Geographically meaningful cluster scanning through weak link correction

Luiz Duczmal, Sabino J. Ferreira, Anderson R. Duarte, Marcus V. Soares,
Eliane D. Gontijo, André L.F. Cançado, Ricardo H.C. Takahashi.
Universidade Federal de Minas Gerais, Brazil

OBJECTIVE

Irregularly shaped clusters occur naturally in disease surveillance, but they are not well defined. The number of possible clusters increases exponentially with the number of regions in a map. This concurs to reduce the power of detection, motivating the utilization of some kind of penalty function to avoid excessive freedom of shape. We introduce a weak link based correction which penalizes inconsistent clusters, without forbidding the presence of the geographically interesting irregularly shaped ones.

BACKGROUND

Many heuristics were developed recently to find arbitrarily shaped clusters (see review [1]). The most popular statistic is the spatial scan [2]. Nevertheless, even if all cluster solutions could be known, the problem of selecting the best cluster is ill posed. This happens because other measures, such as geometric regularity [3-5] or topology [6] must be taken into consideration. Most cluster finding methods does not address this last problem. A genetic multi-objective algorithm was developed elsewhere to identify irregularly shaped clusters [5]. That method conducts a search aiming to maximize two objectives, namely the scan statistic and the regularity of shape (using the *compactness* concept). The solution presented is a Pareto-set, consisting of all the clusters found which are not simultaneously worse in both objectives. The significance evaluation is conducted in parallel for all the clusters in the Pareto-set through a Monte Carlo simulation, determining the best cluster solution.

METHODS

The concept of *weak links*, or under-populated regions which disconnect a cluster when removed from it, was defined in [3]. It was argued that the presence of weak links impacts the power of detection of clusters. In the present paper we quantify such notion and introduce the definition of *cohesion*, as a function of the strength of a cluster, based on the presence or absence of weak links (Figure 1). We use this function as the second objective of a multi-objective scan genetic algorithm (replacing the compactness function above), and study its power to detect spatial clusters, its sensitivity and positive predicted value.

RESULTS

We ran the weak link algorithm for the benchmarks defined in [3] and compared it with the compactness based algorithm [5] and the mono-objective algorithm [4] without penalization. Our algorithm has significantly more power, about the same sensitivity and slightly less specificity compared to the other methods. An application is presented for Chagas' disease incidence in puerperal women in Brazil.

CONCLUSIONS

The weak link correction is more effective compared to the geometric compactness correction, and produce solutions which are more geographically meaningful.

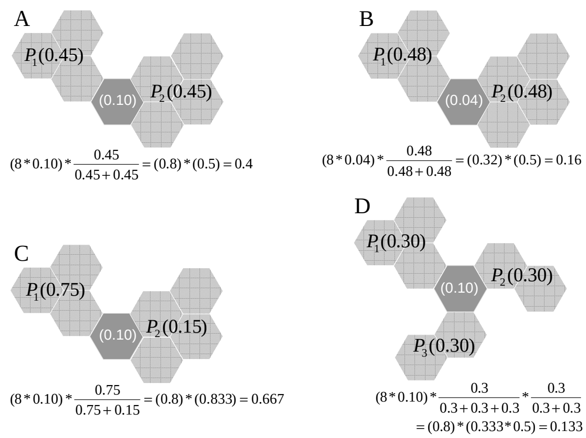

Figure 1 – Weak links are undesirable because they compromise cluster stability, making it easy to break the cluster into pieces. Dark grey regions are weak links for all 8-region clusters A-D, having less than average population (0.125), and also disconnecting the clusters when removed. The weak link in B is worse than in A due to its smaller population. The weak link in A is worse than in C because it breaks the cluster more evenly. Finally, the weak link in D is worse than in A, breaking the cluster into three parts.

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, p. 1-24. (2008, to appear)

[2]Kulldorff M, 1997. A Spatial Scan Statistic, *Comm. Statist. Theory Meth.*, 26(6), 1481-1496.

[3]Duczmal L, Kulldorff M, Huang L. Evaluation of spatial scan statistics for irregularly shaped disease clusters, *J. Comput. Graph. Stat.*, 2006;15, 428-442

[4]Duczmal L, Cançado, ALF, Takahashi RHC, Bessegato LF. A genetic algorithm for irregularly shaped spatial scan statistics, *Comput. Stat. Data Anal.*, 2007; 52, 43-52

[5]Duczmal L, Cançado ALF, Takahashi RHC. Geographic Delineation of Disease Clusters through Multi-Objective Optimization, *J. Comput. Graph. Stat.*, 2008; 17, 243-262

[6]Yiannakoulias N, Rosychuk RJ, Hodgson J. Adaptations for finding irregularly shaped disease clusters, *IJ Health.G.*, 2007;6,28

Corresponding author: duczmal@ufmg.br