6.8 Surface analysis

In GIS, the term 'surface' refers to a scalar field, a variable represented as a function of location.

A surface represents elevation, temperature distribution, land price, etc.

Theoretically, a surface can be either continuous or discrete. We can easily imagine continuous surfaces, but what is discrete surface?

Actually, we frequently use discrete surfaces in GIS: numerical variables calculated on a tessellation, a set of spatial units such as administrative units and square cells. Typical examples include population density and land use ratios. The surface data based on a tessellation are often called 'area data'.

The term ‘area data’ is used for not only numerical but also categorical variables.

Strictly speaking, categorical area data are not a scalar field, so they are not surface data.
Nevertheless, categorical and numerical area data share the methods of analysis. Thus, in practice, it is convenient to treat both types of area data together.

In this section I will first explain analysis of surfaces, both continuous and discrete surfaces, the latter is numerical area data.

6.8.1 Trend surface analysis

Trend surface analysis is to represent a surface by a simple polynomial function so that we can easily understand its global structure.

A polynomial function is fitted to a surface and coefficients are used to analyze the original surface. Trend surface analysis is also used in (global) spatial smoothing.

Properties of trend surface analysis

Trend surface analysis describes the spatial structure of a surface by a small set of variables (coefficients).

It is useful when the spatial structure is simple, that is, a surface of a simple shape. On the other hand, trend surface analysis does not work when a surface has a very complicated shape, say, when a great local variations exists. In such a case it is difficult to interpret the result of analysis, meanings of the coefficients.
6. Spatial Analysis

6.8.2 Geostatistics

Geostatistics, which is used in spatial interpolation, is also useful for analyzing surfaces.

S: Study region of area A
f(x): Surface function defined in S

Variogram function

Variogram function evaluates the spatial variation of a surface, from which we can see how f(s) fluctuates in S:

\[ \gamma(h) = \frac{\int_{x \in S} \int_{t \in S \setminus x \setminus \delta} (f(x) - f(t))^2 \, dx}{2 \int_{x \in S} \int_{t \in S \setminus x} \, dx} \]

The variogram function is the average (square) difference of the surface value f(s) between two points of distance h.

\[ \gamma(h) = \frac{\int_{x \in S} \int_{t \in S \setminus x \setminus \delta} (f(x) - f(t))^2 \, dx}{2 \int_{x \in S} \int_{t \in S \setminus x} \, dx} \]

The variogram function is the average (square) difference of the surface value f(s) between two points of distance h.

The anisotropic variogram, covariogram, and correlogram are also useful in spatial analysis, because they summarize the spatial structure of surfaces by simple functions.

Topological method focuses on the topological, rather than metric structure of surfaces.

Topological method classifies all the points of surfaces into one of the four categories: peaks, pits, cols, and slopes.
At peaks, pits, and cols, a surface is even with the ground. The derivative of the surface function with respect to $x$ is zero, so we can easily find peaks, pits, and cols in GIS.

At slope points, on the other hand, the derivative is not zero, so the steepest descent direction can be defined at any slope point. Tracing the steepest descent vectors, we have a set of lines that connect peaks and pits. Among the lines we choose those passing through cols.

At each col exactly four lines meet, two are connected to peaks and the other to pits. The network that consists of extracted lines is useful in surface analysis because it summarizes the spatial structure of the original surface. It naturally defines the ‘neighborhood’ relationship between peaks, pits, and cols.
Moreover, we can use network analysis methods to analyze the surface structure. We can evaluate, for example, the connectivity and accessibility of the surface network.

6.9 Spatial autocorrelation: analysis of area data

Until now we have focused on the spatial aspect of spatial distributions - point distributions, spatial relationship between points and lines, network analysis, and so forth.

In this section we discuss both the spatial and aspatial (attribute) aspects of spatial objects simultaneously.

Tobler’s First Law of Geography

Waldo Tobler, a famous geographer in the United States, wrote “everything is related to everything else, but closer things are more closely related” in his paper Tobler (1970). This is called ‘Tobler’s First Law of Geography.’


Actually, we have seen several examples of Tobler’s Law. Typical variogram functions increase monotonically with $h$, the distance between two points. This indicates that closer locations have closer surface values, a concrete example of Tobler’s Law.

When we look at spatial and attribute data simultaneously, one of our objectives is to confirm Tobler’s Law. We ask ourselves “Do spatial objects located closely have similar attributes?”

Concept of spatial autocorrelation

The term ‘spatial autocorrelation’ refers to the correlation of locational and attribute similarities among spatial objects.
6. Spatial Analysis

Positive spatial autocorrelation:
Spatial objects that are located more closely have more similar attributes. In short, similar objects are closely located.

Negative spatial autocorrelation:
Spatial objects that are located more closely have more different attributes. Similar objects tend to avoid spatially with each other.

Both positive and negative spatial autocorrelations support Tobler’s First Law of Geography. The former suggests that a positive effect exists among spatial objects while the latter implies a negative effect.

When no autocorrelation exists, we think spatial interaction is not strong within the objects. Their attributes may be affected by other spatial objects.

Spatial tessellations

The concept of spatial autocorrelation typically appears in area data defined on a spatial tessellation.

A spatial tessellation represents a set of spatial units that extensively cover a certain region such as census tracts and administrative units.

A spatial tessellation often has area data, either numerical or categorical, the number of households, population density, land use category, etc..

References - Spatial autocorrelation

Among various variables, we should pay special attention to ‘density’ type variables. We calculate them by dividing a variable by the area of spatial units, so if we use different spatial units we obtain different area data.

‘Density’ type variables depend on the definition of spatial units, in other words, they are functions of spatial units. This is a kind of scale-related (frame-dependent) problem frequently found in spatial analysis, say, the quadrat method, choropleth maps, and so forth.

We can consider ‘spatial autocorrelation’ in not only numerical but also categorical area data on a spatial tessellation, because it only describes the correlation of locational and attribute similarities among spatial objects.

Moreover, we can even imagine spatial autocorrelation in variables defined on other types of spatial objects such as points, lines, and polygons.
6.9.1 Join count statistic

Join count statistic, in its simplest form, treats ‘black and white pictures’ type of spatial tessellation data, that is, binary variables defined on a square lattice.

We are interested in whether white and black cells are clustered within the same colors (positive spatial autocorrelation), or cells of the same color tend to avoid with each other (negative spatial autocorrelation).

Notation

\[ a_i \]: The number of cells adjacent to cell \( i \).

Adjacency is defined by the 4-connected neighborhood.

\[ A = \frac{1}{2} \sum_i a_i \]

\[ A \approx 40 \]

\[ BB \]: The number of adjacent black cells
\[ WW \]: The number of adjacent white cells
\[ BW \]: The number of adjacent pairs of white and black cells.

These variables are called ‘join count statistics’. They obviously satisfy

\[ BB + WW + BW = A \]
If cells of the same color are clustered, $BB$ and $WW$ become large while $BW$ is small.

Join count statistics describe the degree of spatial autocorrelation in a binary raster data. In practice, however, it is obviously redundant to use all the statistics because they are highly correlated with each other.

In general, we use only $BW$ as a measure of spatial autocorrelation.

Evaluation of $BW$

Significance test of an observed $BW$ statistic, whether it is significantly large (small), requires the probability distribution of the statistic $BW$ under complete randomness.

We compare the observed $BW$ with its probability distribution under complete randomness, and accept either the null or alternative hypothesis.

Free sampling

Null hypothesis $H_0$:
Cells are assigned either $B$ or $W$ independently.

Alternative hypothesis $H_1$:
Assignment of cell color is spatially correlated.

In point pattern analysis, the null hypothesis considers points under CSR (Complete Spatial Randomness), that is, a homogeneous Poisson distribution.

In spatial autocorrelation analysis, there are at least two ways of defining spatial randomness: free sampling and nonfree sampling (randomization). Consequently, there are two types of null hypotheses.

In the null hypothesis, $B$ ($W$) is assigned to a cell with the probability $p_B$ ($p_W$), which is given by the ratio of black (white) cells:

$$p_B = \frac{n_B}{n} \quad p_W = \frac{n_W}{n}$$

$n_B$: The number of black cells
$n_W$: The number of white cells
$n$: The total number of cells ($n_B + n_W = n$)
6. Spatial Analysis

Obviously, \( p_B + p_W = 1 \).

In free sampling, the number of black (white) cells is not always equal to \( n_B \) (\( n_W \)), because \( B \) or \( W \) is assigned independently to each cell. Someone think the null hypothesis of free sampling scheme is not realistic.

**Nonfree sampling (randomization)**

Null hypothesis \( H_0 \):
Cells are assigned either \( B \) or \( W \) with the constraint that the numbers of black and white cells are equal to \( n_B \) and \( n_W \), respectively.

Alternative hypothesis \( H_1 \):
Assignment of cell color is spatially correlated.

In nonfree sampling, assignment of cell color is not independent even under the null hypothesis. The null hypothesis assumes that the cell attribute (\( B \) or \( W \)) is randomized among cells in the observed data, so it is often called 'randomization'.

Nonfree sampling assures that the number of black (white) cells is always equal to \( n_B \) (\( n_W \)).

**BW under free sampling**

The probability distribution of \( BW \) under free sampling is approximated by a normal distribution \( N(\mu, \sigma^2) \).

We define two variables \( A_1 \) and \( A_2 \):

\[
A_1 = \frac{1}{2} \sum a_i
\]

\[
A_2 = \frac{1}{2} \sum a_i^2
\]

The mean and variance of the probability distribution is then given by

\[
\mu = 2A_1 p_B p_W
\]
\[
\sigma^2 = 2A_1 p_B p_W - 4(A_1^2 - A_1^2 + 2A_1) p_B^2 p_W^2
\]

**BW under nonfree sampling**

The probability distribution of \( BW \) under nonfree sampling is also approximated by a normal distribution.
6. Spatial Analysis

Limitations of the join count statistic

The (ordinary) join count statistic has the following limitations.

1. Adjacent cells are defined only by the 4-connected neighborhood.
2. Spatial units on which a variable is defined are limited to square lattices.
3. A variable has to be binary, that is, it can take only one of two values.

6.9.2 Extensions of join count statistic

The first and second limitations can be overcome by changing the definition of adjacent cells. Join count statistic can be extended to any spatial tessellation.

Let us remember the definition of $BW$ and its probability distributions under the null hypothesis.

The statistic $BW$ is the number of adjacent black-white cells, and its probability distributions are represented as functions of only $a_i$, $n_B$, and $n_W$ ($a_i$ is the number of cells adjacent to cell $i$).

This indicates that join count statistic can be applied to any spatial structure only if spatial adjacency relationship is clearly defined so that we can count $BW$ pairs and calculate $a_i$.

Therefore, to overcome the first limitation of the ordinary join count statistic, we have only to define adjacent cells by the 8-connected neighborhood.

In this case interior cells have eight adjacent cells, while boundary cells have three or five adjacent cells.
The second limitation can also be overcome in a similar way. We count the number of polygons adjacent to each unit and count $BW$ pairs for adjacent polygons.

Actually, ‘adjacent polygons’ do not have to be really spatially adjacent in a tessellation. We can arbitrarily define the adjacency relationship between polygons, independently of their spatial relationship. We may regard two polygons as adjacent if they are connected by a rapid transit.

To overcome the third limitation of ordinary join count statistic, that is, it treats only binary variables, we first consider a categorical variable of more than two classes. This type of variable is often called a multicolor map, in contrast to ‘black and white pictures’.

**Notation**

Suppose a categorical variable that can take one of $k$ values (classes).

- $n$: Total number of polygons
- $n_i$: The number of class $i$ polygons
- $p_i$: The ratio of class $i$ polygons ($=n_i/n$)
Instead of BW, we count the number of adjacent polygons of classes i and j for every pair of (i, j). 

\( N_{ij} \): The number of adjacent polygons of classes i and j.

If \( i=j \), \( N_{ii} \) corresponds to BB and WW, while \( N_{ij} \) for \( i \neq j \) is equivalent to BW. Consequently, small \( N_{ii} \) for \( i \neq j \) implies positive spatial autocorrelation.

Evaluation of \( N_{ij} \)

Significance test of an observed \( N_{ij} \) requires the probability distribution of the statistic \( N_{ij} \) under complete randomness.

As well as ordinary join count statistic BW, \( N_{ij} \) can be evaluated from two different viewpoints, that is, free sampling and nonfree sampling.

Free sampling

Null hypothesis \( H_0 \):
One of the \( k \) classes is independently assigned to each polygon. Class i is assigned to a polygon with the probability \( p_i \).

Alternative hypothesis \( H_1 \):
Class assignment is spatially correlated.

Nonfree sampling (randomization)

Null hypothesis \( H_0 \):
Cells are assigned One of the \( k \) classes with the constraint that the number of class i polygons is equal to \( n_i \) for any \( i \).

Alternative hypothesis \( H_1 \):
Class assignment is spatially correlated.

The probability distribution of \( N_{ij} \) under free sampling is approximated by a normal distribution \( N(\mu, \sigma^2) \).

If \( i=j \),
\[
\mu = A_i p_i \]
\[
\sigma^2 = A_i p_i^2 + 2(A_i - A) p_i^2 - (A_i^2 + 2A_i - A) p_i^4
\]

If \( i \neq j \),
\[
\mu = 2A_i p_i p_j
\]
\[
\sigma^2 = 2A_i p_i p_j + 2(A_i - A) p_i p_j \left(p_i + p_j\right) - 4\left(A_i^2 + 2A_i - A\right) p_i^2 p_j^2
\]

The probability distribution of \( N_{ij} \) under nonfree sampling is also approximated by a normal distribution \( N(\mu, \sigma^2) \).

If \( i=j \),
\[
\mu = \frac{A_i(n_i - 1)}{n(n-1)}
\]
\[
\sigma^2 = \frac{A_i(n_i - 1)}{n(n-1)} \left( \frac{2(A_i - A)n_i(n_i - 1)(n_i - 2)}{n(n-1)(n-2)} \right) - \frac{\{A_i^2 + 2A_i \}n_i(n_i - 1)(n_i - 2)(n_i - 3)}{n(n-1)(n-2)(n-3)} - 2 \left[ \frac{A_i(n_i - 1)}{n(n-1)} \right]^2
\]

If \( i \neq j \),
\[
\mu = 2A_i p_i p_j
\]
\[
\sigma^2 = 2A_i p_i p_j + 2(A_i - A) p_i p_j \left(p_i + p_j\right) - 4\left(A_i^2 + 2A_i - A\right) p_i^2 p_j^2
\]
6. Spatial Analysis

If \( i \neq j \),

\[ \mu = \frac{2 \sum_{i \neq j} a_{ij} n_{i} n_{j}}{n(n-1)} \]

\[ \sigma^2 = \frac{A(n) - A(n-1)}{n(n-1)(n-2)(n-3)} \]

\[ + \frac{4k(k+1)(n+1)(n+2)}{n(n-1)(n-2)(n-3)} \]

\[ A(n) = \sum_{i} a_{ii} n_i^2 \]

\[ A(n-1) = \sum_{i} a_{ii} n_i (n_i - 1) \]

6.9.3 Moran’s I statistic

Moran’s I statistic is an extension of the join count statistic developed by Moran (1948).


Geary’s C

Geary’s C statistic, which is also an extension of the join count statistic, was developed in parallel with Moran’s I.


Moran’s I and Geary’s C statistics are essentially equivalent. They are interchangeable, so only Moran’s I statistic is explained in detail in this lecture.

Advantages of Moran’s I statistic

Join count statistic has several limitations, but some of them can be easily overcome as discussed earlier.

Nevertheless, there are still two problems to be solved.

1. Numerical variables other than binary variables cannot be treated.

2. Spatial relationship between spatial objects is limited to adjacent and non-adjacent. Metric measures such as the distance between objects are not considered. Because of this, join count statistic can treat only spatial tessellations.

Moran’s I statistic resolves both of the two problems.
1. Moran’s $I$ statistic can treat metric variables other than binary variables, that is, ordinal, interval, and ratio scale variables.

2. The relationship between spatial objects can be determined flexibly by the user. For example, we can incorporate the distance between spatial objects into account. This makes it possible to treat various types of spatial objects other than spatial tessellations.

Variables used in Moran’s $I$ statistic

- $n$: Number of spatial objects (points, polygons, etc.)
- $x_i$: Attribute value of object $i$ (numerical variable)
- $a_{ij}$: Accessibility between objects $i$ and $j$

Definition of accessibility between objects

Accessibility is a measure that indicates how accessible a spatial object is from another spatial object. It is thus a concept opposite to adjacency used in join count statistic.

Accessibility is typically defined as a decreasing function of the distance between spatial objects, such as the inverse and negative exponential functions.

Various distances

The distance is defined by, for example,

1. Euclidean distance between centroids,
2. square Euclidean distance between centroids,
3. minimum distance between objects,
4. average distance between objects,
5. binary variable representing adjacency of objects, and
6. length of common boundary between polygons.
6. Spatial Analysis

Definition of Moran's I and Geary's C statistics

Once we define the accessibility between objects \( i \) and \( j \) for all \((i, j)\) pairs, we can calculate Moran’s \( I \) statistic:

\[
I = \frac{n}{\sum_{i} \sum_{j \neq i} a_{ij}} \sum_{i} \sum_{j \neq i} a_{ij} (x_i - \bar{x})(x_j - \bar{x})
\]

Geary’s \( C \) statistic is also defined by using \( n, x_i, \) and \( a_{ij} \):

\[
C = \frac{n-1}{2 \sum_{i} \sum_{j \neq i} a_{ij}} \sum_{i} \sum_{j \neq i} a_{ij} (x_i - \bar{x})^2
\]

Moran’s \( I \) statistic shows a large positive value if a strong positive spatial autocorrelation exists. Moran’s \( I \) statistic becomes negative if spatial autocorrelation is negative.

The domain of Moran’s \( I \) statistic is

\[-1 \leq I \leq 1\]

Moran’s \( I \) statistic is quite similar to the Pearson’s correlation coefficient, so its interpretation is easy: a value close to 1 indicates positive spatial autocorrelation, while a value close to –1 indicates negative spatial autocorrelation.

Significance test

To test the significance of an observed Moran’s \( I \) statistic, we compare it with its probability distribution under complete spatial randomness. As seen in join count statistic, there can be several ways of defining complete spatial randomness. In Moran’s \( I \) statistic, we consider two definitions of complete spatial randomness: independent normality and random permutation. They provide two different null hypotheses.
Independent normality

Null hypothesis $H_0$:
Attribute value of each object independently follows an identical normal distribution.

Alternative hypothesis $H_1$:
Attribute values are spatially correlated.

The null hypothesis corresponds to that considered in join count statistic.

The mean and variance of the normal distribution assumed under the null hypothesis are estimated from observed data.

\[
\hat{\mu} = \frac{1}{n} \sum x_i \\
\hat{\sigma}^2 = \frac{1}{n-1} \sum (x_i - \bar{x})^2
\]

Moran’s $I$ statistic under independent normality

The probability distribution of Moran’s $I$ statistic under independent normality is approximated by a normal distribution $N(\mu, \sigma^2)$:

\[
\mu = -\frac{1}{n-1} \\
\sigma^2 = \frac{n^2 S_1 - n S_2 + 3S_0^2}{(n-1)(n+1)S_0^2} - \frac{1}{n-1}
\]

Random permutation

Null hypothesis $H_0$:
Each object is randomly assigned one of the observed attribute values without duplication.

Alternative hypothesis $H_1$:
Assignment of attribute value is spatially correlated.

Random permutation is equivalent to nonfree sampling in join count statistic. It considers all the permutations of attribute values, and evaluate the significance of an observed Moran’s $I$ statistic.
6. Spatial Analysis

Moran’s I under random permutation

The probability distribution of the Moran’s I under random permutation is also approximated by a normal distribution \( N(\mu, \sigma^2) \):

\[
\begin{align*}
\mu &= -\frac{1}{n-1} \\
\sigma^2 &= \frac{n\{n^2-3n+3\}S_1-nS_2+3S_3}{(n-1)(n-2)(n-3)S_1} \\
&\quad - \frac{1}{(n-1)^2}
\end{align*}
\]

where

\[
K = \frac{\sum (x_i - \bar{x})^4}{\left( \sum (x_i - \bar{x})^2 \right)^2}
\]

Note

Moran’s I statistic heavily depends on how we define accessibility between spatial objects represented by \( a_{ij} \).

Different definition yields different I value and consequently may lead to different statistical conclusions. This is a kind of scale-dependency problem, which is also found in quadrat method, nearest neighbor distance method, etc., because the definition of accessibility reflects the scale of analysis.

Generalized Moran’s I

Moran’s I statistic can be generalized as follows.

The original Moran’s I statistic is given by

\[
I = \frac{\sum_{i \neq j} a_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum_{i} (x_i - \bar{x})^2}
\]

We first replace \( x_i \) with \( x_i - \bar{x} \):

\[
I = \frac{\sum_{i \neq j} a_{ij} x_i x_j}{\sum_{i} a_{ij} x_i^2}
\]

In random permutation test of an observed Moran’s I statistic, the number of spatial objects \( n \) and the denominator of the right hand are constant. Therefore, we can remove \( n \) and the denominator of the right hand without loss of generality. Moran’s I statistic then becomes

\[
I = \sum_{i \neq j} a_{ij} x_i x_j
\]
This statistic is sometimes called the generalized Moran's \( I \) statistic. From this statistic we can define various statistics; some spatiotemporal statistics are equivalent to the generalized Moran's \( I \) statistic.

Any statistic based on the generalized Moran's \( I \) statistic can be statistically analyzed in a way similar to that of Moran's \( I \) statistic. The probability distribution under the null hypothesis (random permutation) is approximated by a normal distribution.