Recent Advances in Software for Space-Time Data Analysis

Sergio J. Rey
GeoDa Center for Geospatial Analysis and Computation
School of Geographical Sciences and Urban Planning
Arizona State University

Future Directions in Spatial Demography
Santa Barbara, CA
December 12, 13  2011
Acknowledgments

• Economic Development Administration
• National Institutes of Health
• National Institute of Justice
• National Science Foundation
Outline

• Evolution of space-time analysis software
• PySAL: spatial dynamics
• Challenges
Evolution
Space-Time in GIScience

- Representation
- Data Modeling
- Geovisualization
- Spatialization
- Geostatistics
Space-Time Domains

- Tracking
- Change Detection
- Polygon Coverages
- Agent Based Models
- Cellular Automata
- Events

Goodchild, M.F. (2010) GISRUK Keynote
Space-Time Identified as Future Theme

- Dynamics of spatial clustering
- Clustering of temporal co-movements
- No specialized packages in existence
STARS
Space-Time Analysis of Regional Systems
Brushing and Linking
Space-Time Path and Time Traveling
Distributional Leap Frogging and Spatial Travel
Spatial Markov
History

- Spatial Analysis Laboratory (UIUC)
- Regional Analysis Laboratory (SDSU)
- STARS
- GeoDa
- Various other projects
Uses of PySAL

- Platform agnostic
- Shell
- Desktop Applications
  - GeoDaSpace
  - STARS
- Plug-ins (ArcGIS, QGIS)
- Distributed services, Web apps
Pedagogic Goals

- Code as text
- no black boxes
- replicability
- Extensive documentation
  - tutorials/API
  - cultural shift
Performance: Weights Creation

Computation time to create lattice (queen) weights

- R
- Spreg

Minutes

N

0 500000 1000000 1500000 2000000 2500000 3000000
ESDA

• Measures of spatial autocorrelation
  • Moran’s I, Geary’s c, join counts

• Map Classification
  • Natural breaks, Fisher Jenks, equal interval, more

• Rate smoothing
  • Empirical Bayes, age adjusted, excess risk, more
Inequality

• Theil Index
  • Entropy based measure of spatial inequality
  • Regional decompositions
    • Interregional inequality
    • Intraregional inequality
Regionalization

- max-p (Duque, Anselin, Rey 2012)
  - Given n areas, form the maximum number (p) of regions respecting contiguity and threshold constraints

- Random Regions
  - Randomly construct regions given various constraints
Spatial Dynamics

- Markov transition matrices
  - Classic, spatial, LISA
- Space-time interaction tests (1.2)
  - Knox, Mantel, Jacquez
- Space-time Rank mobility tests
- Space-time LISA
- Directional LISA
  - (Rey, Murray, Anselin 2011)
Directional LISA
Directional Moran Scatter
Origin Standardized
<table>
<thead>
<tr>
<th>Segment</th>
<th>Count</th>
<th>Expected</th>
<th>s</th>
<th>z</th>
<th>p-norm</th>
<th>p-rand</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>19</td>
<td>18.157</td>
<td>2.356</td>
<td>0.358</td>
<td>0.360</td>
<td>0.432</td>
</tr>
<tr>
<td>2</td>
<td>13</td>
<td>9.141</td>
<td>1.940</td>
<td>1.989</td>
<td>0.023</td>
<td>0.041</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>4.587</td>
<td>1.412</td>
<td>-1.124</td>
<td>0.131</td>
<td>0.233</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>6.947</td>
<td>1.720</td>
<td>-2.876</td>
<td>0.002</td>
<td>0.010</td>
</tr>
<tr>
<td>5</td>
<td>7</td>
<td>1.924</td>
<td>1.467</td>
<td>3.460</td>
<td>0.000</td>
<td>0.005</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>0.543</td>
<td>0.720</td>
<td>2.024</td>
<td>0.021</td>
<td>0.092</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>1.223</td>
<td>1.019</td>
<td>-0.219</td>
<td>0.413</td>
<td>0.638</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>5.478</td>
<td>2.060</td>
<td>-2.174</td>
<td>0.015</td>
<td>0.013</td>
</tr>
</tbody>
</table>

Table 1: Conditional randomization tests of directionality
Bivariate LISA
Bivariate LISA

Space-Time Correlation

We can think of the plot on the left above as an example of **inward diffusion** (the neighbors now, in July 1987, to the core in the future, August 1987).

We can switch the roles of the variables on each axis and compare the neighbors in the future to the value now. This is called **outward diffusion**. Note that the value for Moran’s I is now 0.4192.

Hallahan, C. (2009) SIGSTAT
Bivariate LISA

- Consistent with diffusion
  - inward
  - outward
- Also consistent with stable spatial autocorrelation
- Does not distinguish between
  - apparent diffusion/contagion
  - true diffusion/contagion
LISA Markov
LISA Markov

- LISA = Local Indicator of Spatial Association (Anselin, 1995)

- LISA Markov (Rey and Janikas 2006)
LISA Markov

- 4 states for the chain: HH, LH, LL, HL
- 16 possible transitions over one time interval
- characterize spatial dynamics
  - diffusion/contagion
  - directionality
<table>
<thead>
<tr>
<th>$t_1$</th>
<th>$t_2$</th>
<th>Dynamics</th>
</tr>
</thead>
<tbody>
<tr>
<td>HH</td>
<td>HH</td>
<td>stability</td>
</tr>
<tr>
<td>HH</td>
<td>LH</td>
<td>own suppression</td>
</tr>
<tr>
<td>HH</td>
<td>LL</td>
<td>concurrent suppression</td>
</tr>
<tr>
<td>HH</td>
<td>HL</td>
<td>other suppression</td>
</tr>
<tr>
<td>LH</td>
<td>HH</td>
<td>inwards contagion</td>
</tr>
<tr>
<td>LH</td>
<td>LH</td>
<td>stability</td>
</tr>
<tr>
<td>LH</td>
<td>LL</td>
<td>outwards suppression</td>
</tr>
<tr>
<td>LH</td>
<td>HL</td>
<td>inwards displacement</td>
</tr>
<tr>
<td>LL</td>
<td>HH</td>
<td>concurrent increase</td>
</tr>
<tr>
<td>LL</td>
<td>LH</td>
<td>potential inwards</td>
</tr>
<tr>
<td>LL</td>
<td>LL</td>
<td>stability</td>
</tr>
<tr>
<td>LL</td>
<td>HL</td>
<td>potential outwards</td>
</tr>
<tr>
<td>HL</td>
<td>HH</td>
<td>outwards contagion</td>
</tr>
<tr>
<td>HL</td>
<td>LH</td>
<td>outwards displacement</td>
</tr>
<tr>
<td>HL</td>
<td>LL</td>
<td>inwards suppression</td>
</tr>
<tr>
<td>HL</td>
<td>HL</td>
<td>stability</td>
</tr>
</tbody>
</table>
Crime Analytics for Space-Time
Cluster Transition Categories

<table>
<thead>
<tr>
<th>Before Transition</th>
<th>After Transition</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

Options for 1.
- Hotspot
- Coldspot
- Hot-Cold
- Cold-Hot
- Not Significant

Options for 2.
- Hotspot
- Coldspot
- Hot-Cold
- Cold-Hot
- Not Significant
Challenges
MAUP in Space-Time

• Most (all?) MAUP attention on cross-sectional case
  • Aggregation and zoning components

• In space-time: more complex
  • appearance of new counties
  • annexations
  • split/merging of census tracts
Responses

- Longitudinal Studies (common)
  - areal interpolation to time-consistent and exogenous boundaries

- Endogenous boundaries (future)
  - space no longer exogenous container
  - predicting tract splits/merger
  - predicting redistricting
  - predicting state formation
Software

• CyberInfrastructure
  • Enormous potential in hpc/parallelization
  • Substantial refactoring required

• GUI - Putty/Clay
  • Need for extensible/flexible toolkits
  • New methods will likely be required
  • Scientist as producer rather than consumer
PySAL

• Next release: **Jan 31, 2012 (1.3)**

• Google Code
  • Feature requests
  • Bug reports

• Get involved
  • Feature requests - what would spatial demographers want/need?
http://pysal.org

geodacenter.asu.edu
twitter.com/GeoDaCenter
www.facebook.com/geodacenter
PySAL 1.3+
Python 3.x

• Changes
  • Python 2.x series ends with Python 2.7
  • Many backward incompatible changes in Python 3.x
  • New Python functionality only in 3.x

• PySAL was written for Python 2.x
  • Tests of current PySAL code base shows broad compatibility with Python 3.x
Parallel PySAL

• Integration with CyberGIS project
  • plisa

• Focus at ASU
  • Examining parallel mechanisms in Python
  • Mapping of PySAL spatial analytical components to alternative parallel mechanisms
  • Multiple implementations for delivery
Contrib Module

- New in 1.3
- Leverage third party libraries
- Avoid core dependencies
- Libraries
  - Shapely
  - proj4
```python
>>> import pysal
>>> 
>>> w = pysal.open("stl.gal").read()
>>> data = pysal.open("stl_hom.csv", "r")
>>> y = np.array(data.by_col["HR8893"])

>>> mi = pysal.esda.Moran(y, w)

>>> mi.I
0.24366

>>> mi.p_norm
0.00027147862770937614
```