SMCC Higher Education Research Journal
ISSN Print: 2449-4402 · ISSN Online: 2467-6322
Volume 9 · January 2022 · H Index = 3 from Publish or Perish

Besag-Newell's And Kulldorff's Spatial Scan Methods in Detecting Spatial Clustering

DANILO G. LANGAMIN
https://orcid.org/0000-0002-1690-2282
boylangamin@yahoo.com
Department of Mathematics and Statistics
Mindanao State University – Iligan Institute of Technology
9200 Iligan City, Philippines

ROSADELIMA V. LOPEZ
https://orcid.org/0000-0001-9040-4479
rosadelimalopez22@gmail.com
Department of Mathematics and Statistics
Mindanao State University – Iligan Institute of Technology
9200 Iligan City, Philippines

Gunning Fog Index: 14.20 • Originality: 99% • Grammar Check 99%
Flesch Reading Ease: 37.06 • Plagiarism: 1%

ABSTRACT

The Besag-Newell and Kulldorff's Spatial Scan are used to detect the spatial clustering of malaria incidence of the 177 barangays in the 11 island municipalities of Tawi-Tawi province. ClusterSeer and SaTScan are the two (2) specialized clustering software that processed the data. The data analysis for the Besag-Newell's method resulted in 51 local clusters using ClusterSeer and the same result of 51 clusters using Kulldorff's Spatial Scan method. The QGIS visualizes these results, thus showing the geographical locations of malaria incidence in Tawi-Tawi province. It was recommended that for those who are fascinated to study more about detecting spatial clustering using Spatial

and Space-time Scan statistics, the researcher recommends that if ever they preferred to use these methods, they have to use some statistical software for the analysis. Also, encourage them to provide another way of interpreting the test's significance aside from the Monte Carlo randomization method to lead them more to an appropriate test.

KEYWORDS

Cluster detection, spatial statistics, Philippines

INTRODUCTION

The study of geographic patterns of disease is part of the characteristic in descriptive epidemiology of "time, person, and place." This is important because it contains powerful, versatile, and useful tools in disease control and prevention. Nowadays, the existing spatial methods are widely known in the statistical community, and some have been in use. But some epidemiologists and public health specialists are not so familiar with these spatial methods (Sankoh & Becher, 2002).

The unfamiliarity of spatial methods is now remedied by accessing newly published textbooks on the subject matter through the internet or updated libraries (Han, 2001). Thus, detection of clusters of diseases can now be done through studies to assist Public Health personnel in their pursuit to identify the location(s) where the outbreak occurs. Hence, the objective of this study is to expose some selected spatial clustering methods, like Besag-Newell for Spatial and Spatial Scan Statistics for Spatial and Spatio-Temporal kind of clustering centering on local type of clustering.

Two cluster detection methods will be used in this study: the Besag – Newell (1991) and Kulldorff's Spatial Scan (Kulldorff 1997; Kulldorff 1998) cluster detection methods. Their cluster detection tests are both based on the Poisson Distribution Model. In the Poisson Model, the cases under the null hypothesis are assumed to be generated from the inhomogeneous Poisson process. The expected number of cases in each area is proportional to its population size or the person-years (Gómez-Rubio et al., 2005).

OBJECTIVE OF THE STUDY

The researchers used a dataset on malaria cases using two different clustering detection methods using specialized clustering software and illustrated different approaches to achieve a specific result of the study. Thus, the main objective of the study was to introduce Besag-Newell's and Kulldorff's Spatial Scan as spatial cluster detection methods

METHODOLOGY

Besag and Newell's Method

This method scans the data to collect cases that appear to be unusual clusters. To do so, it centers a circular window on each region in turn. This window is then expanded to include neighboring regions until the total number of cases in the window reaches a user-specified threshold, k. Then, the population size inside the window is compared to the expected under an average or expected disease frequency. This method can detect local or global spatial clusters in group-level data. The researchers use Besag-Newell analysis in software and get both local and global analysis output. It also describes the extent of local clustering, the number of regions needed to aggregate at least k cases. Thus, use l to evaluate local scale clustering. This method is designed for case and population-at-risk count data aggregated into regions with small populations.

Besag and Newell's Method: Statistics

H_o: *The number of cases in an area follows a Poisson distribution with a common rate, lambda (λ).*

H_1: *For some areas, the number of cases exceeds lambda (λ).*

The null hypothesis is that there is no clustering so that a common Poisson disease rate exists across the study area. Thus, the population at risk inside the window should be proportional to the case count. Otherwise, the null hypothesis can be rejected.

Kulldorff's Spatial Scan Method

Kulldorff's Spatial Scan method can detect local space clusters in group-level data. The scan statistic uses a circular window like Besag-Newell to identify excess cases in space. A circular window increases in size at each spatial location until it reaches an upper limit. The scan statistic provides a measure of whether the observed number of cases is unlikely for a window of that size, using reference values from the entire study area.

Kulldorff's Spatial Scan Method: Statistic

H_o: *The spatial model is an inhomogeneous Poisson point process with intensity λ, proportional to the population-at-risk.*

H_1: *In some locations in the space, the number of cases exceeds that predicted under the null model.*

The likelihood function value is ranked with the maximum likelihood ratio from the Monte Carlo replications. These ranks are called p-values. For a cluster to be statistically significant, its p-value should be less than one. The cluster with the smallest p-value is the most likely to cluster, which has occurred not by chance.

To perform an analysis, the essential requirement is the data. Besag-Newell, (1991) and Kulldorff s Spatial Scan (Kulldorff 1997 & Kulldorff 1998) methods were used in this study to detect clusters of disease outbreaks. The data on the malaria case was gathered from Integrated Provincial Health Office (IPHO) and of Tawi-Tawi Province from Regional Health Unit (RHU).

The researcher selected Tawi-Tawi malaria cases for 2003, 2008, 2013, and 2013-2014 for Bongao malaria cases. The researcher examined 177 barangays in 11 island municipalities of Tawi-Tawi province for 2014 with a total population–at–risk approximately 360,000 and with malaria cases of about 1800 people. Kulldorff's Spatial Scan and Beasg-Newell's methods were used to detect spatial pattern, specifically, local spatial. The 92 malaria cases of Bongao municipality with 35 barangays were used for space-time clustering.

Statistical Analysis

To detect clusters of disease outbreaks, statistical analysis is done using ClusterSeer and SaTScan. ClusterSeer is used for the Besag-Newell method and SaTScan for the Spatial Scan method. In the input tab, data imported to this two software were transformed into text files form to execute, read, and analyze the data to obtain the needed statistics.

RESULTS AND DISCUSSION

Besag and Newell's Method Result Using ClusterSeer

Using ClusterSeer, Besag-Newell's Method is run restricting the spatial search of clusters in the whole study region (177 barangays), including 680 cases for the entire one-year period (2013). The implicit settings of the program are used: the researcher chose the cluster cut-off size (k) to detect is 51, the number of Monte Carlo simulations performed = 999, and default alpha level 0.05 is used. After successfully running the software, the analysis revealed fifty-one (51) local clusters.

Table 1. Null Hypothesis and Monte Carlo Distribution

Significant Local Clusters	Expected Size of (R) Under Null Hypothesis	Upper-tail P-value for r Monte Carlo Distribution (999 simulation runs)
51	4.27363	0.001000

Table 1 explains that 51 significant local clusters have assimilated with the expected size of (R) under null hypothesis = 4.27363 and with an upper-tail p-value for r Monte Carlo Distribution (999 simulation runs) to be 0.001000 is less than compared to alpha 0.05. Thus, significant local clusters of 51 are statistically significant, and the occurrences are not by chance.

Spatial Visualization Analysis in ClusterSeer

ClusterSeer calculated the total number of significant local clusters (r) and the expected size of (R) under the null hypothesis and its significance for all clusters using the method from Waller et al. (1994). An alpha of 0.05 provides the threshold for significance, and the software gave the list of all clusters with a probability less than the alpha.

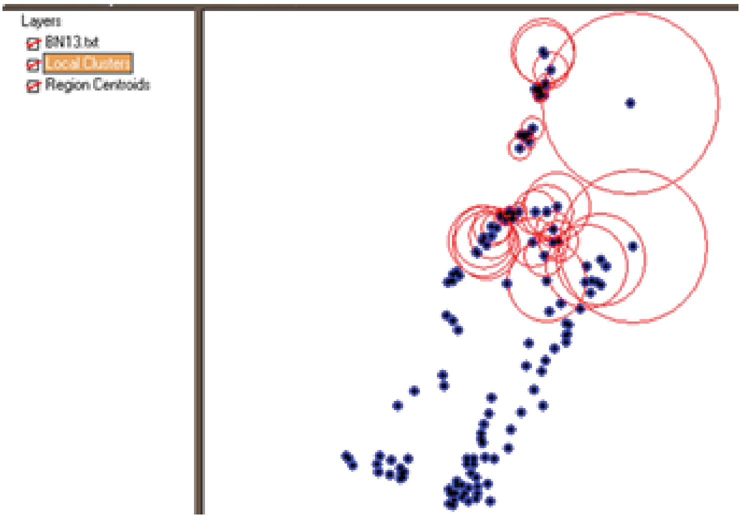

Figure 1: Map View of Malaria Data, Region and Test Statistic

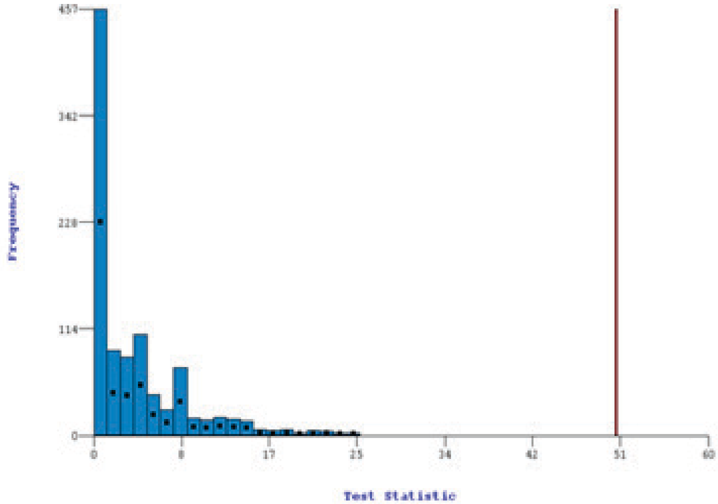

Figure 2: ClusterSeer: the Frequency Centroid and Local Clusters

Kulldorff's Spatial Scan Method Result Using SaTScan

Datasets 2003, 2008, and 2013 were used in SaTScan, and the file was saved as twt.case, twt.pop, and twt.geo. SaTScan supports the extension for case files containing the location ID, case count of each location and year, for a population that contains location ID, year and population-at risk of each location, and coordinates that contain the location ID, latitude, and longitude information about each location. The study period was specified according to the period specified in the dataset and time precision to be year and coordinates latitude/longitude.

Table 2 Details of Detected Clusters (Malaria Case 2013)

CLUSTER	LOC_ID	LATITUDE	LONGITUDE	RADIUS	START_DATE	END_DATE	NUMBER_LOC	LLR	P_VALUE	OBSERVED	EXPECTED	ODE	REL_RISK	GINI_CLUST
1	39782	5.173262	120.49916	33.3406	1/1/2013	12/31/2013	51	669.146458	1.00E-17	555	118.50379	4.6833946	21.037667	F
2	39698	5.086466	119.88982	0.0000	1/1/2013	12/31/2013	1	9.442273	6.70E-03	17	4.78704	3.5512525	3.616669	F
3	39746	4.852321	119.46955	0.0000	1/1/2013	12/31/2013	1	3.958119	6.65E-01	9	2.97775	3.0224142	3.049540	F

1. There are 51 locations identified hot-spot clusters included in the first cluster are 39782, 39791, 39783, 39798, 39784, 39785, 39799, 39795, 39788, 39797, 39800, 39777, 39796, 39780, 39778, 39786, 39787, 39793, 39794, 39789, 39779, 39824, 39815, 39807, 39822, 39810, 39814, 39811, 39819, 39817, 39818, 39792, 39808, 39816, 39790, 39716, 39721, 39823, 39727, 39813, 39719, 39820, 39821, 39728, 39726, 39781, 39812, 39723, 39718, 39809, 39806 with number of cases observed as 555 and expected number of cases 118.5037 with likelihood ratio to be 669.1465 approximately, first highest among detected clusters. It is statistically significant because its p-value is 0.00000000000000001 which shows that its occurrence is not by chance.

Figures are the Spatial displays, viewed the malaria case map 2013 with locations, detected clusters and coordinate of each location and likewise.

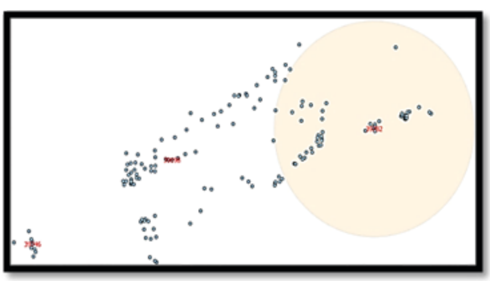

Figure 3. Spatial Display of Detected Clusters combining coordinates of each Location of Malaria Case Year 2013

Figure 4. Spatial Display of Detected Clusters combining Map and Coordinates of each Location of Malaria Case Year 20013

The reasons for having a high number of outbreaks in the areas related to this cluster are as follows: Most areas are rural and remote, so control activities are quite difficult. People lack awareness about the prevention and control of malaria disease. Malaria is caused by the plasmodium parasite. The parasite can be spread to humans through the bites of infected mosquitoes. Infected individuals by malaria disease coming from other places could transmit the malaria disease to the people living in the particular area through the bites of infected mosquitos when they come to visit their kinsfolks or maybe to come to do business.

Table 3. BN's and Kulldorff's SS methods in detecting spatial clusters using two specialized software (ClusterSeer and SaTSCan)

The researchers used the 2013 malaria data of Tawi-Tawi as the entire study population comprising 177 regions/barangays to ascertain the use of Besag-Newell's and Kulldorff's Spatial Scan methods in detecting spatial clusters using two specialized software (ClusterSeer and SaTSCan). The researcher's objective is to see if methods based on circles can also detect non-circular clusters. Two different methods have been used to locate and identify the clusters. The purpose of this study is to evaluate the performance of two methods, BN and KSS, for local cluster detection for similar types of spatial clusters (local type of clustering) using two specialized software for cluster detection.

It further shows all the dataset values, including 51 detected clusters and their respective circular spatial scan statistics. All p-values of the test are based upon the distribution of likelihood ratio test statistic with (999) of Monte Carlo replications of the data set generated under the null hypothesis. The alpha level used is the default alpha 0.05.

CONCLUSIONS

Besag-Newell's and Kulldorff's Spatial Scan methods were used to obtain result of analysis of Tawi-Tawi malaria case in year 2013 using ClusterSeer and SaTScan to detect spatial clusters. Default number of Monte Carlo simulation run (999), and alpha 0.05 levels were used to determine statistical significance of detected clusters. Statistical software had generated the detail of clusters information for the basis of interpreting the significance of the test presented in this paper.

The main objective of this study is not to compare the result of Besag-Newell's and Kulldorff's Spatial Scan methods evaluated by this two specialized clustering software but instead to examine on how these two statistical methods perform in detecting spatial clustering and to identify statistically the significance occurrences of detected clusters whether they occurred by chance or not by chance.

With the aid of QGIS with its standalone packages and programs, geographical visualization of the results has been done. It is helpful in identifying hot-spot clusters having prevalent disease of malaria. With the geographic visualization method maps were generated to present the spreading of cases, density of cases within each location or county and to show risk different significant clusters.

Hence, Besag-Newell's and Kulldorff's Spatial Scan methods are compelling methods to detect emerging clusters in space and space-time. These methods could be used not only in rural but also could be used efficiently in urban studies and planning to detect areas where certain kind of disease is endemic that could lead to a loss in the quality of life of the people.

RECOMMENDATIONS

For those who are fascinated to study more about detecting spatial clustering using Spatial and Space-time Scan statistics, the researcher recommends that if ever, they prefer to use these methods to use some statistical software for the analysis. Also, encourage them to provide another way of interpreting the test's significance aside from the Monte Carlo randomization method to lead them more to an appropriate test.

LITERATURE CITED

Block, R. (2007). Software review: scanning for clusters in space and time: a tutorial review of SatScan. Social Science Computer Review, 25(2), 272-278. Retrieved on November 10, 2021, from https://bit.ly/3HwDKrb

Durbeck, H., Greiling, D., Estberg, L., Long, A., Jacquez, G., Pallicaris, Y., & Hinton, S. (2012). ClusterSeer Software for The Detection and Analysis of Event Clusters, User Manual Book 2, version 2.5. BioMedware geospatial research and software.

Retrieved on November 21, 2021, from https://bit.ly/3skrROB

Gandhi, U. (2018). QGIS Tutorials and Tips. Last updated on Apr, 30. Retrieved on November 12, 2021, from https://bit.ly/3J4asQY

Gómez-Rubio, V., Ferrándiz-Ferragud, J., & López-Quílez, A. (2005). Detecting clusters of disease with R. Journal of geographical systems, 7(2), 189-206. Retrieved on November 12, ,2021, from https://bit.ly/3AZnnAS

Han, J. (2001). Spatial clustering methods in data mining: A survey. Geographic data mining and knowledge discovery, 188-217. Retrieved on January 10, 2021, from https://bit.ly/3gtt68w

Mala, S., & Sengupta, R. (2013). Geo-visual approach for spatial scan statistics: an analysis of dengue fever outbreaks in Delhi. Editorial Preface, 4(10). Retrieved on November 10, 2021, from https://bit.ly/3J8fUCw

Sankoh, O. A., & Becher, H. (2002). Disease cluster methods in epidemiology and application to data on childhood mortality in rural Burkina Faso. Informatik Biometrie und Epidemiologie in Medizin und Biologie, 33(4), 460-472. Retrieved on November 10, 2021, from https://bit.ly/3gtRDKG

Tango, T., & Takahashi, K. (2005). A flexibly shaped spatial scan statistic for detecting clusters. International journal of health geographics, 4(1), 1-15. Retrieved on November 12, 2021, from https://bit.ly/3B3ZBUd

Tuia, D., Kaiser, C., Da Cunha, A., & Kanevski, M. (2007). Socio-economic cluster detection with Spatial Scan Statistics. Case study: services at intra-urban scale. In Geocomputation 2007, National University of Ireland, Maynooth, 3-5 September 2007. Retrieved on November 12, 2021, from https://bit.ly/3gpasyT