

**GEOINFORMATIC SURVEILLANCE FOR HOTSPOT DETECTION AND PRIORITIZATION  
Innovation with Epsilon Machines, Formal Language Measures, Upper Level Set Scans,  
Partially Ordered Set Prioritizations, Decision Support Systems,  
and Virtual Situation Room Servers**

**Table of Contents**

The following Opening Information pages cover:

(1) Introduction and Motivation.....	2-3
(2) Information Technology and Its Novel Application....	3-4
(3) Proposed Research Group.....	5-6
(4) PI, Co-PI, and Co-I.....	7
(5) Theory and Practice Advisory Council.....	8
(6) Biosketches.....	9-10
(7) Case Studies.....	11-13
(8) Decision Support System.....	14-15
(9) Management Plan and Structure.....	16-17

**G. P. Patil**

**Center for Statistical Ecology and Environmental Statistics**

**Department of Statistics**

**The Pennsylvania State University**

**University Park, PA 16802**

**<http://www.stat.psu.edu/~gpp>**

**Research, Outreach, and Advisory Team**

**Partnership on PP 5--8**

**GEOINFORMATIC SURVEILLANCE FOR HOTSPOT DETECTION AND PRIORITIZATION  
Innovation with Epsilon Machines, Formal Language Measures, Upper Level Set Scans,  
Partially Ordered Set Prioritizations, Decision Support Systems,  
and Virtual Situation Room Servers**

## 1. Introduction and Motivation

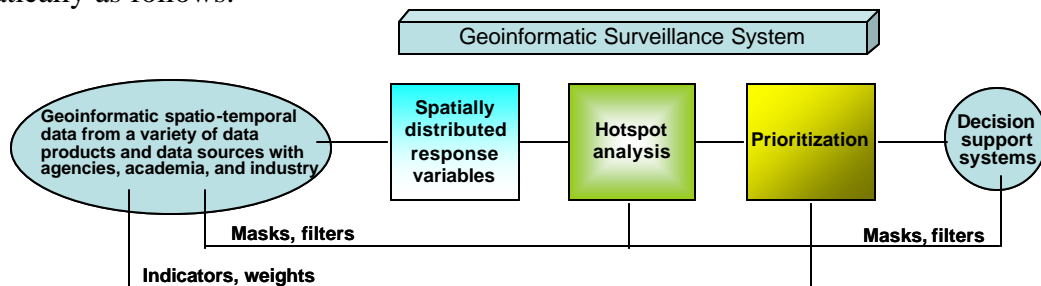
A primary purpose of this proposal is to invent, implement, and interface innovative information technology (IT) for the much needed geoinformatic surveillance decision support system for hotspot detection and prioritization in the project-networked virtual situation room capable of online interaction, cross-cutting solution, and dynamically updated communication with application partners, educational users or decision makers involved in a real situation.

Geographic surveillance for hotspot detection and delineation has become an important area of investigation both in geospatial ecosystem studies and in geospatial public health studies. In order to find critical areas based on synoptic cellular data, geospatial ecosystem investigations applied recently discovered echelon tools (Myers et al 1997, 1999). In order to find elevated rate areas based on synoptic cellular data, geospatial public health investigations apply recently discovered SaTScan, circle-based spatial scan statistic tool (Kulldorff, 1997; Kulldorff and Nagarwalla, 1995). The PI (Patil, 2003; Patil, Balbus et al. 2003; Patil, Bishop et al., 2002, 2003) has conceptualized a joint role for these together in the spirit of a cross-disciplinary cross-fertilization to accomplish more effective and efficient geographical surveillance for hotspot detection, and early warning system.

Clearly, clusters are clusters. They can be of any shape, and cannot be captured only by circles. This is likely to give more false alarms and more false negatives than warranted. What we need is the capability to detect arbitrarily shaped clusters and the ability to handle network-based as well as cell-based data. The upper level set scan system innovation will fill this need and provide a timely next generation hotspot detection and delineation system (Patil, 2002; Patil, Myers et al 2002; Myers, Kurihara et al 2002, 2003; Patil, Brooks et al 2001, 2002). Also see Patil, Balbus et al. (2003) for a broad perspective of multiscale advanced raster map analysis system of which hotspot detection is a part.

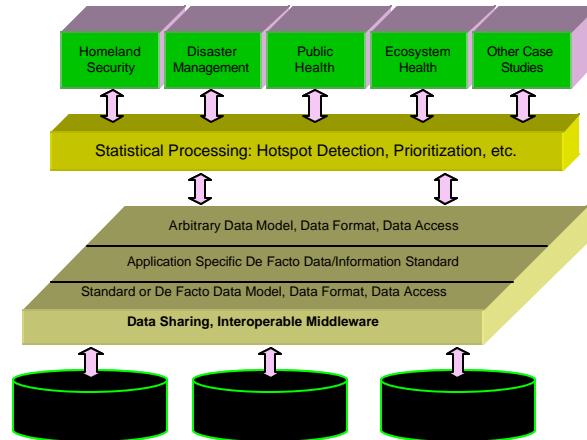
The significance and the timeliness become clearer as we witness various reports and action plans of various federal, state, and local agencies and prestigious foundations and academies, suggesting geographic and network surveillance for arbitrarily shaped hotspots, using next generation of sophisticated hotspot detection and prioritization tools. For example, a recent NRC report on making the nation safer: the role of science and technology in countering terrorism.

While the proposed research derives its particular significance within the context of national homeland security, it has powerful place within the much broader infrastructure of science and technology. Major information flows in the geoinformatic surveillance can be represented schematically as follows:



The case studies in this project address a broad range of national applications such as homeland security, biosecurity, disaster management, public health, ecosystem health, water management,

carbon budget, coastal management, community infrastructure, etc. The geographic information sharing middleware will provide the component to support distributed, dynamic data-driven applications and case studies, and enhance the system security and stability. This middleware will access appropriate databases for supporting the case-studies (see Figure below).



Our team involves researchers with solid track records in several complementary areas that are at the core of this project. We will integrate the resulting advances into a prototype system applied to a rich set of large-scale case studies. Project goals and results will be achieved in a well-integrated disciplinary and cross-disciplinary effort coupled with matching educational abilities, leading to an emergent software system. It will help strengthen the methodology and technology infrastructure needed nationally for hotspot detection and prioritization across geographic regions and across networks in the 21<sup>st</sup> century, and provide a basic foundation to an envisioned National Center serving society's need for geoinformatic surveillance.

## 2. Information Technology and Its Novel Application

Current methods to organize, represent, and process large bodies of complex information spread over space and time are inadequate for today's decision making needs, especially in a time of crisis. Advances are needed in methods of quickly and accurately recognizing and prioritizing critical changes in important parameters that are masked by fluctuations. We propose research that will address these needs in crisis situations, as well as the non-crisis infrastructure needs of science and technology. Our project will conduct fundamental information science and technology research and its novel application to geoinformatic surveillance for hotspot detection and prioritization. A hotspot means something unusual—an anomaly, aberration, outbreak, elevated cluster, critical area, etc. The declared need may be for monitoring, etiology, management, or early warning. Responsible factors may be natural, accidental, or intentional.

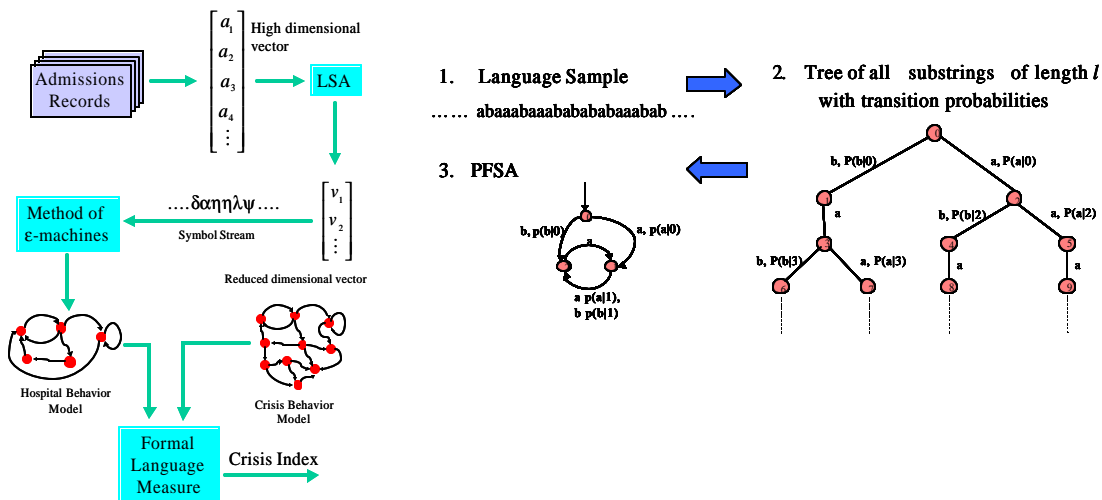
The most innovative aspect of this research develops *upper level set scan statistic theory* to recognize arbitrarily shaped hotspots. Spatio-temporal data are integrated with a new level of accuracy providing more sensitive indicators of changes in critical parameters. The technique applies not only to physical space, but also to connected collections of objects or regions, i.e. networks. A second innovation is the development of partially ordered set prioritization theory to rank hotspots without having to integrate multiple indicators into a single index. A third is a new method of automated knowledge acquisition in the form of *behavior recognition technology* built on the concept of *e-complexity* and *e-machines* from Statistical Physics and a *formal language measure* from Discrete Event Control Theory. Local behaviors can now be compared to known

behaviors using traditional pattern-matching techniques for classification. Behaviors are represented symbolically by formal languages in a form that can be used directly for automated decision aides in the form of discrete event controllers.

Our research consists of three parts. First, fundamentally new information technologies are developed from advances in statistics, statistical physics and control theory. A new level of sensitivity is attained for recognizing and responding to critical changes in noisy, chaotic environments. Second, the technological advances are proven in test cases covering a broad range of critical situations. The range of applications demonstrates the fundamental nature of the new technologies. Third, we will move our advances into society by building prototype *situation room servers*. The servers will integrate complex distributed data sources for selected applications. These servers and the new tools they make available will revolutionize crisis prediction and management.

Toward the end of the grant, we will find interested agencies and make the technology available in an ongoing, operational capacity. This will ensure that the benefits of our research will have a long-term impact on society. The project will also have a strong educational component with technology transfer, outreach, and built-in evaluation.

\*\*\*\*\*



**Figure 1. (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  $\epsilon$ -machine procedure for converting an event stream into a parse tree and finally into a probabilistic finite state automaton (PFSA).

\*\*\*\*\*

### **3. Proposed Research Group**

#### **PI:**

G. P. Patil, PSU

Geographical Hotspot Detection, Prioritization, and Improvement; Mathematical and Environmental Statistical Systems; Technology Transfer and Outreach

#### **Co-PI: Information Technology**

Menas Kafatos, GMU

Earth Observing and Space Research, Database Management, Decision Support System

Shashi Phoha, PSU

Information Science and Technology, Computer and Electrical Engineering, Mobile Sensor Networks, e -Machines, Computational Mechanics

#### **Co-PI: Science and Technology**

Martin Kulldorff, UCONN

Public Health, Spatial Scan Statistic Software, Syndromic Surveillance

Wayne Myers, PSU

Landscape Ecology, Remote Sensing, Critical Area Detection, Software

#### **Senior Computer Scientists and Information Technology Experts**

Raj Acharya, PSU

Computer Science and Engineering, Data Mining, Information Fusion, and Visualization

John Yen, PSU

Intelligent Agents, Team-Based Intelligent Agents, Soft Computing

#### **Case Study Scientists at Penn State**

1. Paul Backman, Plant Pathology
2. Robert Brooks, Forest Resources
3. Barry Evans, PASDA
4. David Friedlander, Applied Research Lab
5. Amy Glasmeier, Geography
6. Eugene Lengerich, Epidemiology
7. Henry Lin, Crop and Soil Sciences
8. Elizabeth Marshall, Spatial Economics
9. David Mortensen, Agronomy
10. Gary Petersen, Remote Sensing
11. Stephen Rathbun, Statistics
12. Charles Taillie, Statistics
13. Denice Wardrop, Forest Resources

**Case Study Scientists at George Mason**

1. Zafer Boybeyi, Computational Science and Environment
2. Yuechen Chi, Decision Support System
3. Menas Kafatos, Virtual Situation Room
4. Xue Liu, Computational Science and Environment
5. Phil Yang, Emerging Infectious Disease Cluster Detection

**Case Study Scientists at Other Academic Institutions**

1. Charles Hopkinson, MBL
2. Siamak Khorram, NCSU
3. Ken Kleinman, Harvard
4. Stephen Prince, UMD

**Case Study Scientists at Federal Agencies**

**Present**

1. Robert Knox, NASA
2. Thomas Loveland, USGS
3. Nancy Maynard, NASA
4. Anthony Olsen, EPA
5. Christopher Portier, NIH
6. Phillip Ross, EPA

**Planned**

1. Richard Birdsey, USDA
2. Lawrence Cox, CDC
3. Charles Dull, USDA
4. Christopher Elvidge, NOAA
5. John Kelmelis, USGS
6. Kamlesh Lulla, NASA
7. Dale Quattrochi, NASA
8. K. Thirumalai, DOT

**Case Study Scientists: International**

1. Luiz Duczmal, Brazil
2. Marie-Josée Fortin, Canada
3. Koji Kurihara, Japan
4. Bo Ranneby, Sweden
5. Orazio Rossi, Italy



## **5. Theory and Practice Advisory Council**

### **1. Significant Partner in the Definition and Execution of the Proposed Work**

John Kelmelis, USGS

### **2. Theory and Practice Advisory Council (TAPAC)**

John Kelmelis, USGS

### **3. User Advisory Council (UAC)**

1. Christopher Portier, NIH (Chair)
2. Charles Dull, USDA
3. Chris Elvidge, NOAA
4. John Kelmelis, USGS
5. Robert Launer, DOD
6. Kamlesh Lulla, NASA
7. Nancy Maynard, NASA
8. Phillip Ross, EPA
9. Ashbindu Singh, UNEP
10. K. Thirumalai, DOT

### **4. Science Advisory Board (SAB)**

1. Michael Goodchild, UCSB (Chair)
2. Richard Birdsey, USDA
3. Howard Burkom, JHU
4. Larry Cox, CDC
5. Joel Hersh, PADOH
6. Joseph JaJa, UMD
7. John Kelmelis, USGS
8. Andrew Lawson, USC
9. Thomas H. Mace, NASA
10. Anthony Olsen, EPA
11. Dale Quattrochi, NASA
12. David Rapport, UG

### **5. User Organization and Case Study Partnership**

1. Richard Birdsey, USDA
2. Lawrence Cox, CDC
3. Charles Dull, USDA
4. Richard Heffernan, NYCDOH
5. John Kelmelis, USGS
6. Robert Knox, NASA
7. Anthony Olsen, EPA
8. Christopher Portier, NIH
9. Phillip Ross, EPA
10. Ashbindu Singh, UNEP

## 6. Biosketches

This section gives biographical sketches of the senior project partnership, largely excluding the ones that have sent letters of participation in their role of leadership support. The biographical sketches appear according to the nature of the project participation.

### **PI:**

G. P. PATIL

(Geographical Hotspot Detection, Prioritization, and Improvement; Mathematical and Environmental Statistical Systems; Technology Transfer and Outreach)

### **Co-PI: Information Technology**

MENAS KAFATOS

(Earth Observing and Space Research, Database Management; Online Decision Support System)

SHASHI PHOHA

(Computer and Electrical Engineering, Informatics, Mobile Sensor Networks, Security)

### **Co-PI: Science and Technology**

MARTIN KULLDORFF

(Syndromic Surveillance, Spatial Scan Statistic, SaTScan)

WAYNE L. MYERS

(Hotspot Detection, Critical Area Analysis, Landscape Ecology, Remote Sensing)

### **Senior Computer Scientists and Information Technology Experts**

RAJ ACHARYA

(Computer Science and Engineering, Data Mining, Information Fusion, and Visualization)

JOHN YEN

(Intelligent Agents, Team-Based Intelligent Agents, Soft Computing)

### **Case Study Scientists at Penn State**

PAUL A. BACKMAN

(Biosurveillance and Biosecurity, Plant Pathology)

ROBERT BROOKS

(Wetlands, Forest Resources, Geographical Ecology)

BARRY EVANS

(Remote Sensing, GIS, GeoDecisions, Water Management)

DAVID FRIEDLANDER

(Computer and Electrical Engineering, Informatics, -Machines)

AMY K. GLASMEIER

(Economic Geography and Regional Planning, Poverty Research, Environmental Justice) 26

EUGENE J. LINGERICH

(Syndromic Surveillance, Epidemiology, Bioinformatics)

HANGSHENG (HENRY) LIN

(Soil, Geography, GIS)

ELIZABETH MARSHALL

(Environmental Economics, Spatial Economics)

DAVID A. MORTENSEN

(Weed Ecology, Invasive Species)

GARY W. PETERSEN

(Water Management, Remote Sensing)

STEPHEN RATHBUN

(Geospatial Statistics, Computational Statistics, Landscape Ecology, Environmental Statistics)

CHARLES TAILLIE, JR.

(Computational Statistics and Statistical Computing)

DENICE HELLER WARDROP

(Systems Analysis, Watershed Assessment, Biological Impairment)

**Case Study Scientists at George Mason**

ZAFER BOYBEYI

(Computer Science, Weather Prediction)

YUECHEN CHI

(Computer Science, Decision Support System)

XUE LIU

(Geomatics, Remote Sensing and Mapping)

PHIL YANG

(Web Development and Telecommunications)

**Case Study Scientists at Other Institutions**

CHARLES S. HOPKINSON, JR.

(Coastal Management, Water Management and Conservation)

SIAMAK KHORRAM

(Computer and Electrical Engineering, Remote Sensing)

KEN P. KLEINMAN

(Disease Surveillance, Crisis Prediction)

STEPHEN PRINCE

(Regional Earth Systems Science Applications, Coastal Studies)

ROBERT KNOX

(Forest Ecosystems Dynamics, Analysis, and Models: Environmental Resource Evaluation)

ANTHONY R. OLSEN

(Environmental Monitoring and Assessment)

**Case Study Scientists: International**

LUIZ DUCZMAL

(GIS, Geospatial Urban Crime Analysis, Simulated Annealing)

MARIE-JOSEE FORTIN

(Geospatial Boundary Analysis, Spatially Restricted Clustering)

KOJI KURIHARA

(Regional Analysis, Echelon Analysis, Spatial Scan Statistics)

BO RANNEY

(Geomatics of Natural Resources, Critical Area Detection)

ORAZIO ROSSI

(Landscape Ecology, Biodiversity, Change Detection)

ASHBINDU SINGH

(Social Geography, Regional Planning)

**Graduate Students**

JOSEPH A. BISHOP

(GIS, Landscape Analysis, Freshwater Stream Network)

YUEMEI WANG

(Computer Science, Chemistry, Computational Statistics)

## 7. Case Study Descriptions

The proposed case studies will involve ongoing project investigations of interested project scientists with support from their sponsors to conduct research requiring hotspot detection and/or prioritization in their work. The circle-based scan statistic may be in use. The composite-index-based ranking may be in use. Or the investigations maybe simply waiting to identify appropriate techniques and tools. The following case studies will thus serve the purpose of development, refinement, demonstration, and validation of the proposed upper level set scan statistic, the partially ordered set prioritization, and the  $\epsilon$ -machine, as appropriate.

The proposed case studies cover important national applications such as: homeland security, biosecurity, disaster management, sensor networks, public health, syndromic surveillance, geographic disease surveillance, ecosystem health, coastal management, carbon budgets, climate change, invasive species, community infrastructure, environmental justice, urban heat islands, urban crime spots, etc.

The following tabulation gives the proposed case studies grouped according to their national applications. Items in the **Case Studies** column are hyperlinked: on-line viewers may use CTRL+CLICK to jump to the case study in question.

<b>National Application</b>	<b>Case Studies</b>
<b>Biosurveillance</b>	<a href="#">PCS 10</a> , <a href="#">UCS 11</a>
<b>Carbon Management</b>	<a href="#">PCS 6</a> , <a href="#">UCS 2</a>
<b>Coastal Management</b>	<a href="#">PCS 6</a> , <a href="#">PCS 7</a> , <a href="#">UCS 3</a> , <a href="#">UCS 6</a>
<b>Community Infrastructure</b>	<a href="#">UCS 3</a> , <a href="#">UCS 5</a>
<b>Disaster Management</b>	<a href="#">PCS 7</a> , <a href="#">PCS 14</a> , <a href="#">PCS 15</a> , <a href="#">UCS 10</a>
<b>Disease Surveillance</b>	<a href="#">PCS 16</a>
<b>Ecosystem Health</b>	<a href="#">PCS 1</a> , <a href="#">ICS 3</a> , <a href="#">ICS 5</a>
<b>Environmental Justice</b>	<a href="#">UCS 13</a>
<b>Environmental Management</b>	<a href="#">ICS 2</a> , <a href="#">ICS 4</a>
<b>Environmental Policy</b>	<a href="#">PCS 12</a> , <a href="#">ICS 6</a>
<b>Homeland Security</b>	<a href="#">PCS 3</a> , <a href="#">PCS 9</a> , <a href="#">PCS 15</a> , <a href="#">UCS 4</a>
<b>Poverty Policy</b>	<a href="#">PCS 11</a>
<b>Public Health</b>	<a href="#">UCS 12</a>
<b>Public Health and Environment</b>	<a href="#">UCS 5</a> , <a href="#">UCS 8</a> , <a href="#">UCS 9</a>
<b>Syndromic Surveillance</b>	<a href="#">PCS 9</a> , <a href="#">PCS 13</a>
<b>Urban Crime</b>	<a href="#">ICS 1</a>
<b>Water Management</b>	<a href="#">PCS 1</a> , <a href="#">PCS 2</a> , <a href="#">PCS 4</a> , <a href="#">PCS 8</a> , <a href="#">UCS 1</a> , <a href="#">UCS 7</a>

The listing on the next page gives the proposed case studies grouped according to the following categories:

- Prototype Case Studies ([PCS](#))
- User Organization Case Studies ([UCS](#))
- International Case Studies ([ICS](#))

Items in the listing are hyperlinked: on-line viewers may use CTRL+CLICK to jump to the case study in question.

## **Listing of Case Studies**

### **Prototype Case Studies (PCS)**

- PCS 1: Denice Wardrop, Wayne Myers, Ganapati Patil, and Charles Taillie  
Network-Based Analysis of Biological Integrity in Freshwater Streams
- PCS 2: Robert Brooks, Wayne Myers, G. P. Patil, and Charles Taillie  
Prioritization Model for Diagnosis of Watershed Impairment and Vulnerability
- PCS 3: Paul Backman, Barry Evans, and Gary Petersen  
Development of Remote Sensing Methods for Crop Bioterrorism
- PCS 4: Henry Lin, Ray Bryant, and Douglas Goodlander  
Identifying "Hotspots" Impacted by P-Based Regulations and P-Loss in Pennsylvania
- PCS 5: David Mortensen and Stephen Rathbun  
Early Detection and Delineation of Outbreaks of Invasive Plant Species
- PCS 6: Siamak Khorram, Thomas Loveland, Thomas Mace, and Ganapati Patil  
Accuracy Assessment of Regional Land Use and Land Cover Produced from Remotely Sensed Data
- PCS 7: Menas Kafatos and Yuechen Chi  
Oil Spill Hotspot Detection and Prioritization
- PCS 8: Henry Lin, Stephen L. Rathbun, Gary Petersen  
Soil Moisture Hotspot Detection and Prioritization for Water Management
- PCS 9: David Friedlander, G. P. Patil, Shashi Phoha, and Charles Taillie  
Emerging Hotspot Detection Using Networked Patient Records for Crisis Prediction
- PCS 10: Ken Kleinman  
Detection of Unusual Clusters of Disease, including Biological Terrorism Events, through Real-Time Analysis of Electronic Records Data
- PCS 11: Amy Glasmeier, Ganapati Patil, and Charles Taillie  
Examining Persistent Poverty Using Spatio-Temporal Typology of Hotspots
- PCS 12: Elizabeth Marshall, Ganapati Patil, and James Shortle  
Prioritization and Ranking for Environmental Policy Using Indicators Resulting from Data Envelopment Analysis and Stochastic Frontier Analysis
- PCS 13: Martin Kulldorff, David Friedlander, Eugene Lengerich, Ganapati Patil, and Charles Taillie  
Syndromic Surveillance for Biosurveillance and Biosecurity
- PCS 14: Zafer Boybeyi and David Friedlander  
The AVE-SESAME I: Atmospheric Variability Experiment—Severe Environmental Storms and Mesoscale Experiment
- PCS 15: Zafer Boybeyi, Menas Kafatos, David Friedlander, and Ganapati Patil  
Geoinformatic Surveillance for Chemical and Biological National Security Program—Salt Lake City Experiment
- PCS 16: Menas Kafatos and Phil Yang  
Mosquito Control, Surveillance Strategy, and Disease Prevention

### **User Organization Case Studies (UCS)**

- UCS 1: Anthony Olsen  
Hot Spot Identification of Stream Condition within a State
- UCS 2: Robert Knox  
Prototyping Simplified Methods of Mapping Priority Hotspots of Disturbance Using EOS Data

UCS 3: Stephen Prince

Mid-Atlantic Regional Earth Systems Science Applications Center and Hotspot Analysis

UCS 4: Shashi Phoha, Ganapati Patil, David Friedlander, Charles Taillie, and John Yen

Tasking of a Self-Organizing Oceanic Surveillance Mobile Sensor Network

UCS 5: Phillip Ross, G. P. Patil, and Charles Taillie

Assessing the State of Environment and Emerging Problems Using Multiple Indicators and Measures

UCS 6: Charles Hopkinson, James Morris, and Raymond Torres

Network-Based Analysis of Integrity of Coastal Intertidal Wetlands

UCS 7: Charles Hopkinson and Ganapati Patil

N Processing in the Ipswich River Basin as Revealed by Hotspot Identification and Analysis

UCS 8: Christopher Portier, William Farland and Ganapati Patil

Health and Environment Effects of Agent Orange and Dioxin in Vietnam

UCS 9: Christopher Portier, Dale Quattrochi and Ganapati Patil

Urban Heat Islands, Urban Sprawl, and Disease Evaluation

UCS 10: Charles W. Dull

Wildland Fire Management

UCS 11: Martin Kulldorff, Richard Heffernan, and Farzad Mostashari

Dead Bird Clustering: Comparison of ULS with the Cylinder-Based Space-Time Scan Statistics as Early Warning Systems for West Nile Virus Activity

UCS 12: Martin Kulldorff, Eugene Lengerich, Linda Pickle, and Ganapati Patil

Geography of Cancer: Clusters of Cancer Mortality

UCS 13: Chris Portier, Amy Glasmeier, and Ganapati Patil

Environmental Justice Studies in the Appalachian Region and Mississippi Delta

### **International Case Studies (ICS)**

ICS 1: Luiz Duczmal and Renato Assuncao, Brazil

Detecting Crime Hotspots in Brazil Using a Simulated Annealing Scan Statistic

ICS 2: Marie-Josée Fortin, Canada

Local Spatial Analyses of Simulated Forested Landscapes

ICS 3: Koji Kurihara, Japan

Spatial Structure and Hotspot of Japanese Ecosystem Health based on Spatial Scan and Echelon Systems

ICS 4: Bo Ranneby, Sweden

Detection and Location of Spatially Sparse Forest Events

ICS 5: Orazio Rossi, Italy

Map of Italian Nature: Critical Area Detection and Prioritization

ICS 6: Ashbindu Singh (UNEP), Ganapati Patil, and Charles Taillie

Prioritization involving Nationwide Human Environment Interface Worldwide at UNEP

## 8. Decision Support System

**Online Decision Support System:** This component of the project focuses on the development of software system to assist scientists to apply the hotspot detection and prioritization technology and application users to use the system. With dramatic technology advances in online storage and internet networking, more and more data are immediately and easily available and accessible. This enables the development of online data services in the data system, in which data system processes the user's request and delivers the results to the user directly. Our partner, George Mason University (GMU), has been developing a software system for Earth science data management and online services. A key function of the system is to enable service providers to define and register a data processing service in the system and their users to use the service to act upon the data represented in the system. GMU also develops the standard of data process development, and the communication protocol between the system and the data service.

The project will use this software system to collect and manage the information of data and data processes. The software system has user interfaces of Web, API, and query language. We will develop the web-based user interface to provide wide interactive access to the system and timely update the dynamical information in the underlying database. We will also use and further develop the software system to enable scientists to define and register their hotspot detection and prioritization data processes as data services in the system, and to enable the application users to use the system in support of their decisions in response to the hotspots. This will form an online Geoinformatic surveillance decision support system.

In the data system, we will create database tables to record data usage and data service activities to track system usage for performance metrics measurement. The system will also record user comments and satisfaction levels for our decision support system.

**Computational Structure, System Integration, and Database Management:** This component of the project focuses on the development of efficient data structures and algorithms coupled with efficient visualization techniques for hotspot detection and prioritization using upper level set scans and partially ordered set prioritizations. In fact, recently the problem has been for quickly identifying regions for multivariate maps for which a number of geospatial parameters satisfy certain conditions. See JaJa and Shi (2001), K.-S. Yang et al. (2001), and R. Yang et al. (2001). This project will extend these techniques in directions dictated by the proposed scanning techniques and prioritization tools.

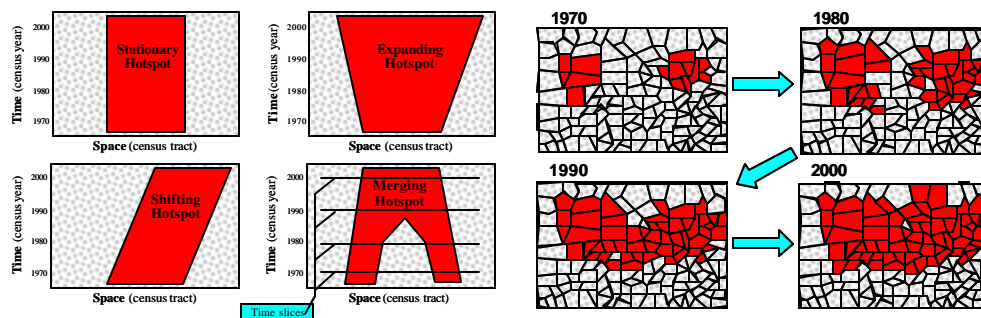
**Information Visualization, User Interface Design and GIS Linkage:** A key goal of this effort is to develop a visualization interface integrated with software tools based on various statistical models and techniques developed in this project. Information visualization and interface design are critical to making effective use of the various models and techniques. Our goal will be to promote the discovery of inherent structures and patterns, build and test hypotheses, enable the detailed study of particular facets and dimensions of the data, and provide means to visually assess the utility and accuracy of the statistical and computational techniques developed. Our approach will be to work with applications partners to identify their needs and frequent tasks. A phased implementation will allow us to implement simple algorithms at first and then embed more sophisticated algorithms. In conjunction with case-study validations and initial demonstrations, we will conduct usability tests to refine the interfaces and demonstrate efficacy.

**Online Information System:** This component of the project focuses on the development of an information system for communication among the partners and for communication with the general users. We will develop the project web site, which will describe significant information products involving the project and its progress. It will provide central access point to locate such information, and identify opportunities to provide comments and other feedbacks during the development of some of these products. The main capabilities of the information system include:

- Tight interactions with operational users over several phases, e.g., project formulation, defining the products and services, testing-validation-refinement, and technology transfer. On-going community involvement supported through a web interface for communicating about hotspots and an interactive system for knowledge sharing and discovery.
- Support for spatial, temporal, and/or subject matter queries of hotspot data products, associated metadata, comments, and documents.
- Solicitation and archiving of responses, comments, etc., tagged with links to the relevant products, locations, documents, and/or time points (product date(s), as well as the usual header information for the communication).
- Development of a path for feedback about hotspot data and priority rankings to flow back to how indicators are ranked and combined.
- Evaluation of the hotspot interface using different views, depending on application area and role in the project (investigator, application partner, education partner, student, resident/citizen of a particular geographical region, etc.).

**Online Virtual Situation Room:** Public and educational involvement in evaluating hotspot products will be a vital complement to more formal or structured feedback from application partners. The information system will be accessible not only by project partners and target application users but also widely distributed general public users. This information system will form the virtual situation room for distributed partners and general users benefit from the timely interaction and communication involving the real life situations at their end.

\*\*\*\*\*



**Figure 3. The four diagrams on the left depict different types of space-time hotspots. The spatial dimension is shown schematically on the horizontal and time is on the vertical. The diagrams on the right show the trajectory (sequence of time slices) of a merging hotspot.**

\*\*\*\*\*

## 9. Management Plan and Structure

The overall management of the project will be the responsibility of the PI, Dr. G. P. Patil. He brings a substantial administrative experience in managing crossdisciplinary research initiatives and large projects. The management team of the project will consist of the PI and the co-PIs who will set short-term and long-term directions and goals, implementation plans, assess project progress, and establish collaborative mechanisms among the participating investigators. The management and research team enjoys the kind of cross-disciplinarity the current solicitation would like to see. The goals stated in the section on Project Outcomes, Partner Synergy, and Workplan will provide initial guidance for the management team. The management team will be in constant communication through email and phone calls, and will meet twice a year in conjunction with the two annual workshops planned for the project. These meetings will focus on assessing progress, adjusting goals and directions as appropriate, and setting new goals.

Moreover, each member of the management team will lead the coordination of the research efforts in a thrust area as follows:

- Geoinformatic Surveillance – Dr. Patil
- Information Science and Technology – Dr. Phoha, Dr. Acharya, and Dr. Yen
- Data Mining, Information Fusion, and Visualization – Dr. Acharya
- Upper Level Set Detection System – Dr. Taillie
- Partially Ordered Set Prioritization System – Dr. Taillie
- Public Health Applications – Dr. Kulldorff
- Environmental and Ecological Applications – Dr. Myers
- Decision Support System – Dr. Kafatos
- Virtual Situation Room – Dr. Kafatos, Dr. Patil, and Dr. Phoha  
 Disaster and Dispersion: Kafatos; Syndromic Surveillance: Friedlander; Sensor Networks: Phoha; Disease Clusters: Kulldorff; Ecosystem Critical Areas: Wardrop; Poset Prioritization: Taillie
- Education, Dissemination, and Outreach – Dr. Patil

The two planned workshops, one in the middle of the academic year and the other in the summer, will be an integral part of the project, which will allow investigators to describe their work and how it relates to the overall goals of the project, and to assist the investigators to better integrate their expertise to help evolve the upper level set scan statistic system and the partially ordered set prioritization system with prototype case studies, and the national and international refinement and validation case studies. The summer workshop will include special tutorial sessions covering the system and the methodologies.

Most of the proposed case studies have been carefully selected as ongoing funded projects whose successful completion would significantly benefit from application of the proposed toolkit for hotspot detection and prioritization. Thus, the sponsor will have served as catalytic sponsor of methodology and technology directed toward current and future opportunities and challenges.

The proposed project has versatile partnership. It also has versatile leadership. It will have Theory and Practice Advisory Council (TAPAC) consisting of Science Advisory Board (SAB) and User Advisory Council (UAC). It will ensure that science and user interest are represented

throughout the project. It will also be a strong source of relevant domain expertise. It is fortunate to have fitting triple leadership with John Kelmelis, Chief Scientist for Geography at USGS as the TAPAC Chair; Mike Goodchild, National Academy Member and Director of the NSF Center for Spatially Integrated Social Science as SAB Chair; and Chris Portier, Director of the National Environmental Toxicology Program at NIH as UAC Chair. The TAPAC will meet once a year during the summer workshop, evaluate project progress, and advise the management team on future direction.

Partner federal agencies include CDC, DOD, DOT, EPA, NASA, NIH, NOAA, USDA, and USGS with USGS as the Coordinating Agency. The TAPAC will meet within the first ninety days of the Project to help initiate a Partners Plan for suitable funding and strategic goals for the envisioned National Center for Geosurveillance upon Project completion. The Project will play the role of embryonic National Center during its fourth and fifth year duration. The Partners Plan will help match the requested project budget with costshare and kindshare direct with PI, Co-PI, Co-Investigators, and/or Case Study Scientists. Agency needs for this type of toolkit are sufficiently pressing, and we will look forward to full operational adoption, once its efficacy has been demonstrated. Based on the feedback we have, we are optimistic. We also expect partnership and leadership to grow to serve this timely mission.

Finally, the concept of mobility and interactive visits will be fully explored and implemented between participating faculty, graduate students, and postdocs across the three universities. Each university group will have weekly miniseminar(s) on relevant themes of the project involving local faculty, graduate students, and post docs. These will be carefully strengthened from time to time with visiting collaborators from participating institutions to keep the individual and collective momentum and synergy in progressive development. Every effort will be made to iteratively accomplish the upward spiral of horizontal and vertical research and training integration.

**International Collaboration.** This project will have an impressive international collaboration with scientific leaders responsible for major geospatial landscape level spatio-temporal programs in their countries such as Brazil, Canada, India, Italy, Japan, and Sweden. The international synergies and benefits to be gained from these collaborations include the considerable enrichment of our case studies that will strengthen the empirical and validation aspects of our project. The PI has already given a series of training courses to the Italian Map of Nature and will be giving a start off workshop in Sweden in the summer to help initiate their case studies as an integral part of the Swedish Program of Remote Sensing for the Environment. The Supplement to this proposal (Section I) contains support letters and describes the international collaborations.

**Performance Metrics and Built-in Evaluation.** Beyond reporting project inputs, outputs, and the number and kinds of applications supported, and their status, we will develop ways of measuring and reporting the individual project outcomes and the impact of the project. This will be done in a timely manner on an annual basis.