Journal of Information Technology and Digital World (ISSN: 2582-418X)
www.irojournals.com/itdw/

Getis-Ord (Gi*) based Farmer Suicide Hotspot Detection

Amisha Bharti[1*] and Sonajharia Minz[2]

[1,2]School of Computer and Systems Sciences, Jawaharlal Nehru University, New Delhi, India

**E-mail: ***amisha3360@gmail.com

Abstract

Farmer suicidal hotspot detection proposed in this paper aims to reduce the death of the farmers. Using geographical information system is vital in predicting potential hotspots for farmer suicide. This study has collected and analyzed data on farmer suicide in India, using state-wise information from the National Crime Records Bureau and has determined the recent higher rate of farmer suicide. Spatial statistics analysis tools that address average nearest neighbor analysis has been used. Global analysis through Moran's Index, analyzed that the farmer suicides have a clustered pattern and plotted a farmer suicidal hotspot map using Getis-Ord (Gi*) analysis. The results show the highest farmer suicide index is in Maharashtra and hence, farmer suicidal hotspot has been found district wise. There are four farmer suicidal factors such as, number of farmer suicide, the population density of farmers, climate, and income. This hotspot geographical region helps to identify future suicidal risk by studying the hotspot map. Moreover, government policy may suggest a hotspot zone to help the overall development of the country's growth.

Keywords: Farmer suicidal factor, Moran's Index, Hotspot Analysis, Population pattern

1. Introduction

Farmer suicide is a grim issue for our society and has also found an increasing growth rate over the last two decades. It is susceptible that the farmer suicidal death increases. In 2019-2020, the farmer suicidal rate increased by 10%. This paper aims to detect the hotspot region, which helps to take new progressive initiatives and policies to reduce the risk of suicide.

India is an agriculture-based country that has very diversified geomorphological areas. In India, 70% [12] of the population has their livelihood on farming, and agriculture

DOI: https://doi.org/10.36548/jitdw.2022.2.002
Received: 18.05.2022, received in revised form: 11.06.2022, accepted: 26.06.2022, published: 06.07.2022

also contributes to approximately 15% of GDP (Gross Domestic Product) to our country's growth. Agriculture also participates in the employment of 54% of our country's labor. Indian agriculture contains three sections for working; first is named as farming, where farmers work on their land, the second one is labor, and the last is farming sector employees. The risk of farmer suicide is very high in small and marginal farmers.

A hotspot is a region where the occurrence of a particular event is very high. The National Crime Records Bureau (NCRB) [2] data has been used to build a model to analyze farmer suicidal hotspots. A spatial hotspot is one of the methods that analyze farmers' suicide. Some particular places are witnessing more farmer suicide than others. For computing farmer suicidal hotspot, four factors are considered state-wise to determine the hotspot: the number of farmer suicide, population density, environmental factor, and economic stress such as the income of farmers, loans, and expenditure.

In this paper, Getis-Ord (Gi*) [20] is used for hotspot analysis and further spatial autocorrelation Moran′s Index [19] is applied to map farmer suicidal hotspot zones. After applying the Hotspot analysis technique, it is found that Maharashtra, Karnataka, and Andhra Pradesh have the highest number of farmer suicide cases. The district-wise data from the criminal investigation department (CID) or State Crime Records Bureau (SCRB) [13] has been used. In Maharashtra, farmer suicide has followed an increasing trend in the last decades. NCRB Report has published that, 21474 farmers have committed suicide from 2013 to 2020 [2]. Using division-wise statistical techniques in Maharashtra hotspot detection, it is found that Amaravati and Aurangabad have the highest number of farmer suicides. There are many applications of hotspot detection in medical field [3, 4, 6, 7].

The rest of the paper is arranged as follows. Section 2 contains the study area and methodology. Section 3 summarizes the experiments and results, and Section 4 concludes this paper.

2. Study Area and Methodology

2.1 Study Area

India with a population of 1,369 million people stands at the world's second rank. In the Asian continent, the geographical location of India lies in between 08◦04′N to 37◦06′N and 68◦07′E to 97◦25′E, with 28 states and 8 union territories. It is the seventh-largest country globally, with a territorial area of 3.288 million square kilometers with 60.43% land used for agriculture purposes according to the year 2018. About 58% of India's population primarily depends on agriculture and allied activities for their livelihood.

2.2 Methodology for farmer suicide

The following method has been implemented to find the farmer suicide hotspot. The number of farmer suicide, population density, environmental factor, and economic factors [1] are used to find hotspot detection using the following methods.

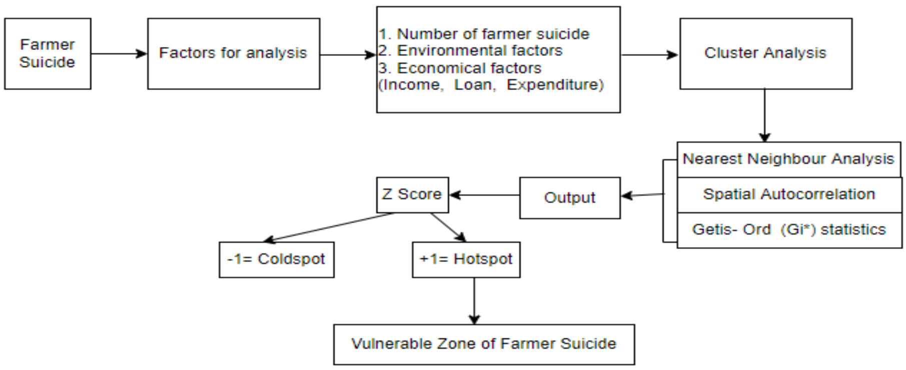

Figure 1. The methodology used for farmer suicide hotspots detection.

2.3 Nearest Neighbour Analysis

Nearest Neighbour Analysis (NNA) [10] is the statistical technique used for mapped point pattern analysis. It measures the distance between the sample data point and its nearest neighbor. There are two methods to measure the distance to the nearest neighbor, such as 'sample located to nearest point' and 'point to nearest point'. The 'sample located to nearest point' randomly selects a location in the study area and measures the distance to the nearest point. The 'point to nearest point' approach randomly selects one point from all the points and measures the distance to the nearest neighbor. NNA uses observed mean distance and expected mean distance. The observed mean distance calculates as the mean of the nearest neighbor's distance. The expected mean distance derived from the Poisson probability distribution assumes that the sample area is a circle with a radius *r*. The mean of *r* is the expected mean distance to measure the nearest neighbor *E(r)* in a pattern generated by the Poisson probability distribution process, as defined in the equation (1).

$$E(r) = \frac{1}{2(\sqrt{\lambda})} \quad (1)$$

where, λ is the point per unit area, the standard error of E(r) is given as in equation (2), and N is the number of distance measurements made. To test the significance of departure from the Poisson model, equation (3) is used.

$$\sigma_r = \frac{0.2613}{\sqrt{N\lambda}} \tag{2}$$

$$Z = \frac{\bar{r} - E(r)}{\sigma_r} \tag{3}$$

where, Z is the standard variable of the normal curve, and the observed mean distance to the nearest neighbor is $\bar{r}$. The p-value validates the hypothesis against the observed value. The relation between the p-value and z score is shown in Fig. 2.

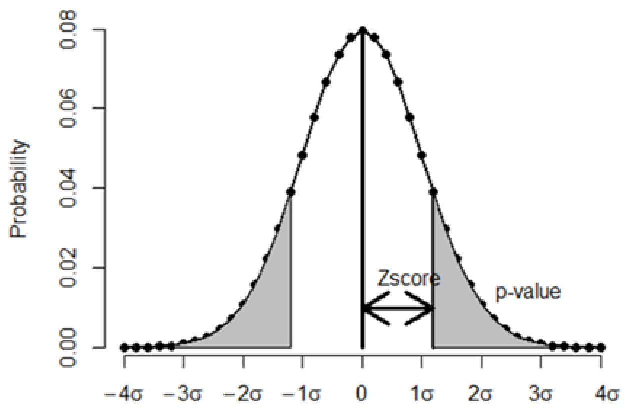

Figure 2. Z score and p-value

For NNA, the Nearest Neighbor Ratio (NNR), which is the ratio of expected mean distance and observed mean distance has been computed as in equation 4.

$$NNR = \frac{Expected\ Mean\ Distance}{Observed\ Mean\ Distance} \tag{4}$$

NNR scale has defined the property of data points association as the cluster, random, and uniform distribution.

2.4 Spatial Autocorrelation

Spatial autocorrelation [11] is defined as the relationship between spatial objects and their neighbors. It depends on the relative distribution of spatial objects of interest and their nearest neighbor. The advantage of multi-dimension spatial autocorrelation is that it performs in multi-dimensional space. These types of spatial autocorrelation [16] are associated with positive, random, and negative. The positive autocorrelation occurs when the object has a high association in a cluster. When a spatial object with a particular property shows no pattern of association, it is known as random autocorrelation. The negative autocorrelation occurs when spatial objects with a particular property are distributed evenly over a large geographical space. For calculating spatial autocorrelation, Moran's Index method is used.

2.4.1 Moran's Index

Moran's Index (Moran's I) [19] is a multi-dimensional spatial autocorrelation statistical technique proposed by P. Moran. This method follows a global measure of spatial autocorrelation based on the location and attribute value. Moran's I use network autocorrelation analysis techniques to determine the index value for each point and check whether the pattern is clustered, dispersed, or random.

A mathematical explanation of Moran's I statistical is in equation 5.

$$Moran's\ I = \frac{n}{D_\beta} \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} H_{ij}(y_i - \bar{x})(y_j - \bar{x})}{\sum_i (y_i - \bar{x})} \quad (5)$$

$$D_\beta = \sum_i \sum_j H_{ij} \quad (6)$$

where, $\bar{x}$ = mean of a variable.

n = number of the segment.

y_i = value of variable a on segment i.

H_{ij} = spatial weight matrix of i and j.

D_β = collection of spatial weights.

$i =$ 1, 2, 3, ..., n and $j =$ 1, 2, 3, ….., n.

Based on the Moron's I, the statistical significance test is based on normal frequency distribution by using equation (7)

$$Z = \frac{I - E(I)}{S_{error\ (I)}} \quad (7)$$

$$E(I) = \frac{-1}{n-1} \quad (8)$$

where, i is the sample calculated value of Moron's I, and E(I) is the expected value.

The value of I lies between -1 to +1. The low and high value of Z shows negative and positive autocorrelation, respectively. The autocorrelation can be the same throughout the space, thus, making it less realistic. To overcome this problem, a local measure of spatial autocorrelation has been used.

2.5 Getis – Ord (Gi*)

Getis – Ord (Gi*) [21] is a statistical technique based on a local measure of spatial autocorrelation. Getis and Ord have developed this concept. For calculating Gi*, equation 9 is used.

$$\text{Gi*} = \frac{\sum_{j=1}^{n} y_{ij} x_j - \bar{X} \sum_{j=1}^{n} y_{ij}}{s\sqrt{\frac{[n\sum_{j=1}^{n} y_{ij}{}^2 - (\sum_{j=1}^{n} y_{ij})^2]}{n-1}}} \tag{9}$$

where, x_j is the feature value, $y_{i,j}$ is used as the spatial weight between

feature i and j, and the total number of features is written as n.

$$\bar{X} = \frac{\sum_{j=1}^{n} x_i}{n} \tag{10}$$

$$s = \sqrt{\frac{\sum_{j=1}^{n} x_j{}^2}{n} - (\bar{X})^2} \tag{11}$$

The large value of Gi*[5] indicates a high-value cluster together. If Gi* is low, it shows a low-value cluster together. It is a potential coldspot when the Gi* value is less than the expected value. The Gi* [9] greater than expected value indicates the potential hotspot.

3. Results and Discussion

The farmer suicide data has been used for hotspot detection, which shows the states with individual cases of farmer suicides.

3.1 Dataset

Farmer suicidal data from 2005 to 2020 [2] are collected from the National Crime Records Bureau (NCRB). The farmer population density in a state is the ratio of the total number of farmer population [12] [17] to the corresponding total area of the state. The farmer population and total population data are collected from the official database of the Registrar General and Census Commissioner [14, 15]. Environmental factors like rainfall, drought, and flood data are collected from the Indian Meteorological Department [8]. Agricultural economic sources are farmer income, loans, and farming expenditure [8]. For Maharastra, district-wise data are collected on farmer suicide reported per the government portal. Suicidal data are collected from the State Crime Records Bureau (SCRB) of the Criminal Investigation Department (CID) [13]. The geographic information system database and the spatial distribution for the Maharashtra are used. Shapefile are created and mapped for Maharashtra district-wise farmer suicide.

3.2 Average Nearest Neighbour for farmer suicide

The average nearest neighbor value, which shows the computed distance of the observed mean 33.6m is used, and 47.7m is the distance of the expected mean. The average NNR is 1.42, which is greater than 1.0, indicates a cluster as a polygon of farmer suicide. The z-score is 4.50, which is greater than -2.58. The average nearest neighbor computed value represents the results of farmer suicide, as shown in Fig. 3.

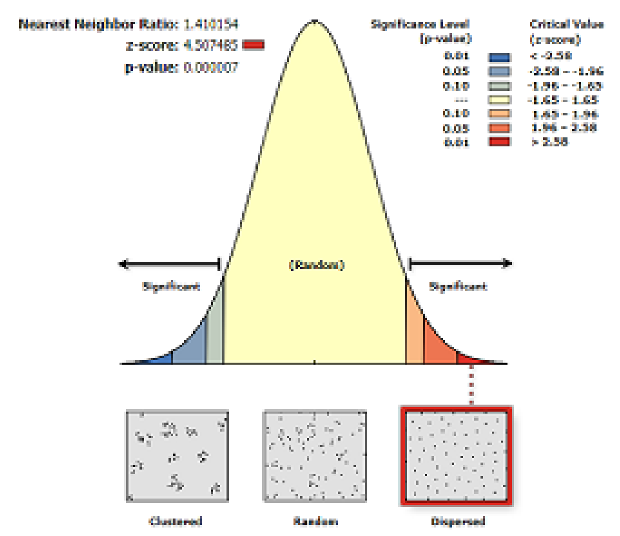

Figure 3. Nearest neighbor ratio

3.3 Moran's Index using Spatial autocorrelation for farmer suicide

Spatial autocorrection is a method that finds a similar feature in the nearest distance. Moran's I statistic [19] uses four factors to determine the clustered hotspot analysis for farmer suicide.

In statistical analysis, the average Moran I [18] is 0.1092, as given in the table below. This value shows that the probability of random chance is less than 1%. Therefore, this reveals that each pattern has similar values that follow a similar distribution. It can be low as well as high. Farmer suicide value shows the Moran's I [19], expected Index, z value, and p-value for 2019-2020 data.

Table 1. Spatial analysis for farmer suicide in year 2019-2020

	Moran I	Expt. Index	z score	p-value
Farmer Suicide	0.1092	-0.0285	2.4131	0.0158
Pop_Density	0.0157	-0.0285	0.9384	0.3480
Env Factors	0.1042	-0.0285	2.3200	0.0203
Economic fact.	0.0594	-0.0285	1.9466	0.0515

3.4 Getis-Ord (Gi*) based Hotspot Detection for farmer suicide

Getis-Ord (Gi*) [21] is used as a spatial network structure, and the weight matrix is used to calculate the (Gi*) value by using the above equation (9). Computed values measure the region's high and low concentrations for hotspot detection. Gi* statistics uses the threshold distance value of 550008, which is used for Euclidean distance value. Its value always lies between +2.56 to -2.56. The map scaling uses 1:25,000,000 in Arc GIS. It uses the z value to identify the hotspot. The p-value and z score values are positive values with high values that are more spatially clustered, defined as a hotspot (figure 4).

3.5 Hotspot area

In figure 4, the result shows that Maharashtra, Karnataka, and Andhra Pradesh are the hotspot of farmer suicide, with Degree of Confidence of 99%, 95%, and 90%, respectively.

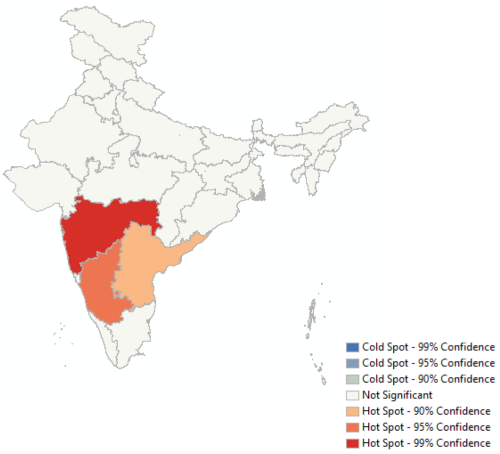

Figure 4. Farmer suicide hotspot

In the district-wise analysis of Maharashtra, there are six divisions and 35 districts in Maharashtra [3]. After statistical hotspot mapping, the Amravati has the highest number of farmer suicide, followed by the Aurangabad division in Maharashtra. Figure 5 shows Maharashtra district-wise hotspot detection. The red color indicates the highest number of farmer suicide in this region, the orange color shows a smaller number of farmer suicide compared to red, and the least is in the blue region on the Maharashtra map.

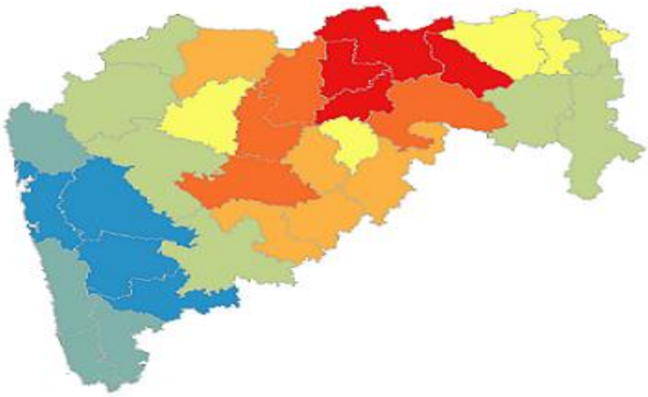

Figure 5. Maharashtra Farmer Suicide hotspot.

4. Conclusion

This study has collected and analyzed state wise data on farmer suicide in India. In this paper, spatial analysis techniques that help address the Average Nearest Neighbor are used, and Moran's I is used to analyze global patterns. The computed value on farmer suicidal data has mentioned clustered pattern for graphical map representation of farmer suicidal hotspot using Getis-Ord techniques. The results represent the highest number of farmer suicide indexes in Maharashtra. Small and marginal farmers are more at risk of suicide due to the loss of farming sunk costs. Farmer suicidal hotspot detection helps to reduce the risk of the same by using new initiatives in agriculture techniques.

References

[1] https://www.fao.org/india/fao-in-india/india-at-a-glance/en/

[2] https://ncrb.gov.in/

[3] Ade, Vikas V.: "Trends and causes of farmers suicide in Maharashtra State, India." International journal of scientific research in science, engineering and technology 8.1 pp248-261(2021).

[4] Jeefoo, Phaisarn, Nitin Kumar Tripathi, and Marc Souris. "Spatio-temporal diffusion pattern and hotspot detection of dengue in Chachoengsao province, Thailand." International journal of environmental research and public health 8.1 pp51-74(2011)

[5] Ord, J. K. and Getis, A. (1995) Local spatial autocorrelation statistics: distributional issues and an application. Geographical Analysis 27, 286-306.

[6] Parvin, Farhana, et al. "Spatial prediction and mapping of the COVID-19 hotspot in India using geostatistical technique." Spatial Information Research 29.4 pp479-494(2021).

[7] Chen, Chien-Yuan, and Qi-Hua Yang. "Hotspot Analysis of the Spatial and Temporal Distribution of Fires." GISTAM. 2018.

[8] https://mausam.imd.gov.in/imd latest/contents/index rainfall state new.php

[9] Manepalli, Uday RR, Ghulam H. Bham, and Srinadh Kandada. "Evaluation of hotspots identification using kernel density estimation (K) and Getis-Ord (Gi*) on I-630." 3rd International Conference on Road Safety and Simulation. 2011

[10] Okabe, Atsuyuki, Barry Boots, and Kokichi Sugihara. "Nearest neighbourhood operations with generalized Voronoi diagrams: a review." International Journal of Geographical Information Systems 8.1 (1994): 43-71.

[11] Cliff, Andrew D., and Keith Ord. "Spatial autocorrelation: a review of existing and new measures with applications." Economic Geography 46.sup1 (1970): 269-292.

[12] https://www.fao.org/india/fao-in-india/india-at-a-glance/en/

[13] https://mahacid.gov.in/

[14] https://mospi.gov.in

[15] https://censusindia.gov.in/census.website/

[16] Pearson, Karl. "Determination of the coefficient of correlation." Science 30.757 (1909): 23-25.

[17] https://eands.dacnet.nic.in/Agricultural

[18] Waldhör, Thomas. "The spatial autocorrelation coefficient Moran's I under heteroscedasticity." Statistics in Medicine 15.7- 9 (1996): 887-892.

[19] Moran. P. (1948) The interpretation of statistical maps. Journal of' the Royal Statistical Society lob. 243-25 1.

[20] Getis, A. Spatial interaction and spatial autocorrelation: a cross-product approach. Environment and Planning A 23, 1269- 1277.