From the Geo-dipole to the Geo-multipole: A Framework of Applying Multiple-Point (Geo)Statistics for Geographic Field Analysis

Rui Zhu

STKO Lab, Department of Geography
University of California, Santa Barbara

April, 2017; Boston

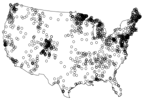
Outline

Outline

Spatial Patterns

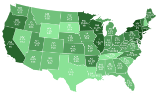
- Types of spatial patterns

Spatial Point Pattern



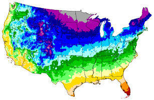
National Dams

Spatial Areal Pattern



State Tax Burdens

Spatial Field Pattern



National Mean Temperature

Spatial Network Pattern

National Highway System

- How to quantify spatial pattern?
  1. First-order or environmental effects: is the attribute mean stationary across the domain?
     - e.g., local intensity and local mean of the attribute value.
  2. **Second-order or interaction effects**: is there any correlation between locations?
     - e.g., spatial association

Second-order Interactions

How to model spatial interactions?

1. Single distance class:
   - Geary's C

$$C = \frac{(N-1)\sum_i \sum_j \mathbf{w_{ij}}(X_i - X_j)^2}{2\sum_i \sum_j \mathbf{w_{ij}} \sum_i (X_i - \bar{X})^2}$$

   - Moran's I

$$I = \frac{N}{\sum_i \sum_j \mathbf{w_{ij}}} \frac{\sum_i \sum_j \mathbf{w_{ij}}(X_i - \bar{X})(X_j - \bar{X})}{\sum_i (X_i - \bar{X})^2}$$

2. Multiple distance classes:
   - Semivariance

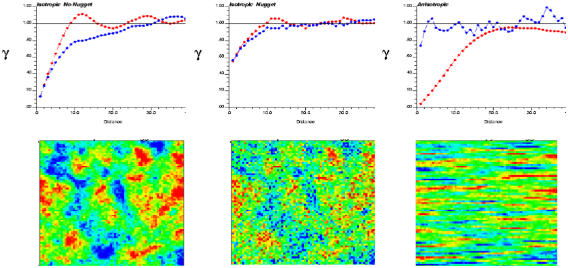
$$\hat{\gamma} = \frac{1}{2N(\mathbf{h})} \sum_{i=1}^{N(\mathbf{h})} [X(s_i + \mathbf{h}) - X(s_i)]^2$$

Sencod-order Interactions

Are these second-order statistics robust enough to analyze all spatial patterns?

- In some cases, they *are:*

- Blue line: north-south direction
- Red line: west-east direction

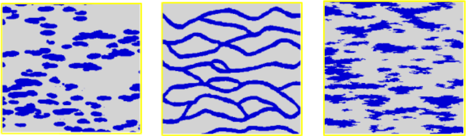


Second-order Interactions

Are these second-order statistics robust enough to analyze all spatial patterns?

- In some cases, they *are not:*

proportion of blue pixels = 0.28

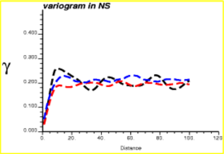


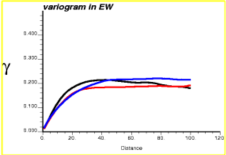
1 2 3

North-south Direction

Directional semivariograms

East-west Direction





Second-order Interactions

What are the reasons that make the second-order interaction (e.g.,semivariogram) fail?

1. The interaction between location A and location B may be dependent on a third, or even more nearby location(s) C.
2. Spatial patterns may sometimes have evident geometrical properties.

Therefore, models that could quantify **higher-order interactions** are needed!

Higher-order Interactions

- What it is **(conceptualization)**: interactions among multiple (i.e., more than two) locations are modeled *simultaneously*.
- How to model it:
    - Similar to variograms, use the sample data to model the spatial pattern. However, the size of sample data sets is typically small for such multiple-point inference;
    - **Multiple-Point (Geo)Statistics**: The *training image* is applied as an analogue, or more generally a prior model of the target pattern; then a *template* is used for modeling the higher-order interactions (statistics).

Outline

Geographic Conceptualization

Goodchild et al., (1999, 2007)'s general theory of geographic representation in GIS:

- **Geo-atom**:

$$< x, Z, z(x) >$$

where a spatial location x is associated with an attribute feature Z via the function mapping $z(x)$.

- **Geo-dipole**:

$$< x, x', Z, z(x, x') >$$

where the interaction of variables between two locations x and x' is described via the two-point mapping $z(x, x')$.

Geo-multipole

Following Goodchild's geo-dipole $< x, x', Z, z(x, x') >$, we propose a generalized conceptualization for modeling spatial interactions:

Geo-multipole: $\langle x, t_N, Z, z(x, t_N)\rangle$
where $t_N = \{x_1, .., x_N\}$ are the N neighbors of x.

Geo-multipole

Both the geo-atom and geo-dipole could be described by the geo-multipole:

- When t_N is empty, the geo-multipole equals to the geo-atom, which is a *single-point* data model;
- When $t_N = \{x'\}$, the geo-multipole equals to the geo-dipole, which is a *two-points* data model;
- The most generalized version of geo-multipole, where $t_N = \{x_1, .., x_N\}$, is a *multiple-points* data model.

Geo-dipole VS. Geo-multipole

The key difference between the geo-dipole and the geo-multipole:

- **Geo-dipole**: interactions are considered in *pairs* under the conceptualization of the geo-dipole despite that multiple pairwise interactions could be combined in *sequence*;
- **Geo-multipole**: locations $x_1, .., x_N$ in t_N are *simultaneously* considered (along with the corresponding attribute values) when modeling their interactions with x.

Mathematically, $f(z(x, x_1), ..., z(x, x_N))$ does not necessarily imply $z(x, t_n = \{x_1, ..., x_N\})$.

Statistical Perspectives

Geographic fields are frequently assumed to be generated from stochastic processes, and are thus regarded as realizations of a random field. Therefore, the generalized geo-multipole, with its two special cases, are translated to their statistical perspectives (i.e., modeled as probability density/mass functions).

- **Single-point data model** (i.e., geo-atom)
- **Two-points data model** (i.e., geo-dipole)
- **Multiple-points data model**

Statistical Perspective (Single-point Data Model)

Probability density function:

$$f(z, x) = prob(Z(x) = z \pm \epsilon)$$

- Statistics:
  - Continuous attributes: mean, standard deviation, quantiles ...
  - Categorical attributes: mode, the proportion of one specific category ...
- Examples:
  - Univariate ($f(z, x)$): could only be estimated by experts using physical models or experience;
  - Multivariate ($f(z, z', x) = prob(Z(x) = z|Z'(x))$): linear and non-linear models could be used for modeling the relation between z and z' that are co-located at location x.

Statistical Perspective (Two-points Data Model)

Probability density function:

$$f(z, x, Z(x')) = prob(Z(x) = z|Z(x'))$$

- Statistics:
  - Continuous attributes: semivariogram, cross-semivariogram ...
  - Categorical attributes: indicator semivariogram, transition probabilities ...
- Examples:
  - Univariate ($f(z, x, Z(x'))$): interpolation (e.g., IDW, Kriging) ...
  - Multivariate ($f(z, z', Z(x'), Z'(x)) = prob(Z(x) = z|Z(x'), Z'(x))$): contextual classification with multi-spectral remotely sensed images...

Statistical Perspective (Multiple-points Data Model)

Probability density function:

$$f(z, x, Z(t_N)) = prob(Z(x) = z|Z(t_N))$$

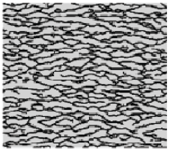
- Statistics: higher-order conditional probabilities ... $\rightarrow$ But, how do we model it?
- Examples: interpolation, classification and simulation...

Outline

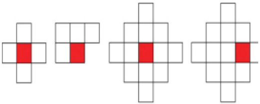
Building Blocks

- **Training images (*TI*)** that are assumed to contain the target spatial patterns;
- **Template (*T*)** for scanning training images;
- **Data events (*dev*(x))**, which are simultaneous (joint) combinations of attribute values at the template configuration.

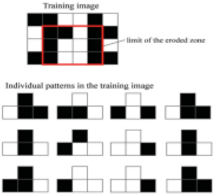
Training Image



Templates



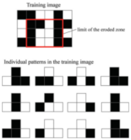
Data Events



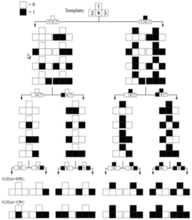
Algorithms of Multipole-Point (Geo)Statistics

- Extended Normal Equations Simulation (ENESIM): Guardiano and Srivastava (1993)
- Simple Normal Equation Simulation (SNESIM): Strebelle (2002)
- Pattern-based Simulation (SIMPAT): Arpat (2005)
- Filter-based Simulation (FILTERSIM): Zhang et al. (2006)
- Direct Sampling (DS): Mariethoz et al. (2010)
- Parallel Multiple-point algorithm using a list approach (IMPALA): Straubhaar et al. (2011)

Simple Normal Equation Simulation (SNESIM)



Step 1



Step 2

1. Scan the training image using the template;
2. Store the data events in a tree structure;
3. Estimate the conditional probability:

$$prob(z(x) = c_k | t_N) = \frac{\#(z(x)=c_k | t_N)}{\sum_{i=1}^{K} \#(z(x)=c_i | t_N)}$$

where c_k is the k^{th} class, and $t_N = \{x_1, .., x_N\}$ are the N neighbors of x

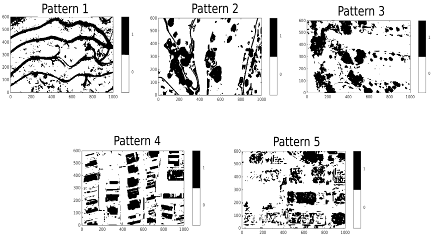
Outline

Case Study

Do we really need the geo-multipole to model real world geographic field information?

- Use two-point statistics and multiple-point statistics respectively to:
    - describe the spatial pattern;
    - simulate the spatial pattern.

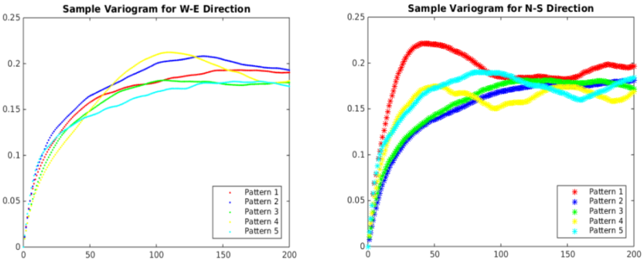
Remotely sensed images (1km)



Binary spatial patterns (600-by-1000)

Description of the Pattern

Variogram-based analysis:

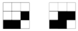
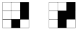
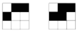


Directional semivariograms for the five patterns

Description of the Pattern

MPS-based analysis: A data event's conditional probability is calculated as the frequency of occurrence:

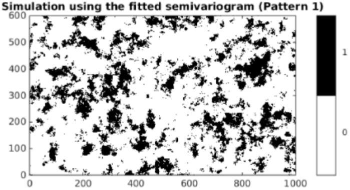
$$prob(z(x) = 0|t_N) = \frac{\#(z(x) = 0|t_N)}{\#(z(x) = 0|t_N) + \#(z(x) = 1|t_N)}$$

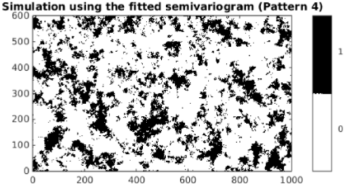
Templates	Data Events	P(x| n1, ... , n8)			
		Pattern 1		Pattern 4	
		P(x=0| n_1, ... , n_8)	P(x=1| n_1, ... , n_8)	P(x=0| n_1, ... , n_8)	P(x=1| n_1, ... , n_8)
$\begin{matrix} n_1 & n_2 & n_3 \\ n_4 & \times & n_5 \\ n_6 & n_7 & n_8 \end{matrix}$		0.000	**1.000**	**0.600**	0.400
		**0.600**	0.400	0.300	**0.700**
		0.500	0.500	**0.625**	0.375
		0.500	0.500	0.000	**1.000**
		**0.530**	0.470	0.458	**0.542**
		**0.610**	0.390	0.487	**0.513**

Conditional multiple-point probabilities for patterns 1 and 4

Simulation of the Pattern

Variogram-based simulation: Unconditional moving average simulation via the Fast Fourier Transform (FFT) was used; the resulting continuous images were then thresholded using suitable cutoff values so as to reproduce the same proportion of black pixels as the corresponding original binary images.

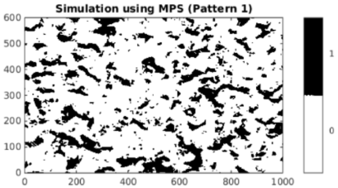


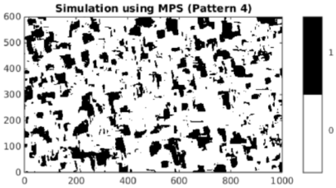


Variogram-based simulations (pattern 1 and pattern 4)

Simulation of the Pattern

MPS-based simulation: Unconditional SNESIM was used; template was set to 80×80 squares.





MPS-based simulations (pattern 1 and pattern 4)

Observations

- Multiple-point (geo)statistics (MPS) could better distinguish spatial patterns that are visually different compared to variograms;
- The two simulations using MPS show obviously different patterns with pattern 1 showing more curvilinearity along the west-east direction, and pattern 4 showing more polygonal geometries;
- The MPS-based simulations are more similar to the ones from the original images compared to variogram-based simulation.

Therefore, the geo-multipole, specifically the MPS, indeed plays a role in spatial analysis.

Conclusions

- Discussed motivations for higher-order interactions;
- Proposed a generalized geographic conceptualization, with higher-order interaction included;
- Introduced an algorithm to model the higher-order interactions;
- Showed the strength of higher-order interaction in describing and simulating spatial patterns using real world geographic field information.

Outline

Future Work

- The application details of multiple-point (geo)statistics for quantifying spatial patterns in geographic phenomena should be further explored;
- Contextual classification and spatial simulation using higher-order interactions could be studied and improved;
- Methods of modeling higher-order interactions for other geographic information could be investigated.

Acknowledgement

Special thanks to:
Dr. Krzysztof Janowicz, Dr. Phaedon C. Kyriakidis,
and my colleagues at STKO!

Zhu, R., Kyriakidis, P., and Janowicz, K. (2017). Beyond Pairs: Generalizing the Geo-dipole for Quantifying Spatial Patterns in Geographic Fields. AGILE 2017 Conference on Geographic Information Science.