Geographical Disease Surveillance

- Discrete response
- Hotspot detection and tree-structured SatScan
- Hotspot delineation and hot-spot rating
- Multiple hotspot detection and delineation
- Hotspot prioritization and poset ranking
- Space-time detection and early warning
- Continuous response
- User friendly software and downloadable website
Breast Cancer by ZIP Code
New York State, 1993-1997

Simple SIRs as observed/expected

<table>
<thead>
<tr>
<th>SIR (maximum likelihood estimate)</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>more than 100% above expected</td>
<td>28</td>
</tr>
<tr>
<td>50% to 100% above expected</td>
<td>93</td>
</tr>
<tr>
<td>15% to 49% above expected</td>
<td>279</td>
</tr>
<tr>
<td>within 15% of expected</td>
<td>471</td>
</tr>
<tr>
<td>15% to 50% below expected</td>
<td>338</td>
</tr>
<tr>
<td>more than 50% below expected</td>
<td>100</td>
</tr>
<tr>
<td>very sparse data</td>
<td>104</td>
</tr>
</tbody>
</table>
Geospatial Analysis for Disease Surveillance—1

Case Event Point Data & Areal Unit Count Data

- Geospatial surveillance
- Cluster detection and evaluation
- Spatial scan statistics
- Choice of zonal parameter space
- Candidate zones as circular windows of expanding size
- Elliptical windows: long island breast cancer study
- Hyperclusters, echelon trees, upper surface sets defined by thresholds-based nodes
Geospatial Analysis for Disease Surveillance—2
Case Event Point Data & Areal Unit Count Data

- Spatio-temporal surveillance
- Cylinders-based spatio-temporal scan statistics
- Three-dimensional echelons and echelon trees
- Candidate zones as upper surface sets defined by thresholds-based nodes
- Temporal persistence and patterns
- Cluster alarms, suspect clusters, and their evaluation
Geospatial and Spatiotemporal Patterns of Change

- Multiple cancer mortality statistics and maps
- Multiple disease incidence statistics and maps
- Across United States over years
- Across individual states over years
- Pooling over types of cancer/disease
- Pooling over types of people
- Change detection and change analysis
- In space, in time, in space-time
- Structure and behavior of chance
- Persistence and patterns of elevated areas
Spatial and Spatiotemporal Scan Statistics—1
SaTScan

• To evaluate reported spatial or spatiotemporal disease clusters
• To see if they are statistically significant
• To test whether a disease is randomly distributed
• To perform geographical surveillance of disease
• To detect areas of significantly high or low rates
Spatial and Spatiotemporal Scan
Statistics—2
SaTScan

• Poisson model, where the number of events in an area is Poisson distributed under the null hypothesis
• Bernoulli model, with 0/1 event data such as cases and controls
• The program adjusts for the underlying inhomogeneity of a background population
• With the Poisson model, the program can also adjust for any number of categorical variates provided by the user
• Tessellation of a geographic region

Region $R$, Tessellation $T = \{a\}$ of $R$

Cell $a$, Response $Y_a$, Cell “Size” $A_a$

Two distributional settings:

– $Y_a$ is Binomial $(N_a, p_a)$, $A_a = N_a$, $p_a = \text{cell rate/intensity}$
– $Y_a$ is Poisson $(\lambda_a A_a)$, $\lambda_a = \text{cell rate/intensity}$

Cell sizes $A_a$ are known and fixed

Cell responses $Y_a$, $a \in A$, are independent
• $G_a = \frac{Y_a}{A_a}$ empirical cell intensity
determines a cellular (piece-wise constant) surface
defined over the tessellated region

• Zones $Z$ are connected unions of cells from tessellation
  $\Omega = \text{collection of all possible zones}$

• SatScan hotspot model:
  – Zone $Z$ such that
  – $p_a = p_1$ for all $a \in Z$
  – $p_a = p_0$ for all $a \in R - Z$
  – $p_1 > p_0$
SATScan Setup – 3
Hotspot Detection

- **Hypothesis testing approach:**
  \[ H_0: \text{There is no hotspot; } p_a \text{ constant for all cells } a \]
  \[ H_1: \text{There is a hotspot } Z \in \Omega; \ Z \text{ unknown} \]

- **Parameter space for full model:**
  \[ \{(Z, p_1, p_0) : Z \in \Omega, 0 < p_0 \leq p_1 < 1 \} \]

- **For fixed** \( Z \), expression for likelihood and MLE for \( p_1, p_0 \) are straightforward

- **Profile likelihood for** \( Z \):
  \[ L(Z) = \text{Max} \ \{ L(Z, p_1, p_0) : 0 < p_0 \leq p_1 < 1 \} , \quad Z \in \Omega \]
SATScan Setup – 4
Hotspot Estimation

• $\Omega$ is finite but large

• Maximizing $L(Z), Z \in \Omega$, by exhaustive search impractical

• Possible optimization methods:
  – Stochastic optimization (annealing, GA, etc.)
  – Reduction of $\Omega$ to $\Omega_0$ with $\Omega_0 \subset \Omega$
    and $\Omega_0$ small enough for exhaustive search

• Possible reductions of $\Omega$ to $\Omega_0$:
  – Expanding circles
    An a priori reduction which depends only on the tessellation
    and not on the data
  – Upper level sets of empirical intensity surface (Tree-Structured SATScan)
    An adaptive, data-dependent, reduction allowing flexible shapes for
    zones in $\Omega_0$. Data dependence must be incorporated into simulations.
Hotspot Testing

- Test $H_0$ versus $H_1$ for significance of hotspot
- Use reduced parameter space $\Omega_0$
- $\hat{Z} =$ hotspot MLE, maximizes $L(Z)$, $Z \in \Omega_0$
- Likelihood ratio test
  Test statistic: $LR = L(H_0) / L(\hat{Z})$
  Reject $H_0$ when $LR$ is small
- Nonstandard likelihood situation
  ($\Omega_0$ is finite discrete; parameter $Z$ is non-identifiable under $H_0$)
- Asymptotic chi-squared not applicable
- Null distribution of LR to be determined by simulation
  - Eliminate nuisance parameters $p_1, p_0$ under $H_0$ by conditioning on the total response $\sum_a Y_a$
SATScan Setup – 6

• **Goal:** Identify geographic zone(s) in which a response is significantly elevated relative to the rest of a region

• A list of candidate zones $Z$ is specified *a priori*.
  – This list becomes part of the parameter space and the zone must be estimated from within this list.
  – Each candidate zone should generally be spatially connected, e.g., a union of contiguous spatial units or cells.
  – Longer lists of candidate zones are usually preferable
  – Expanding circles or ellipses about specified centers are a common method of generating the list
• **Question:** Are there data-driven (rather than \textit{a priori}) ways of selecting the list of candidate zones?

• **Motivation for the question:** A human being can look at a map and quickly determine a reasonable set of candidate zones and eliminate many other zones as obviously uninteresting. Can the computer do the same thing?

• **A data-driven proposal:** Candidate zones are the connected components of the upper level sets of the response surface. The candidate zones have a tree structure, which may assist in automated detection of multiple, but geographically separate, elevated zones.

• **Null distribution:** If the list is data-driven (i.e., random), its variability must be accounted for in the null distribution. A new list must be developed for each simulated data set.
Tree-Structured SATScan

- Data-adaptive approach to reduced parameter space $\Omega_0$
- Zones in $\Omega_0$ are connected components of upper level sets of the empirical intensity function $G_a = Y_a / A_a$
- Upper level set (ULS) at level $g$ consists of all cells $a$ where $G_a \geq g$
- Upper level sets may be disconnected. Connected components are the candidate zones in $\Omega_0$
- These connected components form a rooted tree under set inclusion.
  - Root node = entire region $R$
  - Leaf nodes = local maxima of empirical intensity surface
  - Junction nodes occur when connectivity of ULS changes with falling intensity level
Upper Level Set (ULS) of Intensity Surface

Intensity $G$

Region $R$

Hotspot zones at level $g$
(Connected Components of upper level set)
Changing Connectivity of ULS as Level Drops

Intensity $G$

$g$

$g'$

Region $R$
ULS Connectivity Tree

Intensity $G$

Schematic intensity “surface”

N.B. Intensity surface is cellular (piece-wise constant), with only finitely many levels
A, B, C are junction nodes where multiple zones coalesce into a single zone
• Ingredients:
  – Tessellation of a geographic region:
  – Intensity value $G$ on each cell. Determines a cellular (piece-wise constant) surface with $G$ as elevation.

• Imagine surface initially inundated with water
• Water evaporates gradually exposing the surface which appears as islands in the sea
• How does connectivity (number of connected components) of the exposed surface change with falling water level?
Think of the tessellated surface as a landform

Initially the entire surface is under water

As the water level recedes, more and more of the landform is exposed

At each water level, cells are colored as follows:
- Green for previously exposed cells (green = vegetated)
- Yellow for newly exposed cells (yellow = sandy beach)
- Blue for unexposed cells (blue = under water)

For each newly exposed cell, one of three things happens:
- New island emerges.
  Cell is a local maximum. Morse index=2. Connectivity increases.
- Existing island increases in size.
  Cell is not a critical point. Connectivity unchanged.
- Two (or more) islands are joined.
  Cell is a saddle point Morse index=1. Connectivity decreases.
ULS Connectivity Tree -- 3

Newly exposed island

Island grows

ULS Tree

- a
- b, c
ULS Connectivity Tree -- 4

Second island appears

Both islands grow

ULS Tree

New leaf node (local maximum)
Islands join – saddle point

Exposed land grows

ULS Connectivity Tree -- 5

ULS Tree

Junction node

Root node
Comparison of Tree-Structured and Circle-Based SATScan

- Agreement/Disagreement regarding hotspot locus
  \[ \Pr[ \hat{\mathcal{Z}}_{ULS} \cap \hat{\mathcal{Z}}_{circles} \neq \emptyset ] \]

- Comparative plausibility and accuracy of hotspot delineation
  \[ \Pr[ L(\hat{\mathcal{Z}}_{ULS}) \geq L(\hat{\mathcal{Z}}_{circles}) ] \]

- Execution time and computer efficiency
Poor Hotspot Delineation by Circular Zones

Circular zones may represent single hotspot as multiple hotspots
Hotspot Delineation and Hotspot Rating -- 1

• Determine a confidence set for the hotspot
• Each member of the confidence set is a zone which is a statistically plausible delineation of the hotspot at specified confidence
• Confidence set lets us rate individual cells a for hotspot membership
• Rating for cell a is percentage of zones in confidence set that contain a. (More generally, use weighted proportion.)
• Map of cell ratings:
  – Inner envelope = cells with 100% rating
  – Outer envelope = cells with positive rating
Estimation Uncertainty in Hotspot Delineation

- MLE
- Outer envelope
- Inner envelope
Confidence Set Determination

• Confidence set is all null hypotheses that cannot be rejected

• As hypotheses, use

\[ \tilde{H}_0 : \text{hotspot } Z = Z_0 \]
\[ \tilde{H}_1 : \text{hotspot } Z \neq Z_0 \]

where \( Z_0 \in \Omega_0 \) is a given zone.

• Confidence set is all \( Z_0 \in \Omega_0 \) for which \( \tilde{H}_0 \) cannot be rejected.

• Likelihood ratio test:

Test statistic: \( LR = \frac{L(Z_0)}{L(\hat{Z})} \) where \( \hat{Z} \) = MLE under \( \tilde{H}_0 \cup \tilde{H}_1 \)

Reject \( H_0 \) when \( LR \) is small

• Null distributions have to be determined by simulation
Hotspot Delineation and Hotspot Rating -- 3
Review of LR Confidence Set Determination

• \( H_0 : \theta = \theta_0 \)
  \( H_1 : \theta \neq \theta_0 \)

\( \alpha = \) significance level, \( c = \) confidence level

Test statistic: \( LR = \frac{L(\theta_0)}{L(\hat{\theta})} \)

\( \Pr[ LR \geq t(\theta_0) \mid H_0 : \theta = \theta_0 ] = 1 - \alpha = c \)

\( t(\theta_0) = t, \) critical point approximately free of \( \theta_0 \)

• Null distribution free of parameter (approximately)
• Simulation of null distribution at endpoints of confidence interval
LR Confidence Intervals

\[ LR(\theta_0) \]

\( t \)

\( t' \)

\( \hat{\theta} \)

Disconnected CI

CI at threshold \( t \)
Confidence Region on ULS Tree

MLE

Junction Node

Alternative Hotspot Delineation

Tessellated Region $R$

Nodes in Confidence Set

Extremity Node
Hotspot Delineation and Hotspot Rating -- 6
Tree-Structured SATScan

• How is the null distribution to be simulated for given $Z_0 \in \Omega_0$?

• What is the analogue of extemity or boundary of the confidence set, when the parameter set $\Omega_0$ is finite?

• How do we handle and interpret multimodality of LR giving rise to disconnected confidence set for the hotspot?
Hotspot Delineation and Hotspot Rating -- 7
Tree-Structured SATScan

- Is the null distribution fairly constant across much of the tree?
- Assignment of $p$-value to every LR value and hence to every node in the ULS tree
- Secondary hotspots, $p$-values *versus* pseudo $p$-values
Multiple Hotspot Detection and Delineation-1
Tree-Structured SatScan

$H_0 : p_a$ constant, $a \in T$. No hotspot

$H'_1 : \exists M > 0$ and $\{Z_i : i = 1,2,\ldots,M; \text{separated}\}$, such that

\[
p_a = p_i, \quad a \in Z_i
\]

\[
p_a = p_0, \quad a \in R - \bigcup Z_i = Z_0
\]

\[
p_i > p_0, \quad i = 1,2,\ldots,M
\]

Under $H'_1$, the following parameters:

$M; Z_1, Z_2, \ldots, Z_M; p_0, p_1, p_2, \ldots, p_M$. 
Multiple Hotspot Detection and Delineation-2
Tree-Structured SatScan
Maximized Profile Likelihood

\[
L(M, Z_1, Z_2, \cdots, Z_M) = \max_{p_0, \cdots, p_M} \{ L(M; Z_1, \cdots; Z_M; p_0, p_1, \cdots; p_M) \}
\]

\[
= L(M; Z_1, \cdots, Z_M; \hat{p}_0, \hat{p}_1, \cdots, \hat{p}_M)
\]

where \( \hat{p}_i = \frac{\sum Y_a}{\sum N_a}, a \in Z_i \).

Exhaustive search to maximize \( L \) over \( \Omega = \{M; Z_1, \cdots, Z_M\} \)
\( \Omega \) enormously large.

Apply stochastic optimisation or
choose manageable subset \( \Omega' \subset \Omega \) and exhaustive search over \( \Omega' \)
Multiple Hotspot Detection and Delineation -- 3
Tree-Structured SATScan

- Three parameter space reduction schemes:
  - Falling waterline model (modes of intensity function on ULS tree)
  - Modes of LR on ULS connectivity tree
  - Sequential determination of secondary hotspots
Likelihood Function Defined on ULS Tree
Multiple Hotspots

Numerical labels are values of the likelihood function.
Highlighted nodes are local maxima (modes) of the likelihood function over the tree. Each is a candidate for a hotspot locus.
Node \( x \) is the MLE (global maximum).
Hotspot Prioritization and Poset Ranking

• Multiple hotspots with intensities significantly elevated relative to the rest of the region
• Ranking based on likelihood values, and additional attributes: raw intensity values, socio-economic and demographic factors, feasibility scores, excess cases, seasonal residence, atypical demographics, etc.
• Multiple attributes, multiple indicators
• Ranking without having to integrate the multiple indicators into a composite index
Breast Cancer by ZIP Code
New York State, 1993-1997

Simple SIRs as observed/expected

<table>
<thead>
<tr>
<th>SIR (maximum likelihood estimate)</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>more than 100% above expected</td>
<td>28</td>
</tr>
<tr>
<td>50% to 100% above expected</td>
<td>93</td>
</tr>
<tr>
<td>15% to 49% above expected</td>
<td>279</td>
</tr>
<tr>
<td>within 15% of expected</td>
<td>471</td>
</tr>
<tr>
<td>15% to 50% below expected</td>
<td>338</td>
</tr>
<tr>
<td>more than 50% below expected</td>
<td>100</td>
</tr>
<tr>
<td>very sparse data</td>
<td>104</td>
</tr>
</tbody>
</table>
Ranking Possible Disease Clusters in the State of New York

Data Matrix

<table>
<thead>
<tr>
<th>cluster</th>
<th>SIR</th>
<th>LL</th>
<th>Young Cases</th>
<th>Multiple Cancers</th>
<th>Atypical Demographics</th>
<th>Late Stage of Diagnosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>LF2</td>
<td>2.09</td>
<td>10.36</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>LM14</td>
<td>1.5</td>
<td>36</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>LM4</td>
<td>2.04</td>
<td>19.21</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>LF7</td>
<td>1.51</td>
<td>15.43</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>B2</td>
<td>1.21</td>
<td>31.3</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>B4</td>
<td>1.25</td>
<td>28.4</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>LM1</td>
<td>2.32</td>
<td>21.91</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>LM3</td>
<td>2.13</td>
<td>21.26</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>LM7</td>
<td>2.12</td>
<td>13.33</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
</tbody>
</table>

* LF = lung, female; LM = lung, male; B = breast
Multiple Criteria Analysis
Multiple Indicators
Partial Ordering Procedures

- Cells are objects of primary interest, such as countries, states, watersheds, counties, etc.
- Cell comparisons and rankings are the goals
- Suite of indicators are available on each cell
- Different indicators have different comparative messages, i.e., partial instead of linear ordering
- Hasse diagrams for visualization of partial orders. Multi-level diagram whose top level of nodes consists of all maximal elements in the partially ordered set of objects. Next level consists of all maximal elements when top level is removed from the partially ordered set, etc. Nodes are joined by segments when they are immediately comparable.
HUMAN ENVIRONMENT INDEX
LAND, AIR, WATER INDICATORS

for land - % of undomesticated land, i.e. total land area - domesticated (permanent corps and pastures, built up areas, roads, etc.)
for air - % of renewable energy resources, i.e. hydro, solar, wind, geothermal
for water - % of population with access to safe drinking water

<table>
<thead>
<tr>
<th>RANK</th>
<th>COUNTRY</th>
<th>LAND</th>
<th>AIR</th>
<th>WATER</th>
<th>HEI</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Sweden</td>
<td>69.01</td>
<td>35.24</td>
<td>100</td>
<td>0.68</td>
</tr>
<tr>
<td>2</td>
<td>Finland</td>
<td>76.46</td>
<td>19.05</td>
<td>98</td>
<td>0.65</td>
</tr>
<tr>
<td>3</td>
<td>Norway</td>
<td>27.38</td>
<td>63.98</td>
<td>100</td>
<td>0.64</td>
</tr>
<tr>
<td>5</td>
<td>Iceland</td>
<td>1.79</td>
<td>80.25</td>
<td>100</td>
<td>0.61</td>
</tr>
<tr>
<td>13</td>
<td>Austria</td>
<td>40.57</td>
<td>29.85</td>
<td>100</td>
<td>0.57</td>
</tr>
<tr>
<td>22</td>
<td>Switzerland</td>
<td>30.17</td>
<td>28.10</td>
<td>100</td>
<td>0.53</td>
</tr>
<tr>
<td>39</td>
<td>Spain</td>
<td>32.63</td>
<td>7.74</td>
<td>100</td>
<td>0.47</td>
</tr>
<tr>
<td>45</td>
<td>France</td>
<td>28.34</td>
<td>6.50</td>
<td>100</td>
<td>0.45</td>
</tr>
<tr>
<td>47</td>
<td>Germany</td>
<td>32.56</td>
<td>2.10</td>
<td>100</td>
<td>0.45</td>
</tr>
<tr>
<td>51</td>
<td>Portugal</td>
<td>34.62</td>
<td>14.29</td>
<td>82</td>
<td>0.44</td>
</tr>
<tr>
<td>52</td>
<td>Italy</td>
<td>23.35</td>
<td>6.89</td>
<td>100</td>
<td>0.43</td>
</tr>
<tr>
<td>59</td>
<td>Greece</td>
<td>21.59</td>
<td>3.20</td>
<td>98</td>
<td>0.41</td>
</tr>
<tr>
<td>61</td>
<td>Belgium</td>
<td>21.84</td>
<td>0.00</td>
<td>100</td>
<td>0.41</td>
</tr>
<tr>
<td>64</td>
<td>Netherlands</td>
<td>19.43</td>
<td>1.07</td>
<td>100</td>
<td>0.40</td>
</tr>
<tr>
<td>77</td>
<td>Denmark</td>
<td>9.83</td>
<td>5.04</td>
<td>100</td>
<td>0.38</td>
</tr>
<tr>
<td>78</td>
<td>United Kingdom</td>
<td>12.64</td>
<td>1.13</td>
<td>100</td>
<td>0.38</td>
</tr>
<tr>
<td>81</td>
<td>Ireland</td>
<td>9.25</td>
<td>1.99</td>
<td>100</td>
<td>0.37</td>
</tr>
</tbody>
</table>
Hasse Diagram
(all countries)
Hasse Diagram
(western Europe)
An Example

Poset (Hasse Diagram)

Some linear extensions

Jump Size: 3 1 5 4 2

Jump or Imputed Link (-------) is a link in the ranking that is not implied by the partial order
In the example from the preceding slide, there are a total of 16 linear extensions, giving the following frequency table.

<table>
<thead>
<tr>
<th>Element</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>9</td>
<td>5</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>16</td>
</tr>
<tr>
<td>b</td>
<td>7</td>
<td>5</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>16</td>
</tr>
<tr>
<td>c</td>
<td>0</td>
<td>4</td>
<td>6</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>16</td>
</tr>
<tr>
<td>d</td>
<td>0</td>
<td>2</td>
<td>4</td>
<td>6</td>
<td>4</td>
<td>0</td>
<td>16</td>
</tr>
<tr>
<td>e</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>3</td>
<td>6</td>
<td>6</td>
<td>16</td>
</tr>
<tr>
<td>f</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>10</td>
<td>16</td>
</tr>
<tr>
<td>Totals</td>
<td>16</td>
<td>16</td>
<td>16</td>
<td>16</td>
<td>16</td>
<td>16</td>
<td>16</td>
</tr>
</tbody>
</table>

- Each (normalized) row gives the rank-frequency distribution for that element
- Each (normalized) column gives a rank-assignment distribution across the poset
Ranking Partially Ordered Sets – 3a
Rank-Frequency Distributions

Element a

Element b

Element c

Element d

Element e

Element f
Ranking Partially Ordered Sets – 5

Poset (Hasse Diagram)

Linear extension decision tree

Jump Size: 1 3 3 2 3 5 4 3 3 2 4 3 4 4 2 49 2
An Example of the Procedure

In the example from the preceding slide, there are a total of 16 linear extensions, giving the following cumulative frequency table.

<table>
<thead>
<tr>
<th>Element</th>
<th>Rank</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>(a)</td>
<td>9</td>
<td>14</td>
<td>16</td>
<td>16</td>
<td>16</td>
<td>16</td>
</tr>
<tr>
<td>(b)</td>
<td>7</td>
<td>12</td>
<td>15</td>
<td>16</td>
<td>16</td>
<td>16</td>
</tr>
<tr>
<td>(c)</td>
<td>0</td>
<td>4</td>
<td>10</td>
<td>16</td>
<td>16</td>
<td>16</td>
</tr>
<tr>
<td>(d)</td>
<td>0</td>
<td>2</td>
<td>6</td>
<td>12</td>
<td>16</td>
<td>16</td>
</tr>
<tr>
<td>(e)</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>4</td>
<td>10</td>
<td>16</td>
</tr>
<tr>
<td>(f)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>16</td>
</tr>
</tbody>
</table>

Each entry gives the number of linear extensions in which the element (row label) receives a rank equal to or better that the column heading.
The curves are stacked one above the other and the result is a linear ordering of the elements: \( a > b > c > d > e > f \)
Cumulative Rank Frequency Operator – 7

An example where \( \mathcal{F} \) must be iterated

Original Poset (Hasse Diagram)

\[
\begin{array}{c}
\bullet a \\
\bullet f \\
\bullet e \\
\bullet b \\
\bullet g \\
\bullet c \\
\bullet h \\
\bullet d
\end{array}
\]

\[
\begin{array}{c}
\mathcal{F} \\
\bullet a \\
\bullet f \\
\bullet e \\
\bullet b \\
\bullet g \\
\bullet c \\
\bullet h \\
\bullet d
\end{array}
\]

\[
\mathcal{F}^2 \\
\bullet a \\
\bullet f \\
\bullet e \\
\bullet b \\
\bullet g \\
\bullet c \\
\bullet h
\]

52
An example where \( F \) results in ties

Original Poset (Hasse Diagram)

\[ a \quad b \quad c \quad d \]

\( F \)

\[ a \quad b, c \text{ (tied)} \quad d \]

- Ties reflect symmetries among incomparable elements in the original Hasse diagram
- Elements that are comparable in the original Hasse diagram will not become tied after applying \( F \) operator
Space-Time Detection and Early Warning-1

Tree-Structured SatScan

• The traditional space-time scan statistic employs cylinders as the candidate zones in the reduced parameter space. In many instances, the cylindrical shape may be a poor approximation to actual space-time hotspots, whereas the ULS approach is able to adapt its shape to the actual hotspot.

• Since the ULS tree is derived from the adjacency matrix, the same software will work once the notion of adjacency has been specified for space-time cells.
Some Space-Time Hotspots and Their Cylindrical Approximations

Cylindrical approximation sees single hotspot as multiple hotspots.
Hotspot Detection, Delineation, and Prioritization
Tree-Structured SatScan

- Continuous response
- Cancer survival
- And in many other situations

\[ Y_a \sim \text{Gamma}(k, \beta) \]

- Additivity with the index parameter \( k \) suggests that we take \( k \) proportional to size:

\[ k_a = A_a / c. \]
Multiple Criteria Analysis
Multiple Indicators and Choices
Health Statistics
Disease Etiology, Health Policy, Resource Allocation

• First stage screening
  – Significant clusters by SaTScan and/or upper level sets
• Second stage screening
  – Multicriteria noteworthy clusters by partially ordered sets and Hass diagrams
• Final stage screening
  – Follow up clusters for etiology, intervention based on multiple criteria using Hass diagrams