions).

Aggregation effect: is the variability in statistical results obtained within a set of modifiable units as a function of the various ways these units can be grouped at a given scale, and not as a result of the variation in the size of those areas



Scale



Aggregation

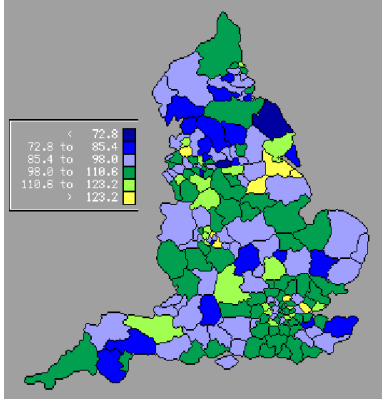
69

6. Areas

Visualization...

- Grouping
 - Equal intervals
 - Quantiles
 - Standard deviation

- Be careful!
 - Color mapas can lead to wrong interpretation



Breast cancer in England (1985-1989)

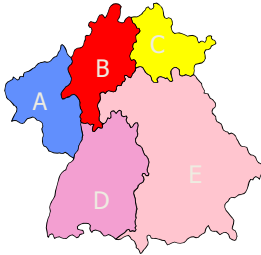
Source: Bailey and Gattrel, 1995

70

6. Areas

Spatial Proximity Matrix

- Matrix $W(n \times n)$, where each elements w_{ij} represents a measure of nearness between O_i and O_j
- Criteria:
 - $w_{ij} = 1$, if O_i touches O_j
 - $w_{ij} = 1$, if $\text{distance}(O_i, O_j) < h$



	A	B	C	D	E
A	0	1	0	1	0
B	1	0	1	1	1
C	0	1	0	0	1
D	1	1	0	0	1
E	0	1	1	1	0

71

6. Areas

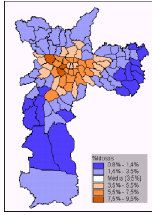
Moving Averages

- Local smoothing of attribute values

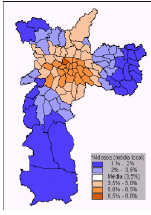
$$\hat{\mu}_i = \frac{\sum_{j=1}^n W_{ij} y_j}{\sum_{j=1}^n W_{ij}} \quad i = 1, 2, \dots, n$$

where:

- ▶ W_{ij} is the spatial weights matrix.
- ▶ y_j is the attribute value for each area.
- ▶ n is the number of areas



Standard Deviations



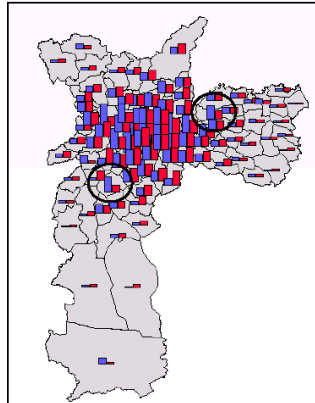
Moving Averages

Proportion of population aged 70 or older, São Paulo, 1991

72

6. Areas

Moving Averages



Source: INPE, Cămara

Regions where there is a large difference between the original value and the local mean

Indicates places of spatial transitions

73

6. Areas

Indicators of spatial autocorrelations

- Generic formulation:

$$\Gamma_i^{\text{local}} = \sum_j^n w_{ij} a_{ij}$$

$$\Gamma^{\text{global}} = \sum_i^n \sum_j^n w_{ij} a_{ij}$$

where

w_{ij} : spatial proximity between i and j

a_{ij} : measured relation between object and its neighbors

74

6. Areas

Global Indicators of Spatial Autocorrelation

- Moran

$$I = \frac{n}{\sum_{i=1}^n \sum_{j=1}^n w_{ij}} \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} (y_i - \bar{y})(y_j - \bar{y})}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

onde:

- n number of areas,
- y_i attribute value in area i ,
- mean value in study region
- w_{ij} spatial weights matrix.

- How to interpret the above equation?

Values of I go from -1 to +1.
 ⇒ -1 negative spatial autocorrelation
 ⇒ 0 no spatial autocorrelation
 ⇒ +1 positive spatial autocorrelation

75

6. Areas

Random or Clustered?

Values of I in the simulated landscapes

Value of I in the observed landscapes
I = 0.7578

- Testing Moran's I
 - Permutate the spatial values 999 times
 - Obtain a probability distribution
 - Locate the "real" value in the distribution
 - In this case, I = .45 (very significant!)

Reject null hypothesis (Complete Spatial Randomness), i.e. the variable doesn't depend of it's spatial location.

76

6. Areas

The Moran's I statistic for spatial autocorrelation is given as:

$$I = \frac{n \sum_{i=1}^n \sum_{j=1}^n w_{i,j} z_i z_j}{S_0 \sum_{i=1}^n z_i^2} \quad (1)$$

where z_i is the deviation of an attribute for feature i from its mean ($x_i - \bar{X}$), $w_{i,j}$ is the spatial weight between feature i and j , n is equal to the total number of features, and S_0 is the aggregate of all the spatial weights:

$$S_0 = \sum_{i=1}^n \sum_{j=1}^n w_{i,j} \quad (2)$$

The z_I -score for the statistic is computed as:

$$z_I = \frac{I - E[I]}{\sqrt{V[I]}} \quad (3)$$

where:

$$E[I] = -1/(n - 1) \quad (4)$$

$$V[I] = E[I^2] - E[I]^2 \quad (5)$$

[http://webhelp.esri.com/arcgisdesktop/9.3/index.cfm?TopicName=How%20Spatial%20Autocorrelation:%20Moran's%20I%20\(Spatial%20Statistics\)%20works](http://webhelp.esri.com/arcgisdesktop/9.3/index.cfm?TopicName=How%20Spatial%20Autocorrelation:%20Moran's%20I%20(Spatial%20Statistics)%20works)

$I > -1/(n-1)$, ou $z_I > 0$

Cluster of high values or low values; Positive autocorrelation

$I < -1/(n-1)$, ou $z_I < 0$

Dispersion; negative autocorrelation

77

6. Areas

Local Indicators of Spatial Autocorrelation (LISA)

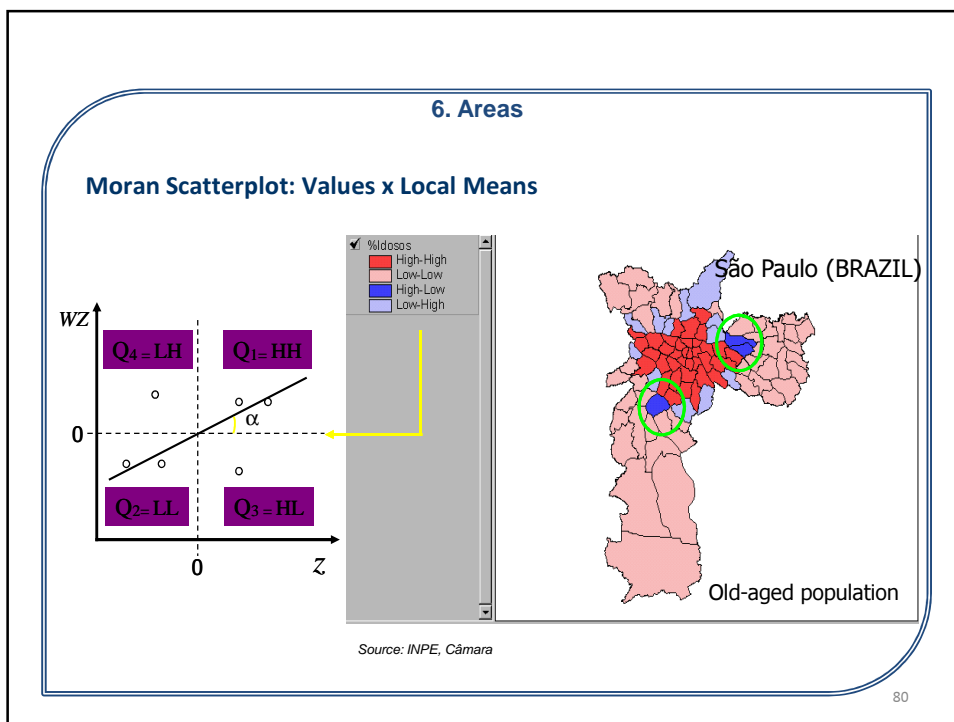
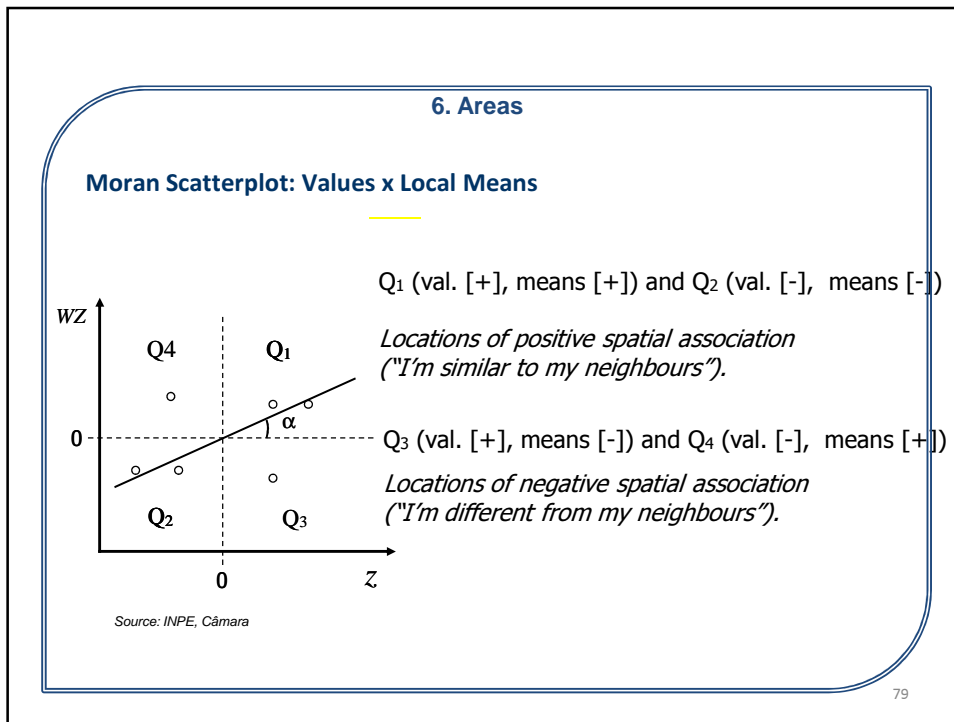
What if we want to find out the spatial correlation of each area?

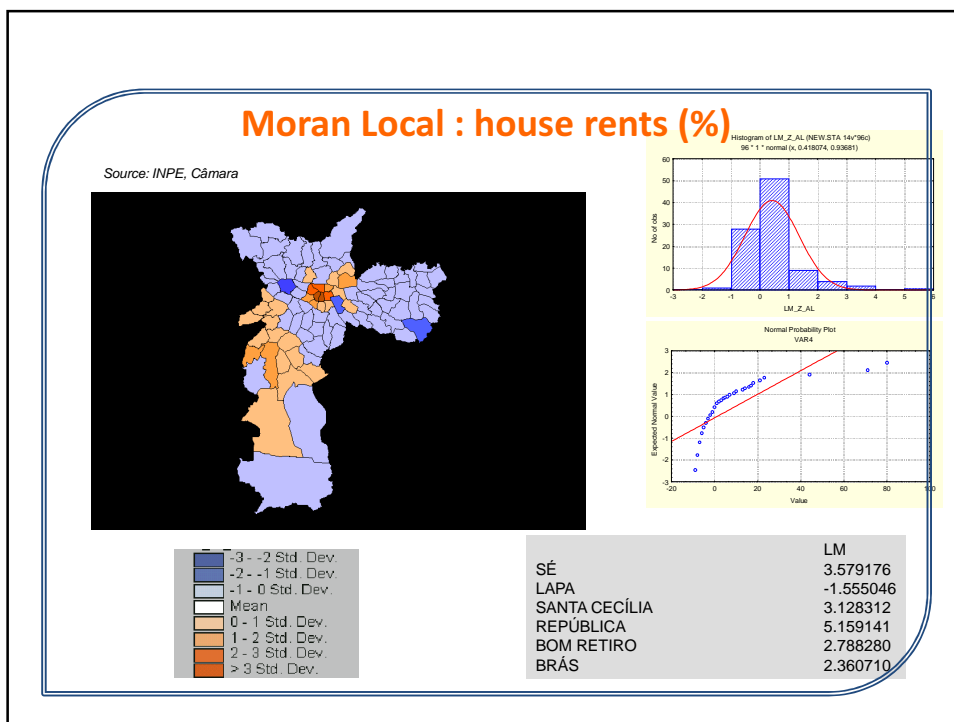
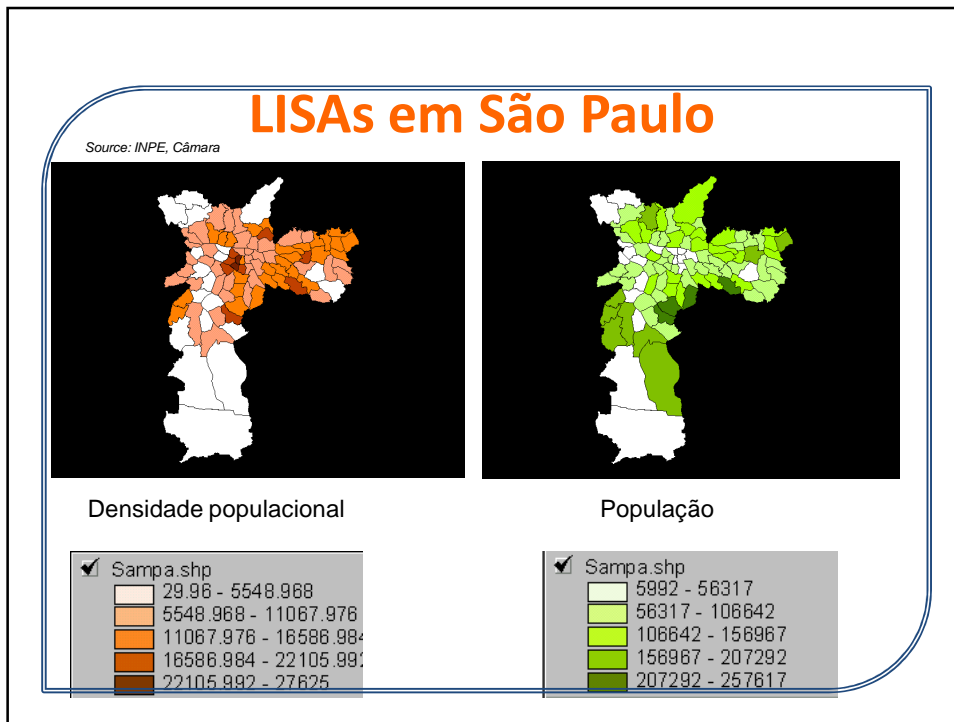
Use a local indicator
Compares local value to that of its neighbours

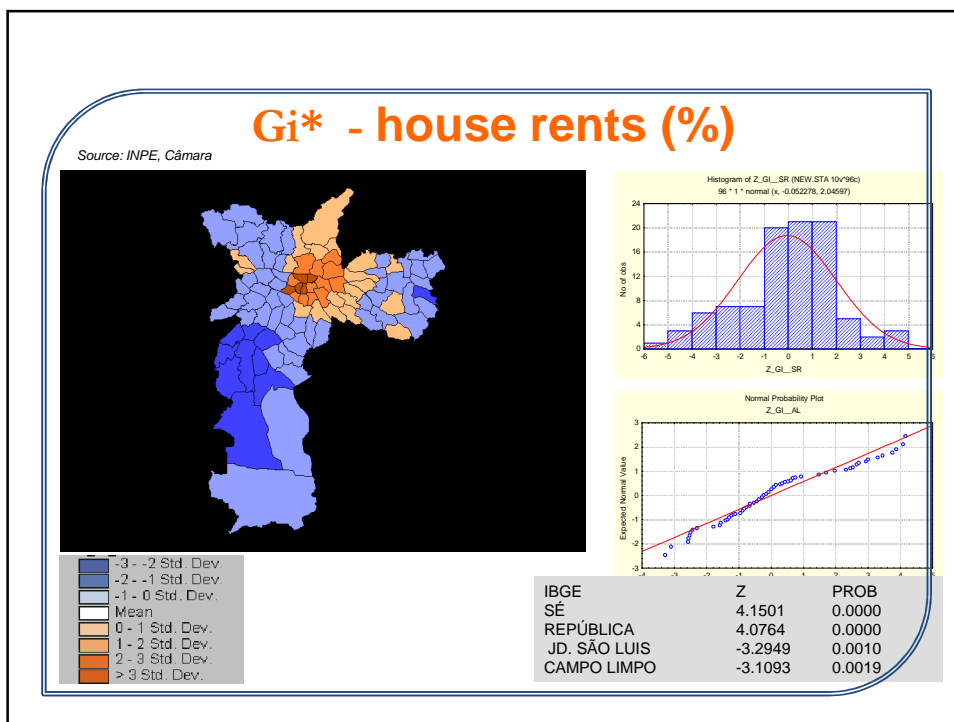
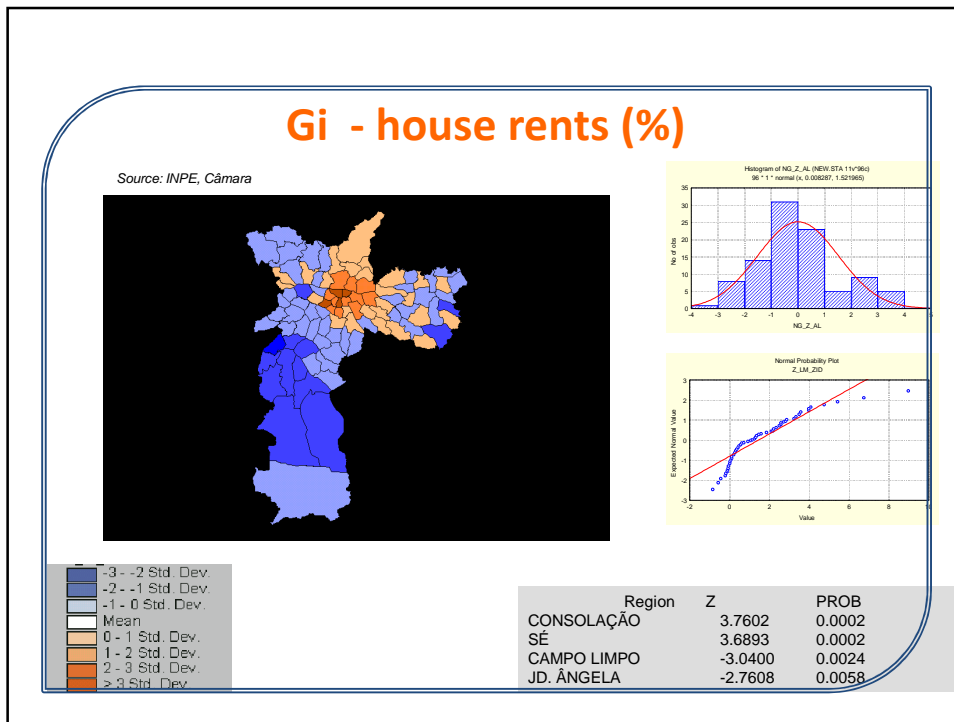
$$I_i = \frac{(y_i - \bar{y}) \sum_{j=1}^n w_{ij} (y_j - \bar{y})}{\sum_{j=1}^n (y_j - \bar{y})^2 \cdot n}$$

- **Identify Hot Spots**
 - significant local clusters in the absence of global autocorrelation
 - significant local outliers
- **Indicate Local Instability**
 - local deviations from global pattern of spatial autocorrelation

78







7. Continuous Surfaces

2. Spatial Data analysis

Surfaces ❖ The basic interest now is spatial interpolation of sample data.

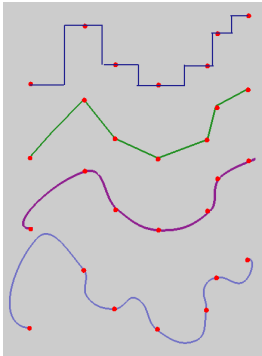
Spatial interpolation requires two basic input:
Known points and an **interpolation method**.

The problem of interpolation

Find a function that passes through a set of points:

curve: $y = f(x)$
 surface: $z = f(x,y)$
 volume: $w = f(x,y,z)$
 Dynamic volume : $w = f(x,y,z,t)$

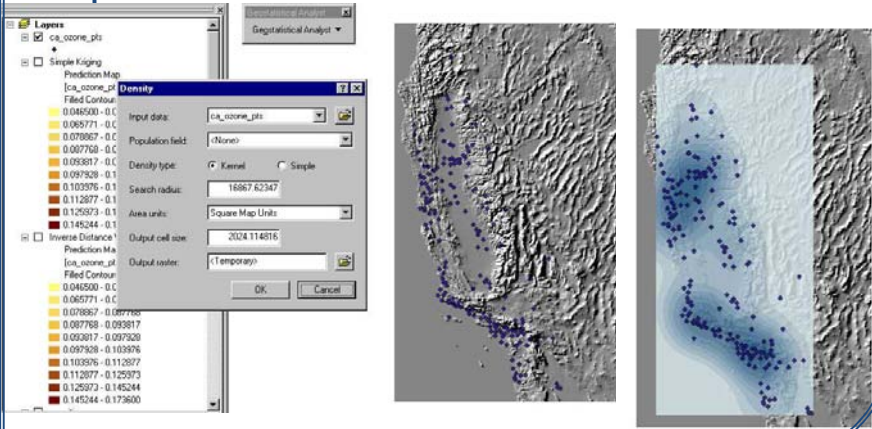
The problem has no unique solution!



7. Continuous Surfaces

2. Spatial Data analysis

Example

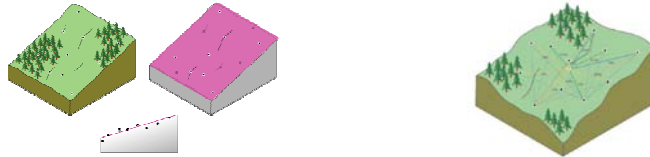


86

7. Continuous Surfaces

2. Spatial Data analysis

Global		Local	
Deterministic	Stochastic	Deterministic	Stochastic
Trend surface (inexact)*	Regression (inexact)	Thiessen (exact) Density estimation (inexact) Inverse distance weighted (exact) Splines (exact)	Kriging (exact)



87

THANK YOU VERY MUCH FOR YOUR ATTENTION!



<http://degge.fc.ul.pt/>

Cristina Catita
 Faculdade de Ciências da Universidade de Lisboa
 Departamento de Engenharia Geográfica, Geofísica e Energia
 Bloco C8
 Rua Ernesto Vasconcelos
 1749-016 Campo Grande, Lisboa
 Portugal
 Tel: +351 21 750 08 31 (extensão: 28146)
 email: cmcatita@fc.ul.pt

88