Spatial Autocorrelation: Introduction to Concepts

Geographical Information System (GIS)
[IT60111]

Department of Computer Science and Engineering
Indian Institute of Technology, Kharagpur
West Bengal, India

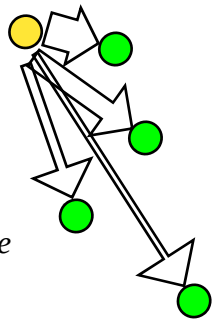
Contents

- What is Spatial Autocorrelation?
- Why Spatial Autocorrelation is Important?
- How to Measure Spatial Autocorrelation?
- Examples

WHAT IS SPATIAL AUTOCORRELATION?

SPATIAL AUTOCORRELATION

- Spatial Autocorrelation is a special *property of geospatial data*.
- It is the formal property that measures the *degree to which near and distant things are related*
  - It is a *statistical test* of match between locational similarity and attribute similarity
  - It is a property that is often exhibited by variables which are sampled over space
  - It is based on ***Tobler's 1st law of geography***.

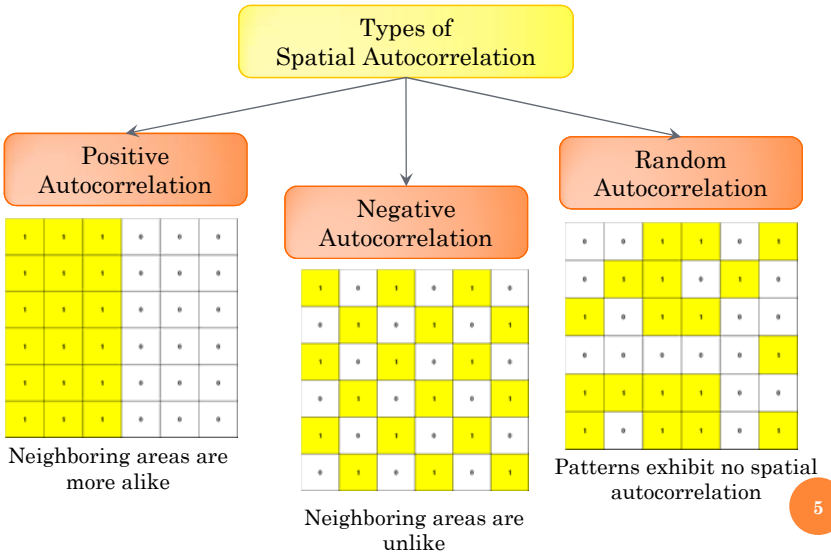


- **Tobler's 1st law of geography:**

  "*All places are related but nearby places are more related than distant places*"

- **Examples:**
  - Temperature values of two locations near to each other will be similar.

Types of Spatial Autocorrelation



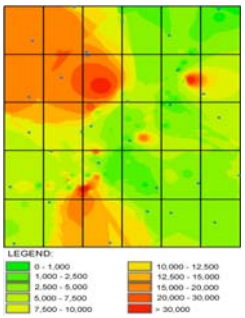
5

Why Spatial Autocorrelation is Important?

6

Importance of Spatial Autocorrelation

- Most *statistics are based on the assumption that* the values of observations in *each sample are independent of one another*.
- If the samples were taken from nearby areas, then *positive spatial autocorrelation may violate this*.



Goals:

- To *Measure the strength* of spatial autocorrelation in a map
- *Test the assumption* of independence or randomness
- To explore *whether there is any clustering pattern* in the data or is it just a random data

7

How to Measure Spatial Autocorrelation?

8

MEASURING SPATIAL AUTOCORRELATION

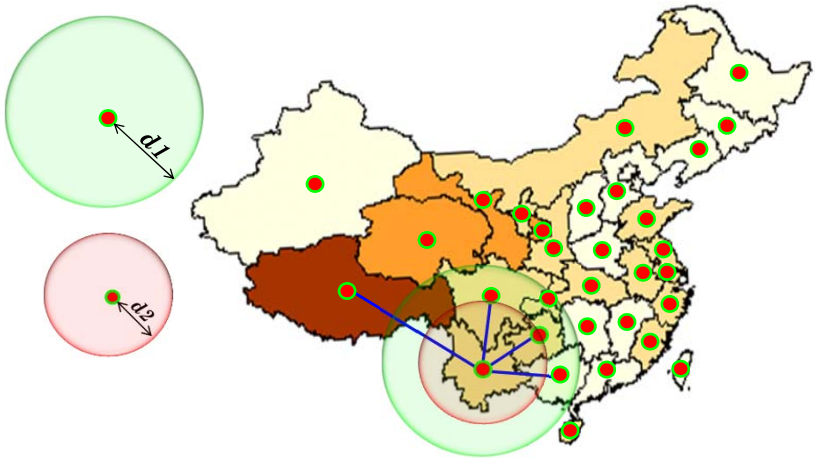
Steps in determining the extent of spatial autocorrelation:

- ***Step-1:*** Find out which areas are linked to one another
  - Choose a **neighborhood criterion**
- ***Step-2:*** Assign weights to the areas that are linked
  - Create a **spatial weights matrix**
- ***Step-3:*** Run **statistical test**, using weights matrix, to examine spatial autocorrelation

9

NEIGHBORHOOD CRITERIA

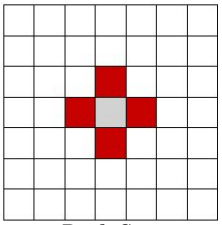
- Contiguity (common boundary)
- Distance (K-nearest neighbors, distance band)
  - ***How many*** "neighbors" to include, ***what distance*** do we use?



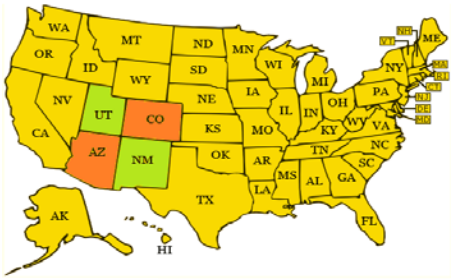
10

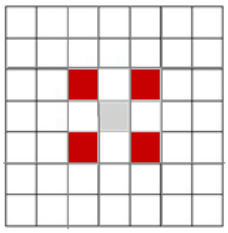
CONTIGUITY

- Adjacency---Sharing a border/boundary or point
- **For Regular Polygons**

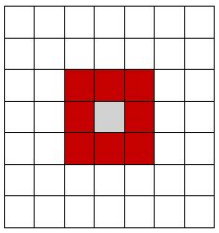


Rook Case





Bishop Case



Queen Case

- **For Irregular polygons**

All polygons that share a common border or have a centroid within the circle defined by the average distance to centroids of polygons that share a common border.

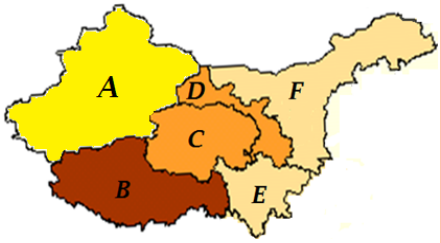
11

SPATIAL WEIGHT MATRIX

- **Weights based on Contiguity**
  - If zone *j* is adjacent to zone *i*, the interaction receives a weight of 1, otherwise it receives a weight of 0 and is essentially excluded

w_{ij} =

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



- **Weights based on Distance**
  - Uses a measure of the actual distance between points or between polygon centroids.
  - Most common choices are:
    - *inverse (reciprocal)* : $w_{ij} = 1/d_{ij}$
    - *inverse of squared distance* : $w_{ij} = 1/d_{ij}^2$
    - *negative exponential :* e^{-d} *or* e^{-d^2}
    - *length of shared boundary*: w_{ij} = *length (i, j)/length(i)*

12

Statistical Tests to Examine Spatial Autocorrelation

Statistical Tests for presence of spatial autocorrelation

- **Global Tests**
  - ***Moran's I***
  - Geary's C
- **Local Tests**

  (LISA – Local Indicators of Spatial Autocorrelation)
  - ***Local Moran's I***
- **Other tests** that are more simple:
  - Chi-square Test
  - ***Join Count Statistic***

13

Global Moran's I

Product of the deviation from the mean for all pairs of adjacent regions (w_{ij}=1)

$$I = \frac{n\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{(\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij})\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

Sum of the weights (count of all adjacent pairs)

A measure of variance across the regions

Where,

n : the number of regions

$\bar{x}$: the mean of the variable

x_i : the variable value at a particular location i

w_{ij} : a weight indexing location of i relative to j

- Moran's I Typically ranges from -1 to 1
- Indices close to zero, indicate random pattern
- Indices toward +1 indicate a tendency toward clustering
- Indices toward -1 indicate a tendency toward dispersion

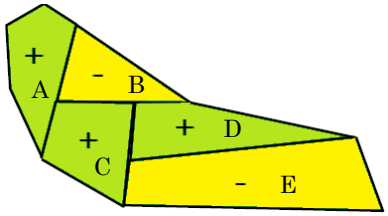
14

EXAMPLES

15

EXAMPLE: MORAN'S I

Research question: Is the areal pattern of + and - values randomly distributed amongst the polygons?



$$W = \{w_{ij}\} =$$

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

j rows

i columns

Polygon	Value
A	20
B	10
C	15
D	16
E	9
Mean	14
Std. Dev.	4.53

$$I_{(\text{Global})} = \frac{n\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{(\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij})\sum_{i=1}^{n}(x_i - \bar{x})^2} = -0.2806$$

I value is less than 0. Therefore, the areal pattern may be dispersed. Z-test is required.

For *i*=C, $I_{i\,(\text{Local})} = \frac{n(x_i - \bar{x})}{\sum_{j=1}^{n} W_{ij}\sum_{j=1}^{n}(x_i - \bar{x})^2}\sum_{j=1}^{n} w_{ij}(x_j - \bar{x}) = 0$

16

REFERENCES

- Getis, A. and Ord, J.K. (1992) *The analysis of spatial association by use of distance statistics* Geographical Analysis, *24(3) 189-206*
- Ord, J.K. and Getis A. (1995) Local Spatial Autocorrelation Statistics: distributional issues and an application Geographical Analysis, *27(4) 286-306*
- O'Sullivan and Unwin *Geographic Information Analysis* Wiley 2003
- Shekhar, Shashi, and Sanjay Chawla. Spatial databases: a tour. Vol. 2003. Upper Saddle River, NJ: prentice hall, 2003.

17

THANK YOU!!!

18