Scan statistics for multiple spatial clusters to investigate geographical disparities of air pollution data

Mohamed-Salem Ahmed[1,2], Michael Genin[1], Matthieu Marbac[3]

[1]Univ. Lille, CHU Lille, ULR 2694 - METRICS: Évaluation des technologies de santé et des pratiques médicales, F-59000 Lille, France.
[2]Socit d'Alicante, Seclin, France.
[3]Univ. Rennes, Ensai, CNRS, CREST - UMR 9194, F-35000 Rennes, France

E-mail for correspondence: matthieu.marbac-lourdelle@ensai.fr

Abstract: Motivated by the investigation of spatial disparities of air pollution data, we develop a new method for detecting multiple spatial clusters and testing their significance. In this context, a geographical cluster has a general parametric shape (allowing for elliptic and rectangular clusters) and is defined by a change of the conditional mean assessed by a regression model of the target variable (here the concentration of particles PM10) given the spatial coordinates and other covariates. We introduce a numerical approach to detect the potential clusters avoiding the use of suboptimal or exhaustive (but unfeasible for large samples) approaches. We present a new Monte-Carlo procedure used for assessing quantiles of the scan statistics under the null hypothesis. We adress the consistency and asymptotic efficiency of the procedure. Contrary to the standard approach, the method permits all the alternative hypothesis to be detected. Finally, the procedure provides a data-driven selection of the number of clusters. The proposed approach is used for analyzing air pollution data given by the European Environmental Agency that the standard spatial scan methods fail to analyze due to the data dimension and to their sub-optimality for detecting some alternative hypothesis that leads to fewer but far away larger detected clusters.

Key words: Cluster detection; Quasi likelihood; Spatial scan statistics.