Remote Sensing of Environment 266 (2021) 112687

Contents lists available at ScienceDirect

Remote Sensing of Environment

journal homepage: www.elsevier.com/locate/rse

A simple multiscale layer detection algorithm for CALIPSO measurements

Feiyue Mao [a,b], Zhenxing Liang [b,*], Zengxin Pan [c], Wei Gong [b,d], Jia Sun [e], Tianhao Zhang [b], Xin Huang [a], Lin Zang [f], Xin Lu [b], Jia Hong [b]

[a] School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430079, China
[b] State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China
[c] Institute of Earth Sciences, Hebrew University of Jerusalem, Jerusalem 91904, Israel
[d] Electronic Information School, Wuhan University, Wuhan 430079, China
[e] School of Geography and Information Engineering, Chinese University of Geosciences, Wuhan 430079, China
[f] Chinese Antarctic Centre of Surveying and Mapping, Wuhan University, Wuhan 430079, China

ARTICLE INFO

Editor: Dr. Menghua Wang

Keywords:
CALIOP
Aerosol and cloud
Layer detection
Multiscale

ABSTRACT

The Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observations (CALIPSO) is unique in its ability to perform profiling measurements of aerosol and cloud layers globally. Detecting the layer boundaries of aerosols and clouds is a crucial step in CALIPSO data retrieval. The CALIPSO team uses the selective iterated boundary location (SIBYL) algorithm based on threshold arrays to find aerosol and cloud layers at different horizontal resolutions. However, threshold arrays could obstruct the detection of optically tenuous layers at a high resolution and may cause overestimation when averaging signals of layer and clear air at a low resolution. Here, a multiscale algorithm using a series of sliding window sizes without threshold setting is proposed based on a predefined probability. The results over land and marine areas show that the multiscale algorithm detected 37.41% and 16.36% more layer area than the SIBYL at 1–80 km resolutions at daytime and 1–5 km resolutions at night time, respectively. This indicates that the multiscale algorithm does not need a threshold array, allowing more tenuous layers to be detected, especially at low signal to noise ratios (SNRs). In contrast, the SIBYL detects 4.40% more layer area than the multiscale algorithm at 1–80 km resolutions at nighttime, mainly caused by the large proportion of layer area detected by SIBYL at 20 and 80 km resolutions. This implies possible noteworthy overestimation by the SIBYL at low resolutions. Additionally, the evaluation using the depolarization ratio of ice clouds shows that the extra detected layers by the multiscale algorithm are reliable. Besides, simulation tests show that the multiscale and SIBYL algorithms achieve a 100% true detection rate when SNR is approximately 2 and 4, respectively. The new multiscale algorithm could upgrade the resolution and accuracy of the layer detection of space lidars and reduce the underestimation of layer optical depth due to missing layers.

1. Introduction

Atmospheric aerosols and clouds play essential roles in climate change by affecting the radiation budget and circulation pattern of the Earth-atmosphere climate system through the absorption and reflection of solar radiation (Bi et al., 2014; Chen et al., 2021; Li et al., 2019; Ma et al., 2018; Ramanathan et al., 2001). The vertical and horizontal extent of clouds and aerosols should be identified accurately to facilitate the understanding of their properties (Xie et al., 2013; Zhao and Garrett, 2015). Many advanced tools have been developed to address this issue, of which lidar is a powerful instrument for active remote sensing (Bian et al., 2020; Guo et al., 2016; Liu et al., 2019a; Wang and Zhao, 2017; Winker et al., 2010; Winker et al., 2003; Xie et al., 2017; Zhang et al., 2019). The Cloud-Aerosol Lidar with Orthogonal Polarization (CALIOP), which is carried on the Cloud-Aerosol Lidar Pathfinder Satellite Observation (CALIPSO) satellite, has been conducted observations for more than ten years. A considerable amount of critical information has been provided by CALIPSO for studying global climate and environmental changes (Berry and Mace, 2014; Feofilov et al., 2015; Fu et al., 2017; Li et al., 2017; Winker et al., 2007).

It is critical to detect the aerosol and cloud layer boundaries in lidar data accurately, because their uncertainties will be passed to the subsequent layer classification and extinction retrieval (Ma et al., 2018). Over the years, researchers have proposed many methods for locating

* Corresponding author.
E-mail address: liangzhenxing@whu.edu.cn (Z. Liang).

https://doi.org/10.1016/j.rse.2021.112687
Received 11 February 2021; Received in revised form 30 June 2021; Accepted 2 September 2021
Available online 14 September 2021

the layer boundaries in lidar data. Most layer detection algorithms used by the lidar community fall into two categories, namely, slope and threshold methods (Vaughan et al., 2005; Zhao et al., 2014). The slope method uses the "zero-crossing" behavior of the first derivative of the original signal, of which calibration is not needed to identify the base of the layer (Comerón et al., 2013; Pal et al., 1992). However, the slope method requires that the signal is strong enough and continually increasing (or decreasing) to determine whether the slope is sequentially positive or negative. Otherwise, strong instrument noises will adversely affect the detection. Thus, the slope method is inapplicable for space lidar because of its low signal-to-noise ratio (SNR) (Vaughan et al., 2005).

The threshold method is the most widely-used method for space lidars to determine a layer by detecting whether a signal in a certain height range exceeds a series of thresholds (Clothiaux et al., 1998; Mao et al., 2018; Melfi et al., 1985; Wang et al., 2021; Winker and Vaughan, 1994). Vaughan et al. (2009) proposed the selective iterated boundary location (SIBYL) algorithm based on the threshold method and Attenuated Scattering Ratio (ASR). The SIBYL comprises two physical models: the first is a single-profile scanner based on the adaptive threshold technique, and the second is a multi-profile averaging (1, 5, 20, and 80 km) mechanism for detecting tenuous layers. By applying the single profile scanner to the profiles generated by the multi-profile averaging mechanism, layer detection results at different resolutions can be obtained. The SIBYL can successfully find a strong layer, in which the signal continuously exceeds the threshold array. However, the SIBYL may miss the tenuous layers and layer edges with a weak signal increase near the top and base of the layers in its detection (Balmes and Fu, 2018; Davis et al., 2010; Thorsen et al., 2017; Yorks et al., 2011). The missing detection is due to that the noise suppresses weak signals continuously exceeding the threshold.

Averaging multiple signals can improve the SNR; however, this process tends to overestimate the layer coverage (averaging a signal of clear air with that of a layer is considered equivalent to a layer covering the entire averaged space), thereby hindering balancing the spatial resolution and accuracy of layer detection (Cesana et al., 2016; Hagihara et al., 2010b). Further, many studies have reported that the CALIPSO AOD was substantially underestimated due to the missing layers compared with various observations, such as ground sunphotometer and MODIS (Kacenelenbogen et al., 2014; Kim et al., 2013; Redemann et al., 2011; Thorsen et al., 2017). Vaillant de Guélis et al. (2021) and Hagihara et al. (2010b) proposed advanced two-dimensional detection methods, which are applied to a 2D lidar scene instead of a 1D lidar profile, but still relied on the threshold array. Mao et al. (2021) proposed a detection method based on automatic segmentation with a minimum cost function, as an improvement of the SIBYL, which also requires the threshold array.

It is necessary to consider a new detection method that is insensitive to noise and independent of the threshold array, while yielding complete and accurate layer information at a high resolution of 1 or 5 km. Such a method would avoid the need to average a large number of profiles, which can reduce the chances of false detection, particularly at 20 km and 80 km resolutions. Previously, we have proposed a multiscale layer detection algorithm for ground-based lidar signals (Mao et al., 2011). This method first determines the increasing and decreasing trends of each range bin of uncalibrated ground-based lidar at different scales (i.e., a series of sliding window sizes), and then defines the layer based on multiscale detections (i.e., combines detections on different scales). The advantage of this method is its relative insensitivity to noise. Meanwhile, the layer boundary obtained by combining multiscale information is more accurate. However, considering the extremely low SNR of a space lidar, the algorithm that directly performs layer detection based on the signal difference between adjacent bins is inapplicable.

Overall, the detection of optically tenuous layers is still challenging for space lidars. Thus, we proposed a multiscale algorithm based on probability computations in this study. First, we calculated the probabilities that each bin belonged to clear air at each scale. Second, we labeled a bin as a layer bin when its probability of belonging to clear air was very small at each scale. Finally, we combined the multiscale information to use both of the advantages of the robust detection of the large scales and the specialty of the small scales in locating the layer edges. The new algorithm can detect tenuous layers more effectively than the previous algorithms because the probabilities could determine that given bins are layer bins, before their slope or ASR are continuously changed or larger than the threshold array.

2. Data and methods

2.1. Attenuated scattering ratios (ASR)

Consistent with the SIBYL, the new algorithm was based on the ASR array, which was calculated from the total attenuated backscatter coefficients (from the CALIPSO level-1 version-4 datasets) and meteorological data (from the CALIPSO "Met Manager" datasets), as follows (Vaughan et al., 2009; Winker et al., 2009):

$$ASR(z) = \frac{\beta'_{Total}(z)}{\beta'_{air}(z)} = \left(1 + \frac{\beta_p(z)}{\beta_m(z)}\right) \cdot T_p{}^2(z) \tag{1}$$

where z is the altitude, β'_{Total} represents the measured total attenuated backscatter coefficients, and β'_{air} represents the model of the attenuated backscatter coefficients that would be expected in ideal clear air. β_p and β_m represent the backscatter coefficients of the particles and molecules, respectively, and $T_p{}^2$ refers to the two-way transmittance of the particles.

The ideal ASR (ASR_{ideal}) represents the attenuated scattering ratio of the ideal clear air atmosphere without noise, which initial value should be equal to one according to Eq. (1) because the particle attenuation in the clear regions can be disregarded. If a new ith layer is detected, then the ideal ASR should be updated by the following:

$$ASR_{ideal}(i+1) = ASR_{ideal}(i) \cdot T^2_{layer}(i) \tag{2}$$

where T^2_{layer} represents the two-way transmittance of the last detected layer, which can be calculated as follows (Vaughan et al., 2005):

$$T^2_{layer} = 1 - 2 \cdot \gamma'_{layer} \cdot S_{reasonable} \tag{3}$$

where γ'_{layer} is the layer-integrated attenuated backscatter, which can be estimated by the following (Vaughan et al., 2005):

$$\gamma'_{layer} = g - \left(\frac{1}{2} \cdot (z_{top} - z_{base}) \cdot (\Re_{top} + \Re_{base})\right) \tag{4}$$

where z_{top} and z_{base} represent the top and base heights of the layer, respectively. $\Re_k$ and g can be calculated as follows, here, k indicates the position of the current bin:

$$\begin{gathered} \Re_k = \beta_m \cdot ASR(z_k) \\ g = \frac{1}{2} \cdot \sum_{k=top+1}^{base} (z_{k-1} - z_k) \cdot (\Re_{k-1} + \Re_k) \end{gathered} \tag{5}$$

where $S_{reasonable}$ represents the reasonable lidar ratio. In the SIBYL, the $S_{reasonable}$ is a given empirical value, which may cause large uncertainty in the ASR_{ideal} update. In this study, we adopted the iterative algorithm proposed by Mao et al. (2018), which can more accurately estimate the $S_{reasonable}$ of a transparent layer. The main idea is to iteratively try different lidar ratio S_j (j = 1, …, num, num is iteration number) to generate different $T_{layer}{}^2(i, S_j)$ and $ASR_{ideal}(i, S_j)$ below the layer i based on Eq. (3). Then, the reasonable lidar ratio is determined when the following $Fun(S_j)$ is equal to one:

$$Fun(S_j) = ASR_{ideal}(i, S_j) \bigg/ \overline{\sum_{z_{base}}^{z_{next-top}} ASR(z)} \tag{6}$$

where $\overline{\sum_{z_{base}}^{z_{next-top}} ASR(z)}$ refers to the mean value of the measured $ASR(z)$ of the clear air from the current layer base to the next layer top detected based on $ASR_{ideal}(i, S_j)$. One essential advantage of the technology of Mao et al. (2018) is that the updated ASR_{ideal} will optimally fit the signal of the clear region under the overlying layers mathematically, which avoids causing errors in the underlying layer detection due to inaccurate update of ASR_{ideal} using a fixed lidar ratio. More details can be found in Mao et al. (2018).

2.2. Definition of the probability

A measured lidar signal could be considered to consist of an ideal pure signal and Poisson noise, which approximately follows the Gaussian distribution for a sufficiently strong signal (Vaughan et al., 2005). Therefore, the probability that a measured ASR is greater or smaller than the ideal ASR is considered the same (i.e., ½). For selected m bins, we simply assume that each bin is independent, although there is a certain relationship between the bins (Vaillant de Guélis et al., 2021). In probability and statistical theory, the behave of the m selected bins is equivalent to m repeated Bernoulli experiments (Basu and Pereira, 1990), that is, obeying the binomial distribution with parameters m and p. For m bins, the probability that exactly u bins are larger than the ideal ASR, is:

$$\begin{aligned} P(u) &= C(m,u)\cdot p^u\cdot(1-p)^{m-u} \\ C(m,u) &= \frac{m!}{u!\cdot(m-u)!} \end{aligned} \tag{7}$$

where $C(m, u)$ represents the coefficient of the binomial distribution, the p is equal to ½ in this study. Thus, for the measurement of clear air, the probability that greater than or equal to u bins have ASR values larger than the ideal ASR can be calculated as follows:

$$P_{Clear}(m,u) = \sum_{u}^{m} C(m,u)\cdot\left(\frac{1}{2}\right)^m \tag{8}$$

The calculation results of $P_{Clear}(m, u)$ under different numbers of selected bins are shown in Fig. 1. For a fixed m, the probability that the selected bins belong to clear air decreases with the increase of their ASR values compared with the ideal ASR. Therefore, a very small $P_{Clear}(m, u)$ indicates that this situation is unlikely to happen if the air is clear, i.e., the selected bins are more likely to belong to a layer.

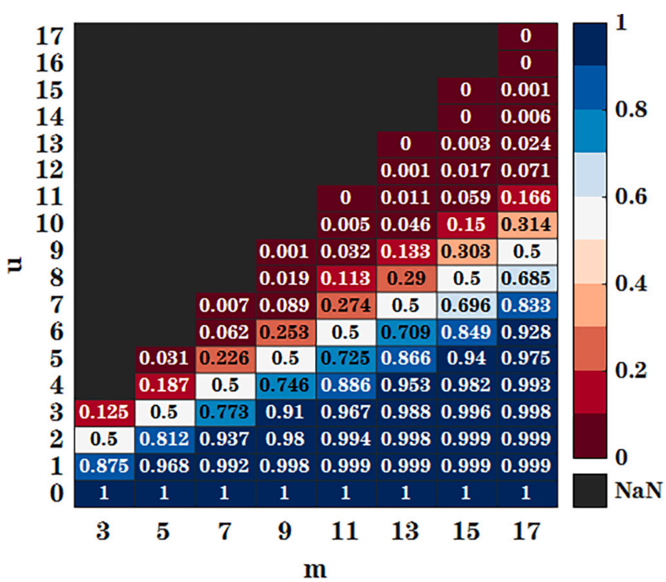

Fig. 1. The probability, $P_{Clear}(m, u)$, under different numbers of selected bins.

2.3. Multiscale detection mechanism

In order to detect the layers for an entire ASR profile based on the $P_{Clear}(m, u)$, a moving window (with an odd number of bins) strategy is utilized here. The moving windows move one bin every time, and the detection result will be marked only for the central bin. Consequently, if a layer bin (i.e., a bin belonging to a layer) is included in the moving window, the central bin is more likely to be determined and flagged as a layer bin. For the example of the moving window with a size of 3 in Fig. 2, since the $n + 2$ bin belongs to a layer, the $n + 1$ bin is more likely to be detected and marked as a layer bin. Thus, although only the $n + 2$ to $n + 5$ bins are true layer bins, all the $n + 1$ to $n + 6$ bins could initially be marked as layer bins, as shown in Fig. 2. Here, we note that the maximum bias of the layer top and base detection is half of the window size (i.e., (m-1)/2). Therefore, to remove the extra detected layer bins (such as the $n + 1$ and $n + 6$ bins) near the boundaries, we conservatively rejected (m-1)/2 bins of the layer near the layer base and top for the detection result at the current window size.

For the above moving window strategy, a large window size could produce robust detection, but removing the bins near the boundaries will sacrifice the accuracy in locating the boundaries. A small window size is beneficial to locate the layer edges accurately but is susceptible to noises. Therefore, a multiscale mechanism is utilized here to integrate the advantages of each scale (i.e., window size) to improve the detection ability. At any scale (except scales 3 and 5), the default label of each bin is clear air, and a central bin is labeled as a layer bin if the probability of the selected bins belonging to clear air is less than 0.01. At scales 3 and 5, since the probability, $P_{Clear}(m, u)$, is never less than 0.01 as shown in Fig. 1, a central bin is labeled as a layer bin if all the selected bins are larger than the ideal ASR. Finally, an integrated layer-label profile is obtained by labeling a vertical bin as a layer bin if it was labeled as a layer bin at any scale. The above detection will cause false alarms with the probability of about 0.01, most of which can be safely removed by the requirements of minimum layer thickness and minimum layer integrated attenuated backscatter (see Section 2.4).

In this study, the minimum scale used is 3, because scale 1 will detect half of the clear air bins as layer bins, which is inefficient. Theoretically, since we conservatively rejected (m-1)/2 bins of the layer near the layer base and top, a large scale will not introduce extra false alarms. However, in a multiple-layer profile, a clear air region between two layers may be improperly marked as a layer by a very large window. To avoid this issue, we set the maximum scale here to a relatively small size (i.e., 17), i.e., we choose odd numbers from 3 to 17 as the scales.

2.4. Multiscale layer detection process

Considering various strong and tenuous layers that may exist, based on the above definition of the $P_{Clear}(m, u)$ and the multiscale detection technology, a single-profile scanner can be defined as follows.

(1) Calculate the ASR based on the total attenuated backscatter coefficients and set the initial ideal ASR sequence to 1, as shown in Fig. 3(a). For convenience, we directly use CALIOP surface elevation data to remove surface echoes when performing surface detection.
(2) Calculate the $P_{Clear}(m, u)$ by using Eq. (8) at different scales through moving windows, such as those shown in Fig. 3(b) for the measured ASR in Fig. 3(a).
(3) According to the multiscale detection mechanism, at any scale (except scales 3 and 5), a bin is labeled as a layer bin if the probability is less than 0.01. At scales 3 and 5, because the probability is never less than 0.01, a central bin is labeled as a layer bin if all the selected bins are larger than the ideal ASR. The

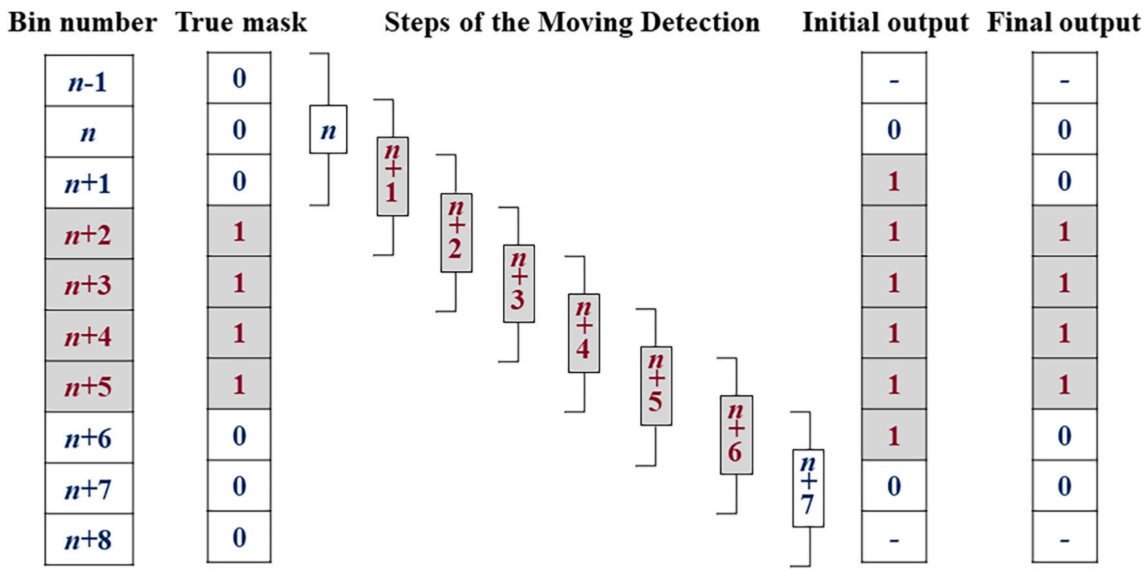

Fig. 2. An example of layer detection through a moving window, when the window size is three. For the mask, 0 and 1 mean clear air and layer, respectively.

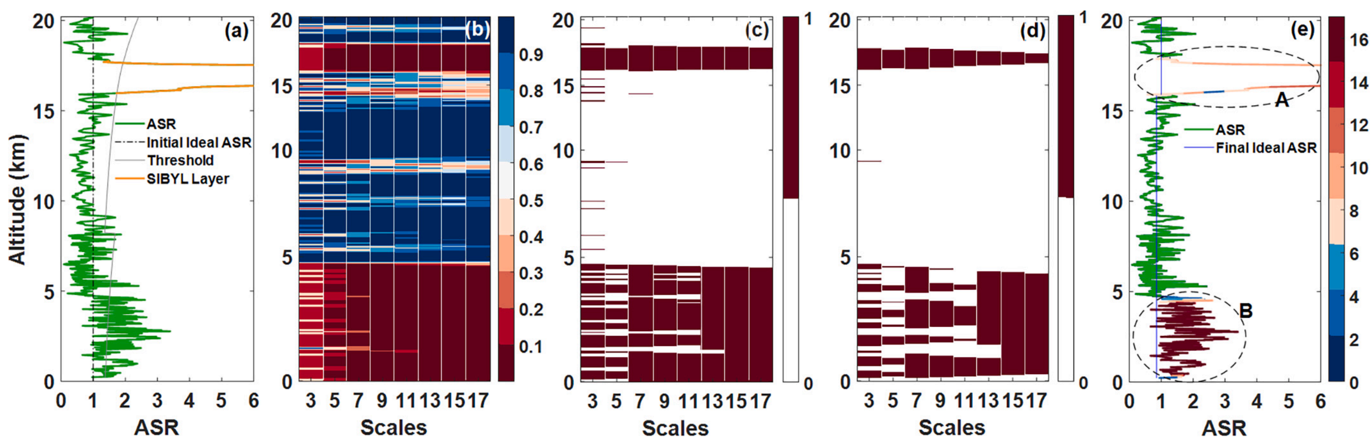

Fig. 3. An example to illustrate the process of the multiscale algorithm: (a) Measured ASR profile of 5 km resolution, threshold array, and layer detected by the SIBYL as well as the initial ideal ASR of the multiscale algorithm. (b) The probability of each bin at different scales for the signal. (c) The layer label obtained according to the probability, where 0 and 1 represent clear air and layer, respectively. (d) The layer label after rejecting half of the window size near the layer edge at each scale. (e) The layers detected by the multiscale algorithm with their maximal labeled scales marked by graduated colour, which show that layer-edge detection relies on small scales.

obtained layer label of each vertical bin at different scales is shown in Fig. 3(c). Fig. 3(d) shows the layer labels at each scale after conservatively rejecting half of the window size of the layer bins near the layer base and top. Finally, an integrated layer-label profile is obtained by labeling a vertical bin as a layer bin if it was labeled as a layer bin at any scale. In other words, once a bin has been identified as a layer at one scale, the identification will not be revoked by subsequent scans at other scales.

(4) Search from the high altitude starting bin of the current ASR profile to the low altitude bins with the integrated layer-label profile. Determine the first layer (label A in Fig. 3(e)) based on the filter conditions as used by the SIBYL, such as the minimum layer integrated attenuated backscatter (6.54×10^{-4} at night and 1.5×10^{-3} sr^{-1} at day), minimum layer thickness (540, 240, and 180 m at the lower stratosphere, upper troposphere, and lower troposphere, respectively) and minimum spike thickness (360, 120, and 90 m at the lower stratosphere, upper troposphere, and lower troposphere, respectively) (Vaughan et al., 2005).

(5) Update the ideal ASR according to Eq. (2) based on the detected layer, and repeat steps 1–5 to process the signal from the base of the last detected layer to the end. In the case of Fig. 3, a second layer was detected, as shown by label B in Fig. 3(e), which was missed by the SIBYL as shown in Fig. 3(a) because the signal intensity fluctuates randomly around the threshold, and does not present continually greater than the threshold array. Please note that the base of the last detected layer could be relocated to a lower one when repeating steps 1–5 based on the updated ideal ASR.

(6) Until the entire profile is traversed, merge the layers based on the closing gaps (0.4 km) (Vaughan et al., 2005), and terminate the layer detection of the current lidar profile.

For a scene, similar to the SIBYL (Vaughan et al., 2005), we apply the above single-profile scanner to the 5 km resolution profiles first, which are averaged by the 15 original profiles. Then, we scan the 1 and 1/3 km resolution profiles based on the 5 km resolution detection. It should be noted that we only scan 1/3 km of data below 8.2 km as the SIBYL. After clearing the boundary-layer cloud detected at 1/3 km resolution, the attenuation correction of the signal beneath those clouds is not performed, and the single-profile scanner is performed again at 5 km resolution. Then, the layers detected at 5 km resolution were removed, and the attenuation was corrected for the ASR. Finally, the profiles of 20 and 80 km resolutions were scanned to obtain tenuous layers.

3. Results and discussion

The purpose of this section is to report the findings of the application of the multiscale algorithm to the CALIPSO measurements. The CALIPSO level-2 layer products (Standard-V4.2), which were detected by the SIBYL, were used to verify the performance of the new algorithm.

3.1. Single profile performance

The SIBYL algorithm only determines those points continually greater than the threshold array, as well as meet the requirements of the minimum layer thickness and minimum integrated attenuation backscatter. Those rules of the SIBYL will lead to missed parts of layers near the edges with weak enhanced signals, as the geometrically thick layers denoted by the blue circle in Fig. 4(b1), (c1), and (d1). Besides, the missing part of layers near the edges will lead to missing the entire geometrically-thin layers, as without the missing parts, the layers will not meet the requirements of the minimum layer thickness and minimum integrated attenuation backscatter, as shown by the red circle in Fig. 4(b1), (c1), and (d1). Vaughan et al. (2009) also stated that the SIBYL was subjected to the influence of the threshold setting, which can result in the non-detection of optically tenuous layers. However, the multiscale algorithm does not rely on a threshold setting but is based on the likelihood of the measured ASR to the ideal ASR. Therefore, the multiscale algorithm can detect the layers more completely, which could avoid the geometrically thin layers to be filtered out, as shown in Fig. 4 (b2), (c2), and (d2). The above comparisons were performed at 5 km resolution, but most of the missing layers could be detected by the SIBYL at coarse resolutions after averaging multiple profiles, as shown in Section 3.2.

3.2. Scene performance

For scene detection, the horizontal averaging strategy of the multiscale algorithm is the same as the SIBYL, as we described in Section 2.4. Given that the purpose of the detection at 1/3 km resolution was to remove the boundary-layer clouds with strong scattering by an extremely large threshold, rather than to fully detect the layers. Thus, we used the same method as that of the SIBYL to clear the boundary-layer clouds, and no comparison of 1/3 km resolution profiles was performed afterward.

The SIBYL can successfully detect strong layers of a scene. However, only parts of the layers were found, as shown in Fig. 5(b), which is the detection results of the scene in Fig. 5(a) at 5 km resolution. Meanwhile, the multiscale algorithm acquired additional layers, and the layers were detected more completely than by the SIBYL (Fig. 5(d) and (f)). The detection results (i.e., Vertical Feature Mask, VFM) of the SIBYL at 1–80 km resolutions (Fig. 5(c)) show that the layer (top and base) detected by the SIBYL at 1 km resolution was notably less (inaccurate) than that at 5 km resolution. The optically tenuous layers were essentially not detected. However, the layers acquired by the multiscale algorithm at 1 km resolution were essentially comparable with the layers acquired at 5 km resolution, as shown in Fig. 5(e). That means the multiscale algorithm can obtain more accurate layers than the SIBYL at high resolutions. Furthermore, the SIBYL did not detect layers completely at 1 and 5 km resolutions, which led to overestimations of the layer coverage in the subsequent 20 and 80 km resolution profiles (blue ellipse in Fig. 5(g)).

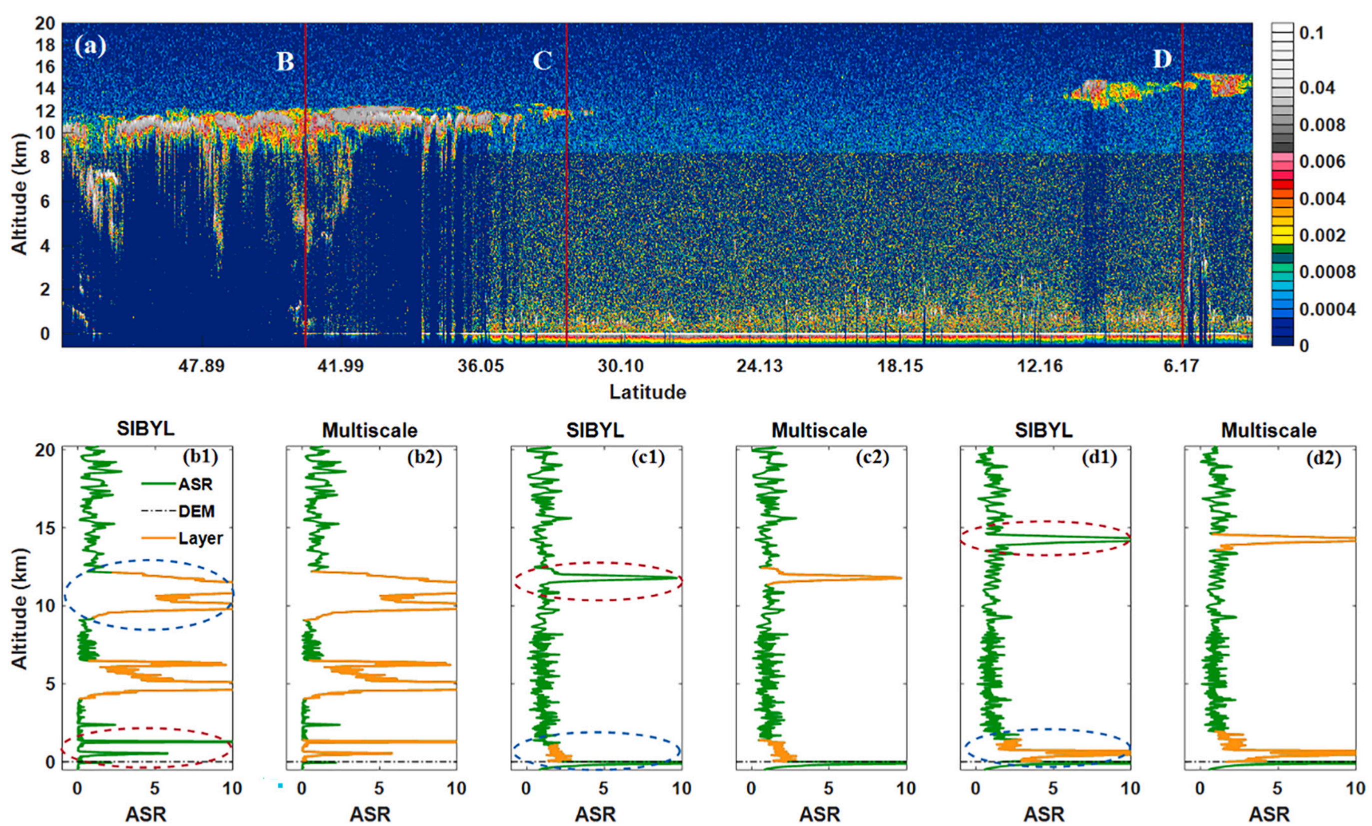

Fig. 4. (a) Total attenuated backscatter coefficients of a CALIPSO scene. (b1, b2), (c1, c2), (d1, d2) represent the detection results of the SIBYL and multiscale algorithms at 5 km resolution for the profiles of B, C, and D marked by the red lines in (a), respectively. We use the digital elevation model (DEM) to represent the height of the surface. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

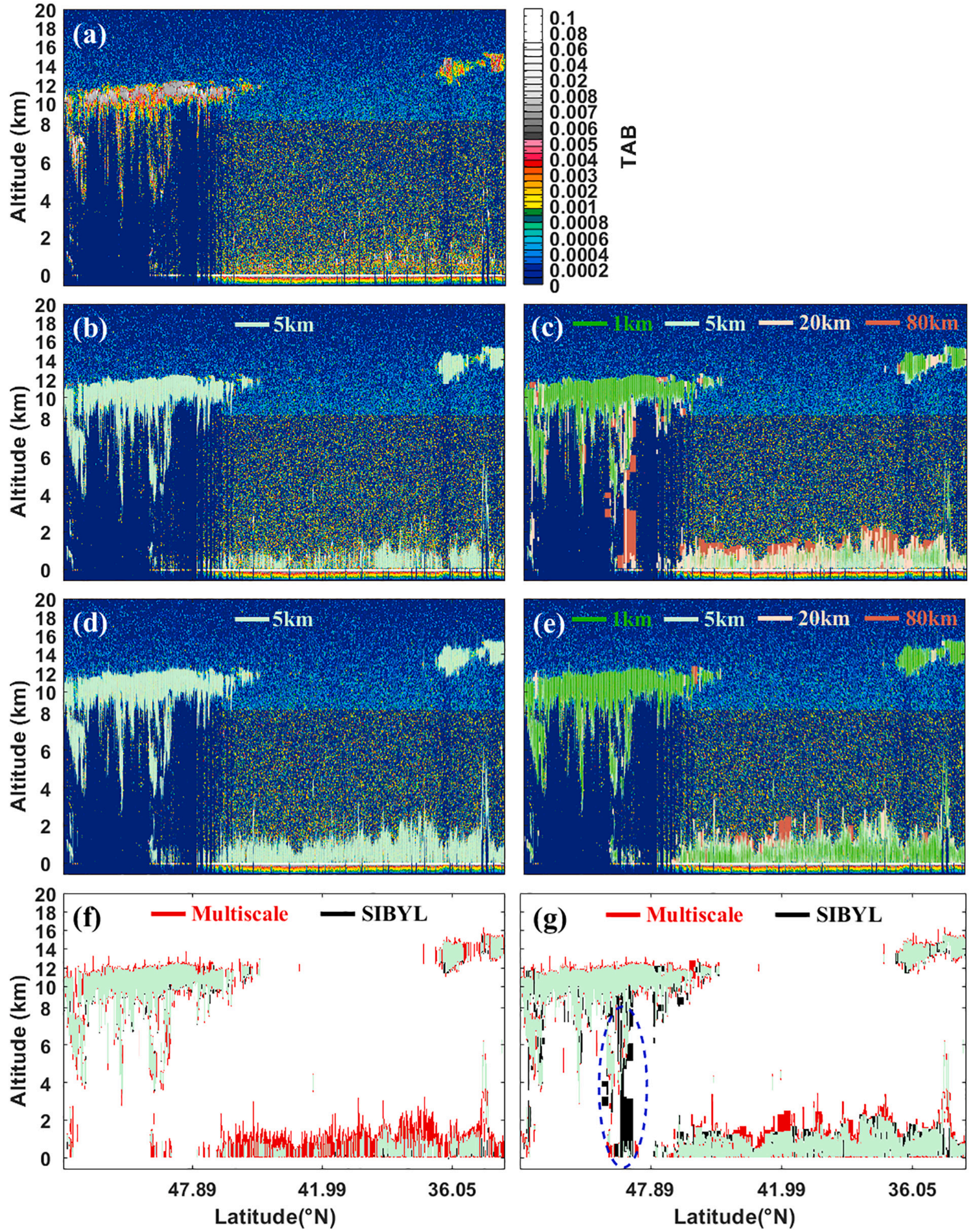

Fig. 5. Layer detection results of the two algorithms in a CALIPSO scene. (a) Total attenuated backscatter coefficients of the scene. (b) and (c) are the detection results (i.e., vertical feature mask, VFM) for the SIBYL at 5 km and 1–80 km resolution, respectively. (d) and (e) are the same as (b) and (c) but for the multiscale algorithm. (f) and (g) represent the VFM difference of the two algorithms at 5 km and 1–80 km resolution, respectively. The black (red) area indicates that the SIBYL (the multiscale algorithm) identifies a layer that undetected by the multiscale algorithm (the SIBYL), and the green area indicates the layers that are detected by both of the two algorithms. The blue ellipse in (g) marks the regions where the detection results of the two algorithms are quite different. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3.3. Regional statistical performance

To compare the performance of the two algorithms, we selected data from two different regions in December 2017 for testing (Fig. 6). Region A (3°–54° N, 73°–136° E) contains 526 CALIPSO scenes (Daytime: 263 and Nighttime: 263; more than 2.9 million profiles at 1 km resolution), which is vast in territory and complex in terrain, and the pollution propagation is complicated in winter. Region B (10°–60° S, 55°–110° E) is over the southern latitude ocean with a relatively clear atmosphere, which contains 614 CALIPSO scenes (Daytime: 306 and Nighttime: 308; more than 3.4 million profiles at 1 km resolution).

In the two regions (A and B), the total layer area detected by the multiscale algorithm was 11.93% more than that of the SIBYL (Fig. 7(a)). The total layer area detected by the multiscale algorithm is 1.31 and 1.52 (0.60 and 0.84) times as that detected by the SIBYL at 1 and 5 km (20 and 80 km) resolutions, respectively (Fig. 7(a)). By separating the detected layers into daytime and nighttime, results show that the multiscale algorithm detected 37.41% and 16.36% more layer area than the SIBYL at 1–80 km resolutions at daytime (Fig. 7(b)) and 1–5 km resolutions at nighttime (Fig. 7(c)), respectively. This is because the SNR is very low at 1–80 km resolutions at daytime and 1–5 km resolutions at nighttime, and the SIBYL requires very large thresholds to suppress the noise effects, obstructing the detection of optically tenuous layers. However, the multiscale algorithm does not need a threshold array, allowing more tenuous layers to be detected successfully when the SNR is low.

Furthermore, though the multiscale algorithm is expected to detect more layers than the SIBYL at nighttime, similar to the case at daytime, the result is the opposite (Fig. 7(c)). The result shows that the multiscale algorithm detected 4.40% less layer area than the SIBYL at nighttime, mainly caused by the large proportion of layer area detected by the SIBYL at 20 and 80 km resolutions. This may be due to the overestimation of the SIBYL caused by averaging layer and clear air profiles together at a low resolution (Balmes and Fu, 2018; Cesana et al., 2016; Hagihara et al., 2010a), as the blocky, rectangular layers shown in Fig. 5(c). Additionally, this could also be caused by the overestimations of the SIBYL in an underlying layer detection due to inaccurate ASR_{ideal} update using a fixed lidar ratio to the upper layers (Mao et al., 2018). Based on the results, we can infer that the multiscale algorithm could effectively reduce the underestimation of the official CALIPSO AOD due to the missing layers at daytime (Kacenelenbogen et al., 2014; Kim et al., 2013; Redemann et al., 2011; Thorsen et al., 2017), but the official CALIPSO AOD at nighttime should have much less bias than at daytime.

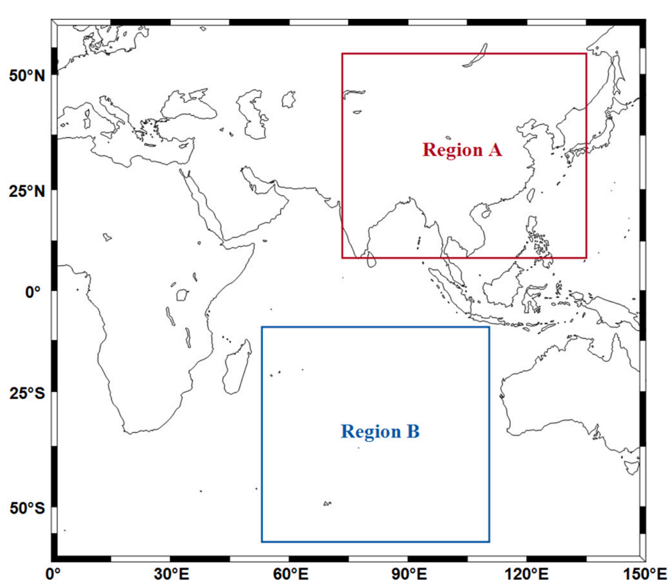

Fig. 6. Study regions A (3°–54° N, 73°–136° E) and B (10°–60° S, 55°–110° E).

The number of layers obtained by the multiscale algorithm in regions A (Fig. 8(a)–(c)) and B (Fig. 8(d)–(f)) at 1 and 5 km resolutions showed a similar distribution trend and performance at daytime and nighttime. Generally, the multiscale algorithm detected more tenuous layers near the surface than at high altitudes, because aerosols were dominant near the surface, which was more tenuous and easier to be missed than the clouds at high altitudes.

3.4. Verification of detection reliability and efficiency

The layer depolarization ratio (DR) is an important property for layer classification (Liu et al., 2019b; Omar et al., 2009). Ideally, the DR of clear air is about 0.0035, but the DR of ice clouds is generally greater than 0.05 (Liu et al., 2005; Vernier et al., 2011). Therefore, we can utilize the integrated volume DR (IVDR) characteristic of the extra detected ice clouds by the multiscale algorithm to evaluate the reliability of the extra detected layers. We considered that the extra detected layer bins adjacent to the ice clouds of the official product belong to ice cloud bins. We extracted the IVDR of those extra detected ice cloud bins and a neighboring clear air bin for each of them. Further, we calculated the mean IVDR of the extracted ice cloud and clear air bins at every 80 km horizontal interval for all the studied scenes, respectively. We calculated the mean IVDR to reduce noise interference as did in CALIPSO layer classification (Liu et al., 2019b; Omar et al., 2009). The results show that the mean IVDR of clear air does fluctuate around 0.0035, and 86.1% and 81.3% of the mean IVDR of extra detected ice clouds is greater than 0.05 in regions A and B, respectively (Fig. 9). These results indicate that the extra detected layer bins by the multiscale algorithm are reliable because they are not clear air bins but layer bins with a very high probability.

To compare and verify the detection capability of the multiscale algorithm and the threshold method, we conducted a large number of simulation tests. Since the purpose of the simulation test here is to ideally explore the efficiency difference between the detection capabilities of the two algorithms under different SNRs. Thus, we simply simulated the Normal distribution with different SNRs but did not simulate the signal based on the iterative Poisson distribution (i.e., a Neyman Type-A) used by the official CALIPSO simulator (Liu et al., 2006; Powell et al., 2006). We first simulated a clear air signal and then added simulated layer signals with different signal-to-noise ratios (n) at a certain height. The simulated layer signal follows a normal distribution as $Signal \sim N(1 + n\sigma)$, i.e., with a mean value of $1 + n\sigma$ and a standard deviation of σ. We set n from 0 to 5 with the step size as 0.1 and σ set to 1. For each n, we simulate 10,000 signals and then use the two algorithms to perform layer detection.

The result shows that the true detection rate (i.e., the ratio of the number of detected true layer bins to the total number of layer bins) of the multiscale algorithm is significantly higher than that of the SIBYL when the SNR is low (Fig. 10). The multiscale and SIBYL algorithms achieve a 100% true detection rate when n is about 2 and 4, respectively. The false detection rate is much less than 1%, which is defined as the probability of determining a clear air bin as a layer bin. The simulation tests further prove that the multiscale algorithm could detect layers more efficiently than the SIBYL under the same conditions. Additionally, please note that, for a real CALIPSO observation, the false detection rate is considerable, which could be caused by the uncertainty of instrument defects, calibration, averaging scheme, meteorological factors, background aerosols, etc. Those complex factors are not considered in this simulation yet but are very interesting to be discussed in the future.

4. Conclusions

A multiscale layer detection algorithm without threshold setting is proposed for CALIPSO measurements. The new algorithm compares the measured ASR with the ideal ASR, determines layers based on a pre-

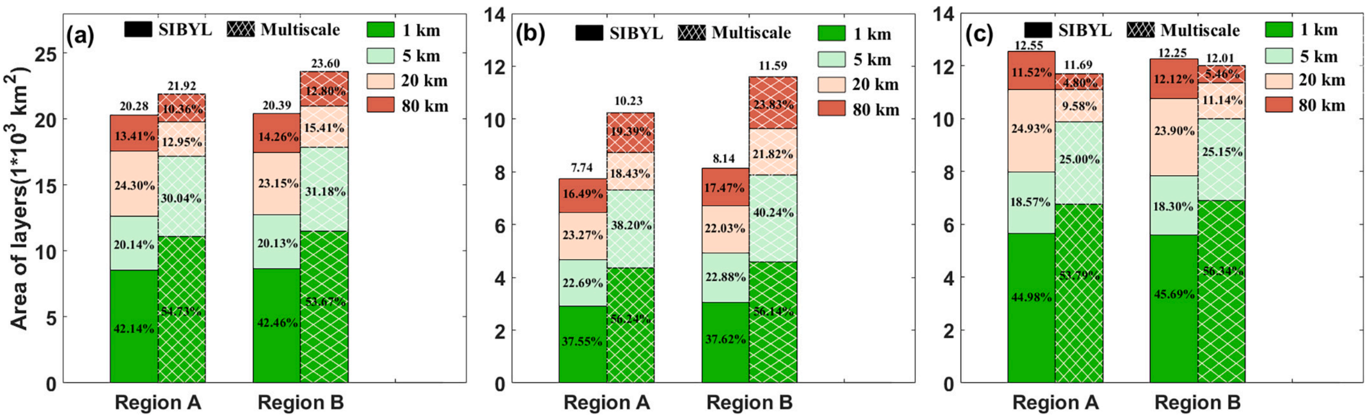

Fig. 7. Statistical results of the two algorithms applied to CALIPSO measurements in Dec 2017 over regions A and B. (a) the total area of layers detected by the SIBYL and multiscale algorithms at daytime and nighttime in regions A and B at different resolutions. (b) and (c) are the same as (a) but for daytime and nighttime, respectively. In each sub-figure, the SIBYL detected area is considered as 100%, the percentage of the detected area by the multiscale algorithm is relative to the SIBYL detected area.

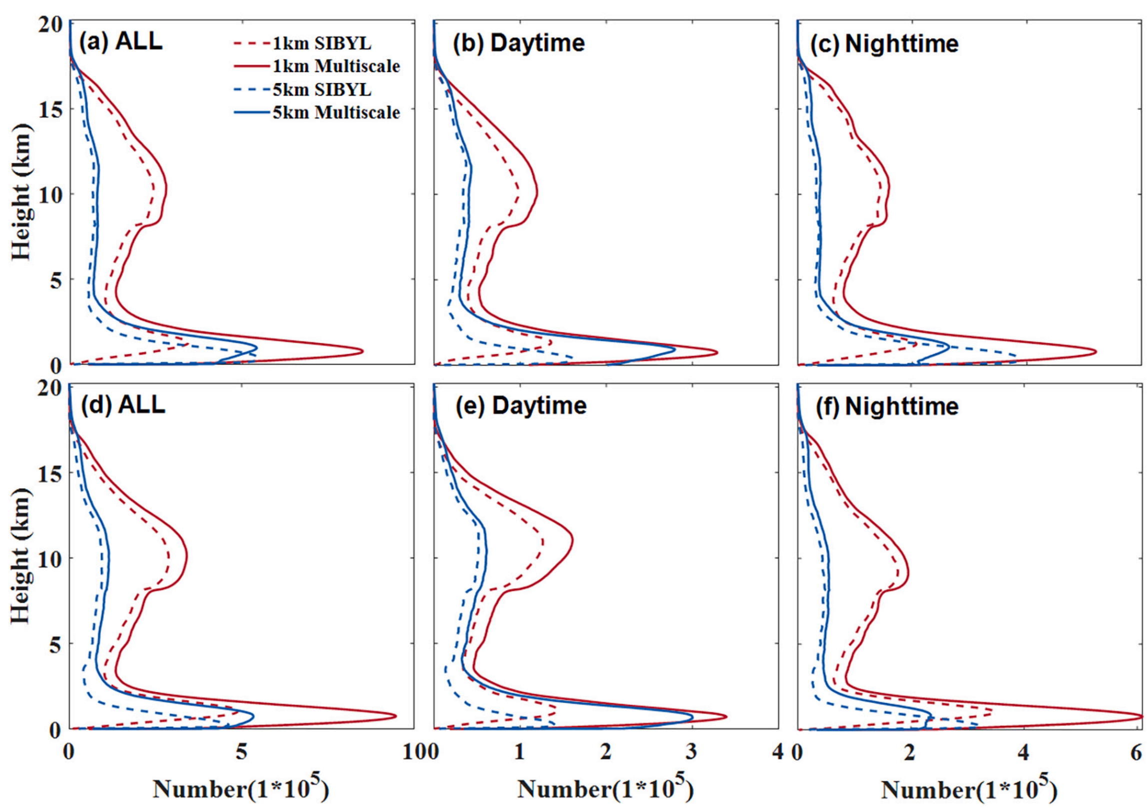

Fig. 8. Distribution of layers detected by the multiscale algorithm and SIBYL in region A and B. (a) the number of layer bins detected at 1 and 5 km resolutions at all time in region A. (b) and (c) are the same as (a) but for daytime and nighttime, respectively. (d)-(f) are the same as (a)-(c) but for region B. Note that the detection number at 5 km resolution is multiplied by 5 and deducted the detection number at 1 km resolution.

defined probability, and then combines the information on multiple scales to reduce the effects of noise.

(1) The profile and scene cases detection results show that the multiscale algorithm could detect optically tenuous layers and layer edges effectively, and avoid the overestimation of the SIBYL caused by averaging layer and clear air profiles together at a low resolution.
(2) Overall, the multiscale algorithm detected 11.93% more layer area than the SIBYL over the experimental regions. The multiscale algorithm detected 37.41% and 16.36% more layer area than the SIBYL at 1–80 km resolutions at daytime and 1–5 km resolutions at night time, respectively. This indicates that the multiscale algorithm could detect much more tenuous layers than the SIBYL when the SNR is low because a threshold array used by the SIBYL could significantly obstruct tenuous layers from being detected.
(3) The SIBYL detected 4.40% more layer area than the multiscale algorithm at nighttime, which is mainly contributed by the large proportion of layer area detected by the SIBYL at 20 and 80 km resolutions. This implies that there may be noteworthy overestimation by the SIBYL at low resolution.

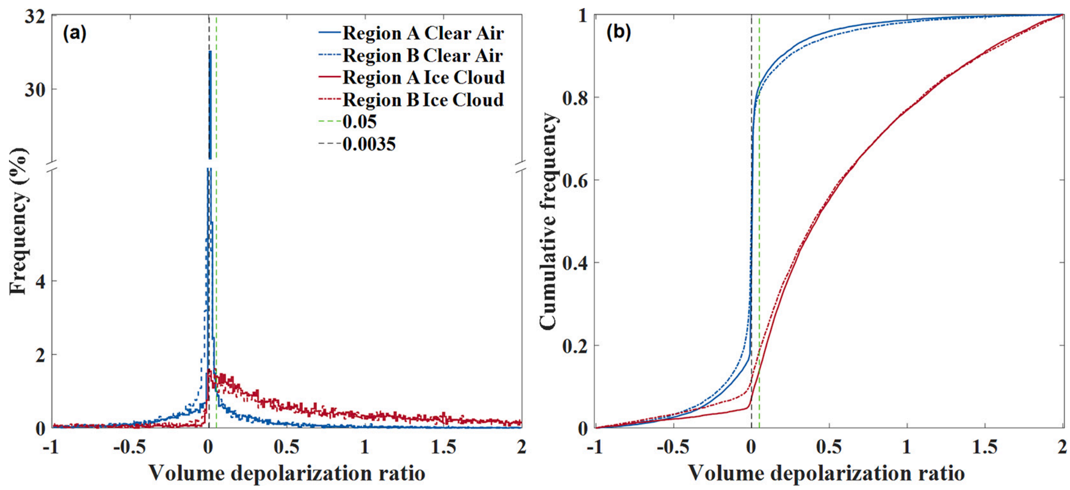

Fig. 9. Reliability evaluation of the extra detected layers based on depolarization ratio. (a) and (b) are the frequency and cumulative frequency of the mean layer integrated volume depolarization ratio (with 0.01 as interval) of clear air and the extra detected ice clouds by the multiscale algorithm in regions A and B, respectively.

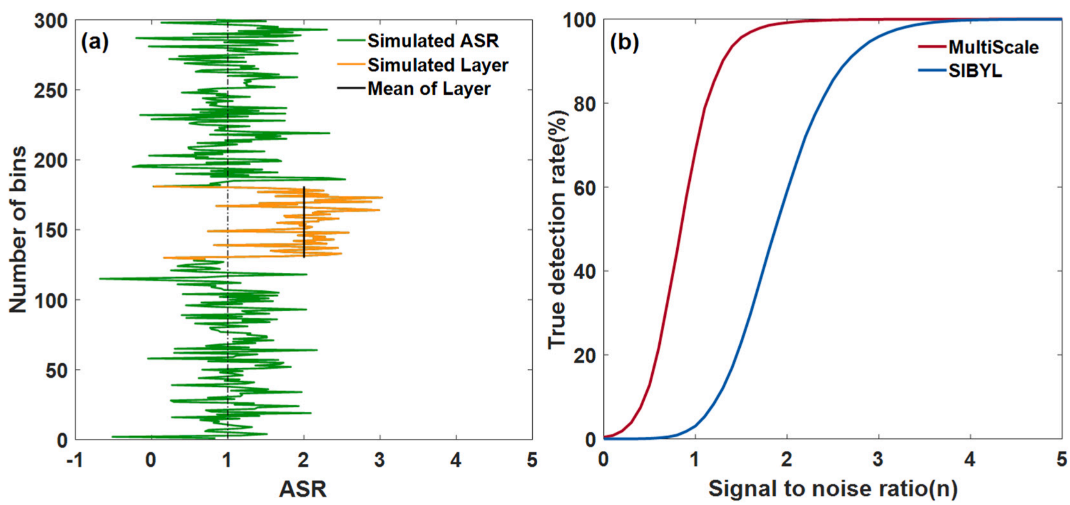

Fig. 10. Simulation tests for the detection rate of multiscale and SIBYL algorithms. (a) A simulated signal, (b) the true detection rate of the two algorithms with different signal-to-noise ratios (n). Note that the false detection rate is equal to the true detection rate at $n = 0$.

(4) The evaluation using the depolarization ratio shows that the found missing layers by the multiscale algorithm are reliable. Besides, the simulation tests show that the true detection rate of the multiscale algorithm is significantly higher than that of the SIBYL under different SNRs.
(5) We infer that the multiscale algorithm could effectively reduce the underestimation of the official CALIPSO AOD due to the missing layers at daytime, but the official CALIPSO AOD at nighttime should have much less bias than at daytime.

In this study, we used a multiscale mechanism instead of the threshold array mechanism, but the other part of the scheme was similar to the SIBYL. In the future, a more comprehensive multiscale algorithm than the one proposed in this paper may be considered to reduce several one-size-fits-all determinations of the SIBYL. Additionally, more useful information could be provided after performing the classification and extinction retrieval in the future.

Declaration of Competing Interest

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

Acknowledgments

This work was supported by the National Natural Science Foundation of China (grant nos. 41971285, 41627804, and 41701381) and the National Key Research and Development Program of China (grant no. 2016YFC0200900). The CALIPSO dataset was obtained from the NASA Langley Research Centre and Atmospheric Sciences Data Centre (eosweb.larc.nasa.gov). We would like to deeply thank one of the reviewers named Mark Vaughan from NASA. He thinks this study proposed a very interesting and cool method that may become an important contribution to the community, and strongly supports the publication of this paper.

References

Balmes, K.A., Fu, Q., 2018. An investigation of optically very thin ice clouds from ground-based ARM Raman Lidars. Atmosphere 9, 445.
Basu, D., Pereira, C.A.D.B., 1990. Blackwell sufficiency and bernoulli experiments. Braz. J. Probab. Stat. 4, 137–145.
Berry, E., Mace, G.G., 2014. Cloud properties and radiative effects of the Asian summer monsoon derived from A-train data. J. Geophys. Res.-Atmos. 119, 9492–9508.

Bi, J., Huang, J., Hu, Z., Holben, B.N., Guo, Z., 2014. Investigating the aerosol optical and radiative characteristics of heavy haze episodes in Beijing during January of 2013. J. Geophys. Res.-Atmos. 119.

Bian, Y., Xu, W., Hu, Y., Tao, J., Kuang, Y., Zhao, C., 2020. Method to retrieve aerosol extinction profiles and aerosol scattering phase functions with a modified CCD laser atmospheric detection system. Opt. Express 28, 6631–6647.

Cesana, G., Chepfer, H., Winker, D., Getzewich, B., Cai, X., Jourdan, O., Mioche, G., Okamoto, H., Hagihara, Y., Noel, V., 2016. Using in-situ airborne measurements to evaluate three cloud phase products derived from CALIPSO. J. Geophys. Res.-Atmos. 121, 5788–5808.

Chen, T., Li, Z., Kahn, R.A., Zhao, C., Rosenfeld, D., Guo, J., Han, W., Chen, D., 2021. Potential impact of aerosols on convective clouds revealed by Himawari-8 observations over different terrain types in eastern China. Atmos. Chem. Phys. 21, 6199–6220.

Clothiaux, E.E., Mace, G., Ackerman, T., Kane, T.J., Spinhirne, J., Scott, V., 1998. An automated algorithm for detection of hydrometeor returns in micropulse lidar data. J. Atmos. Ocean. Technol. 15, 1035–1042.

Comerón, A., Sicard, M., Rocadenbosch, F., 2013. Wavelet correlation transform method and gradient method to determine aerosol layering from lidar returns: some comments. J. Atmos. Ocean. Technol. 30, 1189–1193.

Davis, S., Hlavka, D., Jensen, E., Rosenlof, K., Yang, Q., Schmidt, S., Borrmann, S., Frey, W., Lawson, P., Voemel, H., Bui, T.P., 2010. In situ and lidar observations of tropopause subvisible cirrus clouds during TC4. J. Geophys. Res.-Atmos. 115.

Feofilov, A.G., Stubenrauch, C.J., Delanoë, J., 2015. Ice water content vertical profiles of high-level clouds: classification and impact on radiative fluxes. Atmos. Chem. Phys. 15, 12327–12344.

Fu, Y., Chen, Y., Li, R., Qin, F., Xian, T., Yu, L., Zhang, A., Liu, G., Zhang, X., 2017. Lateral boundary of cirrus cloud from CALIPSO observations. Sci. Rep. 7, 14221.

Guo, J., Liu, H., Wang, F., Huang, J., Xia, F., Lou, M., Wu, Y., Jiang, J.H., Xie, T., Zhaxi, Y.J.A.R., 2016. Three-dimensional structure of aerosol in China: a perspective from multi-satellite observations, 178, 580–589.

Hagihara, Y., Okamoto, H., Yoshida, R., 2010a. Development of a combined CloudSat-CALIPSO cloud mask to show global cloud distribution. J. Geophys. Res.-Atmos. 115. –.

Hagihara, Y., Okamoto, H., Yoshida, R., 2010b. Development of a combined CloudSat-CALIPSO cloud mask to show global cloud distribution. J. Geophys. Res.-Atmos. 115.

Kacenelenbogen, M., Redemann, J., Vaughan, M.A., Omar, A., Russell, P.B., Burton, S.P., Rogers, R.R., Ferrare, R.A., Hostetler, C.A., 2014. An evaluation of CALIOP/CALIPSO's aerosol-above-cloud detection and retrieval capability over North America. J. Geophys. Res. 119, 230–244.

Kim, M., Kim, S.W., Yoon, S., Omar, A., 2013. Comparison of aerosol optical depth between CALIOP and MODIS-aqua for CALIOP aerosol subtypes over the ocean. J. Geophys. Res. 118.

Li, Z., Guo, J., Ding, A., Liao, H., Liu, J., Sun, Y., Wang, T., Xue, H., Zhang, H., Zhu, B., 2017. Aerosol and boundary-layer interactions and impact on air quality. Natl. Sci. Rev. 4, 810–833.

Li, Z., Wang, Y., Guo, J., Zhao, C., Cribb, M.C., Dong, X., Fan, J., Gong, D., Huang, J., Jiang, M., Jiang, Y., Lee, S.-S., Li, H., Li, J., Liu, J., Qian, Y., Rosenfeld, D., Shan, S., Sun, Y., Wang, H., Xin, J., Yan, X., Yang, X., Yang, X.-Q., Zhang, F., Zheng, Y., 2019. East Asian study of tropospheric aerosols and their impact on regional clouds, precipitation, and climate (EAST-AIRCPC). J. Geophys. Res.-Atmos. 124, 13026–13054.

Liu, Z., Omar, A., Hu, Y., Vaughan, M., Winker, D., Poole, L., Kovacs, T., 2005. CALIOP algorithm theoretical basis document. Part 3: scene classification algorithms. In: NASA-CNES document PC-SCI-203.

Liu, Z., Hunt, W., Vaughan, M., Hostetler, C., Mcgill, M., Powell, K., Winker, D., Hu, Y., 2006. Estimating random errors due to shot noise in backscatter lidar observations. Appl. Opt. 45, 4437–4447.

Liu, D., Zheng, Z., Chen, W., Wang, Z., Li, W., Ke, J., Zhang, Y., Chen, S., Cheng, C., Wang, S., 2019a. Performance estimation of space-borne high-spectral-resolution lidar for cloud and aerosol optical properties at 532 nm. Opt. Express 27, A481–A494.

Liu, Z., Kar, J., Zeng, S., Tackett, J., Vaughan, M., Avery, M., Pelon, J., Getzewich, B., Lee, K.-P., Magill, B., 2019b. Discriminating between Clouds and Aerosols in the CALIOP Version 4.1 Data Products.

Ma, P.L., Rasch, P.J., Chepfer, H., Winker, D.M., Ghan, S.J., 2018. Observational constraint on cloud susceptibility weakened by aerosol retrieval limitations. Nat. Commun. 9, 2640.

Mao, F., Gong, W., Zhu, Z., 2011. Simple multiscale algorithm for layer detection with lidar. Appl. Opt. 50, 6591–6598.

Mao, F., Pan, Z., Wang, W., Li, S., Wei, G., 2018. Iterative method for determining boundaries and lidar ratio of permeable layer of a space lidar. J. Quant. Spectrosc. Radiat. Transf. 218, 125–130.

Mao, F., Zhao, M., Gong, W., Chen, L., Liang, Z., 2021. Layer detection algorithm for CALIPSO observation based on automatic segmentation with a minimum cost function. J. Quant. Spectrosc. Radiat. Transf. 261, 107498.

Melfi, S., Spinhirne, J., Chou, S.-H., Palm, S., 1985. Lidar observations of vertically organized convection in the planetary boundary layer over the ocean. J. Clim. Appl. Meteorol. 24, 806–821.

Omar, A.H., Winker, D.M., Vaughan, M.A., Hu, Y., Trepte, C.R., Ferrare, R.A., Lee, K.-P., Hostetler, C.A., Kittaka, C., Rogers, R.R., 2009. The CALIPSO automated aerosol classification and lidar ratio selection algorithm. J. Atmos. Ocean. Technol. 26, 1994–2014.

Pal, S.R., Steinbrecht, W., Carswell, A.I., 1992. Automated method for lidar determination of cloud-base height and vertical extent. Appl. Opt. 31, 1488–1494.

Powell, K.A., Liu, Z., Hunt, B., 2006. Simulation of Random Electron Multiplication in Calipso Lidar Photomultipliers.

Ramanathan, V., Crutzen, P., Kiehl, J., Rosenfeld, D., 2001. Aerosols, climate, and the hydrological cycle. Science 294, 2119–2124.

Redemann, J., Vaughan, M.A., Zhang, Q., Shinozuka, Y., Russell, P.B., Livingston, J.M., Kacenelenbogen, M., Remer, L.A., 2011. The comparison of MODIS-aqua (C5) and CALIOP (V2 & V3) aerosol optical depth. Atmos. Chem. Phys. 12, 3025–3043.

Thorsen, T.J., Ferrare, R.A., Hostetler, C.A., Vaughan, M.A., Fu, Q., 2017. The impact of lidar detection sensitivity on assessing aerosol direct radiative effects. Geophys. Res. Lett. 44, 9059–9067.

Vaillant de Guélis, T., Vaughan, M.A., Winker, D.M., Liu, Z., 2021. Two-dimensional and multi-channel feature detection algorithm for the CALIPSO lidar measurements. Atmos. Meas. Tech. 14, 1593–1613.

Vaughan, M.A., Winker, D.M., Powell, K.A., 2005. CALIOP algorithm theoretical basis document, part 2: feature detection and layer properties algorithms. Rep. PC-SCI 202, 87.

Vaughan, M.A., Powell, K.A., Winker, D.M., Hostetler, C.A., Kuehn, R.E., Hunt, W.H., Getzewich, B.J., Young, S.A., Liu, Z., McGill, M.J., 2009. Fully automated detection of cloud and aerosol layers in the CALIPSO lidar measurements. J. Atmos. Ocean. Technol. 26, 2034–2050.

Vernier, J.P., Thomason, L.W., Kar, J., 2011. CALIPSO detection of an Asian tropopause aerosol layer. Geophys. Res. Lett. 38 (7).

Wang, Y., Zhao, C., 2017. Can MODIS cloud fraction fully represent the diurnal and seasonal variations at DOE ARM SGP and Manus sites? J. Geophys. Res.-Atmos. 122, 329–343.

Wang, N., Shen, X., Xiao, D., Veselovskii, I., Zhao, C., Chen, F., Liu, C., Rong, Y., Ke, J., Wang, B., Liu, D., 2021. Development of ZJU high-spectral-resolution lidar for aerosol and cloud: feature detection and classification. J. Quant. Spectrosc. Radiat. Transf. 261, 107513.

Winker, D., Vaughan, M., 1994. Vertical distribution of clouds over Hampton, Virginia observed by lidar under the ECLIPS and FIRE ETO programs. Atmos. Res. 34, 117–133.

Winker, D.M., Pelon, J.R., McCormick, M.P., 2003. CALIPSO mission: spaceborne lidar for observation of aerosols and clouds. In: Lidar Remote Sensing for Industry and Environment Monitoring, III. International Society for Optics and Photonics, pp. 1–11.

Winker, D.M., Hunt, W.H., McGill, M.J., 2007. Initial performance assessment of CALIOP. Geophys. Res. Lett. 34.

Winker, D.M., Vaughan, M.A., Omar, A., Hu, Y., Powell, K.A., Liu, Z., Hunt, W.H., Young, S.A., 2009. Overview of the CALIPSO mission and CALIOP data processing algorithms. J. Atmos. Ocean. Technol. 26, 2310–2323.

Winker, D.M., Pelon, J., Coakley, J.A., Ackerman, S.A., Charlson, R.J., Colarco, P.R., Flamant, P., Fu, Q., Hoff, R.M., Kittaka, C., Kubar, T.L., Le Treut, H., Mccormick, M. P., Mégie, G., Poole, L., Powell, K., Trepte, C., Vaughan, M.A., Wielicki, B.A., 2010. The CALIPSO mission: A global 3D view of aerosols and clouds. Bull. Am. Meteorol. Soc. 91, 1211–1230.

Xie, S., Liu, X., Zhao, C., Zhang, Y., 2013. Sensitivity of CAM5-simulated arctic clouds and radiation to ice nucleation parameterization. J. Clim. 26, 5981–5999.

Xie, H., Zhou, T., Fu, Q., Huang, J., Huang, Z., Bi, J., Shi, J., Zhang, B., Ge, J.J.O.E., 2017. Automated Detection of Cloud and Aerosol Features with SACOL Micro-Pulse Lidar in Northwest China, p. 25.

Yorks, J.E., Hlavka, D.L., Vaughan, M.A., McGill, M.J., Hart, W.D., Rodier, S., Kuehn, R., 2011. Airborne validation of cirrus cloud properties derived from CALIPSO lidar measurements: spatial properties. J. Geophys. Res.: Atmos. 116.

Zhang, Y., Zhao, C., Zhang, K., Ke, J., Che, H., Shen, X., Zheng, Z., Liu, D., 2019. Retrieving the microphysical properties of opaque liquid water clouds from CALIOP measurements. Opt. Express 27, 34126–34140.

Zhao, C., Garrett, T.J., 2015. Effects of Arctic haze on surface cloud radiative forcing. Geophys. Res. Lett. 42, 557–564.

Zhao, C., Wang, Y., Wang, Q., Li, Z., Wang, Z., Liu, D., 2014. A new cloud and aerosol layer detection method based on micropulse lidar measurements. J. Geophys. Res.-Atmos. 119, 6788–6802.