



Optimizing lightning location accuracy: A study of propagation velocity and time of arrival in long-range lightning location algorithms

Jie Li^a, Lin Song^{b,*}, Qilin Zhang^{a,*}, Shudong Wang^c, Jing Yang^a, Quanbo Ge^a

^a Collaborative Innovation Center on Forecast and Evaluation of Meteorological Disasters (CIC-FEMD)/ Key Laboratory of Meteorological Disaster, Ministry of Education (KLME)/Research Institute of Intelligent-Sensing and Disaster Prevention for Extreme Weather/Jiangsu Collaborative Innovation Center on Atmospheric Environment and Equipment Technology (CICAET), Nanjing University of Information Science and Technology, Nanjing 210044, China

^b Qingdao Ecological and Agricultural Meteorological Center, Qingdao 266003, China

^c China Meteorological Administration Public Meteorological Service Center, Beijing 100081, China

ARTICLE INFO

Keywords:

Long-range lightning location
Time of arrival
Propagation velocity
Location accuracy

ABSTRACT

The propagation velocity (v) and time of arrival (T_a) are the primary factors affecting long-range lightning location. In this paper, we have enhanced the Very Low Frequency Long-range Lightning Location Network (VLF-LLN) in China in terms of both pulse pairing methods and location algorithms. Then, the algorithm utilized in the VLF-LLN, and four other long-range lightning location algorithms were compared using lightning strike accidents, simulated lightning strike, and other lightning data as the reference values. Building on this foundation, the discussion centers on the effect of v and T_a on location accuracy, with the results showing that an appropriate T_a yields a more significant improvement in location accuracy compared to v . VLF-LLN uses the peak time of the ground wave as T_a avoids the significant lag caused by sky wave, which is present in other algorithms, and therefore has the best location accuracy with a median location accuracy of 1.81 km.

1. Introduction

The radio atmospheric signal, also known as a “sferic,” refers to a broadband electromagnetic pulse generated by natural lightning strike discharges [1,2]. The VLF component of a sferic can propagate thousands of kilometres in the Earth–ionospheric waveguide (EIWG) [3,4], providing the basis for long-range lightning location [5]. The long-range lightning location network can achieve widespread lightning detection with fewer stations and has significant advantages in detecting lightning in uninhabited land areas and over oceans [6–9]. Meanwhile, lightning location has important applications in lightning physics, short-range warning, and disaster assessment [10–14], making it a crucial tool for understanding and mitigating the impact of lightning strikes.

The sferics detected by the long-range lightning detection network propagation a more extended distance in the EIWG before their arrival at the stations compared to those received by the three-dimensional lightning location network, making the sferics more significantly affected by ground propagation effects, such as the soil conductivity and propagation topography [15–17]. Moreover, the sferics received at the station is a mixture of ground waves and sky waves reflected by the ionosphere, and its characteristics are highly correlated with the

ionospheric characteristics of the Fresnel zone [18–20]. The factors such as changes in the geomagnetic field, lightning strike discharges in the lower part of the Fresnel zone, solar activity all affect the properties of the Fresnel zone, and thus indirectly affect the characteristics of the sferic [18,19,21]. Therefore, the accuracy of long-range lightning location has been at the kilometer level [22–25].

Before lightning location, it is essential to achieve accurate pairing of sferics generated by the same lightning strike discharge received at multiple stations. While crucial in three-dimensional lightning location [26–28], this task is often overlooked or simply handled in long-range lightning location. In the World Wide Lightning Location Network (WWLLN), sferics with similar time of arrival (T_a) are simply assumed to be generated by the same lightning strike discharge [5]. This method overlooks the possibility of multiple lightning strikes occurring in proximity in time. Another long-range lightning locator network, ZEUS, searches for sferics among the stations that are likely the same-stroke candidates based on the degree of similarity between the waveforms recorded by the stations [29,30]. Due to the potential for a “combinatorial explosion” in this method, the ZEUS network was divided into two parts, with five stations in one part and seven stations in the other [29].

The three most common lightning electromagnetic radio-frequency-

* Corresponding authors.

E-mail addresses: qdsonglinyy@163.com (L. Song), qlzhang@nuist.edu.cn (Q. Zhang).

<https://doi.org/10.1016/j.measurement.2024.114754>

Received 30 January 2024; Received in revised form 28 March 2024; Accepted 19 April 2024

Available online 26 April 2024

0263-2241/© 2024 Elsevier Ltd. All rights reserved.

locating techniques are magnetic direction finding (MDF), time of arrival (TOA), and interferometry [31]. The long-range lightning location algorithms are mostly based on TOA, which find the lightning locations by measuring the T_a of sferics and solving the inverse problem with the estimated or empirical propagation velocity (v) of sferics [32]. Currently, there are significant differences in the approach to these factors among the existing long-range lightning algorithms [5,29,33,34]. However, there is a lack of sufficient research to demonstrate which method is superior and to report how much location bias is introduced by differences in the way these factors are handled.

One approach to determine v or T_a is to use theoretical or semi-empirical methods to obtain relatively accurate velocities for lightning location. Dowden et al. proposed the time of group arrival (TOGA) method [5,35], which calculates the T_a by performing a least-squares linear fit on the phase spectrum and has been successfully applied to the WWLLN. An approximate v of about 0.9922 times the speed of light in vacuum was used for all sferics, propagation paths, and frequencies at all times, as recommended by Watt [36]. This value corresponds to the representative group velocity in the middle of the detected frequency range. However, this fixed v cannot fully reflect the time- and space-dependent properties of the EIWG. Furthermore, Xu et al. combined the International Reference Ionosphere (IRI) model with the Long Wavelength Propagation Capability (LWPC) to calculate the v of each sferic and found that a more accurate v leads to higher location accuracy [31]. Another typical long-distance lightning location network, Global Lightning Dataset 360 (GLD360), used a waveform bank to estimate the propagation distance and exact T_a , avoiding the effects of from the EIWG [37]. This method requires extensive data accumulation and accurate propagation modelling, so GLD360 has been upgraded several times [38,39].

Another approach is to use variable v in the location process, where v and lightning location are determined by fitting. This method does not rely on complex theoretical derivations or large amounts of observational data, is easy to implement, and is suitable for use in the initial construction of detection networks. Details can be found in the papers by Liu et al., Wang et al., and Li et al. [24,33,34]. The distinction between the location algorithms presented in these papers lies in whether each sferic is allocated an independent v and the approaches employed to determine T_a . Although there are differences in location accuracy between algorithms in these papers, it is not possible to confirm what factors are responsible.

During 2021, the VLF-LLN was established in the Chinese region and recorded the waveform of the sferics [34]. Based on this, different location algorithms can be compared with each other. In the VLF-LLN, a simulated waveform bank was utilized to extract the time of ground wave peak arrival of the sferics as T_a , which was then combined with the variable v for lightning location. This work has two main shortcomings: the use of an outdated pulse pairing method that requires significant computational resources, and the application of an identical v for each sferic. To address these shortcomings, this paper proposes a new pulse pairing method of measured data based on a simulated waveform bank, while also introducing an independent variable v for each sferic. Subsequently, several long-range lightning location algorithms were applied to the sferics recorded by VLF-LLN, and their location accuracies were compared using both real lightning strike accidents and simulation strikes. Finally, a discussion is presented on the potential introduction of location bias resulting from various methods to handling v and T_a .

2. Data and method

2.1. VLF-LLN

The VLF-LLN used in this paper was established in 2021 and expanded to 26 stations in 2023 [34]. The stations measure vertical electric field within the VLF band (5 Hz ~ 30 kHz), with baseline lengths between stations ranging from hundreds to thousands of

kilometres. These stations are distributed across various regions in China, ensuring that the detection range of the lightning location network covers the entirety of China as well as parts of East Asia and Southeast Asia. The detection range of the lightning location network, ranging from 75°E to 135°E, and from 5°N to 55°N.

2.2. Pulse pairing method

In this paper, we propose a pulse pairing method based on the waveform bank to reduce the time complexity of the lightning location algorithm. The waveform bank utilized in this paper aligns with the waveform bank employed in the ground wave peak extraction of the location algorithm [34,40]. Put simply, we utilize same waveform bank for both pulse pairing and T_a identification. The construction of the waveform bank and the work on T_a identification have been described in detail in the paper by Li et al [34]. In contrast to the waveform bank utilized by Said et al. in GLD360 [37], the waveform bank employed in this paper is very simple.

The simulated waveform bank is constructed using a two-dimensional Finite-Difference Time-Domain (FDTD) method and contains typical daytime and nighttime sferic waveforms at 100 km intervals from 100 km to 3000 km [34]. The size of the FDTD simulation domain is 3200 km × 100 km, the space step is set to $\Delta r = \Delta z = 500$ m, and the time step is set to $\Delta t = 1$ μs to ensure the same time resolution as the observed data. The lightning channel is a vertical dipole model with a height of 10 km. Since the VLF electromagnetic wave has a long wavelength and is much larger than the height of the lightning channel, the lightning channel can be considered as a dipole and the currents are equal throughout the channel. The current source is the dipole model proposed by Cummer [41]. The soil conductivity was set to 0.001 S/m and the geomagnetic field strength was set to 5×10^{-5} T. A two-parameter exponential function expresses the electron density distribution in the D-layer of the ionosphere (~60–90 km). The electron density distributions in the E and F layers are calculated using the International Reference Ionosphere – IRI 2016 model. In addition, the model considers the effect of the Earth's curvature on the simulation results by correcting for the atmospheric refractive index [42]. The simulation results are passed through a Butterworth filter with frequencies from 5 Hz to 30 kHz. Finally, for the accuracy of the results of the FDTD simulation, one can refer to the work of Hou et al [40,43].

The pulse-pairing method is based on the concept that the characteristics of the sferic waveform are more closely correlated with the propagation distance rather than the lightning strikes [44], especially within the VLF band. To enhance the credibility of this concept, Fig. 1 illustrates simulated and measured waveforms at varying distances, along with the distribution of correlation coefficients between the measured and simulated waveforms at each distance. In this paper, the correlation coefficient is used to express the degree of similarity between two waveforms, and it is the absolute value of the Pearson correlation coefficient [45], regardless of the direction of the relationship, ranging from 0 (no linear relationship) to 1 (perfect linear relationship). Fig. 1 (a), 1(c), and 1(e) display the simulated and measured sferic waveforms for daytime and nighttime at distances of 500 km, 1500 km, and 2500 km from the lightning strike point, respectively. The correlation coefficients between the 500 km and 2500 km waveforms with the 1500 km waveform during both daytime and nighttime are 0.56 and 0.57, and 0.52 and 0.64, respectively. Fig. 1(b), (d), and (f) illustrate the distribution of correlation coefficients between the measured and simulated waveforms. The measured waveform datasets comprise sferic waveforms located within a 10 km radius of the difference between the distance received at each station and the distance of the simulated waveform on the left. The dataset spans a one-month period, during which we have collected over 20,000 waveforms for each distance to ensure the representativeness of the results. A correlation coefficient threshold of 0.65 was utilized to distinguish sferic from noise. Obviously, the correlation coefficient between the measured and simulated

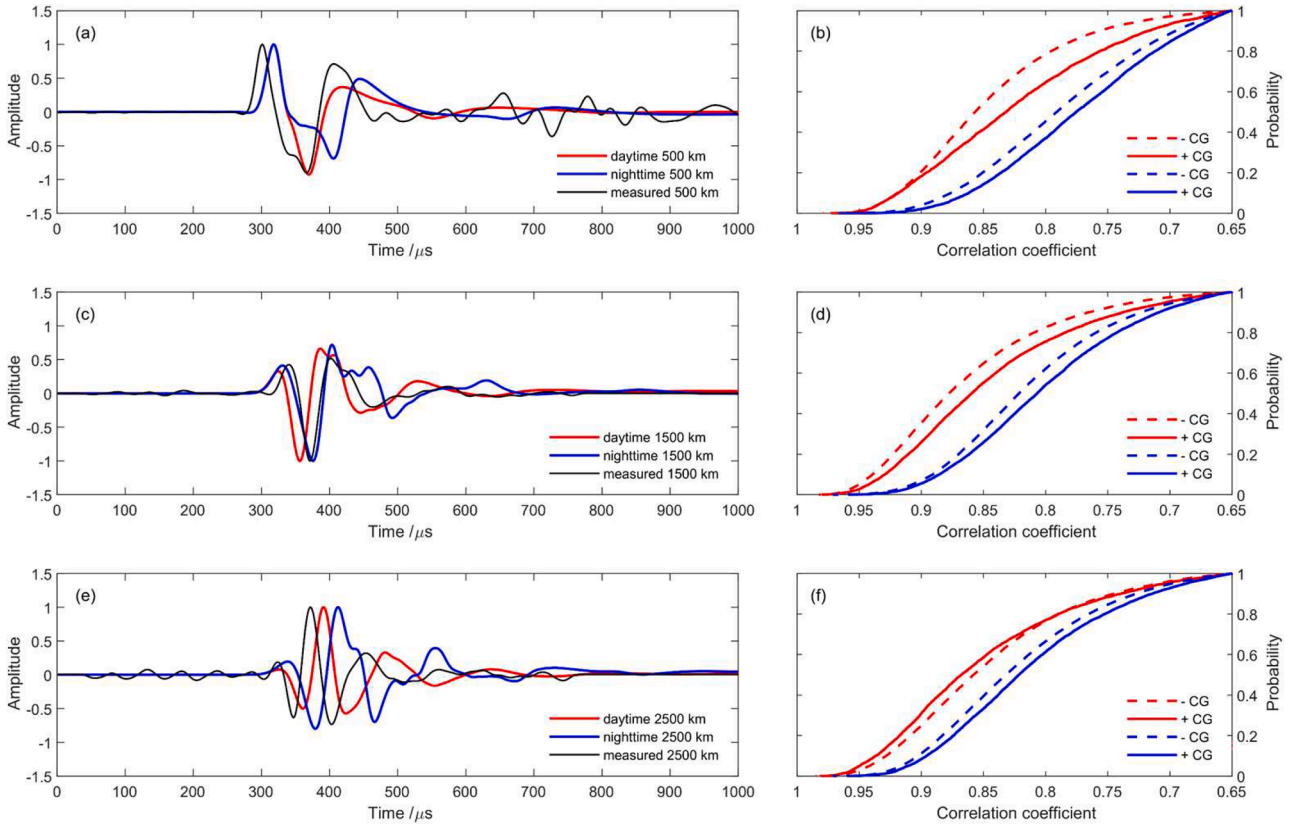


Fig. 1. Simulated and measured waveforms at varying distances and the distribution of correlation coefficients between the measured and simulated waveforms at each distance.

waveforms at the same distance is significantly higher than that between simulated waveforms at different distances, regardless of time, polarity, propagation distance, and path. Therefore, by matching with the simulated waveforms bank, an approximate propagation distance, referred to as D_m , can be determined for each sferic.

Then the relationship in Eq. (1) should be satisfied for the sferic produced by the same lightning discharge process received at two different stations.

$$|\Delta D_m - (\Delta T_a^* \nu)| \leq Th \quad (1)$$

where ΔD_m is the difference between the approximate propagation distances of two sferic, ΔT_a is the difference between the time of arrival of two sferic, and ν is set to c . Th represents the error threshold, which is associated with the accuracy of the waveform bank. As the parameter settings for the waveform bank are fixing and ideal, it is not possible for the simulation waveforms to be consistent with all measured waveforms under different conditions. Inevitably, there will be deviation between the D_m determined by the simulated waveform bank and the actual propagation distance. A suitable error threshold can therefore ensure that more homologous sferics are included in a group without including non-homologous sferics. For the waveform bank in this paper, Th is set to 500 km based on empirical data analysis from August 7th, 2022, demonstrating its effectiveness in sustained operations. Determining whether a sferic belongs to a certain sferic group is simply a matter of determining whether it satisfies Eq. (1) with other sferics within the sferic group.

2.3. Independent variable ν for each sferic

In lightning location algorithms, the initial solution is generally computed by the TOA method and then optimized using the Levenberg-Marquardt algorithm [46,47]. In the early location algorithm, the

optimal solution is obtained by adopting a variable ν in the lightning location, minimizing the location error caused by various factors in the VLF-LLN [34]. The parameters to be optimised include the latitude, longitude and altitude of the lightning strike point and the ν for sferic. However, an identical ν was used for all sferics, and the objective function is shown in Eq. (2) [33].

$$F = \sqrt{\frac{1}{N} \sum_{n=1}^N \left(\Delta t_n - \frac{\Delta d_n}{\nu} \right)^2} \quad (2)$$

where, N is the number of stations, Δt_n denotes the difference in the time of arrival between the N th station and the reference station, Δd_n represents the difference in distance from the lightning strike point to the N th station and the reference station, and ν is from 0.950 c to 1.005 c .

The optimal solution obtained through this method can be significantly affected by individual stations, potentially resulting in negative optimization outcomes [48]. Furthermore, assuming that the ν for all sferics is identical does not reflect realistic conditions. Hence, in the upgraded algorithm, the objective function is updated to Eq. (3).

$$F = \sqrt{\frac{1}{N} \sum_{n=1}^N \left(\left(\Delta t_n - \frac{\Delta d_n}{\nu_n} \right) \cdot \nu_n \right)^2 + h^2} \quad (3)$$

where each sferic is assigned an independent ν . The variable h represents the altitude of lightning. It is assumed that the long-range lightning location network only detects strong cloud-to-ground lightning. This setup enables a fast search for the optimal solution using the method of least squares within an appropriate range. The surface height of each region is ignored, and the two-dimensional resulting error in location is negligible for long-range lightning location networks. To standardize the unit of each term, the time difference in the formula is multiplied by the scalar ν . At the end of the location algorithm, a quality control module is

added to remove possible outliers. When the magnitude of v reaches the boundary value, this location result will be discarded.

2.4. VLF-LLN location results

Fig. 2 illustrates the unfiltered sferic waveforms received by the VLF-LLN within a 10,000 μ s time window. Assuming that sferic propagates at the speed of light, the propagation distance of 10,000 μ s is approximately 3,000 km, which closely aligns with the effective detection range of the station. In other words, all the sferic in this time window could have been generated by the same lightning strike discharge. This is obviously not possible as the same station receives two sferics. And these sferics arrive at various stations at different intervals, indicating that they are not from the same source. For this situation, the pulse pairing method described in this paper successfully separated the sferics generated by these two lightning strikes from a bunch of complex signals. The sferics generated by Lightning 1 and Lightning 2 are shown in red and blue respectively, and the sferics that were discarded are shown in grey.

The information about these two lightning strikes is presented in Table 1, while their respective locations are depicted in Fig. 3. The background of Fig. 3 shows cloud-top temperature (CTT) data from the Feng Yun Satellite Data Centre (<https://satellite.nsmc.org.cn>), which can reflect the strength of convective activity to some extent. From Table 1, it can be observed that these two lightning strikes occurred in different regions, with only a 653 μ s difference in their occurrence times. The sferics they generate arrive at each station at similar times, but due to the different locations at which they occur, each station receives the sferics generated by the two lightning in a different order and at different time intervals. Although Lightning 1 occurred earlier, because it was located outside the VLF-LLN, the sferics it generated reached stations other than the PE stations later than the sferics generated by Lightning 2.

Fig. 3 also illustrates distribution of lightning strikes on 20 July 2023 between 09:00 and 09:15 UTC. Although the stations of VLF-LLN are in mainland China, it still has a good location effect on lightning occurring in southeast Asia and the Pacific Ocean. The location results are all distributed in the region of low CTT probed by Fengyun4A, which is consistent with existing studies [49,50]. Combined with parallel computing, real-time location of lightning in the detection area can be achieved on a single Intel 12700 CPU.

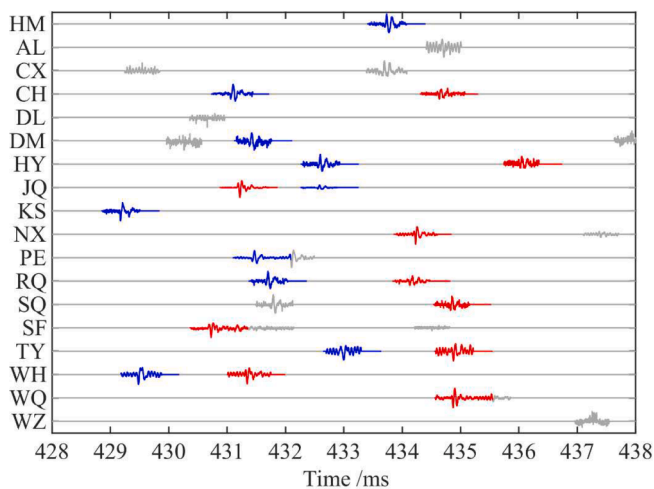


Fig. 2. The unfiltered sferic waveforms received by all stations within a 10000 μ s time window. Red, blue, and grey represent sferics generated by Lightning 1, Lightning 2 and discarded sferics, respectively. The y-axis label is the name of the station.

Table 1

The information of two lightning strikes.

Lightning	Time	Longitude	Latitude	Polarities	Number of stations
1	29.427681	92.9282	28.8371	negative	10
2	29.428334	105.2601	33.4637	positive	10

3. Result

Thanks to the VLF-LLN recording of sferic raw waveforms received at the station, several commonly used long-range lightning location algorithms are compared and evaluated in this section. Since the primary focus of the paper was to identify the optimal location algorithms that do not rely on a substantial accumulation of data but can be directly applied, the location algorithm of GLD360 was not replicated [37]. Table 2 briefly describes the location algorithms compared in this paper. All algorithms are abbreviated in the form of the author's initials followed by the year, for example, the location algorithm of WWLLN is referred to as D08 [5]. In the subsequent discussion, each location algorithm will be applied to sferics that have been identified as originating from the same lightning strike using the pulse matching method employed in this paper. The location accuracy of each algorithm in locating lightning strikes will be assessed using both real accidents and simulated lightning.

A brief account of the above location algorithm is provided below. The M07 and D08 algorithms filter sferics with frequencies between 6 kHz and 22 kHz, using a group velocity of 0.9922 c [5,29] as v . The L16, W20, and L23 algorithms use a variable v during the location process, with the distinction that the L16 algorithm utilizes an identical v for all sferics [33,34]. The L23 algorithm is the one used in this paper.

In the approach to handling T_a , the M07 algorithm uses cross-correlations to ascertain the time difference in T_a among different sferics, as opposed to exact values of T_a . Although the cross-correlation method does not precisely identify homologous sferics, pulse pairing is conducted prior to this in the paper. Therefore, it has also been included in the comparative study. The TOGA, which is used in the D08 algorithm, is calculated by adding sliding time windows to the signal and determining the slope of the phase with respect to frequency within each time window [35]. The T_a in the L16 algorithm is determined by the first peak, which is the nearest local maximum that precedes the absolute maximum in the complex trace in the sferic [33]. The W20 algorithm calculates the Hilbert envelope of the sferic and determines its peak time as T_a [24].

3.1. Assessment of lightning strike accidents

Lightning strike accidents can accurately indicate the location of a lightning strike and provide sufficient evidence to evaluate the effectiveness of a lightning location algorithm, despite posing a significant threat to human life. This section seeks information on four lightning strike accidents (see Table 3) to evaluate the effectiveness of the location algorithms. Accident 1 was a fatal lightning strike that was widely reported in the news and had the highest level of confidence in its location, while the remaining accidents were lightning strikes reported by businesses and received by the local weather service.

Lightning strikes within a 50 km radius of the strike point during the minute when the lightning strike accident occurred were re-located by the five algorithms mentioned above, and the results are shown in Fig. 4. From a macro perspective, all lightning strikes from the L23 algorithm are situated in the region above 35 dBZ. In contrast, the results from the other algorithms are partially located in the low reflectivity region. Referring to the superposition figure of lightning location results from local lightning location networks with radar reflectivity, it can be said that the location accuracy of VLF-LLN is comparable to those networks from a macroscopic perspective [51].

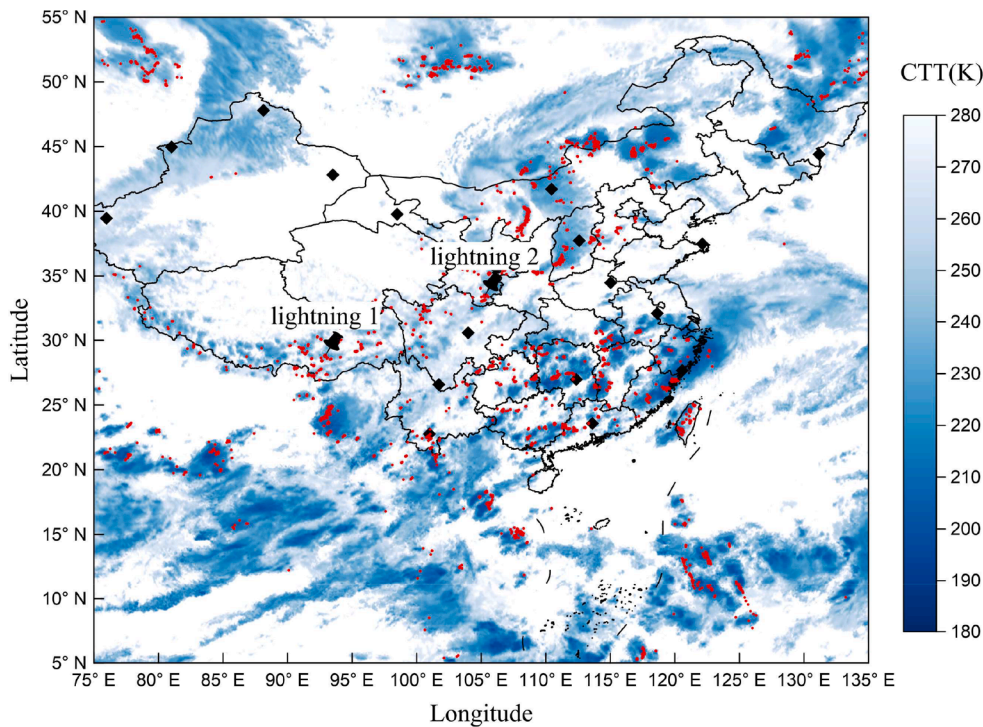


Fig. 3. The distribution of lightning strikes on 20 July 2023 between 09:00 and 09:15 UTC. The black diamonds depict stations presented in Fig. 2, while the small red circles indicate lightning strikes detected by at least five stations. The background signifies the Cloud Top Temperature during the relevant timeframe.

Table 2
The information of the different location algorithms.

Acronyms	T_a	v
M07	cross-correlation	0.9922 <i>c</i>
D08	TOGA	0.9922 <i>c</i>
L16	the first peak	identical
W20	the peak of the envelope	independent
L23	the ground wave peak	independent

Since the reported times of the lightning strike accidents were only accurate to the minute, it was impossible to determine which lightning strikes had caused them. Therefore, the location results within a 10 km radius around the lightning strike accident point are considered as

potential strikes that may have caused the lightning strike accident, and the statistical description of these strikes is shown in Table 3. Except for accident 4, the L23 algorithm outperforms the other algorithms in terms of the number of potential lightning strikes and the nearest distance. For accident 2, only the lightning located by the L23 algorithm was within 10 km of the lightning strike accident. Similarly, in accident 4, the L23 algorithm also outperforms the W20 algorithm in terms of the number of potential lightning strikes, and the superior overall location accuracy of L23 can be seen in Fig. 4(d). The median distance between all potential strikes and the lightning accident point is within the range of 1 to 5 km, with the nearest distance being within 1 km.

Table 3
The potential strikes located by different location algorithms during lightning strike accidents.

Date (UTC)	Latitude	Longitude	Algorithm	Number of Potential Strikes	Nearest Distance (km)	Median Distance (km)
2022/08/26 11:25	30.9190	118.3406	M07	9	3.787	6.697
			D08	8	3.313	8.186
			L16	8	1.636	3.001
			W20	1	4.145	4.145
			L23	16	0.577	1.206
2023/05/22 21:10	22.7308	113.4364	L23	1	3.030	3.030
2023/05/31 21:48	22.5317	113.5331	M07	5	2.824	7.052
			D08	4	5.891	7.580
			L16	3	5.695	6.216
			W20	2	6.956	7.846
			L23	7	1.285	4.328
2023/05/31 22:39	22.3339	113.4639	M07	2	5.082	6.705
			D08	1	9.013	9.013
			L16	5	1.675	5.334
			W20	5	1.136	1.445
			L23	7	2.183	3.957

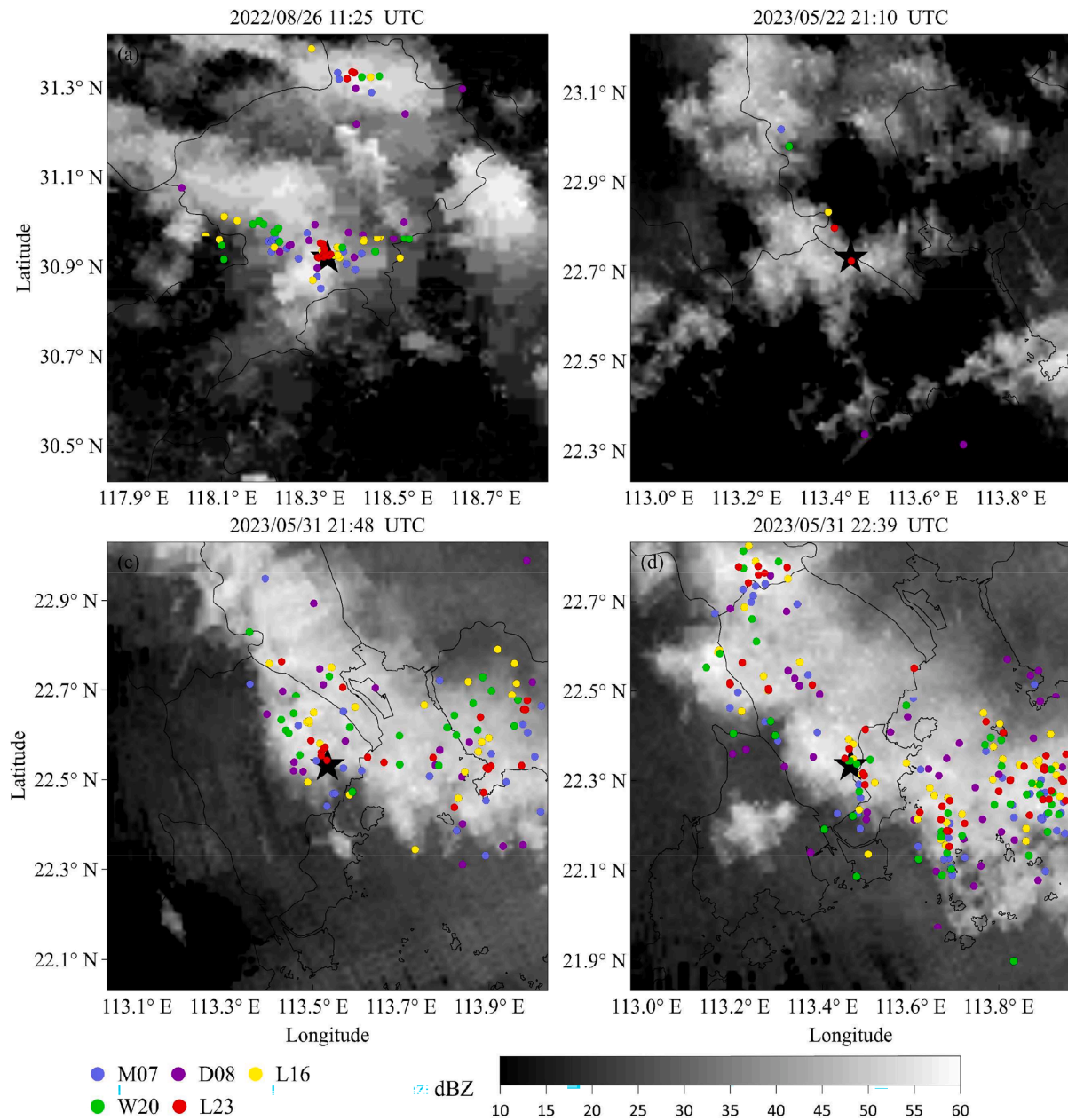


Fig. 4. The relocated result of the lightning within a 50 km radius of the strike point was determined by the above five algorithms during the minute of the lightning accident.

3.2. Assessment of simulated lightning strikes

In this section, the sferic arriving at each station from the above lightning strike accident points are modelled using FDTD and located using the five algorithms. Due to the proximity of the locations of the last three lightning strikes, the modelling was restricted to the first and third strikes. The simulated stations consist of a total of 21 stations, including those used to locate the lightning strikes described above (10 for the first strike and 11 for the third strike). The spatial step size of the simulation domain is 500 m, which means it is not possible to perfectly reproduce the exact lightning strike point. However, the location results still hold significant reference value. Since the Earth is treated as an Ortho sphere in the FDTD model, the simulated location is solved using the Ortho sphere model, while the actual work is done with the WGS84 model.

Fig. 5 shows the location results and deviations of the five algorithms for the simulated lightning strikes mentioned above. The L23 location algorithm used in this paper produces location results closest to the true

location, which is consistent with the assessment based on the actual lightning strike accidents mentioned earlier. Applying independent variable v to the sferic received at each station will lead to better location results, as expected [34]. Despite utilizing the independent variable v , the W20 algorithm does not exhibit a significant advantage in terms of location accuracy over other algorithms. This could possibly be attributed to the method used by the W20 algorithm to handle T_a .

Fig. 6 illustrates the T_a determined by each of the algorithms (excluding M07) for the sferics of the segments at near, middle, and far distances. The M07 algorithm uses cross-correlation and therefore cannot give an accurate T_a . To maintain consistency in the timeline of the data, the 300th point of the plot data is taken as the time node at which the sferic propagates to the station at the speed of light. On the left side, the simulated waveform is displayed, while the measured waveform is shown on the right side, visually illustrating the consistency between the simulated and measured data. The variation of T_a with the propagation distance is minimized by the L23 algorithm employed in

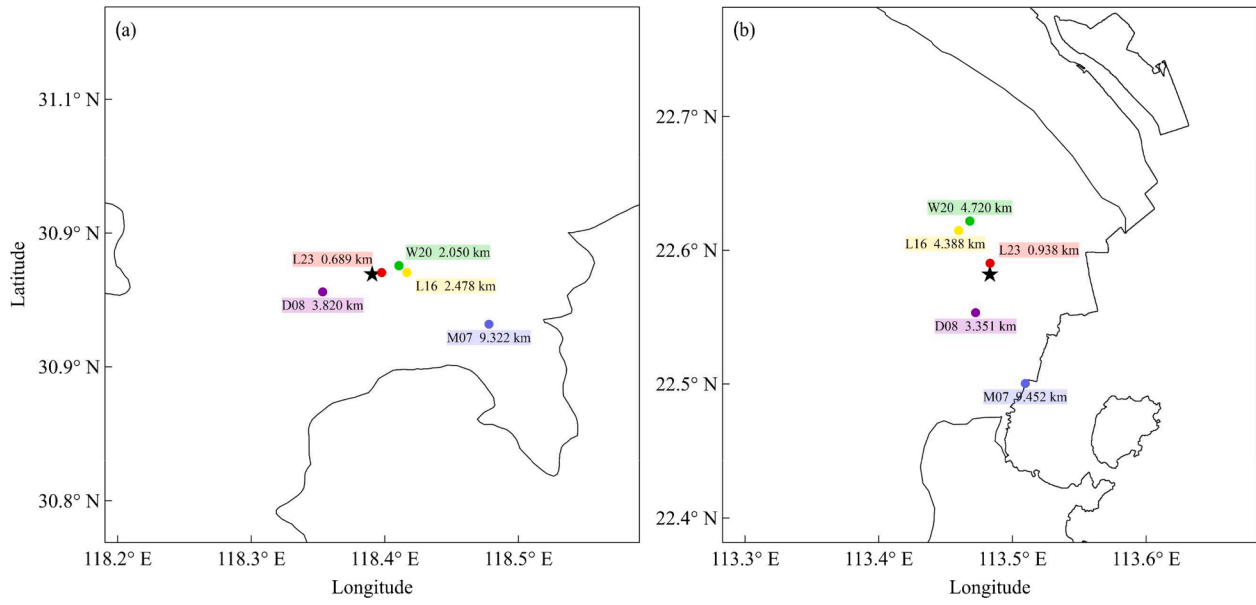


Fig. 5. The location results and deviations of the above five algorithms for simulated lightning strikes. The black star is the simulated lightning strike point.

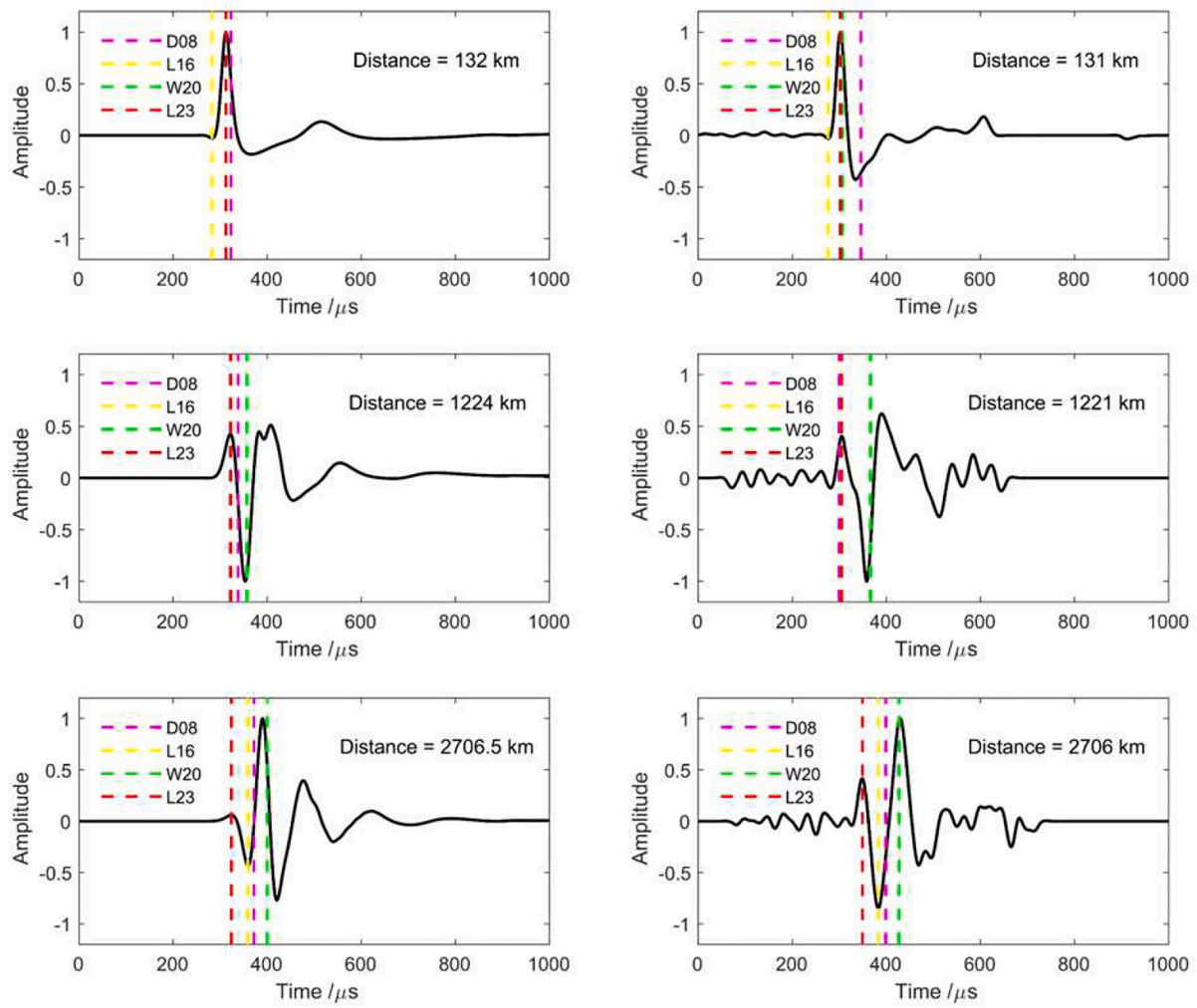


Fig. 6. The T_a determined by different location algorithms are illustrated in the figure. Purple, yellow, green, and red colours correspond to D08, L16, W20, and L23, respectively. Simulated waveforms are shown on the left, while measured waveforms are displayed on the right.

VLF-LLN. The T_a identified by the L16 algorithm at the middle distance aligns with the result obtained from the L23 algorithm and is therefore disregarded. At the other two distances, T_a is determined as the trough of the low-frequency oscillations caused by filtering and as the peak of the first-hop sky wave, respectively. With increasing propagation distance, the proportion of skywave components in the sferic increases, causing the T_a identified by the D08 and W20 algorithms to be pushed later, with the variation being maximum for the W20 algorithm. This may explain the lack of significant improvement in location accuracy despite the use of the independent variable v in the W20 algorithm.

To demonstrate that reducing variation in T_a improves positioning accuracy, we controlled the distance of the simulated data involved in location. Case 1 includes data from five sferics at distances ranging from 688.5 km to 1237.5 km, while Case 2 includes data from six sferics at distances ranging from 834.5 km to 1336.5 km. Although it will reduce the number of stations involved in location, the variation in T_a shrinks. This step makes the T_a determined by the L16 and L23 algorithms the same, so that the difference in the location result between the two is then caused by use of the identical v and the use of independent v .

Before starting the analysis, it should be noted that the D08 algorithm will no longer be involved in the discussion. After reducing the number of stations involved in location, the location error substantially increases for the D08 algorithm. This raises the question of whether the T_a determined by the D08 algorithm is accurate. Analysis of the data indicates that because the TOGA method operates in the frequency domain [35], it is significantly affected by filtering. Simply increasing the order of the Butterworth filter from 1st to 2nd order results in a positioning error of more than ten kilometres. When fewer stations are involved in location, the phase change due to filtering can substantially impact the D08 algorithm's accuracy. However, the analyses presented thus far are not invalidated. However, the analyses already presented are not useless, and this tells us that to reproduce the D08 algorithm, it is necessary to be consistent in the antenna parameters, without being able to pass the filtering of the data processing part at a later stage.

Fig. 7 demonstrates that several location algorithms exhibit significant improvements in accuracy when using only sferics within above distance range. A comparison between the L16 and W20 algorithms reveals that utilizing T_a , which from sferics with similar distances, results in a greater enhancement in accuracy than using an independent v for each sferic. It is important to note that while better location results

can be achieved by using only sferics with similar distances, in practice, the goal is to utilize as many successfully paired sferics as possible. This approach helps to mitigate location errors stemming from noise interference, GPS errors, and other unexpected factors [52].

4. Discussion

4.1. T_a error from sky wave

This section explains why using similar distance sferics for location reduces T_a error. Ideally, if the signal is not affected by propagation effects, TOA can locate the source very accurately. Since the 300th point in the simulated waveform bank in this paper is set to represent the time at which the sferic propagates to the site at the speed of light, the difference between the T_a identified by the different methods and this point can be interpreted as an error introduced by the characteristics of the sferic itself, the propagation effect, and the method of identifying T_a . However, if this difference is constant at all distances, it does not affect the accuracy of the T_a difference used for location, i.e. the spatial location of the source remains accurate, only the time of occurrence is in error.

Fig. 8 illustrates the differences between the T_a of the simulated waveform at each distance and the T_a of the 100 km simulated waveform as determined using the M07, L16, W20, and L23 algorithms, respectively. As analysed in the previous paragraph, the flatter the curve on the Fig. 8, the smaller the location error generated when the T_a determined by the algorithm is used for location. Before analysing, a correction is made to the problem that the L16 algorithm above would judge the low-frequency oscillation point to be T_a for close sferic. If the L16 algorithm considers the low-frequency oscillation point to be, then is corrected to the time of the ground wave peak. These curves corresponding to M07, L16, and W20 algorithms all have distinct jump points, reflecting the influence of skywave on T_a . In daytime, the lower equivalent reflection height for the sferic of the lower ionosphere causes the skywave component arrive at the station earlier [21,34,53]. As a result, the M07 and L16 algorithms, which rely more on the overall waveform characteristics of the signal to identify T_a , have earlier jump points compared to night.

On the other hand, the jump points of the curve corresponding to the W20 algorithm may be attributed to the fact that, after 500 km of

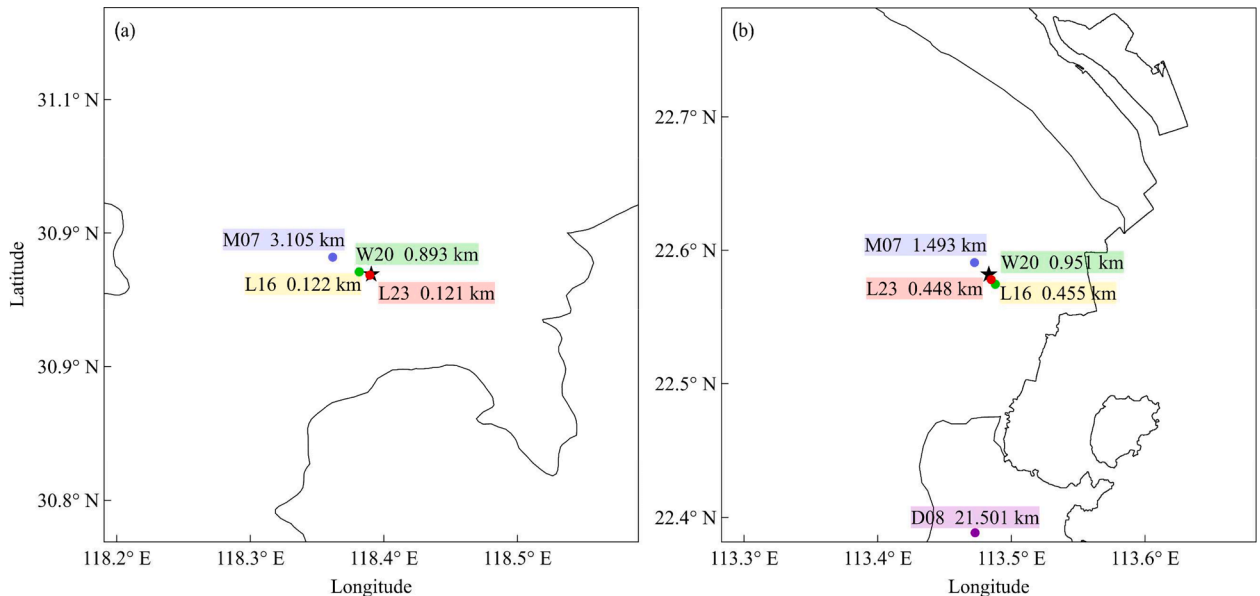


Fig. 7. The location results and deviations of the five algorithms mentioned above for locating sferics at similar distances. The black star is the simulated lightning strike point.

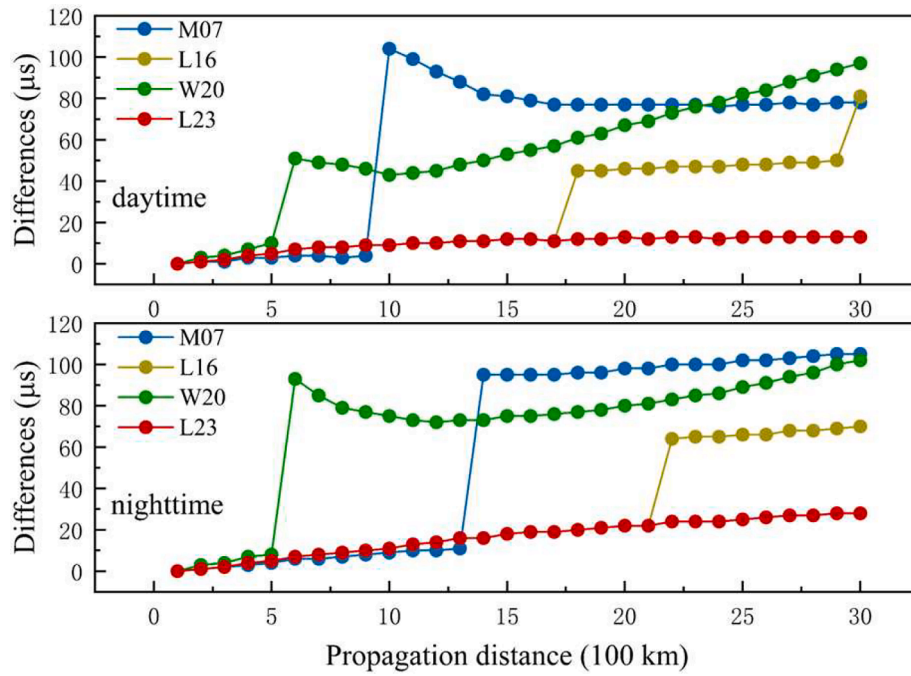


Fig. 8. The differences between the T_a of the waveform at each distance and the T_a of the 100 km waveform as determined using the M07, L16, W20, and L23 algorithms, respectively.

propagation, the energy of the ground wave has decayed to a level lower than that of the sky wave. Due to the sky wave component being more backward at night (see Fig. 1), the T_a difference after the jump is also larger in comparison to daytime. Analysing the propagation process of sferics is a highly complex task, but it is evident that the L23 algorithm corresponds to the minimum T_a difference.

4.2. Location bias due to v and T_a

In the paper, simulation studies revealed that employing T_a (which minimally affected by sky wave) and using an independent v for each sferic both contribute to improved location accuracy. However, in practical applications, the exact extent of the improvement in location accuracy resulting from these methods is not yet known. Therefore, the

location results obtained from the L23 algorithm are considered as reference points, and two control algorithms are established to investigate the impact of v and T_a on location bias. Control Algorithm 1 (CA1) utilizes the time of the ground wave peak as T_a and applies a uniform v for all sferics. It is used to examine the impact of v on location bias. Control Algorithm 2 (CA2) employs the peak time of the envelope as T_a and assigns an independent v for each sferic. Please note that the CA2 algorithm mentioned here is the same algorithm as the W20 algorithm mentioned above, and it is used to explore the effect of T_a on location bias. The relocated data covers all lightning occurrences in Summer 2023, totalling 4,364,341, which is highly representative. To mitigate the impact of station layout on the results, the study area was confined to latitudes 20°N to 40°N and longitudes 95°E to 120°E.

The median location bias for the v and T_a are 0.076 km and 6.843 km,

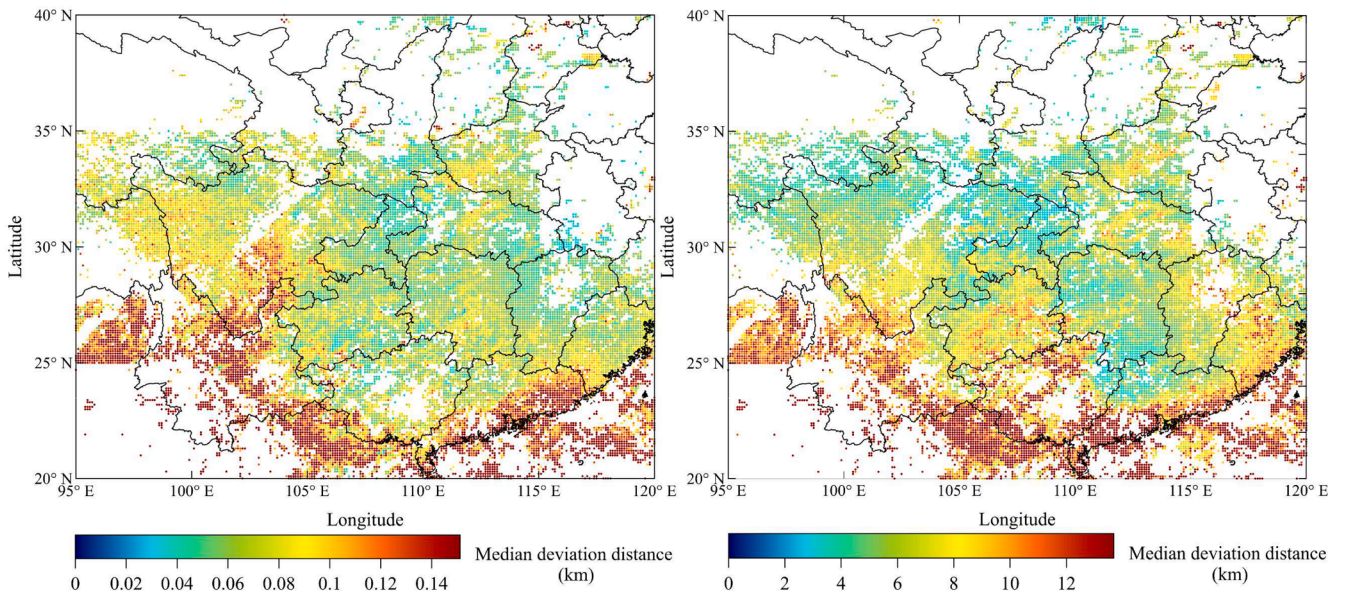


Fig. 9. The distribution of median location bias between the control algorithm and the L23 algorithm, with CA1 on the left and CA2 on the right.

respectively. The study area was divided into grids with a step size of 0.1° . Median location bias was calculated for grids with more than 50 lightning strikes within a grid point, and the results are presented in Fig. 9. At the macroscopic level, there is a gradual widening trend in the deviations between the control and L23 algorithms from the center of the network outwards. In detail, however, the location bias is not evenly distributed, which may be due to the layout of the stations [52]. From the results so far, it appears that the location bias due to differences in v and T_a are inconsistently affected by the station layout, as the deviation distribution is not uniform across different parts of Fig. 9.

4.3. VLF-LLN location accuracy

In this section, we will evaluate the location accuracy of the VLF-LLN using the advanced direction-time lightning detection system (ADTD), which has been in widespread use in China for decades. The location accuracy of the ADTD, which is 500 m [24], is significantly larger than the potential location bias caused by v . It may not be appropriate to utilize the ADTD to assess the superiority of the L23 and CA1 algorithms. Therefore, in this section, only the ADTD data is used to evaluate the location accuracy of L23 and CA2 algorithm.

The data for ADTD are for the whole day of 30 July 2023 BJT, the study area is the same as in Section 4.2, and there are 63,767 lightning flashes after clustering and matching. Clustering involves grouping strikes that meet both temporal and distance thresholds within the same lightning location network into a single flash. Matching involves considering flashes in different lightning location networks that meet both the temporal and distance thresholds as the same flash. Because the supplied ADTD data was only accurate to the second, the time threshold was set to the same second instead of the commonly used 0.5 s [51]. However, the distance thresholds for clustering and matching are consistent with commonly used values, which are 10 km and 30 km, respectively.

Fig. 10 shows the distribution of positioning accuracy for the L23 and CA2 algorithms. Despite the different station layout in VLF-LLN compared to APLLN, the location of the replicated CA2 algorithm in VLF-LLN aligns with the 5 km to 10 km location accuracy reported by Wang et al. for APLLN, with a median value of 5.998 km [24]. This also demonstrates the successful algorithm reproduction in this paper. The median location accuracy of VLF-LLN using the L23 algorithm is 1.814 km, which aligns with the previous study, demonstrating that using a T_a that minimally affected by sky wave significantly improves the location accuracy.

5. Conclusions

This paper presents a pulse-pairing method in a long-range lightning location algorithm that is based on the waveform bank. Five different location algorithms are compared by using lightning strike accidents and simulation strikes. Additionally, two control algorithms are constructed, and the effects of the parameters v and T_a on the location results are discussed. The main conclusions of this paper are as follows:

The characteristics of the sferic depend more on the propagation distance than on the lightning source. The Pearson correlation coefficient between the measured and simulated waveforms at the same distance is significantly higher than that between simulated waveforms at different distances, regardless of time, polarity, propagation distance, and path. Consequently, the approximate propagation distance of the sferic can be determined using a simulated waveform bank. Subsequently, the pulse pairing method based on this propagation distance and the ground wave peak time can successfully group the sferics generated by two lightning strikes occurring almost simultaneously at different locations.

By implementing the control algorithm, it was observed that using different values for v and T_a resulted in a location bias of 0.076 km and 6.843 km, respectively. Comparative results with ADTD demonstrate

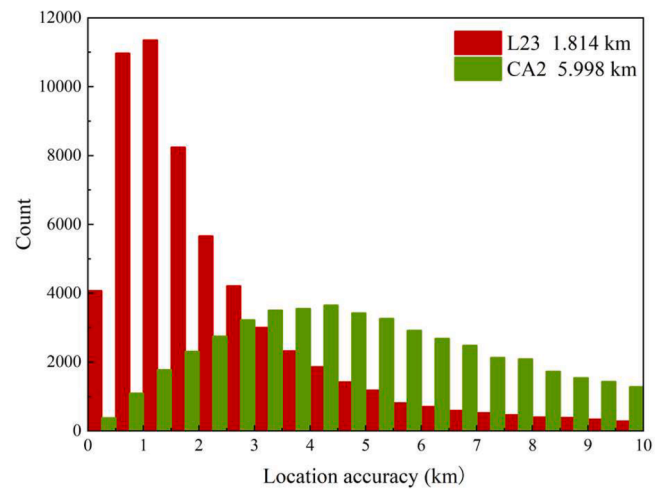


Fig. 10. The distribution of location accuracy for the L23 algorithm and the CA2 algorithm.

that using a T_a which minimally affected by sky wave significantly improves location accuracy. In terms of median, the location accuracy of VLF-LLN is 1.814 km.

The pulse-pairing method presented in this paper has the potential to further improve pairing efficiency through methods such as machine learning, especially as a large amount of observational data is accumulated. Additionally, the calibration of the VLF-LLN is ongoing due to the increasing number of new stations, with the future expectation of providing the electrical parameters of each lightning strike. Furthermore, our next research focus will be on the application of VLF-LLN data in the field of thunderstorm forecasting [54].

CRedit authorship contribution statement

Jie Li: Methodology, Conceptualization, Software, Writing – original draft. **Lin Song:** Resources, Data curation. **Qilin Zhang:** Writing – review & editing, Resources, Conceptualization. **Shudong Wang:** Data curation. **Jing Yang:** Methodology. **Quanbo Ge:** Resources.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

The authors would like to thank all those involved in the construction and maintenance of the VLF-LLN and the staff of the meteorological services who provided the meteorological data.

References

- [1] D.A. Smith, K.B. Eack, J. Harlin, et al., The Los Alamos Sferic Array: a research tool for lightning investigations[J], J. Geophys. Res. Atmos. 107 (D13) (2002), <https://doi.org/10.1029/2001JD000502>. ACL-1–ACL 5-14.
- [2] A. Ohkubo, VLF/ELF sferic evidence for in-cloud discharge activity producing sprites[J], Geophys. Res. Lett. 32 (4) (2005) L04812, <https://doi.org/10.1029/2004GL021943>.
- [3] A.T. Pessi, S. Businger, K.L. Cummins, et al., Development of a long-range lightning detection network for the Pacific: construction, calibration, and performance[J], J. Atmos. Oceanic Tech. 26 (2) (2009) 159–168, <https://doi.org/10.1175/2008JTECHA1132.1>.

- [4] H. Zhang, G. Lu, X. Qie, et al., Locating narrow bipolar events with single-station measurement of low-frequency magnetic fields[J], *J. Atmos. Sol. Terr. Phys.* (2016) 88–101, <https://doi.org/10.1016/j.jastp.2016.03.009>.
- [5] R.L. Dowden, R.H. Holzworth, C.J. Rodger, et al., World-wide lightning location using VLF propagation in the earth-ionosphere waveguide[J], *IEEE Antennas Propag. Mag.* 50 (5) (2008) 40–60, <https://doi.org/10.1109/MAP.2008.4674710>.
- [6] X. Qie, T.Y. Zhou, Lightning activity on the central Tibetan Plateau and its response to convective available potential energy[J], *Chin. Sci. Bull.* (2003), <https://doi.org/10.1007/BF03183302>.
- [7] L. Pan, X. Qie, D. Liu, D. Wang, J. Yang, The lightning activities in super typhoons over the Northwest Pacific, *Sci. China Ser. D Earth Sci.* 53 (8) (2010) 1241–1248, <https://doi.org/10.1007/s11430-010-3034-z>.
- [8] H. Iwasaki, Climatology of global lightning classified by stroke energy using WWLLN data[J], *Int. J. Climatol.* 35 (14) (2016) 4337–4347, <https://doi.org/10.1002/joc.4291>.
- [9] X. Qie, K. Qie, L. Wei, K. Zhu, Z. Sun, S. Yuan, R. Jiang, H. Zhang, C. Xu, Significantly increased lightning activity over the Tibetan Plateau and its relation to thunderstorm genesis, *Geophys. Res. Lett.* 49 (16) (2022), <https://doi.org/10.1029/2022GL099894>.
- [10] X. Qie, R. Zhu, T. Yuan, et al., Application of total-lightning data assimilation in a mesoscale convective system based on the WRF model[J], *Atmos. Res.* 145 (2014) 255–266, <https://doi.org/10.1016/j.atmosres.2014.04.012>.
- [11] F. Wang, X. Qie, D. Liu, et al., Lightning activity and its relationship with typhoon intensity and vertical wind shear for Super Typhoon Haiyan (1330) [J], *J. Meteorol. Res.* 30 (1) (2016) 117–127, <https://doi.org/10.1007/s13351-016-4228-x>.
- [12] H. Zhang, G. Lu, F. Lyu, et al., First measurements of low-frequency sferics associated with Terrestrial Gamma-ray Flashes produced by equatorial thunderstorms[J], *Geophys. Res. Lett.* (2020), <https://doi.org/10.1029/2020GL089005>.
- [13] F.J. Pérez-Invernón, H. Huntrieser, T. Erbertseder, et al., Quantification of lightning-produced NO_x over the Pyrenees and the Ebro Valley by using different TROPOMI-NO₂ and cloud research products[J], *Atmos. Meas. Tech.* 15 (11) (2022) 3329–3351, <https://doi.org/10.5194/amt-15-3329-2022>.
- [14] H. Wang, Y. Tan, Z. Shi, et al., Diurnal differences in the effect of aerosols on cloud-to-ground lightning in the Sichuan Basin[J], *Atmos. Chem. Phys.* 23 (4) (2023) 2843–2857, <https://doi.org/10.5194/acp-23-2843-2023>.
- [15] J.R. Jöhler, C.M. Lilley, Ground-conductivity determinations at low radio frequencies by an analysis of the sferic signatures of thunderstorms[J], *J. Geophys. Res.* 66 (10) (1961) 3233–3244, <https://doi.org/10.1029/JZ066i010p03233>.
- [16] R. Said, M. Gólkowski, V. Harid, Empirical parameterization of broadband VLF attenuation in the earth-ionosphere waveguide[J], *J. Geophys. Res. Space Phys.* 128 (4) (2023) e2022JA030742, <https://doi.org/10.1029/2022JA030742>.
- [17] W. Hou, M. Azadifar, M. Rubinstein, et al., On the propagation of lightning-radiated electromagnetic fields across a mountain[J], *IEEE Trans. Electromagn. Compat.* 62 (5) (2020) 2137–2147, <https://doi.org/10.1109/TEMC.2019.2947095>.
- [18] X. Zhou, J. Wang, Q. Ma, et al., A method for determining D region ionosphere reflection height from lightning skywaves[J], *J. Atmos. Sol. Terr. Phys.* 221 (2021) 105692, <https://doi.org/10.1016/j.jastp.2021.105692>.
- [19] Z. Qin, Probe the ionospheric D region with lightning sferics based on a ray-tracing earth-ionosphere wave guide model[J], 2019. <https://theses.lib.polyu.edu.hk/handle/2000/10028>.
- [20] J.C. McCormick, M.B. Cohen, N.C. Gross, et al., Spatial and temporal ionospheric monitoring using broadband sferic measurements[J], *J. Geophys. Res. Space Phys.* 123 (4) (2018) 3111–3130, <https://doi.org/10.1002/2017JA024291>.
- [21] X.M. Shao, E.H. Lay, A.R. Jacobson, Reduction of electron density in the night-time lower ionosphere in response to a thunderstorm[J], *Nat. Geosci.* 6 (1) (2013) 29–33, <https://doi.org/10.1038/ng1668>.
- [22] A.R. Jacobson, R. Holzworth, J. Harlin, et al., Performance assessment of the world wide lightning location network (WWLLN), using the Los Alamos Sferic Array (LASA) as ground truth[J], *J. Atmos. Oceanic Tech.* 23 (8) (2006) 700–705, <https://doi.org/10.1175/JTECH1902.1>.
- [23] D. Abreu, D. Chandan, R.H. Holzworth, et al., A performance assessment of the World Wide Lightning Location Network (WWLLN) via comparison with the Canadian Lightning Detection Network (CLDN)[J], *Atmos. Meas. Tech.* 3 (4) (2010) 1143–1153, <https://doi.org/10.5194/amt-3-1143-2010>.
- [24] J. Wang, Q. Ma, X. Zhou, et al., Asia-Pacific Lightning Location Network (APLLN) and preliminary performance assessment[J], *Remote Sens. (Basel)* 12 (10) (2020) 1537, <https://doi.org/10.3390/rs12101537>.
- [25] T. Zhang, J. Wang, Q. Ma, et al., Improving the detection effect of long-baseline lightning location networks using PCA and waveform cross-correlation methods [J], *Remote Sens. (Basel)* 16 (5) (2024) 885, <https://doi.org/10.3390/rs14174242>.
- [26] F. Lyu, S.A. Cummer, R. Solanki, et al., A low-frequency near-field interferometric-TOA 3-D Lightning Mapping Array[J], *Geophys. Res. Lett.* 41 (22) (2015) 7777–7784, <https://doi.org/10.1002/2014GL061963>.
- [27] J. Wang, Y. Zhang, Y. Tan, Z. Chen, D. Zheng, Y. Zhang, Y. Fan, Fast and fine location of total lightning from low frequency signals based on deep-learning encoding features, *Remote Sens.* 13 (2021) 2212, <https://doi.org/10.3390/rs13112212>.
- [28] W. Yanhui, M. Yingchang, L. Yali, et al., A non-time-synchronized lightning positioning method and its preliminary application[J], *Atmos. Res.* 285 (2023) 106641, <https://doi.org/10.1016/j.atmosres.2023.106641>.
- [29] T.G. Chronis, E.N. Anagnostou, Evaluation of a long-range lightning detection network with receivers in Europe and Africa[J], *IEEE Trans. Geosci. Remote Sens.* 44 (6) (2006) 1504–1510, <https://doi.org/10.1109/TGRS.2006.871217>.
- [30] C.A. Morales, E.N. Anagnostou, E. Williams, et al., Evaluation of peak current polarity retrieved by the ZEUS long-range lightning monitoring system[J], *IEEE Geosci. Remote Sens. Lett.* 4 (1) (2007) 32–36, <https://doi.org/10.1109/LGRS.2006.883528>.
- [31] V.A. Rakov, Electromagnetic methods of lightning detection, *Surv. Geophys.* 34 (2013) 731–753, <https://doi.org/10.1007/s10712-013-9251-1>.
- [32] X. Xu, L. Huang, S. Wang, et al., VLF/LF lightning location based on LWPC and IRI models: a quantitative study[J], *Remote Sens. (Basel)* 14 (22) (2022) 5784, <https://doi.org/10.3390/rs14225784>.
- [33] Z. Liu, K.L. Koh, A. Mezentsev, et al., Variable phase propagation velocity for long-range lightning location system[J], *Radio Sci.* 51 (2016), <https://doi.org/10.1002/2016RS006058>.
- [34] J. Li, B. Dai, J. Zhou, J. Zhang, Q. Zhang, J. Yang, Y. Wang, J. Gu, W. Hou, B. Zou, et al., Preliminary Application of long-range lightning location network with equivalent propagation velocity in China, *Remote Sens.* 14 (2022) 560, <https://doi.org/10.3390/rs14030560>.
- [35] R.L. Dowden, J.B. Brundell, C.J. Rodger, VLF lightning location by time of group arrival (TOGA) at multiple sites[J], *J. Atmos. Sol. Terr. Phys.* 64 (7) (2002) 817–830, [https://doi.org/10.1016/S1364-6826\(02\)00085-8](https://doi.org/10.1016/S1364-6826(02)00085-8).
- [36] A.D. Watt, VLF Radio Engineering [M]. Pergamon, 1967. doi: 10.1016/C2013-0-02069-5.
- [37] R.K. Said, U.S. Inan, K.L. Cummins, Long-range lightning geolocation using a VLF radio atmospheric waveform bank[J], *J. Geophys. Res. Atmos.* 115 (D23) (2010), <https://doi.org/10.1029/2010JD013863>.
- [38] R. Said, A. Nag, Method to Improve Location Accuracy of the GLD360[C]//23rd International Lightning Detection Conference and 5th International Lightning Meteorology conference, 2014, <https://api.semanticscholar.org/CorpusID:15333960>.
- [39] R. Said, M. Murphy, GLD360 upgrade: Performance analysis and applications[C]//24th international lightning detection conference, International Lightning Detection Conference and International Lightning Meteorology Conference, San Diego, CA, 2016, <https://api.semanticscholar.org/CorpusID:53645357>.
- [40] W. Hou, M. Azadifar, M. Rubinstein, et al., The polarity reversal of lightning-generated sky wave[J], *J. Geophys. Res. Atmos.* (2020), <https://doi.org/10.1029/2020JD032448>.
- [41] S.A. Cummer, An analysis of new and existing FDTD methods for isotropic cold plasma and a method for improving their accuracy[J], *IEEE Trans. Antennas Propag.* 45 (3) (1997) 392–400, <https://doi.org/10.1109/8.558654>.
- [42] W. Hou, M. Azadifar, M. Rubinstein, et al., An efficient FDTD method to calculate lightning electromagnetic fields over irregular terrain adopting the moving computational domain technique[J], *IEEE Trans. Electromagn. Compat.* 62 (3) (2019) 976–980, <https://doi.org/10.1109/TEMC.2019.2917282>.
- [43] Z. Qin, M. Chen, B. Zhu, et al., An improved ray theory and transfer matrix method-based model for lightning electromagnetic pulses propagating in Earth-ionosphere waveguide and its applications[J], *J. Geophys. Res. Atmos.* 122 (2) (2017) 712–727, <https://doi.org/10.1002/2016JD025599>.
- [44] J. Wang, F. Xiao, S. Yuan, et al., A novel method for ground-based VLF/LF single-site lightning location[J], *Measurement* 196 (2022) 111208, <https://doi.org/10.1016/j.measurement.2022.111208>.
- [45] I. Cohen, Y. Huang, J. Chen, et al., Pearson correlation coefficient[J], *Noise Reduct. Speech Process.* (2009) 1–4, https://doi.org/10.1007/978-3-642-00296-0_5.
- [46] P.M. Bitzer, H.J. Christian, M. Stewart, et al., Characterization and applications of VLF/LF source locations from lightning using the Huntsville Alabama Marx Meter Array[J], *J. Geophys. Res. Atmos.* 118 (8) (2013) 3120–3138, <https://doi.org/10.1002/jgrd.50271>.
- [47] Z. Ma, R. Jiang, X. Qie, et al., A low frequency 3D lightning mapping network in north China[J], *Atmos. Res.* 249 (2021) 105314, <https://doi.org/10.1016/j.atmosres.2020.105314>.
- [48] J. Zhang, J. Zhou, J. Li, et al., Location accuracy improvement of long-range lightning detection network in China by compensating ground wave propagation delay[J], *Remote Sens. (Basel)* 14 (14) (2022) 3397, <https://doi.org/10.3390/rs14143397>.
- [49] F.T. São Sabbas, D.D. Sentman, Dynamical relationship of infrared cloud top temperatures with occurrence rates of cloud-to-ground lightning and sprites[J], *Geophys. Res. Lett.* 30 (5) (2003), <https://doi.org/10.1029/2002GL015382>.
- [50] Y. Tian, X. Qie, Y. Sun, et al., Total lightning signatures of thunderstorms and lightning jumps in hailfall nowcasting in the Beijing area[J], *Atmos. Res.* 230 (2019) 104646, <https://doi.org/10.1016/j.atmosres.2019.104646>.
- [51] A. Srivastava, Y. Tian, X. Qie, et al., Performance assessment of Beijing Lightning Network (BLNET) and comparison with other lightning location networks across Beijing[J], *Atmos. Res.* 197 (2017) 76–83, <https://doi.org/10.1016/j.atmosres.2017.06.026>.
- [52] D.R. Poelman, W. Schulz, R. Kaltenboeck, et al., Analysis of lightning outliers in the EUCLID network[J], *Atmos. Meas. Tech.* 10 (11) (2017) 4561–4572, <https://doi.org/10.5194/amt-10-4561-2017>.
- [53] E.H. Lay, X.M. Shao, High temporal and spatial-resolution detection of D-layer fluctuations by using time-domain lightning waveforms[J], *J. Geophys. Res. Space Phys.* 116 (A1) (2011), <https://doi.org/10.1029/2010JA016018>.
- [54] Z. Chen, J. Sun, X. Qie, et al., A method to update model kinematic states by assimilating satellite-observed total lightning data to improve convective analysis and forecasting[J], *J. Geophys. Res. Atmos.* 125 (22) (2020) e2020JD033330, <https://doi.org/10.1029/2020JD033330>.