Geospatial Health
Health Applications in Geospatial Science

Università degli Studi di Napoli Federico II

Dipartimento Medicina Veterinaria Produzioni Animali

pagepress

Appendix

Geospatial Health #696

Examining the impact of the number of regions used in cluster detection methods: an application to childhood asthma visits to a hospital in Manitoba, Canada

The CSS, FSS, and BYM spatial focused cluster detection methods are outlined below.

Circular Spatial Scan Statistic (CSS)

Kulldorff proposed the spatial scan statistic (Kulldorff, 1997), which has been used in a variety of ways in the field of epidemiology. A circular window S is fixed on each region in the circular spatial scan statistic. The radius of the circle ranges from zero to a pre-specified maximum distance d or a pre-specified maximum number of regions J to be included in the cluster. The window composed of the $(j-1)$-*th* closest neighbours to region i is defined as $S_{i:j}(j = 1,2,\ldots,J)$. Then, $S_1 = \{S_{i:j}; i = 1,2,\ldots,m; j = 1,2,\ldots,J\}$ denotes the set of all windows the circular spatial scan statistic is to examine. Based on the number of observed and expected cases inside and outside the circle, a likelihood ratio statistic is calculated for each circle. Here, L_0 signifies the likelihood under the null hypothesis, where it is assumed no cluster exists in region i under the null hypothesis. The alternative hypothesis implies there is a cluster in region i based on the j-*th* nearest neighbours, where $L_i(i = 1,2,\ldots,m)$ represents the likelihood under the alternative hypothesis. Then the likelihood ratio statistic is given by

$$\max_i \frac{L_i}{L_0} = \left(\frac{C_i}{E_i}\right)^{C_i} \left(\frac{N - C_i}{N - E_i}\right)^{N - C_i} I(C_i > E_i), \qquad (1)$$

where the observed and expected number of cases within the circle are denoted by C_i and E_i, respectively. Similarly, the observed and expected number of cases outside of the circle are represented by $(N - C_i)$ and $(N - E_i)$, respectively. The indicator function $\mathrm{I}(\mathrm{C_i} > \mathrm{E_i})$ is equal to 1 when $C_i > E_i$ and 0 otherwise. Potential clusters are identified as circles with high likelihood ratios (Kullforff, 1997).

The SaTScan (Kulldorff *et al.*, 1998) or FleXScan (Takahashi *et al.*, 2006) software can be used to implement the CSS method. Usually, J is chosen to include 50% of the population at risk in a potential cluster. In this study, $J = 15$ was used as this is the FleXScan default. Our study used aggregate data and therefore, the radius of the circle had to include the region centroid for that region to be considered part of the cluster.

Flexible Spatial Scan Statistic (FSS)

The flexible spatial scan statistic (Tango and Takahashi, 2005) performs in a similar manner to the circular scan statistic, although, the shape of the potential cluster in the FSS method is flexible, while still limited to a small neighbourhood of each region. With the flexible spatial scan statistic, an irregularly shaped window S is placed on each region by connecting the neighbouring regions. For any region i, the set of irregularly shaped windows of length j, containing j attached regions including region i, can vary from 1 to the previously set maximum length of the cluster J. Furthermore, to prevent improbable clusters the joined regions are restricted to the subsets of the regions i and $(J-1)$-*th* closest neighbours of region i. Then, $S_2 = \{S_{i:j(k)}; i = 1,2,\dots,m; j = 1,2,\dots,k_{ij}\}$ denotes the set of all windows the flexible spatial scan statistic is to examine. There are J circles examined by the circular spatial scan statistic, whereas the flexible spatial scan statistic observes J circles in addition to all the sets of connected regions whose centroids are found within the J-*th* largest concentric circle. Hence, the size of S_2 is larger than S_1, which is at most mJ. The likelihood ratio test for the FSS test statistic, under the Poisson assumption can be obtained by the equation given in (1), where the circle defined in (1) refers to S_2 instead of S_1. FleXScan software (Takahashi *et al.*, 2006) is used to perform the FSS method, under the default setting $J = 15$. The circles with the highest likelihood ratio values are identified as potential clusters, which is analogous to the circular spatial scan statistic.

Bayesian Disease Mapping (BYM)

First used by Besag *et al.* (1991), the Bayesian approach offers a flexible and robust method for spatial analysis and disease mapping. This method is done via a Bayesian framework using Markov chain Monte Carlo (MCMC). The cases are assumed to follow a Poisson distribution with an area specific parameter $\theta_i E_i$ in the first part of the BYM method

$$C_i \sim Poisson(\theta_i E_i),$$

where the observed number of cases in region i is denoted by C_i and expected number of cases in region i is given by E_i. The second part of the model is then expressed as

$$log(\theta_i) = \mu + \eta_i,$$

where the disease ratio (relative risk) in region i is given by θ_i, μ is the overall log of the mean ratio over the region and η_i signifies spatially correlated random effects. The spatial random effects η_i are modeled using a proper conditionally autoregressive (CAR) model. A variety of CAR models may also be used by taking a collection of mutually compatible conditional distributions $p(\eta_i|\eta_{-i}), i = 1,2,\dots,m,$ where $\eta_{-i}: j \neq i, j \in \delta_i$ and the set of neighbours for the *i-th* region is given by δ_i. The following general model for the spatial random effects η_i is given

$$\eta = (\eta_1, \dots, \eta_m)' \sim N(0, \Sigma_\eta),$$
$$\Sigma_\eta = \sigma_\eta^2 (I - \lambda_n D)^{-1} P^{-1},$$

where P is a $m \times m$ diagonal matrix with elements $P_{ii} = e_i$ with e_i as the number of regions adjacent to region i; D is a $m \times m$ matrix with elements $D_{ij} = 1/e_i$ if region i and j are neighbouring and $D_{ij} = 0$ otherwise; the spatial dispersion parameter is given by σ_η^2; the spatial association is measured by λ_η. Using non-informative priors, the parameters can be predicted in the Bayesian framework (MCMC). The posterior distributions for the parameters in the model are then produced. It is noted that the gamma distribution was used for the inverse of the variance components σ_η^2 with shape and scale parameters 0.001 and a Normal distribution was used with mean 0 and variance 10^6 for the fixed effects. Also, the uniform distribution was used as a prior for λ_η. For the BYM method, a cluster is defined as a region where the estimated disease ratio is significantly larger than two (in terms of their credibility sets). WinBUGS software (Spiegelhalter *et al.*, 2004) was used to apply this method in order to compute the estimated disease ratio values.

References

Besag JE, York JC, Mollìe A, 1991. Bayesian image restoration with two applications in spatial statistics (with discussion). Ann Inst Statist Math 43:1-59.

Kulldorff M, 1997. A spatial scan statistic. Comm Statist: Theor Meth 26:1481-96.

Kulldorff M, Rand K, Gherman G, Williams G, DeFrancesco D, 1998. SaTScan V2.1: Software for the Spatial and Space-time Scan Statistics. Bethesda, MD: National Centre Institute.

Spiegelhalter D, Thomas A, Best N, Lunn D, 2004. WinBUGS version 1.4 User Manual. London, UK: MRC Biostatistics unit, Institute of Public Health.

Takahashi K, Yokoyama T, Tango T, 2006. FleXScan: Software for the Flexible Scan Statistic. Japan: National Institute of Public Health.

Tango T, Takahashi K, 2005. A flexibly shaped spatial scan statistic for detecting clusters. Int J Health Geogr 4:1-15.