

Evaluation of Fire Ignition Probability Model by Using Spatial Autocorrelation Method: A Case Study in Yunnan Province

Wei WANG, Zhanqiang CHANG, Xiaomeng LIU, Jie ZHU & Wen YU

College of Resource, Environment & Tourism, Capital Normal University, Beijing, China

Key Lab. of 3D Information Acquisition of Education Ministry of China, Beijing 100048, China

Keywords: spatial autocorrelation analysis; model residuals; Yunnan Province

Abstract: The forests in Yunnan province are dense and precious. Meanwhile, the forests fire happens frequently. There are ten models generated with two modeling methods in the study area, Yunnan province. In this paper, the spatial autocorrelation method was applied to analyze and evaluate the performance of fire ignition probability models. The results show that: 1) the models generated by geography weight logistic regression method are better than the ones created by binary logistic regression method; 2) compared with other models, the Model 7 is optimal in terms of evaluating by the quantitative spatial autocorrelation method. The results are of significance for making decisions of establishing early warning systems.

1 Introduction

During the Spring Festival in 2013, forest fires broke out 39 times totally throughout China, including 22 times in Yunnan province. Due to the frequent forest fires in Yunnan province, the fire ignition probability model is especially fatal for prevention work of forest fires. Forest fires study starts late in China (Di et al. 1993). After forest fires of the Greater Hinggan in 1987, China started the many projects of researching on fire prevention and extinguishing (Hu et al. 2005). The aspects of extracting fire point, monitoring forest fires and assessing disaster losses, etc., have achieved good results and applications (Shu et al. 1998).

In view of the logistic regression model based on multivariate forecasting ability for the specified event and the input data of the model without the need of normal distribution (Legendre et al. 1998), logistic regression is a method which has been widely applied in fire simulation (Keane et al. 2010). In the previous study, ten models were generated by logistic regression method. This paper is designed to focus on evaluating performance of the fire ignition probability model by using spatial autocorrelation method.

2 Study area and related data

2.1 Study area

Located in the southwest border of China, Yunnan province has dense forests and enjoys subtropical monsoon climate. Many rare, ancient, unique and endangered animals and plants exist in the dense virgin forests of Yunnan province. These precious animals and plants are not renewable biological treasure for human beings. However, the dense forests of Yunnan province are very prone to fires seasonally. Therefore, it is vital for fire prevention of the dense forests to build a fire ignition probability model.

2.2 The related data

The former and the latter five models were generated by binary logistic regression and geography weight logistic regression method, respectively. Each model is made up of many points. So there are eleven documents of shapefile format including a polygon file representing the border of Yunnan province and ten multipoint files stand for monitoring points of ten models. For simplicity, we only take three property fields (X value, Y value and residual error) of every multipoint file. The attribute table of the first model show in Table 1.

Table 1 Attribute table of Model 1

Point number	X	Y	Residual
	coordinate	coordinate	
	m	m	
1	-247,227	2,946,879	0.131670
2	165,525	2,945,575	0.884693
...
560	112,175	2,638,732	0.092817

The basic form of statistical model is represented by the Eq.1:

$$Y=AX+R. \quad (1)$$

where Y = dependent variable, A = constant, X = independent variables and R = random residual errors of every model. As is known to all, the more random residuals are, the better performance model has. Hence, we can evaluate these models by estimating the randomness of the distribution of the residuals in the spatial autocorrelation method.

3 Spatial autocorrelation method

Spatial correlation exists when there is a significant similarity or dissimilarity between the values of a variable Z at all pairs of adjacent locations (Li et al. 2009). If these attribute values (values of variable Z) along with the shrinking of distance between two points are close or diverging, the attribute shows significantly positive or negative spatial correlation. If these attribute values have no obvious change trend, the spatial distribution of this attribute of points is random. Generally, Moran's Index (Moran's I) is used to measure spatial correlation.

3.1 Global Moran's I statistic

Global Moran's I is used to describe the overall situation of a certain attribute. The maximum values of global Moran's I is 1, which means absolute positive spatial autocorrelation, for example, high values or low values cluster together. The values of global Moran's I is 0, which means a random pattern. And the minimum values of global Moran's I is -1, which indicates an absolute negative spatial autocorrelation, for example, high values cluster with low values (Li et al. 2009).

3.2 Anselin Local Moran's I statistic

Since Global Spatial Autocorrelation cannot point out the correlation in which areas explicitly, Local Spatial Autocorrelation is introduced. Anselin Local Moran's I measure the related degree between value of each unit and its adjacent unit, and can be used to evaluate local spatial agglomeration and variation of high and low values (ESRI 2010).

4 Analysis by global and local spatial autocorrelation

4.1 Global Spatial Autocorrelation

In order to analyze these data effectively and comprehensively, this study mainly adopted the Incremental Spatial Autocorrelation tool of ArcMap10.1 to carry on the global spatial autocorrelation analysis. The Incremental Spatial Autocorrelation tool runs the Spatial Autocorrelation tools for a battery of increasing distances, measuring the intensity of spatial clustering or space differentiation for each distance (ESRI 2010). However, the beginning distance and distance increment should be identified and set up before using the Incremental Spatial Autocorrelation tool.

Therefore, first of all, Near tool was used to calculate the biggest distance from adjacent points of each model. And then the maximum value of this biggest distances was set to the beginning distance. Secondly, through a survey for the scope of the Yunnan province, we decided to enlarge the neighborhood size twenty times, and set up 5000 meters as distance increment to analyze each model. Finally, the Incremental Spatial Autocorrelation tool was used for processing residuals of each model and generating data. The Table 2 is an example for the results of Model 7 processed by the Incremental Spatial Autocorrelation tool.

Table 2 the Results of Model 7 Processed by the Incremental Spatial Autocorrelation Tool

ID	Distance km	Moran's I	P value
1	99.4	0.002187	0.608515
2	104.4	0.002110	0.598042
...
20	194.4	-0.000485	0.819266

Ten tables of results processed by Incremental Spatial Autocorrelation Tool drawn in Figures 1:

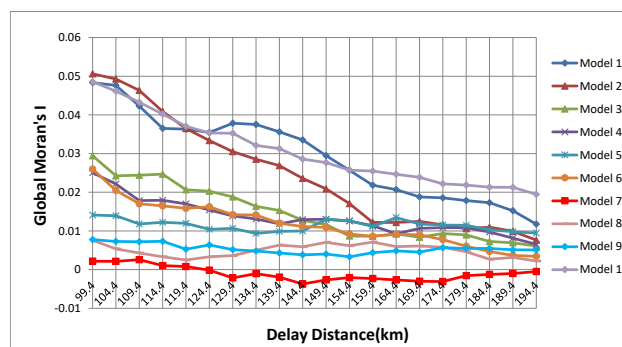


Figure 1 Global Moran's Indexes of Ten Model's Residuals in Twenty Delay Distances

These Global Moran's indexes of ten models' residuals are closing to zero gradually with the enlargement of the research scope. That is to say, the residuals of ten models are random in the scope of the whole Yunnan province. Among them, the Global Moran's I of the residuals of Model 7, Model 8 and Model 9 are very close to the zero from beginning to end. Specially, the Moran's I of Model 7 in twenty distances are always fluctuating up and down around zero, and their absolute values are not more than 0.004. In other words, Model 7 may be a perfect model, because its residuals are most random compared with other models in any scope of research.

Although the Global Moran's I of Model 7 are the most close to zero, the Moran's Indexes of the other models are also very small. Furthermore, the indexes of spatial correlation of ten models are approximating zero in the wake of enlargement of the research scope. Therefore, only rely on the numerical size of Global Moran's I, it is not precise to evaluate the performances of models in this study. We can further test these P values of ten Global Moran's indexes, and confirm whether the Global Moran's indexes are caused by the null hypothesis. These P values in twenty delay distances were depicted with a line chart (Fig. 2).

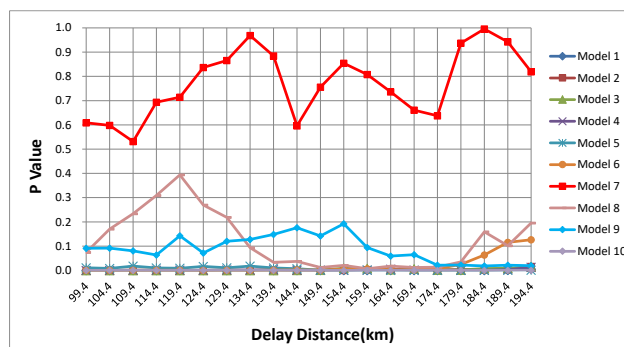


Figure 2 P Values at Ten Delay Distances

As shown in the line chart, obviously, the P value of the seventh model is greater than 0.5, thus the Moran's I of Model 7 do not reject the null hypothesis. In the next place, the P values belonging to Model 8 and Model 9 are greater than 0.1 at many delay distances, thus the two models' Global Moran's indexes do not reject the null hypothesis too. It means these residuals of Model 7, Model 8 and Model 9 are random spatial distribution. For the rest of the seven models, their P values are far less than 0.1. Therefore, the Global Moran's Indexes of the remainder seven models should refuse the null hypothesis, and it means that residuals of the seven models is either statistical agglomeration or variation.

4.2 Local Spatial Autocorrelation

In this process of Local Spatial Autocorrelation analysis, Cluster and Outlier Analysis tool was used to calculate the Local Moran's I of the ten models' residuals. And then different symbols are represented different local spatial correlations.

There are four distinct symbols in any one picture (Fig. 3). The four symbols identify four types of local spatial association between a point and its neighbors: (HH) a high residual error with high residual error neighbors; (LH) a low residual error surrounded by high residual error neighbors; (LL) a low residual error surrounded by low residual error neighbors; (HL) a high residual error with low residual error neighbors (Li et al. 2009). All the points of distinct symbols represent the regions of possessing the local spatial agglomeration or variation. While, there are many inconspicuous points that represent the regions without local spatial correlation in every figure.

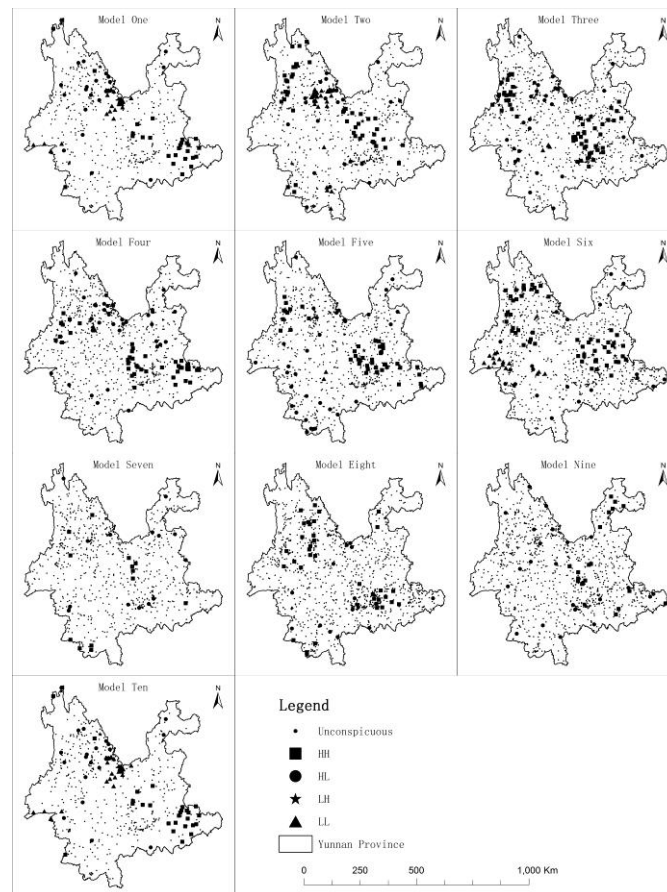


Figure 3 Local Spatial Autocorrelation of Ten Models

In order to compare intuitively, the information of the ten figures was compiled into Table 3.

Table 3 Statistical Table of Correlation Points in the Ten Models

Model Name	Number of Correlation Points	Total of Points	Percentage
Model 1	82	559	14.67%
Model 2	114	728	15.66%
Model 3	100	928	10.78%
Model 4	85	862	9.86%
Model 5	79	869	9.09%
Model 6	97	947	10.24%
Model 7	27	787	3.43%
Model 8	51	1121	4.55%
Model 9	42	1055	3.98%
Model 10	202	1071	18.86%

On the basis of the above table, the amount of local spatial correlation points is the least in the Model 7. And the local spatial correlation points of Model 7 also account for the minimum percentage. In the second place, the amount and the percentages of local spatial correlation points are also very small in the Model 9 and Model 8. In other words, compared with other models, the three models have more points of the nature of randomness.

5 Discussion and conclusions

In the Global Spatial Autocorrelation analysis, the Model 7 is optimal model. The residuals of the Model 7 not only have lower Global Moran's I compared with other models in each research scope, but also cannot refuse to null hypothesis. In the second place, the Global Moran's Indexes of residuals of Model 8 and Model 9 are smaller than 0.01 in twenty delay distances. And their P values are too big to accept null hypothesis. Therefore, the residuals of Model 7, Model 8 and Model 9 are random spatial distribution. Because the fire ignition probability models are statistical model, the more random spatial distribution residual errors of model are, the better performance model has. Thus, the performances of Model 7, Model 8 and Model 9 are relatively good.

Through the analysis of the Local Spatial Autocorrelation, the amount of local spatial correlation points is the minimum in Model 7 and devotes a smallest percentage compared with other models. That is to say, the regions of local spatial randomness account for the biggest percentage. Therefore, the Model 7 is optimal model. In accordance with the results of the Global Spatial Autocorrelation analysis, Model 8 and Model 9 are great models that are slightly worse than Model 7.

Apart from that, the ten fire ignition probability models are reviewed. How was each model generated? "The former and the latter five models were generated by binary logistic regression and geography weight logistic regression method, respectively" shows that Model 7, Model 8 and Model 9 were generated by geography weight logistic regression method. We can say that, in this study at least, the models generated by using geography weight logistic regression method are better than ones by using binary logistic regression method.

Acknowledgements

This research work was supported by the National Natural Science Foundation of China [Grant No. 41671417]

References

- [1] Di X Y, Wang H L, Yao S R. 1993. Forest fire forecast. Harbin: Northeast Forestry University Publishing Company.
- [2] Hu H Q. 2005. Forest fire ecology and management. Beijing: China Forest Press.
- [3] Shu L F, Xiang A M, Tian X R. 1998. 3S in forest fire management. *Forest Science* 24(1):46-51.
- [4] Legendre P, Legendre L. 1998. 'Numerical Ecology.' 2nd English edn. Amsterdam: Elsevier Science BV.
- [5] Keane R E, Drury S A, Karau E C, Hessburg P F, Reynolds K M. 2010. A method for mapping fire hazard and risk across multiple scales and its application in fire management. *Ecological Modelling* 221(1):2-18.
- [6] Li Fei, Zhou Chenghu. 2009. Spatial Autocorrelation Analysis on Regional Economic Disparity of Northeast Economic Region in China. *Chinese Journal of Population, Resources and Environment* 7(2):27-31.
- [7] Environmental Systems Research Institute. Information on <http://resources.arcgis.com/en/help/main/10.1/index.html>