remote sensing

MDPI

Article

Extraction of Factors Strongly Correlated with Lightning Activity Based on Remote Sensing Information

Haochen Zhang [1], Yeqiang Deng [1], Yu Wang [1], Lei Lan [1,*], Xishan Wen [1], Chaoying Fang [2] and Jun Xu [2]

[1] School of Electrical Engineering and Automation, Wuhan University, Wuhan 430072, China; zhanghc21@whu.edu.cn (H.Z.); dengyeqiang@whu.edu.cn (Y.D.); yuwang@whu.edu.cn (Y.W.); xswen@whu.edu.cn (X.W.)

[2] State Grid Fujian Electric Power Research Institute, Fuzhou 350007, China; fang_chaoying@fj.sgcc.com.cn (C.F.); xu_jun2@fj.sgcc.com.cn (J.X.)

* Correspondence: lanlei@whu.edu.cn

Abstract: Thunderstorms are a common natural phenomenon posing significant hazards to power systems, structures, and humans. With technological advancements, protection against lightning is gradually shifting from passive to active measures, which require the prediction of thunderstorm occurrences. Current research on lightning warning relies on various data sources, such as satellite data and atmospheric electric field data. However, these studies have placed greater emphasis on the process of warning implementation, overlooking the correlation between parameters used for lightning warning and lightning phenomena. This study relied on the ERA5 dataset and lightning location dataset from 117.5°E to 119.5°E longitude and 24.5°N to 25.5°N latitude during 2020–2021, utilizing Kriging interpolation to standardize the spatiotemporal precision of different parameters. After that, we conducted preliminary screening of the involved parameters based on the chi-squared test and utilized the Apriori algorithm to identify parameter intervals that were strongly associated with the occurrence of lightning. Subsequently, we extracted strong association rules oriented towards the occurrence of lightning and analyzed those rules with respect to lightning current amplitude, types, and ERA5 parameters. We found that thunderstorm phenomena are more likely to occur under specific ranges of temperature, humidity, and wind speed conditions, and we determined their parameter ranges. After that, we divided the target area into regions with different levels of lightning probability based on the strong association rules. By comparing the actual areas where lightning phenomena occurred with the areas at high risk of lightning based on ERA5 parameters, we validated the credibility of the obtained strong association rules.

Keywords: chi-squared test; Apriori algorithm; association analysis; lightning warning; Kriging interpolation

Citation: Zhang, H.; Deng, Y.; Wang, Y.; Lan, L.; Wen, X.; Fang, C.; Xu, J. Extraction of Factors Strongly Correlated with Lightning Activity Based on Remote Sensing Information. *Remote Sens.* **2024**, *16*, 1921. https://doi.org/10.3390/rs16111921

Academic Editor: Richard Müller

Received: 23 March 2024
Revised: 23 May 2024
Accepted: 24 May 2024
Published: 27 May 2024

1. Introduction

Lightning is a common physical phenomenon in nature, characterized by both random distribution at the individual storm level and regular patterns in overall distribution. The high voltage generated by lightning strikes and the high-intensity light/heat effects of lightning severely affect the regular operation of power systems [1]. To reduce the harm caused by lightning, we should transform passive lightning protection measures into active ones [2–4]. For example, we can forecast the occurrence and development of lightning and establish response measures before it occurs [5,6]. To date, there has been a great deal of lightning warning research that has made use of various data sources. However, research on the correlation analysis between lightning and multiple data sources is still lacking. This leads to the occurrence of irrelevant data in the lightning warning process, reducing the weight of effective data and, ultimately, lowering the accuracy of the warning.

Research on lightning warning can be classified according to data and methods. Warning based on atmospheric electric field data can be divided into two categories: thundercloud charge structure warning and atmospheric electric field intensity threshold

warning [7–10]. Warning based on lightning location data generally requires the assistance of other parameters and machine-learning methods to analyze historical data [11–14]. Warning based on radar and satellite data mainly involves obtaining feature thresholds through correlation analysis and calculating the movement direction of lightning areas using the TREC (Tracking Radar Echoes by Correlation) method [15–21]. In addition, there is also research on lightning prediction based on parameters such as cloud top height [22,23], convective available potential energy [24–26], and the upward flux of ice particles [27,28]. All of the above methods have strict requirements for data correlation. If the selected data are poorly correlated with lightning, the accuracy of the final warning results will be disappointing.

In current research on lightning warning, there is often more focus on how to use data for lightning warning, while the process of data filtering is overlooked. This leads to the inclusion of too many low-correlation parameters in the warning process, resulting in inaccurate prediction results. Moon et al. proposed a preprocessing method of undersampling data to improve the predictive performance of imbalanced data [29]. This method was validated to improve the hit rate of predictions. Lin et al. believe that due to the complex evolution process of lightning, unlike stable precipitation weather systems, the movement and development trends of severe convective weather systems often change significantly in a shorter period of time [30]. As the forecast progresses, the difference between the historical trend of lightning and the actual trend gradually improves, limiting the corrective effect of simulation biases to the early stages of the forecast. Based on this, the authors proposed a new attention-focusing mechanism that amplifies the weighting of data from different sources at different stages of prediction. This method can leverage the advantages of multi-source data and improve the accuracy of forecasts. Therefore, it can be seen that preprocessing the data by removing noisy data and then conducting correlation analysis leads to much more reasonable results [31]. How to address the redundancy among data and remove low-correlation data is a crucial step in improving the accuracy of lightning warning. Therefore, further research is needed on the correlation between different data and the weighting of results, which is also the focus of this study.

2. Materials and Methods

2.1. Data Types and Sources

The remote sensing data discussed in this study can be classified into two categories: first, the lightning location system dataset (LLS) acquired from the electromagnetic radiation process during lightning discharges; second, the atmospheric reanalysis dataset (ERA5) obtained from physical models and observational data. The target area in this paper spans from 117.5°E to 119.5°E longitude and 24.5°N to 25.5°N latitude. The research objective was to investigate the correlation between different types of data within this target area. Parameters such as the occurrence time, location, and amplitude of lightning currents were sourced from the lightning location system in Fujian, China. The ADTD Lightning Location System of Fujian consists of more than 10 monitoring stations, including Fuzhou and Xiamen stations. The system achieves a network detection efficiency of over 90% for CG flashes, with accuracy within 500 m and a temporal resolution of approximately 1 ms for network detection; it recorded lightning information for the entire period (2020–2021) in the target research area [32]. For example, the information recorded for a particular lightning event is shown in Table 1 below.

Table 1. Lightning occurrence information.

Lightning Time	Longitude	Latitude	Amplitude/kA	Position	Type
2020-07-15_15:48:46	24.945	117.763	10.2	Fujian Province	Negative ground lightning

The atmospheric reanalysis data were sourced from the ECMWF ERA5 dataset, which provides hourly estimates of a wide range of atmospheric, oceanic, wave, and land surface variables. Before conducting the research in this paper, we performed an initial screening of a large amount of data. Ultimately, we selected a total of 21 parameters, including 100 m U wind component, 100 m V wind component, 10 m U wind component, 10 m V wind component, 2 m temperature, cloud base height, convective precipitation, mean convective precipitation rate, total precipitation, relative humidity on 500/700/850 hPa pressure levels, temperature on 500/700/850 hPa pressure levels, U component of wind on 500/700/850 hPa pressure levels, and V component of wind on 500/700/850 hPa pressure levels as the objects for this paper. The parameters in subsequent figures, tables, and descriptions are all described in abbreviated form. For example, "r850" represents "relative humidity at 850 hPa level", where "r" denotes relative humidity and "850" denotes the results at the 850 hPa level.

2.2. Kriging Interpolation Data Optimization Method

The latitude and longitude resolution of the lightning location system is 0.001°, or approximately 0.1 km, fully meeting the research requirements of this paper. However, the latitude and longitude resolution of the atmospheric reanalysis dataset is 0.25°, or approximately 25 km. According to the research by Bao et al., thunderstorm cloud clusters generally have a range distribution from several kilometers to 20 km, with an average speed of 16.4 m/s [33]. To seek the correlation between different datasets, the current resolution of the atmospheric reanalysis dataset does not meet the requirements. Therefore, we optimized the dataset via Kriging interpolation.

The Kriging interpolation method originated from the inverse distance interpolation method. Its core concept is based on the first law of geography: any attribute in space has correlation, and neighboring entities are more similar. The calculation formula for inverse distance interpolation is as shown in Equation (1):

$$\hat{z} = \sum_{i=1}^{n} \frac{1}{d^{\alpha}} z_i, \tag{1}$$

where $\hat{z}$ represents the calculated result at the target point, d represents the distance between the target point and the known point, and z_i represents the attribute value of the ith known point. Inverse distance interpolation calculates the value of an unknown point by weighting the data from all known points in space, where the weight depends on the inverse (or a power of the inverse) of the distance.

Kriging interpolation further introduces two conditions on this basis, obtaining optimal coefficients, λ_i, for estimating any point in the spatial domain: First, the difference between the estimated value at the target point and the true value is minimized. Second, the expectation of the difference between the estimated value and the true value at any point in space is zero, achieving unbiased estimation. Its mathematical expression is as shown in Equation (2):

$$\begin{array}{l} \hat{z}_o = \sum\limits_{i=1}^{n} \lambda_i z_i \\ \min\limits_{\lambda_i} Var(\hat{z}_o - z_o) \\ E(\hat{z}_o - z_o) = 0 \end{array}, \tag{2}$$

where $\hat{z}_0$ represents the estimated value at the target point, while z_0 represents the true value at the target point. With the assumptions in Equation (2) and properties of covariance, the variance of the estimated value and the true value at any point can be expressed as shown in Equation (3):

$$\begin{aligned} &Var(\hat{z}_o - z_o) \\ &= Var\left(\sum_{i=1}^{n} \lambda_i z_i - z_o\right) \\ &= Var\left(\sum_{i=1}^{n} \lambda_i z_i\right) - 2\mathrm{Cov}\left(\sum_{i=1}^{n} \lambda_i z_i, z_o\right) + \mathrm{Cov}(z_o, z_o) \\ &= \sum_{i=1}^{n}\sum_{j=1}^{n} \lambda_i \lambda_j \mathrm{Cov}(z_i, z_j) - 2\sum_{i=1}^{n} \lambda_i \mathrm{Cov}(z_i, z_o) + \mathrm{Cov}(z_o, z_o) \end{aligned} . \quad (3)$$

To find the optimal set of parameter λ_i, we introduce the semivariance function r_{ij}, as shown in Equation (4):

$$r_{ij} = \sigma^2 - \mathrm{cov}(z_i, z_j) = \frac{1}{2} E\left[(z_i - z_j)^2\right]. \quad (4)$$

After that, we construct the corresponding matrix by Equation (4) and the method of Lagrange multipliers, as shown in Equation (5):

$$\begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} & 1 \\ r_{21} & r_{22} & \cdots & r_{2n} & 1 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ r_{n1} & r_{n2} & \cdots & r_{nn} & 1 \\ 1 & 1 & \cdots & 1 & 0 \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \cdots \\ \lambda_n \\ -\phi \end{bmatrix} = \begin{bmatrix} r_{1o} \\ r_{2o} \\ \cdots \\ r_{no} \\ 1 \end{bmatrix}. \quad (5)$$

To obtain parameter λ_i, we should calculate parameter r_{ij}. For all known points in space, the relationship between r_{ij} and the distance between two points (x_i,y_i and x_j,y_j) can be obtained through fitting, as shown in Equation (6). Based on this function relationship, we can solve the semivariance results at any point in space and then substitute them into Equation (5) to obtain parameter λ_i. Subsequently, we can calculate the attribute value at any point in space using Equation (2).

$$r_{ij} = r(d_{ij}) \quad (6)$$

2.3. Association Analysis Method Based on the chi-Squared Test

Lightning activity is influenced by multiple factors, and its mechanisms are complex. There is no perfect model that can explain the process of lightning's occurrence and development. With the development of neural network algorithms and advances in computer capabilities, forecasting lightning based on strongly correlated parameters has gradually become mainstream [34–36]. In order to enhance the reliability of the forecast and avoid interference from irrelevant factors, we need to extract parameters with strong correlations as the basis for analysis.

The chi-squared test is a non-parametric statistical method based on the chi-squared distribution; it can be used to test whether the distribution of a single parameter conforms to the expected mathematical expectation (chi-squared goodness-of-fit test), or to compare the strength of association between two or more samples (chi-squared test of independence). In this study, we primarily utilized the latter application to assess the correlation between the ERA5 and LLS datasets.

The basic assumption of the chi-squared test is the 0 hypothesis, which assumes that there is no association or difference between the variables under research. When the calculated chi-squared value exceeds the critical value for the chosen degrees of freedom and significance level, the 0 hypothesis is rejected, indicating that there is a significant association between the variables. The calculation formula for the chi-squared test is as shown in Equation (7):

$$\chi^2 = \sum_{i=1}^{q} \frac{(O_i - E_i)^2}{E_i} = \sum_{i=1}^{q} \frac{(O_i - np_i)^2}{np_i}, \quad (7)$$

where q represents the number of distribution intervals for a certain ERA5 datum, n is the total sample size, p_i is the proportion of the data distributed in the ith interval, O_i is the quantity of samples falling into the ith interval, and E_i is the expected quantity within the ith interval. The chi-squared value indicates the degree of deviation between the observed and expected values. A larger chi-squared value means a more significant difference between the two variable categories. The basic steps for conducting the chi-squared test in this research are shown below.

(a) Retrieve ERA5 data for lightning grid points;
(b) Randomly sample ERA5 data from non-lightning grid points;
(c) Divide intervals based on the maximum and minimum values of parameters;
(d) Divide the sample data into intervals and create the contingency table;
(e) Perform the chi-squared test.

To conduct a chi-squared test, it is necessary to match the lightning activity data with the ERA5 data, which can be achieved through two approaches: First, as shown in Figure 1a, directly obtain the ERA5 parameters at the specified spatiotemporal location via the Kriging interpolation methods based on the coordinates of each lightning strike point and a designated grid size. Second, as shown in Figure 1b, divide the selected area and ERA5 parameters into grids (0.01° × 0.01°, approximately 1 km × 1 km). Then, classify lightning strike points into corresponding grid points based on their coordinates. For this research, we chose the second method, which offers advantages for subsequent correlation analysis and feature parameter analysis. Additionally, this method ensures a uniform format for inputs and outputs.

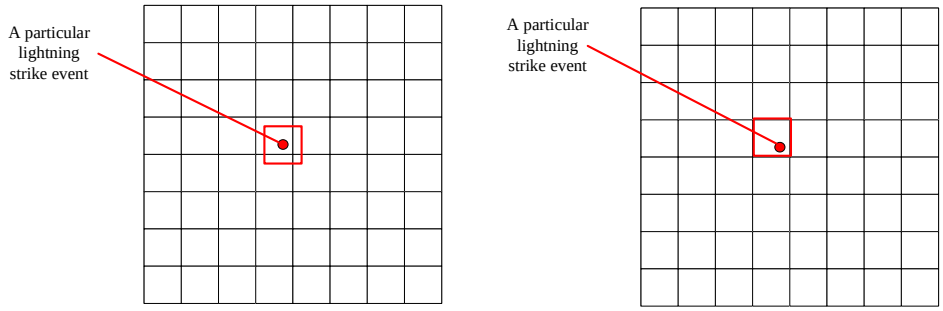


Figure 1. Method to match LLS data with ERA5 data: (**a**) method based on lightning strike locations; (**b**) method based on grid partitioning.

Based on the above method, we finally obtained an ERA5 parameter matrix of 20,301 × 35,136 × n in size, where 20,301 represents the number of grid intervals obtained by dividing the target area into grids of 0.01°, 35,136 represents the number of intervals obtained by dividing the target area into 15 min intervals for the entire year of 2020, and n represents different types of ERA5 parameters. Combined with the LLS data, we created a lightning activity database, as shown in Table 2 below.

Table 2. Lightning activity database.

Time Label	Space Label	Amplitude/kA	Type *
4324	10,107	129.9	PCG
4324	10,505	96.8	IC
4324	10,703	12.8	NCG
4325	10,708	9.4	NCG
4325	11,511	18.1	NCG

* IC: intra-cloud flash; NCG: negative cloud-to-ground flash; PCG: positive cloud-to-ground flash.

2.4. Apriori Algorithm

The Apriori algorithm can analyze the association between LLS data and ERA5 data. This algorithm explores valuable and representative data by identifying frequent itemsets from large datasets. The data uncovered through this process exhibit certain association relationships. The algorithm involves several concepts [37], including items and itemsets, transactions, support, confidence, minimum support and minimum confidence, frequent itemsets, and strong association rules, defined as follows:

(a) Items and itemsets: Items refer to individual elements in the dataset. Itemsets are combinations of one or more items.
(b) Transactions: Transactions represent samples or records in the dataset, typically expressed as sets of itemsets.
(c) Support: The probability that itemsets A and B occur at the same time is called the support degree of the association rule. The expression of support is represented as shown in Equation (8):

$$Support(A \rightarrow B) = P(A \cup B) \tag{8}$$

(d) Confidence: If itemset A occurs, then the probability of itemset B occurring is the confidence level of the association rule. The expression of confidence is represented as shown in Equation (9):

$$Confidence(A \rightarrow B) = \mathrm{P}(B|A) \tag{9}$$

(e) Minimum support and minimum confidence: Thresholds set to filter frequent itemsets and strong association rules based on their support and confidence values.
(f) Frequent itemsets: Itemsets that support no less than the minimum support threshold.
(g) Strong association rules: Association rules with confidence no less than the minimum confidence threshold.

Figure 2 shows the implementation process of the Apriori algorithm. This process is mainly divided into two steps: The first step is to find all frequent itemsets. This process is further divided into two steps, as follows:

Step 1: Connection step. To search for frequent k itemsets, L_k, all frequent $(k-1)$ itemsets should be self-connected to generate k item candidate sets, C_k. Assuming that I_m and I_n belong to frequent $(k-1)$ itemsets, then $I_i[j]$ represents the jth item in I_i. If the items in all itemsets are sorted alphabetically, then the frequent $(k-1)$ itemsets are self-connected.

Step 2: Pruning step. For a particular minimum support threshold, the support of $k-1$ item candidate sets is compared with the minimum support threshold, and itemsets smaller than the threshold are eliminated to obtain $k-1$ frequent itemsets.

The second step is to obtain frequent itemsets from strong association rules. In other words, itemsets that do not exceed the predetermined minimum support threshold are eliminated. If the remaining rules meet the predetermined minimum confidence threshold, then strong association rules are mined.

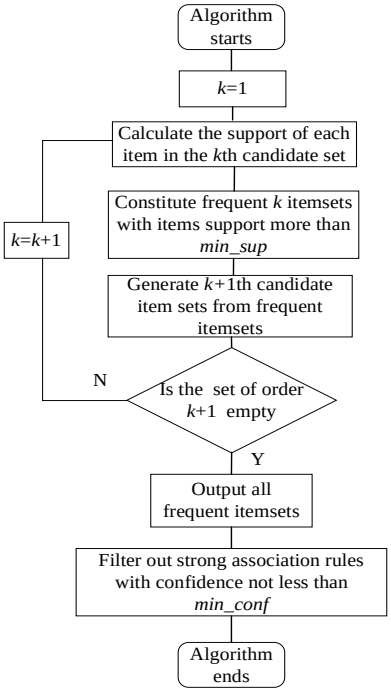
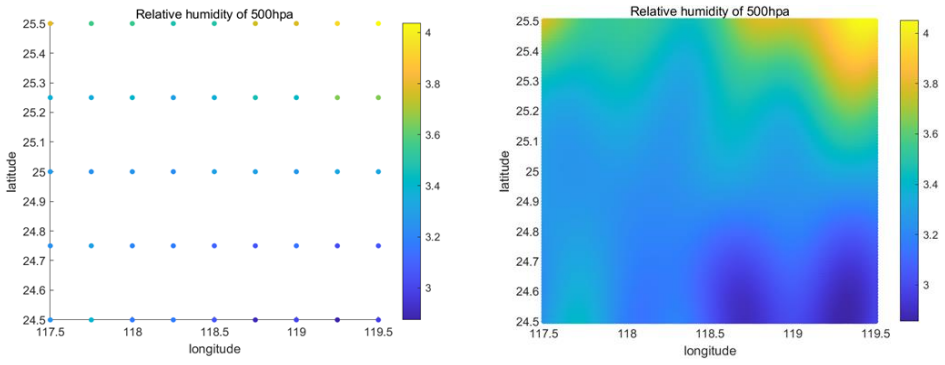


Figure 2. Apriori algorithm flow.

3. Results

3.1. Optimization of ERA5 Data

For the ERA5 data within the range of 117.5°E to 119.5°E longitude and 24.5°N to 25.5°N latitude, this research utilized the data optimization method proposed in Section 2.2 to increase the spatial resolution to 0.01° × 0.01° (approximately 1 km × 1 km) and enhance the temporal resolution to 15 min. For example, considering the relative humidity at 500 hPa pressure at 08:00 on 1 January 2020, Figure 3a illustrates the original resolution, while Figure 3b represents the optimized data resolution.



(a) (b)

Figure 3. Kriging interpolation results for ERA5 parameters: (**a**) original resolution; (**b**) optimized resolution.

For example, we optimized partial typical ERA5 parameters at 08:00 on 1 January 2020, and the results are shown in Figure 4.

The original data have a temporal resolution of 1 h. However, one hour is too long compared to the duration of thunderstorm activity. This leads to data dispersion and, consequently, to poor results in correlation analysis. In this research, considering the entire year of 2020, with a total of 8784 h, we generated 35,136 time labels via the linear interpolation optimization method. To illustrate this, we present the relative humidity at 850 hPa pressure from 02:00 to 05:00 on 1 January 2020, as shown in Figure 5.

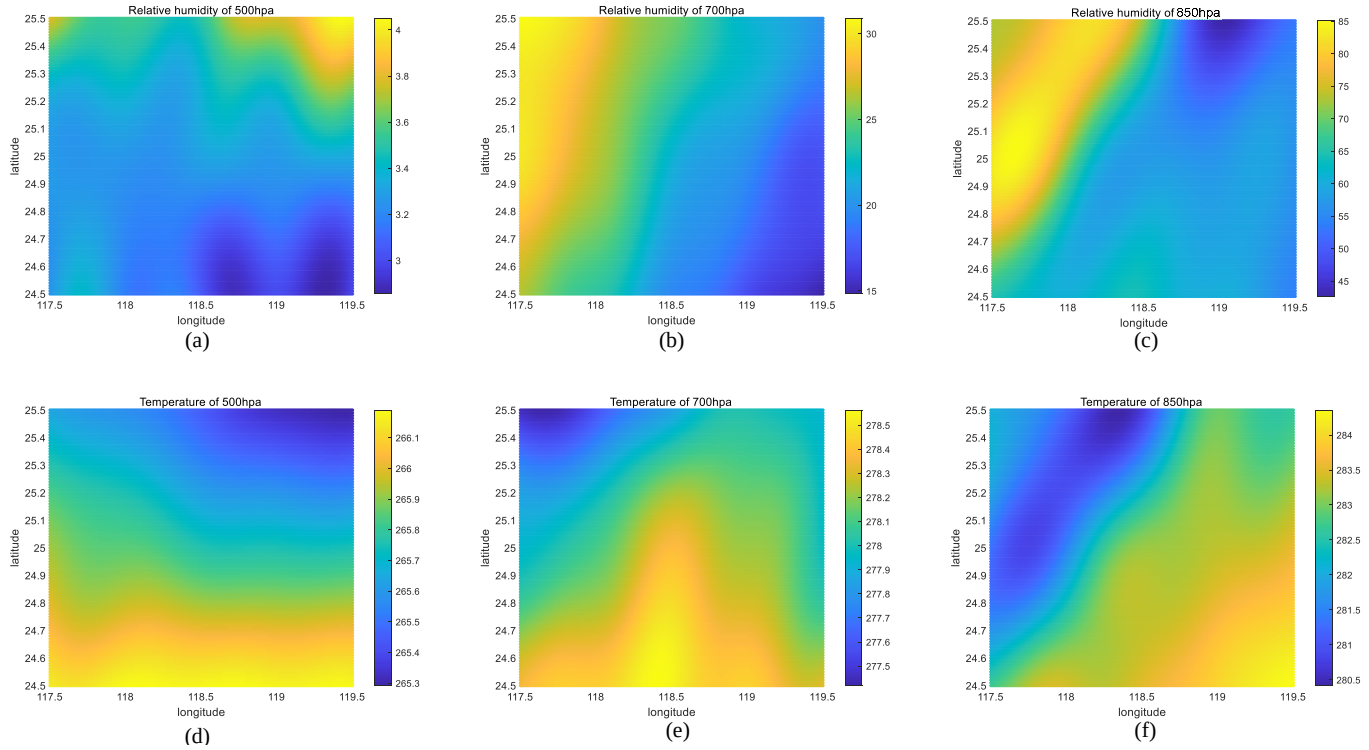


Figure 4. Partial optimization results of ERA5: (**a**) relative humidity of 500 hPa (r500); (**b**) relative humidity of 700 hPa (r700); (**c**) relative humidity of 850 hPa (r850); (**d**) temperature of 500 hPa (t500); (**e**) temperature of 700 hPa (t700); (**f**) temperature of 850 hPa (t850).

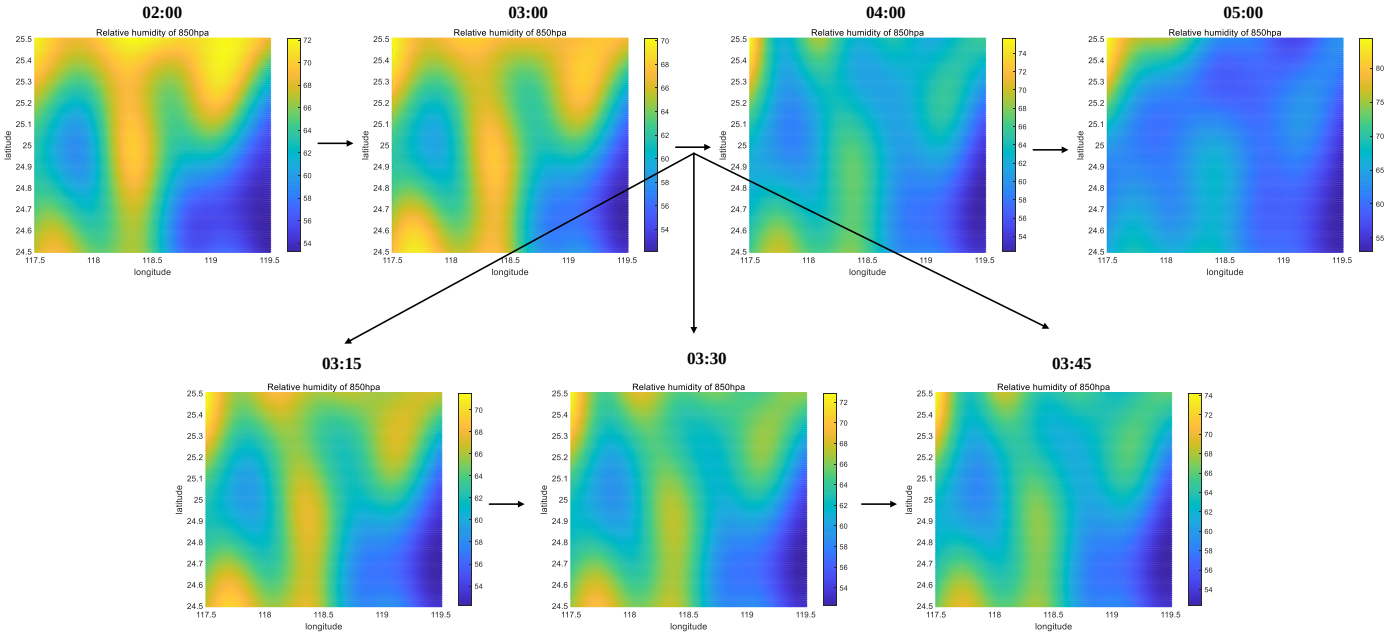


Figure 5. Variation in relative humidity with time at 850 hPa pressure.

3.2. Analysis of the chi-Squared Test Results

In 2020, a total of 13,584 data points were recorded by the lightning location system in the target area. After partitioning the lightning strikes based on time and space intervals, we merge lightning activities occurring at the same time within the same grid point. We obtained a total of 8978 data points. This data volume is significantly smaller compared to the typical grid division of meteorological parameters, which is 20,301 × 35,136. This indicates that there are significant differences in the occurrence of lightning in the natural

environment. The chi-squared test requires samples of both lightning occurrences and non-lightning occurrences. How to select an appropriate ratio of positive to negative samples directly affects the results. Figure 6 shows the chi-squared test results under different sampling ratios.

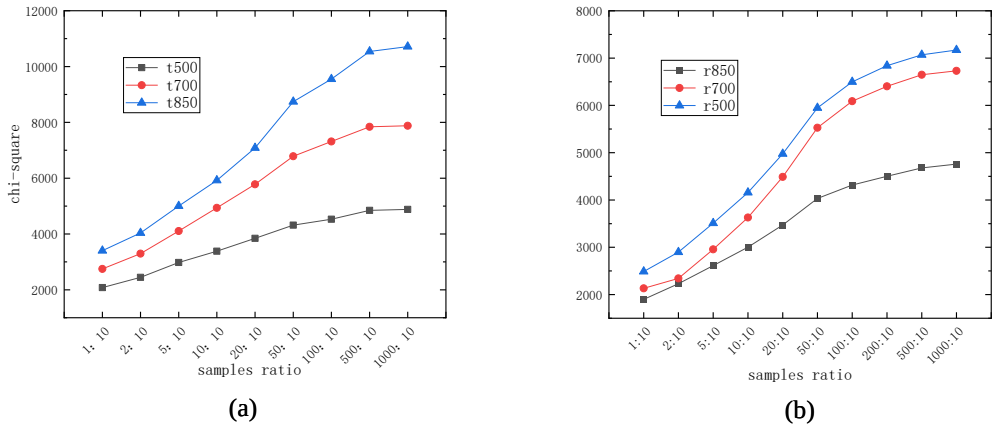


Figure 6. Chi-squared test results under different sampling ratios: (**a**) temperature results at different pressure levels; (**b**) humidity results at different pressure levels.

As can be seen from the above figure, as the sampling proportion increases it leads to an increase in the sample size, and the results of the chi-squared test also gradually increase. This reflects the inherent contradiction in the chi-squared test itself under large sample sizes, between theoretical results and practical applicability. According to the chi-squared expression, assuming that the distribution proportions of the samples remain unchanged, if the total sample size increases tenfold the corresponding chi-squared value will also increase tenfold. In this case, with the significance levels and degrees of freedom unchanged, the judgment results may change. In addition to the sampling ratio, the method of dividing sample intervals also directly affects the accuracy of the chi-squared test. The chi-square test requires ensuring that each interval division contains a sufficient number of samples, otherwise it may lead to deviations from the actual calculation results. When conducting chi-square tests on our parameters, we found that empty intervals occasionally occur when the number of divisions exceeded 10. This occurrence became more frequent when the number of divisions exceeded 15. When the number of divisions is too small, it can also lead to inaccurate analysis of parameter correlations. For example, in Figure 7 the chi-square test results fluctuate for intervals 3–7, while the results for intervals 7–15 tend to stabilize. Based on the above reasons, we chose 10 intervals as the division standard.

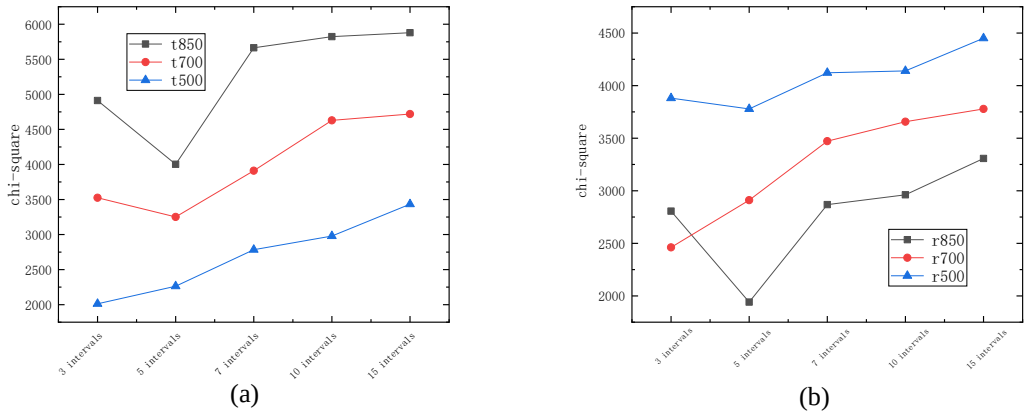


Figure 7. Chi-squared test results under different intervals: (**a**) temperature results at different pressure levels; (**b**) humidity results at different pressure levels.

Based on the analysis of the above results, for the correlation analysis between LLS data and ERA5 data, dividing the data into 10 equally spaced intervals produced better results. Due to the inherent limitations of the chi-squared test, when the sample size is large its significance test loses credibility. It can only judge the strength of the data correlation based on the magnitude of the chi-squared test value. Figure 8 shows the chi-squared test results and differences between adjacent parameters for typical ERA5 parameters, with positive and negative samples extracted in a 1:1 ratio and divided into 10 equally spaced intervals. Figure 8 shows that when the typical parameters are arranged in descending order based on the calculated chi-squared results, it is visually apparent that there is a significant difference between parameter u10 and parameter u700. Using this as a dividing line(marked with a red line in Figure 8), the target parameters can be classified into a set of strongly correlated parameters and a set of weakly correlated parameters.

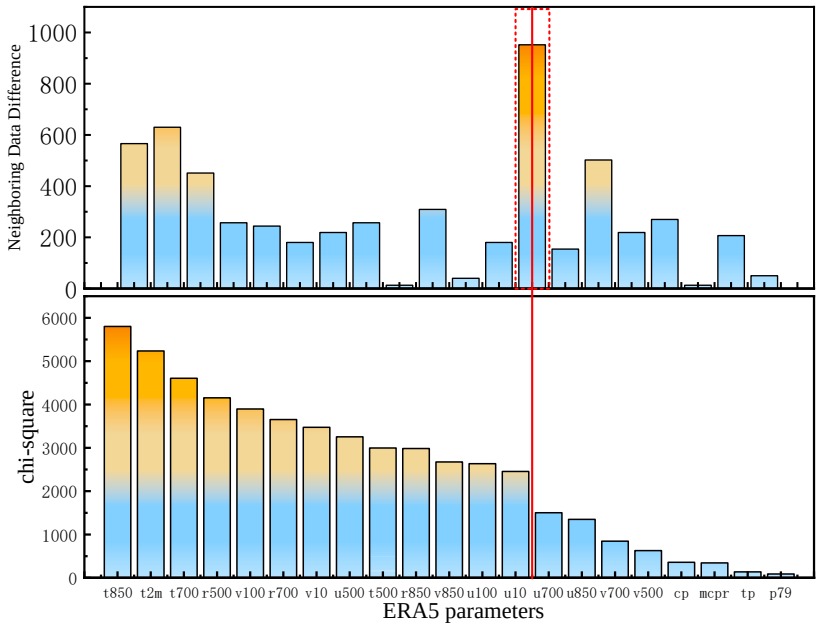


Figure 8. Chi-squared test results between ERA5 data and LLS data.

3.3. Apriori Association Rule Extraction

Based on the analysis in Section 3.2, we ultimately selected the u10 parameter as the critical parameter. We selected the dataset with correlations stronger than this parameter for subsequent analysis. After that, we performed data extraction based on a 1:1 ratio of lightning occurrence samples and non-lightning samples. Each parameter was divided into 10 equally spaced intervals. The partially formed parameter matrix is shown in Table 3 and Appendix A, where A–M correspond to different parameters and numbers correspond to different intervals. The unit of parameters related to temperature is K, parameters related to humidity are given as percentages, and the unit of parameters related to wind is m/s.

Using the Apriori algorithm, we extracted high-frequency data intervals corresponding to each parameter and lightning occurrences, as shown in Figure 9 and Table 4. The confidence of all of the extracted strong association rules was greater than 0.6. It can be considered that when the environmental conditions reflect the intervals in Table 4, lightning phenomena are more likely to occur. The parameters in this research can be classified into three types: A–D are related to temperature, E–G are related to humidity, and H–M are related to wind speed.

Table 3. Part of ERA5 data and their corresponding interval ranges.

	t2m (A) K	r850 (E) %	u500 (I) m/s
1	(272.79, 276.73)	(0.76, 11.01)	(−11.62, −6.18)
2	(276.73, 280.68)	(11.01, 21.26)	(−6.18, −0.74)
3	(280.68, 284.62)	(21.26, 31.51)	(−0.74, 4.7)
4	(284.62, 288.57)	(31.51, 41.76)	(4.7, 10.14)
5	(288.57, 292.51)	(41.76, 52)	(10.14, 15.58)
6	(292.51, 296.46)	(52, 62.25)	(15.58, 21.02)
7	(296.46, 300.4)	(62.25, 72.5)	(21.02, 26.46)
8	(300.4, 304.34)	(72.5, 82.75)	(26.46, 31.9)
9	(304.34, 308.29)	(82.75, 92.99)	(31.9, 37.34)
10	(308.29, 312.23)	(92.99, 103.24)	(37.34, 42.78)

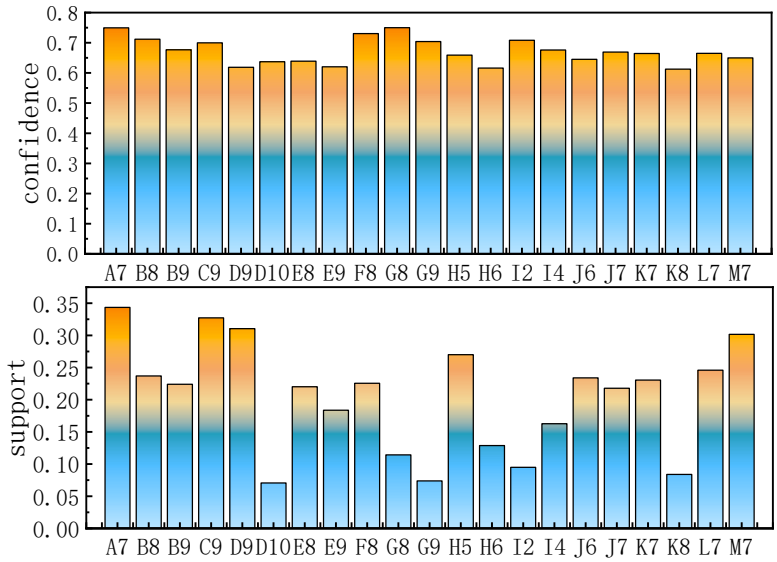


Figure 9. Support and confidence of the rules for lightning occurrence.

Table 4. Range of strong association parameter intervals.

Strong Association Parameter	Range	Strong Association Parameter	Range	Strong Association Parameter	Range
A7	(296.46, 300.4)	E9	(82.75, 92.99)	I4	(4.7, 10.14)
B8	(289.87, 292.65)	F8	(73.55, 84.05)	J6	(−2.47, 0.85)
B9	(292.65, 295.43)	G8	(74.34, 84.89)	J7	(0.85, 4.18)
C9	(283.81, 285.68)	G9	(84.89, 95.45)	K7	(−0.49, 1.74)
D9	(268.3, 270.3)	H5	(0.2, 4.3)	K8	(1.74, 3.97)
D10	(270.3, 272.31)	H6	(4.3, 8.4)	L7	(0.16, 2.8)
E8	(72.5, 82.75)	I2	(−6.18, −0.74)	M7	(−0.63, 1.1)

With a support threshold of 0.05 and lightning occurrence as the target outcome, association rules were derived from the entire database. The ten strongest association rules are presented in Table 5. The occurrences of different parameter intervals within these ten rules are tabulated in Figure 10.

The data involved in this research can generally be categorized into three types: temperature, humidity, and wind speed. When facing lightning events, strong association parameters of the same type often tend to occur simultaneously. The conclusion drawn from the previous analysis suggests that, for a particular area, the higher the number of parameters falling into the strong association intervals, the higher the probability of lightning's occurrence. However, it is common for different parameters of the same type

to simultaneously fall into strong association intervals. In such cases, it is essential not only to consider the number of parameters falling into strong association intervals but also to pay attention to the types of parameters involved. For example, the association rules in Table 5 mainly involve temperature, humidity, and wind speed types. Taking temperature parameters as an example, Figure 11 demonstrates the correlation between temperature-related parameters A–D. The support for each strong association parameter interval generally exceeds 0.3, and the confidence exceeds 0.6. This confirms that strong association intervals of the same type of parameter tend to occur simultaneously. During lightning warning, it is advisable to appropriately reduce the weight of subsequent occurrences of different parameters within the same type. To facilitate the display of the correlation between different parameters, we merged the strong association intervals of the same parameter. For example, intervals B8 and B9 were combined into the {B8, B9} interval.

Table 5. Association rule results for lightning occurrence.

	Antecedent Rule	Consequent Rule	Support	Confidence
1	A7, D9, E9, H5, L6	Yes	0.051	0.971
2	A7, C9, E8, H5, J6, L6	Yes	0.051	0.967
3	A7, C9, E8, H5, L6	Yes	0.051	0.966
4	A7, B9, C9, H5, J6, L6, M7	Yes	0.053	0.965
5	A7, C9, H5, J6, L6, M7	Yes	0.082	0.964
6	A7, D9, E9, J6, H5	Yes	0.050	0.964
7	A7, G8, H5, M7	Yes	0.058	0.963
8	A7, B9, C9, J6, H5	Yes	0.073	0.963
9	A7, B9, C9, H5, L6, M7	Yes	0.054	0.962
10	F8, I4, J7, K7, L6, M7	Yes	0.051	0.962

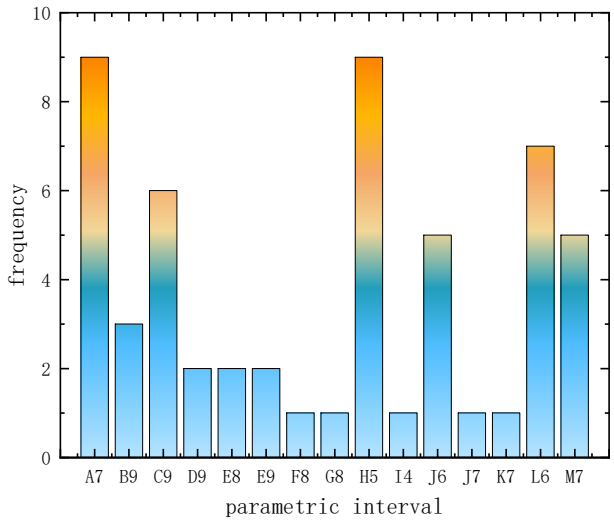


Figure 10. Frequency of strong association intervals' occurrence.

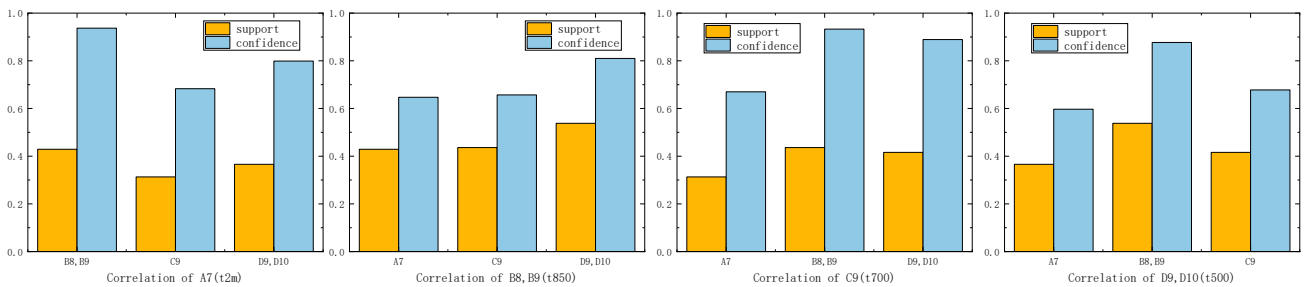
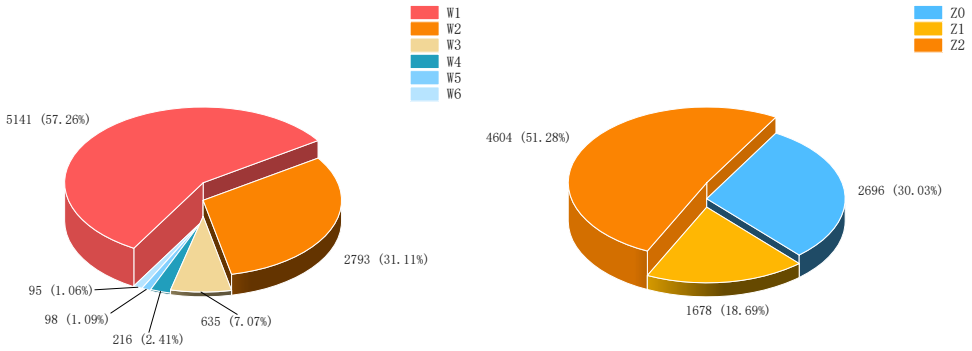


Figure 11. Correlations among data of the same type.

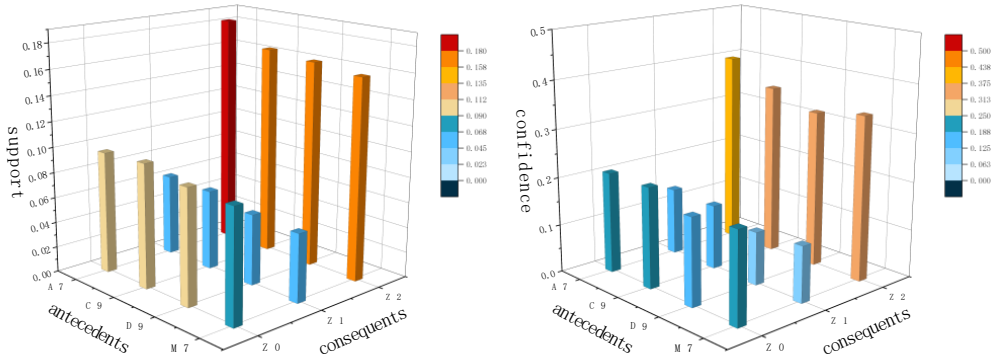
On the basis of all lightning occurrences, we classified lightning strikes based on their amplitude and type. Regarding the amplitude of lightning strikes, we categorized it into six levels (W1–W6), with 20 kA as the threshold. The number and proportion of each type are illustrated in Figure 12a. For lightning strike types, they can be divided into three categories: intra-cloud flash (IC), positive cloud-to-ground (PCG), and negative cloud-to-ground (NCG), labeled as Z0–Z2. The number and proportion of each type are shown in Figure 12b.



(a) (b)

Figure 12. Distribution percentage of lightning currents: (**a**) division of lightning current amplitude ranges; (**b**) division of lightning current type ranges.

The distribution of lightning current amplitude and lightning current type exhibits certain similarities. Taking the lightning type as an example, the support and confidence of the obtained association rules are illustrated in Figure 13 below. The parameters selected in this research are not sensitive to changes in lightning type and amplitude. The same parameter intervals correspond to three different lightning current types, indicating three distinct association rules. For example, the A7 interval simultaneously corresponds to three different types of lightning currents: Z0, Z1, and Z2. The support and confidence of this association rule vary with the distribution proportion of different types of lightning events. The antecedent parameter intervals appearing in the association rules in Figure 13 overlap with the antecedent parameter intervals in Table 4, further demonstrating that the parameters selected in this study are more sensitive to whether lightning currents occur, while lacking discernment regarding changes in lightning amplitude and type.



(a) (b)

Figure 13. Support and confidence of association rules for lightning types: (**a**) support of lightning type; (**b**) confidence of lightning type.

The conclusion that the parameters selected in this research are not sensitive to changes in lightning current amplitude and type can be illustrated by analyzing the correlations between different parameters and lightning amplitude. After reducing the requirements for support and confidence, the results shown in Figure 14 can be obtained. From this

figure, we can conclude that there is no single correspondence between different parameters and lightning current amplitude. Instead, a certain temperature interval corresponds to multiple lightning current amplitude intervals simultaneously. The two graphs on the right in Figure 15 show the support and confidence. The first concerns the antecedent A7 and all lightning amplitude intervals, while the second concerns the consequent W1 and all t2m intervals. It can be observed that as the proportion of intervals W1 to W6 gradually decreases, their corresponding support and confidence also decrease. Meanwhile, in the extracted association rules with W1 as the target result, the support and confidence for the corresponding A7 interval are higher, consistent with the strong association rules obtained in Table 4.

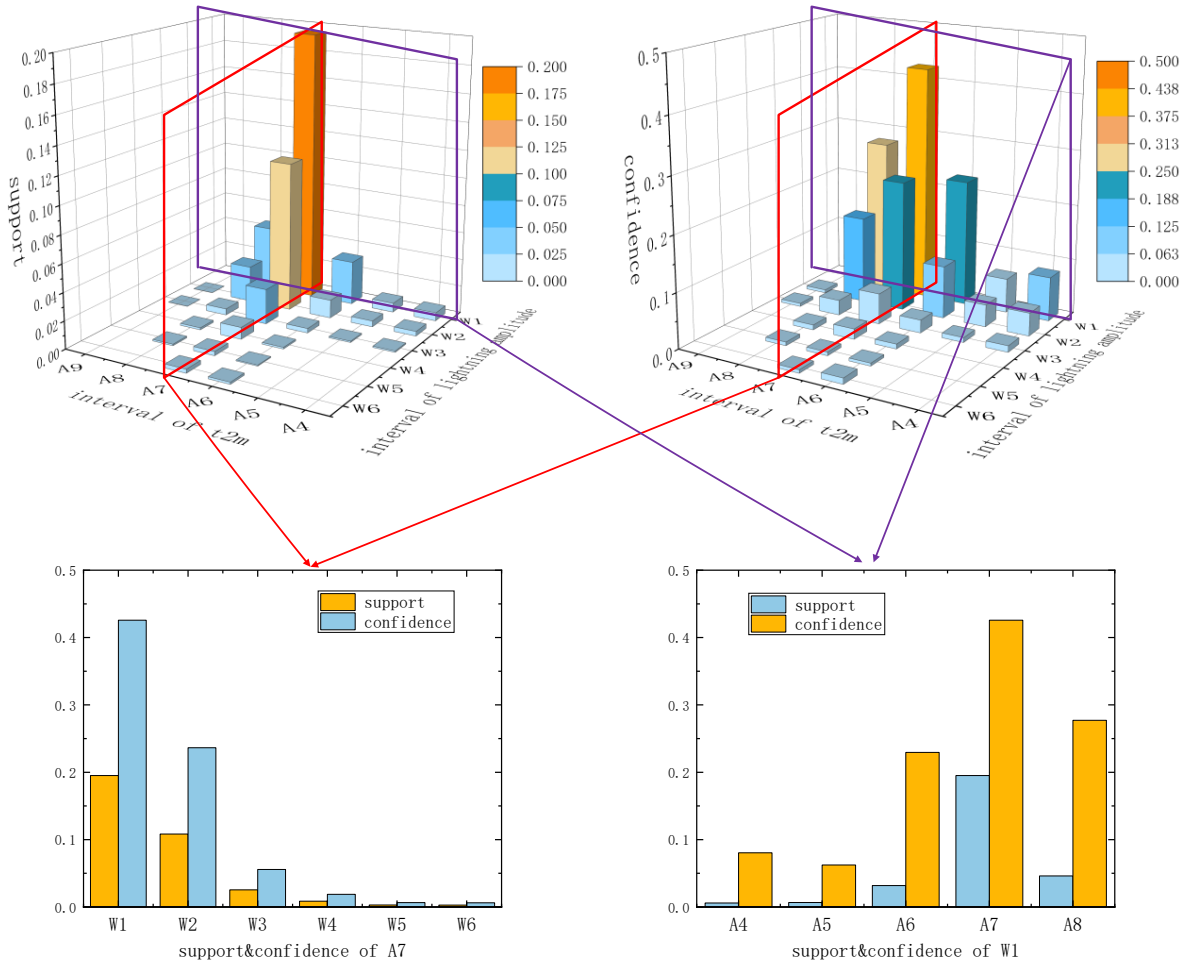


Figure 14. Correlation between the t2m parameter and the distribution of lightning amplitudes.

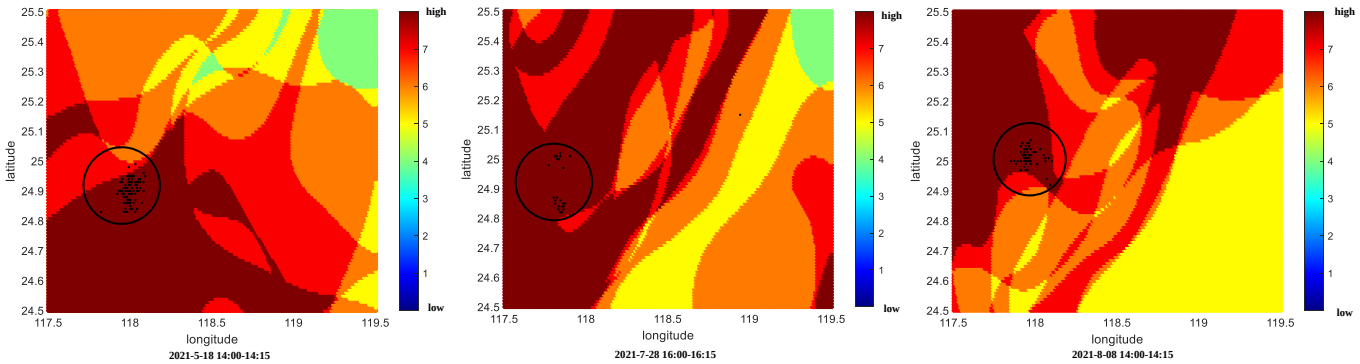


Figure 15. Distribution of lightning events and their probabilities at different times.

4. Discussion

This study established strong correlations between ERA5 parameters and lightning activity, revealing that lightning activity is primarily associated with temperature, humidity, and wind speed.

Research by Chen et al. indicates that the occurrence and development of lightning are closely related to the water vapor content and ice particle concentration [38], which is consistent with the strong correlation between lightning activity and relative humidity found in this study. Research by Zeng et al. suggests that relative humidity greater than 60% in the mid-level (500–340 hPa) atmosphere and greater than 28% in the upper-level (340–130 hPa) atmosphere is conducive to the development of severe thunderstorms [39]. We further investigated and discussed this conclusion by analyzing the relationship between relative humidity at 500, 700, and 850 hPa pressure and lightning activity, resulting in strong association rules. When the antecedents are "E8", "E9", "F8", "G8", and "G9", lightning activity is more likely to occur.

Dai et al.'s research indicates that warm and humid environments are more likely to experience severe storms [40]. This is consistent with the conclusion of this research, which highlights the strong correlation between thunderstorm activity and temperature. When the antecedents are "A7", "B8", "B9", "C9", "D9", and "D10", thunderstorm activity is more likely to occur.

Based on the research conducted in this paper, we classified the 13 strongly correlated parameters in the target area, and we counted the number of parameters falling into the strongly correlated intervals shown in Table 4. Since there were a total of four temperature-related parameters, three humidity-related parameters, and six wind-speed-related parameters, in the statistical process the upper limit for counting was set to seven. This means that situations where more than seven parameters fell into the strongly correlated intervals were not distinguished, representing the scenario where more than half of the parameters are in the strongly correlated intervals and at least two different types of strongly correlated parameters are present. As shown in Figure 15, taking three different instances of lightning strikes in 2021 as examples, the black dots and circles represent the regions where lightning events occurred, while the different-colored blocks represent the probability of lightning's occurrence. The different colors represent the number of parameters falling into strongly correlated intervals. The more types of parameters within a strongly correlated interval that are present in the same grid, the higher we consider the likelihood of a lightning strike occurring at that grid point. The results in this graph visually demonstrate that the areas where lightning occurs are all within high-probability regions. This indicates the consistency between the extracted strong association rules for lightning events and the actual occurrences.

5. Conclusions

The purpose of this research was to eliminate low-correlation redundant data and extract strong correlation rules between meteorological parameters and lightning activity parameters, forming a database suitable for providing mid–short-term lightning warning. The conclusions obtained are as follows:

(a) This study utilized the Kriging interpolation method to enhance the resolution of ERA5 data within the range of 117.5°E to 119.5°E longitude and 24.5°N to 25.5°N latitude. Furthermore, we established grid-point correspondence between lightning activity data and meteorological data for this region in the year 2020.

(b) This study utilized the chi-squared test and the Apriori algorithm to extract strong association rules between lightning activity and meteorological parameters in the target area. For 2 m temperature between 296.46 and 300.4 K, T850 between 289.87 and 295.43 K, T700 between 283.91 and 285.68 K, T500 between 268.3 and 272.31 K, R850 between 72.5% and 92.99%, R700 between 73.55% and 84.05%, R500 between 74.34% and 95.45%, V850 between 0.2 and 8.4 m/s, U500 between −6.18 and 10.14 m/s, V100 between −2.47 and 4.18 m/s, U100 between −0.49 and 3.97 m/s, V10 between

0.16 and 2.8 m/s, and U10 between −0.63 and 1.1 m/s, lightning activity was more likely to occur, and we obtained the ranges of each strong association interval.

(c) According to the results of this paper, variables such as temperature, humidity, and wind speed show higher sensitivity to the occurrence of lightning. Their association rule supports are greater than 0.1, and their confidence levels are all greater than 0.6. However, these variables are not sensitive to lightning amplitude and lightning type.

(d) Based on the strong association parameter intervals extracted from this research, we divided the target area into grid points, and we then counted the number of parameters falling into each grid point within the strong association parameter intervals. Based on that, we formed a distribution map of lightning probability, and the actual lightning occurrences were found to be concentrated in regions with high lightning probability, which is consistent with the conclusions drawn by this research.

Author Contributions: Conceptualization, H.Z. and Y.D.; methodology, H.Z.; software, H.Z.; validation, H.Z., Y.D. and L.L.; formal analysis, H.Z.; investigation, H.Z. and Y.W.; resources, C.F.; data curation, H.Z.; writing—original draft preparation, H.Z.; writing—review and editing, H.Z., L.L. and Y.D.; visualization, H.Z.; supervision, L.L. and X.W.; project administration, Y.W., J.X. and Y.D.; funding acquisition, Y.W., J.X. and X.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Science and Technology Project of State Grid "Research on Disaster Warning and Risk Assessment Technology for Lightning Strike in Large Wind Farms" (No. 5500-202221143A-1-1-ZN).

Data Availability Statement: The data reported in this paper are presented in, archived in, or available from the ECMWF or Climate Data Store (CDS) (https://cds.climate.copernicus.eu/, accessed on 9 May 2024)

Conflicts of Interest: Authors Chaoying Fang and Jun Xu were employed by the companyState Grid Fujian Electric Power Research Institute. The authors declare that this study received funding from [State Grid Corporation of China]. The funder was not involved in the study design, collection, analysis, interpretation of data, the writing of this article or the decision to submit it for publication. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Appendix A

Table A1. Temperature of ERA5 data and their corresponding interval ranges.

	t2m (A) K	t850 (B) K	t700 (C) K	t500 (D) K
1	(272.79, 276.73)	(270.41, 273.19)	(268.8, 270.68)	(252.25, 254.26)
2	(276.73, 280.68)	(273.19, 275.97)	(270.68, 272.55)	(254.26, 256.26)
3	(280.68, 284.62)	(275.97, 278.75)	(272.55, 274.43)	(256.26, 258.27)
4	(284.62, 288.57)	(278.75, 281.53)	(274.43, 276.3)	(258.27, 260.27)
5	(288.57, 292.51)	(281.53, 284.31)	(276.3, 278.18)	(260.27, 262.28)
6	(292.51, 296.46)	(284.31, 287.09)	(278.18, 280.06)	(262.28, 264.29)
7	(296.46, 300.4)	(287.09, 289.87)	(280.06, 281.93)	(264.29, 266.29)
8	(300.4, 304.34)	(289.87, 292.65)	(281.93, 283.81)	(266.29, 268.3)
9	(304.34, 308.29)	(292.65, 295.43)	(283.81, 285.68)	(268.3, 270.3)
10	(308.29, 312.23)	(295.43, 298.21)	(285.68, 287.56)	(270.3, 272.31)

Table A2. Relative humidity of ERA5 data and their corresponding interval ranges.

	r850 (E) **%**	**r700 (F)** **%**	**r500 (G)** **%**
1	(0.76, 11.01)	(0.01, 10.5)	(0.44, 11)
2	(11.01, 21.26)	(10.5, 21.01)	(11, 21.56)
3	(21.26, 31.51)	(21.01, 31.51)	(21.56, 32.11)
4	(31.51, 41.76)	(31.51, 42.02)	(32.11, 42.67)
5	(41.76, 52)	(42.02, 52.53)	(42.67, 53.23)
6	(52, 62.25)	(52.53, 63.04)	(53.23, 63.78)
7	(62.25, 72.5)	(63.04, 73.55)	(63.78, 74.34)
8	(72.5, 82.75)	(73.55, 84.05)	(74.34, 84.89)
9	(82.75, 92.99)	(84.05, 94.56)	(84.89, 95.45)
10	(92.99, 103.24)	(94.56, 105.07)	(95.45, 106.01)

Table A3. Wind speed of ERA5 data and their corresponding interval ranges.

	v850 **m/s**	**u500** **m/s**	**v100** **m/s**	**u100** **m/s**	**v10** **m/s**	**u10** **m/s**
1	(−16.2, −12.1)	(−11.62, −6.18)	(−19.1, −15.78)	(−13.88, −11.65)	(−15.69, −13.05)	(−11.01, −9.28)
2	(−12.1, −8)	(−6.18, −0.74)	(−15.78, −12.45)	(−11.65, −9.42)	(−13.05, −10.41)	(−9.28, −7.55)
3	(−8, −3.9)	(−0.74, 4.7)	(−12.45, −9.12)	(−9.42, −7.19)	(−10.41, −7.76)	(−7.55, −5.82)
4	(−3.9, 0.2)	(4.7, 10.14)	(−9.12, −5.8)	(−7.19, −4.96)	(−7.76, −5.12)	(−5.82, −4.09)
5	(0.2, 4.3)	(10.14, 15.58)	(−5.8, −2.47)	(−4.96, −2.72)	(−5.12, −2.48)	(−4.09, −2.36)
6	(4.3, 8.4)	(15.58, 21.02)	(−2.47, 0.85)	(−2.72, −0.49)	(−2.48, 0.16)	(−2.36, −0.63)
7	(8.4, 12.49)	(21.02, 26.46)	(0.85, 4.18)	(−0.49, 1.74)	(0.16, 2.8)	(−0.63, 1.1)
8	(12.49, 16.59)	(26.46, 31.9)	(4.18, 7.5)	(1.74, 3.97)	(2.8, 5.45)	(1.1, 2.83)
9	(16.59, 20.69)	(31.9, 37.34)	(7.5, 10.83)	(3.97, 6.21)	(5.45, 8.09)	(2.83, 4.55)
10	(20.69, 24.79)	(37.34, 42.78)	(10.83, 14.16)	(6.21, 8.44)	(8.09, 10.73)	(4.55, 6.28)

References

1. Gu, S.; Wang, J.; Wu, M.; Guo, J.; Zhao, C.; Li, J. Study on Lightning Risk Assessment and Early Warning for UHV DC Transmission Channel. *High Volt.* **2019**, *4*, 144–150. [CrossRef]
2. Tovar, C.; Aranguren, D.; Lopez, J.; Inampues, J.; Torres, H. Lightning Risk Assessment and Thunderstorm Warning Systems. In Proceedings of the 2014 International Conference on Lightning Protection (Iclp), Shanghai, China, 11–18 October 2014; pp. 1870–1874.
3. Tao, H.; Gu, S.; Wang, H.; Feng, W.; Guo, J.; Wang, Y.; Zhang, L. Method of Lightning Warning Based on Atmospheric Electric Field and Lightning Location Data. In Proceedings of the 2016 33rd International Conference on Lightning Protection (Iclp), Estoril, Portugal, 25–30 September 2016.
4. Li, X.; Yang, L.; Yin, Q.; Yang, Z.; Zhou, F. Lightning Risk Warning Method Using Atmospheric Electric Field Based on EEWT-ASG and Morpho. *Atmosphere* **2023**, *14*, 1002. [CrossRef]
5. Meng, Q.; Yao, W.; Xu, L. Development of Lightning Nowcasting and Warning Technique and Its Application. *Adv. Meteorol.* **2019**, *2019*, 2405936. [CrossRef]
6. Mansouri, E.; Mostajabi, A.; Tong, C.; Rubinstein, M.; Rachidi, F. Lightning Nowcasting Using Solely Lightning Data. *Atmosphere* **2023**, *14*, 1713. [CrossRef]
7. Montanya, J.; Bergas, J.; Hermoso, B. Electric Field Measurements at Ground Level as a Basis for Lightning Hazard Warning. *J. Electrost.* **2004**, *60*, 241–246. [CrossRef]
8. Lu, Y.; Zhou, Z.; Gu, S.; Wu, D.; Guo, J.; Tao, H. Research on Lightning Warning Method Based on the Characteristics of AtmospheNic Electric Field. In Proceedings of the 2016 33rd International Conference on Lightning Protection (Iclp), Estoril, Portugal, 25–30 September 2016.
9. Aranguren, D.; Montanya, J.; Sola, G.; March, V.; Romero, D.; Torres, H. On the Lightning Hazard Warning Using Electrostatic Field: Analysis of Summer Thunderstorms in Spain. *J. Electrost.* **2009**, *67*, 507–512. [CrossRef]
10. Yan, C. Predict Lightning Location and Movement with Atmospherical Electrical Field Instrument. In Proceedings of the 2019 IEEE 10th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), Vancouver, BC, Canada, 17–19 October 2019; pp. 535–537.

11. Cummins, K.L.; Murphy, M.J. An Overview of Lightning Locating Systems: History, Techniques, and Data Uses, with an In-Depth Look at the US NLDN. *IEEE Trans. Electromagn. Compat.* **2009**, *51*, 499–518. [CrossRef]
12. de Abreu, L.P.; Goncalves, W.A.; Mattos, E.V.; Albrecht, R.I. Assessment of the Total Lightning Flash Rate Density (FRD) in Northeast Brazil (NEB) Based on TRMM Orbital Data from 1998 to 2013. *Int. J. Appl. Earth Obs. Geoinf.* **2020**, *93*, 102195. [CrossRef]
13. Tkachev, I.D.; Vasilyev, R.V.; Belousova, E.P. Cluster Analysis of Lightning Discharges: Based on Vereya-Mr Network Data. *Sol.-Terr. Phys.* **2021**, *7*, 85–92. [CrossRef]
14. Kohn, M.; Galanti, E.; Price, C.; Lagouvardos, K.; Kotroni, V. Nowcasting Thunderstorms in the Mediterranean Region Using Lightning Data. *Atmos. Res.* **2011**, *100*, 489–502. [CrossRef]
15. Zeng, Q.; Wang, Z.; Guo, F.; Feng, M.; Zhou, S.; Wang, H.; Xu, D. The Application of Lightning Forecasting Based on Surface Electrostatic Field Observations and Radar Data. *J. Electrost.* **2013**, *71*, 6–13. [CrossRef]
16. Voormansik, T.; Rossi, P.J.; Moisseev, D.; Tanilsoo, T.; Post, P. Thunderstorm Hail and Lightning Detection Parameters Based on Dual-Polarization Doppler Weather Radar Data. *Meteorol. Appl.* **2017**, *24*, 521–530. [CrossRef]
17. Buechler, D.E.; Goodman, S.J. Echo Size and Asymmetry—Impact on NEXRAD Storm Identification. *J. Appl. Meteorol.* **1990**, *29*, 962–969. [CrossRef]
18. Feng, G.; Qie, X.; Wang, J.; Gong, D. Lightning and Doppler Radar Observations of a Squall Line System. *Atmos. Res.* **2009**, *91*, 466–478. [CrossRef]
19. Rinehart, R.E.; Garvey, E.T. Three-Dimensional Storm Motion Detection by Conventional Weather Radar. *Nature* **1978**, *273*, 287–289. [CrossRef]
20. Dixon, M.; Wiener, G. Titan—Thunderstorm Identification, Tracking, Analysis, and Nowcasting—A Radar-Based Methodology. *J. Atmos. Ocean. Technol.* **1993**, *10*, 785–797. [CrossRef]
21. Bi, K.; Xie, L.; Zhang, H.; Chen, X.; Gu, X.; Tian, Q. Accurate Medium-Range Global Weather Forecasting with 3D Neural Networks. *Nature* **2023**, *619*, 533. [CrossRef]
22. Karagiannidis, A.; Lagouvardos, K.; Lykoudis, S.; Kotroni, V.; Giannaros, T.; Betz, H.-D. Modeling Lightning Density Using Cloud Top Parameters. *Atmos. Res.* **2019**, *222*, 163–171. [CrossRef]
23. Zhang, X.; Yin, Y.; Kukulies, J.; Li, Y.; Kuang, X.; He, C.; Lapierre, J.L.; Jiang, D.; Chen, J. Revisiting Lightning Activity and Parameterization Using Geostationary Satellite Observations. *Remote Sens.* **2021**, *13*, 3866. [CrossRef]
24. Goswami, B.B.; Mukhopadhyay, P.; Mahanta, R.; Goswami, B. Multiscale Interaction with Topography and Extreme Rainfall Events in the Northeast Indian Region. *J. Geophys. Res. Atmos.* **2010**, *115*, D12114. [CrossRef]
25. Pawar, S.; Lal, D.; Murugavel, P. Lightning Characteristics over Central India during Indian Summer Monsoon. *Atmos. Res.* **2012**, *106*, 44–49. [CrossRef]
26. Galanaki, E.; Kotroni, V.; Lagouvardos, K.; Argiriou, A. A Ten-Year Analysis of Cloud-to-Ground Lightning Activity over the Eastern Mediterranean Region. *Atmos. Res.* **2015**, *166*, 213–222. [CrossRef]
27. Leinonen, J.; Hamann, U.; Germann, U. Seamless Lightning Nowcasting with Recurrent-Convolutional Deep Learning. *Artif. Intell. Earth Syst.* **2022**, *1*, e220043. [CrossRef]
28. McCaul, E.W.; Goodman, S.J.; LaCasse, K.M.; Cecil, D.J. Forecasting Lightning Threat Using Cloud-Resolving Model Simulations. *Weather Forecast.* **2009**, *24*, 709–729. [CrossRef]
29. Moon, S.-H.; Kim, Y.-H. Forecasting Lightning around the Korean Peninsula by Postprocessing ECMWF Data Using SVMs and Undersampling. *Atmos. Res.* **2020**, *243*, 105026. [CrossRef]
30. Lin, T.; Li, Q.; Geng, Y.-A.; Jiang, L.; Xu, L.; Zheng, D.; Yao, W.; Lyu, W.; Zhang, Y. Attention-Based Dual-Source Spatiotemporal Neural Network for Lightning Forecast. *IEEE Access* **2019**, *7*, 158296–158307. [CrossRef]
31. Wang, Y.; Yang, Y.; Jin, S. Evaluation of Lightning Forecasting Based on One Lightning Parameterization Scheme and Two Diagnostic Methods. *Atmosphere* **2018**, *9*, 99. [CrossRef]
32. Chen, J.; Zhao, C.; Gu, S.; Xiang, N.; Wang, Y.; Li, M. Present Status and Development Trend of Lightning Detection and Protection Technology of Power Grid in China. *High Volt. Eng.* **2016**, *42*, 3361–3375.
33. Bao, J.; Wang, X.; Zhang, F.; Huang, X.; Sun, P.; Li, Z. A Data-Driven Early Warning Method for Continuous Lightning Trip-out of Compact Power Corridors Based on Lightning Location Data. *Power Syst. Technol.* **2022**, *46*, 1194–1205.
34. Geng, Y.-A.; Li, Q.; Lin, T.; Yao, W.; Xu, L.; Zheng, D.; Zhou, X.; Zheng, L.; Lyu, W.; Zhang, Y. A Deep Learning Framework for Lightning Forecasting with Multi-Source Spatiotemporal Data. *Q. J. R. Meteorol. Soc.* **2021**, *147*, 4048–4062. [CrossRef]
35. Mostajabi, A.; Finney, D.L.; Rubinstein, M.; Rachidi, F. Nowcasting Lightning Occurrence from Commonly Available Meteorological Parameters Using Machine Learning Techniques. *Npj Clim. Atmos. Sci.* **2019**, *2*, 41. [CrossRef]
36. Ivanova, A. International Practices of Thunderstorm Nowcasting. *Russ. Meteorol. Hydrol.* **2019**, *44*, 756–763. [CrossRef]
37. Deng, Y.; Li, M.; Wang, Y.; He, X.; Wen, X.; Lan, L.; Ma, Y.; Pan, H. Relationship Between Lightning Activity and Terrain in High-Altitude Mountainous Areas. *IEEE Trans. Power Deliv.* **2023**, *38*, 3561–3570. [CrossRef]
38. Chen, Z.; Qie, X.; Tian, Y.; Wang, D.; Yuan, S. Assimilation of Lightning Data through Comprehensively Nudging Water Contents at the Cloud-Resolving Scale. *Acta Meteorol. Sin.* **2017**, *75*, 442–459.

39. Zeng, F.-H.; Guo, F.-X.; Lian, C.-H.; Gan, M.-J.; Li, Q.; Liu, Z. The Formation Mechanism of the Lower Positive Charge Center of Thunderstorm in the Inland Plateau of China. *Sci. Technol. Eng.* **2019**, *19*, 25–33.
40. Dai, J. Study on Thunderstorm Development and Evolution, and Its Mechanisms in the Yangtze River Delta Region. Ph.D. Thesis, Nanjing University, Nanjing, China, 2013.