RTI SYMPOSIUM on
HOMELAND and
HEALTH SECURITY

Biosurveillance Geoinformatics of
Hotspot Detection and
Prioritization for Biosecurity

G. P. Patil

November 7, 2003
Research Triangle Institute
This report is very disappointing.
What kind of software are you using?
<table>
<thead>
<tr>
<th></th>
<th>Stone-age data</th>
<th>Space-age data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stone-age analysis</td>
<td>![Orange Symbol]</td>
<td>![Green Symbol]</td>
</tr>
<tr>
<td>Space-age analysis</td>
<td>![Red Symbol]</td>
<td>![Blue Symbol]</td>
</tr>
</tbody>
</table>
Spatially distributed response variables

Hotspot analysis

Prioritization

Geoinformatic Surveillance System

Geoinformatic spatio-temporal data from a variety of data products and data sources with agencies, academia, and industry

Masks, filters

Indicators, weights

Decision support systems

Masks, filters
Issues

Limitations and Needs
- Circles capture only compactly shaped clusters
- Want to identify clusters of arbitrary shape
- Circles handle only synoptic (tessellated) data
- Want to also handle data on a network
- Circles provide point estimate of hotspot
- Want to assess estimation uncertainty (hotspot confidence set)

Poor Hotspot Delineation by Circular Zones

Poor Hotspot Delineation by Space-Time Cylinders
Geospatial Surveillance

Synoptic Data
Hotspots within Hotspots

Data on a River Network

Estimation Uncertainty in
Hotspot Identification
Spatial Temporal Surveillance

Shifting Hotspot

Typology of Space-Time Hotspots

Trajectory of a Merging Hotspot
Syndromic Crisis-Index Surveillance

Probabilistic Finite Automata (PFA)

A PFA is a DFA \((Q, q_0, \Sigma, \delta)\) with a probability attached to each transition such that the sum of the probabilities across all transitions from a given node is unity.

Formally, \(p: Q \times \Sigma \rightarrow [0, 1]\) such that

- \(p(q, \omega) = 0\) if and only if \(\delta(q, \omega) = \text{Blocked}\)
- \(\sum_{\omega \in \Sigma} p(q, \omega) = 1\) for all \(q \in Q\)

Multiplying branch probabilities lets us assign a probability value \(\mu(q, \delta)\) to each string \(\delta\) in \(\Sigma^*\). E.g., \(\mu(q_0, abcd) = (0.8)(0.6)(0.4) = 0.192\)

Distance Between Two PFA

Let \(A\) and \(B\) be two PFAs on the same alphabet \(\Sigma\)

Let \(w(i)\) be a probability distribution across string lengths \(i\)

Let \(\pi_A\) and \(\pi_B\) be the \(w\)-weighted probability measures of \(A\) and \(B\)

Define the distance between \(A\) and \(B\) as the variational distance between the probability measures \(\pi_A\) and \(\pi_B\):

\[
d(A, B) = \| \pi_A - \pi_B \|
\]
Hotspot Prioritization

Prioritization of Disease Clusters with Multiple Indicators

**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

**Ranking Partially Ordered Sets**

Poset (Hasse Diagram)

Linear extension decision tree

**Cumulative Rank Frequency Operator**

The curves are stacked one above the other giving a linear ordering of the elements: $a > b > c > d > e > f$

Geoinformatic surveillance for spatial and temporal hotspot detection and prioritization is a critical need for the 21st century Digital Government. A hotspot can mean an unusual phenomenon, anomaly, aberration, outbreak, elevated cluster, or critical area. The declared need may be for monitoring, etiology, management, or early warning. The responsible factors may be natural, accidental or intentional, with relevance to both infrastructure and homeland security.

This project describes a multi-disciplinary research program based on novel methods and tools for hotspot detection and prioritization, driven by a wide variety of case studies of direct interest to several government agencies. These case studies deal with critical societal issues, such as carbon budgets, water resources, ecosystem health, public health, drinking water distribution system, persistent poverty, environmental justice, crop pathogens, invasive species, biosecurity, biosurveillance, remote sensor networks, early warning and homeland security. The geosurveillance provides an excellent opportunity, challenge, and vehicle for synergistic collaboration of computational, technical, and social scientists.

Our methodology involves an innovation of the popular circle-based spatial scan statistic methodology. In particular, it employs the notion of an upper level set and is accordingly called the upper level set scan statistic, pointing to the next generation of a sophisticated analytical and computational system, effective for the detection of arbitrarily shaped hotspots along spatio-temporal dimensions. We also propose a novel prioritization scheme based on multiple indicator and stakeholder criteria without having to integrate indicators into an index, using revealing Hasse diagrams and partially ordered sets.

Responding to the Government's role and need, we propose a cross-disciplinary collaboration among federal agencies and academic researchers to design and build the prototype system for surveillance infrastructure of hotspot detection and prioritization. The methodological toolbox and the software toolkit developed will support and leverage core missions of federal agencies as well as their interactive counterparts in the society. The research advances in the allied sciences and technologies necessary to make such a system work are the thrust of this five year project.

The project will have a dual disciplinary and cross-disciplinary thrust. Dialogues and discussions will be particularly welcome, leading potentially to well considered synergistic case studies. The collaborative case studies are expected to be conceptual, structural, methodological, computational, applicational, developmental, refinemental, validational, and/or visualizational in their individual thrust.

For additional information, see the webpages:
(1) http://www.stat.psu.edu/~gpp/PDFfiles/Prospectus%2016.pdf
(2) http://www.stat.psu.edu/~gpp/PDFfiles/Prospectus%2016%20overview.pdf

Project address: Penn State Center for Statistical Ecology and Environmental Statistics
421 Thomas Building, Penn State University, University Park, PA 16802
Telephone: (814)865-9442; Email: gpp@stat.psu.edu
National Applications

- Biosurveillance
- Carbon Management
- Coastal Management
- Community Infrastructure
- Crop Surveillance
- Disaster Management
- Disease Surveillance
- Ecosystem Health
- Environmental Justice
- Environmental Management
- Environmental Policy
- Homeland Security
- Invasive Species
- Poverty Policy
- Public Health
- Public Health and Environment
- Syndromic Surveillance
- Urban Crime
- Water Management
A primary purpose of MARMAP System Partnership is to develop sound methodology and appropriate software for the quantitative analysis and interpretation of multi-categorical raster maps and cellular surfaces (inferential geospatial informatics) involving landscape pattern analysis, multiscale landcover landuse change detection, accuracy assessment, critical area detection and delineation, disease mapping and geographic surveillance, prioritization and ranking without having to integrate multiple indicators, and a few more. It will be nice to see you participate in one capacity or the other. The following websites may be of particular interest at this time, giving recent publications together with current exciting events. Please feel free also to share this material with your potentially interested friends and colleagues.

1. MARMAP and MARMAP Prospectus 1, 2, 3, 4, 5, 6, 7.


4. Project MARMAP System Partnership Collaboration with UNEP Division of Early Warning and Assessment on Human Environment Index based on Countrywide Land, Air, and Water Indicators.

5. Project MARMAP Show and Tell Seminar series: EPA ORD NCEA, EPA ORD NERL, EPA OEI, NASA HQ, NASA GSFC, NCHS, NYSDEH; UMD, GWU, UCB, MSU, UM, SUNY SPH. Powerpoint Presentations
http://www.stat.psu.edu/~gpp/PDFfiles/TR2002-0202.pdf

http://www.stat.psu.edu/~gpp/PDFfiles/TR2002-0301.pdf

http://www.stat.psu.edu/~gpp/PDFfiles/TR2002-0501.pdf


10. Finding upper level sets in cellular surface data using echelons and SaTScan. 
http://www.stat.psu.edu/~gpp/PDFfiles/TR2002-0801.pdf
Securing the nation's computer networks from cyber attack is an important aspect of Homeland Security. Project develops diagnostic tools for detecting security attacks, infrastructure failures, and other operational aberrations of computer networks.

Many critical applications of surveillance sensor networks involve finding hotspots. The upper level set scan statistic is used to guide the search by estimating the location of hotspots based on the data previously taken by the surveillance network.

New York City has installed 892 drinking water sampling stations. Currently, about 47,000 water samples are analyzed annually. The ULS scan statistic will provide a real-time surveillance system for evaluating water quality across the distribution system.

Emerging hotspots for disease or biological agents are identified by modeling events at local hospitals. A time-dependent crisis index is determined for each hospital in a network. The crisis index is used for hotspot detection by scan statistic methods.

West Nile virus is a serious mosquito-borne disease. The mosquito vector bites both humans and birds. Scan statistical detection of dead bird clusters provides an early crisis warning and allows targeted public education and increased mosquito control.

Disruption of American agriculture and our food system could be catastrophic to the nation's stability. This project has the specific aim of developing novel remote sensing methods and statistical tools for the early detection of crop bioterrorism.

The scan statistic hotspot delineation and poset prioritization tools will be used in combination with our oil spill detection algorithm to provide for early warning and spatial-temporal monitoring of marine oil spills and their consequences.

This study employs the network version of the upper level set scan statistic to characterize biological impairment along the rivers and streams of Pennsylvania and to identify subnetworks that are badly impaired.
Website Links

1. Prospectus 8: Synoptic Surveillance
   http://www.stat.psu.edu/~gpp/PDFfiles/Prospectus-8.pdf
2. Prospectus 11: Network-Based Surveillance
   http://www.stat.psu.edu/~gpp/PDFfiles/Prospectus-11.pdf
3. Prospectus 10: Classification and Prioritization
   http://www.stat.psu.edu/~gpp/PDFfiles/Prospectus-10.pdf
4. Prospectus 9: Crop Surveillance
   http://www.stat.psu.edu/~gpp/PDFfiles/Prospectus-9.pdf
5. Prospectus Abstract Syndromic Surveillance
   http://www.stat.psu.edu/~gpp/PDFfiles/prospectus-12.pdf
6. Poster for Geographic and Network Surveillance for Hotspots
   http://www.stat.psu.edu/~gpp/PDFfiles/Poster%201.pdf
7. Proof-of-Concept Paper-1
   http://www.stat.psu.edu/~gpp/PDFfiles/TR2002-0501.pdf
   http://www.stat.psu.edu/~gpp/PDFfiles/TR2002-0801.pdf
10. Background Biographics 1
    http://www.stat.psu.edu/~gpp/PDFfiles/Patil-3-page%20bio.pdf
11. Background Biographics 2
Hotspot Detection Innovation
Upper Level Set Scan Statistic

Attractive Features
- Identifies arbitrarily shaped clusters
- Data-adaptive zonation of candidate hotspots
- Applicable to data on a network
- Provides both a point estimate as well as a confidence set for the hotspot
- Uses hotspot-membership rating to map hotspot boundary uncertainty
- Computationally efficient
- Applicable to both discrete and continuous syndromic responses
- Identifies arbitrarily shaped clusters in the spatial-temporal domain
- Provides a typology of space-time hotspots with discriminatory surveillance potential
The Spatial Scan Statistic

- Move a circular window across the map.
- Use a variable circle radius, from zero up to a maximum where 50 percent of the population is included.
A small sample of the circles used
Spatial Scan Statistic: Properties

- Adjusts for inhomogeneous population density.
- Simultaneously tests for clusters of any size and any location, by using circular windows with continuously variable radius.
- Accounts for multiple testing.
- Possibility to include confounding variables, such as age, sex or socio-economic variables.
- Aggregated or non-aggregated data (states, counties, census tracts, block groups, households, individuals).
Detecting Emerging Clusters

- Instead of a circular window in two dimensions, we use a cylindrical window in three dimensions.
- The base of the cylinder represents space, while the height represents time.
- The cylinder is flexible in its circular base and starting date, but we only consider those cylinders that reach all the way to the end of the study period. Hence, we are only considering ‘alive’ clusters.
West Nile Virus Surveillance in New York City

- 2000 Data: Simulation/Testing of Prospective Surveillance System
- 2001 Data: Real Time Implementation of Daily Prospective Surveillance
Major epicenter on Staten Island

- Dead bird surveillance system: June 14
- Positive bird report: July 16 (coll. July 5)
- Positive mosquito trap: July 24 (coll. July 7)
- Human case report: July 28 (onset July 20)
Hospital Emergency Admissions in New York City

- Hospital emergency admissions data from a majority of New York City hospitals.
- At midnight, hospitals report last 24 hour of data to New York City Department of Health.
- A spatial scan statistic analysis is performed every morning.
- If an alarm, a local investigation is conducted.
Candidate Zones for Hotspots

- **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.
Scan Statistic Zonation for Circles and Space-Time Cylinders

Cholera outbreak along a river flood-plain
• Small circles miss much of the outbreak
• Large circles include many unwanted cells

Outbreak expanding in time
• Small cylinders miss much of the outbreak
• Large cylinders include many unwanted cells
ULS Candidate Zones

- **Question:** Are there data-driven (rather than *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 (echelon tree is a subtree), 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.
ULS Scan Statistic

- 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 $R$

$g$

$Z_1$, $Z_2$, $Z_3$

Hotspot zones at level $g$
(Connected Components of upper level set)
Changing Connectivity of ULS as Level Drops
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.
A confidence set of hotspots on the ULS tree. The different connected components correspond to different hotspot loci while the nodes within a connected component correspond to different delineations of that hotspot.
Network Analysis of Biological Integrity in Freshwater Streams
New York City Water Distribution Network

- Water Sampling Station:
  - Nearly 1000 sampling stations throughout New York City
  - Over 47,000 samples were collected and tested in 2001

- Water Testing and Treatment Facilities:
  - Continuous monitoring of water quality flowing into tunnels and aqueducts
  - Water treatment and disinfection
NYC Drinking Water Quality
Within-City Sampling Stations

- 892 sampling stations
- Each station about 4.5 feet high and draws water from a nearby water main
- Sampling frequency increased after 9-11
  Currently, about 47,000 water samples analyzed annually
- Parameters analyzed:
  - Bacteria
  - Chlorine levels
  - pH
  - Inorganic and organic pollutants
  - Color, turbidity, odor
  - Many others
Sampling Station Locations

City-Wide

Manhattan
Due to the nature of water flow, most hotspot investigations of water resources are network-based.

Some possibilities for synoptic investigations:

- Lakes and other large bodies of water.
  - Remotely sensed parameters.
  - Surface thermal characteristics
- Shorelines and coastal regions
  - Algal blooms?
- Groundwater
  - Sampling/data collection issues
Network-Based Surveillance

- Subway system surveillance
- Drinking water distribution system surveillance
- Stream and river system surveillance
- Postal System Surveillance
- Road transport surveillance
- Syndromic Surveillance
Syndromic Surveillance

- Symptoms of disease such as diarrhea, respiratory problems, headache, etc
- Earlier reporting than diagnosed disease
- Less specific, more noise
Syndromic Surveillance

(left) The overall procedure, leading from admissions records to the crisis index for a hospital. The hotspot detection algorithm is then applied to the crisis index values defined over the hospital network.

(right) The $\varepsilon$-machine procedure for converting an event stream into a parse tree and finally into a probabilistic finite state automaton (PFSA).
Experimental Validation

Formal Language Events:
- a – green to red or red to green
- b – green to tan or tan to green
- c – green to blue or blue to green
- d – red to tan or tan to red
- e – blue to red or red to blue
- f – blue to tan or tan to blue

Wall following

Target Behavior

Random walk

Analyze String Rejections
Mapping Priority Hotspots of Vegetative Disturbance for Carbon Budgets
Carbon Transformations from Disturbance

Carbon transformations from disturbance must be included in developing integrated carbon budgets at national, continental, and global scales.

Data at high spatial resolution can characterize the types and intensities of disturbance to estimate carbon releases.

Rapid identification and prioritization of disturbance hotspots would facilitate:

- Timely follow-up imaging to estimate carbon transformations
- Identifying policies that may have contributed to the disturbance or that might mitigate its effects on the global carbon budget
- Developing a time-series of measurements to estimate ecosystem responses and recovery rates
Data Sources, Methods, and Indicators

- **Disturbance Hotspot Detection**
  - Data products from MODIS instruments on Terra and Aqua
  - Corrected and 8-day composited surface reflectance values at red and near-infrared wavelengths (bands 1 & 2, MOD09_L3) will be used to identify vegetated pixels showing both a significant increase in red reflectance and a decrease in near-infrared reflectance
  - ULS scan statistic applied to pixellated adjacency network of detected vegetation change using a sequential image pair

- **Hotspot Prioritization based on:**
  - hotspot area
  - statistical significance
  - type of land cover
  - vegetation index before the change
  - magnitude of reflectance changes
  - hotspot geographical context
Upper Midwest, Siberian Boreal Forest and Carbon Dynamics

- Relatively Undisturbed Carbon Sinks
- Top down LCLUC Modeling with MARMAP Methods
- Socio-economic Drivers and Regional Carbon Dynamics
- State-Controlled Soviet Era and Post Soviet Transitioning Era
Crop Attack Decision Support System

Site Identification Module

Crops → Key Crop Areas → Threat Locations

NOAA Weather

Signature Development Module

Plants
Infected
Non-infected
Sentinel

Ground Cameras
Air/Space Platforms

Data Processing

Hyperspectral Imagery

Signature Library

Anomaly Report

Ground Truthing

Sentinel
Plants
Non-infected
Infected

Crop Attack Decision Support System
Crop Biosurveillance/Biosecurity

Spectral Signature Deformation

Healthy Plant → Photosynthesis Mapping Operator → Diseased Plant

\[ T : V \rightarrow W \]

\( V \) = Healthy Signature
\( W \) = Diseased Signature
Crop Biosurveillance/Biosecurity
Data Processing Module

- Hyperspectral Imagery
- Image Segmentation (hyperclustering)
- Proxy Signal (per segment)
- Disease Signature
- Signature Library
- Similarity Index (per segment)
- Tessellation (segmentation) of raster grid
- Signature Similarity Map
- Hotspot/Anomaly Detection

Crop Biosurveillance/Biosecurity Data Processing Module

- Hyperspectral Imagery
- Image Segmentation (hyperclustering)
- Proxy Signal (per segment)
- Disease Signature
- Signature Library
- Similarity Index (per segment)
- Tessellation (segmentation) of raster grid
- Signature Similarity Map
- Hotspot/Anomaly Detection
Early Detection of Biological Invasions
Mid-Atlantic Invasive Species

- Significant, but often overlooked problem
- Existing strategies are reactive
- Action taken only after detection of invasion/collateral damage
- Due to time-lag, species may already be established
- Such an approach is environmental catch-up
What is a better way?

Predictive Modeling

- Accumulate occurrence data on native range
- Build ecological niche model
- Project niche model to areas of actual or potential invasion
Predictive Modeling:
Genetic Algorithm for Rule-set Prediction (GARP)

- Uses a genetic algorithm, an artificial intelligence application, for choosing rules
- Uses multiple rule types (BIOCLIM, logistic regression, etc.)
- Different decision rules may apply to different sectors of species’ distributions
- Extensive testing indicates excellent predictive ability
Ecological Niche Modeling

Native range locality data

Specimen records

Ecological data

Temperature
Precipitation
Solar radiation
Snow cover
Frost-free days
Native range locality data

Ecological data

GARP

Ecological Niche Modeling

Distributional prediction

Invasive species projection
Sudden Oak Death
(*Phytophthora ramorum*)
Sudden Oak Death
(*Phytophthora ramorum*)
Sudden Oak Death
Native Range?
Emergent Surveillance Plexus (ESP)
Surveillance Sensor Network Testbed
Autonomous Ocean Sampling Network
Types of Hotspots

- Hotspots due to multiple, localized, stationary sources
- Hotspots corresponding to areas of interest in a stationary mapped field
- Time-dependent, localized hotspots
- Hotspots due to moving point sources
Ocean SAMpling MOBILE Network
OSAMON Feedback Loop

- Network sensors gather preliminary data
- ULS scan statistic uses available data to estimate hotspot
- Network controller directs sensor vehicles to new locations
- Updated data is fed into ULS scan statistic system
SAmpling MOBILE Networks (SAMON)

Additional Application Contexts

- Hotspots for radioactivity and chemical or biological agents to prevent or mitigate the effects of terrorist attacks or to detect nuclear testing
- Mapping elevation, wind, bathymetry, or ocean currents to better understand and protect the environment
- Detecting emerging failures in a complex networked system like the electric grid, internet, cell phone systems
- Mapping the gravitational field to find underground chambers or tunnels for rescue or combat missions
We also present a prioritization innovation. It lies in the ability for prioritization and ranking of hotspots based on multiple indicator and stakeholder criteria without having to integrate indicators into an index, using Hasse diagrams and partial order sets. This leads us to early warning systems, and also to the selection of investigational areas.
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
HUMAN ENVIRONMENT INTERFACE
LAND, AIR, WATER INDICATORS

for land - % of undomesticated land, i.e., total land area-domesticated (permanent crops 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 COUNTRY</th>
<th>LAND</th>
<th>AIR</th>
<th>WATER</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Sweden</td>
<td>69.01</td>
<td>35.24</td>
</tr>
<tr>
<td>2</td>
<td>Finland</td>
<td>76.46</td>
<td>19.05</td>
</tr>
<tr>
<td>3</td>
<td>Norway</td>
<td>27.38</td>
<td>63.98</td>
</tr>
<tr>
<td>5</td>
<td>Iceland</td>
<td>1.79</td>
<td>80.25</td>
</tr>
<tr>
<td>13</td>
<td>Austria</td>
<td>40.57</td>
<td>29.85</td>
</tr>
<tr>
<td>22</td>
<td>Switzerland</td>
<td>30.17</td>
<td>28.10</td>
</tr>
<tr>
<td>39</td>
<td>Spain</td>
<td>32.63</td>
<td>7.74</td>
</tr>
<tr>
<td>45</td>
<td>France</td>
<td>28.34</td>
<td>6.50</td>
</tr>
<tr>
<td>47</td>
<td>Germany</td>
<td>32.56</td>
<td>2.10</td>
</tr>
<tr>
<td>51</td>
<td>Portugal</td>
<td>34.62</td>
<td>14.29</td>
</tr>
<tr>
<td>52</td>
<td>Italy</td>
<td>23.35</td>
<td>6.89</td>
</tr>
<tr>
<td>59</td>
<td>Greece</td>
<td>21.59</td>
<td>3.20</td>
</tr>
<tr>
<td>61</td>
<td>Belgium</td>
<td>21.84</td>
<td>0.00</td>
</tr>
<tr>
<td>64</td>
<td>Netherlands</td>
<td>19.43</td>
<td>1.07</td>
</tr>
<tr>
<td>77</td>
<td>Denmark</td>
<td>9.83</td>
<td>5.04</td>
</tr>
<tr>
<td>78</td>
<td>United Kingdom</td>
<td>12.64</td>
<td>1.13</td>
</tr>
<tr>
<td>81</td>
<td>Ireland</td>
<td>9.25</td>
<td>1.99</td>
</tr>
</tbody>
</table>
Hasse Diagram
(all countries)
Hasse Diagram
(Western Europe)
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 2
Cumulative Rank Frequency Operator – 5
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>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>a</td>
<td>9</td>
</tr>
<tr>
<td>b</td>
<td>7</td>
</tr>
<tr>
<td>c</td>
<td>0</td>
</tr>
<tr>
<td>d</td>
<td>0</td>
</tr>
<tr>
<td>e</td>
<td>0</td>
</tr>
<tr>
<td>f</td>
<td>0</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 $F$ must be iterated

Original Poset
(Hasse Diagram)

$\mathcal{F}$

$h$

$\mathcal{F}^2$

$a$

$h$
Incorporating Judgment

Poset Cumulative Rank Frequency Approach

- Certain of the indicators may be deemed more important than the others

- Such differential importance can be accommodated by the poset cumulative rank frequency approach

- Instead of the uniform distribution on the set of linear extensions, we may use an appropriately weighted probability distribution $\pi$, e.g.,

$$\pi(\omega) = w_0 + w_1 n_1(\omega) + w_2 n_2(\omega) + \cdots + w_p n_p(\omega)$$
Mid-Atlantic Highlands Watershed Prioritization Model (WPM)

MAH Region:
- Most of Pennsylvania
- West Virginia
- Western Maryland
- Western Virginia
Watershed Prioritization Model (WPM)

Classify watersheds according to:

- **Disturbance** — observed stressors
- **Vulnerability** — physical characteristics and natural features
- **Feasibility** — economic, social, and political costs as well as technical limitations of protection and restoration
Watershed Prioritization Model (WPM) Schematic View

**Susceptibility To Impairment (Vulnerability)**

- **High**: High priority for protection
- **Low**: Monitor as reference

**Watershed Disturbance**

- **Low**: Low restoration potential
- **High**: High restoration potential

**Feasibility** is an optional third axis
Watershed Prioritization Model (WPM)

Primary Variables for MAH

- **Disturbance** (observed stressors)
  - Excess sediment
  - Riparian degradation
  - Mine drainage
  - Acid deposition
  - Excess nutrients
  - Exotic species
  - Agriculture (esp. on slopes)
  - Road crossings
  - Forest fragmentation
  - Biological impairment (IBI)

- **Vulnerability** (physical characteristics and natural features)
  - Hydrogeomorphology (HGM)
  - Climate
  - Aspect
  - Slope
  - Stream sinuosity
  - Soil type
  - Bedrock
  - Water Source
Space-Time Poverty Hotspot Typology

- Federal Anti-Poverty Programs have had little success in eradicating pockets of persistent poverty.
- Can spatial-temporal patterns of poverty hotspots provide clues to the causes of poverty and lead to improved location-specific anti-poverty policy?
Geoinformatic spatio-temporal data from a variety of data products and data sources with agencies, academia, and industry

Masks, filters

Indicators, weights

Spatially distributed response variables

Hotspot analysis

Prioritization

Geoinformatic Surveillance System

Decision support systems

Masks, filters
Agency Databases

Disaster Management

Public Health

Ecosystem Health

Other Case Studies

Homeland Security

Statistical Processing: Hotspot Detection, Prioritization, etc.

Data Sharing, Interoperable Middleware

Arbitrary Data Model, Data Format, Data Access

Application Specific De Facto Data/Information Standard

Standard or De Facto Data Model, Data Format, Data Access

Agency Databases

Thematic Databases

Other Databases