ISDS

Emerging Health Threats Journal 2011, **4**:s20. **doi:** 10.3134/ehtj.10.020

www.eht-journal.org

ABSTRACT

A Voronoi based scan for space-time cluster detection in point event data

LH Duczmal, GJP Moreira, D Burgarelli, RHC Takahashi, FCO Magalhães, and EC Bodevan

Federal University of Belo Horizonte, Minas Gerais, Brazil
E-mail: duczmal@est.ufmg.br

Objective
We describe a method for prospective space-time cluster detection of point event data based on the scan statistic. Our aim is to detect as early as possible the appearance of an emerging cluster of syndromic individuals because of a real outbreak of disease amidst the heterogeneous population at risk.

Introduction
Scan statistics are highly successful for the evaluation of space-time clusters.[1] Recently, concepts from the graph theory were applied to evaluate the set of potential clusters. Wieland *et al.*[2] introduced a graph theoretical method for detecting arbitrarily shaped clusters on the basis of the Euclidean minimum spanning tree of cartogram transformed case locations, which is quite effective, but the cartogram construction step of this algorithm is computationally expensive and complicated.

Methods
The data set consist of locations of entities (cases/controls) often represented as points in a two-dimensional map space or in a three-dimensional space-time domain. Here, cases refer to individuals with a particular disease of interest, and controls refer to individuals with similar characteristics, but do not have the disease. We build the *Voronoi Diagram* of the case-control observed data space-time set. Assign to each point p_i (case or control) the part of the domain closer to p_i than to any other point. In this construction, each cell contains exactly one case or control. Let $C = \{c_i^t\}$ be the subset representing the cases. We define a weighted complete graph $G(C) = (V, E)$ with vertex set $V = \{c_i^t | c_i^t \in C\}$ and edge set $E = \{(c_i^t, c_j^t) | c_i^t, c_j^t \in C, i \neq j\}$. Edges $(u, v) \in E$ have weight defined by the *case-to-case cell-crossing count* (C^5), that is, the number of Voronoi cells between cases u and v. This *Voronoi metric* is used instead of the usual Euclidean metric. After computing the minimum spanning tree (MST) T of the weighted graph $G(C)$, every potential cluster is defined as a connected subgraph of T. Given a set V of n cases, we evaluate only $2n$-1 potential clusters, by sequentially deleting the longest remaining edge of T. We consider the two newly emergent connected components for each step. Figure 1 shows an example of the spatial projection of the MST.

Results
Our C^5 scan was compared numerically with the elliptic version of the popular prospective space-time scan[1,3] and an identical scan (but using the Euclidean metric) according to power of detection, positive predicted value (PPV) and sensitivity. We used artificial datasets with total population

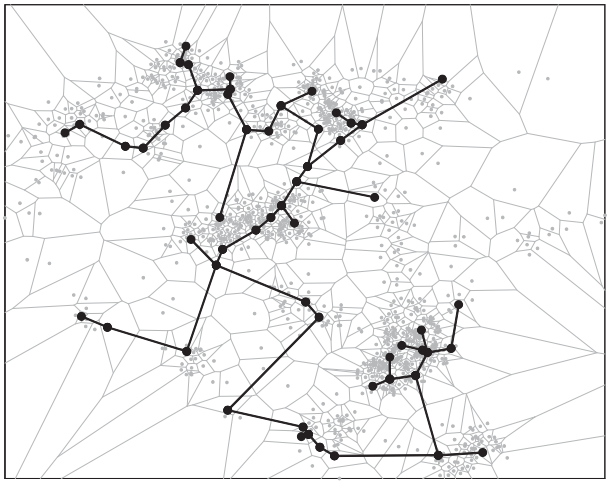

Figure 1 A MST linking cases according to the Voronoi metric.

Table 1 Power, PPV and sensitivity for three space time clusters

Cluster	*Power*		*PPV*		*Sensitivity*	
Model	C^5	*Elliptic*	C^5	*Elliptic*	C^5	*Elliptic*
Cylinder	0.6072	0.4789	0.6763	0.6415	0.6762	0.5447
Cone	0.4540	0.3863	0.6078	0.5822	0.6025	0.4683
L-shape	0.5115	0.3316	0.5980	0.5323	0.6301	0.4530

Abbreviation: PPV, positive predicted value.

at risk of 1000 individuals, 100 cases, for a period of 10 days. Three alternative hypotheses models of space-time clusters with different shapes were simulated; exactly the same sets were used for all algorithms, with 10 000 Monte Carlo simulations for each model (Table 1).

Conclusions

We have proposed a space-time cluster scan for case event data. For all three-shaped clusters our scan has better performance compared with the prospective elliptic scan[1,3] and is much faster than the density-equalizing Euclidean minimum spanning trees algorithm[2] and the elliptic scan. It also is significantly better than the Euclidean metric MST scan (results not shown).

Acknowledgements

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 Kulldorff M. Prospective time periodic geographical disease surveillance using a scan statistic. *JRSS A* 2001;**164:**61–72.
2 Wieland SC, Brownstein JS, Berger B, Mandl KD. Density-equalizing Euclidean minimum spanning trees for the detection of all disease cluster shapes. *PNAS* 2007;**104:**9404–9.
3 Kulldorff M, Huang L, Pickle L, Duczmal L. An elliptic spatial scan statistic. *Stat Med* 2006;**25:**3929–43.