Geoinformatic Hotspot Systems (GHS) for Detection, Prioritization, and Early Warning*

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

ABSTRACT
The five year NSF DGP project has been instrumental to conceptualize surveillance geoinformatics partnership among several interested cross-disciplinary scientists in academia, agencies, and private sector. A declared need is around for statistical geoinformatics and software infrastructure for spatial and spatiotemporal hotspot detection. Our efforts are driven by a wide variety of case studies of potential interest to federal agencies involving critical society issues, such as public health, ecosystem health, biosurveillance, biosecurity, sensor networks, robotic networks, social networks, video mining, homeland security, early warning, and disaster management.

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
Early warning, geosurveillance statistics, hotspot detection, space-time hotspots, surveillance geoinformatics partnership.

1. INTRODUCTION
The primary thrust of the proposed work is now three-fold:

a) To formulate and develop statistical methodology and computational technology for geoinformatic surveillance of hotspot detection and prioritization using upper level set detection

and partially ordered set prioritization methods, software tools, and visualization capabilities.

b) To formulate and initiate individual case study/application area project proposals that will have stronger and speedier performance, utilizing the detection and prioritization methods and software tools of (a) above.

c) To work toward a National Center for Geoinformatic Surveillance, utilizing (a) and (b) above as a synergistic springboard.

2. RESEARCH
Our research activities have focused on planning and software development for the methodological components of the geoinformatic surveillance project. The system has two methodological components: prioritization and hotspot detection. Prioritization-related work includes:

- Visualization software for Hasse diagrams
- Software for exhaustive enumeration of the linear extension decision tree
- Markov chain Monte Carlo (MCMC) algorithms for sampling from the linear extension decision tree.

For the hotspot detection component, we have been developing the following refinements:

- GLM-based covariate adjustments for assessing importance of potential explanatory variables
- Hierarchical spatial models to account for spatial dependence
- Data mining using topdown-bottomup clustering for aggregation of spatial units
- Criminal network analysis

2.1 Prioritization Component
The prioritization component of the geoinformatic surveillance project is concerned with the question of ranking a finite collection of objects when a suite of indicator values is available for each member of the collection.

The goal of the prioritization system is to canonically transform a partial order into a linear order of the objects. We propose a novel prioritization scheme based on multiple indicators that does not require reduction of the data to a single index.

2.2 Hotspot Detection Component
The hotspot component of the geoinformatic surveillance project is concerned with the question of identifying, delineating, and assessing the significance of hotspots.

Our approach to hotspot detection is based on the spatial scan statistic (Kulldorff and Nagarwalla 1995; Kulldorff 1997), which has been widely adopted in the health sciences for disease

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

surveillance. This tends to produce zones that are relatively compact and roughly circular in shape. We propose a novel development of an upper level set (ULS) algorithm for selecting the candidate zones in an adaptive (date-driven) manner. The ULS approach allows for arbitrarily shaped hotspot candidate zones.

Our efforts are driven by a wide variety of case studies of potential interest to Federal agencies 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.

For additional information regarding our project, see http://www.stat.psu.edu/hotspots/, http://www.stat.psu.edu/~gpp/, demo and poster at this conference, and http://www.digitalgovernment.org/news/stories/2004/1104/1104_hotspots_heyman.jsp\.

3. PARTNERSHIPS

CoPrincipal Investigator(s): Raj Acharya; Amy K Glasmeier; Wayne L Myers; Shashi Phoha.

Senior personnel: Robert Brooks; Denice Wardrop; Lance Waller; Elizabeth Middleton; James Shortle; Reza Modarres; Stephen Rathbun; Charles Taillie.

Other collaborators: Howard Burkom; Lawrence Cox; John Kelmelis; Martin Kulldorff; Bo Ranneby; Phil Ross; and others.

3.1 Training and Development Component

Crossdisciplinary classroom for surveillance geoinformatics and multiscale advanced raster map analysis (Ecometrics and Environmetrics). Instructor: G. P. Patil, PI.

An advanced crossdisciplinary graduate course across the academic year for graduate students from different programs on the campus, but with common interest in surveillance geoinformatics and multiscale advanced raster map analysis with emphasis on ecometrics and environmetrics.

3.2 Outreach Component

1. Surveillance Geoinformatics Forum: A partnership in the making.

2. Multiscale Advanced Raster Map Analysis System Partnership: A partnership in the making.

3. Synergistic Outreach

Professor Patil, the PI, was invited to the following multidisciplinary community forums to address on, 'Biosurveillance Geoinformatics of Hotspot Detection and Prioritization for Biosecurity' during the reporting period.

RTI International, Research Triangle Institute, Research Triangle Park, NC. The First RTI Fellows Symposium on Homeland and Health Security.

Washington Statistical Society, Washington, DC. An insightful panel discussion followed with knowledgeable panelists group of ten from interested agencies and academia. Dr. Larry Brandt,Director,NSF Digital Government Program,was the moderator.

George Mason Universtiy, School of Computational Sciences, Fairfax, VA.

Workshop on Statistics and Counterterrorism, National Institute of Statistical Sciences, New York City.

National Institute of Standards and Technology, Information Technology Laboratory, Washington, DC

Cleveland Clinic, Division of Biostatistics and Epidemiology, Cleveland, OH.

Workshop on Surveillance GeoInformatics, International Conference on EcoHealth for Emerging Infectious Diseases, Ecosystems, and Social Systems, Honolulu, HI.

Environmental Statistics Forum at the International Conference on the Future of Statistical Theory, Education, and Practice, Hyderabad, India.

Joint Statistical Meetings, Special Invited Sessions on Surveillance GeoInformatics, New York City, San Francisco, Toronto, Minneapolis.

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] Kulldorff, M. A spatial scan statistic. *Communications in Statistics: Theory and Methods*, 26, (1997), 1481-1496.

[2] Kulldorff M, and Nagarwalla, N. Spatial disease clusters: Detection and inference. *Statistics in Medicine*, 14 (1995), 799–810

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

[4] 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.

[5] Patil, et al. Upper level set scan statistic system for detecting arbitrarily shaped hotspots for digital governance. Poster presentation, dg.o2005, Atlanta, Georgia.

[6] Patil, et al. Geoinformatic surveillance of hotspot detection, prioritization and early warning. Demo, dg.o2005, Atlanta, Georgia