

# Using the Flow of People in Cluster Detection and Inference

Sabino J. Ferreira\*<sup>1</sup>, Francisco S. Oliveira<sup>1</sup>, Ricardo Tavares<sup>2</sup> and Flavio R. Moura<sup>2</sup>

<sup>1</sup>Federal University of Minas Gerais, Belo Horizonte, Brazil; <sup>2</sup>Federal University of Ouro Preto, Ouro Preto, Brazil

## Objective

We present a new approach to the circular scan method [1] that uses the flow of people to detect and infer clusters of regions with high incidence of some event randomly distributed in a map. We use a real database of homicides cases in Minas Gerais state, in south-east Brazil to compare our proposed method with the original circular scan method in a study of simulated clusters and the real situation.

## Introduction

The traditional SaTScan algorithm[1],[2] uses the euclidean distance between centroids of the regions in a map to assemble a connected (in the sense that two connected regions share a physical border) sets of regions. According to the value of the respective logarithm of the likelihood ratio (LLR) a connected set of regions can be classified as a statistically significant detected cluster. Considering the study of events like contagious diseases or homicides we consider using the flow of people between two regions in order to build up a set of regions (zone) with high incidence of cases of the event. In this sense the regions will be closer as the greater the flow of people between them. In a cluster of regions formed according to the criterion of proximity due to the flow of people, the regions will be not necessarily connected to each other.

## Methods

We consider a study map with a number of observed cases and risk population for each region. The original circular scan algorithm randomly chooses one region as the first zone and calculates its respective LLR. In the next step a new zone is created including the first region and the region closest to it according the euclidean distance between their centroids and the respective LLR is calculated. This process is repeated until the zone population exceeds a certain percentage of the total population of the map. In our spatial flow scan algorithm everything works in the same manner except that the degree of proximity of two regions is given by the flow of people between them, the higher the flow between the regions closest one is the other. Instead of considering an order of increasing distances to add a region and create a new zone our algorithm uses a decreasing flow of people. In this way we can obtain a zone/cluster candidate composed of a number of non necessarily connected regions.

## Results

Minas Gerais state is located in Brazil south-eastern region composed of 853 municipalities or regions with an estimated population of 19,150,344 in 2005. All data were obtained from the Brazilian Ministry of Health (WWW.DATASUS.GOV.BR ) and Brazilian In-

stitute of Geography and Statistics (WWW.IBGE.GOV.BR). In the period of 2003 to 2008 were recorded 20,912 homicides at a rate of 22 cases per 100,000. To measure the flow of people between the cities we obtain the data of bus round trips between all the 853 Minas Gerais municipalities from state department of highways (www.der.mg.gov.br ). As a large number of pairs of cities have zero bus trips between them we use a gravity model [3] to estimate the flow of people. We use 30% as upper percentage for a zone population. With the real data of homicides cases the original circular scan found a significant cluster containing the city of Belo Horizonte which is the Minas Gerais state capital and large urban area that include Belo Horizonte and 22 more cities totalizing a population of about 3.5 million people. Our adapted spatial scan algorithm also found a similar cluster including the capital Belo Horizonte but with two small cities less.

## Conclusions

In simulation studies where the real cluster is known we observe that our spatial flow scan algorithm has a performance similar to the circular scan concerning detection power and slightly worse in relation to the positive predicted value (PPV) and the sensitivity when the real cluster is regular. However, the performance of our algorithm is clearly better with regard to the sensitivity and the PPV when the real cluster is irregular and or non-connected.

## Keywords

Spatial scan statistics; flow of people; spatial flow scan algorithm; gravity models

## Acknowledgments

SJF acknowledges the support by Fapemig, MG, Brazil.

## References

- [1] Kulldorff M. A Spatial Scan Statistic, *Comm. Statist. Theory Meth.*, 1997, 26(6), 1481-1496.
- [2] Kulldorff M. SaTScan: Software for the spatial, tem-poral and space-time scan statistics. [www.satscan.org].
- [3] Signorino G.; Pasetto R.; Gatto E.; Mucciardi M.; Rocca M. La; Muso P. Gravity models to classify commuting vs. resident workers. An application to the analysis of residential risk in a contaminated area. *Int. J. of Health Geographics*, 2011. 10:11, pp. 1-10.

\*Sabino J. Ferreira

E-mail: sabjfn@gmail.com

