Geoinformatic Surveillance of Hotspot Detection, Prioritization and Early Warning*

G. P. Patil and S. L. Rathbun
Penn State University
Department of Statistics
University Park, PA 16802 USA
Tel.: 001 814 865 9442
Email: gpp@stat.psu.edu

R. Acharya and P. Patankar
Penn State University
Dept of Computer Sci & Engineering
University Park, PA 16802 USA
Tel: 001 814 865 9505
Email: acharya@cse.psu.edu

Reza Modarres
George Washington University
Department of Statistics
Washington, DC 20052 USA
Tel: 001 202 994 6359
Email: reza@gwu.edu

ABSTRACT

A Geoinformatic Hotspot Surveillance System (GHS) will be demonstrated. This system is comprised of an upper level set scan statistic system for hotspot delineation, and a poset prioritization and ranking system for hotspot prioritization. Surveillance geoinformatics partnership consists of several interested cross-disciplinary scientists from academia, agencies and private sector.

Categories and Subject Descriptors

H. Information systems; H4. Information systems applications; H4.2 Types of systems.

General Terms

Decision support for hotspot detection, prioritization, and early warning.

Keywords

Upper level set scan statistic system, Poset prioritization system, Surveillance geoinformatics partnership.

1. INTRODUCTION

Government agencies often require concise summaries of georeferenced data to support their decisions regarding the geographic allocation of resources. Geoinformatic surveillance for spatial and spatiotemporal hotspot detection and prioritization is a critical need for the 21st century. A hotspot can mean an unusual phenomenon, anomaly, aberration, outbreak, or critical area. Hotspot delineation and prioritization may be required for etiology, management, or early warning. The demo will perform with live geospatial and spatiotemporal data available on some important issues of current interest as listed in Figure 1.

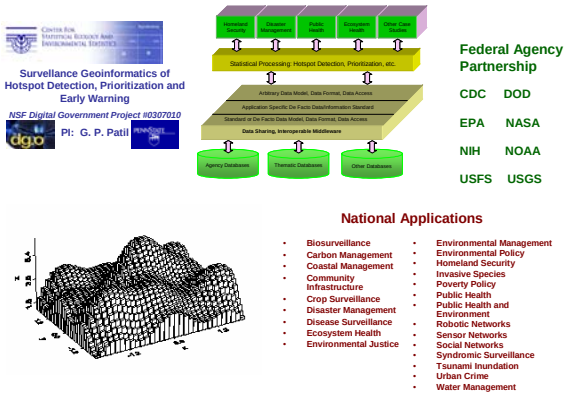

Figure 1. NSF Digital Government surveillance geoinformatics project, federal agency partnership and national applications for digital governance.

With support from the NSF/DG program, an interdisciplinary team has developed a prototype Geoinformatic Hotspot Surveillance (GHS) system for hotspot delineation and prioritization. Our efforts are driven by a wide variety of case studies of potential interest to Federal agencies and involving critical society issues, such as public health, ecosystem health, biosecurity, biosurveillance, robotic networks, social networks, sensor networks, video mining, homeland security, and early warning. The prototype system is comprised of modules for (1) hotspot detection and delineation, and (2) hotspot prioritization.

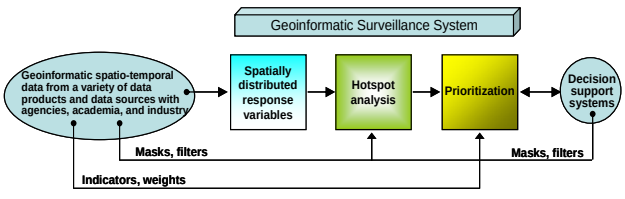

Figure 2. Geoinformatic hotspot surveillance system.

* This paper is dedicated to Charles Taillie, our longtime friend and versatile colleague.

2. UPPER LEVEL SET SCAN STATISTIC SYSTEM

Our approach employs a novel *upper level set scan statistic* to delineate arbitrarily shaped hotspots in both spatial and spatiotemporal dimensions [1]. It features maximum likelihood estimation of candidate hotspots, an upper level set tree, and confidence sets for assessing uncertainty in hotspot delineation.

We introduce, for multidisciplinary use, an innovation of the health-area-popular circle-based spatial and spatiotemporal scan statistic. Success of surveillance rests on potential elevated cluster detection capability. But the clusters can be of any shape, and cannot be captured only by circles. This is likely to give more of false alarms and more of false sense of security. What we need is capability to detect arbitrarily shaped clusters. Currently available circle-based spatial scan statistic software suffers from several limitations. We suggest ways of overcoming these limitations. With suitable modifications, the scan statistic can be used for critical area analysis in fields other than the health sciences. We describe some promising developments that have the following 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

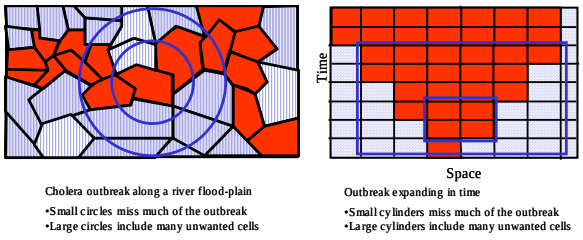

Figure 3. Scan statistic zonation for circles and space-time cylinders.

3. PARTIALLY ORDERED SET PRIORITIZATION SYSTEM

We propose a novel prioritization scheme based on multiple indicators that does not require reduction of the data to a single index. This *poset prioritization and ranking system* features Haase diagrams describing the partial ordering of the data, linear extension decision trees enumerating admissible rankings among hotspots, and cumulative rank functions for hotspot prioritization [2].

Rather than trying to combine indicators, we take the view that the relative positions in indicator space determine only a *partial ordering* and that a given pair of objects may not be inherently comparable. Working with *Hasse diagrams* of the partial order, we study the collection of all rankings that are compatible with the partial order (*linear extensions*). In this way, an interval of possible ranks is assigned to each object. The intervals can be very wide, however. Noting that ranks near the ends of each interval are usually infrequent under linear extensions, a probability distribution is obtained over the interval of possible ranks. This distribution, called the *rank-frequency* distribution, turns out to be unimodal (in fact, log-concave) and represents the degree of ambiguity involved in attempting to assign a rank to the corresponding object.

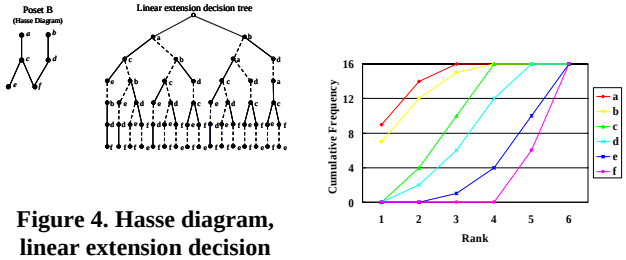

Figure 4. Hasse diagram, linear extension decision tree, and cumulative rank frequency prioritization.

Stochastic ordering of probability distributions imposes a partial order on the collection of rank-frequency distributions. This collection of distributions is in one-to-one correspondence with the original collection of objects and the induced ordering on these objects is called the cumulative rank-frequency (CRF) ordering; it extends the original partial order.

For additional information regarding our project, see http://www.stat.psu.edu/hotspots/ and http://www.stat.psu.edu/~gpp/

4. ACKNOWLEDGMENTS

This material is based upon work supported by (i) the National Science Foundation under Grant No. 0307010, (ii) the United States Environmental Protection Agency under Grant No. CR-83059301 and (iii) the Pennsylvania Department of Health using Tobacco Settlement Funds under Grant No. ME 01324. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the agencies.

5. REFERENCES

[1] Patil, G.P., and Taillie, C. Upper level set scan statistic for detecting arbitrarily shaped hotspots. *Environmental and Ecological Statistics, 11* (2004), 183-197.

[2] Patil, G.P., and Taillie, C. Multiple indicators, partially ordered sets, and linear extensions: Multi-criterion ranking and prioritization. *Environmental and Ecological Statistics 11* (2004), 199-228.