Exploring multi-cluster structures with the multi-objective circular scan

Flávio dos Reis Moura, Luiz Duczmal, Ricardo Tavares, Ricardo H.C. Takahashi

Universidade Federal de Minas Gerais, Brazil

OBJECTIVE

Situations where a disease cluster does not have a regular shape are fairly common. Moreover, maps with multiple clustering, when there is not a clearly dominating primary cluster, also occur frequently. We would like to develop a method to analyze more thoroughly the several levels of clustering that arise naturally in a disease map divided into *m* regions.

BACKGROUND

The spatial scan statistic is the usual measure of strength of a cluster [1]. Another important measure is its geometric regularity [2]. A genetic multi-objective algorithm was developed elsewhere to identify irregularly shaped clusters [3]. A search is executed 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. A significance evaluation is conducted in parallel for all clusters in the Pareto-set through Monte Carlo simulation, then finding the most likely cluster.

METHODS

Instead of using a genetic algorithm, our novel method incorporates the simplicity of the circular scan, being able to detect and evaluate irregularly shaped clusters. We define the circular occupation (CO) of a cluster candidate roughly as its population divided by the population inside the smallest circle containing it. The CO concept, computationally faster and relying on familiar concepts, substitutes here the compactness definition as the measure of regularity of shape. The scan statistic is evaluated for each of the *m* regions of the map taken individually. The regions are ranked accordingly in decreasing order. Let *R(k)* be the set containing the first *k* regions. A multi-objective modification of the circular scan algorithm is successively applied for each set *R(k)*. For each circle, the candidate cluster consists of the regions belonging to *R(k)* within it, and the quotient of the CO calculation takes into account all the regions of the original map inside the circle. In practice we choose only some few *k* values such as *m,m/2,m/4,…,1*. For each value of *k* we build the Pareto-set *P(k)*. We display all the Pareto-sets in a graph, and after joining all of them we compute the global Pareto-set *P(0)* (Figure 1). A Monte Carlo procedure is used for significance evaluation.

RESULTS

The presence of "knees" in the Pareto-sets indicates sudden transitions in the clusters structure, corresponding to rearrangements due to the coalescence of loosely knitted (usually disconnected) clusters. In Figure 1, someone could test the hypothesis that the occurrence of clusters of malaria was related to certain geographic features, and the circular cluster alone (upper left), being very compact, does not provide sufficient clues. The irregularly shaped clusters to the right indicate more geographic detail. As each Pareto-set contains the most likely clusters within a certain level of geographical information, they could be joined to provide an overall complete description.

CONCLUSIONS

The multi-objective circular scan is a fast method that allows peering into the clustering structure of a map. The comparison of Pareto-sets for observed cases with those computed under null-hypothesis provides valuable hints for the spatial occurrence of diseases. The potential for monitoring incipient spatial-temporal clusters at several geographic scales simultaneously is a promising tool in syndromic surveillance, especially for contagious diseases when there is a mix of short and long range spatial interactions.

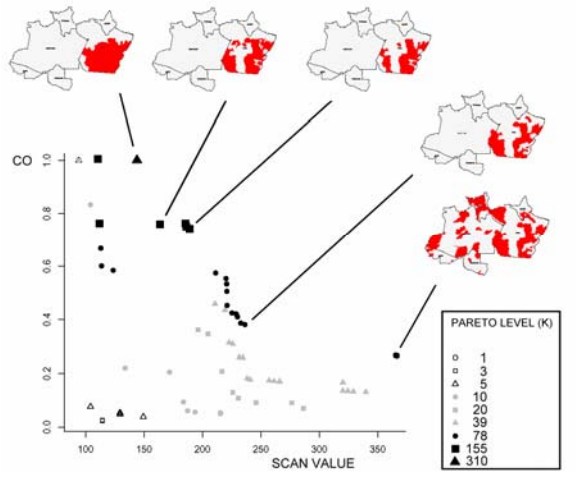

Figure 1 – Several levels of Pareto-sets in a SCAN VALUE versus CO graph for malaria deaths in the Brazilian Amazon (1998-2002).

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.* 15;1-15.

[3] Duczmal L, Cançado ALF, Takahashi RHC, 2006. Delineation of Irregularly Shaped Disease Clusters through Multi-Objective Optimization (*submitted*)

Corresponding author: duczmal@ufmg.br