ADVANCES IN ATMOSPHERIC SCIENCES, VOL. 36, AUGUST 2019, 793–803

• **Original Paper** •

Cloud Classification and Distribution of Cloud Types in Beijing Using Ka-Band Radar Data

Juan HUO*, Yongheng BI, Daren LÜ, and Shu DUAN

Key Laboratory for Atmosphere and Global Environment Observation, Chinese Academy of Sciences, Beijing, China 100029

(Received 27 December 2018; revised 15 April 2019; accepted 23 April 2019)

ABSTRACT

A cloud clustering and classification algorithm is developed for a ground-based Ka-band radar system in the vertically pointing mode. Cloud profiles are grouped based on the combination of a time–height clustering method and the *k*-means clustering method. The cloud classification algorithm, developed using a fuzzy logic method, uses nine physical parameters to classify clouds into nine types: cirrostratus, cirrocumulus, altocumulus, altostratus, stratus, stratocumulus, nimbostratus, cumulus or cumulonimbus. The performance of the clustering and classification algorithm is presented by comparison with all-sky images taken from January to June 2014. Overall, 92% of the cloud profiles are clustered successfully and the agreement in classification between the radar system and the all-sky imager is 87%. The distribution of cloud types in Beijing from January 2014 to December 2017 is studied based on the clustering and classification algorithm. The statistics show that cirrostratus clouds have the highest occurrence frequency (24%) among the nine cloud types. High-level clouds have the maximum occurrence frequency and low-level clouds the minimum occurrence frequency.

Key words: clouds, clustering algorithm, classification algorithm, radar, cloud type

Citation: Huo, J., Y. H. Bi, D. R. Lü, and S. Duan, 2019: Cloud classification and distribution of cloud types in Beijing using Ka-band radar data. *Adv. Atmos. Sci.*, **36**(8), 793–803, https://doi.org/10.1007/s00376-019-8272-1.

Article Highlights:

- A systematic cloud clustering and classification algorithm for ground-based Ka-band radar in the vertically pointing mode is developed.
- A comparison experiment shows that 92% of the cloud profiles are clustered successfully and 85% of the cloud clusters are correctly classified.
- In Beijing, the occurrence frequency of cirrostratus clouds is the highest among the nine cloud types, whereas nimbostratus clouds have the lowest occurrence frequency.

1. Introduction

Clouds have diverse features, shapes and spatial scales, and therefore have different effects on solar and infrared radiation; they are indicative of both short-term weather conditions and long-term climate change (Ramanathan et al., 1989; Molinari and Dudek, 1992; Rossow and Schiffer, 1999; Hamilton, 2006).

Radar has powerful strengths in the exploration of the internal structure of clouds. Large amounts of in-situ observational data (from ground- or air-based sensors) have been analyzed to determine the empirical relations between cloud microphysical parameters and the radar reflectivity factor Z_e (hereinafter referred to simply as reflectivity) for different phases of cloud particles and different cloud types (Sekelsky et al., 1999; Mace and Sassen, 2000; Austin et al., 2009; Heymsfield et al., 2014; Liu et al., 2014). The cloud type determines which relation is suitable for practical retrievals. The Cloud Profiling Radar system (CPR) on CloudSat uses an iterative forward physical model to retrieve microphysical parameters (Rodgers, 1976, 1990, 2000). However, an a-priori set according to cloud type is required to initialize the iteration process, and therefore knowledge of the cloud type is useful in the selection of the initial parameters for the retrieval of cloud microphysical parameters.

Research on the distribution of cloud types is a basic part of meteorological and climatological studies. There have been many cloud classification algorithms developed for different measurement sensors, mostly passive remote sensors such as imagers and radiometers (Ebert, 1987; Baum et al., 1997; Miller and Emery, 1997; Sassen and Wang, 2008; Li et al., 2014). Radar has a number of advantages over passive remote sensing in the accurate measurement of the phys-

* Corresponding author address: Juan HUO
Email: huojuan@mail.iap.ac.cn

ical properties of clouds, such as the determination of cloud height, depth and particle size. Radar therefore provides a potentially improved method for cloud classification, especially for clouds with multiple layers. Despite this advantage, there has been no systematic work on cloud clustering and classification algorithms for ground-based radar systems in the vertically pointing mode. One goal of this study was to develop a cloud clustering and classification algorithm using an 8 mm wavelength radar (Ka-band radar) system.

Section 2 presents a brief introduction to Ka-band radar. Sections 3 and 4 describe the principle and methodology of the cloud clustering and classification algorithm. The performance of the algorithm is presented in section 5 through comparison experiments. A study of the distribution of cloud types in Beijing is presented in section 6, followed by a summary and discussion of our results in section 7.

2. Data

2.1. *Ka-band radar*

A Ka-band polarimetric Doppler radar system was installed on the roof of a laboratory building at the Institute of Atmospheric Physics (39.967°N, 116.367°E) in 2010. This radar system measures the reflectivity, the Doppler velocity, the spectral width and the linear depolarization ratio (Table 1). The Ka-band radar has been working 24 hours a day since October 2012, except for unexpected accidents. It operates in a vertically pointing mode except during special events, such as heavy rain or short-term joint observations with other instruments.

A data quality control approach combining the threshold and median filter methods has been implemented to reduce the effects of clutter and noise on the radar reflectivity (Xiao et al., 2018). Scatter and absorption in gases, water vapor, clouds and precipitation all attenuate radar reflectivity (Kollias et al., 2014). We calibrate the attenuation due to water vapor using the equation

$$Z_{\mathrm{e}}(r) = \frac{Z_{\mathrm{m}}(r)}{e^{-0.2\ln(10)\int_0^f k_{\mathrm{H_2O}}(s)\mathrm{d}s}}, \tag{1}$$

where r is the distance from the radar system, Z_{m} is the measured reflectivity, and $k_{\mathrm{H_2O}}(s)$ is the absorption coefficient of water vapor at the distance of s and pre-calculated for each month based on humidity data obtained from 2015 to 2017. The attenuation in clouds, precipitation and other gases has not yet been calibrated. However, these attenuations are limited because the radar system observes in a vertically pointing mode and the r should be within 16 km. For example, Doviak and Zrnić (1993) reported that two-way gaseous attenuation in a standard atmosphere should be no more than 0.5 dB within 25 km. Evaluations of the Doppler velocity, spectral width and linear depolarization ratio are being performed, but the accuracies are not yet available. They are not used in our algorithm.

2.2. *Temperature*

The temperature has a large effect on the cloud particle phase and size. There is no simultaneous measurement of temperature in the radar data. The temperature data used in this study are from the European Centre for Medium-Range Weather Forecasts (ECMWF) ERA-Interim dataset (http://apps.ecmwf.int/datasets/data/interim-full-daily/levtype=sfc/). The ERA-Interim dataset presents six-hourly temperature profiles with a $0.125° \times 0.125°$ spatial resolution. The temperature profile of the grid (40.0°N, 116.25°E) closest to the radar location is linearly interpolated and used in our classification algorithm.

3. Methodology of cloud clustering

The vertically pointing mode of the radar sequentially records the reflectivity. It is difficult to determine the cloud type of a single cloud profile as a result of the diversity of clouds. Discrete cloud profiles need to be grouped together before classification.

A radar profile is considered to be cloudy if there are more than three bins with a reflectivity higher than −45 dBZ. Cloud clusters usually appear in three patterns in a reflectivity image (Fig. 1): (1) the first cloud cluster ends some time before the appearance of the second cloud cluster (Fig. 1a); (2) the second cloud cluster appears soon after the first cloud cluster, but at a different height layer (Fig. 1b); and (3) the first cluster does not end before the start of the second cluster (Fig. 1c). The first pattern shows a break in the timing of the two cloud clusters and the second pattern illustrates a break in height, whereas the third pattern shows an overlap in both time and height.

According to these three patterns, we built a cloud clustering algorithm based on a combination of the time–height clustering method and the k-means clustering method. Figure 2 shows a flow chart for the clustering algorithm. First, a cloud profile (marked as p_i, where $i = 0, 1\ldots$) is selected to record the time, height and layer number. If the height gap

Table 1. Technical specifications of the Ka-band radar.

Ka-band radar					
Transmitter		Antenna		Receiver	
Frequency	35.075 GHz	Diameter	1.5 m	Noise	5.8 dB
Peak power	29 kW	Gain	54 dB	Noise power	−103 dBm
Pulse length	0.2 μs	Beam width	0.4°		
Type	Magnetron	Scanning mode	Vertically pointing, Plan Position Indicator, Range-Height Indicator	Vertical resolution	30 m

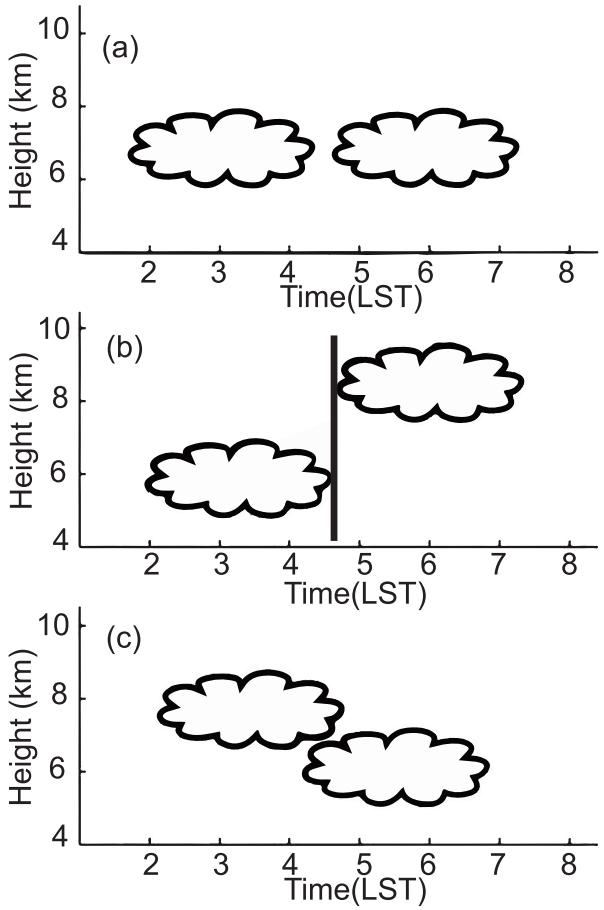

Fig. 1. Three patterns in the distribution of cloud clusters: (a) the first cloud cluster ends before the second cloud cluster appears; (b) the second cloud cluster appears soon after the first cloud cluster, but at a different height layer; and (c) the first cloud cluster overlaps the appearance of the second cluster.

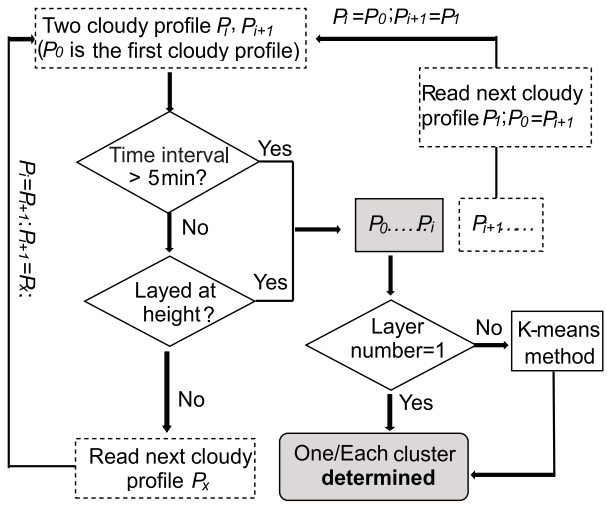

Fig. 2. Flow chart of the cloud clustering algorithm.

between two cloudy bins is > 150 m (five bins), then the cloud is regarded as layered; one height gap indicates two layers, two height gaps correspond to three layers, and so on. If the time interval between two profiles (p_i and p_{i+1}) is more than five minutes (Fig. 1a) or if the cloud heights are layered (Fig. 1b), then this represents the end of a cloud cluster (marked as C_0 and containing profiles $p_0 \dots p_i$). The last profile of C_0 is p_i, whereas p_{i+1} is the first profile of the other cluster. This process groups cloud profiles with a break in time or height. We refer to this process as the time–height clustering method. The start time, end time and maximum layer number are recorded for cluster C_0. If the maximum layer number is 1, then a single cloud cluster is determined (Figs. 1a and b). If the layer number is > 1 (Fig. 1c), the following steps are taken.

For cloud clusters with multiple cloud layers, we use the *k*-means clustering algorithm, which aims to partition *n* points into *k* clusters in which each point belongs to a cluster with the nearest mean distance. Here, the point is a cloudy radar bin. The *k*-means clustering algorithm contains four steps: (1) select *k* points randomly as the initial centers of *k* clusters; (2) calculate the distances of each point from all centers and allocate the point to the cluster with the nearest distance; (3) calculate the centers of the new clusters formed after step (2); and (4) repeat steps (2) and (3) until the centers no longer change. In the *k*-means clustering process, the cloud layer number is used as the a priori of *k*. The time and height are the two input parameters in the distance calculation. After every cloud cluster has been determined, each cluster is input to the classification algorithm for classification of the cloud. Figure 3b shows an example of a cloud cluster marked with different colors in which the black cross indicates the center of each cluster.

Incorrect clustering is inevitable because the features of clouds are variable and radar noise or clutter always exists. According to our analysis, there are two main causes of incorrect clustering. One is the use of an inaccurate layer number in the *k*-means clustering algorithm (Figs. 3c and d). The clouds in Fig. 3c are actually single clusters, but the layer number is mistakenly determined as three due to minor gaps in the clouds. When a wrong layer number is input into the *k*-means clustering algorithm, wrong clustering results are output (Fig. 3d). Another cause of incorrect clustering is the *k*-means algorithm itself, which clusters clouds incorrectly when they show an unexpected distribution (Figs. 3e and f). The clouds in Fig. 3e have three layers and the lowest level clouds only last for three minutes—far shorter than the highest level clouds. The difference in time scale between the two cloud layers outweighs the difference in height, leading to incorrect clustering (Fig. 3f). We performed a statistical analysis using observational data from January to June 2014 and found that 88% of the clouds that required the *k*-means method were clustered successfully.

4. Cloud classification

4.1. *Background*

The World Meteorological Organization has recently published revised principles of cloud classification for ob-

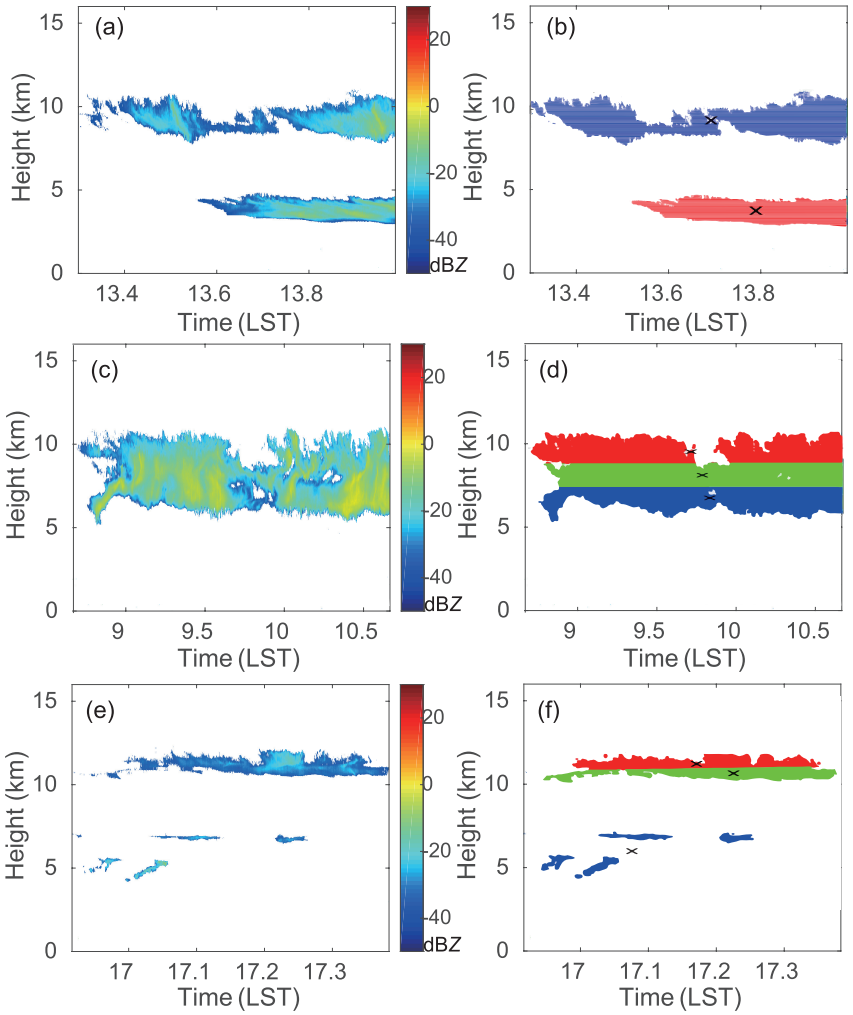

Fig. 3. Three kinds of clustering results: (a), (c) and (e) show radar reflectivity (dBZ); (b), (d) and (f) show clustering results, where the black crosses represent the center of each cluster which is illustrated with different color.

servations of clouds from the Earth's surface (WMO, 2017). These principles use the forms, features, appearance and internal structures of clouds to classify them into ten genera: cirrus (Ci), cirrostratus (Cs), cirrocumulus (Cc), altocumulus (Ac), altostratus (As), stratus (St), stratocumulus (Sc), nimbostratus (Ns), cumulus (Cu) or cumulonimbus (Cb) clouds. Cb clouds are defined as heavy and dense clouds, with a considerable vertical extent, in the form of a mountain or huge towers. The term deep convective (Dc) cloud used in the 1956 edition of the World Meteorological Organization cloud classification principles (WMO, 1956) is no longer used and has been replaced by the term Cb.

The International Satellite Cloud Climatology Project classifies clouds into Cs, Cc, Ac, As, St, Sc, Ns, Cu or Dc clouds based on a combination of the cloud-top pressure and the optical thickness of the cloud (Hahn et al., 2001). Tag et al. (2000) developed a probabilistic neural network cloud classifier using imagery from the Advanced Very High Resolution Radiometer and classified clouds into different types (Ci, Cc, Cs, Ac, As, Sc, St, Cu, Cb, and Cumulus Congestus clouds). These algorithms are developed for passive remote sensors. Williams et al. (1995) developed an algorithm classifying precipitating clouds in the tropics using a 915 MHz wind profile radar system. Wang and Sassen (2001) presented a cloud classification method to classify clouds into St, Sc, Cu, Ns, Ac, As, Dc or high-level clouds using multiple remote sensors, including polarization LiDAR, millimeter-wave radar and infrared radiometers. The cloud classification products (2B-CLDCLASS-lidar) released jointly by the CloudSat and CALIPSO satellites are based on this cloud classification method and clouds are classified into St, Sc, Cu, Ns, Ac, As, Dc or high-level clouds (Sassen and Wang, 2008; Sassen et al., 2008).

These cloud classification methods serve diverse purposes and depend on the specialties of the instrument. There is no unified cloud classification standard suitable for all instruments. Algorithms based on cloud spectral and textural features are suitable for passive sensors due to their broad field of view. Ground-based radar systems have advantages in terms of measurements of cloud height, thickness and internal microphysical characteristics and provide accurate information about the internal structure of clouds. However,

radar in the vertically pointing mode shows weaknesses in detecting the overall shape or appearance of clouds. Ci clouds are defined, based on appearance, as detached clouds in the form of white, delicate filaments or white or mostly white patches or narrow bands (WMO, 2017). Ka radar cannot distinguish Ci clouds by their appearance. Therefore, based on the strengths and limitations of the radar system and scientific requirements, our cloud classification algorithm is designed to classify clouds into nine types: Cs, Cc, Ac, As, St, Sc, Ns, Cu and Cb clouds.

4.2. *Parameters used for classification*

Temperature has a great impact on the phase and size of cloud particles. Atmospheric temperature is closely related to height. Height and temperature are therefore important indicators in cloud classification. Ice crystals dominate high-level clouds, whereas low-level Sc or St clouds are usually referred to as water clouds because liquid particles are in the majority. Mid-level As or Ac clouds are somewhere between low- and high-level clouds. Cumuliform clouds are generally more unstable than stratiform clouds. A non-uniformity coefficient representing the instability of clouds helps to distinguish cumuliform clouds from stratiform clouds. Reflectivity is also related to the cloud particle size and the water (ice) content, which are distributed differently in different types of cloud.

Based on these physical parameters, nine characteristics were selected for cloud classification: the mean cloud-top height ($H_{\rm tp}$), the mean cloud-base height ($H_{\rm bs}$), the maximum cloud depth ($D_{\rm hm}$), the average cloud depth ($D_{\rm av}$), the mean maximum $Z_{\rm e}$ ($Z_{\rm em}$) (the maximum $Z_{\rm e}$ is calculated for each profile), the non-uniformity coefficient ($I_{\rm nh}$), the cloud extent ($C_{\rm hr}$), precipitation ($P_{\rm cp}$), and the cloud temperature ($T_{\rm mp}$). $I_{\rm nh}$ is defined as the standard deviation of the maximum $Z_{\rm e}$ normalized by $Z_{\rm em}$. Precipitation ($P_{\rm cp}$) is flagged when there are more than ten bins with a height < 1 km, $Z_{\rm e} > -10$ dBZ, and a Doppler velocity > 0. The extent of clouds ($C_{\rm hr}$) is defined as the time which the radar uses to detect the cloud cluster. These parameters can be derived directly or indirectly from radar measurements, guaranteeing the portability of the classification algorithm.

4.3. *Methodology of classification*

The threshold-based classification methods are simple, but the classification results are very sensitive to the choice of threshold. The neural network classification method (Bankert, 1994; Tag et al., 2000), which requires many samples for training, is complicated, time consuming and cannot satisfy the requirements of real-time operation. The fuzzy logic method, computing with a "degree of truth" rather than "true or false" (1 or 0), is flexible, tolerant of imprecise data, and based on the experience of experts (Zadeh, 1968; Penaloza and Welch, 1996). The fuzzy logic method is an appropriate option for cloud classification and has been used on CloudSat CPR data to classify clouds (Sassen and Wang, 2008). Our ground-based Ka-band radar system has different specialties and observing modes, and the fuzzy sets for different fuzzy variables need to be changed to apply to the local characteristics of clouds and the specialties of the instrument. In this study, a trapezoid-shaped fuzzy membership function is defined as follows:

$$P(a,A_1,A_2,A_3,A_4)=\left\{\begin{array}{cc} 0, & a\leqslant A_1 \\ \dfrac{a-A_1}{A_2-A_1}, & A_1<a<A_2 \\ 1, & A_2\leqslant a\leqslant A_3 \\ \dfrac{A_4-a}{A_4-A_3}, & A_3<a<A_4 \\ 0, & a\geqslant A_4 \end{array}\right\}, \qquad (2)$$

where P is the degree of membership (between 0 and 1), 0 means definitely no, 1 means definitely yes, and a is the fuzzy variable (referring to the nine characteristic parameters). The fuzzy sets A_1, A_2, A_3 and A_4 determine the degree of membership of each fuzzy variable. A_1–A_4 for the nine characteristic parameters and the nine types of clouds are derived from a combination of the results of the following studies: (1) $H_{\rm tp}$, $H_{\rm bs}$, $D_{\rm hm}$ and $D_{\rm av}$ were obtained from previously published research (Huo and Lu, 2014; Chen et al., 2015; Huo, 2015); (2) typical clouds of the nine different types of cloud were selected and analyzed to determine $Z_{\rm em}$ and $C_{\rm hr}$ and the statistical results were used as reference data (see Fig. 4 for example statistical histogram for $Z_{\rm em}$); and (3) reference values for $I_{\rm nh}$ and $T_{\rm mp}$ were obtained from Sassen and Wang (2008) and Ren et al. (2011), respectively. The final A_1–A_4 values of the nine characteristic parameters for the nine different types of cloud are given in Table 2.

Once a cloud cluster has been distinguished, each characteristic parameter is calculated and input into the fuzzy membership function to obtain the membership value. The rule to obtain the final type is given by:

$$T_i=\sum_{j=1}^{9} w_j P_{i,j}\,, \qquad (3)$$

where the subscript i represents the cloud type, the subscript j represents the characteristic parameter, and w_j is the weight coefficient. We use different weight coefficients for the nine characteristic parameters because they play distinct parts in the classification method. For example, the position of clouds ($H_{\rm tp}$ and $H_{\rm bs}$) determines the cloud level, whereas the cloud depth ($D_{\rm hm}$ and $D_{\rm av}$) is primarily used to distinguish Cb clouds from Cu clouds. The value of $I_{\rm nh}$ represents the homogeneity of clouds. The temperature obtained from the ECMWF dataset may not quite reflect the actual situation. In these causes, w_j is set as 1.0, 1.0, 0.5, 0.8, 1.0, 1.0, 1.0, 0.5 and 0.8, respectively (also listed in Table 2). Based on Eq. (3), the final cloud type is that with a maximum value of T.

5. Application experiments

5.1. *Comparison experiment*

To evaluate the performance of the classification algorithm, a comparison experiment with an all-sky imager (Huo

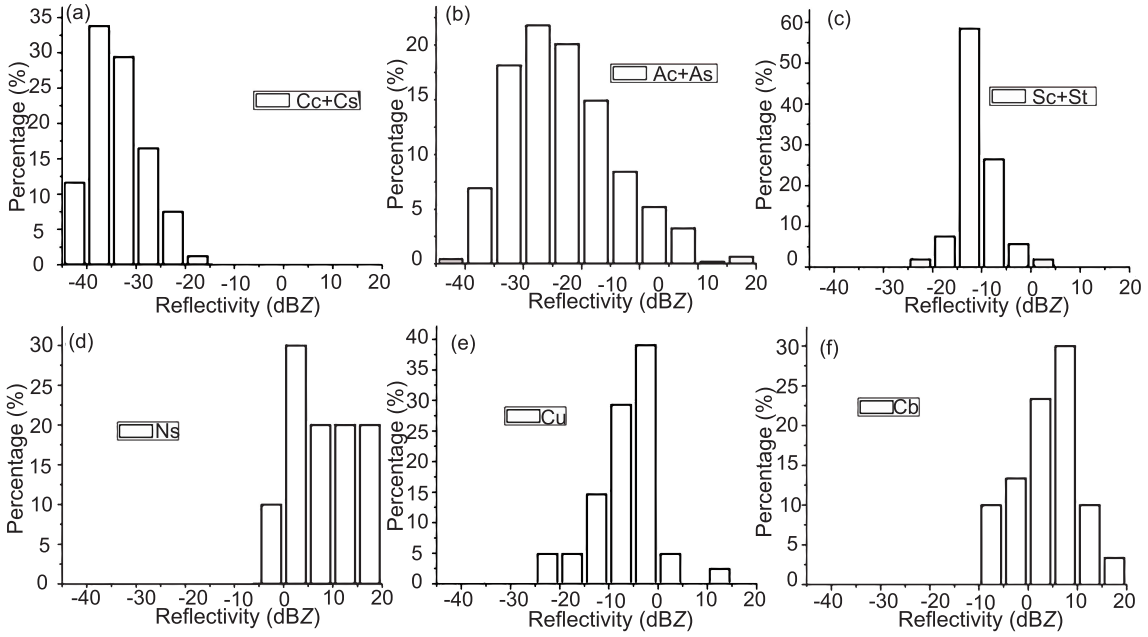

Fig. 4. Histograms of Z_{em} for different cloud types: (a) Cs and Cc; (b) As and Ac; (c) Sc and St; (d) Ns; (e) Cu; and (f) Cb.

Table 2. Fuzzy sets of A_1, A_2, A_3, and A_4.

No.	Parameter	Weight (w)		Cloud types								
				Cc	Cs	Ac	As	Sc	St	Cu	Ns	Cb
1	P (H_{tp}) (km)	1.0	A_4	11	11	11	11	5	4	10	8	10
			A_3	11	11	10	9	4	3	7	5	10
			A_2	7	7	5	4	1	0.5	3	3	8
			A_1	6	6	3	3	0.5	0.5	0.5	1	6
2	P (H_{bs}) (km)	1.0	A_4	10	10	5.5	5.5	2.5	2.5	3.5	2.5	2.5
			A_3	10	10	4.5	4.5	2	2	3	2	2
			A_2	6	6	3	3	1.5	1.5	2	1.5	2
			A_1	5	5	2.5	2.5	1.5	1.5	2	1.5	2
3	P (Z_{em}) (dBZ)	0.5	A_4	−10	−10	15	15	15	15	15	15	15
			A_3	−35	−35	−10	−10	0	0	−2	15	10
			A_2	−35	−35	−30	−30	−15	−10	−10	5	0
			A_1	−35	−35	−45	−45	−18	−18	−19	−10	−10
4	P (I_{nh})	0.8	A_4	0.5	0.35	0.5	0.4	0.65	0.55	0.65	0.55	2
			A_3	0.5	0.3	0.5	0.35	0.65	0.4	0.65	0.45	2
			A_2	0.45	0.3	0.45	0.3	0.55	0.3	0.55	0.4	0.7
			A_1	0.35	0.3	0.35	0.3	0.45	0.3	0.5	0.4	0.5
5	P (D_{hm}) (km)	1.0	A_4	5	5	8	8	5	4	8	8	10
			A_3	4	3	7	6	4	3	7	6	10
			A_2	2	1	3	2.5	1.5	0.7	2	3	10
			A_1	0.5	1	1	0.5	0.7	0.5	1	1	8
6	P (D_{av}) (km)	1.0	A_4	2	2	5	5	2.5	2.3	6	6	10
			A_3	2	1.5	5	3.5	2	1	5	4	10
			A_2	0.5	0.5	1.5	1	0.5	0.5	1.5	2	8
			A_1	0.1	0.5	0.5	0	0.2	0.5	0.5	0.5	6
7	P (P_{cp})	1.0	A_4	1	1	1	1	1	1	1	1	1
			A_3	1	1	1	1	1	1	1	1	1
			A_2	0	0	0	0	0	0	0	1	0
			A_1	0	0	0	0	0	0	0	1	0
8	P (T_{mp}) (°)	0.5	A_4	−18	−18	2	3	10	10	10	14	7
			A_3	−10	−10	−5	−5	10	10	4	7	0
			A_2	−60	−60	−20	−20	−1	−1	−4	−2	−5
			A_1	−60	−60	−30	−30	−10	−10	−10	−10	−15
9	P (C_{hr}) (h)	0.8	A_4	2	3	2.5	4	2	3	4	5	6
			A_3	0.5	3	0.5	4	0.5	3	2	5	4
			A_2	0.3	1	0.3	2.5	0.3	1	1	1.5	1
			A_1	0.3	0.5	0.3	0.5	0.3	0.5	1	1	0.5

and Lu, 2009) accompanying the Ka-band radar is performed. The all-sky imager, located about 8 m from the Ka-band radar system, took all-sky pictures every five minutes during the day from January to June 2014. We distinguished the cloud types from the all-sky images to compare with the clouds detected by the radar system. Because the Ka-band radar detects in the vertically pointing mode, clouds around the zenith in the images are examined closely. The radar system needs much more time than the imager to detect cloud clusters, so cloud types from the all-sky images collected during the radar observation period were counted and the cloud type in the majority was used as the final type (Table 3). A total of 270 clusters was available for comparison and 234 (86.7%) were determined as the same type using both classification methods. The classification accuracy varied from 73.7% (Cu) to 90% (As). The classification algorithm shows good performance through this comparison.

Six months of data from the all-sky imager are what we can use at present for ground-based comparison and evaluation. More data will help verify the evaluation result. Satellite data might be an option. However, the qualities of satellite cloud classification products also need evaluation. In addition, the temporal and spatial resolution and detection mode are very different. Direct evaluation via satellite data is inaccessible. In the future, more comparisons will be made to evaluate and improve the classification algorithm if more suitable data are accessible.

5.2. *Classification of clouds with precipitation*

Clouds with a complex distribution were analyzed to show the performance of the cloud clustering and classification algorithm. Figure 5a shows the radar reflectivity from 0000 LST 6 February to 0810 LST 7 February 2014. The times are given as Local Standard Time (LST, LST = UTC + 8) At 0815 LST 7 February 2014, the radar changed to the plan position indicator mode. Over the 32-hour period in Fig. 5a, the clouds change from high to low, from thin to thick, and from double layer to single layer. The surface temperature varies between 0°C and −4° and the surface wind speed was < 4 m s^{-1}.

Accurate clustering is a prerequisite for successful classification. In Fig. 5a, the clouds from the 16th hour to the 32nd hour (from 1600 LST 6 February to 0810 LST 7 February) challenge the clustering algorithm because there are two cloud layers, which show such large differences in horizontal and vertical extent that the smaller and lower cloud clusters are likely to be neglected. Figure 5b shows a preliminary result of the clustering and classification algorithm. Each individual cloud cluster is correctly separated and the clusters are classified into four types: Cc, Ac, Sc, and "Cb" when the cloud base height (CBH) of precipitation is considered as zero.

Table 3. Comparison of classification using the radar system and an all-sky imager. The rows show the radar results and the columns the imager results.

	Cc	Cs	Ac	As	Sc	St	Cu	Ns	Cb	No.	%
Cc	39	1	3	1	–	–	2	–	–	46	84.8
Cs	9	96	1	2	–	–	–	–	–	108	88.9
Ac	1	–	20	1	1	–	–	–	–	23	87
As	–	–	4	43	–	1	–	–	–	48	90
Sc	–	1	–	–	11	1	–	–	–	13	84.6
St	–	–	–	–	1	8	–	–	–	9	88.9
Cu	1	–	1	–	–	1	14	1	–	19	73.7
Ns	–	–	–	1	–	–	–	3	–	4	75
Cb	–	–	–	–	–	–	–	–	–	–	–
Total										270	

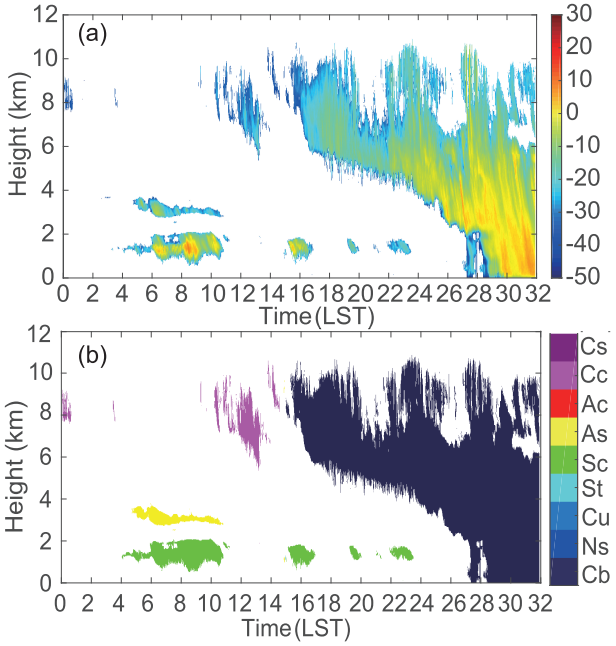

Fig. 5. Application of the clustering and classification algorithm: (a) radar reflectivity (dBZ) detected from 0000 LST 6 February to 0810 LST 7 February 2014; and (b) cloud types colored after clustering and classification.

The "Cb" cloud starts from the 16th hour (1600 LST 6 February) at a height of about 6 km, develops gradually downwards, and reaches the ground at the 27th hour (0300 LST 7 February). Precipitation starts from the 28th hour and the maximum cloud depth is > 10 km. The cloud depth is determined by H_{tp} minus H_{bs}. The preliminary CBH of precipitation for this "Cb" cloud is considered to be zero for simplicity, although the CBH is generally greater than zero. Therefore, the cloud depth must have been overestimated. At present, we have no applicable method to estimate the CBH of precipitation. A combination of reflectivity and the Doppler velocity may be useful in distinguishing droplets from cloud particles, although more studies are required to determine an accurate value of the CBH because precipitation sometimes occurs inside clouds. If profiles with precipitation are not considered, then the "Cb" cloud will be classified as a Cu because there is a large change in H_{bs}. Therefore, cloud clusters with precipitation may be classified differently as a result of the undetermined CBH. Based on our analysis, the cloud classification results may be more reliable if precipitation profiles are removed from the cloud cluster. This might be reasonable because the CBH of the cloud profile immedi-

ately ahead of precipitation is likely to be (or to be close to) the CBH of precipitation.

A meteorological observer needs several minutes to observe cloud types, whereas the all-sky (or whole-sky/total-sky) imager only needs a few milliseconds to take a picture. Ground-based radar in the vertically pointing mode needs comparatively more time to detect cloud clusters, and detection depends on the horizontal scale and speed of movement of the clouds. When clouds are passing over the radar system, radar detection can be considered as a two-dimensional scan of the clouds, whereas when the clouds do not move, radar detection can be considered as recording a process of evolution. Clouds with a long lifetime may undergo several stages of development and it may not be appropriate to group cloud profiles at different stages into one cluster and output only one cloud type. This presents a new problem of how to automatically distinguish each stage of the development process. At present, a manual check of the classification for large-scale clouds with precipitation lasting for more than five hours is suggested to ensure accuracy.

5.3. *Summer and winter cirrus clouds over Beijing*

According to the definition in the Glossary of Meteorology of the American Meteorological Society (Huschke, 1959; Seguin et al., 2012), the term "cirrus" is often used for all types of cirriform clouds (Ci, Cc and Cs clouds), and we use this definition here. Based on this definition of Ci, Cc and Cs clouds and the principle of our classification algorithm, Ci clouds are mostly classified as Cs clouds by our algorithm.

Figure 6 shows two cirrus clouds that occurred over Beijing in summer and winter. The cirrus cloud in Fig. 6a occurred on 4 July 2014 from 1500 to 2000 and lasted for about five hours. The cirrus cloud in Fig. 6b occurred on 21 November 2014 from 0200 to 0800 and lasted for about six hours. The macroscopic and microscopic characteristics of the two examples are very different. The cirrus cloud in Fig. 6a at 8.4 km H_{bs} was higher and thicker than the cloud in Fig. 6b with a height of 6.5 km H_{bs}. The mean Z_e in Fig. 6a is also considerably larger than that in Fig. 6b, indicating a larger cloud particle size or ice water content. Figure 6e shows the temperatures for the two examples. The mean temperatures of both clouds are lower than −20°C, indicating that most of the particles are solid. The winter cirrus cloud has a lower mean temperature than the summer cirrus cloud. These two examples reflect the seasonal variations between the summer and winter in Beijing, which are related to local weather processes and the characteristics of the regional climate. There is more water vapor and stronger convection in summer than in

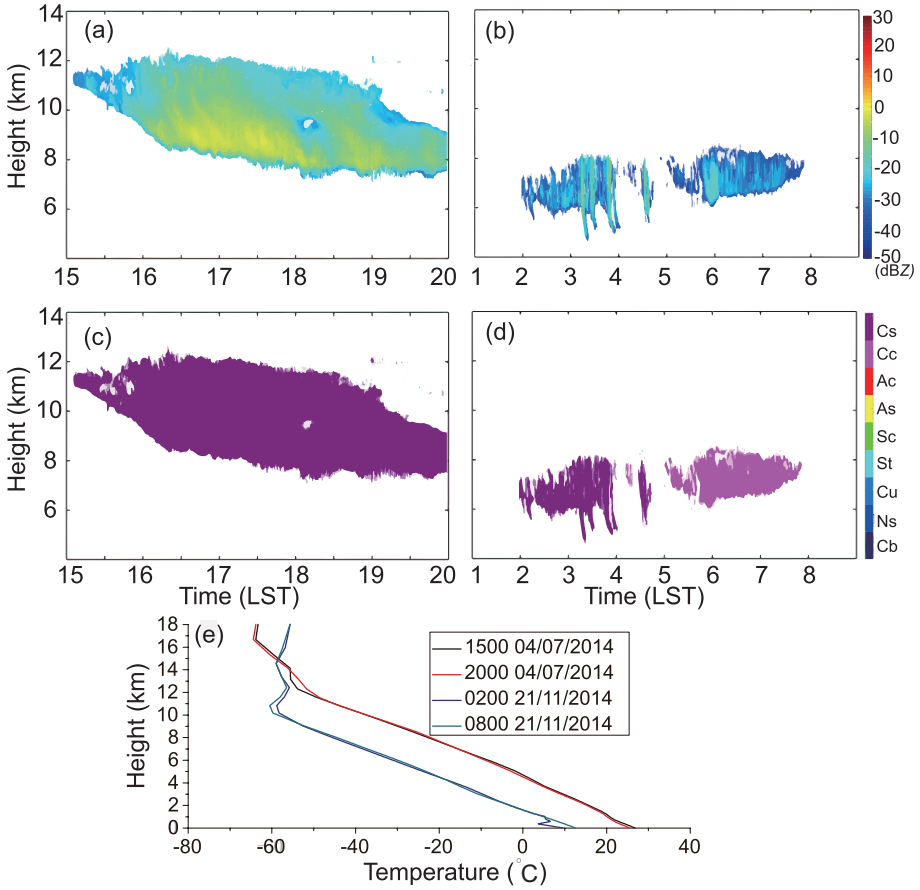

Fig. 6. Cirrus clouds in Beijing in winter and summer: (a) Cirrus cloud from 1500 LST to 2000 LST 4 July 2014; (b) cirrus cloud from 0200 LST to 0800 LST 21 November 2014; (c) classification results of (a); (d) classification results of (b); (e) temperature profiles of the two clouds.

winter. Cirrus clouds in summer in Beijing are usually higher than those in winter as a result of the difference in temperature.

The cloud profiles are grouped into four clusters in Fig. 6d, two of which are classified as Cc clouds, whereas the other two are classified as Cs clouds. The cloud type flips between Cs and Cc depending on $I_{\rm hn}$. The value of $I_{\rm hn}$ is 0.31 and 0.48, respectively, for the two Cc clouds and 0.1 and 0.22, respectively, for the two Cs clouds. The value of $I_{\rm hn}$ determines the cloud type when the clouds have a similar temperature, extent and height.

6. Distribution of cloud types in Beijing

Statistical analysis of the cloud types in Beijing was carried out based on data collected from January 2014 to December 2017. Figure 7 shows the distribution over all four seasons. Figure 7a shows the cloud cluster number histogram of each cloud type, and Fig. 7b shows the time histogram indicating the frequency of occurrence, defined here as the proportion of time one type of cloud was present relative to all types of cloud. Cs clouds accounted for the largest number observed, followed by St clouds. The lowest number was for Ns clouds. Because the average horizontal extent is different for different types of cloud—for example, the average horizontal extent of Ci clouds is generally smaller than that of Ac clouds—the occurrence frequency of clouds is more representative than the actual number of clouds in climatological statistics. The highest occurrence frequency in Fig. 7b is for Cs clouds, which accounted for about 24% of all cloud types. Because the average horizontal length of Ac clouds is relatively large, the occurrence frequency of Ac clouds was the second highest, although the number of clouds was lower than for other types, such as Cc and As clouds. The lowest occurrence frequency was for Ns clouds and the second lowest for Sc clouds, both of which were < 2.5%. The occurrence frequency of Cb clouds was only 6% and this cloud type mostly occurred in summer and autumn. The occurrence frequency of high-level clouds was highest in Beijing, followed by middle-level clouds; low-level clouds had the lowest occurrence frequency.

Cloud types showed a seasonal distribution. High-level clouds had the lowest occurrence frequency in spring (Cc and Cs; see Fig. 7b), but had almost the same occurrence frequency in the other three seasons. Low-level clouds (Sc, St, Cu, Ns and Cb; see Fig. 7b) had the highest occurrence frequency in summer. The occurrence frequency of Ac clouds was slightly higher in winter, but similar in the other three seasons. The occurrence frequency of As clouds is similar in all four seasons. The occurrence frequency of middle-level clouds (Ac and As clouds) had no clear seasonal variation. Cirrus clouds in summer in Beijing are usually higher than those in winter as a result of the difference in temperature and humidity.

Seasonal variation of cloud occurrence frequency in Beijing is related to local weather conditions and climatological characteristics. Beijing is located in the monsoon region with a continental climate of which winter monsoon and summer monsoon dominate alternately. Winter in Beijing is dry and cold while the summer is warm and wet. In addition, the average water vapor in Beijing is relatively poor when compared with the tropics. When water vapor is insufficient, the cloud is mainly formed by the adiabatic expansion and cooling of water vapor in the upward movement. Therefore, high-level clouds occur more than low-level clouds in Beijing. In addition, there is more water vapor and stronger convection in the summer than in winter. This causes obvious seasonal variations of cloud occurrence frequency, especially for low-level clouds.

7. Summary and discussion

The Ka-band radar system at the Institute of Atmospheric Physics has accumulated more than five years of data, and these long-term records support a great deal of cloud data for climatological and weather research in Beijing. Cloud classification is both a basic and necessary part of climatological and weather research and provides key knowledge about cloud types for the retrieval of cloud microphysical parameters based on radar reflectivity.

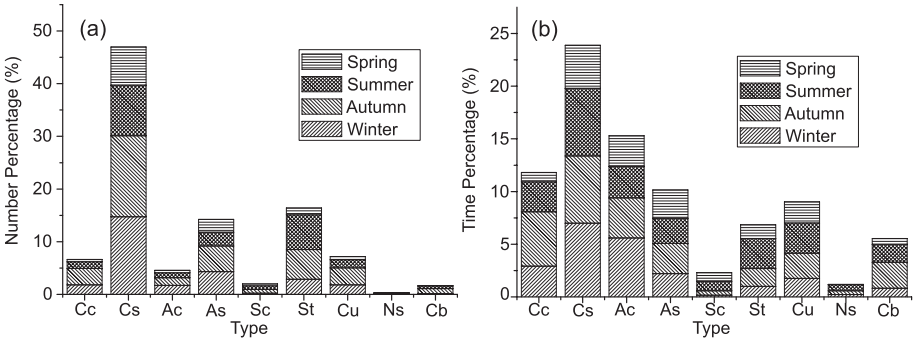

Fig. 7. Distribution of cloud types in spring, summer, autumn and winter in Beijing: (a) number histogram for different cloud types, where the *y*-axis is the percentage of one type of cloud to all nine types of cloud; (b) time histogram for different types of cloud, where the *y*-axis is the percentage of one type of cloud to all nine types of cloud (the occurrence frequency).

We have presented a cloud clustering and classification algorithm for ground-based Ka-band radar. An analysis of clouds observed from January to June 2014 showed that 92% of the data were clustered successfully by this algorithm. Comparison with an all-sky imager showed that the mean classification accuracy was 87%. Cloud clusters with precipitation have an uncertain CBH, and therefore the classification results for clouds with precipitation should be rechecked. In addition, we investigated the distribution of different cloud types in Beijing from January 2014 to December 2017 by applying the clustering and classification algorithm to radar data. Cs clouds had the highest occurrence frequency among the nine cloud types and high-level clouds had the maximum occurrence frequency over Beijing and low-level clouds had the minimum occurrence frequency. Overall, the algorithm presented here is based on, and is suitable for, vertically pointing Ka-band radar, which detects two-dimensional clouds. The algorithm can be transplanted to other waveband radar systems because most of the input parameters for these systems are available.

The clustering and classification algorithm usually needs a longer time than other sensors—such as imagers, radiometers, spectrometers, or meteorological observers—to determine the cloud type in the real-time observation of Ka-band radar in the vertically pointing mode. This is because the Ka radar system needs a longer time, depending on the horizontal extent and speed of movement of clouds, to collect all the cloud cluster profiles. If continuous clouds last for 10 hours or more, then the clustering algorithm will also need the same amount of time to collect all the profiles. The real-time operation of the classification algorithm will therefore depend on the horizontal extent and speed of movement of clouds.

At present, the attenuation calibration of reflectivity is not perfect. However, attenuation has little impact on the clustering and classification algorithm. This is because the attenuation is much lower than the reflectivity and the fuzzy sets of $Z_{\rm em}$ used in the classification algorithm are obtained from the statistics of the radar data, which have inadequate correction for attenuation. In addition, the weight of $Z_{\rm em}$ in the classification algorithm is lower than that of the other parameters. Therefore, inadequate attenuation correction should not overturn the final decision on classification.

Acknowledgements. This work is supported by the National Natural Science Foundation of China (Grant Nos. 41775032 and 41275040). We appreciate the valuable suggestions and insightful instructions from the reviewers. We also acknowledge our Ka radar team for their maintenance service in long-term measurement that made our research possible.

REFERENCES

Austin, R. T., A. J. Heymsfield, and G. L. Stephens, 2009: Retrieval of ice cloud microphysical parameters using the CloudSat millimeter-wave radar and temperature. *J. Geophys. Res.*, **114**, D00A23, https://doi.org/10.1029/2008JD010049.

Bankert, R. L., 1994: Cloud classification of AVHRR imagery in maritime regions using a probabilistic neural network. *J. Appl. Meteor.*, **33**, 909–918, https://doi.org/10.1175/1520-0450(1994)033<0909:CCOAII>2.0.CO;2.

Baum, B. A., V. Tovinkere, J. Titlow, and R. M. Welch, 1997: Automated cloud classification of global AVHRR data using a fuzzy logic approach. *J. Appl. Meteor.*, **36**, 1519–1540, https://doi.org/10.1175/1520-0450(1997)036<1519:ACCOGA>2.0.CO;2.

Chen, C., X. J. Guo, X. B. Qiu, and H. Meng, 2015: Analysis of cloud vertical structure and cloud water content over North China based on satellite remote sensing data. *Journal of Meteorology and Environment*, **31**, 159–164, https://doi.org/10.3969/j.issn.1673-503X.2015.05.023. (in Chinese)

Doviak, R. J., and D. S. Zrnić, 1993: *Doppler Radar and Weather Observations*. 2nd ed., Academic Press, 562 pp, https://doi.org/10.1016/C2009-0-22358-0.

Ebert, E., 1987: A pattern recognition technique for distinguishing surface and cloud types in the polar regions. *J. Appl. Meteor.*, **26**, 1412–1427, https://doi.org/10.1175/1520-0450(1987)026<1412:APRTFD>2.0.CO;2.

Hahn, C. J., W. B. Rossow, and S. G. Warren, 2001: ISCCP cloud properties associated with standard cloud types identified in individual surface observations. *J. Climate*, **14**, 11–28, https://doi.org/10.1175/1520-0442(2001)014<0011:ICPAWS>2.0.CO;2.

Hamilton, K., 2006: High resolution global modeling of the atmospheric circulation. *Adv. Atmos. Sci.*, **23**(6), 842–856, https://doi.org/10.1007/s00376-006-0842-3.

Heymsfield, A., D. Winker, M. Avery, M. Vaughan, G. Diskin, M. Deng, V. Mitev, and R. Matthey, 2014: Relationships between ice water content and volume extinction coefficient from in situ observations for temperatures from 0° to −86°C: Implications for spaceborne Lidar retrievals. *Journal of Applied Meteorology and Climatology*, **53**, 479–505, https://doi.org/10.1175/JAMC-D-13-087.1.

Huo, J., 2015: Physical properties of mid-level clouds based on CloudSat/CALIPSO data over land and sea. *Climatic and Environmental Research*, **20**, 30–40, https://doi.org/10.3878/j.issn.1006-9585.2014.13188. (in Chinese)

Huo, J., and D. R. Lu, 2009: Cloud determination of all-sky images under low-visibility conditions. *J. Atmos. Oceanic Technol.*, **26**(10), 2172–2181, https://doi.org/10.1175/2009JTECHA1324.1.

Huo, J., and D. R. Lu, 2014: Physical properties of high-level cloud over land and ocean from CloudSat-CALIPSO data. *J. Climate*, **27**, 8966–8978, https://doi.org/10.1175/JCLI-D-14-00329.1.

Huschke, R. E., 1959: *Glossary of Meteorology*. American Meteorological Society, 638 pp.

Kollias, P., I. Jo, P. Borque, A. Tatarevic, and K. Lamer, 2014: Scanning ARM cloud radars. Part II: Data quality control and processing. *J. Atmos. Oceanic Technol.*, **31**, 583–598, https://doi.org/10.1175/JTECH-D-13-00045.1.

Li, Y. Y., L. X. Fang, and X. W. Kou, 2014: Principle and standard of auto-observation cloud classification for satellite, ground measurements and model. *Chinese Journal of Geophysics*, **57**, 2433–2441, https://doi.org/10.6038/cjg20140805. (in Chinese)

Liu, L. P., L. Xie, and Z. H. Cui, 2014: Examination and application of Doppler spectral density data in drop size distribution retrieval in weak precipitation by cloud radar. *Chinese Journal of Atmospheric Sciences*, **38**, 223–236, https://doi.org/

10.3878/j.issn.1006-9895.2013.12207. (in Chinese)

Mace, G. G., and K. Sassen, 2000: A constrained algorithm for retrieval of stratocumulus cloud properties using solar radiation, microwave radiometer, and millimeter cloud radar data. *J. Geophys. Res.*, **105**, 29 099–29 108, https://doi.org/10.1029/2000JD900403.

Miller, S. W., and W. J. Emery, 1997: An automated neural network cloud classifier for use over land and ocean surfaces. *J. Appl. Meteor.*, **36**, 1346–1362, https://doi.org/10.1175/1520-0450(1997)036<1346:AANNCC>2.0.CO;2.

Molinari, J., and M. Dudek, 1992: Parameterization of convective precipitation in mesoscale numerical models: A critical review. *Mon. Wea. Rev.*, **120**, 326–344, https://doi.org/10.1175/1520-0493(1992)120<0326:POCPIM>2.0.CO;2.

Penaloza, M. A., and R. M. Welch, 1996: Feature selection for classification of polar regions using a fuzzy expert system. *Remote Sensing of Environment*, **58**, 81–100, https://doi.org/10.1016/0034-4257(95)00260-X.

Ramanathan, V., R. D. Cess, E. F. Harrison, P. Minnis, B. R. Barkstrom, E. Ahmad, and D. Hartmann, 1989: Cloud-radiative forcing and climate: Results from the earth radiation budget experiment. *Science*, **243**, 57–63, https://doi.org/10.1126/science.243.4887.57.

Ren, J. Q., W. Yan, H. L. Yang, and J. K. Shi, 2011: Cloud classification algorithm for CloudSat satellite based on fuzzy logic method. *Journal of PLA University of Science and Technology (Natural Science Edition)*, **12**, 90–96, https://doi.org/10.3969/j.issn.1009-3443.2011.01.017. (in Chinese)

Rodgers, C. D., 1976: Retrieval of atmospheric temperature and composition from remote measurements of thermal radiation. *Rev. Geophys.*, **14**, 609–624, https://doi.org/10.1029/RG014i004p00609.

Rodgers, C. D., 1990: Characterization and error analysis of profiles retrieved from remote sounding measurements. *J. Geophys. Res.*, **95**, 5587–5595, https://doi.org/10.1029/JD095iD05p05587.

Rodgers, C. D., 2000: *Inverse Methods for Atmospheric Sounding: Theory and Practice*. World Scientific Publishing, 256 pp.

Rossow, W. B., and R. A. Schiffer, 1999: Advances in understanding clouds from ISCCP. *Bull. Amer. Meteor. Soc.*, **80**, 2261–2288, https://doi.org/10.1175/1520-0477(1999)080<2261:AIUCFI>2.0.CO;2.

Sassen, K., and Z. E. Wang, 2008: Classifying clouds around the globe with the CloudSat radar: 1-year of results. *Geophys. Res. Lett.*, **35**, L04805, https://doi.org/10.1029/2007GL032591.

Sassen, K., Z. E. Wang, and D. Liu, 2008: Global distribution of cirrus clouds from CloudSat/Cloud-Aerosol Lidar and infrared pathfinder satellite observations (CALIPSO) measurements. *J. Geophys. Res.*, **113**, D00A12, https://doi.org/10.1029/2008JD009972.

Seguin, W., F. Genene, and R. Perriello, 2012: "Cirrus". Glossary of Meteorology. 2nd ed., Amer. Meteor. Soc. [Available online from http://glossary.ametsoc.org/wiki/Cirrus]

Sekelsky, S. M., W. L. Ecklund, J. M. Firda, K. S. Gage, and R. E. McIntosh, 1999: Particle size estimation in ice-phase clouds using multifrequency radar reflectivity measurements at 95, 33, and 2.8 GHz. *J. Appl. Meteor.*, **38**, 5–28, https://doi.org/10.1175/1520-0450(1999)038<0005:PSEIIP>2.0.CO;2.

Tag, P. M., R. L. Bankert, and L. R. Brody, 2000: An AVHRR multiple cloud-type classification package. *J. Climate*, **39**, 125–134, https://doi.org/10.1175/1520-0450(2000)039<0125:AAMCTC>2.0.CO;2.

Wang, Z. E., and K. Sassen, 2001: Cloud type and macrophysical property retrieval using multiple remote sensors. *J. Appl. Meteor.*, **40**, 1665–1682, https://doi.org/10.1175/1520-0450(2001)040<1665:CTAMPR>2.0.CO;2.

Williams, C. R., W. L. Ecklund, and K. S. Gage, 1995: Classification of precipitating clouds in the tropics using 915-MHz wind profilers. *J. Atmos. Oceanic Technol.*, **12**, 996–1012, https://doi.org/10.1175/1520-0426(1995)012<0996:COPCIT>2.0.CO;2.

World Meteorological Organization, 1956: International Cloud Atlas: Abridged Atlas. WMO, 72 pp.

World Meteorological Organization, 2017: International Cloud Atlas. WMO. [Available from https://cloudatlas.wmo.int/home.html]

Xiao, P., J. Huo, and Y. H. Bi, 2018: Ground-based Ka band cloud radar data quality control. *Journal of Chengdu University of Information Technology*, **33**, 129–136, https://doi.org/10.16836/j.cnki.jcuit.2018.02.005.

Zadeh, L. A., 1968: Fuzzy algorithms. *Information and Control*, **12**, 94–102, https://doi.org/10.1016/S0019-9958(68)90211-8.