



# Hot-Spot Geoinformatics for Digital Governance

**G. P. Patil**

*The Pennsylvania State University, USA*

**R. Acharya**

*The Pennsylvania State University, USA*

**R. Modarres**

*George Washington University, USA*

**W. L. Myers**

*The Pennsylvania State University, USA*

**S. L. Rathbun**

*University of Georgia, USA*

## INTRODUCTION

Geoinformatic surveillance for spatial and temporal hot-spot detection and prioritization is crucial in the 21<sup>st</sup> century. A hot spot may be any unusual phenomenon, anomaly, aberration, outbreak, elevated cluster, or critical area. Government agencies require hot-spot delineation and prioritization for monitoring, etiology, management, or early warning. Responsible factors may be natural, accidental, or intentional, with relevance to both infrastructure and security.

This article describes multidisciplinary research based on novel methods for hot-spot detection and prioritization, driven by a diverse variety of case studies of interest to agencies, academia, and the private sector. These case studies concern critical societal issues, such as public health, ecosystem health, biodiversity and threats to biodiversity, emerging infectious disease, water management and conservation, carbon sources and sinks, persistent poverty, environmental justices, crop pathogens, invasive-species management, biosurveillance, biosecurity, disease biogeoinformatics, social networks, sensor networks, hospital networks and syndrome surveillance, video mining, early warning, tsunami inundation, remote sensing, and disaster management.

Our approach has involved an innovation of the popular circle-based spatial scan statistic. In particular, it employs the notion of an upper level set (ULS) and is accordingly called the upper level set scan statistic system, pointing to the next generation of sophisticated

analytical and computational systems, effective for the detection of arbitrarily shaped hot spots along spatiotemporal dimensions. It also involves a novel prioritization scheme based on multiple indicators and stakeholder criteria without having to reduce indicators to a single index using Hasse diagrams and partially ordered sets. It is accordingly called the poset prioritization and ranking system (see Patil & Taillie, 2004a, 2004b).

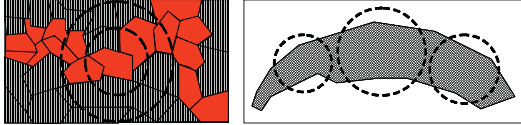
The following Web sites have additional information.

1. <http://www.stat.psu.edu/hotspots/>
2. <http://www.stat.psu.edu/~gpp/>
3. [http://www.digitalgovernment.org/news/stories/2004/1104/1104\\_hotspots\\_heyman.jsp](http://www.digitalgovernment.org/news/stories/2004/1104/1104_hotspots_heyman.jsp)

## UPPER LEVEL SET HOT-SPOT SCAN STATISTIC SYSTEM

Patil and Taillie (2004a, 2004b) introduce an innovation of the health-area-popular circle-based spatial and spatiotemporal scan statistic. It employs the notion of an upper level set, and is accordingly called the upper level set scan statistic, pointing to a sophisticated analytical and computational system as the next generation of the present-day popular SaTScan (Kulldorff, 1997, 2001; Kulldorff & Nagarwalla, 1995; Kulldorff, Rand, Gherman, Williams, & DeFrancesco, 1998; Mostashari, Kulldorff, & Miller, 2002; Waller, 2002).

Figure 1. Limitations of circular scanning windows. (Left) An irregularly shaped cluster—perhaps a cholera outbreak along a winding river floodplain. Small circles miss much of the outbreak and large circles include many unwanted cells. (Right) Circular windows may report a single irregularly shaped cluster as a series of small clusters.



## Background Theory of Scan Statistics

The spatial scan statistic concerns the following situation: A region  $R$  of Euclidian space is tessellated or subdivided into cells, which will be denoted by the symbol  $a$ . Data is available in the form of a count  $Y_a$  on each cell  $a$ . In addition, a size value  $A_a$  is associated with each cell. The cell sizes  $A_a$  are regarded as fixed and known, while the cell counts  $Y_a$  are independent random variables. Two distributional settings are commonly studied:

- **Binomial:** The size  $A_a = N_a$  is a positive integer and  $Y_a \sim \text{Binomial}(N_a, p_a)$ , where  $p_a$  is an unknown parameter attached to cell  $a$  with  $0 < p_a < 1$ .
- **Poisson:** The size  $A_a$  is a positive real number and  $Y_a \sim \text{Poisson}(\lambda_a A_a)$ , where  $\lambda_a > 0$  is an unknown parameter attached to cell  $a$ .

Each distributional model has a simple interpretation. For the binomial,  $N_a$  people reside in cell  $a$  and each person contracts a certain disease independently with probability  $p_a$ . The cell count  $Y_a$  is the number of diseased people. For the Poisson,  $A_a$  is the size (e.g., area or some adjusted population size) of the cell  $a$ , and  $Y_a$  is a realization of a Poisson process with intensity  $\lambda_a$ . In each scenario, the responses  $Y_a$  are independent; it is assumed that spatial variability can be accounted for by cell-to-cell variation in model parameters.

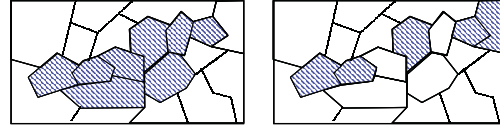
The spatial scan statistic seeks to identify hot spots or clusters of cells having an elevated response with respect to the remainder of the region. Elevated response means large values for the rates (or intensities),

$$G_a = Y_a / A_a,$$

instead of the raw counts  $Y_a$ . The scan statistic easily accommodates other adjustments, such as for age or gender.

A collection of cells from the tessellation should satisfy several geometric properties before it can be

Figure 2. A tessellated region. The collection of shaded cells in the left-hand diagram is connected and, therefore, constitutes a zone in  $\Omega$ . The collection on the right is not connected.



considered as a candidate for a hot-spot cluster. First, the union of the cells should comprise a geographically connected subset of the region  $R$  (Figure 2). Such collections of connected cells will be referred to as zones  $Z$  and the set of all zones is denoted by  $\Omega$ . Second, the zone should not be excessively large. Otherwise, the zone instead of its exterior would constitute the background. This restriction is generally achieved by limiting the search for hot spots to zones comprising of less than, say, 50% of the region.

The notion of a hot spot is inherently vague and lacks any a priori definition. There is no true hot spot in the statistical sense of a true parameter value. A hot spot is instead defined by its estimate, provided the estimate is statistically significant. To this end, the scan statistic adopts a hypothesis testing model in which the hot spot occurs as an unknown zonal parameter in the statement of the alternative hypothesis.

The traditional spatial scan statistic uses expanding circles to determine a reduced list  $\Omega_0$  of candidate zones  $Z$ . By their very construction, these candidate zones tend to be compact in shape and may do a poor job of approximating actual clusters. The reduced parameter space of the circular scan statistic is determined entirely by the geometry of the tessellation and does not involve the data in any way. We propose a scan statistic that takes an adaptive point of view in which  $\Omega_0$  depends very much upon the data. Furthermore,  $\Omega_0$  induces a tree structure useful for visualization and expressing uncertainty of hot-spot clusters in the form of a hot-spot confidence set on the tree.

Although the traditional spatial scan statistic is applicable only to tessellated data, the ULS approach has an abstract graph (i.e., vertices and edges) as its starting point. Accordingly, this approach can also be applied to data defined over networks, such as subway, water, or highway systems. There is complete flexibility regarding the definition of adjacency. For example, one may declare two cells as adjacent if (a) their boundaries have at least one point in common, (b) their common boundary has positive length, or (c) in the case of a drainage network, the flow is from one cell to the next.

7 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

[www.igi-global.com/chapter/hot-spot-geoinformatics-digital-governance/11612](http://www.igi-global.com/chapter/hot-spot-geoinformatics-digital-governance/11612)

## Related Content

---

### Hot Topic Sensing, Text Analysis, and Summarization

Guillaume Bouchard, Stephane Clinchant and William Darling (2014). *Handbook of Research on Advanced ICT Integration for Governance and Policy Modeling* (pp. 216-247).

[www.irma-international.org/chapter/hot-topic-sensing-text-analysis-and-summarization/116665](http://www.irma-international.org/chapter/hot-topic-sensing-text-analysis-and-summarization/116665)

### Information Technology and Administrative Reform: Will E-Government Be Different?

Kenneth Kraemer and John Leslie King (2006). *International Journal of Electronic Government Research* (pp. 1-20).

[www.irma-international.org/article/information-technology-administrative-reform/2009](http://www.irma-international.org/article/information-technology-administrative-reform/2009)

### An Evaluation of Digital Deliberative Democracy in Local Government

Seung-Yong Rho (2008). *Electronic Government: Concepts, Methodologies, Tools, and Applications* (pp. 349-361).

[www.irma-international.org/chapter/evaluation-digital-deliberative-democracy-local/9718](http://www.irma-international.org/chapter/evaluation-digital-deliberative-democracy-local/9718)

### E-Policing: Environmental and Organizational Correlates of Website Features and Characteristics Among Large Police Departments in the United States of America

Melchor C. de Guzman and Matthew A. Jones (2012). *International Journal of Electronic Government Research* (pp. 64-82).

[www.irma-international.org/article/policing-environmental-organizational-correlates-website/64209](http://www.irma-international.org/article/policing-environmental-organizational-correlates-website/64209)

### A Framework for Public eServices Transparency

Rui Pedro Lourenço (2023). *International Journal of Electronic Government Research* (pp. 1-19).

[www.irma-international.org/article/a-framework-for-public-eservices-transparency/317415](http://www.irma-international.org/article/a-framework-for-public-eservices-transparency/317415)