A Space Time Permutation Scan Statistic with Irregular Shape for Disease Outbreak Detection

Marcelo A. Costa, Martin Kulldorff, Renato M. Assunção
Harvard Pilgrim Health Care and Harvard Medical School

OBJECTIVE

This paper describes a methodology for detecting irregular space-time cluster using the space time permutation scan statistic. The methodology includes sequential Monte Carlo simulation and distribution approximation to estimate the error type I.

BACKGROUND

The Space-Time Permutation Scan Statistics [1] is an important tool for early disease outbreak detection surveillance systems. It aims at detecting emerging clusters in regions where a record of the historical number of cases is available. A cylinder with variable radius and height is used to scan the space-time region in order to select the candidate cluster with maximum likelihood. A Monte Carlo simulation is performed to estimate the p-value by means of a data permutation procedure. Changes in the spatial cluster shape have been reported in the literature [4,5] but an extension to space-time data has been a challenge mainly because of the increasing computational effort required to delineate the irregular cluster shape and running the Monte Carlo simulation. A sequential Monte Carlo procedure [3] can replace the conventional method and reduce the number of simulations. Furthermore, an extreme value distribution [2] can be fitted to the simulated statistics and provide a better resolution for the p-value than the standard Monte Carlo estimate.

METHODS

Irregular shapes can be obtained fixing a non-circular geometry search or using a dynamic algorithm that changes the cluster shape as the likelihood maximization process is performed. The second alternative has been widely used since no prior information on the cluster shape is necessary. The objective is to grow the cluster candidate following an interconnected graph structure created previously from geographical features like common geographical boundaries. This purely spatial graph structure can be extended to space-time by connecting a time sequence of spatial layers. Each spatial layer is connected to its previous and posterior layer. In addition, connected areas in space are connected in time with their preceding and subsequent time areas. The irregular cluster scan procedure follows this dense graph structure.

A sequential Monte Carlo has two parameters, the total number of simulations, *n*, and the sequential parameter, *h*. When the number of simulations reaches *h* simulated statistics greater than the observed one it stops. Therefore, the range of simulations is (*h*,*n*). As a consequence, the minimum p-value estimate is (g+1)/n, where g is the number of simulations greater than the observed. To provide more precision to the p-value, a distribution is fitted with the simulated likelihood values.

RESULTS

Historical cases of Upper Gastro-Intestinal disease from Feb. to March 2005 were used to compare the capability of detecting clusters for each method. Both arbitrary in space only and in space and time approaches were tested. Results show that all methods provide outbreak detection but the arbitrary in space and time is less sensitive to time overlapped clusters.

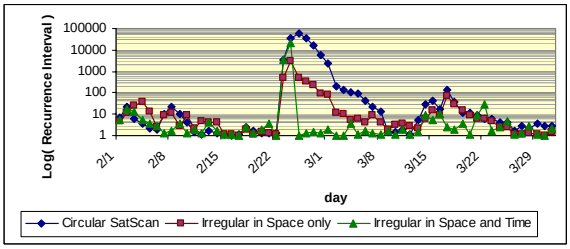

Figure 1 – Logarithm of the Reccurrence Intervals for detected clusters using circular satscan, arbitrary in space only and arbitrary in space and time methods, using Gumbel approximation and Sequential Monte Carlo with h = 50 and n = 1,000.

CONCLUSIONS

We succeeded in implementing a Space Time permutation methodology with irregular shape, improved computational cost and higher p-value resolution. Results show the method's potential to fast detect disease outbreaks.

REFERENCES

[1] Kulldorff M, Heffernan R, Assunção R, Mostashari F, A Space-Time Permutation Scan Statistic for Disease Outbreak Detection, PLoS Medicine. 2005; 2:216-224.

[2] Abrams A, Kulldorff M, Kleinman K. Empirical/Asymptotic P-Values for Monte Carlo-Based Hypothesis Testing: an Application to Cluster Detection Using the Scan Statistic ISDS, 2005.

[3] Besag J, Clifford P, Sequential Monte Carlo p-value, Biometrika. 1991; 78(2):301-304.

[4] Assunção R, Costa M, Tavares A, Ferreira S, Fast Detection of arbitrary shaped disease clusters, Statistics in Medicine. 2006, 25(5) 723-742.

[5] Takahashi K, Kulldorff M, Tango T and Yie K, A Flexible Space-Time Scan Statistic for Disease Outbreak Detection and Monitoring, ISDS, 2006.

Further Information:
Marcelo A Costa, macosta@ufmg.br