International Journal for Housing Science and Its Applications
Publish August 6, 2025. Volume 46, Issue 3 Pages 5633-5642

https://doi.org/10.70517/ijhsa463486

Study on the distribution pattern and spatial prediction of lightning activity based on geographically weighted regression modeling

Amuersana[1], Shi Jin[1,*], Lu Chao[1], Xuan Li[1] and Danping Wang[1]
[1] Meteorological Disaster Prevention Center, Hohhot Meteorological Bureau, Hohhot, Inner Mongolia, 010010, China
Corresponding authors: (e-mail: messidiegoks@126.com).

Abstract Lightning, as a natural discharge phenomenon, is often accompanied by strong convective weather, which can lead to human and animal casualties, facility damage and economic losses. With the socio-economic development, there is an increasing demand for accurate prediction of lightning distribution. In this study, an empirical orthogonal function (EOF) analysis combined with a geographically weighted regression model is used to analyze the distribution pattern of lightning activity in Fujian Province and to make spatial predictions. The study utilizes the actual monitoring data of Fujian Power Grid LLS for the past 40 years (1982-2021), decomposes the ground flash records into mutually orthogonal spatial eigenvectors and time coefficients based on the EOF method, and analyzes the spatial heterogeneity of the lightning activity using a geographically weighted regression model. The results show that the cumulative variance contribution of the first three modes of the EOF decomposition of thunderstorm days in Fujian Province reaches 75.45%, of which the first mode contributes 57.43%, which is mainly characterized by the negative phase distribution in the province; the overshooting lag correlation coefficient between thunderstorm days and ENSO index is 75.07%, of which the correlation coefficient with the El Niño event is as high as 83.53%, which is significantly higher than the correlation coefficient with the La Niña event (35.79%); the correlation coefficient with the ENSO index is 75.07%. The correlation coefficient with El Niño event is 83.53%, which is significantly higher than that with La Niña event (35.79%); 94.67% of the lightning intensities in Fujian are less than 50 kA, 63.46% of the lightning intensities are less than 15 kA, and the interval with the highest frequency of the lightning current amplitude is 7-17 kA. The study concludes that the lightning activity in Fujian Province is mainly affected by the topographic and climatic factors, and the mountainous and hilly areas are the main hotspot of the lightning activity and there is a significant spatial heterogeneity. The geographically weighted regression model can effectively predict the spatial distribution of lightning in complex terrain.

Index Terms Lightning activity, Geographically weighted regression model, Spatial heterogeneity, Empirical orthogonal function, Lightning current magnitude, Spatial prediction

I. Introduction

Lightning is a majestic and intimidating natural phenomenon accompanied by flashes of lightning and thunder, which is active at any given time on Earth. Due to the high temperature in the lower part of the cumulonimbus cloud and the low temperature in the middle and upper part, powerful updrafts and downdrafts are generated [1]. A large number of water vapor condensates such as ice crystals, small and large water droplets, supercooled water droplets, shrapnel and hail within the cloud, through many complex processes such as touch freezing, collision, fragmentation, and melting, electrify the cloud and separate the positive and negative charges, forming positive and negative charge centers in the cloud. When the aggregated charge is large enough, a breakdown discharge occurs between the anisotropic charge centers to produce a spark discharge phenomenon, i.e., the lightning phenomenon [2]. During the discharge, a sudden increase in temperature causes the volume of air to expand dramatically, resulting in a shock wave that causes a strong thunder [3]. Therefore, lightning is a discharge phenomenon in the atmosphere. Lightning is usually generated in cumulonimbus clouds where convection develops vigorously, and is therefore often accompanied by strong gusts of wind and torrential rain, and sometimes by strong convective weather such as hail and tornadoes [4], [5].

In recent years, lightning strikes are frequent, which can not only cause human and animal casualties, forest and building fires, explosions in oil depots and chemical plants, as well as major disasters such as disruption of electricity and communications. And it also seriously interferes with the normal work of electrical equipment, and the instantaneous high potential and transient electromagnetic radiation can lead to the damage of microelectronic equipment. At the same time, it is very easy to cause fire and explosion accidents in flammable and explosive places due to electric sparks, resulting in significant economic losses [6]-[10]. With the rapid development of the social economy and improving the living standards of the people, the impact of lightning disasters on the whole society is becoming more and more serious, so the prediction of lightning activity is of great significance.

And due to the transient and random nature of lightning occurrence, the scientific understanding of lightning climate characteristics relies heavily on the development of lightning detection technology [11]. Among them, the traditional linear regression model has poor performance in explaining the inhomogeneity in the spatial distribution of lightning activity in complex terrain, and its complexity [12]. Geographically weighted regression models, on the other hand, fully consider the spatial autocorrelation and spatial heterogeneity of spatial data, thus improving the accuracy and interpretability of regression results [13].

Lightning is a majestic and intimidating natural phenomenon accompanied by flashes of lightning and thunder, which is active at any given time on earth. Due to the high temperature in the lower part of the cumulonimbus cloud and the low temperature in the middle and upper part, powerful updrafts and downdrafts are generated. A large number of ice crystals, small and large water droplets, supercooled water droplets, shrapnel and hail and other water vapor condensates within the cloud are electrified by touching freezing, collision, crushing and melting and many other complex processes, and the positive and negative charges are separated to form positive and negative charge centers in the cloud. When the gathered electricity is large enough, the opposite charge center will occur between the breakdown discharge and produce spark discharge phenomenon, that is, lightning phenomenon. During the discharge, the sudden increase in temperature causes the volume of air to expand dramatically, thus generating shock waves and leading to strong thunder. Therefore, lightning is a discharge phenomenon in the atmosphere. Lightning is usually generated in cumulonimbus clouds with strong convection development, so it is often accompanied by strong gusts of wind and rainstorms, and sometimes accompanied by strong convective weather such as hail and tornadoes.

In recent years, lightning accidents occur frequently, not only can cause human and animal casualties, causing forests, building fires, oil depots, chemical plants, explosions, as well as power and communication disruptions and other major disasters. And also seriously interfere with the normal work of electrical equipment, and instantaneous high potential, transient electromagnetic radiation can lead to microelectronic equipment damage. At the same time, it is very easy to cause flammable and explosive places due to electrical sparks caused by fire and explosion accidents, resulting in significant economic losses. With the rapid development of socio-economic and improving the living standard of the people, the impact of lightning disasters on the whole society is becoming more and more serious, so the prediction of lightning activity is of great significance.

Due to the transient and random nature of lightning occurrence, the scientific understanding of lightning climate characteristics is largely dependent on the development of lightning detection technology. Among them, the traditional linear regression model has poor performance in explaining the inhomogeneity of the spatial distribution of lightning activity in complex terrain, and its complexity. The geographically weighted regression model, on the other hand, fully takes into account the spatial autocorrelation and spatial heterogeneity of spatial data, thus improving the accuracy and interpretability of the regression results.

This study combines the empirical orthogonal function analysis and geographically weighted regression model, aiming to comprehensively analyze the spatial distribution characteristics of lightning activity and its influencing factors in Fujian Province. The study firstly utilizes the empirical orthogonal function method to decompose the lightning activity data of Fujian Province in the past 40 years, extracts the main spatial modes and temporal coefficients, and identifies the main distribution patterns of lightning activity; secondly, analyzes the correlation between the lightning activity and climatic factors, especially the ENSO events, and explores the differences in the impacts of El Niño and La Niña on the lightning activity of Fujian Province; and then, statistically analyzes the lightning current amplitude and distribution characteristics of the lightning current in Fujian Province, and reveals the accuracy and interpretation of the results. Then, the distribution characteristics are statistically analyzed to reveal the distribution law of lightning intensity; finally, the spatial prediction model of lightning activity is constructed based on the geographically weighted regression model, and the prediction accuracy of the model is improved by the selection of spatial weighting function and the optimization of bandwidth. The results of this study will provide a scientific basis for lightning protection in Fujian Province, and at the same time provide methodological references for the study of lightning activities in other regions.

II. Geographically weighted regression models based on the distribution of lightning activity

II. A. Information and processing methods

In this paper, the actual monitoring information (ground flash occurrence time, latitude and longitude records, etc.) of Fujian Power Grid LLS in the last 40 years is used, and the historical monitoring data from 1982 to 2021 are extracted in the original monitoring dataset with reference to the seasonal characteristics of thunderstorm days in Fujian Province in the past, and the three-dimensional variables of the ground flash records are decomposed into the spatial eigenvectors that are orthogonal to each other and time coefficients that are relative to each other, based on empirical orthogonal functions (EOFs) and thus the typical spatial distribution structure and its evolutionary characteristics are illustrated.

II. A. 1) Data pre-processing

LLS monitoring data are spatially discrete and random. When the grid method lightning day statistics method is used for the analysis of lightning activity in Fujian Province, the LLS cooked data are firstly processed into the day-by-day equal area grid of the ground flash density, and two sets of spatially resolved three-dimensional matrices of 0.1°×0.1° and 0.5°×0.5° are obtained. In view of the seasonal characteristics of lightning activity, the summer and half-year monitoring information of the recent 3a of Fujian Power Grid Lightning Positioning System is selected. Some statistical results of ground flashes in Fujian Province from 1982 to 2021 are displayed as shown in Table 1. The use of 0.1°×0.1° and higher resolution spatial data will make the spatial dispersion of lightning activity increase, and more local monitoring information makes the spatial characteristics of ground flashes tend to be independent, which negatively affects the analysis of spatial characteristics of the change rule of the time series. Therefore, the 0.5°×0.5° day-by-day gridded data of ground flashes over the whole Fujian area during the above time period are selected for analysis and research.

Table 1: Semiflash statistics of Fujian province in 1982~2021

Year	Grid resolution	Sample size/day	Maximum/number	Minimum/times
1982	0.5°×0.5°	185	6589	0
1983	0.5°×0.5°	185	17693	0
…	…	…	…	…
2021	0.5°×0.5°	185	12858	0

II. A. 2) Statistical methods

In order to statistically characterize the spatial features of lightning activity and its temporal change law in Fujian Province in the recent 3a, this paper proposes to use the EOF method [14], [15] to structure the three-dimensional matrix of lightning activity constituted by the gridded data of ground flashes, so as to abstract the main features of its three-dimensional changes.The EOF method takes the time-varying features of the information of a certain two-dimensional spatial meteorological element as an object of analysis, and extracts in multidimensional variables the spatial eigenvectors (or spatial modes) of the meteorological elements are extracted from the multidimensional variables, and the temporal changes corresponding to the spatial eigenvectors are generally referred to as the time coefficients, and thus the EOF analysis is also referred to as the temporal decomposition.

The matrix composition of EOF analysis is as follows:

$$X=\begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1j} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2j} & \cdots & x_{2n} \\ \vdots & \vdots & & \vdots & & \vdots \\ x_{i1} & x_{i2} & & x_{ij} & & x_{in} \\ \vdots & \vdots & & \vdots & & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mj} & \cdots & x_{mm} \end{bmatrix} \tag{1}$$

The matrix X in Eq. (1) represents the number of distance flats of the daily ground flash density in the last 3a summer and half-years, and the matrix dimension is 255×549, which represents the number of grid points of the adopted gridded data and the number of statistical days of the ground flash records, respectively.

The EOF expansion of Eq. (1) transforms the original matrix into the sum of the products of the two parts of the spatial eigenvectors and their corresponding time coefficients, i.e:

$$X = VT \tag{2}$$

In Eq. (2), V and T are the spatial eigenvectors and their time coefficients. When the cumulative variance contribution of the first few spatial eigenvectors is high and passes the significance test, it means that the several eigenvectors have maximally covered the information of the original variable matrix, thus realizing the dimensionality reduction of the multidimensional complex matrix. In general, the decomposition information given by the EOF method is more intuitive than that of the original matrix of variables, and can give targeted physical meaning according to different research categories.

*II. B. **Lightning activity distribution prediction modeling***

II. B. 1) Geographically weighted regression models

In the traditional spatial analysis of geography, for the problem of analyzing the degree of influence of changes in one or more variables on another variable, it is often assumed that the variables are not related to their geographic location, and the relationship between the variables is obtained by constructing the global regression model (OLR) [16] such as the least squares model to determine the regression coefficients of the variables, the obtained regression equations are shown in Equation (3) where β is the coefficient of each variable, ε_i is the coefficient constant, at this time the coefficients of the variables obtained are the average explanatory coefficients of each explanatory variable, which is fixed. However, in the actual study of the problem, the regression coefficients of each parameter often change because of the location of the variable is different, that is, the regression coefficients change with the geographic location of the variable. Namely:

$$Y_i = \beta X_i + \varepsilon_i \tag{3}$$

Due to the existence of different urban geospatial variability such as natural environment and humanistic environment, which leads to the spatial non-stationarity of urban transportation and travel, POI facilities, and other built-up geography, that is, the relationship and structure of these variables are always changing with the geographic location, at this time, if ignoring its spatial non-stationarity using global regression analysis may lead to inconsistent parameters or inaccurate test results. The geographically weighted regression model (GWR) is a better alternative model to capture spatial heterogeneity and can overcome this defect, and this method can effectively reveal the spatial change pattern of impact coefficients in the study area.

(1) Definition of geographically weighted regression

Geographic weighted regression is the process of using the distance decay function to obtain weighted values by bringing in the geographic location of its neighboring regions, and then locally applying the model to these weighted regions. Its mathematical expression is shown below:

$$Y_i = \beta_0\left(u_i, v_i\right) + \sum_{j=1}^{p} \beta_j\left(u_i, v_i\right) X_{ij} + \varepsilon_i \tag{4}$$

where Y_i is the explanatory variable, i.e. the dependent variable, X_{ij} is the regression fit coefficient of the j th explanatory variable at sample point i, $\beta_j\left(u_i, v_i\right)\left(j = 1, 2, 3, 4, \cdots, p\right)$ is the spatial geographic location function, $\left(u_i, v_i\right)$ represents the latitude and longitude coordinates of the sample points in a certain region, geographically weighted regression actually characterizes the results of local regression of each different location, so that the coefficients of the different location fits are dynamically changed in the study area, which allows for a better explanation of the variables.

(2) Parameter estimation of geographic weighting

Since the closer the correlation of things is greater, the sample point $\left(u_0, v_0\right)$ for determining the location is generally solved by the locally weighted least squares method. Where $\omega_i\left(u_0, v_0\right)$ is the geographic weight of the explanatory variable i at the sample point $\left(u_0, v_0\right)$, such that $\beta\left(u_0, v_0\right)$ is the geographic weight of the variable i at the sample point $\beta\left(u_0, v_0\right) = \left(\beta_0\left(u_0, v_0\right), \beta_1\left(u_0, v_0\right), \cdots, \beta_p\left(u_0, v_0\right)\right)^T$, then $\beta\left(u_0, v_0\right)$ at $\left(u_0, v_0\right)$ is localized as follows The least squares estimate is given in Eq. where $X = \left(X_0, X_1, \cdots, X_p\right)$, $X_j = \left(x_{1j}, x_{2j}, \cdots, x_{nj}\right)^T$, and $Y = \left(Y_1, Y_2, \cdots, Y_n\right)^T$, $W(u_0, v_0) = Diag\left(\omega_1(u_0, v_0), \omega_2(u_0, v_0), \cdots, \omega_n(u_0, v_0)\right)$. I.e.:

$$\min \sum_{i=1}^{n} \left[y_i - \sum_{j=1}^{p} \beta_j\left(u_0, v_0\right) x_{ij} \right]^2 \omega_i\left(u_0, v_0\right) \tag{5}$$

$$\hat{\beta}_j\left(u_0, v_0\right) = \left(X^T W\left(u_0, v_0\right) X\right)^{-1} X^T W\left(u_0, v_0\right) X \tag{6}$$

(3) Selection of spatial weighting function

Geographically weighted regression model is to use the weighted value of different sample points in space ω_{ij} to determine the regression coefficient, so the most central element is to determine the spatial weighting matrix, different spatial weighting function characterizes the measure of the distance between the sample points and other data points and quantify the distance attenuation and the relationship between the weighted value of the different methods, when $\omega_{ij}=1$, the weighting value is fixed, and this is an ordinary global regression model. Since the choice of spatial weighting function directly affects the weighting matrix and thus the model effect, determining the spatial weighting function is an important part of constructing a geographically weighted regression model.

The expression of the Gaussian function is shown in equation (7), where X is the regression point, b is the bandwidth, which is the range of each sample point searching the surrounding area in the local weighting, and d_{ij} is used to characterize the distance between the sample point's location (u_i, v_i) and the location of other sample points (u_j, v_j) . I.e:

$$\omega_{ij} = \exp\left(-\frac{1}{2}\left(\frac{d_{ij}}{b}\right)^2\right) \tag{7}$$

In practice, it is often the case that data points that have little or no effect on the regression parameter estimates are truncated and do not participate in the calculation process, i.e., only the weights of sample points inside the bandwidth are calculated. At this point, it is equivalent to replace the Gaussian function with a finite Gaussian function, which is a spatial weight function for the Bi-square function, calculated with a localized Gaussian function when the data points are inside the bandwidth b, and the data points outside the bandwidth are considered to have a weight of 0. That is:

$$\omega_{ij} = \begin{cases} \left[1-\left(\dfrac{d_{ij}}{b}\right)^2\right]^2 & d_{ij} < b \\ 0 & d_{ij} > b \end{cases} \tag{8}$$

(4) Selection of optimal bandwidth

The bandwidth b refers to the non-negative decay parameter that characterizes the functional relationship between the weights and the distance matrix, with the increase of the bandwidth, the weight ω_{ij} decays more slowly with the increase of the distance, and at the same time, the larger the local area of the computed weights is, and the more data points are covered. In the debugging of the actual parameters of geographically weighted regression, the sensitivity of the spatial weight function is lower than that of the bandwidth setting, and the over-setting of the bandwidth will lead to too large a deviation of the regression parameter, while the bandwidth is too small, which is prone to cause too large a variance of the regression parameter. The main criteria commonly used for bandwidth determination are the cross-validation method (CV) and the minimum information criterion (AICc) [17].

The cross-validation method involves dividing the data into multiple sets and determining the final bandwidth selection after multiple sets of calculations with actual observers. The different bandwidth settings are plotted against the obtained CV as a trend curve, and when the CV is smallest, i.e., the deviation between groups is minimized, the bandwidth b_0 at this point is chosen. The formula for this is shown below:

$$CV = \frac{1}{n}\sum_{i=1}^{n}\left[y_i - \hat{y}_{\neq i}(b)\right]^2 \tag{9}$$

$$b_0 = \arg\min_{b>0} CV(h) \tag{10}$$

The minimum information criterion means that when there is a group of model parameters to choose from, the parameter that minimizes the AIC is preferred. This is due to the fact that the size of the AIC depends on the number of independent parameters that make up the model as well as the number of great likelihoods of the model, whereas the smaller the independent parameters are, the more concise the model is, and the smaller the AIC is; and the larger the great likelihoods are, the more accurate the model is, and the smaller the AIC is. it can be seen that when the AIC is smaller, the state characterized by the model is better and better. It can be seen that when the AIC is smaller, the state characterized by the model is better, at this time, only the bandwidth that makes the AIC smallest can be selected. The mathematical expression is shown in the following equation:

$$AIC = 2n\ln(\hat{\sigma}) + n\ln(2\pi) + n\left[\frac{n+tr(S)}{n-2-tr(S)}\right] \tag{11}$$

II. B. 2) Impact factor screening and parameterization

(1) Determination of influencing factors

The geographically weighted regression model is a local weighting treatment based on the global least squares model to obtain the local spatial role of each explanatory variable. Therefore, the explanatory variables with significant correlation with the dependent variable in the global model are still significant in the geographically weighted regression model, so this paper takes the variables that are significant in the global regression model as the explanatory variables in the geographically weighted regression model, to further explore the changes in the spatial degree of influence of each variable.

(2) Parameterization

The selection of spatial weight function and the setting of bandwidth were mentioned in the previous section as the two main factors affecting the effect of geographically weighted regression model. In existing studies, the form center of each traffic district is often used as the regression point of the geographically weighted regression model. Since the research area is divided into the same grid area of 500×500m in the research process of this paper, and the form centers of each grid area are spatially balanced, a fixed bandwidth is adopted in the process of setting the bandwidth, and the spatial weighting function adopts the Gaussian kernel function to construct the model, and at the same time adopts the minimum information criterion (AICc).

III. Lightning activity distribution patterns and spatial prediction results and analysis

III. A. Characterization of the spatial and temporal evolution of the EOF decomposition on thunderstorm days

The orthogonal function method based on the optimal matching monitoring radius is called EOF.In order to further study the spatio-temporal evolution characteristics of thunderstorm days, the first three eigenvectors of the average thunderstorm days calculated for 66 stations in Fujian Province in the past 40 years are decomposed by EOF. The variance contributions of the first three modes are 57.43%, 9.15%, and 8.87%, respectively, and the cumulative variance contribution of the three modes is 75.45%, which can characterize the main spatial and temporal distribution of thunderstorm days. Figures 1 to 3 show the time coefficients and 9-point smoothed curves of the first three modes of EOF for thunderstorm days in Fujian Province from 1982 to 2021, respectively.

The spatial pattern of the first mode of EOF decomposition characterizes the negative phase of thunderstorm days in Fujian Province in the past 40 years, indicating that the spatial distribution of different stations in Fujian Province with the first mode time coefficients is opposite, and the maximum value of the negative phase is concentrated in the west-central and southern parts of Fujian Province, which indicates that the main interannual variation of thunderstorm days in Fujian Province in the past 40 years is characterized by the changes in the west-central and southern parts of the coastal coast. The corresponding time coefficients of the first mode and the 9-point smoothed curves show that the west-central and southern parts of Fujian Province have more lightning activities before 1998 and after 2016, and less lightning activities during 1998-2016.

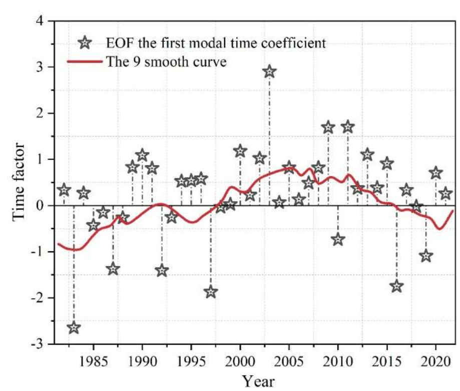


Figure 1: EOF the first modal time coefficient and the 9 smooth curve

The spatial pattern of the EOF decomposition of the second mode indicates a significant opposite phase between the northeast and southwest of Fujian Province. Combining the corresponding time coefficients and the 9-point smoothed curve, it can be seen that the northeast of Fujian Province, particularly the Wuyi Mountain-Pucheng-Songxi area, experienced significantly more lightning activity before 1994 and after 2012, while the southwest had less lightning activity. In contrast, the northeast had a lower number of thunderstorm days than the historical average between 1994 and 2012, while the southwest was above the historical level, with the opposite changes occurring in other years.

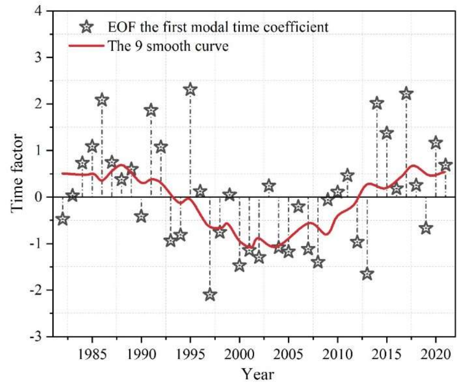


Figure 2: The EOF second modal time coefficient and the 9 smooth curve

The EOF decomposition of the third mode's spatial representation shows a significant antiphase distribution between the inland and coastal areas of Fujian Province. The corresponding time coefficients and the 9-point smoothing curve reflect that before 2012, the inland areas of Fujian experienced relatively active thunderstorms, while starting in 2012, the thunderstorm activity significantly decreased. On the other hand, the southeastern coastal areas had fewer thunderstorm days before 2012, but after 2012, the lightning activity increased relatively.

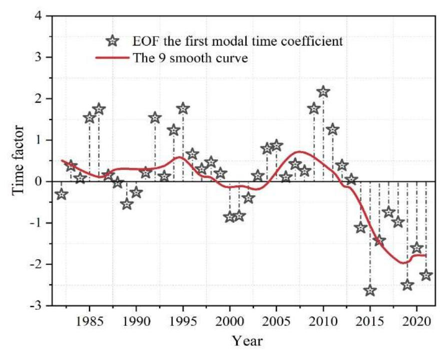


Figure 3: The EOF third modal time coefficient and the 9 smooth curve

Through analysis of the climate thunderstorm days in Fujian Province over the past 40 years, the spatial patterns of EOF decomposition mainly include three types: "province-wide opposite phase," "northeast-southwest region opposite phase," and "inland-coastal region opposite phase." Among these, the first mode has the largest variance contribution, exceeding 50%, and can be considered the primary mode of EOF decomposition.

III. B. Analysis of climate impact factors of thunderstorms in Fujian Province

In order to preliminarily investigate the main influencing factors affecting the thunderstorm days in Fujian Province in the past 40 years and whether the climate activity pattern is affected by ENSO events is worth studying. The overshooting and lagging correlation coefficients of thunderstorm days and ENSO indices in Fujian Province are calculated, and the fitting trends of the two are shown, in which the study periods of thunderstorm days and ENSO are uniformly selected from 1982-2021.

The results of the influence of thunderstorm day activity pattern in Fujian Province from 1982 to 2021 are shown in Fig. 4. It can be seen that the correlation coefficient between thunderstorm days and ENSO index is 75.07%, and the scatter distribution fits the trend better, the stations are basically distributed around the fitted line, of which 15 stations are located above the fitted trend line, and the remaining 16 stations are located below the fitted trend line, which can be preliminarily obtained that there is a certain degree of correlation between the activity of thunderstorms in Fujian Province and the SST in the western Pacific Ocean.

The correlation coefficients of El Niño and La Niña on the thunderstorm days of the corresponding years are calculated, and it is found that the correlation coefficient of El Niño and thunderstorm days is as high as 83.53%, which is more than that of the total ENSO events on the thunderstorms, while the correlation coefficient of La Niña and thunderstorms is only 35.79%. The correlation coefficient between La Niña and thunderstorm days is only 35.79%, which makes El Niño event one of the main factors affecting thunderstorm days in Fujian Province.

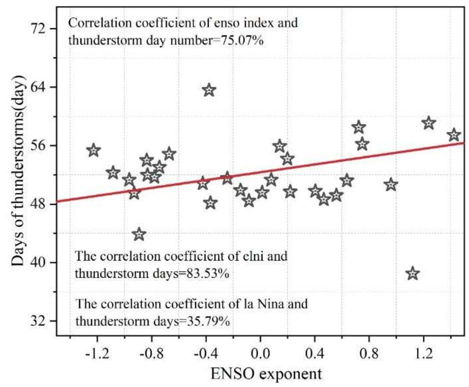


Figure 4: Effect of the daily activity of the thunderstorm day in 1982-2021

The temporal evolution trend of thunderstorm days in Fujian Province from 1982 to 2021 is shown in Fig. 5. The temporal trend of thunderstorm days in Fujian Province during the ENSO events in the last 40 years shows that the positive spatial level above the zero-degree line, which indicates the SST, is an El Niño year, and below the zero-degree line is a La Niña year. There are 19 El Niño years in the last 40 years, and the rest of the years are La Niña years. It is worth noting that the number of lightning days corresponding to El Niño years also have extreme points, and the number of lightning days corresponding to La Niña years are mostly very small points of lightning days.

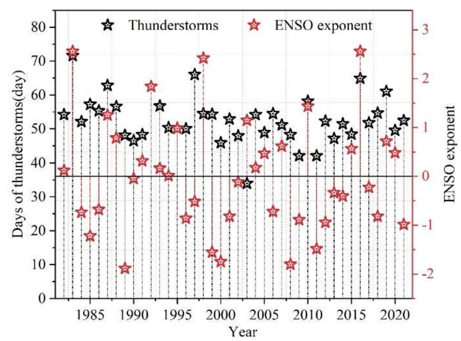


Figure 5: The trend of the time of the thunderstorm in 1982-2021

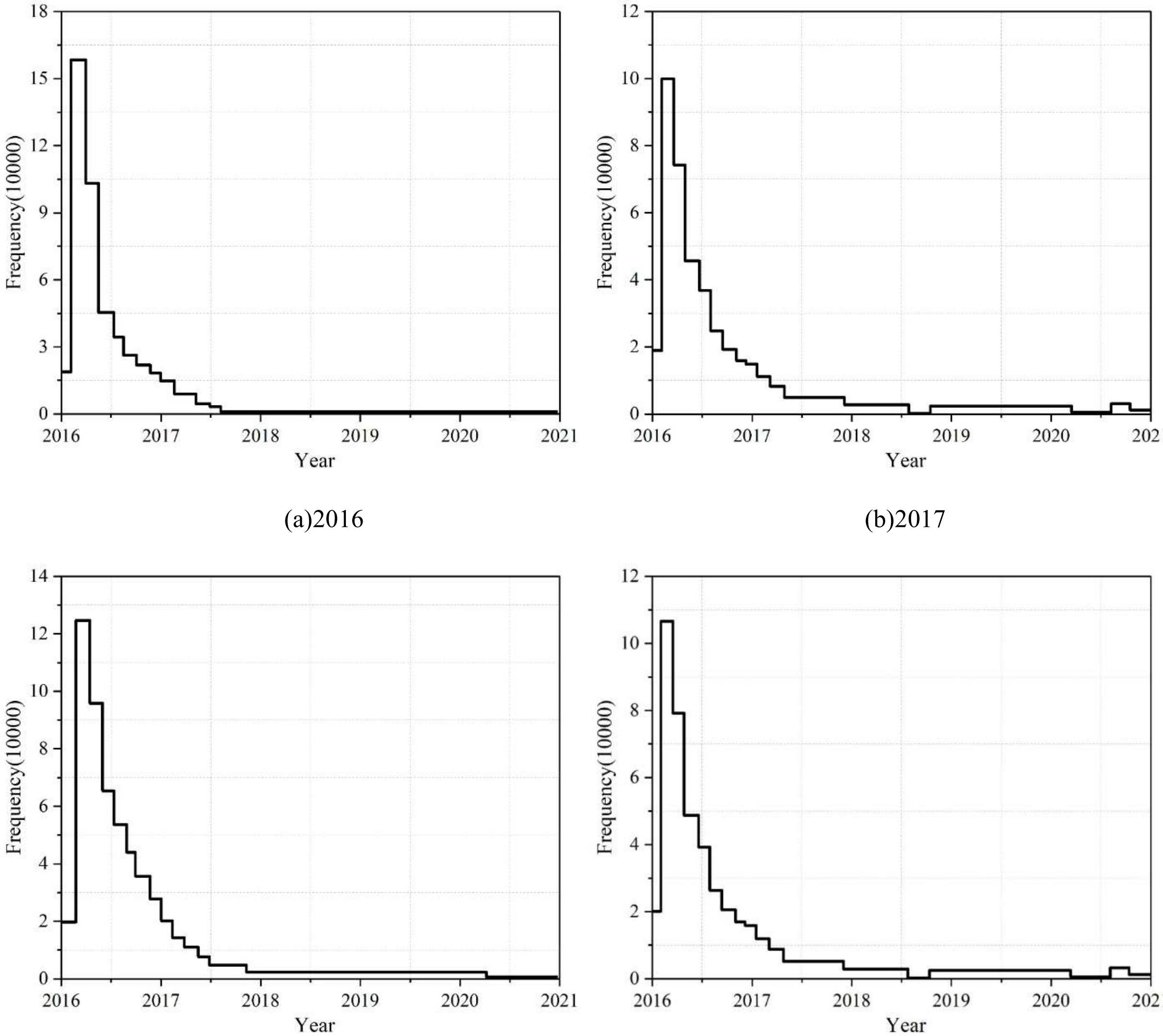
III. C. Lightning current amplitude distribution characteristics

Lightning current magnitude refers to the maximum instantaneous value of the current flowing through the struck object when a lightning strike occurs, and this value is usually expressed in kiloamperes (kA).The frequency distribution of lightning current magnitude in Fujian region from 2016 to 2019 is shown in Fig. 6, in which (a)~(d) represent 2016, 2017, 2018 and 2019, respectively. It can be seen that the highest frequency of lightning current amplitude in Fujian area ranges from 7 to 17kA, of which 7~12kA accounts for the highest proportion, with an annual average frequency of about 110,000 times, accounting for 27.5% of the total frequency of the year. The second highest proportion is in the range of 12-17kA.

III. D. Results of the prediction of the spatial distribution characteristics of lightning activity

Lightning density refers to the number of lightning occurrences per unit area, unit: times/km². The results of cumulative probability distribution of lightning current amplitude in Fujian area from 2016 to 2019 are shown in Figure 7. From the cumulative probability distribution of lightning current amplitude in Fujian area from 2016 to 2019, 94.67% of the lightning intensity is less than 50kA, 91.83% of the lightning intensity is less than 40kA, 81.36% of the lightning intensity is less than 30kA, and 63.46% of the lightning intensity is less than 15kA.

The spatial distribution characteristics of lightning intensity in Fujian are mainly affected by topography and climate, showing obvious differences. Lightning activities are mainly concentrated in the mountainous and hilly areas in the transition from the basin to the plateau, followed by the plains of the basin, while lightning is rare in the plateau area. This distribution characteristic is closely related to the topography, because the uplift of the terrain in mountainous and hilly areas is likely to trigger convective weather, thus promoting the formation and development of thunderstorm clouds. As a result, mountainous and hilly areas have become the main hot spots for lightning activity.



(a)2016 (b)2017

(c)2018 (d)2019

Figure 6: Frequency distribution of lightning flow amplitude in fujian region

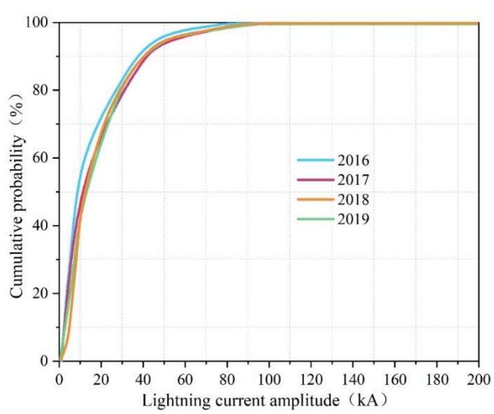


Figure 7: The cumulative probability distribution of the lightning flow amplitude

IV. Conclusion

Based on the empirical orthogonal function (EOF) analysis and geographically weighted regression model, the spatial distribution characteristics of lightning activity in Fujian Province from 1982 to 2021 were systematically studied. The results show that the cumulative variance contribution of the first three modes of the EOF decomposition of thunderstorm days in Fujian Province is 75.45%, which characterizes the main spatial and temporal distribution, among which the contribution of the first mode is 57.43%, which is dominated by the negative phase distribution in the province. There is a significant correlation between thunderstorm days and ENSO index in Fujian Province, with a correlation coefficient of 75.07%, especially with El Niño events, which is as high as 83.53%, indicating that SST anomalies in the western Pacific Ocean are an important factor influencing the lightning activities in Fujian Province. The frequency distribution of lightning current amplitude in Fujian is concentrated in the range of 7-17kA, with 7-12kA accounting for the highest percentage of 27.5%. The cumulative probability distribution of lightning intensity shows that 81.36% of the lightning intensity is less than 30kA, which provides an important parameter for lightning protection design. The results of the geographically weighted regression model analysis showed that the lightning activity in Fujian Province showed obvious spatial heterogeneity, mainly concentrated in the mountainous and hilly areas, followed by the basin-plain area, while the plateau area had less lightning activity. This distribution is closely related to the topographic and climatic conditions, and the topographic uplift favors the development of convective weather, thus promoting the formation of thunderstorm clouds. This study confirms that the geographically weighted regression model can effectively capture the spatial heterogeneity of lightning activity and provide an accurate spatial prediction basis for lightning protection.

Funding

This research was supported by the Inner Mongolia Autonomous Region Natural Science Foundation, Research on Cloud-to-Ground Lightning Characteristics Based on Spatiotemporal Clustering (2022MS04003); Hohhot Meteorological Bureau Innovation Project, Research on the Lightning Monitoring and Early Warning System in Hohhot City (hsqxkjzx202404).

References

[1] Pustovalov, K. N., & Nagorskiy, P. M. (2018). Response in the surface atmospheric electric field to the passage of isolated air mass cumulonimbus clouds. Journal of Atmospheric and Solar-Terrestrial Physics, 172, 33-39.
[2] Goeritno, A., & Azama, I. M. (2024). Explaining the existence of lightning: Literature review and an example of calculating the efficiency value of a detection system. Journal of Applied Science and Advanced Engineering, 2(1), 27-36.
[3] Karch, C., Schreiner, M., Honke, R., & Wolfrum, J. (2018, September). Shock waves from a lightning discharge. In 2018 34th International Conference on Lightning Protection (ICLP) (pp. 1-6). IEEE.
[4] Wilhelm, J., Mohr, S., Punge, H. J., Mühr, B., Schmidberger, M., Daniell, J. E., ... & Kunz, M. (2021). Severe thunderstorms with large hail across Germany in June 2019. Weather, 76(7), 228-237.
[5] Kastman, J. S., Market, P. S., Fox, N. I., Foscato, A. L., & Lupo, A. R. (2017). Lightning and rainfall characteristics in elevated vs. surface based convection in the midwest that produce heavy rainfall. Atmosphere, 8(2), 36.
[6] Yair, Y. (2018). Lightning hazards to human societies in a changing climate. Environmental research letters, 13(12), 123002.
[7] Gomes, C. (2017). Lightning related human risks. Environ Risk Assess Remediat. 2017; 1 (3): 1, 15.
[8] Liu, K., Wu, D., Xia, J., Chen, W., Chen, G., Liu, J., & Li, H. (2021, September). Risk Evaluation Indicator System for Lightning Disaster in Oil Storage Depots. In IOP Conference Series: Earth and Environmental Science (Vol. 859, No. 1, p. 012114). IOP Publishing.
[9] Souto, L., Taylor, P. C., & Wilkinson, J. (2023). Probabilistic impact assessment of lightning strikes on power systems incorporating lightning protection design and asset condition. International Journal of Electrical Power & Energy Systems, 148, 108974.
[10] Song, Y., Xu, C., Li, X., & Oppong, F. (2024). Lightning-induced wildfires: An overview. Fire, 7(3), 79.
[11] Lai, J., Liu, Y., Du, J., & Li, Q. (2019, December). Lightning Detection Technology and Application. In 2019 International Conference on Meteorology Observations (ICMO) (pp. 1-5). IEEE.
[12] Li, C., & Müller, U. K. (2021). Linear regression with many controls of limited explanatory power. Quantitative Economics, 12(2), 405-442.
[13] Raza, O., Mansournia, M. A., Foroushani, A. R., & Holakouie-Naieni, K. (2019). Geographically weighted regression analysis: A statistical method to account for spatial heterogeneity. Archives of Iranian medicine, 22(3), 155-160.
[14] Xie Kang,Meng Fangxiu & Zhang Dongdong. (2023). Regional forecasting of PM².5 concentrations: A novel model based on the empirical orthogonal function analysis and Nadaraya–Watson kernel regression estimator. Environmental Modelling and Software,170,
[15] Guo Panpan,Qian Fang,Zhang Wenyao,Yan Huilong,Wang Qiuwang & Zhao Cunlu. (2021). Radial basis function interpolation supplemented lattice boltzmann method for electroosmotic flows in microchannel.. Electrophoresis,42(21-22),2171-2181.
[16] Geronzi Leonardo,Martinez Antonio,Rochette Michel,Yan Kexin,BelBrunon Aline,Haigron Pascal... & Biancolini Marco Evangelos. (2023). Computer-aided shape features extraction and regression models for predicting the ascending aortic aneurysm growth rate.. Computers in biology and medicine,162,107052-107052.
[17] Chunlu Wang,Yanqing Fan,Renjie He,Jiwu Li,Fa Zhao,Xiaohua Zhou & Zubin Chen. (2025). Automatic arrival picking for microseismic signals based on variational mode decomposition and Akaike information criterion.. The Review of scientific instruments,96(4).