

Practical Point Pattern Analysis

Outline

- Critiques of Spatial Statistical Methods
- Point pattern analysis versus cluster detection
- Cluster detection techniques
- Extensions to point pattern measures
 - Multiple sets of events
 - Space-time analysis

Peter Gould's Critiques

- Geographical data sets are not samples
- Geographical data are not random
- Geographical data are not independent random
- n is always large so results are almost always statistically significant
- A null hypothesis of IRP/CSR being rejected means any other process is the alternative hypothesis

David Harvey's Critiques

- MAUP: altering parameter estimates by changing study region size often can alter conclusions

Problems of PPA in Real World

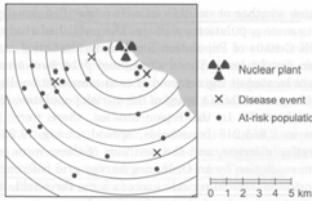
- The application of pure spatial statistical analysis to real-world data and problems is only rarely possible or useful
 - IRP/CSR is rarely an adequate null hypothesis
 - Underlying population distribution
 - Rarely randomly distributed

Global vs. Local

- The pattern-process comparison approach is a **global** technique
 - Concerned with the **overall** characteristics of a pattern and saying little about where the pattern deviates from expectations
 - “*Are there clusters of events in the study area?*”
- We need **local** methods:
 - Identifying locations where there are more events than expected may be an important first step in determining what causes them.
 - “*Where are these clusters of events located in the study area?*”
 - “Cluster Detection”: e.g. hot-spot detection

Detecting Clusters Around Foci

Whether leukemia disease has a connection with nuclear plant?



1. Draw circles centered at the plant (1km, 2 km, ...)
2. Count the number of disease events within 1, 2, ..., 10 km of the nuclear plant
3. Count the total population at risk in these bands
4. Determine the rates of occurrence for each band
5. Compare the rates to expected values generated either analytically or by simulation

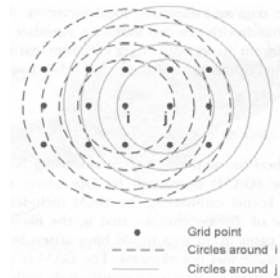
Detecting Clusters Around Foci

- Problems:
 - The boundaries given by the distance bands are arbitrary and because they can be varied, are subject to the **MAUP**
 - The test is a post hoc, after the fact, test. It is **unfair** to choose only the plant as the center of the circles
 - A lot of cases we are looking for the cluster centers

Geographical Analysis Machine (GAM)

- Developed by Stan Openshaw et al (1987)
- A brave but controversial attempt
- Originally developed to study clustering of certain cancers, especially childhood leukemia, around nuclear facilities in England
- GAM is an **automated** cluster detector for point patterns that made an **exhaustive** search using all possible centers of all possible clusters

GAM

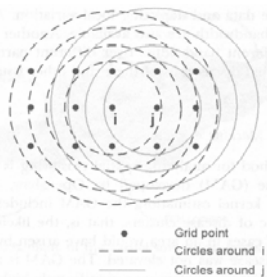


Step 1: lay out a two-dimensional grid over the study region.

Step 2: treat each grid point as the center of a series of search circles

Step 3: generate circles of a defined sequence of radii (e.g. 1.0, 2.0, ..., 20 km)

GAM



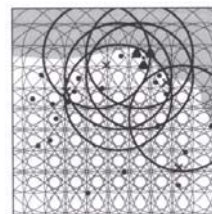
Step 4: for each circle, count the number of events and population at risk falling within it

Step 5: determine whether or not this exceeds a specified density threshold.

Step 6: if the incidence rate in a circle exceeds some threshold, draw the circle on a map.

GAM Case Study: Leukemia

Threshold levels for determining significant circles were determined using **Monte Carlo simulation**



1. Calculate the average incidence rate
2. Randomly assign Leukemia to census enumeration districts (ED) such that each ED has same rate.
3. Run 99 times
4. The actual count of Leukemia cases in each circle was compared to the count that would have been observed for each of the 99 simulated outcomes.
5. Any circle whose observed count was highest among this set of 100 patterns was highlighted ($p=0.01$)

GAM Case Study: Leukemia



- **Clusters detected:** childhood acute lymphoblastic leukemia in a study region in England using GAM (Openshaw et al. 1988)
- Excess risk among children whose fathers worked at the nuclear facilities, especially those fathers who were exposed to a high dose of ionizing before the children's conception (Wakeford 1990).

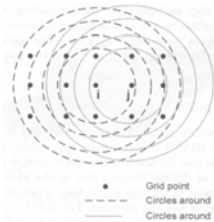
Limitations of GAM

- GAM lacks a clear statistical yardstick for evaluating the number of significant circles that appear on the map.
- Because the circles overlap, many significant circles often contain the same cluster of cases (Poisson tests are not independent)
- The GAM maps often give the appearance of **excess** clustering, with a **high percentage of "false positive" circles**.

Limitations of GAM

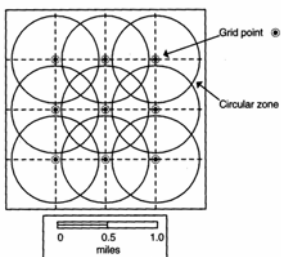
- Two critical issues: attempts to reduce search time
 - Determiner of clusters
 - Exhaustive search → [Smart search](#)

Spatial Scan Statistic



- Developed by Kulldorff (1997), similar to GAM, utilizes a field approach & searches over a regular grid using **circles of different sizes**
- For each circle, computes the likelihood that the risk of disease is elevated **inside** the circle compared to **outside** the circle.
- The circle with the highest likelihood value is the circle that has the highest probability of containing a disease cluster
- Software downloadable from NIH web (<http://dcp.nci.nih.gov/bb/satscan.html>)

Rushton and Lolonis's Method



- 1) It uses a **window of constant size** to scan the study area for clusters
- 2) It provides information about the likelihood that a cluster might have occurred by chance
- 3) **Monte Carlo** procedures are used to simulate possible spatial patterns of health events

Cluster Detection - An Object Oriented Approach

- Besag and Newell (1991) devised a spatial clustering method that **only searches for clusters around cases**.
- Their method adopts an "**object**" approach, treating health events as objects, instead of **field** approach of the previous methods.
- This greatly reduces the amount of spatial search and computation.

Cluster Detection - An Object Oriented Approach

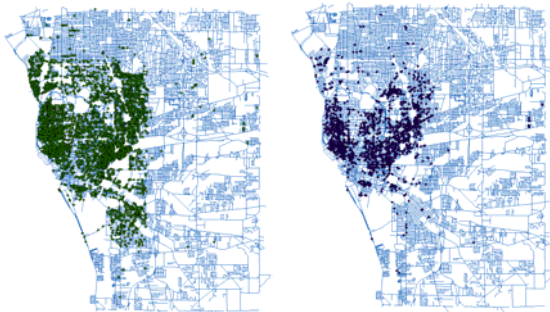


- Step1: Specify k as the minimum event number of a cluster
- Step2: for an event i , find the nearest k events
- Step3: Identify the geographic area M_i that contains these k events
- Step 4: For M_i , calculate the total event number and the population
- Step 5: compare the incidence rate to the average rate (simulation).

Extensions to Point Pattern Measures

- What if there are multiple sets of events?
 - Do the two point patterns differ significantly?
 - Does one point pattern influence another?
 - Does the location of an event in **Pattern 1** has any impact on the location of an event in **Pattern 2**?
- Two approaches
 - Contingency table analysis
 - Distance cross functions

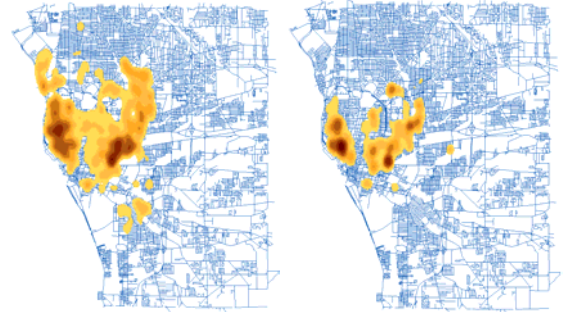
Crime Patterns at Buffalo



Burglary 1996

Drug 1996

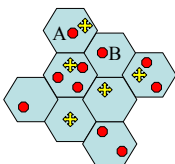
Crime Patterns at Buffalo



Burglary 1996

Drug 1996

Contingency Table Analysis

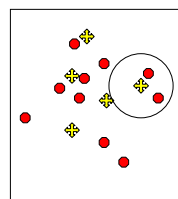


● Event 1 ✚ Event 2

Type 1 events	Type 2 events	
	Absent	Present
Absent	n_{11}	n_{12}
Present	n_{21}	n_{22}

Distance Cross Functions

- $G_{12}(d)$: similar to $F(d)$
- $K_{12}(d)$

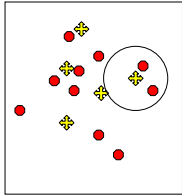


● Event 1 ✚ Event 2

$$K_{12}(d) = \frac{a}{n_1 n_2} \sum_{i=1}^{n_2} no.[S_1 \in C(s_{2i}, d)]$$

$K_{12}(d)$ is large, means two patterns unrelated or related?

Distance Cross Functions



If **Pattern 1** represents **cases of a disease** and **Pattern 2** represents the **at-risk population**

$$D(d) = K_{11}(d) - K_{22}(d)$$

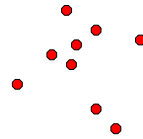
If $D(d) > 0$, then ?

If $D(d) = 0$, then ?

● Event 1 ◆ Event 2

Space-Time Analysis

- **Knox test**: for n events we form the $n(n-1)/2$ possible pairs and find for each their distance in both time and space
- Decide thresholds for time (near-far) and distance (close-distant): could be problematic
- Put them in the contingency table



Proximity in time	Proximity in space	
	Close	Distant
Near	n_{11}	n_{12}
Far	n_{21}	n_{22}

Space-Time Analysis

Space-time K function

$\lambda K(d,t) = E$ (no. events within distance d and time t of an arbitrary event)

If there is no space-time interaction

$$K(d,t) = K(d)K(t)$$

Diggle test (1995)

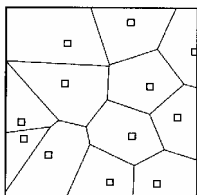
$$D(d,t) = K(d,t) - K(d)K(t)$$

Density and Distance Together

- Proximity Polygon Approach
 - Measure simple properties of polygons to assess the point pattern
 - **Clustered**: small polygons separated by large polygons
 - **Even spacing**: similar polygon size
 - Still fairly new

Proximity Polygon

- The **proximity polygon** of any entity is that region of the space which is closer to the entity than it is to any other



The shape and size of the polygon suggests patterns

Review

- Global vs. Local
- Cluster Detection methods
 - GAM and its variations
- Extension of PPA measures
 - Contingency table analysis
 - Distance cross functions
 - Space-time analysis