

## RESEARCH ARTICLE

10.1002/2014JD021760

## Special Section:

East Asian Study of  
Tropospheric Aerosols and  
Impact on Cloud and  
Precipitation

## Key Points:

- A new reliable MPL-based aerosol and cloud detection algorithm is developed
- The new method detects more clouds with high bases than ARM MPL VAP product
- Retrieved cloud properties are reasonable at both the SGP and the Taihu sites

## Correspondence to:

C. Zhao,  
c Zhao@bnu.edu.cn

## Citation:

Zhao, C., Y. Wang, Q. Wang, Z. Li, Z. Wang, and D. Liu (2014), A new cloud and aerosol layer detection method based on micropulse lidar measurements, *J. Geophys. Res. Atmos.*, *119*, 6788–6802, doi:10.1002/2014JD021760.

Received 17 MAR 2014

Accepted 23 MAY 2014

Accepted article online 28 MAY 2014

Published online 13 JUN 2014

## A new cloud and aerosol layer detection method based on micropulse lidar measurements

Chuanfeng Zhao<sup>1</sup>, Yuzhao Wang<sup>1</sup>, Qianqian Wang<sup>1</sup>, Zhanqing Li<sup>1,2</sup>, Zhien Wang<sup>3</sup>, and Dong Liu<sup>4</sup>

<sup>1</sup>College of Global Change and Earth System Science, Beijing Normal University, Beijing, China, <sup>2</sup>ESSIC and Department of Atmospheric and Oceanic Sciences, Cooperative Institute for Climate Studies, University of Maryland, College Park, Maryland, USA, <sup>3</sup>Department of Atmospheric Science, University of Wyoming, Laramie, Wyoming, USA, <sup>4</sup>Key Laboratory of Atmospheric Composition and Optical Radiation, Anhui Institute of Optics and Fine Mechanics, Chinese Academy of Sciences, Hefei, China

**Abstract** This paper introduces a new algorithm to detect aerosols and clouds based on micropulse lidar measurements. A semidiscretization processing technique is first used to inhibit the impact of increasing noise with distance. The value distribution equalization method which reduces the magnitude of signal variations with distance is then introduced. Combined with empirical threshold values, we determine if the signal waves indicate clouds or aerosols. This method can separate clouds and aerosols with high accuracy, although differentiation between aerosols and clouds are subject to more uncertainties depending on the thresholds selected. Compared with the existing Atmospheric Radiation Measurement program lidar-based cloud product, the new method appears more reliable and detects more clouds with high bases. The algorithm is applied to a year of observations at both the U.S. Southern Great Plains (SGP) and China Taihu sites. At the SGP site, the cloud frequency shows a clear seasonal variation with maximum values in winter and spring and shows bimodal vertical distributions with maximum occurrences at around 3–6 km and 8–12 km. The annual averaged cloud frequency is about 50%. The dominant clouds are stratiform in winter and convective in summer. By contrast, the cloud frequency at the Taihu site shows no clear seasonal variation and the maximum occurrence is at around 1 km. The annual averaged cloud frequency is about 15% higher than that at the SGP site. A seasonal analysis of cloud base occurrence frequency suggests that stratiform clouds dominate at the Taihu site.

## 1. Introduction

Clouds play an essential role in the Earth's climate by modulating the energy budget and water cycle. They can change the Earth's energy balance by reflecting solar radiation and by trapping longwave radiation within the atmosphere. They are also important in the atmospheric transport of water from oceans to land. The representation of clouds remains one of the largest uncertainties in current climate predictions [Intergovernmental Panel on Climate Change, 2007], particularly because of the complicated feedbacks between clouds and other atmospheric constituents. Accurate cloud observations are crucial for both process level studies and climate prediction.

Clouds can be identified using a lidar [Platt et al., 1994; Spinhirne et al., 1995; Clothiaux et al., 1998; Wang and Sassen, 2001; Campbell et al., 2002; Mendoza and Flynn, 2006; Coulter, 2012], a ceilometer [Morris, 2012], a millimeter cloud radar (MMCR) [Clothiaux et al., 1995], or from temperature and water vapor profiles [Zhang et al., 2013]. The micropulse lidar (MPL) is a ground-based optical remote sensing system designed primarily to determine the altitude of clouds overhead, with the unique advantage that it uses a low-pulse energy laser for eye-safe operation. In addition to the real-time detection of clouds, postprocessing of the lidar signal returns can also characterize the extent and properties of aerosols or other particle-laden regions.

Compared to weather radars, MPL lidar systems are particularly sensitive to smaller atmospheric particles due to their enhanced scattering at visible wavelengths [Sassen, 1995]. The MPL usually gives more accurate determination of cloud bases than radar. It is largely immune to widely spaced nonhydrometeorological scatterers and precipitation [Clothiaux et al., 2000] and can detect thin clouds that could be missed by an MMCR. When clouds are thin, MPL measurements can also be used to retrieve cloud optical properties. Furthermore, the MPL is an active remote sensor with high temporal and spatial resolutions, allowing for continuous observation of clouds with high accuracy. The main weakness of the MPL compared to the MMCR

is the quick attenuation of MPL signals by clouds and aerosols which make it impossible to detect tops of optically thick clouds or clouds over a low thick cloud. Therefore, ground-based MPLs can detect most of the low clouds with or without higher clouds above them [Warren *et al.*, 1984], but may significantly underestimate upper multilayer clouds compared with an MMCR or spaceborne lidar or radar measurements [Chang and Li, 2005]. Cloud models with good microphysics and high space and time resolutions, including both one-dimensional models which can simulate the time evolution of cloud top and base heights [e.g., Ćurić and Janc, 1993] and three-dimensional mesoscale cloud models [e.g., Ćurić *et al.*, 2008], can be reliably evaluated using cloud observations from the MPL and the MMCR.

With obtained lidar signal distribution, an algorithm is required to automatically detect clouds. Cloud detection algorithms using MPL are generally based on the difference between the observed lidar-reflected signal and the estimated background signal of ambient light [e.g., Clothiaux *et al.*, 1998; Wang and Sassen, 2001]. When the difference is above a threshold value within a certain layer, the hydrometeors in that layer are classified as clouds. The observed lidar-reflected signals are generally preprocessed with various quality controls. Several methods have been proposed to estimate the background signal of ambient light. For example, Campbell *et al.* [2008] proposed dividing a day into 12 separate periods where a coarse three-point vertical integration scheme seeks out clear profiles for possible background baselines. Wang and Sassen [2001] used molecular signals (reference signals) to measure the layer signal intensity related to molecular scattering. The reference signal is determined based on a log-scale linear interpolation within the nonnoise regions. Here the nonnoise region is defined as the region where the signal is significantly different from the noise. Note that the reference signal is scaled for each layer to account for cloud attenuation effects. This method has been used to generate the Atmospheric Radiation Measurement (ARM) MPL cloud value added product (VAP). The U.S. Department of Energy's (DOE) ARM program provides long-term continuous observations of aerosols, clouds, and radiation [Ackerman and Stokes, 2003].

Most MPL-based cloud detection algorithms use either overlapped corrected lidar signal returns [e.g., Wang and Sassen, 2001] or range-corrected lidar signal returns [e.g., Campbell *et al.*, 2002]. For either method, the key is how to minimize the impact of noise so that aerosol and cloud signatures can be distinguished. Methods like that developed by Wang and Sassen [2001] have used a log scale to decrease the variation in valid lidar signal-to-noise ratios with distance so that potential layers at far distances can be observed. Many other methods carry out a range correction to lidar signal returns before doing the cloud/aerosol detection in order to solve the weak signal-to-noise ratio problem at far distances. Those methods introduce another problem, namely, the enlarging noises in the lidar signal returns. This is a challenging problem that many scientists are trying to solve.

This study presents a new cloud and aerosol detection algorithm using two concepts from image processing, semidiscretization processing (SDP) method [Inesperger and Stepan, 2010], and histogram equalization method [Han *et al.*, 2011]. The SDP method is an efficient numerical method that provides a finite-dimensional matrix approximation of the infinite-dimensional monodromy matrix [Inesperger and Stepan, 2010]. Histogram equalization is a method in image processing of contrast adjustment using the image's histogram. Here we use this concept to process lidar signal returns instead of images and call this method the value distribution equalization (VDE) method. By using these two new concepts, the new cloud detection method can ideally solve the problems mentioned earlier. It can prevent the expansion of noises at far distances, make aerosol/cloud signals stand out clearly from noises, and make signal-to-noise ratios at far distances comparable to those at close distances.

The MPL observations used in this study and the cloud detection algorithm are described in section 2. Section 3 evaluates the new cloud detection algorithm based on a synthetic test and a case study. Section 4 shows the cloud properties detected with an MPL using our new algorithm for both a U.S. site and a China site, which are both located in the midlatitudes.

## 2. Observations and Cloud Detection Algorithm

### 2.1. Observations

Figure 1 shows the two sites considered in this study. One is the U.S. SGP site (36.6°N, 97.5°W) located in Oklahoma and supported by the U.S. DOE's ARM program. The other is Taihu site (116.3°E, 30.4°N) located in



**Figure 1.** Location of the two sites. A is the China Taihu Site and B is the U.S. SGP site.

Jiangsu province in China and was in operation during the East Asian Study of Tropospheric Aerosols and Impact on Regional Climate experiment [Li *et al.*, 2011]. Both sites lie in similar middle latitudes of the Northern Hemisphere. However, the SGP site is located in the continental interior on dry land and Taihu is near a big lake and close to the ocean.

ARM has long-term continuous MPL cloud base and top measurements since 1996 at the SGP site. Fast sampling data are available since 2010. By contrast, the MPL cloud measurements have been made at the Taihu site from March 2008 to April 2009 only. This study uses data from May 2012 to May 2013 at the SGP site and from March 2008 to March 2009 at the Taihu site. Fast sampling data from the SGP site were preferred which is why data from March 2008 to March 2009 at the SGP site were not selected. Considering that the major purpose of this study is to show the application of new MPL cloud detection method and the statistical cloud macrophysical properties at two selected sites, the time period differences here are not critical to our study.

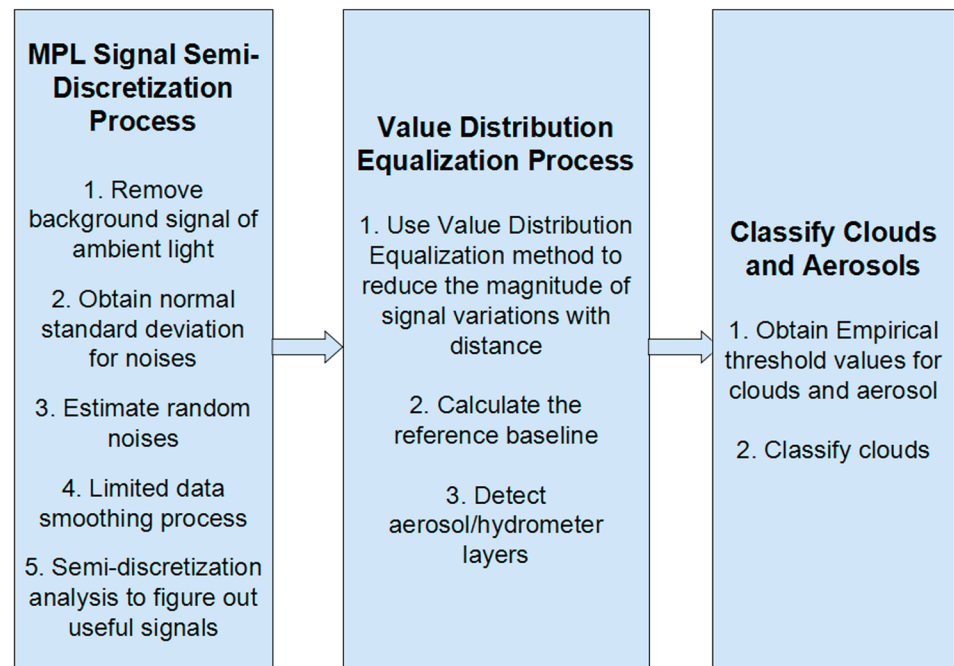
The MPL instruments used at these two sites are both manufactured by the Sigma Space Corporation. Detailed information about the MPL instruments can be found in *Spinhirne* [1993]. MPL data collected at both SGP and Taihu sites have a range resolution (height interval) of 15 m. The maximum range for cloud detection can reach ~18 km. MPL observations have been averaged with time resolutions of 10 s at the SGP site and 3 min at the Taihu site. Potential uncertainties in the MPL measurements have been described in *Coulter* [2012]. For example, there is an inherent calibration uncertainty of the timing electronics of about 2%, which translates directly into an uncertainty of ±2% for all reported distances. Also, reported cloud heights are centered within the range bin, so cloud heights will have an uncertainty of ±7.5 m (half of the range bin resolution). The MPL observations are also subject to the limit of signal attenuation [Spinhirne *et al.*, 1995]. The dead zone for the MPL observations is about 150 m, which means that the results below 150 m cannot be used. A detailed description about the calibration of MPL observations can be found in the instrument manual [Mendoza and Flynn, 2006].

## 2.2. Cloud Detection Algorithm

In general, the size parameter ( $2\pi r/\lambda$ , where  $r$  is the particle radius and  $\lambda$  is the wavelength) of cloud droplets at MPL wavelengths are much larger than 1, so Mie scattering theory can be used. The Mie scattering lidar signal,  $p(z)$ , can be expressed as

$$p(z) = CEO(z) \frac{\beta(z)}{z^2} \exp\left[-2\int_0^z \alpha(z') dz'\right] + N_b + A(z), \quad (1)$$

where  $z$  is the detection distance,  $C$  is a system constant,  $E$  is the pulse energy,  $O(z)$  is the geometry factor,  $\beta(z)$  is the backscattering coefficient of atmospheric particles,  $\alpha(z)$  is the extinction coefficient,  $N_b$  is background noise, and  $A(z)$  is the after-pulse response of the photomultiplier detector. Parameters  $A(z)$ ,  $C$ , and  $O(z)$  are from the system calibration.  $N_b$  can be obtained from the lidar signal returns at far distances, e.g., above 17 km.



**Figure 2.** Schematic diagram showing the steps taken in detecting clouds using MPL measurements.

The cloud detection algorithm is developed mainly based on the Mie scattering described in equation (1). Figure 2 shows the schematic structure of the cloud detection algorithm. It includes three parts. Part 1 is the SDP processing of MPL signal returns based on estimates of noises, which generate useful MPL signals. Part 2 is the MPL signal VDE processing, which can detect most aerosol/hydrometeor layers with high accuracy by reducing the magnitude of signal variations with distance. Part 3 is the classification of clouds and aerosols based on empirical threshold parameters. The three steps are described in detail below.

### 2.2.1. SDP Method

The range correction of lidar signals could enlarge the noises at high altitudes, so no range correction is made in the first cloud detection step. The signal,  $P(z)$ , can now be expressed as

$$P(z) = \frac{p(z) - N_b - A(z)}{CEO(z)}. \quad (2)$$

After removal of the background signal of ambient light  $N_b$ , a large amount of random noise still exists in the lidar signals. This noise could significantly affect the postprocessing of the lidar signals. In most cases, the moving average method is used to process the lidar signals.

For ground-based lidar observations, the received lidar signal from a distance beyond a certain height range (17 km in this study) is treated as the background noise. Therefore,  $N_b$  in equations (1) and (2) can be estimated based on the signal average at that distance range. After subtracting  $N_b$ , the normal standard deviation (sd) for the noise at that distance range is

$$sd = \left( \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \right)^{1/2}, \quad \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (3)$$

where  $x$  is the background noise signal. The noise for the lidar signal  $P(z)$  can be described as

$$\text{noise}(z) = K \cdot sd, \quad (4)$$

where  $K$  is an empirical parameter, which is set as 3 in this study.

In general,  $P(z)$  is smoothly averaged to reduce the random noise. However, too much smoothing will remove information about sharp signal changes due to clouds or aerosols. Here we make limited moving averages. When the vertical resolution is greater than 0.03 km, we use three-point moving average. When the vertical

resolution is less than 0.03 km, a moving average within a window of 0.06 km is done. The resulting lidar signal after moving average is denoted as  $P_s(z)$ .

Next, an SDP analysis of signal  $P_s(z)$  is carried out so that the impact of atmospheric turbulence can be minimized. Starting from the second point of the series  $P_s(z_i)$  ( $i = 2, 3, \dots, N$ ), if the absolute signal difference between  $P_s(z_i)$  and  $P_s(z_{i-1})$  is less than  $\text{noise}(z)$ ,  $P_s(z_i)$  is set as  $P_s(z_{i-1})$ . Otherwise,  $P_s(z_i)$  stays the same. The reason that we make this change is that the change in the signal is useful only when the signal is larger than the noise level. A new lidar return signal,  $PD1(z)$ , is obtained at the end of this procedure. Important signal information is contained in  $PD1(z)$ , such as cloud or aerosol layers and high signal-to-noise ratios.

Meaningless perturbations in the signals have been removed. The same procedure is applied again to the series  $P_s(z_i)$ , but starting from the end ( $i = N - 1, N - 2, \dots, 1$ ). If the absolute signal difference between  $P_s(z_i)$  and  $P_s(z_{i+1})$  is less than  $\text{noise}(z)$ ,  $P_s(z_i)$  is set as  $P_s(z_{i+1})$ . Otherwise,  $P_s(z_i)$  stays the same. A new lidar return signal  $PD2(z)$  is generated. The final signal,  $PD(z)$ , is the average of  $PD1(z)$  and  $PD2(z)$ .

By doing the SDP analysis with moving averages over a limited set of signal points, information about sharp signal changes have been kept and observation data have been made discrete. Compared to other methods using a smoothing method with more data points, this algorithm has sharper signal changes at cloud bases and tops and often causes the derived cloud bases higher and cloud tops lower if same cloud detection threshold values are used.

**2.2.2. VDE Method**

$P(z)$  in equation (2) extends over a large range due to the fact that the atmospheric particle concentration decreases exponentially with height and that the  $R$  square signal decreases as well. Therefore, there is a significant change in clear-sky signal returns with distance in addition to the sharp increase in the observed signal due to clouds or aerosols. To solve this problem, two approaches have been proposed. One involves the use of a log scale for lidar signals to reduce the varying range of the clear-sky signals, and the other directly finds the particle layers by searching for the break points in measured signals. Although the first approach can reduce the signal varying range, it requires a partial moving average in order to get good baseline signals. These moving averages could cause the missing of some thin clouds with weak signal returns. On the other hand, it is possible that clouds exist within the region of lidar noises, and then the first algorithm may result in the application of a log scale to negative values. The second algorithm requires high-quality lidar return signals to accurately search for break points. In other words, lidar signals have to be well processed by removing the noise with a smoothing method. Current smoothing algorithms, like averaging over a certain time range, have their limitations. For example, averaging over a certain time range could make the signal at the cloud/aerosol layer base change unexpectedly if the time range is too large or introduce a large amount of noise if the time range is too small.

The VDE method for lidar signal processing is proposed based on the idea of the first approach mentioned above and the histogram equalization method which is often used in image processing. The histogram equalization method is a technique that dynamically increases the brightness of gray areas representing low values by modifying the distribution of image brightness. It is generally used for discretized image data. The purpose of this technique in image processing is to increase the visibility of the data image. In a similar vein, the purpose in this study is to dynamically increase the visibility of signal-to-noise ratios measured at far distances so that high-level clouds can be more reliably detected. The VDE processing of lidar signals includes four steps.

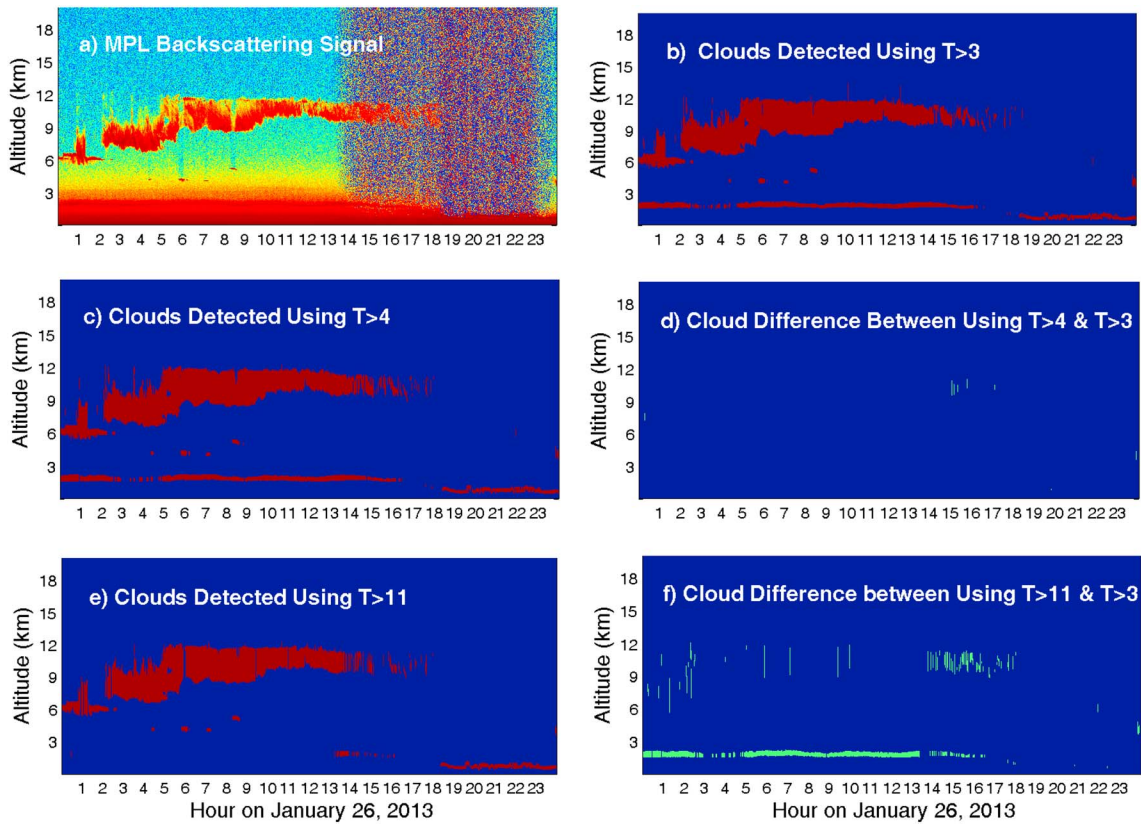
1.  $P_D(z)$  is first reorganized in an ascending order. Assuming there are  $N$  signal returns, the reordered signal is  $R_s(i)$ , ( $i = 1, 2, 3 \dots N$ ). Indices corresponding to the reorganized elements in  $P_D(z)$  are denoted as  $I_s(i)$ , ( $i = 1, 2, 3, \dots N$ ). The maximum and minimum of all signals are denoted as  $M_A$  and  $M_I$ , respectively.
2. The mapping proportion coefficient,  $P_E(i)$ , is then calculated. For each signal  $R_s(i)$ ,  $P_E(i)$  is calculated as

$$P_E(i) = i/N, \quad (i = 1, 2, 3, \dots N). \tag{5}$$

If  $R_s(i) = R_s(i - 1)$ , then  $P_E(i) = P_E(i - 1)$ .

With  $P_E(i)$  and the signal ranges, new ascending data values can be calculated:

$$y(i) = P_E(i) \cdot (M_A - M_I) + M_I, \quad (i = 1, 2, 3, \dots N). \tag{6}$$



**Figure 3.** Sensitivity of the cloud classification method to the threshold values used. (a) MPL backscattering signal returns, (b) cloud classification with  $D < -7$  or  $T > 3$ , (c) cloud classification with  $D < -7$  or  $T > 4$ , (d) the cloud difference between Figures 3c and 3b, (e) cloud classification with  $D < -7$  or  $T > 11$ , and (f) the cloud difference between Figures 3e and 3b. Data are for 26 January 2013 at the SGP site.

Finally, a new data set based on the equalization information calculated in the above steps, is given by

$$P_N(z) = P_N[I_S(i)] = y(i), \quad (i = 1, 2, 3, \dots, N). \quad (7)$$

With this equalization analysis, lidar signals are placed into a slant line. The signals caused by cloud or aerosol particle layers will cause signals to move away from this line. Particle layers can then be identified and clouds can be distinguished from aerosols based on empirical detection parameters.

### 2.2.3. Layer Classification

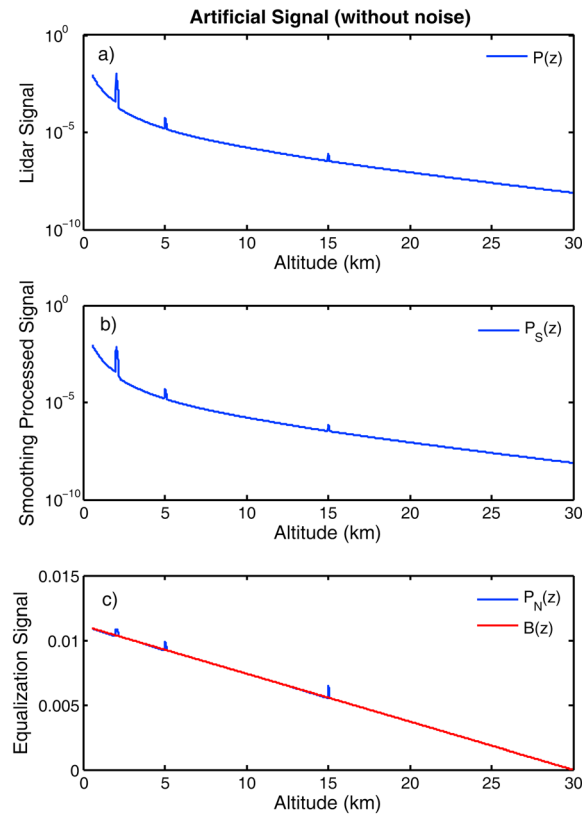
The processed signal  $P_N(z)$  follows the baseline  $B(z)$  with perturbations. The baseline has two end points,  $(Z_1, M_A)$  and  $(Z_N, M_I)$ . By searching for the point where  $P_N(z)$  becomes larger than  $B(z)$  and the closest point that  $P_N(z)$  becomes smaller than  $B(z)$ , the particle layer base and top can be established respectively. In this process, a threshold value for the layer depth can be set to remove some pseudo layers caused by the random signals, such as a minimum depth of 45 m.

Parameters used to classify clouds and aerosols are determined after identifying particle layers. Following the cloud detection method developed by Wang and Sassen [2001], we calculate the slopes,  $F(z)$ , of the signals within the particle layer for both ascending and descending parts as

$$F(z) = \frac{d \ln[P(z) \cdot z^2]}{dz}. \quad (8)$$

The maximum and minimum  $F(z)$  are denoted as  $T$  and  $D$ , respectively. When  $z$  is below 3 km, layers are classified as clouds when  $T > 3$  or  $D < -7$ ; when  $z$  is no less than 3 km, layers are classified as clouds when  $T > 1.5$  or  $D < -7$ . For other cases, particle layers are classified as aerosols. The method used to determine these threshold values is similar to that used in other studies [Clothiaux et al., 1998; Campbell et al., 2002].

The classification method used to distinguish between clouds and aerosols is sensitive to the threshold values used in the algorithm. Figure 3 shows how the cloud classification method responds to different



**Figure 4.** Testing the new cloud detection algorithm using synthetic lidar signals. (a) The simulated radar signal,  $P(z)$ , (b) the lidar signal after the moving average procedure,  $P_s(z)$ , and (c) the SDP and VDE-processed signal,  $P_N(z)$ , and baseline background signal of ambient light,  $B(z)$ . Three layers are clearly identified.

threshold values of  $T$ . Three threshold values (3, 4, and 11) are tested. As  $T$  increases, a smaller number of signals are classified as clouds. It is likely that some aerosol (cloud) layers are misclassified as clouds (aerosols) with the subjectively selected threshold values. For example, the low cloud layers shown in Figure 3b are more likely aerosols instead of clouds. As a side note, the ARM MPL VAP cloud product identifies this layer as aerosols instead of clouds. However, for most cases in our study, as shown in the following sections, this VDE method with the selected threshold values can reasonably classify clouds and aerosols.

### 3. Evaluation

Three methods are used to assess the validity of this cloud detection algorithm: a synthetic test with simulations, an intercomparison with clouds determined manually from MPL signal returns [Campbell et al., 2002], and an intercomparison with clouds detected by the method currently used to generate the ARM MPL VAP [Wang and Sassen, 2001].

#### 3.1. A Synthetic Test

Based on equation (1), lidar signals without noise can be ideally generated with known scattering information about atmospheric gases, aerosols, and cloud particles. In this simulation, the contribution of atmospheric gases is

$$\begin{cases} \beta_m(z, \lambda) = 1.54 \times 10^{-3} \left(\frac{532}{\lambda}\right)^4 \exp\left(-\frac{z}{7}\right), \\ \alpha_m(z, \lambda) = \frac{8\pi}{3} \times \beta_m(z, \lambda) \end{cases}, \quad (9)$$

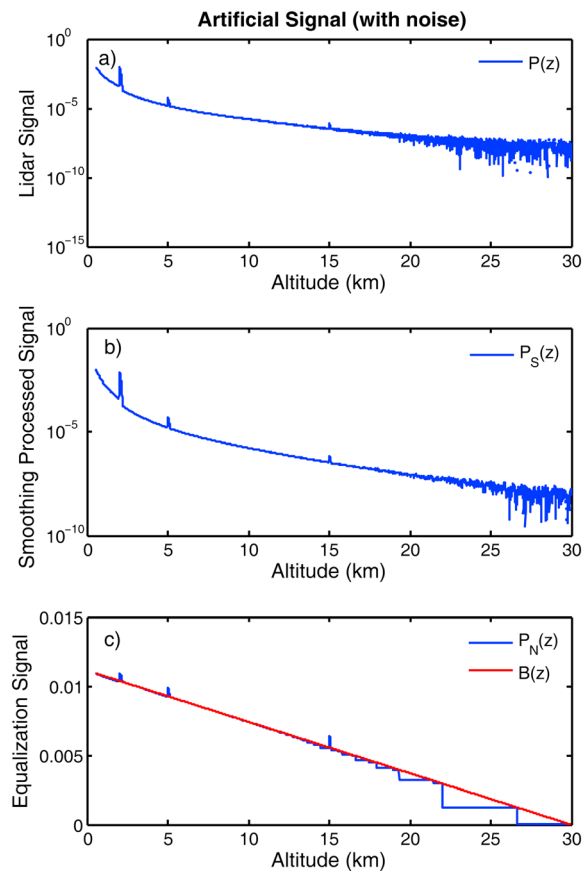
where  $\beta_m$  and  $\alpha_m$  are the backscattering coefficient and the extinction coefficient, respectively, of atmospheric gases,  $\lambda$  is the wavelength, and  $z$  is the altitude. The contribution of continuous atmospheric aerosols is expressed as

$$\begin{cases} \beta_a(z, \lambda) = \left\{ 2.47 \times 10^{-3} \exp\left(-\frac{z}{2}\right) + 5.13 \times 10^{-6} \exp\left[-\left(\frac{z-20}{6}\right)^2\right] \right\} \left(\frac{532}{\lambda}\right), \\ \alpha_a(z, \lambda) = 50 \times \beta_a(z, \lambda) \end{cases}, \quad (10)$$

where  $\beta_a$  and  $\alpha_a$  are the backscattering coefficient and the extinction coefficient, respectively, of aerosols.

We subjectively add three layers of lidar signals from strong to weak at 2 km, 5 km, and 15 km with depths of 200 m, 150 m, and 100 m, respectively. The simulated lidar signal  $P(z)$  is shown in Figure 4a. Figures 4b and 4c show the lidar signal after moving average  $P_s(z)$ , the VDE-processed signal  $P_N(z)$ , and baseline reference signal  $B(z)$ , generated using the method described in section 2. Three particle layers with almost the same depths as given are clearly detected, and they are identified as cloud layers based on the empirical threshold values. Note that the SDP processing is not applied to this case because no noise is present.

To further evaluate this method, noise was added to the simulated signals between 7 km and 30 km (far distance) as shown in Figure 5a. Figures 5b–5d show  $P_s(z)$ ,  $P_D(z)$ , and  $P_N(z)$  and  $B(z)$ . The cloud detection



**Figure 5.** Same as Figure 4, but with noise added to the synthetic lidar signals.

clouds generated by the VDE cloud detection algorithm, and differences between the ARM MPL VAP and VDE retrievals on 24 September 2012 at the SGP site. Both algorithms identify similar cloud boundaries, especially for single-layer clouds. However, two major differences are seen. For high-level clouds over low thin clouds, the VDE algorithm can give more continuous and accurate cloud information than can the ARM MPL VAP. Cases on other days also show this. As shown later in Figure 7, statistical results show considerable (~1–3%) differences between two algorithms for high-level clouds. This is consistent with the characteristics of the VDE retrieval algorithm which can reduce the magnitude of signal-to-noise variations with distance and make it possible for us to identify these clouds located high in the atmosphere. The other major difference is that cloud bases from the VDE algorithm are generally higher than those from the ARM MPL VAP, and cloud tops from the VDE algorithm are generally lower than those from the ARM MPL VAP. This is also consistent with the characteristics of VDE retrieval algorithm which keeps sharp signal changes at cloud bases and tops.

## 4. Results

The new MPL cloud detection algorithm is applied to 1 year of measurements collected at the SGP site and Taihu site. The following shows the cloud detection results and corresponding analyses.

### 4.1. The SGP Site

As indicated earlier, we consider the measurements made at the SGP site from May 2012 to May 2013. Figure 7 shows the diurnal variation in cloud frequency observed by the MPL using the VDE cloud detection algorithm and the ARM MPL cloud product [Wang and Sassen, 2001]. Both retrievals show almost the same diurnal and vertical variations: more clouds occur during the night than during the day; more clouds occur at heights of 3–6 km and 8–12 km during the night; and more clouds occur at heights of 0.5–1.5 km, 3–5 km, and 7–10 km during the day. These diurnal and vertical variations in cloud occurrence are very important for the

results show little sensitivity to the noise added to the simulated signals. The VDE algorithm has made all important signal returns detectable. All important layer information has been captured in Figures 4 and 5. Noise has been minimized in the VDE algorithm, and all sudden changes in signals have been included so that the boundaries of detected particle layers have a high accuracy. This is essential because some moving average methods can cause changes in particle layer boundaries.

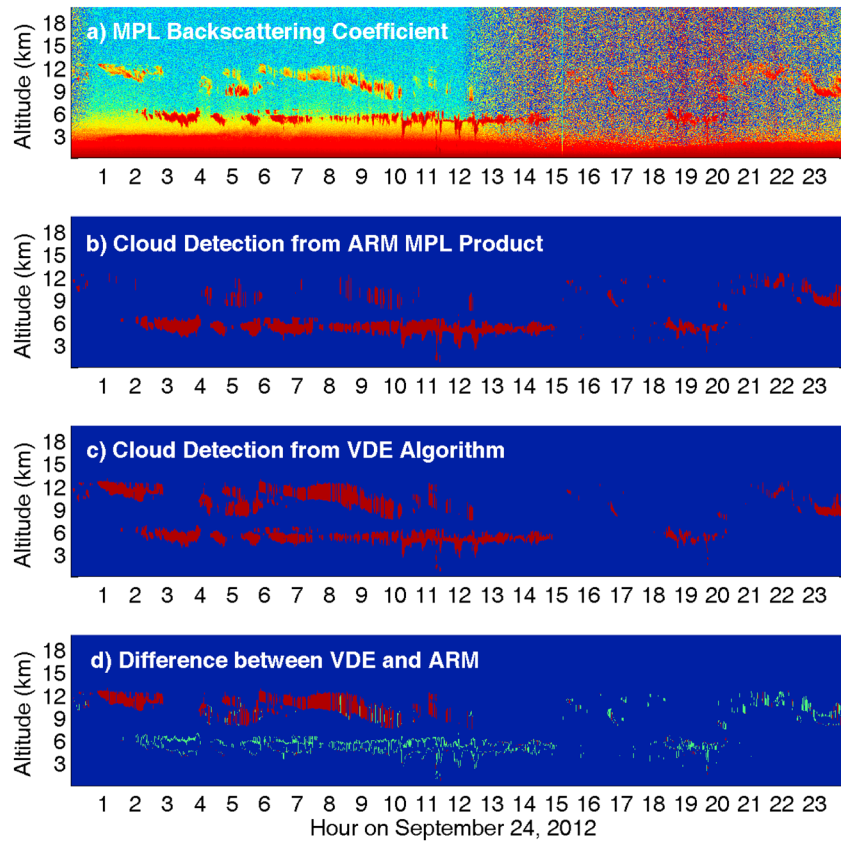
### 3.2. Evaluation With Lidar Signals From Real Clouds

As shown in Figure 3, the new algorithm can detect most clouds successfully and identify cloud boundaries accurately, including some thin multilayer clouds. Even for clouds identified during the noisy period of 1400 to 2300 Universal Time Coordinates (UT), the new algorithm can still identify clouds accurately. However, as indicated in section 2.2.3, some layers could be misclassified due to the threshold values used. For example, the ARM MPL VAP cloud product has classified the low layers between 0000 and 1300 UT as aerosols instead of clouds.

### 3.3. Comparison With the ARM MPL VAP Cloud Product

Figure 6 shows MPL backscattering coefficients, clouds from the ARM MPL VAP, clouds generated by the VDE cloud detection algorithm, and differences between the ARM MPL VAP and VDE retrievals on 24 September 2012 at the SGP site. Both algorithms identify similar cloud boundaries, especially for single-layer clouds. However, two major differences are seen. For high-level clouds over low thin clouds, the VDE algorithm can give more continuous and accurate cloud information than can the ARM MPL VAP. Cases on other days also show this. As shown later in Figure 7, statistical results show considerable (~1–3%) differences between two algorithms for high-level clouds. This is consistent with the characteristics of the VDE retrieval algorithm which can reduce the magnitude of signal-to-noise variations with distance and make it possible for us to identify these clouds located high in the atmosphere. The other major difference is that cloud bases from the VDE algorithm are generally higher than those from the ARM MPL VAP, and cloud tops from the VDE algorithm are generally lower than those from the ARM MPL VAP. This is also consistent with the characteristics of VDE retrieval algorithm which keeps sharp signal changes at cloud bases and tops.

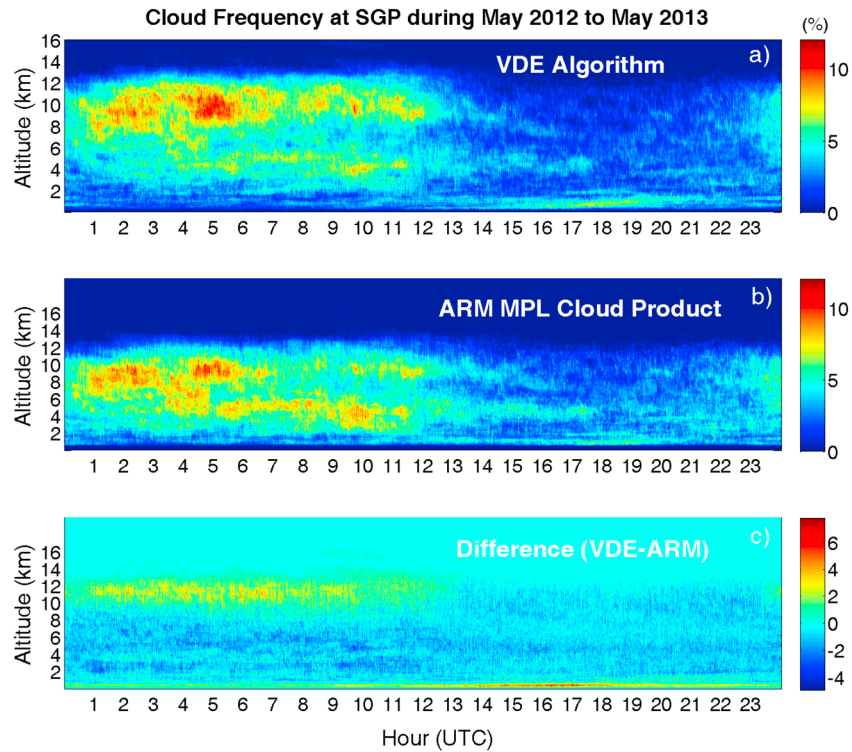




**Figure 6.** (a) MPL signal returns of the backscattering coefficient, (b) clouds from the ARM MPL VAP, (c) clouds detected by the VDE algorithm, and (d) differences in cloud occurrence between VDE retrievals and the ARM MPL VAP. In Figure 6d, the red areas indicate clouds detected by the VDE algorithm but not the ARM MPL VAP, and the green areas indicate clouds detected by the ARM MPL VAP but not the VDE algorithm. Data are for 24 September 2012 at the SGP site.

local radiative budget at the SGP site. More high clouds during the night and low clouds during the day could strengthen both the longwave and shortwave cloud radiative forcing compared to the daily average of cloud occurrence. These features should be considered when studying cloud radiative forcing. The vertical variation in cloud frequency shown in Figure 7 is qualitatively consistent with the findings reported by *Xi et al.* [2009] in that there is a distinct bimodal vertical distribution of clouds at the SGP site. *Xi et al.* [2009] also found an upper peak in cloud frequency between 7.5 km and 10.7 km which varied with season and a lower peak in cloud frequency between 2 km and 3 km which is lower than what is shown in Figure 7. The different study periods may explain this: this study uses data from 1 year while *Xi et al.* [2009] used 10 years of data to carry out their statistical study.

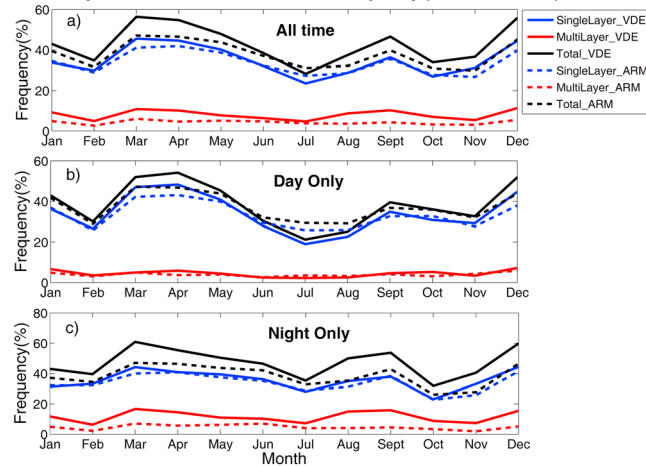
Figure 8 shows the monthly variation in cloud frequency at the SGP site for total, single-layer, and multilayer clouds for all times of the day, daytime only, and nighttime only. The solid lines are for cloud frequencies derived using the VDE method, and the dashed lines are for cloud frequencies obtained from the ARM MPL VAP. The total cloud frequency for all times of the day varies with season, with a maximum value of ~58% (VDE method) or 50% (ARM MPL VAP) in winter and spring and a minimum value of ~35% (both the VDE method and ARM MPL VAP) in summer and fall. This is consistent with other findings [*Dong et al.*, 2006; *Xi et al.*, 2009]. The annual averaged total cloud frequency from the VDE method based on all times of the day is about 50% which is also consistent with the value of 49% found by *Dong et al.* [2006] and 51% found by *Xi et al.* [2009]. By contrast, the annual averaged total cloud frequency from the ARM MPL VAP based on all times of the day is about 45%, less than the values reported by *Dong et al.* [2006] and *Xi et al.* [2009]. MPL-detected clouds are mainly classified as single-layer clouds (80%). This value is much larger than the value of 61.5% found by *Xi et al.* [2009]. Note that *Xi et al.* [2009] used a combination of MPL and MMCR retrievals which differs from the input used to generate the ARM MPL VAP. The higher percentage of single-layer clouds



**Figure 7.** Diurnal variation in cloud frequency observed by the MPL using (a) the VDE algorithm and (b) the ARM MPL VAP. (c) The diurnal variation in cloud frequency differences between the two algorithms at the SGP site.

detected by the MPL may be associated with the strong attenuation of lidar signals when clouds are thick. Therefore, the MPL cloud detection method could have relatively large uncertainties (underestimation) when detecting multilayer clouds. However, the VDE method identifies about 5% more clouds than those from the ARM MPL VAP, especially multilayer clouds at night. This results in more agreeable cloud frequency

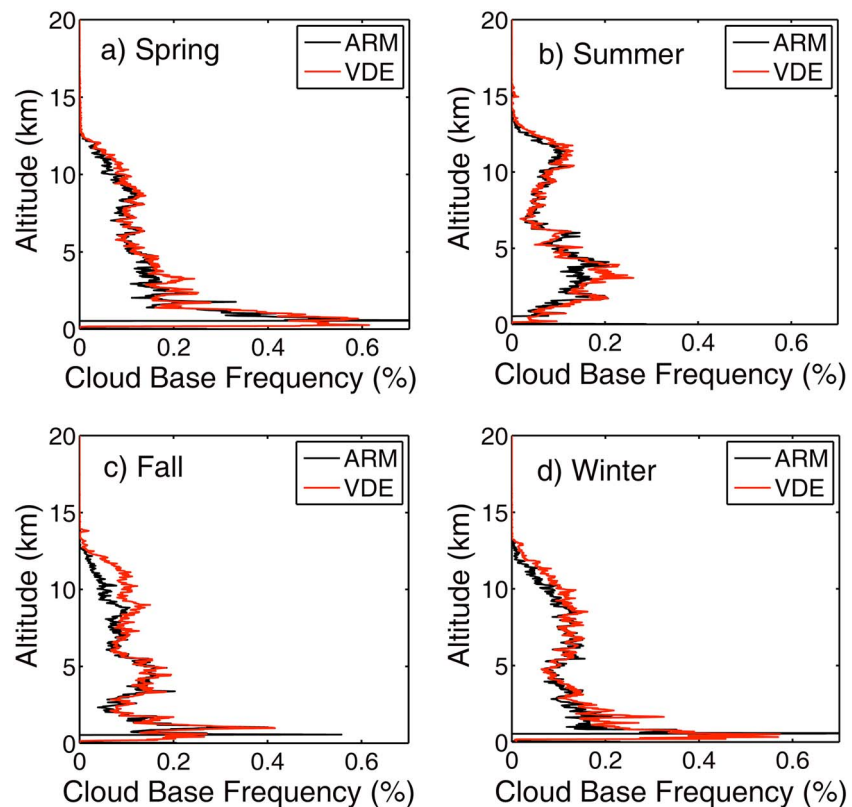
**Monthly Variation of Cloud Occurrence Frequency (SGP, 2012–2013)**



**Figure 8.** Monthly variations in cloud frequency for total, single-layer, and multilayer clouds for (a) all times, (b) daytime only, and (c) nighttime only at the SGP site. Solid lines are for cloud frequencies calculated using the VDE algorithm, and dashed lines are for cloud frequencies from the ARM MPL VAP.

observations with those found by both radar and lidar-based studies [Dong et al., 2006; Xi et al., 2009]. A similar seasonal variation is found for both daytime only and nighttime only cloud frequencies. Consistent with the results found in Figure 7, the total cloud frequency for nighttime only is slightly larger than that for daytime only. Moreover, less multilayer cloud frequencies are observed during the day (<10%) than during the night (>10%). Figure 7 have shown that low-boundary clouds (0.5–1.5 km) occur more often during the day which could result in the detection of less multilayer clouds during the day than at night.

Figure 9 shows vertical distributions of the frequency of cloud base occurrence determined from the VDE algorithm based on MPL observations made during spring (March, April, and May),



**Figure 9.** Vertical variation in the frequency of cloud base occurrence determined from the VDE algorithm based on MPL observations made during (a) spring, (b) summer, (c) fall, and (d) winter from 2012 to 2013 at the SGP site. Black lines are for cloud frequencies obtained from the ARM MPL VAP, and red lines are for cloud frequencies derived from the VDE algorithm.

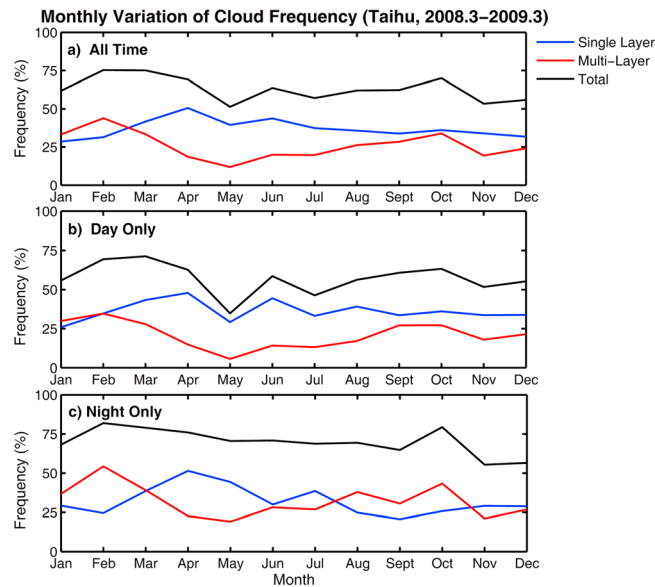
summer (June, July, and August), fall (September, October, and November), and winter (December, January, and February). These vertical distributions vary with season. In spring and winter, boundary layer clouds with bases  $\leq 1$  km dominate. Figure 9 also shows a second weak peak of cloud base frequency around 6–10 km. The dominant low clouds at the SGP site are stratiform clouds in spring and winter as shown by *Zhao et al.* [2014]. In summer, there is a very obvious bimodal vertical distribution of cloud base occurrence. Two peaks lie around 3 km and 11 km which are most likely associated with convective clouds prevalent at the SGP site during the summer [Zhao et al., 2014]. In fall, the vertical distribution of cloud base occurrence exhibits a triple-mode pattern with peaks at 1 km, 4 km, and 8 km. This might be associated with the transition of dominant cloud types from convective to stratiform.

Vertical distributions of cloud base occurrence frequency obtained from the ARM MPL VAP are also shown in Figure 9. The same seasonal and vertical variations as those from data generated by the VDE algorithm are seen. However, the VDE algorithm shows more clouds with high bases, including clouds with bases between 3 and 5 km in summer, and with bases between 7 and 14 km in other seasons. The differences are particularly significant between 7 and 14 km in fall. These differences should be related to their differences in the methods. The VDE method is designed to increase the visibility of lidar signals at far distances so it can reliably identify more clouds higher in the atmosphere than can the ARM MPL VAP.

#### 4.2. The Taihu Site

For the China Taihu site, we consider the whole observation period which is from March 2008 to March 2009. The time resolution is 3 min, and the total valid data number is 128,586.

Figure 10 shows the monthly variation in cloud frequencies for total, single-layer, and multilayer clouds during all times of the day, daytime only, and nighttime only. Different from the SGP site, there are no clear seasonal variations in cloud frequency at the Taihu site while slightly larger (but not significant) values are found during the transition months of February to April. The annually averaged total cloud frequency for all



**Figure 10.** Monthly variations in cloud frequency for total, single-layer, and multilayer clouds during (a) all times, (b) daytime only, and (c) nighttime only at the Taihu site.

times of the day is ~65% with ~20% identified as multilayer clouds. This annually averaged total cloud frequency is about 15% larger than that at the SGP site, which is most likely due to the moist climate of the region surrounding the Taihu site. Zhang *et al.* [2010] analyzed radiosonde measurements from May to December 2008 and found an average cloud frequency of about 78.5% over ShouXian in Anhui province (about 500 km away from the Taihu site). These different findings could be related to the site, instrument, and sampling differences. Clouds separated by a distance of 500 km may be weakly correlated or uncorrelated. In addition, radiosondes are typically launched 4 times a day so the sampling frequency is much less than that of the MPL. Differences may also arise from the different methods used. The method

used to identify cloud layers from radiosonde measurements was trained using the cloud classifications of the Active Remote Sensing of Cloud product [Zhang *et al.*, 2010] so any errors in the latter are transformed into errors in the former. MPL observations also show more multilayer clouds at the Taihu site than at the SGP site. However, large uncertainties could exist in the detection of multilayer clouds by the MPL. Also, clouds at the Taihu site are likely contaminated by aerosols which can result in the overestimation of cloud frequency. Quantifying the impact of aerosols on cloud detection at many heavily aerosol-laden regions in China is challenging. However, since aerosol layer boundaries are not as clearly defined as cloud layer boundaries, the VDE algorithm will pick up the sharp signal variations near cloud boundaries. An accurate quantification of aerosol impacts on cloud detection is beyond the scope of the current study.

Similar monthly variations in cloud frequency are seen for both daytime only and nighttime only clouds at the Taihu site. Single-layer clouds are not as dominant at the Taihu site as they are at the SGP site. At night, single-layer and multilayer clouds have roughly the same occurrence frequency. Considering that the MPL tends to underestimate multilayer clouds due to the attenuation of MPL lidar signals, about 40% or more of detected clouds are multilayer clouds at the Taihu site.

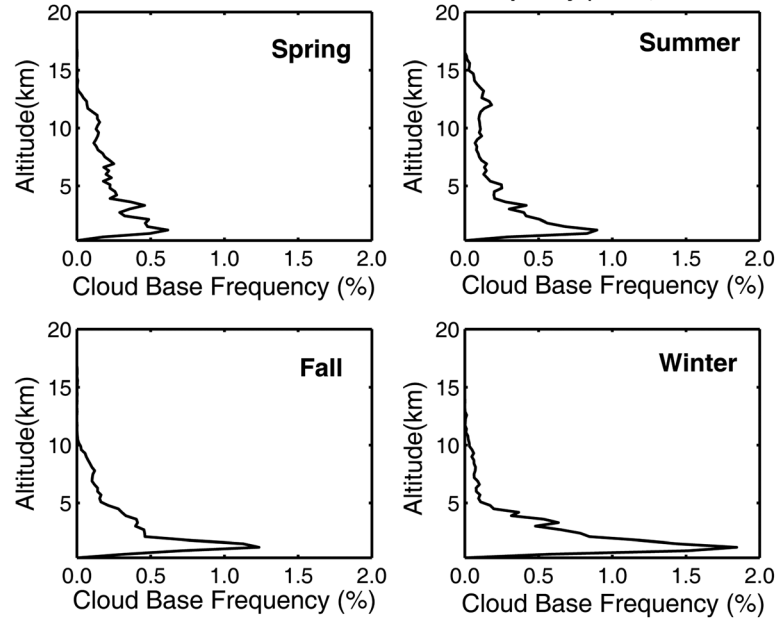
Figure 11 shows the seasonal variation in vertical profiles of cloud base occurrence frequency at the Taihu site. Different from the SGP site, the vertical distribution of cloud base frequencies follows a single mode distribution with a maximum occurrence around 1 km for all four seasons. More clouds with low bases occur in winter than in other seasons. High clouds with bases above 5 km mainly occur in summer and spring. There are almost no clouds with bases above 10 km in fall and winter.

Vertical distributions of cloud frequency during the daytime and during the nighttime are shown in Figure 12. The highest cloud frequencies are found at low altitudes. About 50% or more of these low-level clouds are likely stratiform clouds located between 0.5 and 3 km. High clouds shown in Figure 12 are likely either high-level ice clouds or convective clouds. Figure 12 also shows that more high-level clouds occur at night than during the day. This would trap more longwave radiation within the atmosphere at night and cause weak diurnal variations in temperature.

### 5. Summary

A new cloud and aerosol detection algorithm based on MPL observations has been developed. This new method has two promising characteristics: it keeps the sharp signal variations at cloud bases and tops for

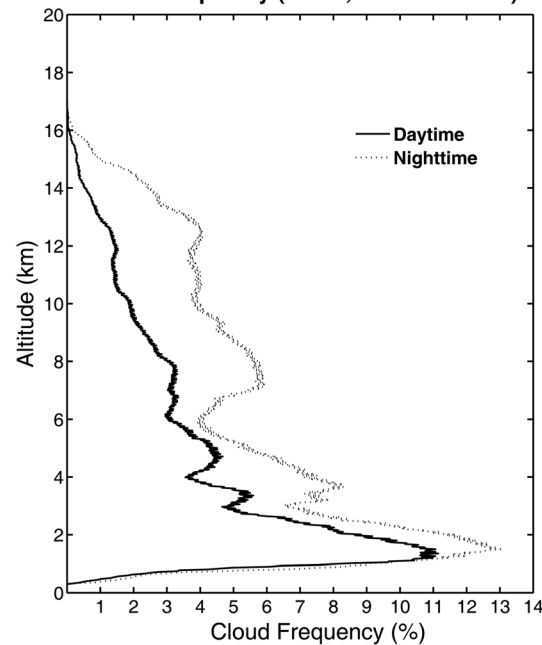
Vertical Distribution of Cloud Base Occurrence Frequency (Taihu, 2008.3–2009.3)



**Figure 11.** Seasonal variations in the vertical distribution of cloud base occurrence frequencies at Taihu. Data are from March 2008 to March 2009.

reliable cloud boundary identification using the SDP technique with limited moving averages, and dynamically increases the relative magnitude of detection signals at far distances using the VDE algorithm. Compared to existing retrieval algorithms, this algorithm can identify most aerosol/cloud layers with high accuracy including those layers with weak signal-to-noise ratios at far distances. Empirical threshold values are used to classify clouds from aerosols. Below 3 km, layers are classified as clouds for  $T > 3$  or  $D < -7$ ; above 3 km, layers are classified as clouds for  $T > 1.5$  or  $D < -7$ . The cloud/aerosol classification scheme is sensitive to selected threshold values, indicating the possibility of misclassification between aerosol and clouds.

Cloud Frequency (Taihu, 2008.3–2009.3)



**Figure 12.** Vertical distributions of annually averaged cloud occurrence frequencies at the Taihu site for daytime only (solid line) and nighttime only (dashed line).

The cloud detection algorithm has been evaluated with a synthetic test, a case study, and a systematic comparison with the ARM MPL VAP. The evaluation shows that the new method can successfully classify clouds with high reliability. Compared to the ARM MPL cloud product, similar results are obtained, especially for low clouds, but considerable differences are seen in clouds at high altitudes. The VDE algorithm can detect continuous middle or high clouds while the ARM MPL cloud product shows scattered broken clouds which appear to be detection artifacts. Same as the large uncertainties in cloud microphysical property retrievals [Zhao et al., 2012], errors could exist in the MPL-detected clouds. For example, the attenuation of lidar signals has limited the observation of multilayer clouds above low thick clouds.

Applying this VDE cloud detection algorithm to 1 year of observations made at the U.S. SGP site and the China Taihu site, the temporal variation and

vertical distribution of cloud occurrence frequencies have been examined. At the SGP site, a bimodal vertical distribution in cloud occurrence frequency is seen, with peaks of occurrence around 3–6 km and 8–12 km. More clouds are detected at night than during the day, which is important for understanding or simulating the diurnal variation of the surface energy balance. A clear seasonal variation in cloud occurrence frequency is found: maximum cloud frequencies occur in winter and spring, and minimum in summer and fall. Analysis of the vertical distribution of cloud base occurrence frequency suggests that the dominant cloud types are stratiform in winter and convective in summer. These results are consistent with findings based on lidar-radar data [Dong *et al.*, 2006; Xi *et al.*, 2009; Zhao *et al.*, 2014].

The temporal variation and vertical distribution of cloud occurrence frequencies at the Taihu site are different. There are no clear seasonal variations in the cloud occurrence frequency. The annual average cloud frequency is around 65%, which is about 15% higher than that at the SGP site and 13% lower than that at another site in China (ShouXian). About 40% or more of detected clouds are multilayer clouds at Taihu site. The vertical distribution of cloud base frequencies follows a single mode distribution with a maximum occurrence around 1 km for all four seasons. Stratiform clouds between 0.5 and 3 km likely dominate at the site. More clouds with low bases occur in winter than in other seasons. High clouds with bases above 5 km mainly occur in summer and spring. More high clouds occur at night than during the day. Further studies, including model simulations, are needed to get a more accurate physical understanding of the findings presented here.

#### Acknowledgments

This work was supported by the Ministry of Science and Technology of China (grants 2013CB955802 and 2012AA120901), the Chinese Program for New Century Excellent Talents in University (NCET), the Chinese National Science Foundation (CNSF; grant 91337103), and the Fundamental Research Funds for the Central Universities. The data used in this study are supported by the Ministry of Science and Technology of China and the U.S. DOE ARM program.

#### References

- Ackerman, T. P., and G. M. Stokes (2003), The Atmospheric Radiation Measurement program, *Phys. Today*, *56*, 38–44.
- Campbell, J. R., D. L. Hlavka, E. J. Welton, C. J. Flynn, D. D. Turner, J. D. Spinhirne, V. S. Scott, and I. H. Wang (2002), Full-time, eye-safe cloud and aerosol lidar observation at Atmospheric Radiation Measurement Program sites: Instruments and data processing, *J. Atmos. Oceanic Tech.*, *19*, 431–442.
- Campbell, J. R., K. Sassen, and E. J. Welton (2008), Elevated cloud and aerosol layer retrievals from micropulse lidar signal profiles, *J. Atmos. Oceanic Tech.*, *25*, 685–700, doi:10.1175/2007JTECHA1034.1.
- Chang, F., and Z. Li (2005), A near-global climatology of single-layer and overlapped clouds and their optical properties from Terra/MODIS data using a new algorithm, *J. Clim.*, *18*, 4752–4771.
- Clothiaux, E. E., M. A. Miller, B. A. Albrecht, T. A. Ackerman, J. Verlinde, D. M. Babb, R. M. Peters, and W. J. Syrett (1995), An evaluation of a 94-GHz radar for remote sensing of cloud properties, *J. Atmos. Oceanic Tech.*, *12*, 201–229.
- Clothiaux, E. E., G. G. Mace, T. P. Ackerman, T. J. Kane, J. D. Spinhirne, and V. S. Scott (1998), An automated algorithm for detection of hydrometeor returns in micro pulse lidar data, *J. Atmos. Oceanic Tech.*, *15*, 1035–1042.
- Clothiaux, E. E., T. P. Ackerman, G. C. Mace, K. P. Moran, R. T. Marchand, M. A. Miller, and B. E. Martner (2000), Objective determination of cloud heights and radar reflectivities using a combination of active remote sensors at the ARM CART Sites, *J. Appl. Meteorol.*, *39*, 645–665.
- Coulter, R. (2012), Micropulse Lidar (MPL) handbook, *ARM Tech. Rep. TR-019*, 12 pp.
- Ćurić, M., and D. Janc (1993), Predictive capabilities of a one-dimensional convective cloud model with forced lifting and a new entrainment formulation, *J. Appl. Meteorol.*, *32*, 1733–1740.
- Ćurić, M., D. Janc, and V. Vučković (2008), Precipitation change from a cumulonimbus cloud downwind of a seeded target area, *J. Geophys. Res.*, *113*, D11215, doi:10.1029/2007JD009483.
- Dong, X., B. Xi, and P. Minnis (2006), A climatology of midlatitude continental clouds from the ARM SGP central facility. Part II: Cloud fraction and surface radiative forcing, *J. Clim.*, *19*, 1765–1783.
- Han, J. H., S. Yang, and B. U. Lee (2011), A novel 3-D color histogram equalization method with uniform 1-D gray scale histogram, *IEEE Trans. Image Process.*, *20*(2), 506–512.
- Inspurger, T., and G. Stepan (2010), *Semi-discretization for time-delay systems—Stability and engineering applications*, Springer, New York.
- Intergovernmental Panel on Climate Change (2007), *Climate Change 2007: IPCC Fourth Assessment Report*, Cambridge Univ. Press, Cambridge, U. K.
- Li, Z., et al. (2011), East Asian studies of tropospheric aerosols and their impact on regional climate (EAST-AIRC): An overview, *J. Geophys. Res.*, *116*, D00K34, doi:10.1029/2010JD015257.
- Mendoza, A., and C. Flynn (2006), *Micropulse Lidar (MPL) Handbook*, ARM TR-019, U.S. Department of Energy, U.S. Department of Energy, Washington, D. C.
- Morris, V. R. (2012), *Vaisala Ceilometer (VCEIL) Handbook*, DOE/SC-ARM-TR-020, U.S. Department of Energy, Washington, D. C.
- Platt, C. M. R., et al. (1994), The experimental cloud lidar pilot study (ECLIPS) for cloud-radiation research, *Bull. Am. Meteorol. Soc.*, *75*, 1635–1645.
- Sassen, K. (1995), Lidar cloud research, *Rev. Laser Eng.*, *23*, 148–153.
- Spinhirne, J. D. (1993), Micro pulse lidar, *IEEE Trans. Geosci. Remote Sens.*, *31*, 48–55.
- Spinhirne, J. D., J. A. R. Rall, and V. S. Scott (1995), Compact eye safe lidar systems, *Rev. Laser Eng.*, *23*, 112–118.
- Wang, Z., and K. Sassen (2001), Cloud type and macrophysical property retrieval using multiple remote sensors, *J. Appl. Meteorol.*, *40*, 1665–1682.
- Warren, S. G., C. J. Hahn, J. London, R. M. Chervin, and R. L. Jenne (1984), Atlas of simultaneous occurrence of different cloud types over land, *NCAR Tech. Note, NCAR/TN-241 + STR*, 209 pp., Natl. Cent. Atmos. Res., Boulder, Colo.
- Xi, B., X. Dong, P. Minnis, and M. Khaiyer (2009), A 10 yr climatology of cloud cover and vertical distribution from both surface and GOES observations over DOE ARM SGP site, *J. Geophys. Res.*, *115*, D12124, doi:10.1029/2009JD012800.
- Zhang, J., H. Chen, Z. Li, X. Fan, L. Peng, Y. Yu, and M. Cribb (2010), Analysis of cloud layer structure in Shouxian, China using RS92 radiosonde aided by 95 GHz cloud radar, *J. Geophys. Res.*, *115*, D00K30, doi:10.1029/2010JD014030.

- Zhang, Z., Z. Li, H. Chen, and M. Cribb (2013), Validation of a radiosonde-based cloud layer retrieval method using ground-based remote sensing methods at multiple ARM sites, *J. Geophys. Res. Atmos.*, *118*, 846–858, doi:10.1029/2012JD018515.
- Zhao, C., et al. (2012), Toward understanding of differences in current cloud retrievals of ARM ground-based measurements, *J. Geophys. Res.*, *117*, D10206, doi:10.1029/2011JD016792.
- Zhao, C., S. Xie, X. Chen, M. P. Jensen, and M. Dunn (2014), Quantifying uncertainties of cloud microphysical property retrievals with a perturbation method, *J. Geophys. Res. Atmos.*, *119*, 5375–5385, doi:10.1002/2013JD021112.