Emerging Health Threats Journal 2011, **4**:s23. **doi:** 10.3134/ehtj.10.023

www.eht-journal.org

ABSTRACT

Weighted non-connectivity for detection of irregular clusters

AR Duarte[1], SB Silva[2], L Duczmal[2], SJ Ferreira[2], and ALF Cançado[3]

[1]*Mathematics Department, Federal University of Ouro Preto, Ouro Preto, Brazil;* [2]*Statistics Department, Federal University of Minas Gerais, Belo Horizonte, Brazil; and* [3]*Statistics Department, University of Brasilia, Brasilia, Brazil*
E-mail: sabjfn@gmail.com.br

Objective

Irregularly shaped clusters in maps divided into regions are very common in disease surveillance. However, they are difficult to delineate, and usually we notice a loss of power of detection. Several penalty measures for the excessive freedom of shape have been proposed to attack this problem, involving the geometry and graph topology of clusters. We present a novel topological measure that displays better performance in numerical tests.

Introduction

Heuristics to detect irregularly shaped spatial clusters were reviewed recently.[1] The spatial scan statistic is a widely used measure of the strength of clusters.[2] However, other measures may also be useful, such as the geometric compactness penalty,[3] the non-connectivity penalty[4] and other measures based on graph topology and weak links.[5,6] Those penalties $p(z)$ are often coupled with the spatial scan statistic $T(z)$, employing either the multiplicative formula maximization $\max_z T(z) \times p(z)$[4] or a multiobjective optimization procedure $\max_z(T(z), p(z))$,[3,6] or even a combination of both approaches.[5] The geometric penalty of a cluster z is defined as the quotient of the area of z by the area of the circle, with the same perimeter as the convex hull of z, thus penalizing more the less rounded clusters. Now, let V and E be the vertices and edges sets, respectively, of the graph $G_z(V, E)$ associated with the potential cluster z. The non-connectivity penalty $y(z)$ is a function of the number of edges $e(z)$ and the number of nodes $n(z)$ of $G_z(V, E)$, defined as $y(z)=e(z)/3[n(z)-2]$. The less interconnected tree-shaped clusters are the most penalized. However, none of those two measures includes the effect of the individual populations.

Methods

Based on the non-connectivity penalty, we propose the concept of weighted non-connectivity $w(z)$, taking into account the non-uniform distribution of the populations of the regions of the cluster z. Let

$$w(z) = \sum\nolimits_{a\in A} q(a)/3\left[\sum\nolimits_{v\in V} p(v) - 2\sum\nolimits_{v\in V} p(v)/n(z)\right]$$

where $p(v)$ is the population of the region associated with the node $v \in V$ and $q(a)$, $a \in E$ is a weight, defined as the average of the populations of the regions linked by the edge a (see Figure 1). The function $w(z)$ is used as the second objective in the multiobjective problem $\max_z(T(z), w(z))$, employing a genetic algorithm. The solution is a Pareto set, consisting of all clusters that are not worse on both objectives simultaneously. The statistical significance evaluation is performed for all clusters in the Pareto set through Monte Carlo simulations, with the help of the attainment function concept,[6] determining the best solution.

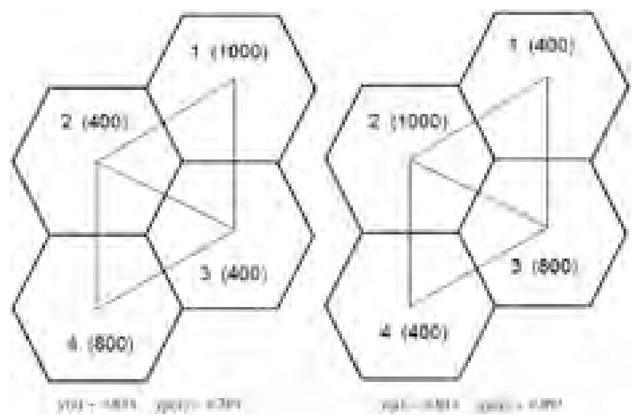

Figure 1 Comparison between measures of non-connectivity and weighted non-connectivity for two hypothetical clusters.

Results

Numerical simulations were conducted to assess the power of detection, sensitivity and positive predictive value. We run the weighted non-connectivity scan for the benchmarks in ref. 3 and compare it with the algorithms defined in refs 3–6. The weighted non-connectivity scan presented significantly higher power of detection and sensitivity, and about the same positive predictive value, compared with the other algorithms.

Conclusion

The weighted non-connectivity scan is more efficient when compared with the geometric, non-connectivity and other topological penalty based scans.

Acknowledgements

SJF acknowledges the support by Fapemig, Brazil. This paper was presented as an oral presentation at the 2010 International Society for Disease Surveillance Conference, held in Park City, UT, USA on 1–2 December 2010.

References

1 Duczmal L, Duarte AR, Tavares R. Extensions of the scan statistic for the detection and inference of spatial clusters. J. Glaz, V. Pozdnyakov, S. Wallenstein. (Org.). *Scan Statistics: Methods and Applications*. Birkhauser, Boston, MA, 2009;157–82.
2 Kulldorff M. A spatial scan statistic. *Comm Statist Theory Meth* 1997;**26**:1481–96.
3 Duczmal L, Cançado ALF, Takahashi RHC. Geographic delineation of disease clusters through multi-objective optimization. *J Comput Graph Stat* 2008;**17**:243–62.
4 Yiannakoulias N, Rosychuk RJ, Hodgson J. Adaptations for finding irregularly shaped disease clusters. *IJ Health G* 2007;**6**:28.
5 Duarte AR, Duczmal L, Ferreira SJ, Cançado ALF. Internal cohesion and geometric shape of spatial clusters. *Environ Ecol Stat* 2010;**17**:203–29.
6 Cançado ALF, Duczmal L, Duarte AR, Ferreira SJ, Fonseca CM, Gontijo ED. Penalized likelihood and multi-objective spatial scans for the detection and inference of irregular clusters. *Int J Health Geogr* 2010;**9**:55.