GEOINFORMATIC SURVEILLANCE DECISION SUPPORT SYSTEM
Geographic and Network Surveillance for Arbitrarily Shaped Hotspots
Next Generation of Geographic Hotspot Detection, Prioritization, and Early Warning with Emerging Hotspots

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

GEOINFORMATIC SURVEILLANCE DECISION SUPPORT SYSTEM
Geographic and Network Surveillance for Arbitrarily Shaped Hotspots
Next Generation of Geographic Hotspot Detection, Prioritization, and Early Warning with Emerging Hotspots

Geoinformatic Surveillance Decision Support System: Geographic and Network Surveillance for Arbitrarily Shaped Hotspots—Next Generation of Geographic Hotspot Detection, Prioritization, and Early Warning with Emerging Hotspots

Abstract

Our proposed system addresses declared needs for geoinformatic surveillance, software infrastructure, and a decision support system for spatial and spatio-temporal hotspot detection and prioritization. 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.

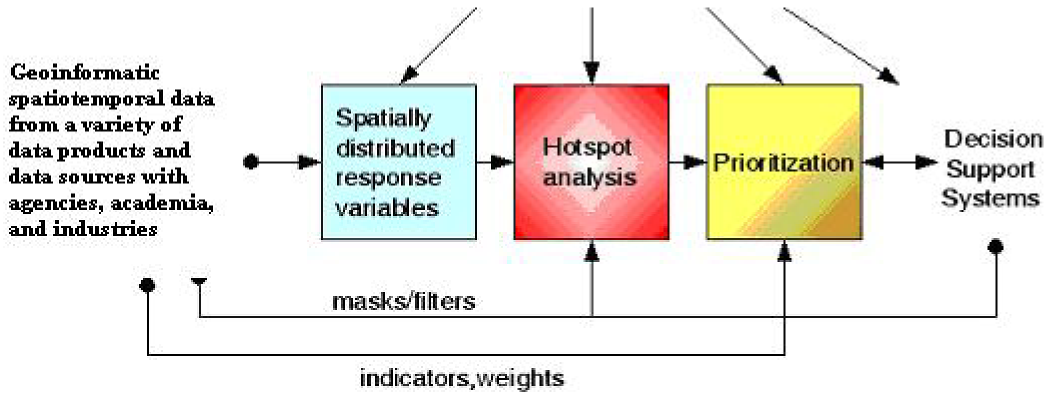
This effort will develop novel methods and tools for hotspot detection and prioritization across geographic regions and across networks, based on cutting-edge statistical research and using various data sources. This innovative investigation will apply, adapt and validate methods and tools for national applications, such as public health, water management and conservation, carbon management, invasive species management, homeland security, among others, leading to a sophisticated next generation analytical and computational decision support system, beyond the present day health-area-based SaTScan. Our innovation employs the notion of upper level set, and is accordingly called the upper level set scan statistic system.

We will also apply, adapt, and validate a prioritization innovation that provides capability for prioritization and ranking of detected hotspots based on multiple criteria indicators without integration of indicators into an index. It employs Hasse diagrams for visualization, and partially ordered sets and MCMC for analysis. Resulting prioritizations and rankings reflect the contrasting requirements of each application and also of varying stakeholders, while employing a common framework for hotspot detection, using data available at high temporal resolutions, and prioritization, using multiple data sources.

To provide wide interactive access to dynamically updated knowledge, the project will have online informational and instructiuonal products bulletin, knowledge discovery, and discussion forum, featuring information visualization, user interface design, and GIS linkage as part of the geoinformatic surveillance decision support system. The project has plans for a strong educational component.

A. Introduction, Relevance, and Approach

Building on important developments in spatial analysis and fundamental advances in statistical research, and catalyzed by the broad outreach of the Penn State Center for Statistical Ecology and Environmental Statistics, a diverse partnership has gathered to develop cross-cutting solutions that address issues common to multiple national applications. The focal point is a toolkit for hotspot detection and prioritization that provides a new way to rigorously use the spatial information available in a variety of national data products. This new information technology will be validated and demonstrated using a suite of case studies. Moreover we will use a software system to represent the new capabilities as registered services for operating on various data sources and to manage and measure the knowledge generated as applications partners and educational users explore the new capabilities and gain insight. Through interfaces with existing decision support processes, demonstrated with case studies, the project will provide decision makers with interactive access to dynamically updated knowledge of the geospatial systems. The major information flows in the geoinformatic surveillance decision support systems can be represented schematically as follows:

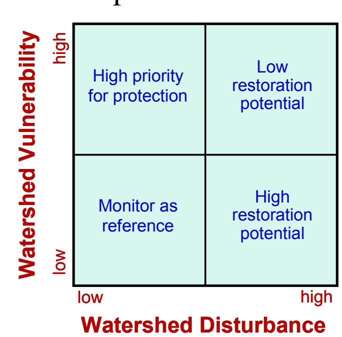


B. Illustrative Applications and Case Studies

Broadly speaking, the proposed geosurveillance project identifies several case studies important for the national applications. In this section, we provide illustrative applications and case studies.

Network analysis of biological integrity in freshwater streams. This study will employ 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. The state Department of Environmental Protection is determining indices of biological integrity (IBI) at about 15,000 sampling locations across the Commonwealth. Impairment will be measured by a complemented form of these IBI values. We will also use remotely sensed landscape variables and physical characteristics of the streams as explanatory variables in order to account for impairment hotspots. Critical stream subnetworks that remain unaccounted for after this filtering exercise become candidates for more detailed modeling and site investigation. See Evans et al. (2003) Hawkins et al. (2000) and Wardrop et al. (2002, 2003).

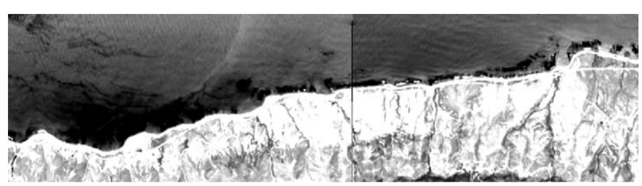
Watershed prioritization for impairment and vulnerability. This study will develop a prioritization model for watersheds (12-digit HUCs) of the Mid-Atlantic Highlands. A suite of indicators will be identified to assess each watershed's susceptibility to impairment (**vulnerability**). A second suite of indicators will measure actual stress or **disturbance** for each watershed. The watersheds will then be ranked according to each of the two separate sets of indicators. The proposed prioritization methodology will be used for ranking purposes. Each watershed is thus assigned a pair of ranks indicating its vulnerability status and its disturbance status. The pairs of ranks yield a scatter plot in the $\text{disturbance} \times \text{vulnerability}$ plane. The four quadrants in this plot have distinctly different management implications, as depicted in the accompanying diagram. Disturbance will be measured by stressor variables such as: excess sediment, riparian degradation, mine drainage, excess nutrients, exotic species,



agriculture (esp. on slopes), road crossings, forest fragmentation, and indices biological impairment. Vulnerability primarily reflects physical characteristics and natural features of the watershed and can be measured by: hydrogeomorphology (HGM), climate, aspect, slope, stream sinuosity, soil type, bedrock, and water source. Products include: a procedure for classifying watersheds by their features and condition, a taxonomy of Mid-Atlantic watersheds, and a set of monitoring and restoration options for each watershed class that can assist managers in developing TMDL (total maximum daily load) plans. See Brooks et al. (1998), Constantz (1997, 2000), Johnson et al. (2001), O'Connell et al. (2000), Patil et al. (2000).

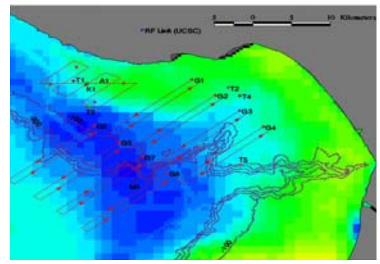
Mapping priority hotspots of vegetative disturbance for carbon budgets. Hotspot detection can complement existing approaches to remote measuring and mapping vegetation disturbance for global change research. Existing data products either strive to reduce 'false alarms' by relying on multi-year comparisons of matched 'best quality' data (see Strahler et al. 1999, Zhan et al. 1999, 2000) or restrict information to one type of disturbance (e.g., MODIS fire products). National and global carbon budgets, at time scales relevant to inversion of atmospheric transport models, require data that are both timelier and more comprehensive. Carbon management is a key area of climate change technology and, for management of carbon sequestration, vegetation disturbance needs to be detected in manner that is timely enough both to inform management decisions and to provide feedback on the consequences of management decisions. (See Wofsy and Harris 2002 for an overview of existing national approaches to inventory carbon stocks.) This case study will work with the USDA Forest Service and USGS to fill this critical gap. The toolkit for hotspot detection and ranking will identify significant disturbance events, providing a 'front-end' to a collaborative system for characterizing their carbon cycle consequences. The case study will sample EOS data streams (primarily from MODIS instruments), test proposed hotspot algorithms for their potential for support of carbon management decisions, identify data sources for hotspot characterization (e.g., GLAS, ETM+, commercial hyperspatial), and develop ways of integrating carbon hotspot detection and prioritization into national carbon inventories and carbon budgets.

Oil spill detection, monitoring, and prioritization. Damage produced by marine oil spills includes soiled beaches, bird and mammal mortality, destruction of fisheries, impaired recreational facilities, and catastrophic impairment to entire ecosystems. Remote sensing can be used for oil spill detection and prevention of further damage. For example, the Exxon Valdez slick was detected through SPOT satellite data, the Ixtoc I well blowout slick in Mexico was detected using GOES and AVHRR on the NOAA polar series satellites, and oiled ice on Gabarus Bay (Kurdistan) was detected using LANDSAT data. We will use hyperspectral image analysis of Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) and Synthetic Aperture Radar (SAR) data to conduct case studies of the Patuxent River in Maryland and the Santa Barbara shoreline of California (see figure) for oil spill detection on sea water and associated mitigation. The main objective of using AVIRIS is to identify, measure, and monitor constituents of the Earth's surface and atmosphere based on molecular absorption and particle scattering signatures. SAR's ability to penetrate cloud cover, to illuminate the Earth's surface with its own signal, and to precisely measure distances, makes it especially useful for detecting and monitoring oil spills. The project's scan statistic hotspot delineation and poset prioritization



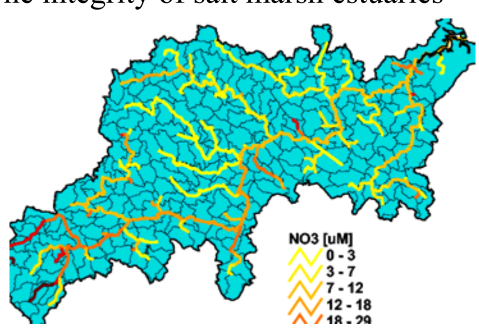
tools will be used in combination with our oil spill detection algorithm to provide for early warning and spatial-temporal monitoring of marine oil spills and their consequences.

Tasking of a self-organizing oceanic surveillance mobile sensor network. The Autonomous Ocean Sampling Network Simulator (AOSN) is used to study coordination and control strategies for high-resolution, spatio-temporally coordinated surveys of oceanographic fields such as bathymetry, temperature and currents using autonomous unmanned undersea vehicles. Currently, the network of mobile sensor platforms is autonomous and self-organizing once given high-level tasking from an external tactical coordinator. This case study proposes to use upper level set scan statistic theory to identify hotspots in data gathered by the sensor network and use this information to dynamically task mobile sensor platforms so that more data can be gathered in the areas of interest. By detecting hotspots and tasking accordingly, network resources are not wasted on mapping areas of little change. The ability of the sensor network to dynamically task its own components expands the network's mission and increases the reactivity to changing conditions in a highly dynamic environment. See Eberbach and Phoha (1999) and Phoha et al. (1997, 1999, 2001).



Early detection of biological invasions. Intentional and unintentional introductions of non-native exotic species have major economic and ecological impacts across the US. A National Academy of Sciences report (2002) estimates the cost of lost crops and containment measures at $137 billion per year. Early detection of invasive weedy plants is the only cost-effective and tractable option for their containment or eradication. But systems for synthesizing on-the-ground observation, spatial data, and newly acquired remotely sensed data are lacking. We will apply the ULS scan statistic and prioritization tools to obtain more efficient surveys for invasive species and to improve the responsiveness of environmental managers to outbreaks. Japanese stiltgrass has become established in forests and waterways in the eastern US and threatens to significantly reduce forest and riparian species diversity, and impede water flow in rivers and streams. Often locally established populations have begun to spread before those populations have been detected and likelihood for successful management is severely compromised. Coupling the data resources with the scan statistic represents a promising approach to preventing the transition of invasive plants from isolated established populations to spreading ones, like that depicted in the photograph. See Mortensen et al. (1993, 2000, 2002).

Network-based analysis of integrity of coastal intertidal wetlands. Salt marsh estuaries provide numerous natural services (Allen and Pye 1992) including: acting as a storm buffer, providing food and habitat for juvenile and larval fish/shellfish, helping to cleanse degraded river runoff, and storing large quantities of carbon in marsh soils. The integrity of salt marsh estuaries is a focus of a long-term ecological research (LTER) study in the Plum Island Estuary (PIE). The PIE LTER is investigating the relations among tidal drainage network configuration, the extent to which marsh drainage networks compare with power law scaling features of fluvial systems, and marsh integrity. Our studies at Plum Island are



designed to test the utility of terrestrially based models of landscape evolution on estuarine channel systems and proposed research will be the first to test the utility of "hotspot" identification to aid in the characterization of wetland integrity and geomorphic condition. Our ultimate goal is to identify indicators of marsh degradation to aid decision makers with current knowledge of the estuarine system. We are approaching the development of geomorphic indicators from two directions. At one level we are combining plot level measures of sediment elevation with characterizations of tidal drainage networks developed from color infrared aerial photography. Through this collaborative proposal, we also plan to quantify the landscape from LIDAR generated DEMs to examine the occurrence of power law scaling within and between channels and networks. LIDAR will provide high-resolution data that has not been available previously for geomorphic studies of estuarine landscapes. In particular, LIDAR, in conjunction with other imagery that can detect flooding water flow paths (e.g., thermal signature of flood waters over warm marsh surface) could help resolve the issue of topographic flow divides with water flow divides. New methods and tools for detecting hotspots of marsh degradation and for prioritization across large regions will greatly promote efforts to identify regions in distress and to devise means to manage nationally important coastal regions. See Day et al. (1977), Morris and Bradley (1999), Pethick (1992) and Rinaldo et al. (1999ab).

Development of remote sensing methods for crop bioterrorism. Testimony to the House National Security Subcommittee this past fall stressed the high probability that crop terrorism could and would occur. The logistics of ground-based monitoring are daunting, and point to a strong need to monitor US cropland by remote sensing. Remote sensing, to be timely and effective, should resolve zones of infection as small as 5m in diameter. Improvements in the next generation of sensors will have to be made, as will the techniques for handling the huge volumes of high resolution data. Another difficulty has been development of high-quality signatures detectable from airborne or space-borne platforms. To-date, reliable detectable signatures have been difficult to obtain. We propose to defeat this impasse by integrating biologists, engineers and statisticians to generate high quality hyperspectral signatures, to determine signal strength on unique portions of these signatures, and to develop goals for instrument resolution from various platforms. Specific goals are: (1) Develop signatures for two key pathogens using ground-based, portable hyperspectral cameras. (2) Determine signal strength vs. noise of plants infected with single or multiple pathogens and/or insects. (3) Enhance hardware and processing algorithms to filter and resolve crop pathogen signals. (4) Provide goals for the next generation of air and space-borne sensors for high-threat pathogens. The diagram depicts the process from signature development though identification of threat areas, signal acquisition/processing to the development of an anomaly report. See Backman and Jacobi (1997), Kang et al. (2002), Madden and Van den Bosch (2002) and Wheelis et al. (2002).