Enhance Analysis of SaTScan Results with Geovisual Analytics

Jin Chen¹, Alan M. MacEachren ¹, Eugene Lengerich ²

¹GeoVISTA Center, Geography Dept, Penn State University
²Department of Health and Evaluation Sciences, Penn State College of Medicine
Outline

• Introduction of Kulldorff’s spatial scan statistic (SaTScan)
• Limitations with SaTScan
• Geovisual analytics enhances analysis of SaTScan results
Scan Statistics

- Detect a local excess or deficiency of events (e.g. death rate due to a disease).
- Employ a moving “window”, collect cases least consistent with null hypothesis (e.g. constant risk of a disease). The cases are most likely clusters.
Kulldorff’s spatial scan static

- Introduced by Kulldorff (1997), known as SaTScan
- Similar to GAM, with advantages
  - deterministic - it reports the location and size of clusters
  - inferential - it evaluate statistical significance of the clusters.
- A result reported by SaTScan

Location IDs included.: 12087, 12086, 12021, ... (i.e. FIPS code)
Coordinates / radius..: (33.957701 N, 91.732284 W) / 377.70 km
Population............: 24785853
Number of cases.......: 1047
Expected cases........: 678.20
Observed / expected.: 1.544

... 

P-value...............: 0.001

The method (and the software) is widely used in epidemiology, crime analysis, etc.
Limitations of SaTScan

- Lack of cartographic output for displaying clusters
- Produces less usable clusters when inappropriately choosing parameters (i.e., heterogeneous contents)
- Sensitive to user controlled parameter choices
  - e.g., the maximum circle size, defined as the percentage of total population at risk.
  - If a circle covers a region that has N% total population, then the circle size is N%.
Visualize SaTScan results

- SaTScan reports clusters in text format. The clusters need to be visualized on a map.
- Visual Inquiry Toolkit (VIT)
  - Visualize SaTScan results without special skills on GIS.
  - Load multiple results
  - Interact with the results
Relative Risk for US Cervical Cancer Mortality

- Measured by Standardized Mortality Ratio (SMR)
  - Equals to observed/expected deaths
  - In theory, a value of 1 means normal risk.

- **Orange (high risk)** ratio $>1.2$

- **White (normal)** ratio $= 0.8-1.2$

- **Blue (low risk)** ratio $< 0.8$
Heterogeneous clusters

- Large, in undesirable scales, less informative
- Mix of high and low risk enumeration units (e.g., counties)
- Core clusters: smaller, more homogeneous high risk

Maximum circle size =50% population

Maximal circle size =40%
How about reduce the circle size

- Reducing circle size produces small, homogeneous clusters
- SaTScan is sensitive to circle size. Which clusters are more reliable?
Reliability Visualization

- Estimation of reliability
  
  \[ R = \frac{C}{S} \]

  - \( R \) - reliability score for a unit (e.g., a county)
  - \( S \) - total number of scans
  - \( C \) - count that a unit is identified as high risk

- Core clusters
  - More reliable across scales
  - **Dark, green color**
  - E.g., B, D, E, F, G.

Based on 8 scans: 4%, 6%, 8%, 10%, 20%, 30%, 40%, 50%

- Reliability - agreement among multiple results. (\( R \)-score)
- Validity - if a cluster is a true high risk region. (\( p \)-value)
High reliable core clusters are
- stable across multiple runs
- smaller in size
- more homogeneous high risk

SMR map displays only clusters with high reliability scores
Both results report same or similar core clusters (e.g., D, E, F, G)

Clusters (H, K) of low R-score vary

4%, 6%, 8%, 10%, 20%, 30%, 40%, 50%

5%, 7%, 9%, 11%, 19%, 29%, 39%, 49%
Compare with other method

- The two methods produce similar results
- A combination of SaTScan and reliability visualization produces more accurate clusters in terms of location and size

Core clusters by reliability visualization

Kafadar’s spatial smoothing method (in GeoDa)
Summary

• No a single optimized maximum circle size.
• Need run multiple scans with different sizes.
• Reliability visualization can help
  • identify smaller, more homogeneous clusters
  • alleviate sensitivity of SaTScan results
Many thanks for your attention.

Questions?