Local Spatial Autocorrelation

Global AC: the extent to which points that are “close together” in space have similar values, on average.

Local AC: the extent to which points that are “close” to a given point have similar values.

LISA: “Local Indicator of Spatial Association” (Anselin 1995), *sensu stricto.*

1. The LISA for each point in space gives an indication of significant spatial clustering of similar or dissimilar values around the point. Significance tests should be possible.

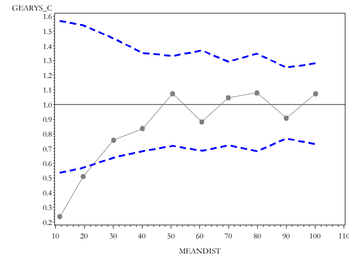
2. The sum of LISAs for all points in a given study area is proportional to a corresponding GLOBAL indicator of spatial association for that area.

Two LISAs:

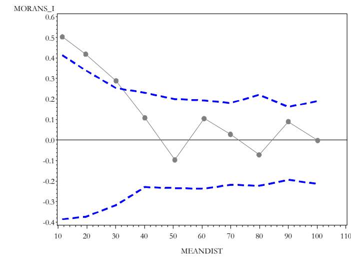
-Local Moran’s I

-Local Geary’s C

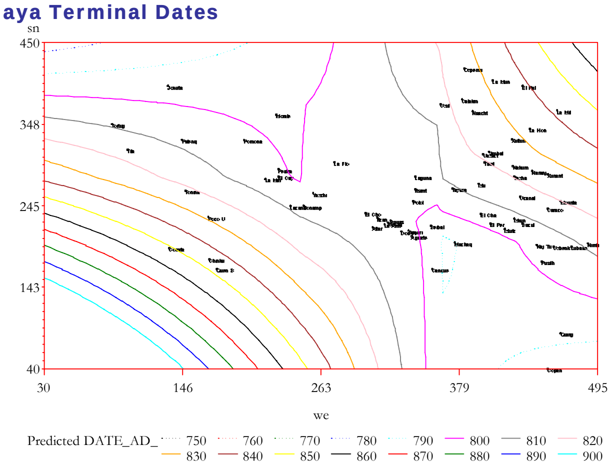
But don’t forget Getis-Ord G!

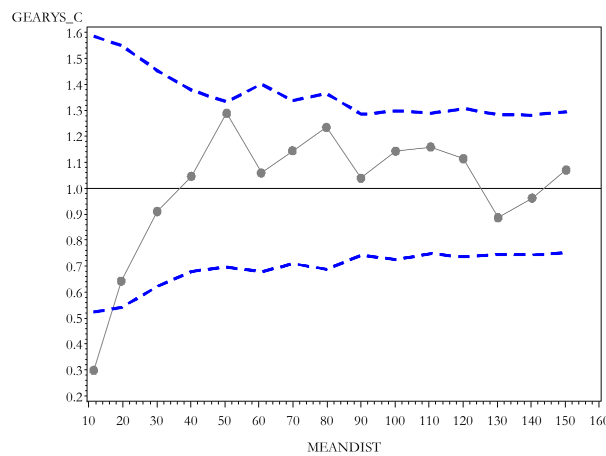


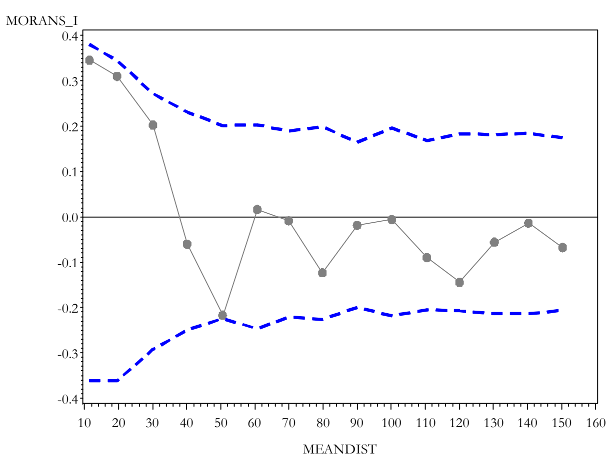
Maya Terminal Dates



Maya Terminal Dates







Moran’s I: Global and Local

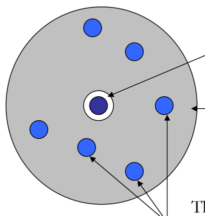
$$I = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}\left(z_i - \bar{z}\right)\left(z_j - \bar{z}\right)}{s_z^2 \sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}}$$

Our old friend – the global version: one estimate of *I* for each spatial lag represented in the ***W*** matrix

$$I_i = \frac{\sum_{j=1}^{n} w_{ij}(z_i - \bar{z})(z_j - \bar{z})}{s_z^{\,2} \sum_{j=1}^{n} w_{ij}}$$

The local version: one estimate of I For each point in the dataset – note the *i* subscript – AND each spatial lag represented in the **W** matrix.

Local Moran's I *

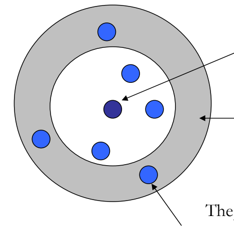


$$I_i = \frac{\sum_{j=1}^{n} w_j (z_i - \bar{z})(z_j - \bar{z})}{s_z^2 \sum_{j=1}^{n} w_{ij}} \rightarrow \begin{cases} (z_i - \bar{z})(z_1 - \bar{z}) + \\ (z_i - \bar{z})(z_2 - \bar{z}) + \\ (z_i - \bar{z})(z_3 - \bar{z}) + \\ (z_i - \bar{z})(z_4 - \bar{z}) + \\ (z_i - \bar{z})(z_5 - \bar{z}) + \\ \ldots + \\ (z_i - \bar{z})(z_j - \bar{z}) \end{cases}$$

*Anselin 1995

Local Moran's I

Neighborhoods that are NOT adjacent to the i'th point are possible – *and* analytically useful. But most software (*e.g.* Geoda, ArcGIS) is not set up to use them.



$$I_i = \frac{\sum_{j=1}^{n} w_j (z_i - \bar{z})(z_j - \bar{z})}{s_z^2 \sum_{j=1}^{n} w_{ij}} \rightarrow \begin{cases} (z_i - \bar{z})(z_1 - \bar{z}) + \\ (z_i - \bar{z})(z_2 - \bar{z}) + \\ (z_i - \bar{z})(z_3 - \bar{z}) + \end{cases}$$

Local Moran's I

$$I_i = \frac{\sum_{j=1}^{n} w_{ij} (z_i - \bar{z})(z_j - \bar{z})}{s_z^2 \sum_{j=1}^{n} w_{ij}} = (z_i - \bar{z}) \sum_{j=1}^{n} w_{ij} (z_j - \bar{z})$$

$$(z_i - \bar{z}) \frac{\sum_{j=1}^{n} w_{ij} (z_j - \bar{z})}{\sum_{j=1}^{n} w_{ij}}$$

The deviation of the i'th point.

The *mean* of the deviation of the j'th points in the neighborhood of the i'th point.

There are 4 types of values for I:

z_i	mean of z_j
H	H (positive)
L	L
L	H (negative)
H	L

Local Moran's I

Use the randomization distribution to test null hypothesis of no local autocorrelation (H0).

- -asymptotic results for the sampling distribution under H0 are bogus.
- -randomization p-values require Bonferroni-esque adjustment.
- -strict Bonferroni (p' = p/n) is probably too conservative.

Values of the statistics are poorly correlated to *p*-values from randomization.

- -do not interpret statistic values w/o p-values. (*c.f.* Premo 2004).

$$I = \frac{\sum_{i=1}^{n} w_i I_i}{\sum_{i=1}^{n} w_i}$$

where w_i is the number of points in the neighborhood of the i'th point. A LISA , *sensu stricto.*

Geary's C: Global and Local

$$c = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} (z_i - z_j)^2}{s_z^2 2 \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}}$$

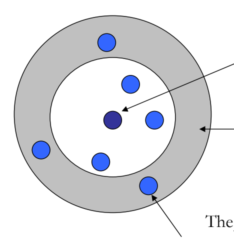
Our old friend – the global version: one estimate of I for each spatial lag represented in the $\boldsymbol{W}$ matrix

$$c_i = \frac{\sum_{j=1}^{n} w_{ij} (z_i - z_j)^2}{s_z^2 2 \sum_{j=1}^{n} w_{ij}}$$

The local version: one estimate of I For each point in the dataset – note the i subscript – AND each spatial lag represented in the **W** matrix.

Local Geary's C

Neighborhoods that are NOT adjacent to the i'th point are possible – *and* analytically useful. But most software (*e.g.* Geoda, ArcGIS) is not set up to use them.



$$c_i = \frac{\sum_{j=1}^{n} w_{ij} (z_i - z_j)^2}{s_z^2 2 \sum_{j=1}^{n} w_{ij}} \rightarrow \begin{cases} (z_i - z_1)^2 + \\ (z_i - z_2)^2 + \\ (z_i - z_3)^2 \end{cases}$$

Local Geary's C

Two types of values for *C*:

z_i	mean of z_i
H	L
L	H

Use the randomization distribution to test null hypothesis of no local autocorrelation (H0).

-asymptotic results for the sampling distribution under H0 are bogus.
-randomization p-values require Bonferroni-esque adjustment.
-strict Bonferroni (p' = p/n) is probably too conservative.

Values of the statistics are poorly correlated to *p*-values from randomization.

-do NOT interpret statistic values w/o p-values.

$$C = \frac{\sum_{i=1}^{n} c_i I_i}{\sum_{i=1}^{n} w_i}$$

where w_i is the number of points in the neighborhood of the *i*'th point.
A LISA , *sensu stricto.*

Getis-Ord G

$$G_i = \frac{\sum_{\substack{j=1 \\ i \neq j}}^{n} w_{ij} z_j}{\sum_{\substack{j=1 \\ i \neq j}}^{n} z_j}$$

The guts of *G*: a ratio of the total of the values in a spatial neighborhood to the global total.

For G_i, the *i*'th point is NOT include in the calculation.

If *i*'th point is included, *G* becomes *G**.

This is not intuitive to interpret, so why not express it as a z-score (which is!)?

A z-score??

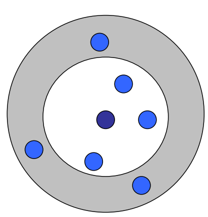
Getis-Ord G

$$z = \frac{x - E\{x\}}{VAR\{X\}^{1/2}}$$ or $$z = \frac{x - \bar{x}}{s}$$

$$E\{G_i\} = \sum_{\substack{j=1 \\ i \neq j}}^{n} w_{ij} \bar{z}_i$$

$$VAR\{G_i\}^{1/2} = \left(\frac{s_i (n-1) \sum_{\substack{j=1 \\ i \neq j}}^{n} w_{ij}^{\,2} - (\sum_{\substack{j=1 \\ i \neq j}}^{n} w_{ij})^2}{n-2} \right)^{1/2}$$

Getis-Ord G



Use to identify "hot spots" and "cold spots."
But remember spots may be doughnuts!

Use the randomization distribution to test null hypothesis of no local autocorrelation (H0).

-asymptotic results for the sampling distribution under H0 are bogus.
-randomization p-values require Bonferroni-esque adjustment.
-strict Bonferroni (p' = p/n) is probably too conservative.
-p-values are identical to whose for local Moran

Values of the statistics are correlated to *p*-values from randomization. So OK to map raw values.

-NOT a LISA, *sensu stricto.*

Why LISAs are cool!

They "decompose" a global results into their local parts.

1. A significant global statistic at a given spatial lag may hide large spatial patches of no autocorrelation – *IID* Gaussian noise. LISA can detect this AND show us the location of BOTH kinds of patches in space.

2. An insignificant global result may hide patches of autocorrelation.

They are NOT a panacea for non-stationarity.

They do not compromise the utility of thinking about spatial data in terms of trends, local-spatial dependence and noise.

Rather they allow us to push that basic model further and make it MORE useful by uncovering hidden, local patterns in data that the global statistics average over.

Powerful potential for (substantive) hypothesis testing.

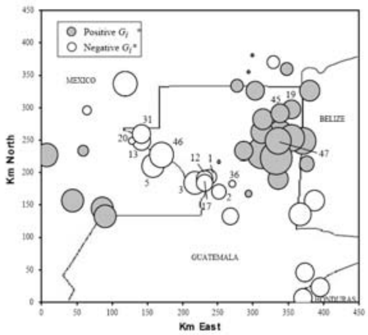


Fig. 3. Bubble graph of standardized G_i^* scores at a lag distance of 75 kilometers. Shaded bubbles represent positive G_i^* scores, white bubbles represent negative G_i^* scores, and bubble area is proportional to $|G_i^*|$. Sites discussed in the text are labeled with their site number from Table 1. Contemporary political boundaries appear in the background.

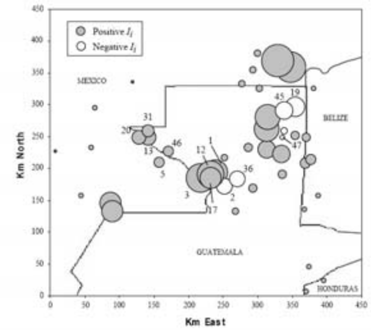


Fig. 2. Bubble graph of standardized local Moran's I_i scores at a lag distance of 75 kilometers. Shaded bubbles represent positive I_i scores, white bubbles represent negative I_i scores, and bubble area is proportional to $|I_i|$. Sites discussed in the text are labeled with their site number from Table 1. Contemporary political boundaries appear in the background.