



Center for **Statistical Ecology** and **Environmental Statistics**

Detection of Spatial Clusters in Maps Equipped with Environmentally Defined Structures

By Luiz Duczmal,¹ G.P. Patil,² Ricardo Tavares,¹ and Andre L.F. Cancado³

¹Univ. Federal de Minas Gerais, Department of Statistics, Brazil

²Center for Statistical Ecology and Environmental Statistics, Department of Statistics,
The Pennsylvania State University, University Park, PA USA

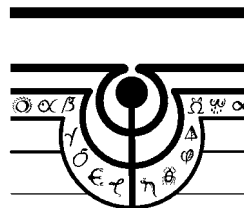
³Univ. Federal de Minas Gerais, Electrical Engr. Department, Brazil

This material is based upon work partially 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) Pennsylvania Department of Health using Tobacco Settlement Funds and 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.

[Presented at the 7th Annual International Conference on Digital Government Research]

Technical Report Number 2006-0527
TECHNICAL REPORTS AND REPRINTS SERIES
May 2006



Department of Statistics
The Pennsylvania State University
University Park, PA 16802

G. P. Patil
Distinguished Professor and Director
Tel: (814)865-9442 Fax: (814)865-1278
Email: gpp@stat.psu.edu

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

<http://www.stat.psu.edu/hotspots>

[DGOnline News](#)

[Environmental and Ecological Statistics-Springer](#)

Detection of spatial clusters in maps equipped with environmentally defined structures

Luiz Duczmal
Univ. Fed. Minas Gerais
Statistics Department
Brazil
55-31-3499-5900
duczmal@est.ufmg.br

Ganapati Patil
Penn State University
Center for Statistical
Ecology and Environmental
Statistics
USA
gpp@stat.psu.edu

Ricardo Tavares
Univ. Fed. Minas Gerais
Statistics Department
Brazil
tavares@est.ufmg.br

André L.F. Cançado
Univ. Fed. Minas Gerais
Electrical Engineering
Department
Brazil
cancado@cpdee.ufmg.br

ABSTRACT

We propose a novel scan statistic algorithm for finding spatial clusters in a map divided in a finite number of regions, whose adjacency is defined by a graph structure. Instead of specifying locations for the possible clusters a priori, as is currently done for cluster finders based on focused algorithms, we alter the usual adjacency induced by the common geographical boundary between regions. In our approach, the connectivity between regions will be reinforced or weakened, according to certain environmental features of interest associated with the map. We build various plausible scenarios, each time modifying the adjacency structure on specific geographic places in the map, and run a multi-objective genetic algorithm to find irregularly shaped clusters. In this manner, the probability of cluster detection is increased or decreased, according to the changes made in the adjacency graph structure, related to the selected environmental features. The clusters with the lowest estimated p-values, along with their corresponding maps of enhanced environmental features, are displayed for comparative analysis. This procedure also provides a tool for testing hypotheses concerning the adequacy of environmentally defined models for clustering of diseases. An application for malaria clusters in Brazil is presented.

Categories and Subject Descriptors

G.3 PROBABILITY AND STATISTICS - Monte Carlo

I.2 ARTIFICIAL INTELLIGENCE - I.2.8. Graph and Tree Search Strategies

J.3 LIFE AND MEDICAL SCIENCES - Health

General Terms

Algorithms, Measurement, Performance.

Keywords

Disease cluster detection, irregularly shaped spatial cluster, multi-objective optimization, spatial scan statistic, adjacency graph map.

1. INTRODUCTION

Tools developed for the detection of spatial disease clusters are currently being used by epidemiologists for etiological studies. When statistically significant, it is important to locate geographically the cluster as precisely as possible. Instead of working with a map devoid of environmental features, it makes sense to try to add as much relevant information as possible. In the literature, focused algorithms rely on a priori information about possible sources of contamination in the map. That geographical information is supposed to be relevant to the appearance of disease clusters.

2. METHODS

In this paper, we explore an alternative approach based on previous knowledge of the environmental features that are suspected of being significant for the spatial distribution of the cases of the specific disease being studied. Differently from the focused algorithms, we do not force the occurrence of clusters in certain specific places, but instead we create more opportunities for their occurrence, enhancing the adjacency structure in the map where we suspect a cluster may be located. In this work, neighboring regions that share common environmental features are defined as strongly connected.

We present a scheme based on multi-objective optimization for selecting the best cluster solution, among the many possible solutions found. By running sequentially the algorithm for various parameter settings, a large number of solutions are obtained. That parameter setting controls several variables, namely the strength of the neighborhood between regions, the choice among several possible scenarios for environmental similarity between regions, and the maximum allowed size (in terms of total population fraction) for the clusters. The algorithm aims to maximize two competing objectives, namely the cluster scan likelihood ratio, and the regularity of cluster shape.

3. RESULTS

For our application example, we analyzed the map of malaria deaths rates in the Brazilian Amazon of Figure 1, looking for clusters related to environmental conditions such as rain precipitation. The usual geographical adjacency graph is depicted in Figure 2. Based in the 2000 year monthly meteorological

records of rainy days, we built 12 scenarios. In each scenario we reinforced the corresponding monthly graph. We have added edges linking second-order neighboring regions where more rainy days have occurred in that month. An illustration of such reinforced graph is shown in Figure 3, for the month of June. Using the 12 scenarios, we applied the multi-objective algorithm to detect clusters of different shapes, from the round, almost circular clusters to the very irregularly shaped ones. The four clusters shown in Figure 4 are representatives of the solutions found by the algorithm. The cluster in the upper left map is the primary cluster, and it has appeared in all the scenarios solution sets, because it is a very conspicuous cluster. The other three clusters are secondary ones, and their appearance depended upon the scenario that we were using in the search. The interesting fact is that all the three secondary clusters found are legitimate ones, as can be checked comparing with Figure 1.

4. CONCLUSIONS

The strategy of reinforcing the adjacency graph, building many different scenarios suggested by environmental conditions presents an alternative to the usual focused methods. In our application for malaria deaths in the Brazilian Amazon, we used the monthly occurrence of rain as the environmental factor to construct the set of scenarios. We observed, through simulations using a multi-objective cluster finder based in the maximization of the scan statistic and the regularity of cluster shape that the primary and secondary clusters were correctly found. Besides, the secondary clusters detection was influenced by the reinforcement applied to each specific scenario, suggesting that the environmental factor selected (rain, in our study) may be related to the occurrence of malaria spatial clusters.

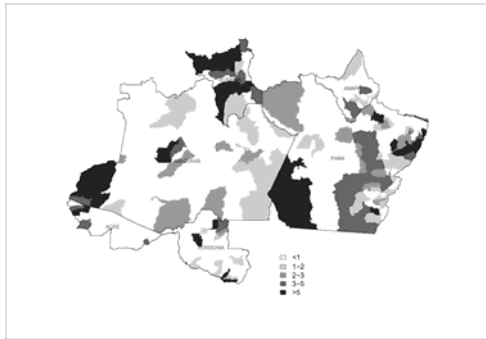


Figure 1 – Malaria deaths in Brazilian Amazon (1998-2002)



Figure 2 – Usual geographical boundary between municipalities.

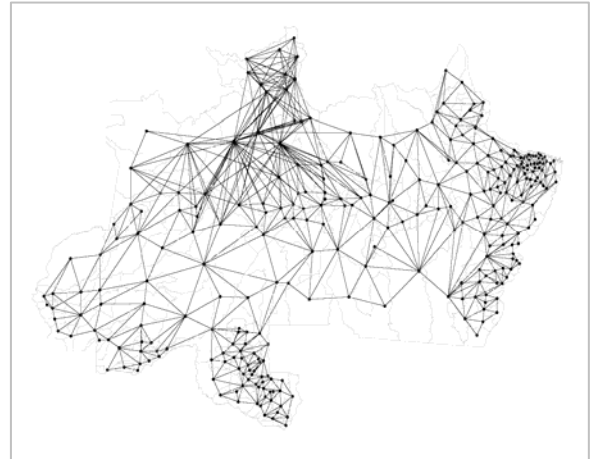


Figure 3 – Adjacency reinforced graph for the June scenario.



Figure 4 – Primary (upper left) and secondary malaria clusters.

5. REFERENCES

- [1] Kulldorff M, 1997. A Spatial Scan Statistic, *Comm. Statist. Theory Meth.*, 26(6), 1481-1496.
- [2] Duczmal L, Kulldorff M, Huang L., 2006. Evaluation of spatial scan statistics for irregularly shaped clusters. *J. Comput. Graph. Stat.* (to appear).
- [3] Duczmal L, Cançado ALF, Takahashi RHC, Bessegato LF, 2006. A genetic algorithm for irregularly shaped spatial scan statistics (*submitted*).
- [4] Duczmal L, Cançado ALF, Takahashi RHC, 2006. Delineation of Irregularly Shaped Disease Clusters through Multi-Objective Optimization (*submitted*)