Preprints.org

Article

Not peer-reviewed version

An Improved Multi-Threshold Data Quality Control Algorithm for W-Band Cloud Radar Based on K-means Clustering

Zhao Shi , LingJiang Huang , Fengyuan Wu , Yong Lei , Huiying Wang * , Zhiya Tang

Posted Date: 28 November 2024

doi: 10.20944/preprints202411.2219.v1

Keywords: W-band cloud radar; Data quality control; K-means; Cloud and fog identification

Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 November 2024 doi:10.20944/preprints202411.2219.v1

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

An Improved Multi-Threshold Clutter Filtering Algorithm for W-Band Cloud Radar Based on K-Means Clustering

Zhao Shi [1,2,†], Lingjiang Huang [1,2,†], Fengyuan Wu [1,2], Yong Lei [4], Huiying Wang [3,*] and Zhiya Tang [1,2]

[1] College of Atmospheric Sounding, Chengdu University of Information Technology, Chengdu 610225, China
[2] Key Laboratory of Atmospheric Sounding, China Meteorological Administration, Chengdu 610225, China
[3] National Meteorological Information Center of CMA, Beijing 100081, China
[4] Meteorological Observation Center of CMA, Beijing 100081, China
* Correspondence: wanghy@cma.gov.cn
† These authors contributed equally to this work.

Abstract: This study investigates the application of an improved multi-threshold method based on the K-means algorithm for clutter filtering in W-band cloud and fog radar observations. Utilizing W-band millimeter-wave cloud and fog radar data collected from March to July 2023 in the Qingdao area, a dataset of cloud and fog echo of different types was constructed and statistically analyzed. Subsequently, a multi-threshold clutter filtering method was proposed to identify and eliminate abnormal interferences such as noise spikes, radial interference, and suspended matter clutter. This method employs the basic data and spatiotemporal information from the cloud radar as feature variables for K-means clustering and dynamically adjusts thresholds based on the clustering results. The clutter-filtered data were further used for the verification analysis of cloud and fog identification. The results demonstrate that the proposed multi-threshold method effectively removes clutter and significantly reduces its impact on cloud and fog echo under weather conditions of clouds, fog, and coexisting cloud–fog, while controlling the loss of cloud and fog echo within the required accuracy range.

Keywords: W-band cloud radar; clutter filtering; K-means; Cloud and fog identification

1. Introduction

Clouds play a crucial role in the energy balance of the climate system and have a profound impact on the atmosphere in various ways [1]. They are not only a key link in the Earth's water cycle but also an important part of numerous chemical reactions in cloud and precipitation processes [2]. Additionally, clouds are the primary targets for weather modification [3]. Clouds that touch the ground can be defined as fog, which is a weather phenomenon wherein the horizontal visibility distance is reduced to within 1 km due to the suspension of numerous water droplets or ice crystal particles in the near-ground layer of the atmosphere [4]. Secondary disasters induced by inland fog and sea fog often cause significant losses to the national economy and social life. Therefore, the study of the formation, dissipation, and evolution characteristics and mechanisms of clouds and fog is one of the key issues in meteorological disaster prevention and reduction, as well as in meteorological research [5].

In recent years, the development of ground-based vertical remote sensing technology and equipment for meteorological observations has been rapidly accelerating [6]. Millimeter-wave cloud radars (MMCRs) have become advanced atmospheric detection instruments for cloud and fog research due to their high sensitivity, high spatial and temporal resolution, and all-weather observation capabilities. Cloud radar can capture the height, thickness, and microphysical parameters of clouds and fog, providing important parameters and observational data for the development of physical process parameterization schemes in numerical cloud forecast models,

© 2024 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 November 2024 doi:10.20944/preprints202411.2219.v1

weather modification, and quantitative estimation of precipitation, thereby enhancing the capabilities of weather forecasting and warning services [7].

Both domestic and international millimeter-wave cloud radars mainly choose the Ka or W bands, with corresponding frequency window ranges of approximately 35 GHz or 94 GHz, respectively. The differences in power, microparticle sensitivity, hardware systems, and platform carrying capabilities of the two bands make them applicable in different scenarios [8,9]. Ka-band cloud radar is mainly based on ground or vehicle platforms, characterized by high power and mature hardware, suitable for observing targets within a particle size range of 0–800 μm, which is usually non-precipitating clouds, weak precipitating clouds, and drizzle [10]. W-band cloud radar, in addition to ground-based observations, is also applied on shipborne, airborne, and satellite platforms [11,12]; its characteristics include small hardware, higher sensitivity, and suitability for observing particle sizes ranging from 0 to 350 μm—typically inland fog, marine fog, and cirrus clouds—thus being capable of refined detection of low-altitude cloud and fog targets. Pazmany et al. developed a 95 GHz dual-polarization radar for cloud research, and its high-frequency characteristics provide a new perspective for detailed observations of cloud structures [13]. Clothiaux et al. evaluated a 94 GHz radar, proving the effectiveness of W-band radar in detecting cloud properties [14]. Sekelsky and McIntosh utilized dual-polarization radar technology to provide new polarimetric measurements for cloud research [15]. Martner and Moran used dual-polarization measurements from cloud radar to distinguish between stratocumulus clouds and insect echoes, which is an important application in the field of meteorological detection [16]. Kollias et al. emphasized the new frontier status of millimeter-wave radar in atmospheric cloud and precipitation research through observation experiments [17], the view also supported by Johannes et al. [18] and Wada et al. [19], who revealed multilayer mixed-phase cloud processes and cirrus cloud observations through millimeter-wave radar. Turk et al. demonstrated the use of near-coincident observations between Global Precipitation Measurement (GPM) and the CloudSat Profiling Radar (CPR) (W-band, or 94 GHz) to extend the ability to represent light rain and cold-season precipitation from dual-frequency precipitation radar (Ku and Ka bands) and the GPM passive microwave constellation sensors [20]. Beni et al. compared radars operating in the X, Ku, and W bands for bridge monitoring and found that, despite their differences, all three sensors are suitable for dynamic structural monitoring, highlighting the role of amplitude variation in phase/displacement measurement [21]. Camplani et al. utilized the 94 GHz CloudSat brightness temperature (TB) product to conduct an extensive comparison with the corresponding simulated signal, aiming to improve snowfall retrieval capabilities, emphasizing the role of W-band radar in addressing the challenges of supercooled liquid water in snowfall [22]. In addition, Chinese research on W-band radar technology has seen notable advancements [23]. Ji Lei et al. compared attenuation correction methods for airborne W-band radar, simulating echo intensities and optimizing the kc–Z relationship [24]. Wu Juxiu et al. analyzed cloud characteristics during light rain in Hefei [25] and compared W-band and Ka-band radar detection capabilities, confirming the W-band radar's stability and ability to detect cloud macro- and microstructures [26]. A new Ka/W-band dual-frequency radar at the Tibet Observatory has been analyzed for its cloud detection performance [27]. Despite progress, domestic fog detection equipment remains limited in range and passive, underutilizing the potential of W-band radar for the observation of small particles. Currently, domestic detection equipment for fog and marine fog mainly consists of forward-scattering visibility meters and fog droplet spectrometers, but the detection range of these two types of equipment for fog is very limited, and both are passive observation methods. The advantages of W-band cloud radar in the observation of small particles have not yet been fully utilized in the detection of marine fog.

W-band cloud radar data's quality is affected by the detection environment, the performance of the equipment itself, and the characteristics of the target objects, resulting in various interference issues in the observed data. The most common interferences in the data observed by W-band cloud radar include isolated noise points, suspended matter clutter, radial interference, and lower tropospheric clear-air echoes. To accurately retrieve the macro- and microphysical properties of clouds and fog, especially for marine fog, it is necessary to conduct clutter filtering for the developed W-band cloud radar. Radar noise and radial interference clutter are related to birds and insects above

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 November 2024 doi:10.20944/preprints202411.2219.v1

the radar, as well as the radar hardware and data processing methods. Median filtering is a common method used to deal with radar noise and radial interference [28–30]. This method replaces data points with the median value within the neighborhood, in order to smooth the data while preserving edge information. Clear-air and suspended matter clutter has weak reflectivity factors and high linear depolarization ratios, which can be filtered out by using a dual-threshold method based on the statistical reflectivity factor (Z) and linear depolarization ratio (LDR)[16]. Scholars have studied various clutter filtering methods based on the characteristics of clutter on the dual threshold. Görsdorf et al. used a reflectivity factor of less than 30 dBZ and a linear depolarization ratio greater than -20 dB as threshold values to eliminate the interference of suspended particles in the radar echo [31]. Zheng JiaFeng et al. statistically analyzed the data of Z and LDR below 2 km in Yangjiang and Naqu, and due to the differences in suspended matter, the LDR range showed a significant gap, using different dual-threshold values to filter out suspended matter clutter [32]. Xiao Pei found that clutter has obvious seasonal distribution characteristics, and by dividing the seasons and setting statistical thresholds for Z-LDR, it can better filter out the clutter in the radar data of Beijing [33]. Wan Xia et al. set thresholds based on the characteristics of radial velocity, in addition to the dual threshold of reflectivity factor and linear depolarization ratio, and used a three-threshold algorithm to filter out 90% of the suspended matter echoes in their observed data [34]. The aforementioned clutter filtering algorithms mainly set thresholds by statistically dividing meteorological and non-meteorological echoes. The effect of the median filtering method is affected by the size of the set window parameter, showing different performance. Therefore, once the threshold and parameters are determined, they will remain unchanged in the algorithm. However, the choice of thresholds will fluctuate due to factors such as the geographical location of the radar, the climate region, and the condition of the radar hardware. In addition, W-band radar is usually carried on moving platforms such as ships and satellites, which are characterized by frequent movement and the need to receive and transmit data in real time under complex conditions. Therefore, determining how to effectively achieve dynamic thresholds to filter out various non-meteorological echoes is a major challenge faced by current W-band cloud radar clutter filtering algorithms. Moreover, in the current research on clutter filtering using cloud radar data, algorithms are mainly constructed based on radar spectral parameters such as the reflectivity factor, radial velocity, and depolarization ratio. Since cloud radar uses a vertical pointing detection, it can obtain temporal information of the echo, so it is also possible to use the height and time distribution information and continuity information of the cloud radar data in vertical observation for feature analysis. However, there are still few algorithms that combine spectral data with dynamic spatiotemporal information to achieve sufficient clutter filtering.

Millimeter-wave radar clutter filtering requires the establishment of feature variables, including the reflectivity factor, radial velocity, radial velocity spectrum width, linear depolarization ratio, range gate continuity count, and time gate continuity count, among other key feature variables. The interrelationships of these feature variables are complex, and it is difficult to statistically determine thresholds that can distinguish cloud and fog echoes from non-meteorological echoes. The in-depth analysis and processing of multidimensional data relationships pose significant challenges. With the rapid development of artificial intelligence technology, especially the successful application of machine learning algorithms in radar detection and signal processing fields, AI algorithms have shown great potential in clutter filtering. Among them, clustering algorithms, as unsupervised learning algorithms, group unlabeled sample points based on the data themselves and their internal information relationships.

Clustering algorithms, as data analysis tools, can identify unique relationships in complex data environments [35]. The idea of the K-means algorithm originated in 1957 with Hugo Steinhaus [36], first used by J. MacQueen in 1967 [37], and the standard algorithm was first implemented by Stuart P. Lloyd in 1957 and published in 1982 [38]. The K-means clustering algorithm, with its unique advantages such as light weight, high computational efficiency, and strong interpretability, performs well in scenarios of multidimensional data processing, providing strong technical support for the identification and in-depth analysis of effective information hidden in massive meteorological data [39–43]. Ubaid et al. used random forests for feature selection and then used K-means clustering for

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 November 2024 doi:10.20944/preprints202411.2219.v1

weather classification to improve the prediction of short-term solar irradiance [44]. Yin et al. used the K-means algorithm for cluster analysis of tropical cyclones generated in the South China Sea and, based on the clustering results, extracted the diversity of paths and characteristics [45]. Tao Fa used an improved neural network algorithm combined with principal component analysis and the K-means algorithm to enhance the identification of clear-air echoes by millimeter-wave cloud radar [46]. The common feature of the above studies is that they all used the clustering results obtained by the K-means algorithm, refined the cluster information, and then used that information to optimize the effects of other research methods.

This study used W-band cloud radar data, analyzed the reflectivity factor characteristics of cloud and fog echoes, and established thresholds to distinguish meteorological from non-meteorological echoes based on the continuity of spatiotemporal distribution information. Clustering information was then introduced to correct the statistical thresholds, proposing an improved multi-threshold clutter filtering method for W-band cloud radar based on the K-means algorithm. Subsequently, we applied the multi-threshold method and the improved multi-threshold method to several typical weather scenarios for an evaluation and comparative analysis of the clutter filtering effect. Finally, the effectiveness of this clutter filtering method was verified by identifying clouds and fog using the clutter-filtered cloud radar data, in combination with data from all-sky imaging instruments.

The structure of this paper is as follows: Section 2 introduces the data sources and processing methods used in this study. Section 3 details the multi-threshold clutter filtering method and the improved multi-threshold clutter filtering method based on the K-means method. Section 4 introduces the selected cloud and fog weather cases for evaluating the improved multi-threshold method. Section 5 introduces the cloud and fog identification results after clutter filtering in this method, with a special focus on the comparison with the all-sky imaging instrument. Section 6 mainly presents the research summary of this paper.

2. Materials and Methods

In this study, authoritative data provided by the Huangdao District Meteorological Bureau in Qingdao, Shandong Province was collected and used as the research basis. We mainly used W-band millimeter-wave cloud radar (hereafter referred to as W-cloud radar) and all-sky imaging instruments to observe weather processes. The radar was well calibrated, providing complete and accurate observation data for the design of the algorithm and enhancement of its analytical work .

The W-cloud radar shown in Figure 1 has a remote observation capability of up to 30 kilometers in different scanning modes, including Time Height Indicator (THI), Range Height Indicator (RHI), and Plan Position Indicator (PPI). In normal operation, this radar mainly works in the THI scanning mode, wherein the 30-kilometer distance refers to the vertical distance, which enables it to effectively monitor the sea fog weather process in part of the Yellow Sea. The radar uses pulse detection technology, making full use of the scattering characteristics of cloud and fog particles in response to short-wavelength electromagnetic waves and the penetration power of electromagnetic waves, thereby achieving high-precision measurement of the vertical profile and spatial distribution of particles in clouds and fog. This technology shows excellent detection performance for cloud, fog, and weak precipitation weather processes. The W-cloud radar has the advantage of a fully solid-state system, combined with its small size, strong power, and low power consumption, ensuring that it can operate continuously under all weather conditions; its main performance indicators are shown in Table 1.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 November 2024 doi:10.20944/preprints202411.2219.v1

Figure 1. Installation location of W-cloud radar.

Table 1. Main performance indicators of W-cloud radar.

Component	Parameter Name	Parameter Specification
Antenna	Antenna gain	53.51 dB
	Horizontal beam width	0.3°
	Vertical beam width	0.3°
	Operation mode	Vertically directed observation
	Polarization mode	Transmit horizontal polarization; simultaneously receive horizontal and vertical polarization echoes
Transmitter	Frequency	93.75 GHz
	Transmit pulse power	15 W
	Transmit pulse width	0.6 us/5 us/20 us
	Repetition interval staggered	2：3
Receiver	Dynamic range	95 dB
	Noise figure	8.9 dB
	Intermediate frequency	60 MHz
Signal processing	FFT accumulation number	128
	Spectral average number	32
	Maximum detection range	30 km
	Reflectivity factor	-50 dBZ to 50 dBZ
	Radial velocity	-15 to 15 m/s
	Spectrum width	0 to 15 m/s

The W-band cloud and fog radar uses alternating horizontal and vertical polarization transmission, while receiving horizontal and vertical polarization echoes, i.e., using an "alternate transmit and simultaneously receive" polarization system. The dual-channel receiver can

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 November 2024 doi:10.20944/preprints202411.2219.v1

simultaneously receive the backscatter echoes of the two polarization components. When the radar beam is projected onto cloud and fog particles, they produce scattering. Part of the backscattered signal returns to the antenna and is received. After processing, parameters such as the equivalent reflectivity factor, radial velocity, radial velocity spectrum width, and linear depolarization ratio of weather targets can be obtained. This section introduces the polarimetric radar parameters defined by the dual-polarization scattering amplitude: the equivalent reflectivity factor of horizontal polarization Z_h, and the equivalent reflectivity factor of vertical polarization Z_v.

The equivalent reflectivity factor Z_h of the particle ensemble in horizontal polarization can be represented by the square of the co-polarized scattering amplitude s_{hh} averaged over a collection; hereafter, when we refer to the reflectivity factor Z, it specifically refers to the measurement value Z_h of the radar's horizontal polarization equivalent reflectivity factor, i.e.,

$$Z_h = \frac{4\lambda^4}{\pi^5 |K_w|^2} \langle |s_{hh}|^2 \rangle \quad (1)$$

where $|K_w|^2$ is the dielectric constant of water, λ is the radar wavelength, and the collection average of s_{hh} is calculated within a unit volume of 1 m^3. Similarly, by replacing s_{hh} with s_{vv}, the equivalent reflectivity factor Z_v in vertical polarization can be obtained via the following formula:

$$Z_v = \frac{4\lambda^4}{\pi^5 |K_w|^2} \langle |s_{vv}|^2 \rangle \quad (2)$$

The linear depolarization ratio, which characterizes the meteorological target's depolarization effect on electromagnetic waves, is expressed in dB as the vertical linear depolarization ratio LDR_{hv}:

$$LDR_{hv} = \frac{\langle |s_{hv}|^2 \rangle}{\langle |s_{vv}|^2 \rangle} \quad (3)$$

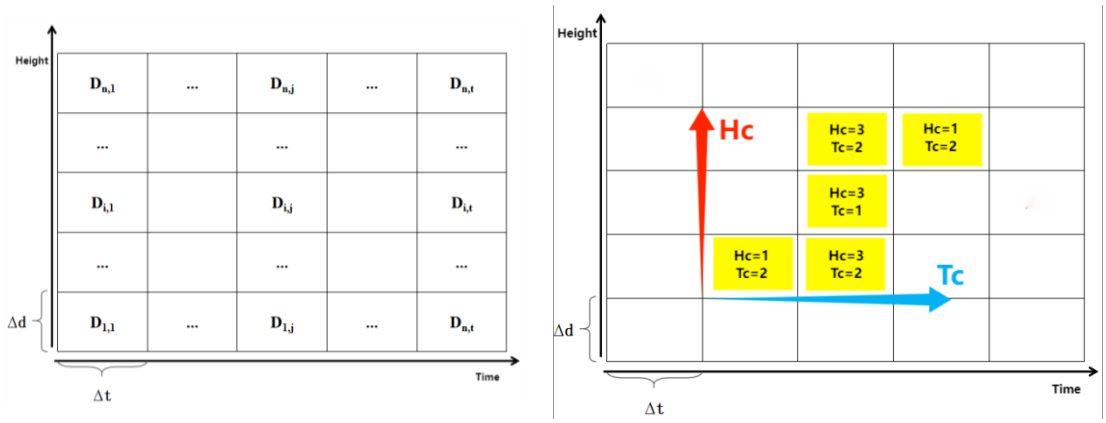
Additionally, the average radial velocity provides information about the speed and direction of airflows in the atmosphere, which is very important for analyzing the motion characteristics of meteorological particles and the vertical velocity of the atmosphere. The spectrum width reflects the diversity and complexity of the internal velocities of the target, which can be used to retrieve products such as water content and particle phase. For millimeter-wave radars, the Doppler spectrum refers to the function of the power of all particle echoes with respect to their velocity in the cloud radar sample library. The actual obtained Doppler velocity spectrum density is S(v), and the first and second moments of the Doppler spectrum density $S_h(v)$ obtained from the horizontal channel are calculated to obtain the corresponding average radial velocity $\bar{v}$(in m/s) and velocity spectrum width σ (in m/s), as expressed below:

$$\bar{v} = \frac{\int_{-\infty}^{\infty} v S_h(v) dv}{\int_{-\infty}^{\infty} S_h(v) dv} \quad (4)$$

$$\sigma = \frac{\int_{-\infty}^{\infty} (v - \bar{v})^2 S_h(v) dv}{\int_{-\infty}^{\infty} S_h(v) dv} \quad (5)$$

The W-band cloud and fog radar performs vertical-pointing observation and, after processing, it obtains THI data of parameters such as $Z_h, \bar{v}$, and σ. The THI echo data are presented in a matrix structure, which is composed of the vertical range gate and the horizontal time gate, referred to as THI data, as shown in Figure 2a. The total number of range gates is n, and the total number of time gates is t. In this matrix, any element Di,j (where 0<i<n, 0<j<t, and i,j are natural numbers) represents a specific datum in the matrix, such as Z_h, $\bar{v}$, σ, LDR_{hv}, etc. In addition, based on the distribution characteristics of the data in the time gate and range gate dimensions, the height range gate continuity count (Hc) and time gate continuity count (Tc) can also be obtained. As shown in Figure 2b, if the yellow area represents valid data, the corresponding range and time gate continuity information will be assigned.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 November 2024 doi:10.20944/preprints202411.2219.v1



(**a**) (**b**)

Figure 2. Schematic diagram of millimeter-wave radar THI data: (a) THI matrix data. (b) Spatiotemporal continuity.

3. Proposed Methodology

The improved multi-threshold clutter filtering algorithm based on K-means clustering proposed in this paper is divided into four steps, as shown in Figure 3. The first step is data collection and statistical echo characteristics, obtaining the distribution information of the reflectivity factor and spatiotemporal features of clouds and fog, and using this information to set multiple threshold classifiers to identify meteorological echoes and eliminate non-meteorological ones; the second step is K-means clustering: preprocessing the data and using the random forest model to select features from the radar data, and then inputting important features for clustering to obtain cluster information of meteorological and non-meteorological echoes; the third step is the improvement of the algorithm with clustering information: extracting cluster information and incorporating the reflectivity factor, time continuity, and range continuity information extracted from the clusters into the multi-threshold algorithm to perform clutter filtering on hourly THI data using the improved threshold; the fourth step is to evaluate the performance of the research method. The steps of the proposed method are detailed below.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 November 2024 doi:10.20944/preprints202411.2219.v1

Figure 3. Research steps of the K-means multi-threshold clutter filtering algorithm.

3.1. Multi-Threshold Clutter Filtering Method

To effectively select meteorological echoes from THI data and exclude non-meteorological echoes, we extracted multidimensional features of meteorological echoes for accurate classification, and data that did not meet the threshold were classified as non-meteorological and clustered echoes. To ensure the accuracy of meteorological echo identification, we needed to set reasonable thresholds for the reflectivity factor. At the same time, due to the vertical detection characteristics of cloud radar, THI data contain information on the continuous time of the echo. On the vertical height, suspended clutter will not show a continuous and patchy distribution. On the other hand, the duration of cloud echoes is significantly longer, which can exceed 30 minutes, and the duration of some stratiform clouds can even reach more than 5 hours. Therefore, when designing the clutter filtering scheme in this method, in addition to using the reflectivity factor threshold for clutter filtering, the height information and time information in THI data were also fully utilized to set thresholds to enhance the accuracy and effectiveness of clutter filtering.

We utilized the observational data from March to July 2023 to establish a cloud and fog weather dataset by manually selecting data samples. Specifically, it was divided into six datasets: high clouds, medium clouds, low clouds, advection fog, radiation fog, and mixed fog.

In the field of meteorology, the classification of clouds is usually carried out based on the key indicator of cloud base height [47]. Specifically, 2 km and 6 km are used as the definite boundaries. If the cloud base height is less than 2 km, the cloud is classified as a low cloud. If the cloud base height falls between 2 and 6 km, such a cloud can be classified as a medium cloud. When the cloud base height reaches 6 km or above, it is classified as a high cloud. Based on this classification criterion, in

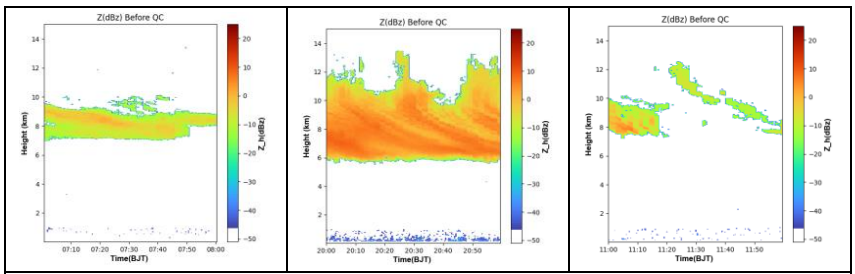
Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 November 2024 doi:10.20944/preprints202411.2219.v1

the subsequent construction of the dataset in this study, different types of clouds could be clearly distinguished, and further exploration of the related meteorological characteristics and the differences among various clouds and fogs could be conducted.

The formation and maintenance mechanisms of heavy fog are mainly described from several perspectives, including the effects of radiative cooling, vertical structural characteristics, water vapor conditions, dynamic structural characteristics, and wind speed conditions. The cases of radiation fog detected by the cloud radar are mostly inland fog in the Qingdao area. Radiation fog is mainly formed by the condensation caused by the radiative cooling and the temperature reduction of water vapor in the near-surface layer, so the height of the fog is relatively low. Due to the decrease in water vapor caused by the downward movement of air during the formation process of radiation fog, the reflectivity factor is relatively low. The interface between upward and downward movements is prone to forming an inversion layer, which effectively prevents the upward transport of low-level water vapor, causing it to accumulate in the near-surface layer, thereby facilitating the formation of heavy fog. By combining the radial velocity data, it can be determined that the height of the inversion layer is also relatively low. The turbulence, enhanced by the radiative cooling effect of radiation fog at night, leads to a relatively wider velocity spectrum compared to that of advection fog. Radiation fog usually appears from midnight to early morning, reaches its peak intensity before sunrise, and gradually dissipates thereafter. Its duration is generally 1-2 hours, ranging from dozens of minutes to ~4-5 hours [48,49].

Advection fog frequently appears in the coastal areas where the radar is deployed. It is characterized by both the height of the fog top and the thickness of the fog being relatively high. This is because of the effect of warm and humid advection, which makes the inversion layer of advection fog higher than that of radiation fog, resulting in the advection fog being higher than radiation fog. Sufficient water vapor supply usually gives advection fog a relatively high reflectivity factor. Meanwhile, the internal turbulence is weak, and the radial velocity spectrum is narrow. Moreover, advection fog has a long duration, and the frequency of its occurrence during the day and at night is very similar. In Qingdao, along the coast, 00:00-08:00 is the main period for fog formation, and 08:00-16:00 is the main period for its dissipation [50].

Mixed fog has the characteristics of both advection fog and radiation fog. When heavy fog has the characteristics of radiation fog, the wet layer is shallow, and the top of the inversion layer is also low. When heavy fog has the characteristics of advection fog, the wet layer is thick, and the top of the inversion layer is also high. The duration of the heavy fog is relatively long, and there will be an obvious conversion process in the height of the fog top and the thickness of the fog. Meanwhile, it also has the characteristics of advection fog and radiation fog in terms of water vapor conditions and dynamic structural characteristics. Through analyzing the reflectivity factor, radial velocity, radial velocity spectrum width, height distribution, and time variation characteristics detected by the radar, data samples were manually selected to obtain the dataset. Examples of the six types of cloud and fog weather datasets are shown in Figure 4.



(**a**)

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 November 2024 doi:10.20944/preprints202411.2219.v1

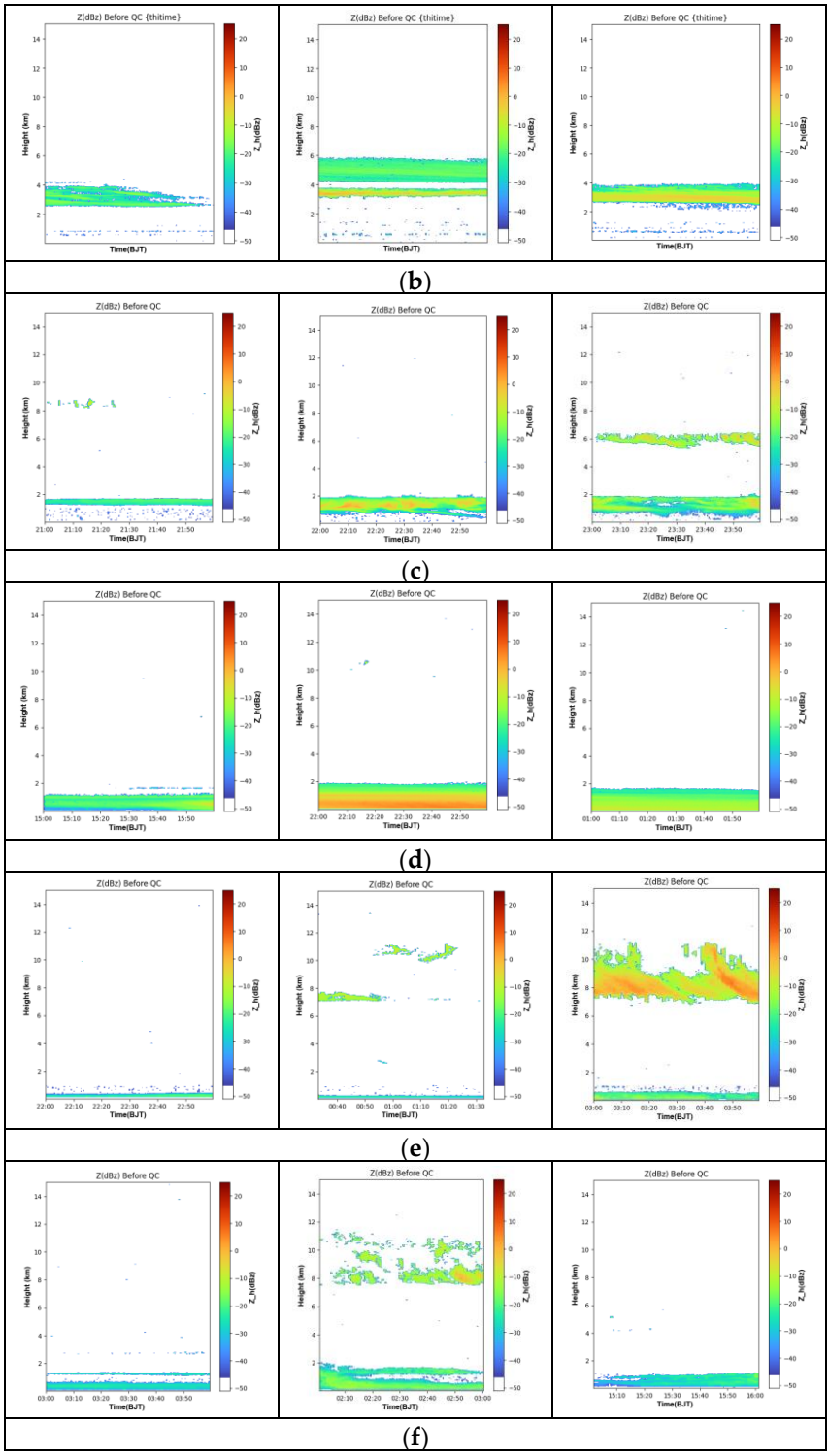


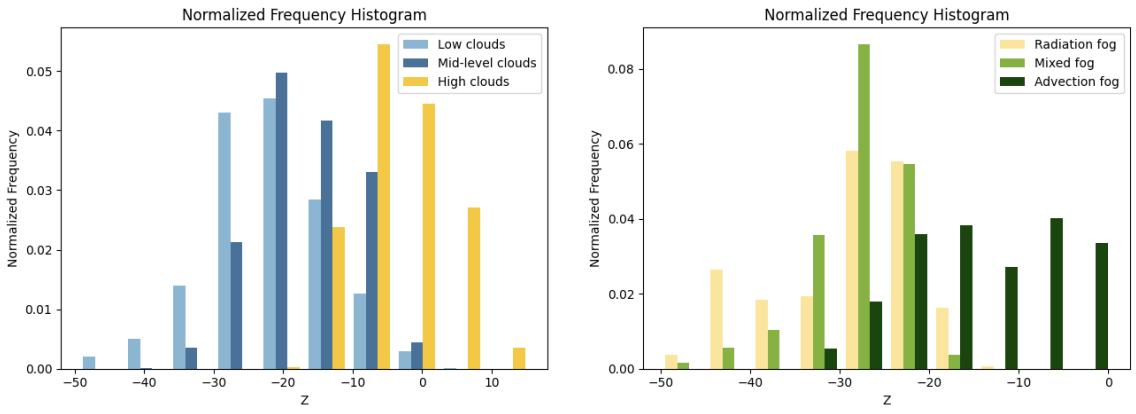
Figure 4. Cloud and fog weather dataset examples in THI scanning mode: reflectivity factor data over a 1-hour period, with the maximum range set to 15 km on the vertical axis. (a) High clouds: Clouds with a base height reaching 6 km or above. (b) Medium clouds: Clouds with a base height between 2 and 6 km. (c) Low clouds: Clouds whose base height is less than 2 km. (d) Advection fog: Frequently observed in coastal areas where the radar is located. (e) Radiation fog: Appears from midnight to early morning, reaching its peak before sunrise and dissipating thereafter. (f) Mixed fog: Combines the characteristics of both advection fog and radiation fog.

Based on the divided datasets, the distribution characteristics of the reflectivity factor of the echo were statistically analyzed. In view of the differences in the total number of reflectivity factor samples,

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 November 2024 doi:10.20944/preprints202411.2219.v1

to ensure the comparability of the data, we used the normalized frequency as the vertical axis to accurately reflect the number of samples corresponding to different reflectivity factors.

Delving into the reflectivity factor distribution characteristics of different cloud layers and fog types, Figure 5 provides an intuitive overview of the data distribution. For low cloud samples, the distribution of the reflectivity factor shows a relatively low concentration trend, mainly concentrated between -50 dBZ and -10 dBZ. Low clouds, due to their loose structure and small droplets or ice crystal particles, result in weaker scattering of the electromagnetic waves emitted by the radar, which is reflected in the lower reflectivity factor. The reflectivity factor distribution of medium cloud samples shifts towards positive values, mainly concentrated between -30 dBZ and 0 dBZ, indicating that the density and particle size of droplets or ice crystals in medium clouds may be increased, thereby enhancing their scattering ability for electromagnetic waves. For high cloud samples, the distribution of the reflectivity factor further shifts towards positive values, ranging from -20 dBZ to 15 dBZ. This distribution characteristic may be attributed to the more complex microphysical structure inside high clouds, including the coexistence of water condensates of different sizes, densities, and phases, which responds more diversely to electromagnetic waves. In terms of fog types, the reflectivity factor of radiation fog samples is mainly concentrated between -50 dBZ and -17 dBZ, showing a lower reflectivity characteristic. The formation of radiation fog echoes often starts from the ground and gradually extends to the upper layers. With the increase in height, the reflectivity factor shows a gradual weakening trend, the height of the echo is usually kept below 1 kilometer, and its ability to scatter radar waves is relatively weak. The reflectivity factor distribution of advection fog samples shows different characteristics, mainly concentrated between -35 dBZ and 0 dBZ. This indicates that advection fog may contain more and larger droplets or ice crystal particles, enhancing its scattering and reflection effects on radar waves. It is worth noting that mixed fog samples have characteristics of both advection fog and radiation fog, so the reflectivity factor is particularly concentrated in the interval of -30 dBZ to -20 dBZ. At the same time, the top and bottom heights of the meteorological echo are recorded from the statistical samples, and the calculation method is the serial number of the distance gate where the top or bottom of the meteorological target echo is located, multiplied by the distance gate resolution of 15 meters.



(**a**) (**b**)

Figure 5. Reflectivity factor distribution of weather processes: (a) Reflectivity factor distribution of low, medium, and high clouds. (b) Reflectivity factor distribution of radiation fog, advection fog, and mixed fog.

Statistical examples of all meteorological echo data and summaries of the reflectivity factor, average echo-top height, and average echo-bottom height data are shown in Tables 2 and 3.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 November 2024 doi:10.20944/preprints202411.2219.v1

Table 2. Statistical characteristics of cloud echo.

Cloud Type	Low Cloud	Medium	High Cloud
Reflectivity (dBZ)	-49.5 to 0.5	-39.5 to 2.5	-21 to 15.5
Average echo-top height (km)	1.91	5.88	12.24
Average echo-bottom height (km)	0.17	2.02	6.01

Table 3. Statistical characteristics of fog echo.

Fog Type	Radiation Fog	Mixed Fog	Advection Fog
Reflectivity (dBZ)	-49.0 to -13.0	-43.0 to -15.5	-34.0 to 0.5
Average echo-top height (km)	0.56	0.95	1.44
Average echo-bottom height (km)	-	-	0.23

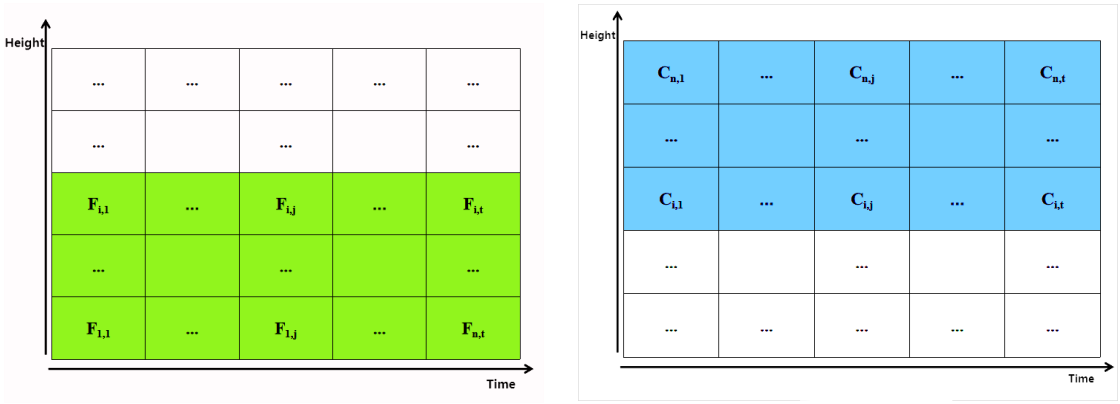
In this study, the observed instances of radiation fog and advection fog all showed the bottom of the fog echo being in direct contact with the ground, so the average echo-bottom height was not considered in the statistics of these two types of fog, and it was uniformly regarded as being in direct contact with the ground. For the identification problem of cloud and fog echoes, the first step is to set the reflectivity factor as the initial screening benchmark value. In view of the short duration of suspended echoes and isolated noise points compared to meteorological echoes, the duration was established as the second identification standard. Through the joint application of these two standards, it is already possible to effectively distinguish cloud and fog echoes from some suspended clutter.

Further analysis revealed a certain degree of overlap between fog echoes and cloud echoes in terms of the reflectivity factor and duration, making it difficult to clearly distinguish between cloud and fog echoes based solely on the reflectivity factor and duration. Observing the data in Tables 2 and 3, there is a clear and significant difference between the clouds and fog in terms of the height where the echo forms. Specifically, in this study, the average top height of fog echoes was up to 1.44 km, while the average top height of cloud echoes was at least 1.91 km. This characteristic difference provides a third key standard for the identification of cloud and fog echoes. Finally, utilizing the continuous distribution characteristics of meteorological echoes on the range gate, this method introduces the continuity count on the height gate—that is, the range gate continuity count—as the fourth judgment threshold.

Therefore, we comprehensively used these four radar parameters with significant statistical characteristics to identify and distinguish cloud and fog echoes.

Using the difference in echo-top height, the THI data were preliminarily processed for the separation of cloud and fog at 1.5 km. The data schematic is shown in Figure 6. For data below 1.5 km, fog echo identification was performed, and for data above 1.5 km, cloud echo identification was performed. Since cloud and fog echoes are weather phenomena that may occur simultaneously, when fog is identified, cloud echo identification should continue above the fog area.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 November 2024 doi:10.20944/preprints202411.2219.v1



(**a**) (**b**)

Figure 6. Schematic diagram of millimeter-wave radar THI data matrix partitioning: (a) Fog partition diagram. (b) Cloud partition diagram.

The next step is to judge the THI data in two parts using threshold classification: First, when the reflectivity factor data are within the threshold range, it is considered that the data may be meteorological echoes, i.e., valid data. Then, the first datum $D_{1,1}$ obtained from the valid data is used to judge whether they are continuous in the range gate direction. If the number of continuous valid range gates Hc exceeds the height judgment threshold (Ht), then $D_{1,1}$ is considered to be a range-valid datum and retained; otherwise, it is considered to be a suspended clutter point and filtered out, and the next datum is used for judgment. Assuming that $D_{1,1}$ is a range-valid datum, judgement continues in the time gate direction. If the number of continuous valid time gates Tc exceeds the time judgment threshold Tc, then $D_{1,1}$ is considered to be a completely valid datum and retained; otherwise, it is still considered to be suspended clutter and filtered out. The above steps are repeated until each datum in the THI data is judged.

In this study, the millimeter-wave radar range gate length was 15 meters, and the time gate interval was 30 seconds. In 2020, Sun Yi et al. statistically found that the average fog top height of the Yellow Sea is around 200 meters, and most of the fog top heights in the Yellow Sea are between 100 m and 40 m [51]. The types of sea fog that occur are mainly radiation fog. To retain as much meteorological echo as possible, the lowest fog top height of 100 meters is taken as the height judgment condition, i.e., 7 range gates are taken as the fog echo range gate continuity judgment condition; 225 meters is taken as the cloud height judgment condition, i.e., 15 range gates are taken as the fog echo range gate continuity judgment condition. For suspended point clutter, its duration is relatively short, while meteorological echoes usually last for a long time, and the duration of clouds, fog, and other weather phenomena can reach more than 2 hours [52]. Finally, to be able to judge the fog echo, 15 minutes (that is, 30 time gates) is taken as the cloud and fog echo time judgment condition. The threshold values used in this paper can be summarized as shown in Table 4.

Table 4. Statistical threshold value table for various parameters.

Weather Type	Cloud	Fog
Reflectivity threshold range (dBZ)	-40 to 15	-40 to 0
Echo-top height (km)	＞1.5	＞0.1 & ＜1.5
Range gate continuity count (counts)	15	7
Time gate continuity count (counts)	≥30	≥30

3.2. K-Means Clustering Method

We used the K-means clustering method to divide the cloud radar data into potential meteorological and non-meteorological echo classes. This section introduces the evaluation criteria

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 November 2024 doi:10.20944/preprints202411.2219.v1

for clustering effects, feature selection for clustering, determination of cluster number K, and integration and extraction of cluster information.

3.2.1. Clustering Effect Evaluation Criteria

The K-means algorithm is used to divide a set of N-dimensional datasets X into K non-intersecting clusters, and each cluster can be described by the mean of the samples in that cluster. This mean is usually referred to as the "centroid" of the cluster. Although these centroids are in the same N-dimensional space, they are usually not points selected from X. The hourly THI feature dataset X can be expressed in matrix form:

$$X_{120000\times5} = [z, v, w, t_c, h_c] \tag{6}$$

where z、v、w、t_c、h_c are the column vectors of the reflectivity factor, radial velocity, radial velocity spectrum width, time gate continuity, and range gate continuity data, respectively. The specific feature parameter selection is outlined in Section 3.2.2. The goal of the K-means algorithm is to ensure a "small intra-cluster difference and large inter-cluster difference", so the distance between each data point in each cluster and the centroid of that cluster, as well as the centroids of other clusters, can be used to measure the difference inside and outside the cluster. The result obtained by calculating the sum of the distances of each point in each cluster is usually called the inertia index. Therefore, the K-means algorithm can reflect the effects of K-means clustering by minimizing the inertia index. The expression for this process is as follows:

$$\sum_{i=0}^{n} \min_{n\subset C_i} (\|x_i - \mu_i\|^2) \tag{7}$$

where C_i is the i-th class, μ_i is the clustering center of cluster C_i, and x_i is a data object in the dataset X. In this paper, a single x_i object and the clustering center μ_i of each cluster all have 5-dimensional information, i.e.,

$$\mu_i = \left(\overline{z_{C_1}}, \overline{v_{C_1}}, \overline{w_{C_1}}, \overline{t_{c_{C_1}}}, \overline{h_{c_{C_1}}}\right) \tag{8}$$

$$x_i = \left(z_i, v_i, w_i, t_{c_i}, h_{c_i}\right) \tag{9}$$

where $\overline{z_{C_1}}, \overline{v_{C_1}}, \overline{w_{C_1}}, \overline{t_{c_{C_1}}}, \overline{h_{c_{C_1}}}$ are the means of the five parameters of cloud radar data in cluster C_i, while $z_i, v_i, w_i, t_{c_i}, h_{c_i}$ are the data points of the current object x_i in the five dimensions of cloud radar data.

In addition, the silhouette coefficient is also one of the most commonly used clustering algorithm evaluation indicators; it is defined for each sample, enabling it to simultaneously measure the similarity of the sample to other samples in its own cluster to samples in other clusters. Therefore, the silhouette coefficient can be expressed as follows:

$$S_i = \frac{b_i - a_i}{\max(a_i, b_i)} \tag{10}$$

where a_i is the average distance between sample point x_i and all other points in the same cluster, while b_i is the average distance between sample point x_i and all points in the nearest cluster. Therefore, the silhouette coefficient ranges from -1 to 1. A value of S_i approaching 1 indicates that the sample has a high degree of similarity with other samples in its own cluster and a significant difference from samples in other clusters. When the similarity of the sample to those outside the cluster exceeds the similarity to those in its own cluster, the silhouette coefficient will be negative.

We used the elbow method corresponding to the inertia index and the silhouette coefficient to select the number of clusters, providing information support for subsequent clustering analysis and multi-threshold algorithm optimization [53]. The determination of the optimal value of K is based on the following considerations: Firstly, for the elbow method, which relies on calculating the sum of the squared errors (SSE), as the number of clusters K increases, the sample division becomes finer

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 November 2024 doi:10.20944/preprints202411.2219.v1

and the cohesion of each cluster gradually improves, leading to a natural decrease in the SSE. When K is less than the optimal number of clusters, an increase in K will significantly enhance the cohesion of each cluster, so the decline range of SSE will be large. When K reaches the optimal number of clusters, the return in terms of cohesion obtained by further increasing K will rapidly decrease, resulting in a sharp reduction in the decline range of SSE, and then it tends to be flat as K continues to increase. In other words, the relationship graph between SSE and K takes on an elbow shape, and the K value corresponding to the elbow is considered to be the optimal number of clusters for the data. Secondly, S_i combines the cohesion and separation of clusters to evaluate the clustering effect. This coefficient ranges from -1 to 1, and a larger value indicates a better clustering effect. When the distances between samples within a cluster are closer and the distances between samples of different clusters are farther, the average silhouette coefficient becomes larger, signifying better clustering. In the case of clustering with millimeter-wave cloud radar data as inputs, the closer the distances between samples within a cluster, the closer the radar characteristic parameters and spatiotemporal distribution characteristics of that range bin. Thus, the K with the largest average silhouette coefficient is regarded as the optimal number of clusters.

Through a large number of sample experiments in this study, the optimal number of clusters was determined to be 3. Moreover, considering the ultimate goal of distinguishing between meteorological and non-meteorological echoes, in Section 3.3, we detail how the algorithm utilizes the information of valid data points to enable these 3 clusters to separate meteorological echoes from non-meteorological ones and extract cluster information for correcting statistical thresholds.

3.2.2. Cloud Radar Clustering Data Preprocessing

Data preprocessing of cloud radar data is a key link before the K-means clustering step. This step aims to ensure the high quality and standardization of the input data, thereby optimizing the effect of training the clustering model. The preprocessing steps specifically include data normalization and feature selection.

Data normalization: When dealing with raw data from different scales or different sources, data standardization can greatly reduce the negative impact of differences in the data's scale and range on model performance. For example, the cloud radar reflects a large difference in reflectivity factors when facing precipitation weather processes and cloud and fog processes; the distribution ranges of radial velocity and velocity spectrum width are different. Normalization is a method of converting a numerical column to a standard scale, with the purpose of preventing certain specific features from having a negative impact on the model due to differences in their value ranges. In this study, we used the Min–Max scaler to rescale each feature to the interval [-1, 1]. That is, the minimum value in the dataset was mapped to -1, the maximum value was mapped to 1, and all other values were adjusted proportionally. Using this scaling method can preserve the original distribution of the data, ensuring that the relationships between values remain unchanged.

Feature selection: Not all features in the K-means model training can make a good contribution to the model's predictive ability. Some features may affect the training effect of the model and may even be redundant. The addition of these features may even lead to substandard training performance. Therefore, it is necessary to add feature selection work before using cloud radar data for clustering. The random forest is a supervised learning algorithm that constructs multiple decision trees and provides an average prediction value for each regression task. This method is very effective in feature selection, and one of its main advantages is the ability to calculate feature importance scores to show the contribution of each feature to the prediction [54]. The built-in estimation function of the random forest can measure the importance of each feature to the target variable [55]. We constructed a random forest model to perform feature selection for the K-means training model. We initially imported a suite of eight characteristic variables into the random forest model for the purpose of importance ranking: the reflectivity factor Z, radial velocity V, velocity spectrum width W, linear depolarization ratio LDR, height information H, temporal information T, range gate continuity count Hc, and time gate continuity count Tc. The outcome of this ranking is depicted in Figure 7. Finally, we selected the top five features with the highest importance as the inputs for the K-means clustering

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 November 2024 doi:10.20944/preprints202411.2219.v1

model. Although these five features were considered in the K-means clustering process, we noted in the cloud and fog feature statistics process that radial velocity V and velocity spectrum width W did not show obvious statistical differences, indicating that their effectiveness in using single parameters to distinguish different cloud and fog types is limited. Therefore, when correcting the statistical threshold range, we chose to only use the reflectivity factor Z, height range gate continuity count Hc, and time gate continuity count Tc information extracted from the K-means clustering.

The clustering information of the K-means algorithm contains the individual information of the current THI data, combining the clustering information that shows the characteristics of a single data feature with the statistical information of the data based on the overall sample. Therefore, the information extracted in the previous section was used in the next step to correct the statistical thresholds in Table 4, improving the fixed multi-threshold algorithm.

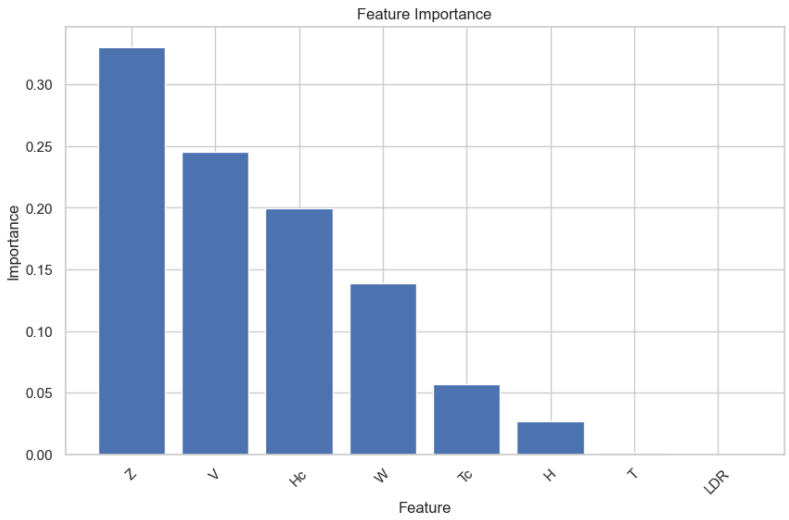


Figure 7. Model feature importance ranking (related to W-cloud radar parameters).

3.3. K-Means Improved Multi-Threshold Clutter Filtering Algorithm

3.3.1. K-Means Clustering Information Extraction

We used the non-meteorological echo class information from the K-means clustering results for extraction, and the process is shown in Figure 8, with specific steps as follows:

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 November 2024 doi:10.20944/preprints202411.2219.v1

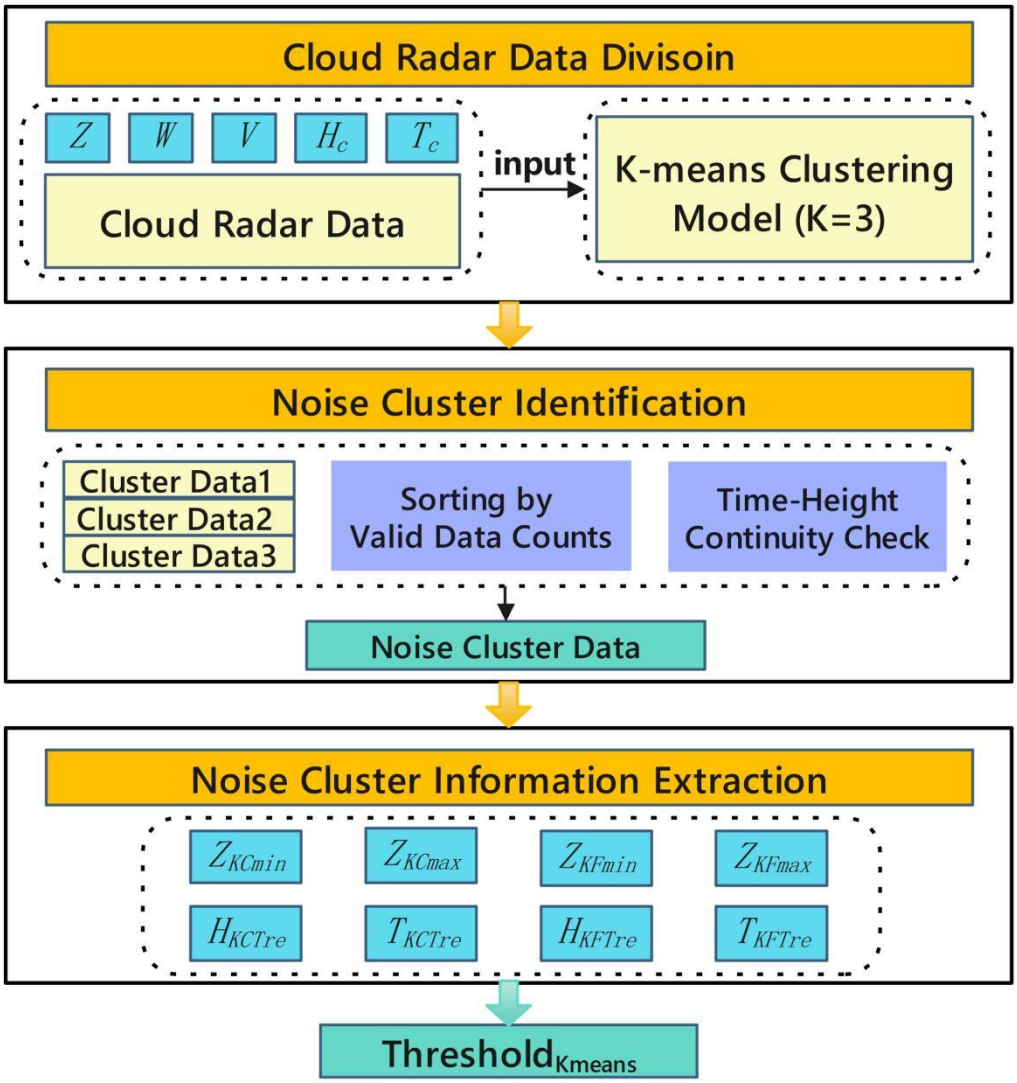


Figure 8. Flowchart for extracting K-means non-meteorological echo cluster information.

Cluster division: Using the elbow method and the silhouette coefficient method, we adopted the optimal number of clusters K = 3 for most data. Through the random forest model, the 8 features mentioned in the previous section were sorted by importance, and the top five features with the greatest contribution to the model prediction were selected as the inputs for the K-means clustering model. The K-means model divides the reflectivity factor data of the cloud radar into 3 clusters.

Extraction of non-meteorological echo clusters: For the THI reflectivity factor data, the distance gate with echo was considered to represent valid data. First, we sorted the valid data points of the 3 clusters extracted in the first step, and we selected the cluster with the fewest valid data points as the potential non-meteorological echo cluster data, denoted as THI_{min}. Then, we checked the time and range gate continuity T_C and H_C for the clusters with the next two valid data points THI_{min} and THI_{second}. The specific method was to calculate the total sum of the continuity of the top 100 data points in time and range gate continuity for these two clusters, denoted by Ct_{min} and Ct_{second}, with the following calculation formula:

$$Ct_{min} = \sum_{i \subset M_{Top100}} Tc_i + Hc_i \quad (11)$$

$$Ct_{second} = \sum_{i \subset S_{Top100}} Tc_i + Hc_i \quad (12)$$

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 November 2024 doi:10.20944/preprints202411.2219.v1

where M_{Top100} and S_{Top100} are the top 100 data points in time and range gate continuity for the two clusters, respectively, and Tc_i and Hc_i are the time gate continuity and range gate continuity of the i-th data point, respectively. This was aimed at preventing the meteorological echo clusters from being erroneously judged as non-meteorological echoes on account of the insufficient number of valid data points under certain circumstances. If $Ct_{min} > Ct_{second}$, it indicates that although THI_{second} has more valid data points, its spatiotemporal continuity is smaller; this implies that it behaves more like non-meteorological echoes, which are typically less continuous in nature compared to genuine meteorological echoes. Moreover, having a smaller spatiotemporal continuity means that it may not follow the regular patterns that we would expect from clouds or fog in terms of how the echoes are distributed over time and across different range gates. Therefore, we exchanged the data of THI_{second} and THI_{min}, and the newly generated THI_{min} was the cluster of non-meteorological echoes.

Extraction of non-meteorological echo information: Statistically, according to the THI data obtained in the second step, we extracted the distribution range of the reflectivity factor, the median of the range gate continuity, and the median of the time gate continuity in the non-meteorological echo cluster, where Z_{Cmin} and Z_{Cmax} are the minimum and maximum values of the THI reflectivity factor above 1 km in the cloud area, respectively; Z_{Fmin} and Z_{Fmax} are the minimum and maximum values of the THI reflectivity factor below 1 km in the fog area, respectively; H_{CTre} and H_{FTre} are the height judgment thresholds for the cloud area and fog area, respectively; and T_{CTre} and T_{FTre} are the time judgment thresholds for the cloud area and fog area, respectively.

In the previous two chapters, we discussed in detail the importance of cloud radar data in meteorological detection, along with the necessity of clutter filtering. We also analyzed the limitations of traditional multi-threshold clutter filtering methods in depth, including the unsuitability of thresholds after moving the radar position, as well as the loss of internal information in the current data. It is against this background that we propose an improved multi-threshold clutter filtering method based on K-means clustering. This method mainly involves several key steps: Firstly, it extracts relevant thresholds such as reflectivity factor thresholds, height continuity judgment thresholds, and time continuity judgment thresholds from the K-means clustering data. Then, these extracted thresholds are utilized to correct the original ones. After that, with the corrected thresholds, the method determines valid data segments and calculates the top height of the echo for each profile in order to distinguish between cloud and fog areas and perform clutter filtering. Figure 9 illustrates the specific process of the multi-threshold clutter filtering algorithm improved by K-means clustering.

The reflectivity factor thresholds Z_{KCmin}, Z_{KCmax}, Z_{KFmin}, and Z_{KFmax}, height continuity judgment thresholds H_{KCTre} and H_{KFTre}, and time continuity judgment thresholds T_{KCTre} and T_{KFTre} are extracted from the K-means clustering data. The thresholds extracted from the clustering results can be used to correct the original thresholds, effectively filtering out suspended clutter while retaining as many of the meteorological echo characteristics as possible. The final thresholds obtained by the correction method are as follows:

$$threshold_{QC} = \frac{threshold_{Kmeans} + threshold_{statistic}}{2} \quad (13)$$

where $threshold_{QC}$ represents the final threshold values, $threshold_{Kmeans}$ refers to the thresholds extracted from the K-means clustering results in Section 3.2, and $threshold_{statistic}$ refers to the thresholds based on the statistics in Table 4.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 November 2024 doi:10.20944/preprints202411.2219.v1

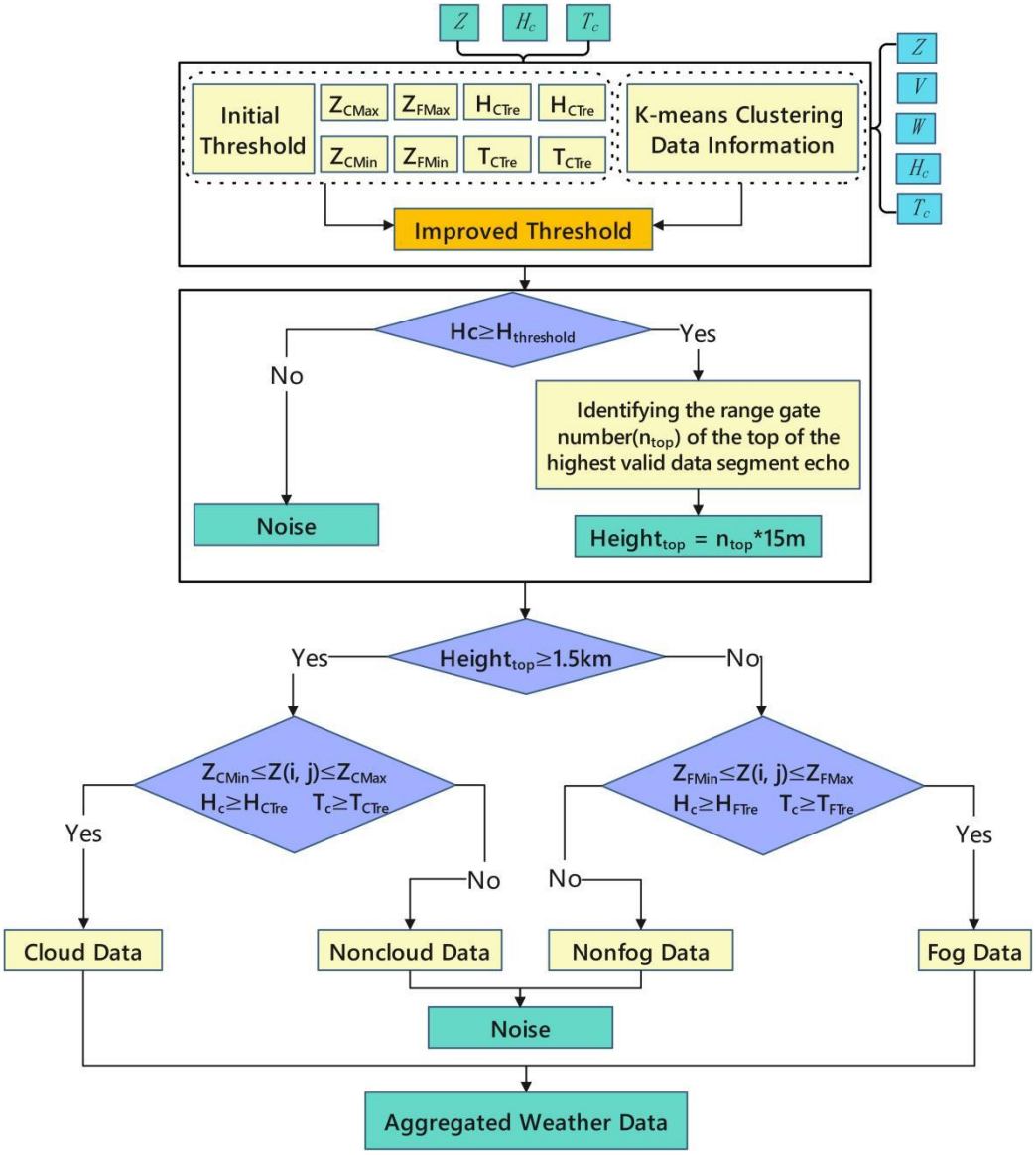


Figure 9. Flowchart of the K-means multi-threshold clutter filtering algorithm.

After the clustering information is extracted, the first step is to determine the valid data segments and calculate the top height of the echo for each profile. This step serves two purposes: to perform clutter filtering processing for cloud and fog areas, and to use the top height of the echo to distinguish low-level clouds and fog. For the fog area echoes below 1.5 km, the top height of the echo segment with the highest range gate continuity and reaching the fog area's height continuity threshold is taken as the valid echo top height ($Height_{top}$). If Heighttop reaches 1.5 km, the part of the data in the fog area will be merged into the cloud area for threshold judgment, and the second-highest top height of the valid data segment in this profile will serve as the fog top height. For the cloud area echoes above 1.5 km, the top height of the echo segment with the highest range gate continuity and reaching the cloud area's height continuity threshold is taken as the valid echo-top height ($Height_{top}$). In the next step, the cloud area echoes and fog area echoes that meet the requirements are selected according to the corrected thresholds for the fog area and cloud area, respectively, as the identification results of cloud and fog weather phenomena. Finally, the cloud area echoes and fog area echoes are merged to generate complete cloud and fog weather echo data. At the same time, the data that do not meet the threshold requirements will be judged as non-meteorological echoes, serving as the identification results for noise.

4. Experiments

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 November 2024 doi:10.20944/preprints202411.2219.v1

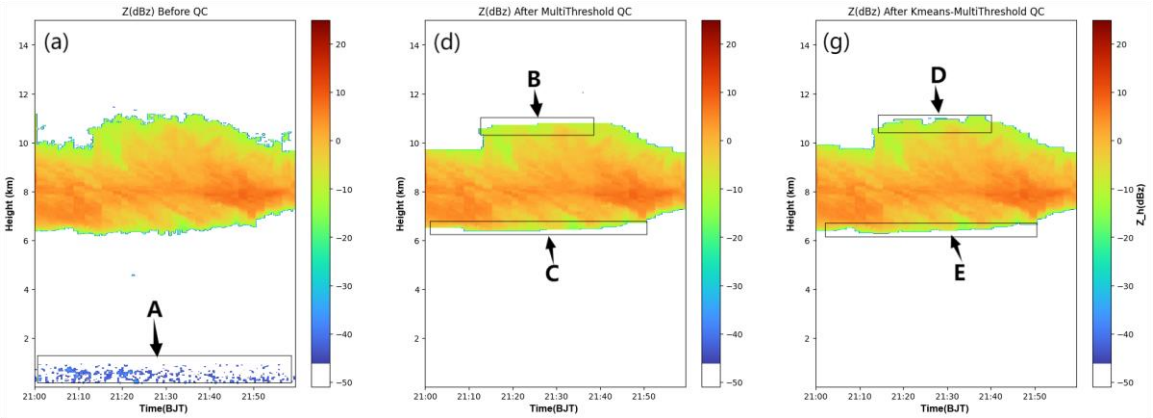
Using the identification method described in Section 3.3, the cloud and fog echoes were identified based on the reflectivity factor, range gate continuity, and time gate continuity. Three typical cases were used to verify the reliability of the clutter filtering method proposed in this paper, and the performance differences between the multi-threshold method and the K-means improved multi-threshold method were compared.

4.1. Cloud Case

Figure 10 shows the effect comparison of the cloud case clutter filtering before and after the multi-threshold method, and after the K-means improved multi-threshold method, at 21:00-22:00 on July 6, 2023, observed by the W-band cloud radar over Qingdao, Shandong.

From the reflectivity factor data before clutter filtering in Figure 10 (left), it can be seen that there is a layer of non-cloud and fog echoes below 1 km at location A. The statistical results of the filtered echoes show that the reflectivity factor intensity of this part of the echoes is concentrated between -48 dBZ and -29.5 dBZ; the radial velocity is concentrated between -2.8 m/s and 1.2 m/s; the radial velocity spectrum width is concentrated between 0.1 and 0.19 m/s; the time gate continuity is mostly below 5 gates, calculated as one time gate of 15 seconds, indicating that the majority of non-meteorological echoes received by the radar do not last more than 2 minutes; and the range gate continuity is mostly below 7 range gates, with the radar's range gate resolution of 15 meters, indicating that the continuous height distribution of the bottom non-meteorological echoes does not exceed 100 meters, and only a few echoes have a range gate continuity exceeding 10 gates.

After the multi-threshold clutter filtering, the main body of the cloud echo is retained, and the non-cloud echoes and isolated scattered point echoes below the cloud layer are effectively filtered out. However, there is some loss of cloud top and bottom edge information at locations B and C in Figure 10 (middle). After the K-means clustering improved multi-threshold clutter filtering, while filtering out non-cloud echoes and isolated scattered point echoes, more echoes at the cloud top and bottom locations are retained, and more cloud boundary information is obtained, as seen at locations D and E in Figure 10 (right). The reflectivity factor intensity of the remaining echoes after this method's clutter filtering is concentrated between -19.5 dBZ and 9 dBZ, the radial velocity is mainly distributed between -3 m/s and -1.4 m/s, and the radial velocity spectrum width is mainly between 0.08 m/s and 0.9 m/s, indicating that, under the combined action of the vertical movement speed of air and particles, the cloud echoes show more upward movement compared to the filtered bottom non-cloud echoes, and the speed range fluctuates more.



Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 November 2024 doi:10.20944/preprints202411.2219.v1

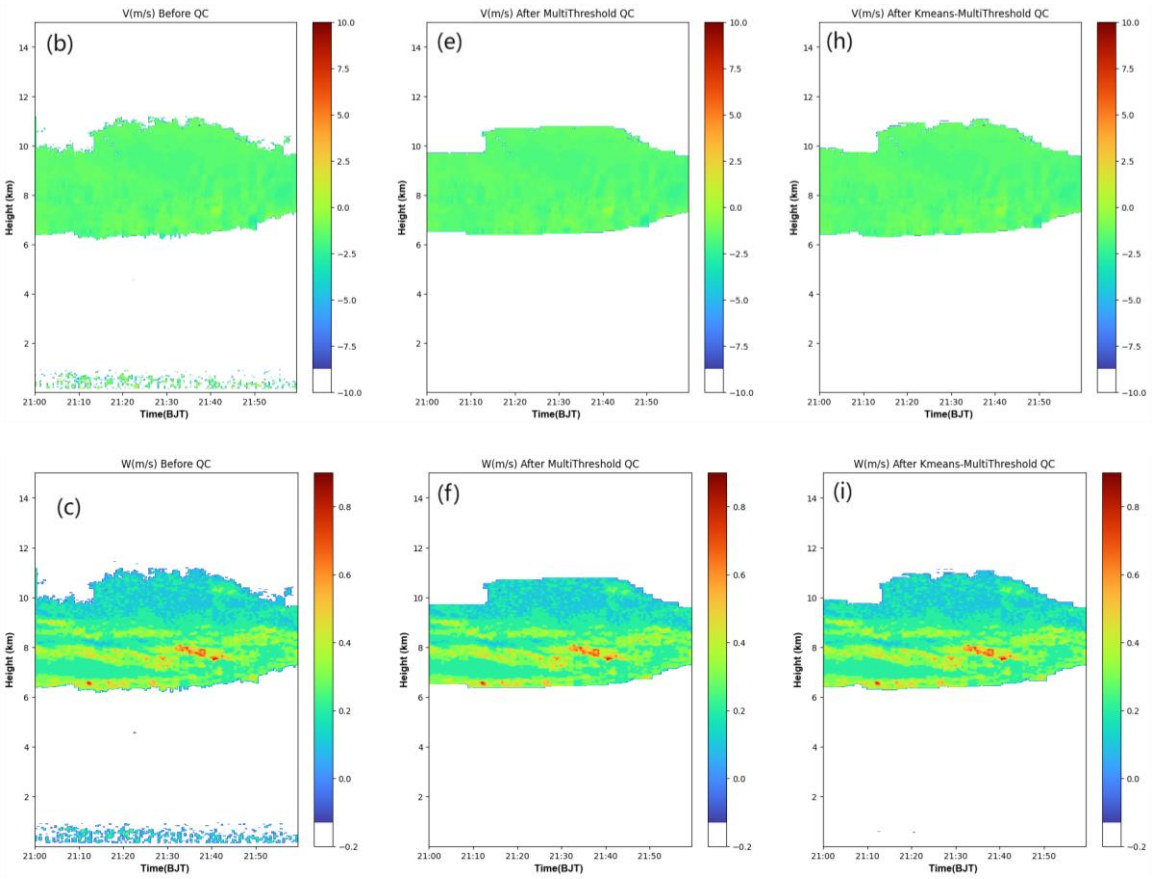


Figure 10. Comparison of echoes before and after multi-threshold clutter filtering and K-means improved clutter filtering for a cloud case observed by a W-band cloud radar over Qingdao, Shandong, from 21:00 to 22:00 on July 6, 2023: (a,b,c) Before clutter filtering: Location A shows non-cloud and fog echoes (<1 km). Echo traits such as reflectivity, velocity, spectrum width, and gate continuities are presented. (d,e,f) After multi-threshold clutter filtering: Main cloud echo retained, but losses in cloud top/bottom edge information at locations B and C. (g,h,i) After K-means improved clutter filtering: At locations D and E, more top/bottom echoes retained, and better boundary information obtained.

4.2. Marine Fog Case

Figure 11 presents a comparison of the effects of the clutter filtering of fog echo data before and after the application of the multi-threshold method, and after the application of the multi-threshold method improved by K-means clustering, at 22:00-23:00 on July 3, 2023, observed by the W-band cloud radar over Qingdao, Shandong.

From location A in Figure 11 (left), it can be seen that there is a layer of non-fog echoes above the fog top. These echoes have ~20 time slots directly connected to the fog top, affecting the calculation of its height. The statistical results of this part of the echoes show that the reflectivity factor intensity is distributed between -47 dBZ and -10.5 dBZ; the radial velocity is distributed between -3.4 m/s and 3.2 m/s; the velocity spectrum width range is 0.1 m/s to 0.2 m/s; the time gate continuity of the clutter below 3 km is distributed between 1 and 4 gates, i.e., the radar receives echoes with a duration of no more than 3 minutes and 30 seconds; and the height gate continuity is concentrated below 9 range gates, where the echoes with a height gate continuity between 7 and 9 gates have a time gate continuity of 1.

After the multi-threshold clutter filtering, the fog echoes are retained, and the non-cloud echoes and isolated scattered point echoes above the fog layer are effectively filtered out. However, there is some loss of fog top edge information at location B. From 21:53:30 to 22:59:30, the fog top echoes

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 November 2024 doi:10.20944/preprints202411.2219.v1

between 405 m and 435 m were filtered out as non-meteorological echoes because the echo time continuity threshold did not reach the preset range gate threshold.

As shown at location C in Figure 11, the K-means improved multi-threshold method extracted the spatiotemporal continuity of the clutter. In this case, the corrected range gate continuity threshold was rounded up to 7, and the corrected time gate continuity threshold was 7. This method separated and filtered out the non-cloud and fog echoes and isolated scattered point echoes, and the additional 30 meters of fog top echo information at location B was retained. The reflectivity factor of the remaining fog echoes after this method's clutter filtering was concentrated between -37.5 dBZ and -13 dBZ, the radial velocity was distributed between -0.4 m/s and 0.2 m/s, and the velocity spectrum width was distributed between 0.1 m/s and 0.3 m/s. The quality-controlled echoes showed more concentrated reflectivity factor intensity and more stable radial velocity (close to 0 m/s) compared to the filtered echoes.

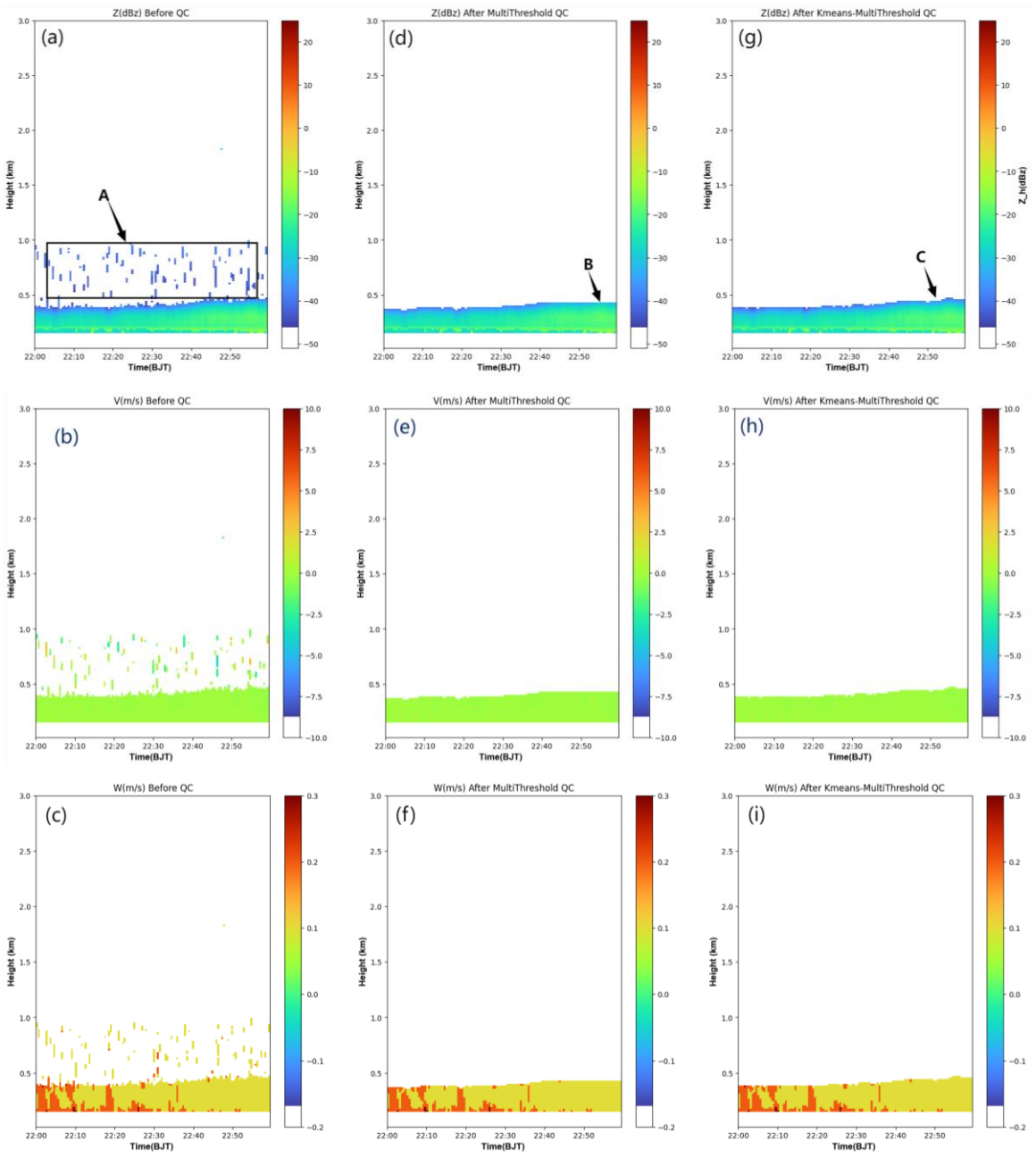


Figure 11. Comparison of echoes before and after multi-threshold clutter filtering and K-means improved clutter filtering for a fog case observed by a W-band cloud radar over Qingdao, Shandong, from 22:00 to 23:00 on July 3, 2023: (a,b,c) Before clutter filtering: The presence of a layer of non-fog

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 November 2024 doi:10.20944/preprints202411.2219.v1

echoes can be seen above the fog top at location A. Echo characteristics include specific ranges of reflectivity factor, radial velocity, and velocity spectrum width, along with details on time and height gate continuities, impacting the height calculation of the fog top. (d,e,f) After multi-threshold clutter filtering: At location B, fog echoes are retained, while non-cloud and isolated scattered point echoes above the fog layer are filtered out. However, there is some loss of fog top edge information due to the echo time continuity threshold not meeting the preset range gate threshold. (g,h,i) After K-means improved multi-threshold clutter filtering: At location C, it can be seen that this method extracts the spatiotemporal continuity of the clutter; it corrects thresholds and retains additional fog top echo information compared to location B.

4.3. Cloud and Marine Fog Coexistence Case

Figure 12 shows the effect comparison of the cloud–fog coexistence echo clutter filtering before and after the multi-threshold method, and after the K-means improved multi-threshold method, at 16:00-17:00 on June 17, 2023, observed by the W-band cloud radar over Qingdao, Shandong.

As shown at location A in Figure 12 (left), there is a layer of scattered non-cloud and non-fog echoes at the height where the echoes exist. Echoes at location B and the cloud boundary also have suspended isolated echoes, among which ~15 time slots are directly connected to the fog top echoes. The statistical results of this part of the echoes show that the reflectivity factor intensity is distributed between -42.5 dBZ and -7 dBZ, the radial velocity range is -6.4 m/s to 6.4 m/s, the velocity spectrum width is between 0.1 m/s and 0.6 m/s, the time gate continuity of the clutter below 3 km is mainly below 20 gates, and the height gate continuity is mainly below 40 range gates.

After the multi-threshold clutter filtering, as shown in Figure 12 (middle), the main body of the two layers of cloud echoes and the low-altitude fog echoes are retained, and the non-cloud echoes and isolated scattered point echoes are effectively filtered out. However, while filtering out the non-meteorological echoes, more cloud top and bottom edge information is lost at locations C and D. At location D, which is key to inferring the base height of the first layer of clouds in this case, 464 range gates of cloud bottom echoes at 4.83 km height were lost from 16:00 to 16:40, and the average echo loss height during this period reached 87 m.

As shown in Figure 12 (right), after the K-means improved multi-threshold clutter filtering, this method also filtered out non-meteorological echoes and isolated scattered point echoes, and the cloud top, middle, and bottom information was better retained, as seen at locations E and F in Figure 12. Among them, at location F, the average echo loss height from 16:00 to 16:40 was reduced to 32 m compared to the original echo data. For the fog top information below 1 km, this method retained an additional 118 range gates of fog top echoes—an average of ~15 m more retained over the hour. The reflectivity factor of the low-altitude fog echoes remaining after this method's clutter filtering was concentrated between -39 dBZ and -18 dBZ; the radial velocity was distributed between -0.6 m/s and 0.6 m/s; the velocity spectrum width was distributed between 0.1 m/s and 0.5 m/s; the minimum time gate continuity was only 5 gates, which is the fog top echo at 600 m in the fog area, and it was retained after the threshold correction by the clustering information. The reflectivity factor of the cloud echoes above after this method's clutter filtering was concentrated between -19.5 dBZ and 3.5 dBZ; the radial velocity was distributed between -1.8 m/s and 0.4 m/s; the velocity spectrum width was distributed between 0.1 m/s and 1 m/s; the spatiotemporal continuity distribution of the cloud echoes during this hour was relatively large—5 to 82 time gates and 239 range gates, i.e., 2 to 41 minutes, and 75 m to 3.59 km. The reflectivity factor intensity distribution interval of the remaining cloud and fog echoes after clutter filtering was more concentrated, corresponding to the statistical characteristics of clouds and fog, respectively. The vertical movement of air and particles in the fog area consisted of both ascending and descending movements close to 0 m/s. While the radial velocity in the cloud area was positive, the overall movement trend of cloud particles was downward, which is consistent with the observed changes in the cloud bottom. The relative radial velocity of water particles in the clouds was greater than that in the fog area and non-meteorological echoes, broadening the spectrum and resulting in a relatively larger spectrum width, with the maximum spectrum width reaching 1 m/s.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 November 2024 doi:10.20944/preprints202411.2219.v1

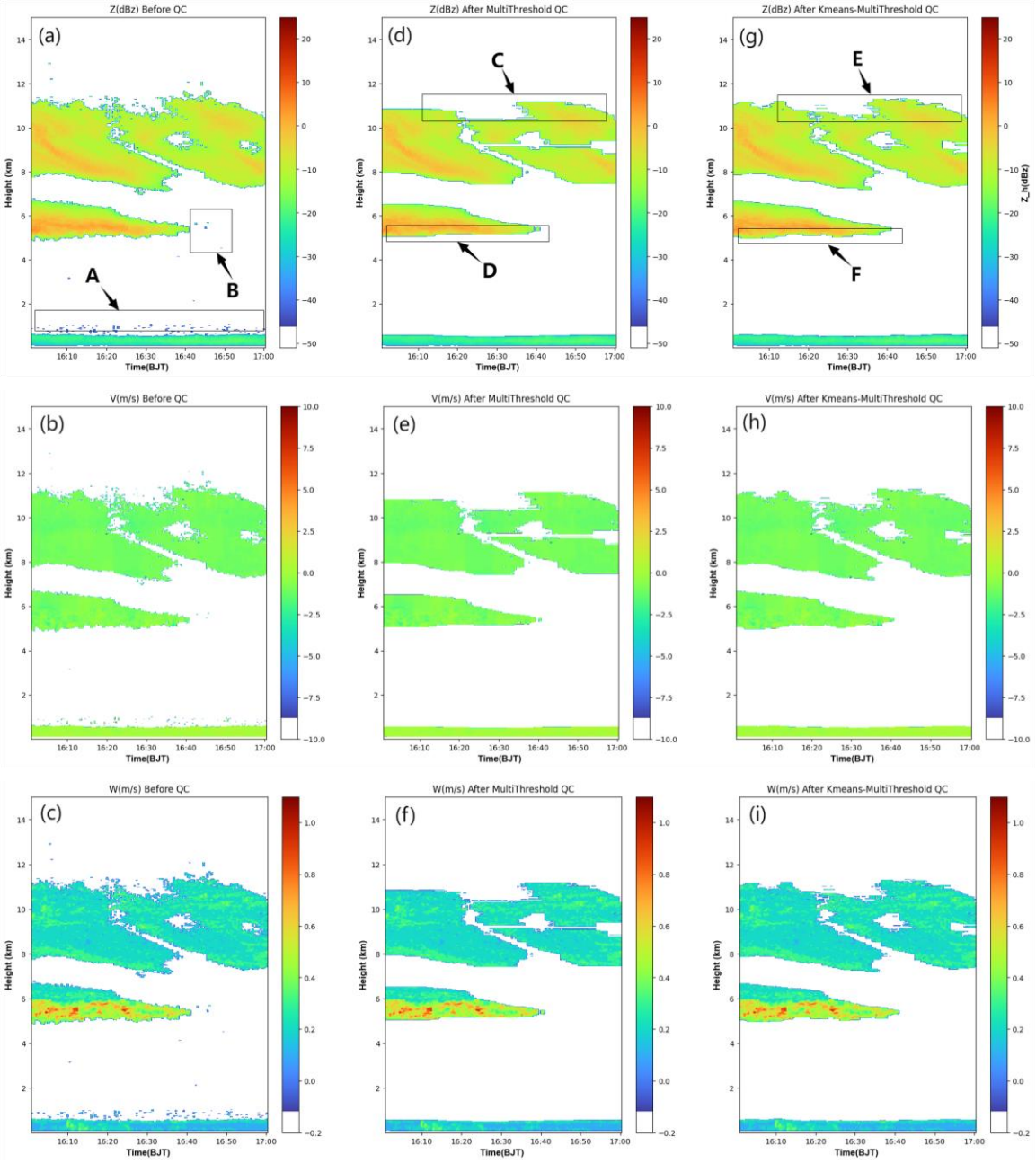


Figure 12. Comparison of echoes before and after multi-threshold and K-means improved clutter filtering for a cloud–fog coexistence case observed by a W-band cloud radar over Qingdao, Shandong from 16:00 to 17:00 on June 17, 2023: (a,b,c) Before clutter filtering: Location A shows scattered non-cloud and non-fog echoes. At location B and the cloud boundary, there are suspended isolated echoes linked to the fog top. The echo details include the reflectivity factor, radial velocity, spectrum width, and gate continuities. (d,e,f) After multi-threshold clutter filtering: Main cloud and low-altitude fog echoes retained, yet significant cloud top/bottom edge information lost at locations C and D (especially D), with notable echo height loss affecting the cloud base height inference. (g,h,i) After K-means improved multi-threshold clutter filtering: Better retention of cloud top/middle/bottom information at locations E and F. Reduced echo loss height. More echoes retained for the fog top.

5. Cloud and Fog Identification After Clutter Filtering

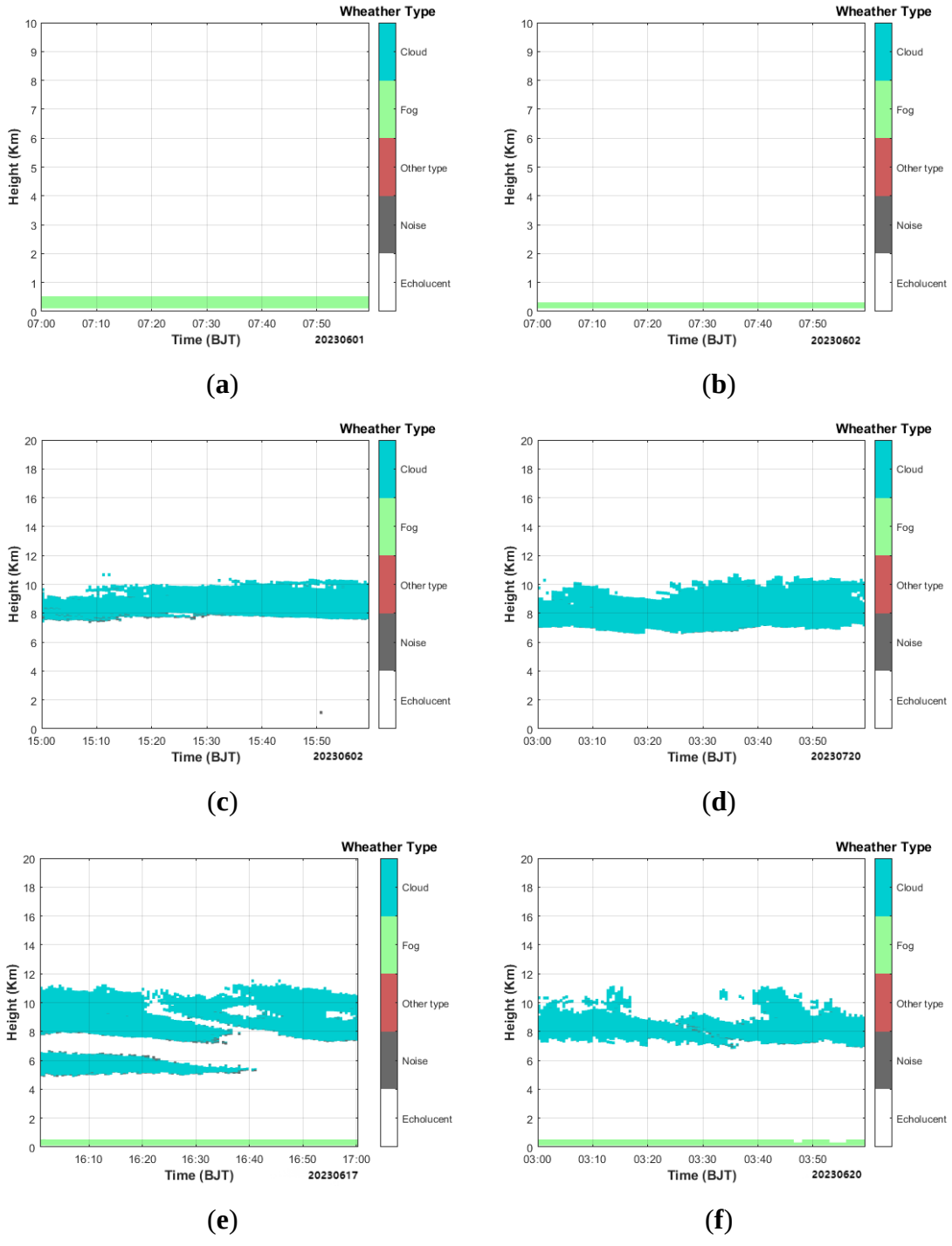
This section validates the cloud and fog echoes identified during the clutter filtering process. First, we use some cases to illustrate the identification results, as shown in Figure 13, and then Section 5.2 compares the identification results of the THI data from March to May 2023 with the observations of the observers and of the all-sky imaging instrument, in order to analyze the accuracy of the

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 November 2024 doi:10.20944/preprints202411.2219.v1

algorithm's identification results. This study classifies the echoes in the THI data into four categories: cloud, fog, noise, and other possible weather echoes that are not identified as cloud and fog echoes. Echolucent refers to the part where the radar does not receive echoes in actual observation, i.e., the area without echoes, which is not strictly a type of echo.

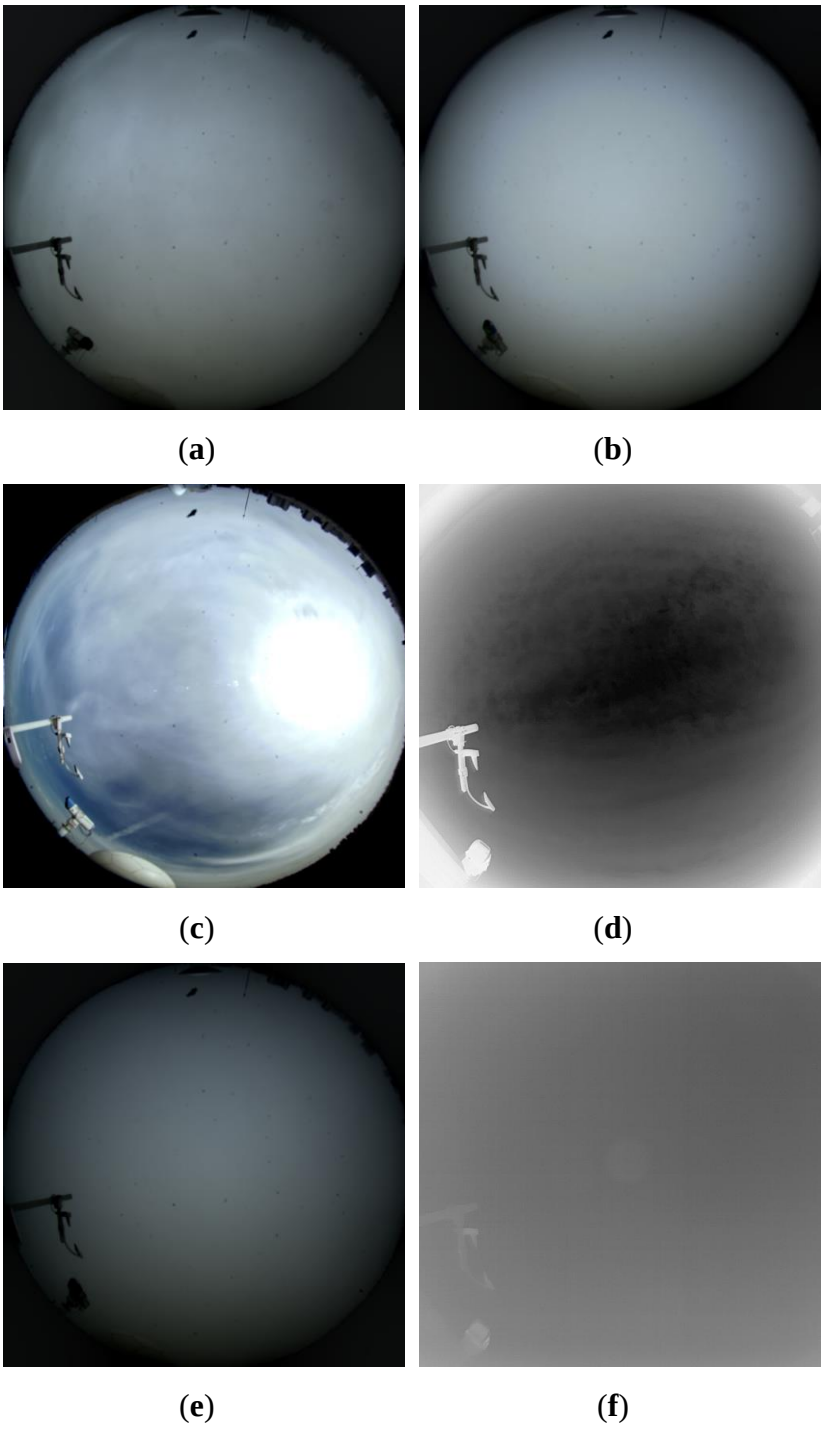
5.1. All-Sky Imaging Instrument Verification

Figure 13a,b show different moments of fog weather processes, and Figure 13c,d show cloud weather processes at different moments. Figure 13e,f show different moments containing both cloud and fog weather processes. The all-sky imaging instrument's observation results are shown in Figure 14, and the manual observation results for the corresponding moments are shown in Table 5.



Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 November 2024 doi:10.20944/preprints202411.2219.v1

Figure 13. Results of meteorological echo identification: (a) June 1st at 07:00; (b) June 2nd at 07:00; (c) June 2nd at 15:00; (d) July 20th at 03:00; (e) June 17th at 16:00; (f) June 20th at 03:00.

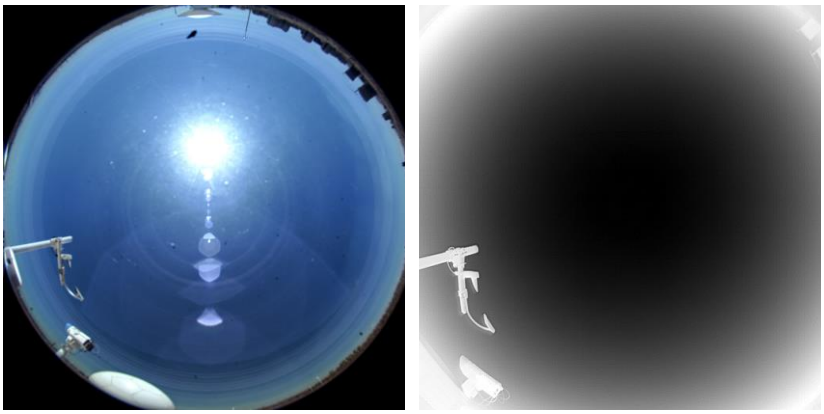


(a) **(b)**

(c) **(d)**

(e) **(f)**

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 November 2024 doi:10.20944/preprints202411.2219.v1



(g) (h)

Figure 14. All-sky imager observations in different weather conditions (Beijing time): (a) Visible-light image on June 1st at 07:00 under foggy conditions. (b) Visible-light image on June 2nd at 07:00 under foggy conditions. (c) Visible-light image on June 2nd at 15:00 under cloudy conditions. (d) Infrared image on July 20th at 03:00 under cloudy conditions. (e) Visible-light image on June 17th at 16:00 under conditions of both fog and clouds. (f) Infrared image on 20 June at 03:00 under conditions of both fog and clouds. (g) Visible-light image on August 23th at 12:00 under clear-sky conditions. (h) Infrared image on August 23th at 23:00 under clear-sky (nighttime) conditions.a.

For the moments at 7:00 on June 1, 2023, 7:00 on June 2, 2023, and 16:00 on June 17, 2023, fog echoes could be identified from the corresponding THI data, with all fog top heights below 1 km and durations reaching 1 hour. The following observations from the all-sky imaging instrument's visible-light images show that the visibility was low at 7:00 on June 1, 2023, 7:00 on June 2, 2023, and 16:00 on June 17, 2023, and there were no fine, dispersed, filamentous, or feathery cloud features, as shown in Figure 14a,b,e. At 3:00 in the early morning of June 20, 2023, clear visible-light images could not be obtained, so the all-sky imaging instrument used infrared images for observation (Figure 14f). To provide a clear contrast, we added two all-sky imaging instrument images taken under cloudless and fog-free conditions. Figure 14g shows the visible-light image taken under clear-sky conditions at noon on August 23, 2023. Figure 14h shows the infrared image taken on the same night at 23:00, with no clouds or fog present. The images clearly show that the imaging instrument's lens was completely covered by fog, and the contours of surrounding objects could not be clearly distinguished. At the same time, the observer's manual observation results also showed that there were fog processes at these four moments, as presented in Table 5. These results were consistent with the algorithm's identification results.

Table 5. Manual observation of weather conditions in relation to the all-sky imager.

Moment	Weather Conditions
7:00 on June 1, 2023	Fog
7:00 on June 2, 2023	Fog
15:00 on June 2, 2023	Cinot
16:00 on June 17, 2023	As op、Fog
3:00 on June 20, 2023	Sc tra、Fog
3:00 on July 20, 2023	Ci dens

For the moments at 15:00 on June 2, 2023 and 3:00 on July 20, 2023, cloud echoes could be identified from the THI data, with cloud base heights around 7.5 km and cloud top heights of ~10 km. The durations were more than 60 minutes, and the reflectivity factor was between -25 dBZ and -5 dBZ. The all-sky imaging instrument could only observe obvious cloud weather texture features at

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 November 2024 doi:10.20944/preprints202411.2219.v1

15:00 on June 2 and 3:00 on July 20 (Figure 14c,f). This is because the fog at 16:00 on June 17 and 3:00 on June 20 was relatively thick, causing visibility to decrease, and the imaging instrument's lens was obscured by fog, making it impossible to distinguish whether there was a cloud process. Therefore, it was necessary to rely on manual observation to obtain the actual weather conditions at the corresponding moments. The manual observation results for the above four moments were pseudo-cirrus (Ci not), overcast stratocumulus (As op), altostratus translucidus (Sc tra), and dense cirrus (Ci dens). Pseudo-cirrus clouds usually have a large area and dense cloud body. Overcast stratocumulus clouds have a thick cloud layer, from below which it is difficult to distinguish the Sun and Moon. Altostratus translucidus clouds are thinner, and the cloud layer is evenly distributed. Combining the all-sky images and manual observation results, the algorithm's identification results for the above four moments were essentially correct.

5.2. Manual Observation Record Comparison

Tables 6 and 7 show the identification results for cloud weather and fog weather, respectively. As mentioned earlier, only the data from 8:00, 14:00, and 20:00 each day were statistically analyzed. At the same time, since manual observation is greatly influenced by the subjectivity of the observer, in order to verify the actual observation data and exclude the influence of subjective factors, only the moments when the manual observation results were the same as the weather conditions observed by the all-sky imaging instrument were considered as actual observation data—i.e., the data statistically analyzed in "Combined observation" in the tables. When the algorithm's identification result was consistent with the combined observation result, it was considered to have achieved correct identification (PI, positive identification). When the combined observation result observed cloud or fog processes that the algorithm did not identify, it was considered to be missed identification (MI). When the combined observation result did not observe cloud or fog processes but the algorithm identified corresponding weather processes, it was considered to be misidentification (MIS).

Table 6. Cloud weather identification results.

Date	Manual Observation	All-Sky Imaging Instrument	Combined Observation	Positive Identification (PI)	Missed Identification (MI)	Misidentification (MIS)
2023.03	44	55	37	28	2	7
2023.04	52	58	47	40	4	3
2023.05	36	43	31	22	1	8
Total	132	156	115	90	7	18

Table 7. Fog weather identification results.

Date	Manual Observation	All-Sky Imaging Instrument	Combined Observation	Correct Identification (PI)	Missed Identification (MI)	Misidentification (MIS)
2023.03	41	10	10	7	1	2

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 November 2024 doi:10.20944/preprints202411.2219.v1

2023.04	25	15	13	10	1	2
2023.05	15	14	12	9	0	3
Total	81	39	35	26	2	7

In meteorology, the Probability of Detection (POD) and Missed Alarm Rate (MAR) are usually used to evaluate the results. POD and MAR can be expressed as follows:

$$POD=\frac{PI}{PI+MI}\times 100\% \quad (14)$$

$$MAR=\frac{MI}{PI+MI}\times 100\% \quad (15)$$

According to the above formulae, the POD for the algorithm in cloudy weather conditions is 92.78%, and the MAR is 7.22%. In foggy weather conditions, the POD of the algorithm is 92.86%, and the MAR is 7.14%. It is worth noting that the verification demonstrated herein is based on a relatively small data sample. In the experimental phase, constrained by the limited availability of data, our analyses could only be carried out on the data at specific time points (i.e., 8:00, 14:00, and 20:00 each day) and under specific conditions (where the results of the manual observations were consistent with the weather conditions observed by the all-sky imaging instrument). Consequently, this verification is essentially rather symbolic. Substantially expanding the number of cases for further verification would be of great significance for reinforcing the robustness and generalizability of our findings. Nevertheless, given the existing data limitations, we are fully aware of the constraints of this verification, and we look forward to enlarging the dataset in future research endeavors, so as to offer a more comprehensive and conclusive assessment of the algorithm's performance.

6. Discussion

This paper used W-band millimeter-wave cloud radar data obtained from the Huangdao District Meteorological Bureau in Qingdao, Shandong from March to July 2023. We aimed to address the data quality issues of non-meteorological echoes in cloud and fog detection by W-band cloud radar. An improved multi-threshold clutter filtering method based on the K-means algorithm was proposed, and the identification of cloud and fog detection was evaluated in detail. The main conclusions are as follows:

1. Non-meteorological echoes in the Qingdao Huangdao area are characterized by lower reflectivity factors and poor continuity in time and space. The cloud area and fog area were identified with thresholds of $-25 < Z < 15$ dBZ, $Hc \geq 15$, $Tc \geq 30$ and $-40 < Z < -5$ dBZ, $Hc \geq 7$, $Tc \geq 30$, respectively. The thresholds were corrected using the information of cloud and fog echoes obtained from the K-means clustering, which can filter out the majority of non-meteorological echoes. For cases where non-meteorological echoes are adjacent to cloud and fog echoes, this method can effectively separate non-meteorological echoes and retain echoes that perform in line with the characteristics of cloud and fog echoes. Validation with typical cloud and fog weather cases shows that the clutter filtering method is effective.
2. The K-means algorithm uses information from the reflectivity factor, radial velocity, spectrum width, height gate continuity count, and time gate continuity count to obtain the non-meteorological echo cluster. It effectively extracts the distribution of non-meteorological echoes in Z, Hc, and Tc, which can better guide the correction of thresholds based on statistics. This makes the clutter filtering method capable of filtering out clutter while retaining more edge information of cloud and fog echoes.
3. Through the evaluation using relevant formulae, in cloudy weather conditions, the algorithm showed a Probability of Detection (POD) of 92.78% and a Missed Alarm Rate (MAR) of 7.22%. In foggy weather conditions, the POD of the algorithm was 92.86%, and the MAR was 7.14%. It should be emphasized that the verification presented here relies on a relatively small data

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 November 2024 doi:10.20944/preprints202411.2219.v1

sample. During the experimental process, due to the limited availability of data, we were restricted to analyzing data at specific timepoints (8:00, 14:00, and 20:00 each day) and under specific conditions (when the manual observation results matched the weather conditions observed by the all-sky imaging instrument). As a result, this verification is rather symbolic in nature.

This study only used the W-band cloud radar data in the clustering process. In the future, more physical constraints—such as the radar power spectrum, temperature, and relative humidity data from automatic weather stations—could be incorporated for analysis. The K-means clustering combined with multiple thresholds has shown potential in the W-band millimeter-wave cloud radar clutter filtering method to better utilize the individual and statistical information of the echoes. Whether it is to better use the results of K-means clustering to correct thresholds, or to use other data analysis methods to more thoroughly utilize the information contained in the cloud radar data, is still a problem worth studying in the future. Meanwhile, it is worth noting that, in the exploration of this algorithm's performance in different weather conditions, such as cloud and fog, the differences in identification rates between them were found to reflect the impact of sample size. Specifically, the collection of sea fog sample cases is relatively difficult, resulting in a relatively small overall sample number in the calculation. Hence, slight changes in the number of missed or misidentified samples can significantly affect the calculation results of the corresponding identification rates. To enhance the reliability and generalizability of our findings regarding the clutter filtering algorithm, it will be crucial to increase the total number of samples with a wider variety of cases in future research, which would facilitate a more comprehensive and conclusive evaluation of the algorithm's performance and help us to obtain more accurate identification rates, as well as to better understand its effectiveness under diverse weather scenarios.

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

Funding: This work was supported by the National Key R&D Program of China (2021YFC3090203) and the National Meteorological Information Center surplus fund project (NMICJY202304).

Data Availability Statement: Data from W-band cloud radar were provided by the China Meteorological Administration; they are accessible to the authors but not publicly available.

Acknowledgments: The authors are deeply indebted and would like to extend their heartfelt gratitude to the Meteorological Bureau of Huangdao District, Qingdao City, Shandong Province for graciously providing the valuable data that served as the cornerstone for this paper. Their generous contribution was of utmost significance and enabled us to conduct in-depth research and analysis. Furthermore, we are sincerely appreciative of the reviewers for their astute and constructive comments, as well as their invaluable editorial suggestions. These inputs were instrumental in refining and enhancing the quality of this paper, guiding us through the process of improvement and ensuring that our work meets the high standards of academic rigor. Their efforts and insights have been truly indispensable and have played a crucial role in the final outcome of this research.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Wei, Z.; Lin, H.; Xin, M. Millimeter-wave meteorological radar's capability in cloud measurement. *Acta Meteorol. Sin.* **1985**, *3*, 378–383. DOI: 10.11676/qxxb1985.050.
2. Wang, J.; Wei, M.; Zhang, Q.; Li, X. Research progress of W-band millimeter-wave cloud radar technology. *Meteorol. Sci. Technol.* **2017**, *5*, 765–775. DOI: 10.19517/j.1671-6345.20160526.
3. Zhong, L.; Liu, L.; Ge, R.; Zhou, X. Introduction to the research and application of millimeter-wave cloud radar (HMBQ) in China. In *Proceedings of the 27th Annual Meeting of the Chinese Meteorological Society - Radar Technology Development and Application*; Chinese Meteorol. Soc., Ed.; [Unspecified Publisher]: [Publisher Location], China, **2010**; pp. 114–125.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 November 2024 doi:10.20944/preprints202411.2219.v1

4. Li, Z.H. Studies of fog in China over the past 40 years. *Acta Meteorol. Sin.* **2001**, *5*, 616–624. DOI: 10.11676/qxxb2001.065.
5. Tan, J.; Qian, C.; Huang, B.; et al. The Current Status and Future Perspectives of Sea Fog Monitoring and Forecasting in China. *Adv. Meteorol. Sci. Technol.* **2024**, *14* (2), 17–23, 32. DOI: 10.3969/j.issn.2095-1973.2024.02.003.
6. Li, H. Application of Ground-Based Remote Sensing Vertical Observation System in Meteorological Observation Operations. *Henan Sci.* **2023**, *42* (22), 115–119. DOI: 10.19968/j.cnki.hnkj.1003-5168.2023.22.022
7. Kollias, P.; Albrecht, B. The Turbulence Structure in a Continental Stratocumulus Cloud from Millimeter-Wavelength Radar Observations. *J. Atmos. Sci.* **2000**, *57*, 2417–2434. DOI: https://doi.org/10.1175/1520-0469(2000)057<2417:TTSIAC>2.0.CO;2
8. Liu, G.; Huang, S.; Liang, Y.; et al. Application of millimeter-wave radar in port sea fog observation and visibility inversion. *Arid Meteorol.* **2019**, *37* (6), 993–1004.DOI:10.11755/j.issn.1006-7639(2019)-06-0993
9. Hu, S.; Wang, Z.; Zhang, X.; et al. Analysis of sea fog echo characteristics and visibility inversion using millimeter-wave radar. *Meteorol.* **2022**, *48* (10), 1270–1280.DOI: 10.7519/j.issn.1000-0526.2022.042601
10. Gao, L. Design and Implementation of W-Band Cloud Radar System. National University of Defense Technology, 2018.
11. Wu, J.; Yang, L.; Dou, F.; et al. Research on the ice cloud detection capability of spaceborne terahertz dual-frequency cloud radar. *Infrared Millim. Waves* 2020, 39(06), 718–727.DOI:10.11972/j.issn.1001-9014.2020.06.009
12. Ge, J.; Wang, J.; Wang, J.; et al. Development and application of millimeter-wave meteorological radar. *J. Nanjing Univ. Aeronaut. Astronaut.* **2018**, *50*(05), 577–585. DOI: 10.16356/j.1005-2615.2018.05.001
13. Pazmany, A.L.; McIntosh, R.E.; Kelly, R.D.; et al. An airborne 95 GHz dual polarization radar for cloud studies. *IEEE Trans. Geosci. Remote Sens.* **1994**, *32*(4), 731–739.DOI: 10.1109/36.298002
14. Clothiaux, E.E.; Miller, M.A.; Albrecht, B.A.; et al. An evaluation of a 94 GHz radar for remote sensing of cloud properties. *J. Atmos. Oceanic Technol.* **1995**, *12*(2), 201–229.DOI: 10.1175/1520-0426(1995)012<0201:AEOAGR>2.0.CO;2
15. Sekelsky, S.M.; McIntosh, R.E. Cloud observations with a polarimetric 33 GHz and 95 GHz radar. *Meteorol. Atmos. Phys.* **1996**, *59*(1), 123–140.DOI:10.1007/BF01032004
16. Martner, B.E.; Moran, K.P. Using cloud radar polarization measurements to evaluate stratus cloud and insect echoes. *J. Geophys. Res.* **2001**, *106*(D5), 4891–4897.DOI:10.1029/2000JD900623.
17. Kollias, P.; Clothiaux, E.E.; Miller, M.A.; et al. Millimeter-wavelength radars: new frontier in atmospheric cloud and precipitation research. *Bull. Am. Meteorol. Soc.* **2007**, *88*(10), 1608–1624.DOI: https://doi.org/10.1175/BAMS-88-10-1608
18. Johannes, V.; Rambukkange, M.P.; Clothiaux, E.E.; et al. Arctic multilayered, mixed-phase cloud processes revealed in millimeter-wave cloud radar Doppler spectra. *J. Geophys. Res.: Atmos.* **2013**, *118*(23), 13199–13213. DOI :https://doi.org/10.1002/2013JD020183
19. Wada, E.; Hashiguchi, H.; Yamamoto, M.K.; et al. Simultaneous observations of cirrus clouds with a millimeter-wave radar and the MU radar. *J. Appl. Meteorol.* **2005**, *44*(3), 313–323.DOI:https://doi.org/10.1175/JAM2191.1
20. Turk, F.J.; Ringerud, S.E.; Camplani, A.; Casella, D.; Chase, R.J.; Ebtehaj, A.; Gong, J.; Kulie, M.; Liu, G.; Milani, L.; et al. Applications of a CloudSat-TRMM and CloudSat-GPM Satellite Coincidence Dataset. *Remote Sens*. **2021**, *13*, 2264. https://doi.org/10.3390/rs13122264
21. Beni, A.; Miccinesi, L.; Pagnini, L.; Cioncolini, A.; Shan, J.; Pieraccini, M. Interferometric Radars for Bridge Monitoring: Comparison among X-Bands, Ku-Bands, and W-Bands. *Remote Sens*. **2024**, *16*, 3323. https://doi.org/10.3390/rs16173323
22. Camplani, A.; Sanò, P.; Casella, D.; Panegrossi, G.; Battaglia, A. Arctic Weather Satellite Sensitivity to Supercooled Liquid Water in Snowfall Conditions. *Remote Sens*. **2024**, *16*, 4164. https://doi.org/10.3390/rs16224164
23. Liu, L. Reviews on retrieval methods for microphysical and dynamic parameters with cloud radar. *Torrential Rain Disasters* **2021**, *40*(3), 231–242.DOI: 10.3969/j.issn.1004-9045.2021.03.002
24. Ji, L.; Wang, Z.; Huang, X.; Zhang, P. Comparison study of attenuation correction methods for airborne W-band radar in different cloud types. *Trop. Meteorol.* **2018**, *34*(02), 260–267. DOI:10.16032/j.issn.1004-4965.2018.02.012
25. Wu, J.; Ma, C.; Chen, Q.; Liu, W.; Zhang, Q.; Wang, X. Analysis on cloud physical characteristics of a light rain process in Hefei detected by W-band cloud radar. *Meteorol.* **2018**, *44*(03), 416–424. DOI: 10.16032/j.issn.1004-4965.2018.02.012
26. Wu, J.; Wei, M.; Su, T.; Wang, X.; Li, Y.; Fan, Y. Comparison of the echoes detected by W-band and Ka-band cloud radars. *J. Mar. Meteorol.* **2017**, *37*(2), 57–64. DOI: 10.19513/j.cnki.issn2096-3599.2017.02.007
27. Bi, Y.; Huo, J.; Lü, D.; Su, T.; Wang, X.; Liu, B. Performance and observation analysis of the Ka and W dual-frequency millimeter-wave cloud radar in Yangbajing, Tibet. *Infrared Millim. Waves* **2022**, *41*(2), 470–482. DOI: 10.11972/j.issn.1001-9014.2022.02.014

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 November 2024 doi:10.20944/preprints202411.2219.v1

28. Boren, T. A.; Cruz, J. R. An artificial intelligence approach to doppler weather radar velocity de-aliasing. In Proceedings of the 23rd Conference on Radar Meteorology, Snowmass, CO, USA, (1986).
29. Cui, Z.; Cheng, M.; Wu, Q.; et al. A Technique of Fast Median Filtering and Its Application to Data Quality Control of Doppler Radar. *Plateau Meteorol.* **2005**, (05), 727–733.DOI:doi:10.5772/39000
30. Ma, Z.Y.; Zhu, C.Q.; Liu, X.M.; et al. Preliminary study on CINRAD radar data quality control methods. *Meteorol.* **2010**, *36*(8), 134–141. DOI: 10.7519/j.issn.1000-0526.2010.8.019
31. Görsdorf, U.; Lehmann, V.; Bauer-Pfundstein, M.; et al. A 35-GHz polarimetric Doppler radar for long-term observation of cloud parameters—description of system and data processing. *J. Atmos. Oceanic Technol.* **2015**, *32*, 675–690.DOI:https://doi.org/10.1175/JTECH-D-14-00066.1
32. Zheng, J.; Liu, L.; Zeng, Z.; et al. Ka-Band millimeter-wave cloud radar data quality control methods. *Infrared Millim. Waves* **2016**, *35*(6), 748–757.DOI:10.11972/j.issn.1001-9014.2016.06.018
33. Xiao, P. Ka-Band cloud radar data quality control and statistical analysis of vertical structure of clouds in Beijing area. Ph.D. Thesis, Chengdu University of Information Technology, Chengdu, China, 2018.
34. Wan, X.; Xu, G.R.; Wan, R.; et al. Analysis of the vertical structure characteristics of non-precipitating clouds observed by Yun radar on the east side of the Tibetan Plateau. *Torrential Rain and Disasters,* **2020**, 39(05), 442–450.DOI: 10.3969/j.issn.1004-9045.2020.05.002
35. Pang, N.T.; Michael, S.; Vipin, K. Introduction to Data Mining, 1st ed.; Addison-Wesley Longman Publishing Co., Inc.: USA, 2005.
36. Steinhaus, H. Sur la division des corps matériels en parties. *Bull. Acad. Polon. Sci*. **1957**, *4*(12), 801–804. MR0090073. Zbl 0079.16403.
37. MCQUEEN, J. Some methods for classification and analysis of multivariate observations. In Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, 1967, pp. 281–297.
38. Lloyd, S. P. Least squares quantization in PCM. *IEEE Trans*. Inf. Theory **1982**, *28*, 129–136.
39. Zhou, X.T.; Chu, X.; Yao, Z.P. A real-time air temperature dynamic quality control method based on K-means clustering. Meteorology 2012, 38(10), 1295–1300.DOI:10.7519/j.issn.1000-0526.2012.10.016
40. Chen, B.; Li, H.; Han, K. Short-term photovoltaic power forecasting method based on K-means algorithm and spiking neural networks. In Proceedings of the 8th Renewable Power Generation Conference (RPG 2019), Shanghai, China, 2019, pp. 1–6. DOI:10.1049/cp.2019.0278
41. Huang, Y.; Hou, Y.F.; Zhao, Z.Q. Application of clustering algorithm in thunderstorm warning of four supercell convective processes. *Mod. Inf. Technol*. **2024**, *8*(06), 145–148. DOI:10.19850/j.cnki.2096-4706.2024.06.031.
42. Kang, X.; Li, B.; Wu, L.; et al. Discrimination method of precipitation data from wind profiler radar based on K-means clustering analysis. *Meteorol. Sci. Technol.* **2013**, *41*(05), 818–824. DOI:10.19517/j.1671-6345.2013.05.005.
43. Li, Z.F.; Li, X.H.; Chen, K.H.; et al. Application of cluster analysis in phase identification of X-band dual-polarization radar. *Meteorol. Sci. Technol*. **2021,** *49*(03), 315–321. DOI:10.19517/j.1671-6345.20200203.
44. Ahmed, U.; Khan, A.R.; Mahmood, A.; Rafiq, I.; Ghannam, R.; Zoha, A. Short-term global horizontal irradiance forecasting using weather classified categorical boosting. *Appl. Soft Comput*. **2024**, *155*, 111441. DOI:10.1016/j.asoc.2024.111441.
45. Yin, Y.; Yong, Y.; Qi, S.; et al. Cluster analyses of tropical cyclones with genesis in the South China Sea based on K-means method. Asia-Pac. J. *Atmos. Sci*. **2023**, *59*, 433–446. DOI:10.1007/s13143-023-00322-8.
46. Tao, F. Research on particle spectrum retrieval and application of millimeter-wave cloud radar. Ph.D. Thesis, Nanjing University of Information Science and Technology, Nanjing, China, 2021.
47. Fang, L.; Li, Y.; Sun, G.; et al. Horizontal and vertical distributions of clouds of different types based on CloudSat-CALIPSO data. Clim. Environ. Res. 2016, 21(5), 547–556. DOI:10.3878/j.issn.1006-9585.2016.15240
48. Liu, C.; Gao, L.; Wang, X.; Yu, J.; Li, J.; Ye, W. Climatic characteristics analysis of heavy fog in Shandong Province. Shandong Meteorol. 2014, 34(2), 26–31. https://doi.org/10.19513/j.cnki.issn1005-0582.2014.02.005.
49. Lv, B.; Jia, B.; Han, F.; Xu, J.; Wang, J. Formation and maintenance mechanism of a persistent heavy fog in central and western Shandong. *Arid Meteorol.* **2014**, *32*(5), 830–836.DOI: 10.11755/j.issn.1006-7639(2014)-05-0830
50. Liang, W.; Hou, Z. Characteristics and forecast of heavy fog in Qingdao. Shandong Meteorol. 2001, 32(2), 12–17. DOI:10.19513/j.cnki.issn1005-0582.2001.02.005
51. Sun, Y.; Yang, Y.; Zhen, Q. The characteristics of the top height of sea fog over the Yellow Sea in spring and summer based on CALIPSO satellite data. *Mar. Forecast*. **2020**, *37*(3), 54–61.DOI:CNKI:SUN:HYYB.0.2020-03-007
52. Wang, X. X.; Liu, J. C.; Li, Y. Comparative study of winter fog and summer fog in the Yellow Sea. Meteorol. Res. Appl. 2022, 43(1), 96–100.DOI:10.19849/j.cnki.CN45-1356/P.2022.1.17
53. Syakur, M. A.; Khotimah, B. K.; Rochman, E. M. S. Integration of K-means clustering method and elbow method for identification of the best customer profile cluster. In Proceedings of the 2nd International Conference on Vocational Education and Electrical Engineering: ICVEE 2017, Surabaya, Indonesia, 9 November 2017; Institute of Physics: Surabaya, Indonesia, 2018; pp. 112–117.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 November 2024 doi:10.20944/preprints202411.2219.v1

54. Kursa, M. B.; Rudnicki, W. R. Feature selection with the Boruta package. *J. Stat. Softw*. **2010**, *36*, 1–13.
55. Boutahir, M. K.; Farhaoui, Y.; Azrour, M.; Zeroual, I.; El Allaoui, A. Effect of feature selection on the prediction of direct normal irradiance. *Big Data Min. Anal.* **2022**, *5* (4), 309–317.DOI: 10.26599/BDMA.2022.9020003

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.