Article

Improving the Detection Effect of Long-Baseline Lightning Location Networks Using PCA and Waveform Cross-Correlation Methods

Ting Zhang [1,2,3,4], Jiaquan Wang [5], Qiming Ma [5] and Liping Fu [1,3,4,*]

1   National Space Science Center, Chinese Academy of Sciences, Beijing 100190, China; zhangting221@mails.ucas.ac.cn
2   University of Chinese Academy of Sciences, Beijing 100049, China
3   Beijing Key Laboratory of Space Environment Exploration, Beijing 100190, China
4   Key Laboratory of Science and Technology on Environmental Space Situation Awareness, CAS, Beijing 100190, China
5   Institute of Electrical Engineering, Chinese Academy of Sciences, Beijing 100190, China; wangjiaquan@mail.iee.ac.cn (J.W.); maqiming@mail.iee.ac.cn (Q.M.)
*   Correspondence: fuliping@nssc.ac.cn

Abstract: Ultra-long-distance and high-precision lightning location technology is an important means to realize low-cost and wide-area lightning detection. This paper carried out research on the high-precision location technology of very-low-frequency (VLF) lightning electromagnetic pulse based on the Asia-Pacific Lightning Location Network (APLLN) deployed in 2018. Two key technologies are proposed in this paper: one is the calculation method of signal arrival time using very-low-frequency lightning electromagnetic pulse waveform, and the other is the compression transmission technology of lightning electromagnetic pulse waveform based on a signal principal component analysis. The results of a comparison and evaluation of the improved APLLN with the ADTD system show that the APLLN has a relative location efficiency of 69.1% and an average location error within the network of 4.5 km.

1. Introduction

Thunderstorms or lightning are common severe convective weather phenomena in nature, with an average of tens to hundreds of lightning flashes per second worldwide, including cloud and cloud-to-ground flashes. Accurate and efficient detection of lightning event information can help people to obtain information about thunderstorms, avoid lightning accidents, and reduce economic losses.

Lightning emits a significant number of electromagnetic pulse signals. The signals cover a wide frequency range, from a few hertz (long continuing currents) to 10^{20} Hz (hard X-rays) [1]. Measuring electromagnetic pulses is the fastest and most effective way to detect lightning events. Considering the propagation attenuation of electromagnetic signals in the Earth–ionosphere waveguide, ground-based lightning location systems generally operate in the extremely-low- to very-high-frequency band. A ground-based lightning location network usually consists of at least four electromagnetic pulse detection sites, a data processing center, and a network transmission link. The distance between the two stations is called the baseline. Baseline sizes typically range from tens to thousands of kilometers depending on the detection range. The lightning location network uses the time difference and direction parameters of the signal to locate each event.

Very-high-frequency lightning electromagnetic pulse signals carry richer information, and the steep peaks of electromagnetic pulses are easily detected by sensors, thus achieving precise lightning imaging [2–6]. However, the propagation distance of high-frequency

LEMP signals is short, and the amplitude rapidly decreases after a long distance. The effective detection range of lightning electromagnetic pulse sensors working in high-frequency bands is usually within several dozen kilometers.

At present, the regional commercial lightning location systems are working in the VLF and LF bands, such as the U.S. National Lightning Detection Network (NLDN), European Cooperation for Lightning Detection (EUCLID), and China Advanced Direction-time Lightning Detection (ADTD). These systems are usually composed of hundreds of lightning detection sites; the baseline of the sites is generally about 100 km, which can be better compatible with the detection efficiency of cloud-to-ground (CG) lightning and intracloud (IC) lightning, and the location accuracy is better than 500 m. The signals in the VLF band could propagate long distances in the Earth–ionosphere waveguide, and long-range lightning detection can be realized by using VLF signals. Considering the long-distance propagation of VLF signals in Earth–ionosphere waveguides [7–11], more and more studies have begun to focus on using VLF lightning electromagnetic pulse signals to achieve low-cost, large-area, and high-precision lightning detection.

The World Wide Lightning Location Network (WWLLN) [12–14] currently operates more than 70 detection sites in the frequency band of 3 kHz to 30 kHz. WWLLN uses lightning electromagnetic pulse signals detected by at least five sites for localization calculation. The lightning strokes could be located by calculating the time difference between the signals arriving at each site based on the time of the group arrival method. Recent research indicates that the detection efficiency of WWLLN for strokes with about 30 kA is approximately 30% globally.

The Institute of Electrical Engineering of the Chinese Academy of Sciences has deployed the Asia-Pacific Lightning Location Network (APLLN), which operates in the VLF band. The average distance between the detection sites is about 1000 km. Wang et al. [15] proposed a calculation method of signal arrival time based on envelope peak, and the location accuracy in the network is around 10 km.
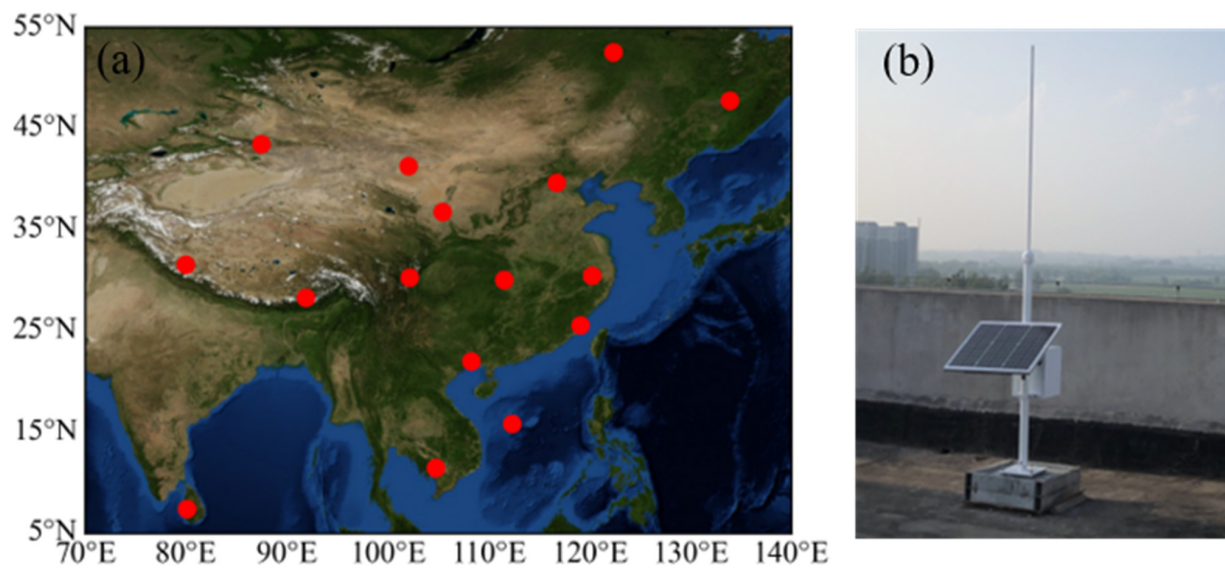
Wang et al. [16] adopted the time-difference location method to realize lightning location. The network boasts a baseline distance of approximately 2000 km and operates in the VLF band. The Gauss–Newton iteration method was utilized to find the optimal fitting point of the arrival time difference, thereby enhancing both the efficiency and accuracy of the location algorithm. Comparisons with lightning location results in the Jiangsu area of the China Power Grid demonstrate a detection accuracy that can reach 12 km.

Overall, the baseline of the VLF lightning location systems is about 1000 km, and the location accuracy is about 10 km. Compared with VLF/LF lightning location system whose location accuracy is better than 500 m, the location accuracy of 10 km is not ideal. To enhance the detection performance of the long-baseline lightning location system, this paper presents an improved localization method based on the lightning electromagnetic pulse waveform.

The rest of this article is arranged as follows: Section 2 provides a brief overview of the APLLN and its components. Section 3 introduces the data processing and localization methods used in the study. Section 4 describes the location results. Section 5 compares and evaluates the results of the new method. Finally, Section 6 summarizes the article.
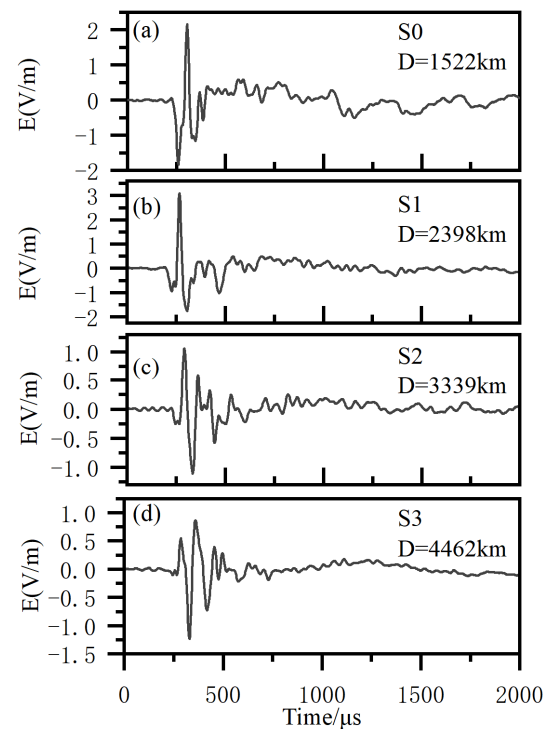
2. Network and Data

Since 2018, the APLLN [15] has established 16 long-baseline LEMP detection sites in China and neighboring countries, which have been in stable operation for five years. The detection sites are distributed between 75 and 134°E and 7 and 53°N, with a baseline range of 800~1500 km. The deployment of the APLLN detection site and the appearance of the electromagnetic pulse detection equipment are shown in Figure 1. In the case of at least four sites detecting signals simultaneously, the real-time two-dimensional location of lightning activity in China and the Asia–Pacific region can be realized based on the time difference of arrival (TDOA).

Figure 1. (**a**) The deployment of the APLLN detection site; the red dots indicate the locations of the sites. (**b**) The appearance of the electromagnetic pulse detection equipment.

Each detection site of the APLLN is mainly composed of a whiplash electric field antenna, signal amplifier, filter, analog and digital conversion circuit, signal processor, and high-precision GPS/Beidou timing module. The detection equipment adopts the trigger sampling method to record the LEMP signal with a duration of 2 ms. The sampling rate of the device is 500KSPS, which fully meets the sampling requirements of signals in the 3–30 kHz frequency band according to the Nyquist sampling theorem. Figure 2 shows the LEMP waveform data recorded by a device. One single records 1000 sampling points, and each sampling point is stored with 4-byte floating-point data, so the memory needs at least 4K bytes of storage space to record a single sample of data.
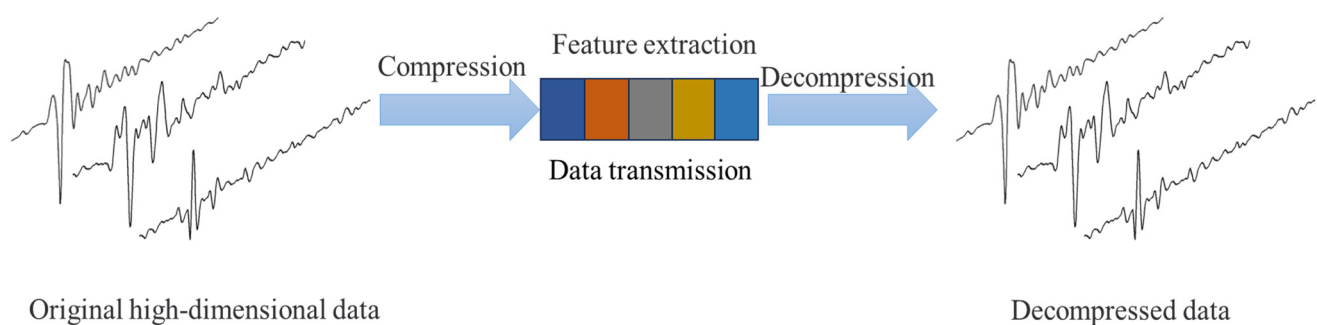

Figure 2. (**a**–**d**) The LEMP waveform data from different distance.

3. Method

3.1. Real-Time Data Compression

Before describing the lightning location method, it is necessary to introduce the compression technology of LEMP waveform data. In recent years, the research team has upgraded the data processing system of APLLN. The detection system can work at a lower trigger threshold. A lower trigger threshold would result in the detection system capturing weaker LEMP signals, thus improving the lightning detection efficiency. However, this also brings new challenges, as data transmission requires larger bandwidth resources.

When processing large amounts of LEMP waveform data, data compression is an effective way to save transmission bandwidth and improve transmission efficiency [17]. The conventional lossless compression method has the disadvantages of slow computation speeds and low compression rates, which cannot meet the real-time LEMP data processing. This paper presents a method of real-time lossy compression of LEMP waveform data based on principal component analysis (PCA) [18,19] and transmits the compressed LEMP waveform data back to the data processing center in real-time. As far as we know, this is the first time that PCA method has been applied in the field of LEMP waveform compression transmission. The data processing center decompresses the data and performs subsequent data processing and location calculation. The process diagram of data compression and decompression using the PCA method is shown in Figure 3.

Figure 3. Data compression transmission flow diagram.

The main idea of the PCA method is to find the most significant components in the data and replace the original data with the most significant components to achieve the purpose of data dimensionality reduction [19]. To minimize the loss of data information, we removed the correlation of the data and preprocessed the data in a lower-dimensional space in the case of high-dimensional space.

For an LEMP waveform data set x of length n, the mean value of x is:

$$\overline{x} = \frac{1}{n}\sum\nolimits_{i=1}^{n} x_i \tag{1}$$

Decentralization of LEMP waveform data:

$$d_i = x_i - \overline{x} \tag{2}$$

The covariance matrix of the LEMP waveform data sample is:

$$C_{n*n} = \frac{1}{n-1}\sum\nolimits_{i=1}^{n} d_i d_i^{T} \tag{3}$$

Using the covariance matrix of the LEMP waveform data, we obtained the eigenvalues and eigenvectors. Then, we sorted the eigenvalues of the LEMP waveform data samples from large to small and selected the top K values [20]. Then, the corresponding K eigen-
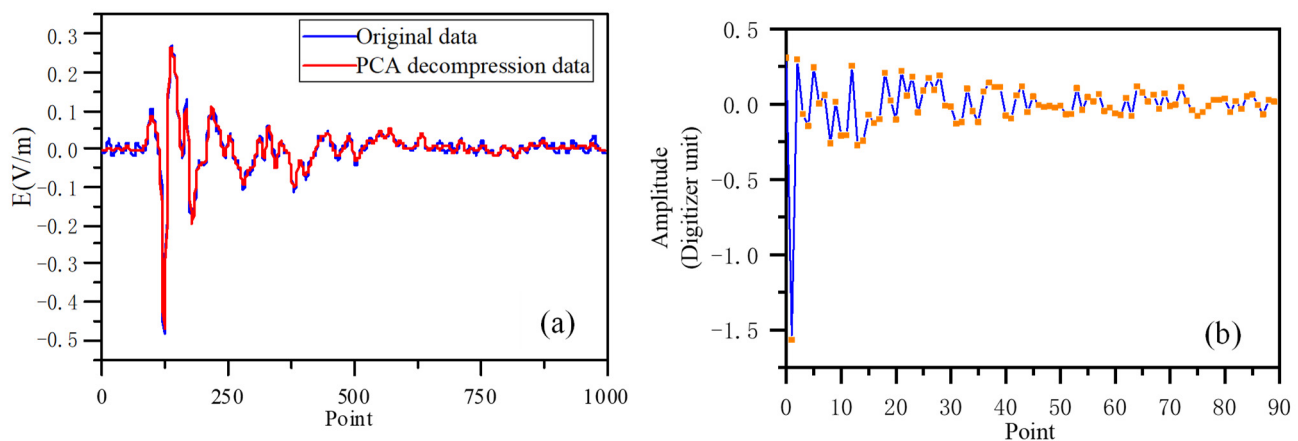
vectors are formed into the eigenvector matrix P_K in the form of row vectors. With the eigenvector set to P_K, LEMP waveform data x can be projected into the eigenspace y_K:

$$y_K = P_K{}^T(x - \bar{x}) \tag{4}$$

The LEMP waveform data x are compressed, transmitted back, and finally decompressed. The decompressed LEMP waveform data can be expressed as:
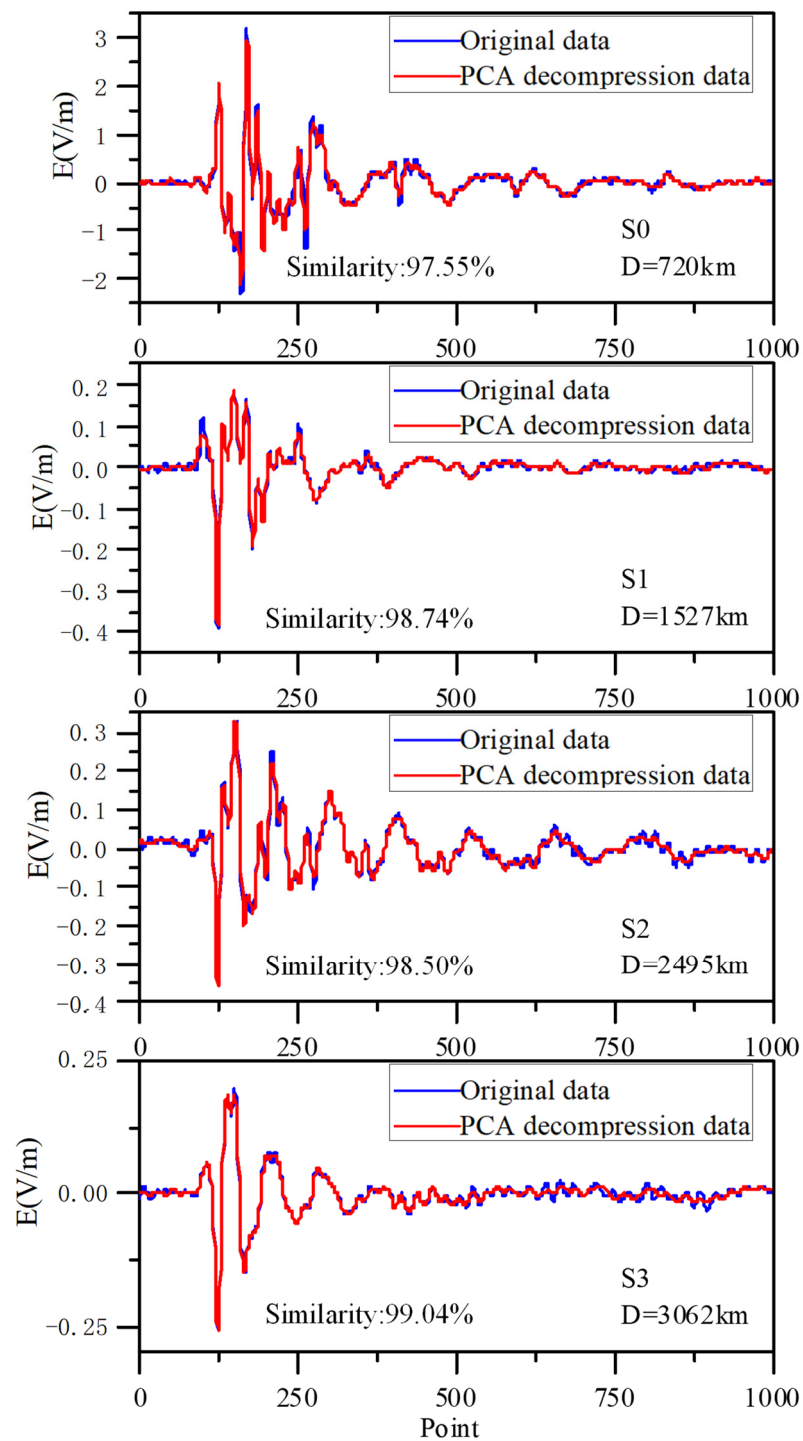
$$x = \bar{x} + \sum_{K=1}^{n} y_K P_K \tag{5}$$

Taking an example of the LEMP waveform data recorded by the APLLN detector as a reference, the blue line in Figure 4a represents the waveform of the original data, and the red line represents the waveform data obtained after PCA compression and decompression. The dimension of the LEMP waveform data is 1000. Through experiments, it was found that when the dimensionality of the data was reduced to 90, i.e., compressed from the original 1000 sampling points to 90 feature points, the correlation between the transformed data and the original data was 98.87%. If only 90 feature points need to be transmitted through the network instead of all 1000 sampling points, it can fully meet the real-time transmission and processing of LEMP data. Figure 4b shows the results of compressing the original data using the PCA method. It can be seen from Figure 4 that the original data and the extracted data are generally consistent in the time domain.

Figure 4. (**a**) The blue line represents the waveform of the original data, and the red line represents the waveform data obtained after PCA compression and decompression. (**b**) The feature data compressed using PCA method, and each colored point represents a feature.

In practical applications, P_K and $\bar{x}$ are fitted from a large amount of historical data and stored in the memory of detection devices and data processing centers, respectively, for the compression and decompression of LEMP waveform data. Considering the presence of a large amount of electromagnetic noise during the propagation process, as well as factors such as propagation distance and terrain that can affect the waveform, we analyzed the LEMP data received from multiple detection sites. All LEMP signals were generated by the same lightning event. Figure 5 below shows a set of homologous LEMP signals collected from four different sites, with propagation distances ranging from 720 km to 3062 km. It can be seen that the PCA conversion results can effectively restore the original LEMP data. Using the method of signal cross-correlation to calculate the similarity of signals, the average similarity is 98.46%. This indicates that the PCA method can effectively restore the principal components of the original LEMP data without causing significant information loss due to electromagnetic noise.

Figure 5. Original data of LEMP waveform with different propagation distance and PCA conversion data. The data are all from the same lightning event and were recorded by different detection stations. Each subplot is labeled with the propagation distance of the LEMP signal.

This paper compares the time consumption and compression ratio of bzip2 [21,22] lossless compression and PCA lossy compression for different amounts of data. It can be seen from Table 1 that the compression rate of PCA lossy compression is much higher than that of bzip2 lossless compression, which greatly reduces the bandwidth occupation of the channel during data transmission. The time consumed by PCA lossy compression is also much less than that consumed by bzip2 lossy compression, and using PCA lossy compression can improve the efficiency of data transmission.

Table 1. Comparison of bzip2 lossless compression and PCA lossy compression.

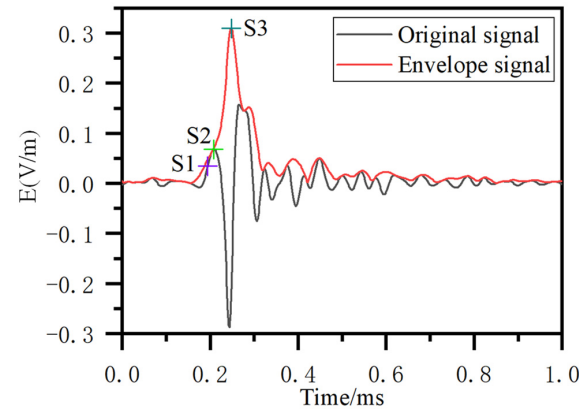| Data Number	bzip2 Lossless Compression		PCA Lossy Compression	
	Consumption of Time (s)	Compressibility (%)	Consumption of Time (s)	Compressibility (%)
100	0.84	0.63	0.0882	0.09
200	1.83	0.63	0.1764	0.09
300	2.75	0.66	0.2646	0.09
400	3.76	0.63	0.3528	0.09
500	4.48	0.60	0.441	0.09
600	5.57	0.63	0.5292	0.09
700	6.49	0.64	0.6174	0.09
800	8.55	0.72	0.7056	0.09

3.2. Location Algorithm

Based on the efficient transmission method of LEMP data explained in Section 3.1, this section mainly discusses the second key issue of VLF lightning detection: how to improve the location accuracy of VLF lightning detection? In general, the mainstream lightning location systems use the TDOA method for location [23–25]. The TDOA method locates the radiation source according to the time difference between the LEMP signals arriving at each detection site. The number of sites ranges from four to more, typically. The long-baseline lightning location system has a large baseline over 1000 km, which is different from the short- and medium-baseline location system with the baseline between 30 and 150 km. The LEMP signals received by the long-baseline lightning location system have a large attenuation of ground wave propagation due to the long distance between the signal source and the detection site. Then, the signal received by the detection site could not calculate the arrival time of the ground wave accurately, which leads to the poor location accuracy of the long-baseline lightning detection system.

In the past few years, APLLN has used a localization method that calculates the envelope of LEMP using the Hilbert transform and then calculates the arrival time of the signal based on the peaks of the envelope [15]. At the same time, the location result of lightning was calculated by the TDOA method, and an improved Levenberg–Marquardt nonlinear least squares iterative algorithm was used to optimize the lightning location results. This location algorithm is collectively referred to as the envelope peak method in the following content.
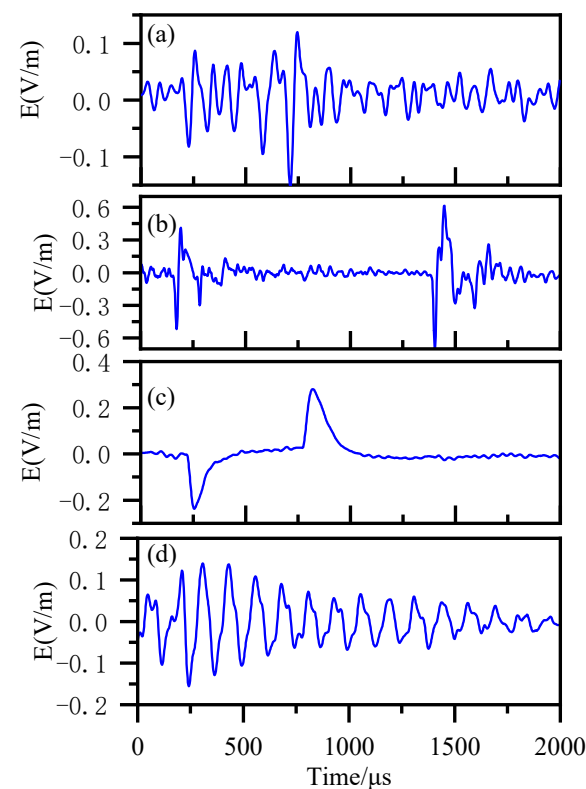
However, the arrival time of the LEMP calculated by the envelope peak method is biased with respect to the true value. As shown in Figure 6, the first pulse peak S_2 of the gray waveform can be regarded as the arrival time of the LEMP signal. The envelope peak S_3 of the red waveform is the signal arrival time determined by the envelope peak method. S_1 is the theoretical arrival time of the LEMP signal, which is the actual time it takes for the signal to travel from lightning to the detection station. From Figure 6, it can be seen that the arrival time of the real signal is earlier than the time of the envelope peak. When the distance between the detection site and the lightning is nearer, such as less than 100 km, S_1 is approximately S_2. As the signal propagation distance increases, S_2 gradually lags behind S_1.

Moreover, the anti-noise ability of the envelope peak method is weak. Figure 7 lists several typical electromagnetic pulse signals that may affect the accuracy of VLF lightning location. The duration of an LEMP signal ranges from tens of microseconds to hundreds of microseconds. Therefore, we generally assumed that there is only one LEMP signal within a sampling window. As shown in Figure 2, the collected signals only have one LEMP signal, which is the case handled by most calculation processes. However, in practical applications, signals such as those in Figure 7a,b may also be encountered, where Figure 7a is a dense lightning electromagnetic pulse signal emitted by cloud flash radiation. Figure 7b shows the presence of multiple LEMP signals within a sampling window. Although both cases collect LEMP signals, the use of signal's envelope or peak methods cannot accurately determine lightning homology, resulting in incorrect location results. Figure 7c,d show two common

electromagnetic pulse interference signals, which are usually similar to LEMP and cannot be accurately removed. At the same time, due to the large amount of interference signals generated in a short period of time, it seriously affects the location calculation. When using signal's envelope or peak methods, it can also produce incorrect location results.

Figure 6. Gray line represents the original signal waveform, and red line represents the envelope waveform after the Hilbert transformation.

Figure 7. Typical waveform of LEMP and interference signals. (**a**) Electromagnetic pulse signals from cloud flash. (**b**) Two LEMPs within the same sampling window. (**c**) Pulse interference signal from switching power supply. (**d**) Radio interference signal.

In order to improve the location capability of the APLLN, the arrival time of LEMP was calculated using the cross-correlation method [26–29] of waveforms in this paper. The waveform is used to find homologous events, remove noise interference, and reduce the influence of noise signals on location. Previously, very-low-frequency lightning positioning systems were limited by data transmission rates and typically calculated the arrival time at the detection equipment end. Due to the arrival time ambiguity caused by the long-distance

propagation of signals in the Earth–ionosphere waveguide, this method cannot accurately calculate the arrival time of signals. Therefore, after using the PCA method to solve the real-time compression and transmission of signals, the signal cross-correlation method was applied in the field of very-low-frequency lightning detection. This location algorithm is collectively referred to as the cross-correlation method in the following content.

The APLLN obtains the real-time high-precision time through the GPS/Beidou satellite timing system and marks the trigger time of the original waveform data. The detected data less than the set threshold are defined as homologous data. The detailed steps of the improved signal arrival time algorithm are as follows:

For a set of homologous data sequences $\{x_1, x_2, x_3, \ldots x_n\}$, the earliest time LEMP waveform data x_1 were selected as the reference data. The time t_{x1} of the pulse peak of LEMP waveform data x_1 was taken as the reference time.
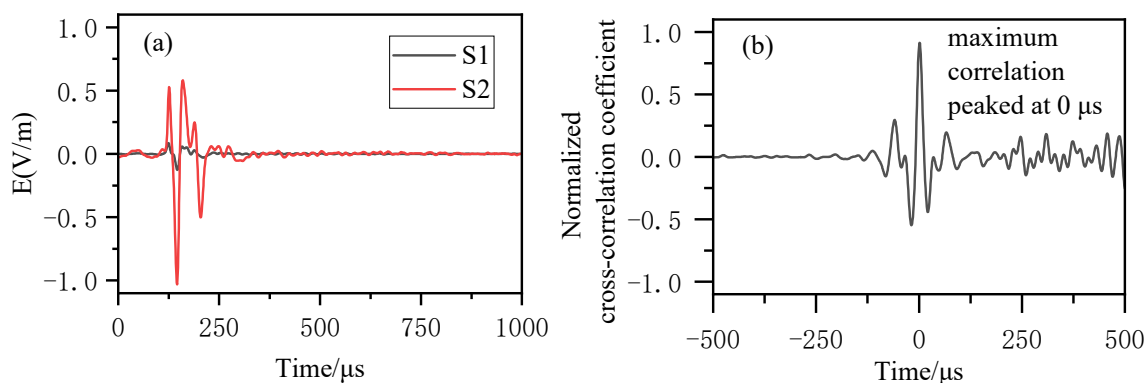
The dimension of the LEMP waveform data is 1000, which is expressed as l. We extended the dimension of x_1 according to Formula (6).

$$x'_1 = \{m, x_1, m\} \tag{6}$$

The dimension of m is 500, and the values are zeros. We calculated the value of the correlation between $x'_1[a : l + a]$ and x_2. Then, we obtained the correlation coefficient sequence $C = \{C_1, C_2, C_3 \ldots C_a\}$, where a is the first point of the data x'_1, $a \in [1, 2l]$.

The index value corresponding to C_{max} with the largest correlation coefficient can be used to calculate the arrival time t_{x2} of the electromagnetic pulse waveform data x_2. By traversing the homologous data sequence $\{x_1, x_2, x_3, \ldots x_n\}$, the time of each waveform arriving at the detection site can be found.

Figure 8a shows the electromagnetic pulse waveform data received by the detection sites S1 and S2. According to the above method, the LEMP waveform data received by the detection sites S1 and S2 were cross-correlated, and the results were obtained as shown in Figure 8b. It can be seen from Figure 8b that at time 0, the maximum correlation coefficient is obtained when the two waveforms are not translated [28].

Figure 8. (**a**) The LEMP signals from different detection sites. (**b**) Cross-correlation results between the LEMP signals from two different detection sites.

3.3. Simulated Analysis

The simulation analysis is mainly used to evaluate the performance of the location algorithm and lightning location network. This paper used Monte Carlo simulation method [30–32]. The geographical range of the simulation is 10°S~70°S, 40°E~165°E. The simulation area was divided into grids with the size of 0.1° × 0.1°. We supposed there are 1000 lightning events in each grid. We calculated the arrival time t_{si} of the electromagnetic pulse signal radiated by the lightning source to each detection site. For the
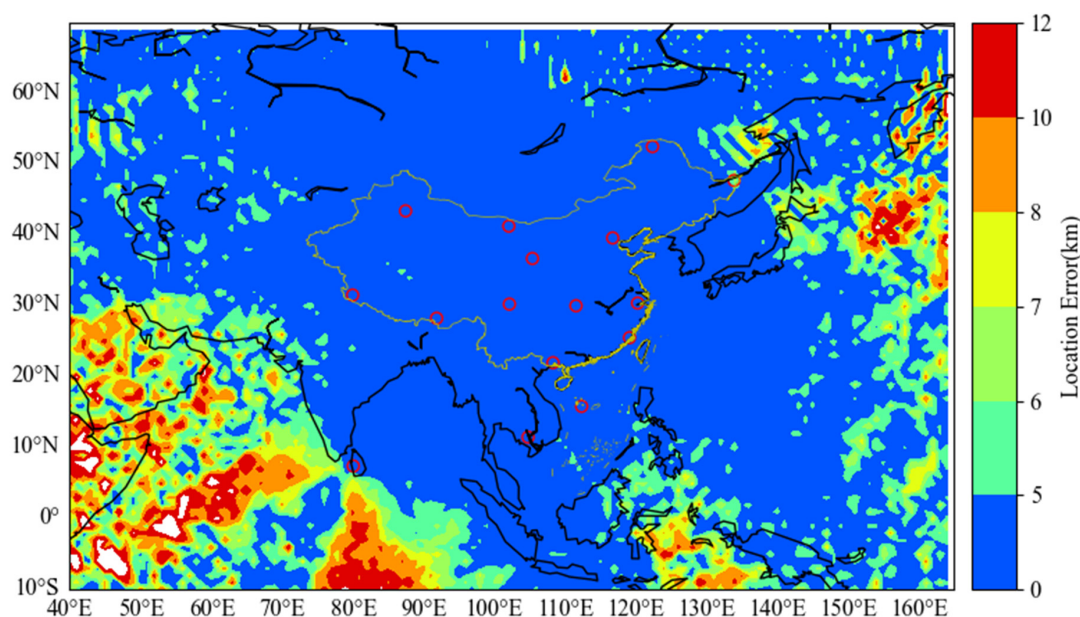
convenience of calculation, we calculated the arrival time of the ground wave signal as t_{si}, which is

$$t_{si} = t_0 + \frac{1}{c}D_{si} \tag{7}$$

t_0 is the assumed time of lightning occurrence; D_{si} is the great circle distance from the lightning to the ith detection site, and c is the speed of light. Considering the calculation error caused by signal propagation factor in real environment, random error was added to the arrival time t_{si} in the simulation process, which was defined as $\Delta\varepsilon_{si}$ in this paper. For 1000 simulation calculations within each grid, we set the mean of the random error to 0 and the variance to 1. Regarding the range of error $\Delta\varepsilon_{si}$, ref. [33] used a smaller error range of -0.4~0.4 μs.

Due to the fact that the TDOA calculation method ultimately assumes signal propagation along the Earth, when using envelope peak or signal peak methods to calculate signal arrival time, the position represented by the peak is generally sky waves rather than ground waves. Therefore, there is an error in the calculation of the signal arrival time. If the signal propagation distance is over a thousand kilometers, $\Delta\varepsilon_{si}$ can reach several tens of microseconds. On the other hand, the propagation of very-low-frequency LEMP signals in the Earth–ionosphere waveguide is affected by the height changes in the D region of the ionosphere, and the arrival time of sky waves will also have varying degrees of lag or lead, which mainly affects the propagation of signals at longer distances. Based on these factors, the time error introduced by APLLN is larger, and simulation requires a larger error range, which is completely different from lightning location systems with a baseline of around 100 km. Similar to [34], this article set the simulation error range to $[-5, 5]$ μs, making the simulation results closer to the real situation.

The distance between each location result and the simulated lightning location was calculated, and the location error of lightning in each grid was obtained. Figure 9 shows the results of location error using the Monte Carlo simulation method. The results show that the location error is less than 5 km for lightning events in APLLN network by using the location method proposed in this paper. With the increase in the signal propagation distance, the location error would be larger, and the location error is about 5 thousandths of R, where R is the average distance from the lightning source to the detection site. At the same time, it can also be seen that due to the uneven deployment of detection sites, large location errors may occur in the Indian Ocean, the Pacific Ocean, and other regions.
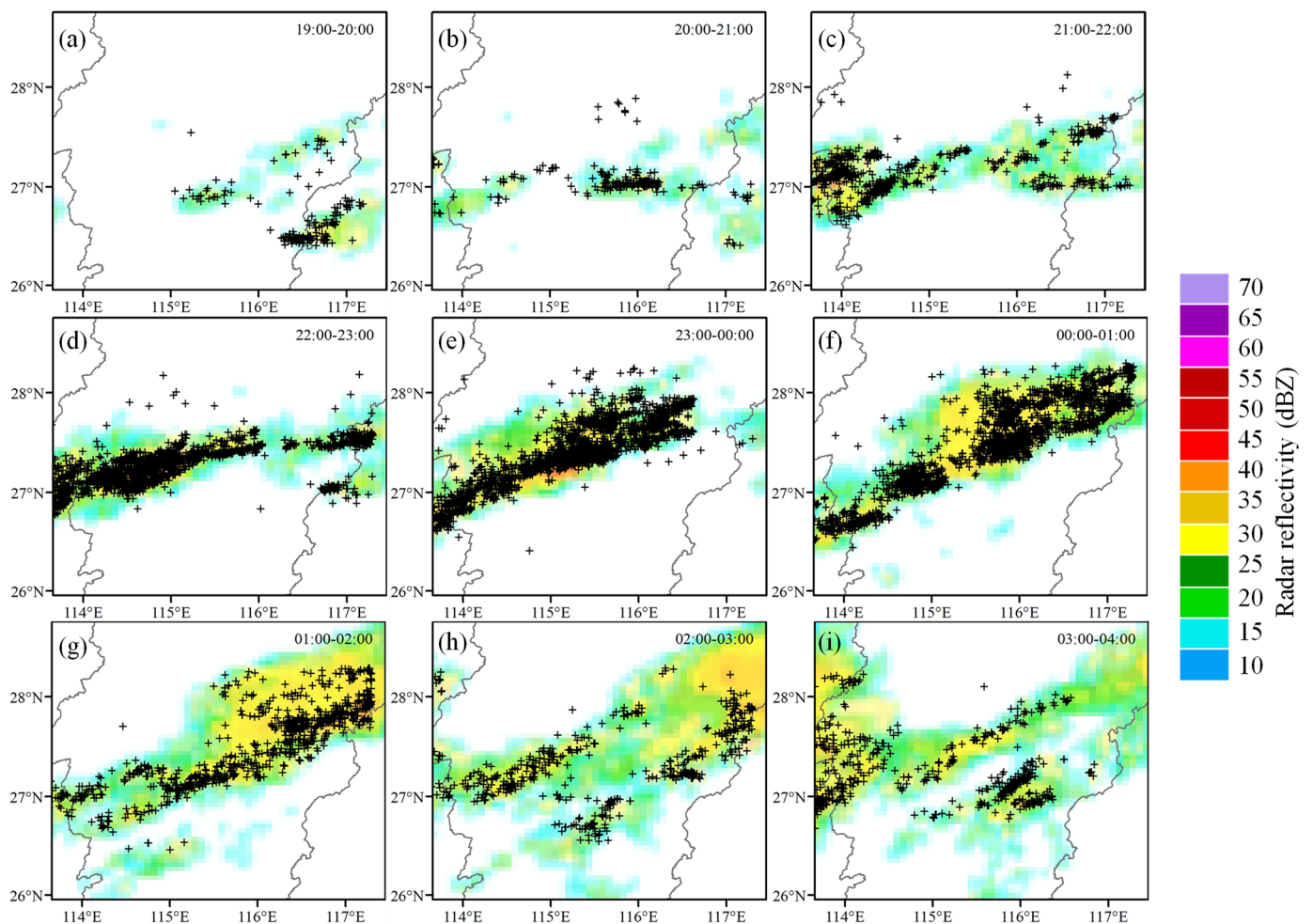

Figure 9. Simulation results of lightning location error. The red circles represent the geolocation of the detection sites.
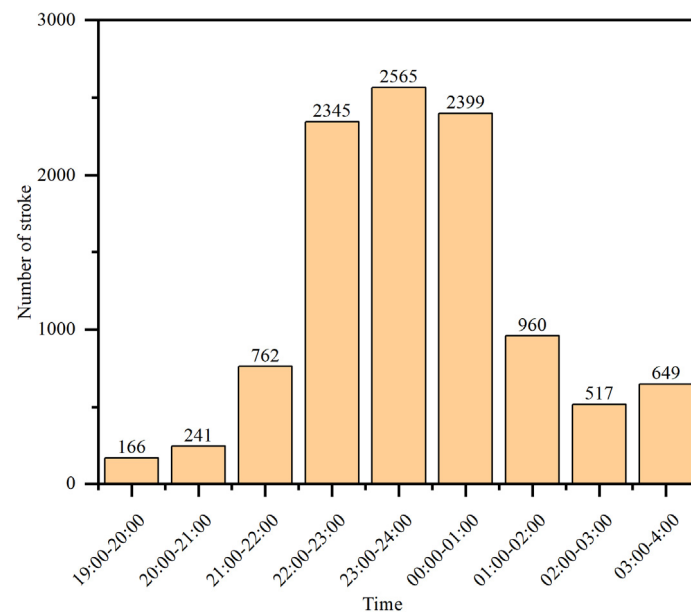
4. Results

4.1. The Location Result of Cross-Correlation Method

This article selected a thunderstorm process in southern China for data comparison and analysis. The thunderstorm process lasted for 9 h. The thunderstorm took place from 19:00 on 22 March 2023 to 4:00 on 23 March 2023, Beijing time. By analyzing the data detected by APLLN, we could visually observe the development process of thunderstorms. The process of this thunderstorm was analyzed in the range of 113.5~117.3°E and 26.4~28.3°N. The location results of the APLLN (cross-correlation method) showed that a total of 10,604 lightning strokes were detected.

As can be seen from Figure 10, the thunderstorm began at 19:00 on 22 March 2023. In the next several hours, the range of thunderstorms gradually expanded and the intensity gradually increased. The pattern displayed a westward trend, reaching its strongest within 23:00–24:00 (00:00) on 22 March 2023. Then, the thunderstorm began to weaken. The weather radar reflectivity is superimposed in Figure 10. It can be seen that the lightning location results are strongly consistent with the areas with high radar reflectivity, which verifies the accuracy of the detection results. Figure 11 shows the number of lightning events detected by the APLLN (cross-correlation method) over different periods. As can be seen from Figure 11, the maximum number of lightning strokes was 2565 at around 23:00–24:00 on 22 March 2023. The whole thunderstorm process developed rapidly and lasted a long time.

Figure 10. The location result of the APLLN by using the method proposed in this paper. Subplots (**a**–**i**) show the lightning location results for each hour, respectively. The symbol '+' indicates the location of each lightning event.
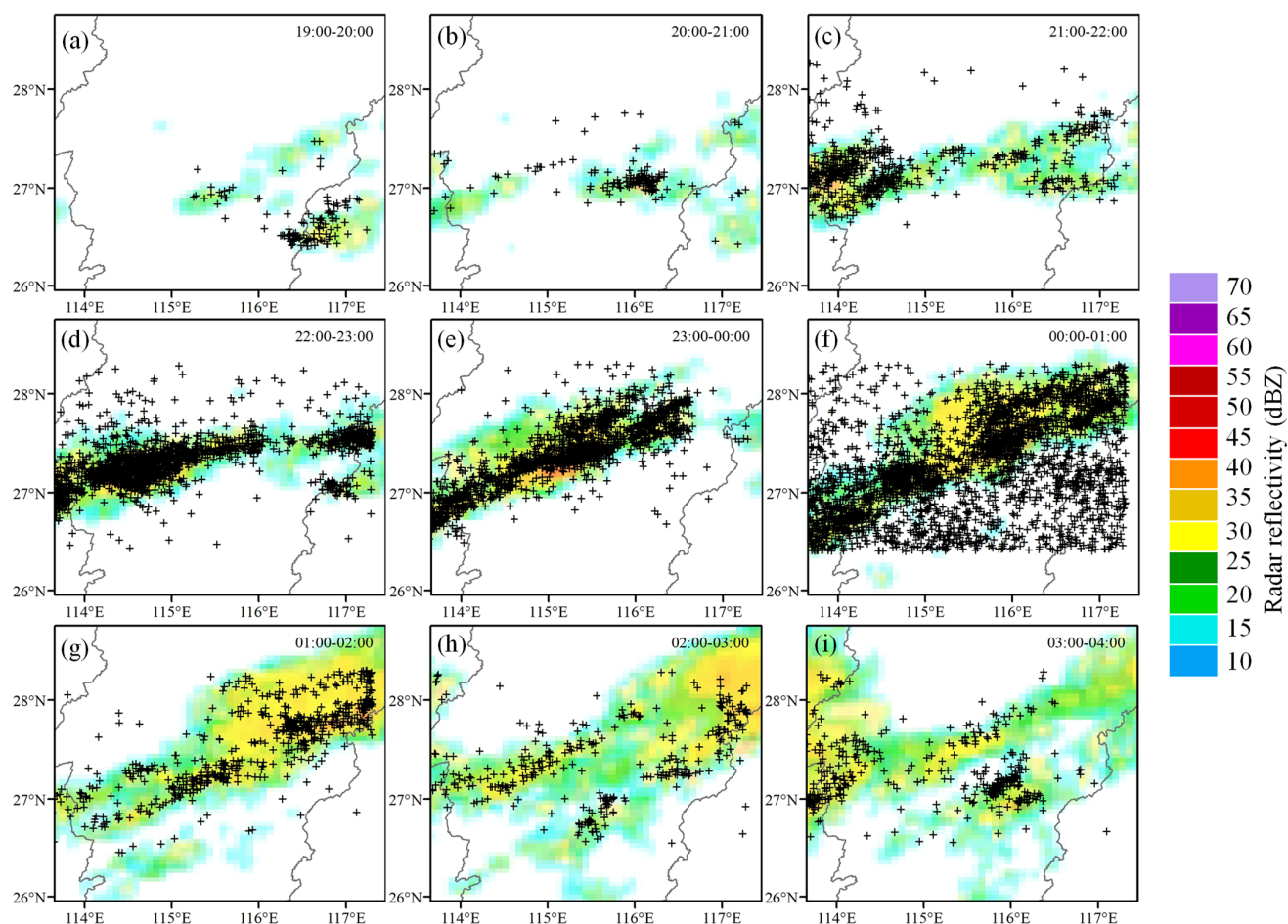
Figure 11. The number of lightning events detected by APLLN at different times using PCA and cross-correlation methods.
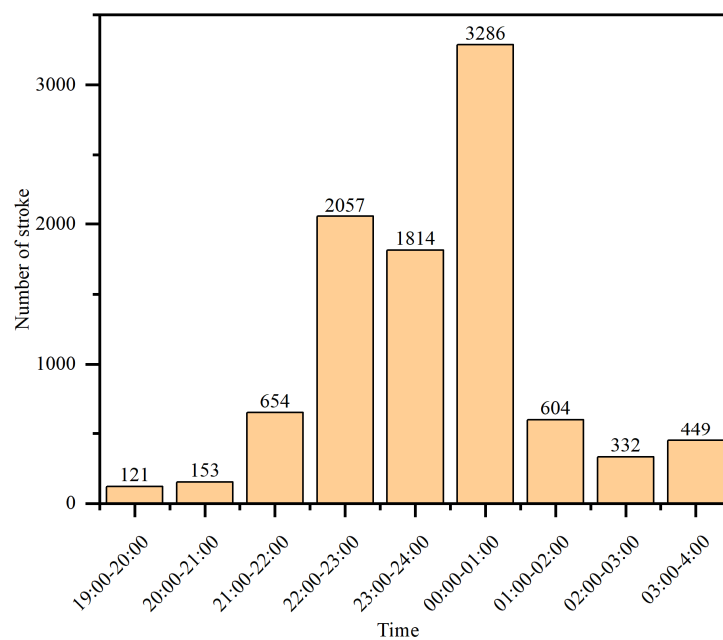
4.2. The Location Result of Envelope Peak Method

Within the same time, latitude, and longitude as described in Section 4.1 above, the location results of the APLLN (envelope peak) showed that a total of 9470 lightning strokes were detected. Figure 12 shows the lightning scatter distribution by hour. It can be seen that the development trend of thunderstorm activity monitored by the above cross-correlation method is consistent. However, from 00:00 to 01:00, it can be seen that the anti-noise ability of the envelope peak method is poor, and the monitored data are doped with a lot of noise signals.

Figure 13 shows the number of lightning events detected by the APLLN (envelope peak method) over different periods. As can be seen from Figure 13, the maximum number of lightning strokes was 3286 from 00:00 to 01:00 on 23 March 2023. The whole thunderstorm process developed rapidly and lasted a long time. Comparing Figure 10 to Figure 13, it can be observed that using the envelope peak localization method resulted in a large amount of discrete data. Among them, the most obvious period is from 00:00 to 1:00, with scattered data covering almost all the analysis areas. These scattered points did not coincide with the radar echo reflection area, which fully indicates that this is an incorrect lightning location result. In order to analyze the reasons for the incorrect results, we analyzed the incorrect lightning location data and the original LEMP data. Figure 14 shows a set of raw LEMP data participating in localization calculations. Figure 14a shows typical types of LEMP data. During thunderstorms, two or even more LEMP data points may appear within a sampling time window. If the envelope method is used for calculation, the detection site will only select the signal with the strongest amplitude for calculation. This method cannot guarantee that the data involved in localization are homologous. An incorrect location result arises when there is at least one non-homologous LEMP data point present. When using the signal cross-correlation method for localization calculation, the algorithm first selects homologous LEMP signals from the received LEMP data. Taking Figure 14 as an example, conducting a cross-correlation analysis between Figure 14a–d can exclude Figure 14a from the candidate queue. Several other typical electromagnetic interference signals are also listed in Figure 7. It should be emphasized that Figure 7c,d, which are electromagnetic interference signals, usually generate continuously. If cross-correlation calculations are not performed, the phenomenon shown in Figure 12f is inevitable.
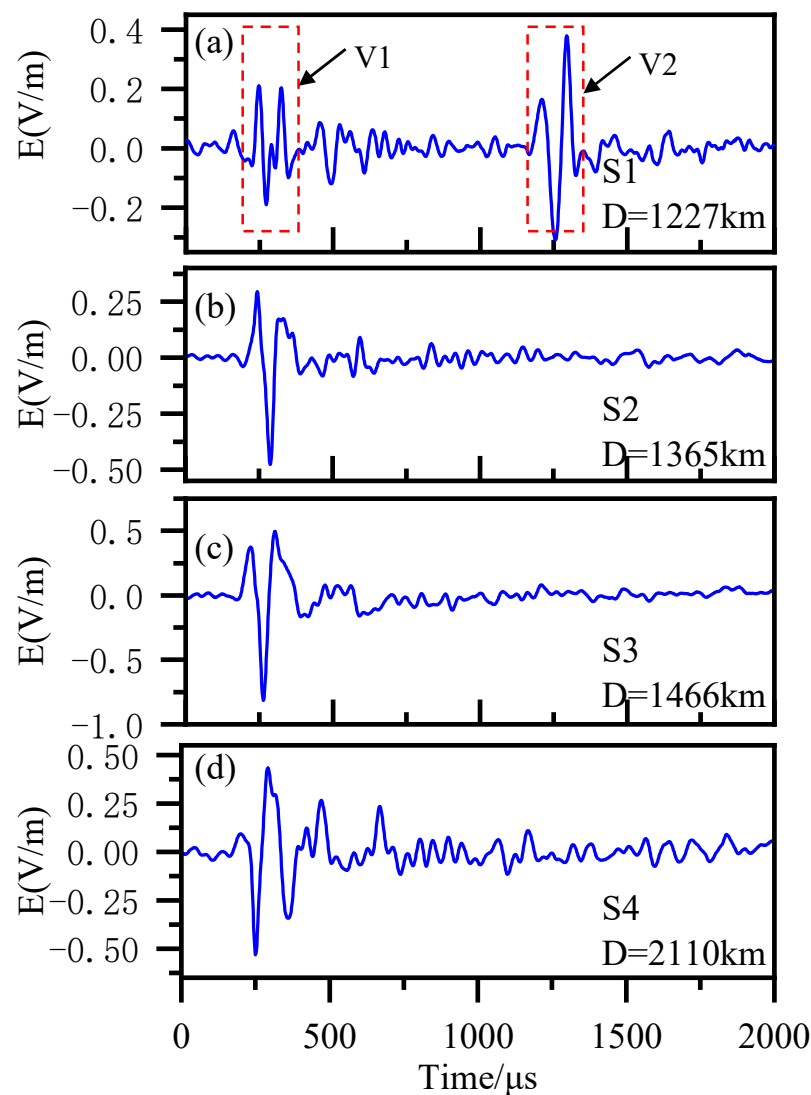
Figure 12. The location result of the APLLN by using envelope peak method. Subplots (**a**–**i**) show the lightning location results for each hour, respectively. The symbol '+' indicates the location of each lightning event.

Figure 13. The number of lightning events detected by APLLN (envelope peak method) at different times.

Figure 14. A set of LEMP waveform data for anomaly localization results. Subplots (**a**–**d**) represent the LEMP data received by each site. In subplot (**a**), two LEMPs are labeled with V1 and V2.

It is undeniable that using the cross-correlation method can improve the accuracy of the signal arrival time calculation and increase the success rate of localization. The statistical results from Figures 11 and 13 demonstrate this point well. But we also found that from 00:00 to 1:00, using the envelope peak method, the number of lightning localization was significantly higher than that using the signal cross-correlation method, which was significantly opposite to other times. One obvious factor is that there were 953 incorrect location results, accounting for 29% of the total. On the other hand, according to statistics, the envelope peak method may result in multiple localization results for the same lightning event due to its inability to accurately confirm the homology of events. These types of data account for approximately 3% of all data. Excluding the abnormal or erroneous data caused by these reasons, 2234 data points were located using the envelope peak method, which is less than the location results using the cross-correlation method. This also explains the abnormal situation from 00:00 to 1:00 very well.
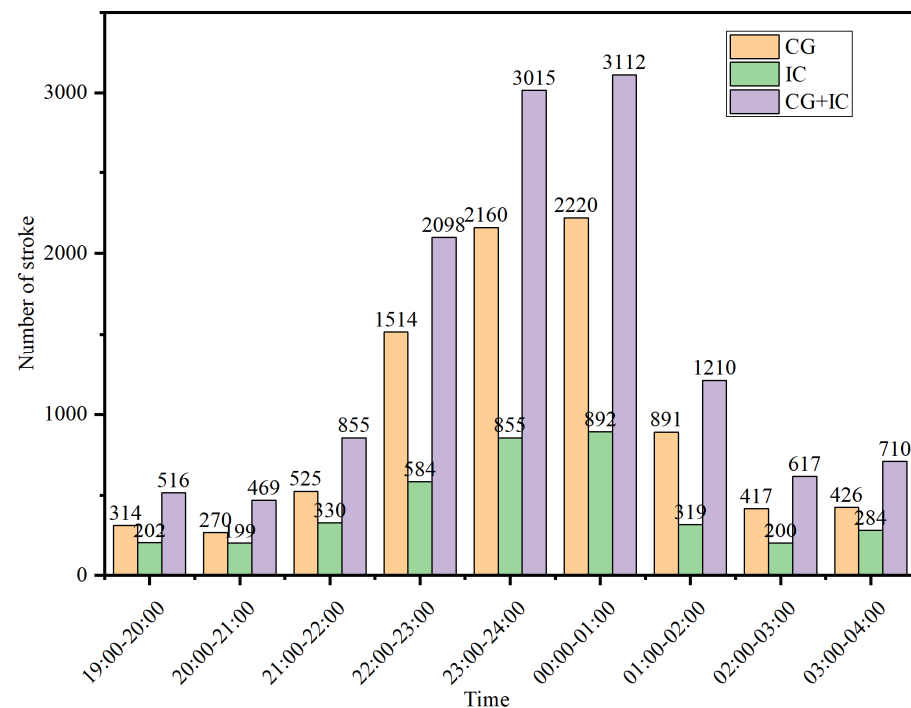
5. Discussion

5.1. Relative Detection Efficiency

In order to compare and evaluate the data, this paper also used the three-dimensional lightning location network (advanced direction-time lightning detection system, ADTD)

data of the Institute of Electrical Engineering, Chinese Academy of Sciences [35]. The ADTD is a short-baseline lightning detection network, with an average distance between stations of about 100 km. The detection range covers most of China and some countries in Southeast Asia. It has been running stably for more than ten years; the lightning location accuracy in the network is better than 500 m, and the cloud-to-ground flash detection efficiency is better than 99%.

Within the same time, latitude, and longitude as described in Section 4.1 above, the ADTD has detected a total of 12,602 lightning strokes, including 8737 CG strokes and 3865 IC strokes. CG strokes accounted for 69.3% of all detected lightning strokes. IC strokes accounted for 30.7% of all lightning. As can be seen from Figure 15, the development trend of LEMP waveform data detected by the ADTD is consistent with that detected by the APLLN.
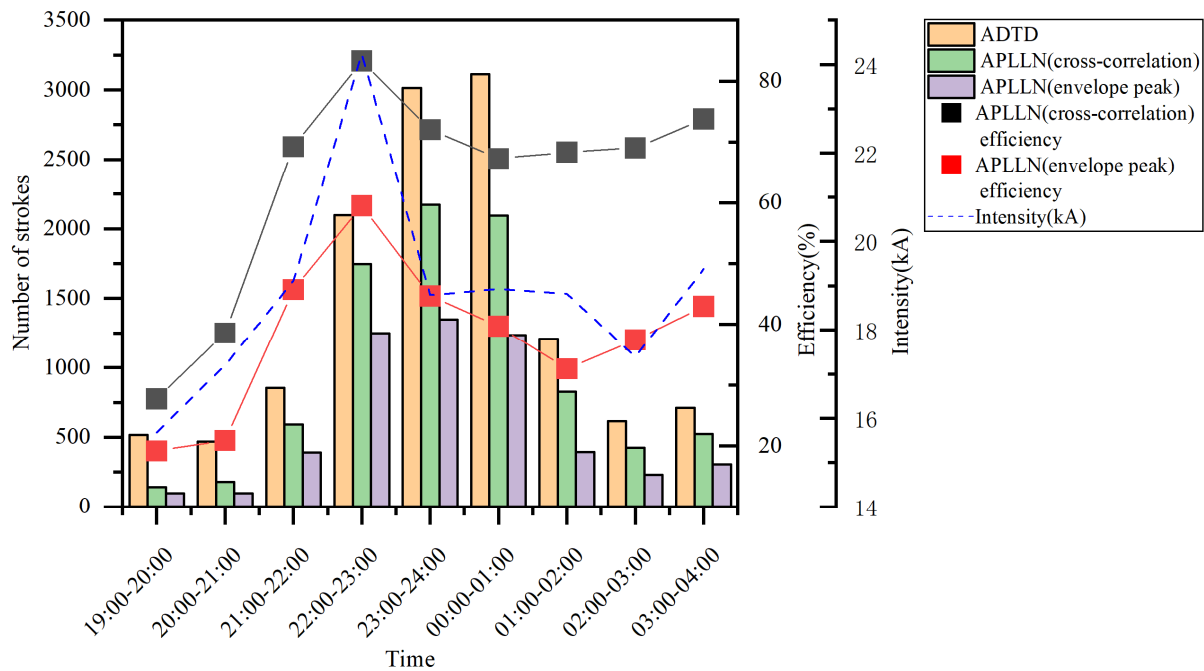

Figure 15. Distribution of LEMP waveform data detected by ADTD from 19:00 on 22 March 2023 to 4:00 on 23 March 2023. Purple means all types of lightning; orange means CG, and green means IC.

From the above discussion, it is evident that within the same latitude and longitude region at the same time, the ADTD detected a total of 12,602 lightning strokes. The APLLN (cross-correlation method) detected a total of 10,604 lightning strokes (the APLLN cross-correlation method detection rate is 84.1% of the ADTD). The APLLN (envelope peak method) detected a total of 9470 lightning strokes (the APLLN envelope peak method detection rate is 75.1% of the ADTD). In order to determine the relative detection efficiency of two detection nets, we needed to calculate the homologous events between the two detection nets. For different detection networks, there are different definitions of homologous events. Based on the characteristics of the APLLN detection network, Wang et al. [15] defined the detection distance of events less than 50 km and the time difference less than 0.5 ms as homologous events. In this paper, the same determination method was used to calculate homologous events. The method for calculating the relative detection efficiency (RDE) of APLLN is:

$$RDE = \frac{Shared\ events\ located\ by\ APLLN\ and\ ADTD}{Events\ located\ by\ ADTD} \tag{8}$$

The APLLN (cross-correlation method) detected a total of 8700 homologous lightning events with the ADTD. It accounted for 69% of the total ADTD detected. The APLLN (envelope peak method) detected a total of 5345 homologous lightning events with the ADTD. It accounted for 42.4% of the total the ADTD detected. The detection results of the APLLN (cross-correlation method) and the APLLN (envelope peak method) relative to the ADTD in different periods are shown in Figure 16.

Figure 16. Distribution of lightning quantity and detection efficiency in different periods. Orange indicates the number of lightning strokes detected by ADTD. Green indicates the number of homologous data detections of APLLN (cross-correlation method) relative to ADTD. Purple represents the number of homologous data detections of APLLN (envelope peak method) relative to ADTD. The black square represents the detection efficiency of APLLN (cross-correlation method) relative to ADTD. The red square represents the detection efficiency of APLLN (envelope peak method) relative to ADTD. The blue dashed line represents the average peak current intensity of lightning detected by ADTD within each hour.

The detection efficiency is related to the peak current intensity and propagation distance of lightning. As the peak current intensity becomes stronger, the signal propagation distance becomes farther in the Earth–ionosphere waveguide and could be detected by more detection sites. From Figure 16, it can be seen that from 19:00 to 20:00, the relative detection efficiency of the APLLN is significantly lower. During this period, the average peak current intensity of lightning detected by the ADTD is 15.7 kA, which is significantly lower than other time periods. This also indicates that the improved the APLLN geolocation algorithm has a lightning detection efficiency of over 60% for lightning above 20 kA.

5.2. Relative Location Accuracy

The results of the ADTD were used as a reference in this paper. The location results of the APLLN (cross-correlation method) and the APLLN (envelope peak method) were compared with those of the ADTD. Within the same latitude and longitude region at the same time selected above, all the data located by each detection network were used for the calculation. The calculation method for relative location accuracy (RLA) can be represented by the following equation:
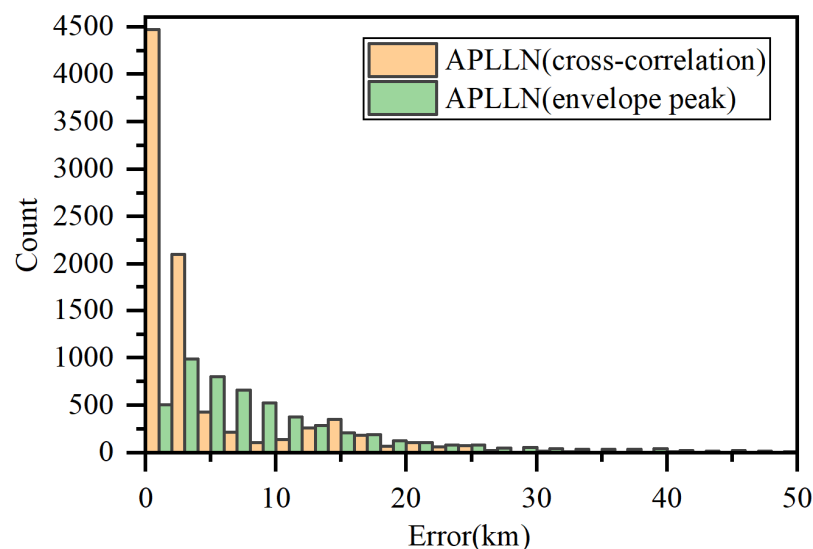
$$RLA = \frac{1}{n}\sum_1^n |Calculated\ value - Reference\ Value| \tag{9}$$

where the reference value and calculated value were selected from the localization results of the same lightning events in ADTD and APLLN, respectively. Considering the existence of errors, the likelihood of identical results between the two systems is relatively low. Therefore, it was necessary to determine an appropriate time tolerance. Similar to the calculation method for relative detection efficiency in Section 5.1, a time tolerance of 0.5 ms was set here. As long as the time error of the location results of the ADTD and APLLN was within 0.5 ms, it was considered as the location result of a lightning event. Finally, we calculated the mean of all errors to eliminate the influence of a small amount of data with larger errors.

To assess the location performance of the cross-correlation method and evaluate the improvement of location accuracy and detection efficiency of the cross-correlation method compared with the envelope peak method, we compared the results of the cross-correlation method and envelope peak method with those of the ADTD. The results are shown in Table 2. The APLLN (cross-correlation method) has a location accuracy of 4.5 km. This is consistent with the Monte Carlo simulation result in Figure 9. The APLLN (envelope peak method) has a location accuracy of 9.9 km. Figure 17 shows the distribution of location errors obtained using the envelope peak and the cross-correlation methods. It is obvious that the location error of the cross-correlation method is concentrated in 0~3 km, accounting for 75.6% of the total quantity. However, using the envelope method, only 27.8% of the data have a location error of less than 3 km. Table 2 summarizes the relative detection efficiency and location error of the APLLN, and it can be seen that the detection efficiency and detection accuracy of the APLLN (cross-correlation method) are greatly improved compared with the APLLN (envelope peak method).

Table 2. Performance evaluation of lightning location accuracy using different algorithms.

	Cross-Correlation Method	Envelope Peak Method
Stroke number	10,604	9470
Homologous events with ADTD	8700	5345
Detection efficiency relative to ADTD	69%	42.4%
Location accuracy relative to ADTD	4.5 km	9.9 km

Figure 17. Distribution diagram of relative location error between envelope peak and cross-correlation method.

6. Conclusions

In this paper, we present an improved algorithm for the long-baseline lightning location network, which aims to achieve better LEMP detection efficiency and location accuracy:

(1) An LEMP waveform compression method based on PCA is proposed and applied for the first time, which realizes real-time compression and the efficient transmission of LEMP waveform data. The compression time for each data point is less than 1 ms.

(2) The cross-correlation technique of the long-baseline LEMP waveform is proposed. On the one hand, the influence of noise on location accuracy is minimized, and on the other hand, the accuracy of calculating the signal arrival time difference is improved.

(3) The relative detection efficiency and relative location accuracy of the waveform cross-correlation method and envelope peak method were evaluated with ADTD data. The detection performance of the long-baseline lightning location network can be further improved by using waveform cross-correlation technology. The lightning location accuracy can be better than 4.5 km, and the relative detection efficiency can reach 69%. It should be emphasized that the location accuracy of the proposed method is twice higher than that of the envelope peak method.

References

1. Nag, A.; Murphy, M.J.; Schulz, W.; Cummins, K.L. Lightning locating systems: Insights on characteristics and validation techniques. *Earth Space Sci.* **2015**, *2*, 65–93. [CrossRef]
2. Kawasaki, Z.; Mardiana, R.; Ushio, T. Broadband and narrowband RF interferometers for lightning observations. *Geophys. Res. Lett.* **2000**, *27*, 3189–3192. [CrossRef]
3. Dong, W.; Liu, X.; Zhang, Y.; Zhang, G. Observations on the leader-return stroke of cloud-to-ground lightning with the broadband interferometer. *Sci. China Ser. D Earth Sci.* **2002**, *45*, 259–269. [CrossRef]
4. Qiu, S.; Zhou, B.-H.; Shi, L.-H.; Dong, W.-S.; Zhang, Y.-J.; Gao, T.-C. An improved method for broadband interferometric lightning location using wavelet transforms. *J. Geophys. Res. Atmos.* **2009**, *114*, D18211. [CrossRef]
5. Liu, H.; Dong, W.; Wu, T.; Zheng, D.; Zhang, Y. Observation of compact intracloud discharges using VHF broadband interferometers. *J. Geophys. Res. Atmos.* **2012**, *117*, D01203. [CrossRef]
6. Zhang, G.; Wang, Y.; Qie, X.; Zhang, T.; Zhao, Y.; Li, Y.; Cao, D. Using lightning locating system based on time-of-arrival technique to study three-dimensional lightning discharge processes. *Sci. China Earth Sci.* **2010**, *53*, 591–602. [CrossRef]
7. Budden, K.G.I. The propagation of a radio-atmospheric. *Lond. Edinb. Dublin Philos. Mag. J. Sci.* **1951**, *42*, 1–19. [CrossRef]
8. Barr, R.; Jones, D.L.; Rodger, C.J. ELF and VLF radio waves. *J. Atmos. Sol.-Terr. Phys.* **2000**, *62*, 1689–1718. [CrossRef]
9. Dowden, R.L.; Brundell, J.B.; Rodger, C.J. VLF lightning location by time of group arrival (TOGA) at multiple sites. *J. Atmos. Sol.-Terr. Phys.* **2002**, *64*, 817–830. [CrossRef]
10. Said, R.K.; Inan, U.S.; Cummins, K.L. Long-range lightning geolocation using a VLF radio atmospheric waveform bank. *J. Geophys. Res. Atmos.* **2010**, *115*, D23108. [CrossRef]
11. Pessi, A.T.; Businger, S.; Cummins, K.L.; Demetriades, N.W.S.; Murphy, M.; Pifer, B. Development of a Long-Range Lightning Detection Network for the Pacific: Construction, Calibration, and Performance. *J. Atmos. Ocean. Technol.* **2009**, *26*, 145–166. [CrossRef]

12. Rodger, C.J.; Werner, S.; Brundell, J.B.; Lay, E.H.; Thomson, N.R.; Holzworth, R.H.; Dowden, R.L. Detection efficiency of the VLF World-Wide Lightning Location Network (WWLLN): Initial case study. *Ann. Geophys.* **2006**, *24*, 3197–3214. [CrossRef]

13. Rodger, C.J.; Brundell, J.B.; Holzworth, R.H.; Lay, E.H. Growing detection efficiency of the World Wide Lightning Location Network. *Amer. IOP Conf. Proc.* **2009**, *1118*, 15–20. [CrossRef]

14. Rodger, C.J.; Brundell, J.B.; Hutchins, M.; Holzworth, R.H. The world wide lightning location network (WWLLN): Update of status and applications. In Proceedings of the 2014 XXXIth URSI General Assembly and Scientific Symposium (URSI GASS), Beijing, China, 16–23 August 2014; pp. 1–2. [CrossRef]

15. Wang, J.; Ma, Q.; Zhou, X.; Xiao, F.; Yuan, S.; Chang, S.; He, J.; Wang, H.; Huang, Q. Asia-Pacific Lightning Location Network (APLLN) and Preliminary Performance Assessment. *Remote Sens.* **2020**, *12*, 1537. [CrossRef]

16. Wang, T.; Chen, F.; Zhang, C.; Sun, Y.; Zhu, W.; Qu, X. Source Locating Algorithm for Ultra-long Baseline Lightning Detection System Based on TDOA. *High Volt. Eng.* **2020**, *46*, 1807–1813. [CrossRef]

17. Wessel, P. Compression of large data grids for Internet transmission. *Comput. Geosci.* **2003**, *29*, 665–671. [CrossRef]

18. Jolliffe, I.T.; Cadima, J. Principal component analysis: A review and recent developments. *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.* **2016**, *374*, 20150202. [CrossRef] [PubMed]

19. Ringnér, M. What is principal component analysis? *Nat. Biotechnol.* **2008**, *26*, 303–304. [CrossRef] [PubMed]

20. Cadima, J.; Cerdeira, J.O.; Minhoto, M. Computational aspects of algorithms for variable selection in the context of principal components. *Comput. Stat. Data Anal.* **2004**, *47*, 225–236. [CrossRef]

21. Pankratius, V.; Jannesari, A.; Tichy, W.F. Parallelizing Bzip2: A Case Study in Multicore Software Engineering. *IEEE Softw.* **2009**, *26*, 70–77. [CrossRef]

22. Bentley, J.L.; Sleator, D.D.; Tarjan, R.E.; Wei, V.K. A Locally adaptive data-compression scheme. *Commun. ACM* **1986**, *29*, 320–330. [CrossRef]

23. Lee, A.C.L. An experimental study of the remote location of lightning flashes using a VLF arrival time difference technique. *Q. J. R. Meteorol. Soc.* **1986**, *112*, 203–229. [CrossRef]

24. Betz, H.D.; Schmidt, K.; Laroche, P.; Blanchet, P.; Oettinger, W.P.; Defer, E.; Dziewit, Z.; Konarski, J. LINET—An international lightning detection network in Europe. *Atmos. Res.* **2009**, *91*, 564–573. [CrossRef]

25. Biagi, C.J.; Cummins, K.L.; Kehoe, K.E.; Krider, E.P. National Lightning Detection Network (NLDN) performance in southern Arizona, Texas, and Oklahoma in 2003–2004. *J. Geophys. Res. Atmos.* **2007**, *112*, D05208. [CrossRef]

26. Liu, B.; Shi, L.; Qiu, S.; Liu, H.; Dong, W.; Li, Y.; Sun, Z. Fine Three-Dimensional VHF Lightning Mapping Using Waveform Cross-Correlation TOA Method. *Earth Space Sci.* **2020**, *7*, e2019EA000832. [CrossRef]

27. Stock, M.G.; Akita, M.; Krehbiel, P.R.; Rison, W.; Edens, H.E.; Kawasaki, Z.; Stanley, M.A. Continuous broadband digital interferometry of lightning using a generalized cross-correlation algorithm. *J. Geophys. Res. Atmos.* **2014**, *119*, 3134–3165. [CrossRef]

28. Lyu, F.; Cummer, S.A.; Solanki, R.; Weinert, J.; McTague, L.; Katko, A.; Barrett, J.; Zigoneanu, L.; Xie, Y.; Wang, W. A low-frequency near-field interferometric-TOA 3-D Lightning Mapping Array. *Geophys. Res. Lett.* **2014**, *41*, 7777–7784. [CrossRef]

29. Wu, T.; Wang, D.; Takagi, N. Lightning Mapping With an Array of Fast Antennas. *Geophys. Res. Lett.* **2018**, *45*, 3698–3705. [CrossRef]

30. Bitzer, P.M.; Christian, H.J.; Stewart, M.; Burchfield, J.; Podgorny, S.; Corredor, D.; Hall, J.; Kuznetsov, E.; Franklin, V. Characterization and applications of VLF/LF source locations from lightning using the Huntsville Alabama Marx Meter Array. *J. Geophys. Res.-Atmos.* **2013**, *118*, 3120–3138. [CrossRef]

31. Koshak, W.J.; Solakiewicz, R.J.; Blakeslee, R.J.; Goodman, S.J.; Christian, H.J.; Hall, J.M.; Bailey, J.C.; Krider, E.P.; Bateman, M.G.; Boccippio, D.J.; et al. North Alabama Lightning Mapping Array (LMA): VHF source retrieval algorithm and error analyses. *J. Atmos. Ocean. Technol.* **2004**, *21*, 543–558. [CrossRef]

32. Rodger, C.J.; Brundell, J.B.; Dowden, R.L. Location accuracy of VLF World-Wide Lightning Location (WWLL) network: Post-algorithm upgrade. *Ann. Geophys.* **2005**, *23*, 277–290. [CrossRef]

33. Wang, Y.; Qie, X.S.; Wang, D.F.; Liu, M.Y.; Su, D.B.; Wang, Z.C.; Liu, D.X.; Wu, Z.J.; Sun, Z.L.; Tian, Y. Beijing Lightning Network (BLNET) and the observation on preliminary breakdown processes. *Atmos. Res.* **2016**, *171*, 121–132. [CrossRef]

34. Dowden, R.L.; Holzworth, R.H.; Rodger, C.J.; Lichtenberger, J.; Thomson, N.R.; Jacobson, A.R.; Lay, E.; Brundell, J.B.; Lyons, T.J.; O'Keefe, S.; et al. World-Wide Lightning Location Using VLF Propagation in the Earth-Ionosphere Waveguide. *IEEE Antennas Propag. Mag.* **2008**, *50*, 40–60. [CrossRef]

35. Wang, J.; Huang, Q.; Ma, Q.; Chang, S.; He, J.; Wang, H.; Zhou, X.; Xiao, F.; Gao, C. Classification of VLF/LF Lightning Signals Using Sensors and Deep Learning Methods. *Sensors* **2020**, *20*, 1030. [CrossRef] [PubMed]