



Article Estimating Boundary Layer Height from LiDAR Data under Complex Atmospheric Conditions Using Machine Learning

Zhenxing Liu ^{1,2,3}, Jianhua Chang ^{1,2,*}, Hongxu Li ⁴, Sicheng Chen ¹ and Tengfei Dai ¹

- ¹ Jiangsu Key Laboratory of Meteorological Observation and Information Processing, Nanjing University of Information Science and Technology, Nanjing 210044, China; 20171100005@nuist.edu.cn (Z.L.); 20191219004@nuist.edu.cn (S.C.); 20211118003@nuist.edu.cn (T.D.)
- ² Collaborative Innovation Center of Atmospheric Environment and Equipment Technology, Nanjing University of Information Science and Technology, Nanjing 210044, China
- ³ Department of Information Technology, Taizhou Polytechnic College, Taizhou 225300, China
- ⁴ School of Electronic Information Engineering, Wuxi University, Wuxi 214000, China;
 - 20181132088@nuist.edu.cn
- * Correspondence: jianhuachang@nuist.edu.cn

Abstract: Reliable estimation of the atmospheric boundary layer height (ABLH) is critical for a range of meteorological applications, including air quality assessment and weather forecasting. Several algorithms have been proposed to detect ABLH from aerosol LiDAR backscatter data. However, most of these focus on cloud-free conditions or use other ancillary instruments due to strong interference from clouds or residual layer aerosols. In this paper, a machine learning method named the Mahalanobis transform K-near-means (MKnm) algorithm is first proposed to derive ABLH under complex atmospheric conditions using only LiDAR-based instruments. It was applied to the micro pulse LiDAR data obtained at the Southern Great Plains site of the Atmospheric Radiation Measurement (ARM) program. The diurnal cycles of ABLH from cloudy weather were detected by using the gradient method (GM), wavelet covariance transform method (WM), K-means, and MKnm. Meanwhile, the ABLH obtained by these four methods under cloud or residual layer conditions based on micropulse LiDAR data were compared with the reference height retrieved from radiosonde data. The results show that MKnm was good at tracking the diurnal variation of ABLH, and the ABLHs obtained by it have remarkable correlation coefficients and smaller mean absolute error and mean deviation with the radiosonde-derived ABLHs than those measured by other three methods. We conclude that MKnm is a promising algorithm to estimate ABLH under cloud or residual layer conditions.

Keywords: LiDAR; atmosphere boundary layer height; cloud; residual layer; machine learning

1. Introduction

The atmospheric boundary layer (ABL) is the lowest part of the troposphere near the Earth's surface. Moreover, it is the layer within which energy, water, momentum, and matter in the atmosphere are mixed and exchanged with the surface [1–3]. Pollutants from the ground are dispersed and trapped within this layer, and fog also occurs in the ABL. It evolves throughout the day and is season influenced by solar radiation and other factors. Therefore, the atmospheric boundary layer height (ABLH) is a crucial parameter in numerical weather prediction models, climate change studies and air quality assessments [4–7].

Many instruments have been implemented to detect ABLH, such as tethered balloons [8], microwave radiometers [9], radiosonde (RS) [10–13], LiDAR [14–16], and ceilometer [17]. RS can estimate ABLH with high precision from the vertical profiles of measured temperature and humidity. Thus, it is the most reliable instrument, and its estimate is often used as a reference value to compare with the ABLH obtained by other instruments. However, the radiosonde test is labor intensive and the limited launch frequency cannot



Citation: Liu, Z.; Chang, J.; Li, H.; Chen, S.; Dai, T. Estimating Boundary Layer Height from LiDAR Data under Complex Atmospheric Conditions Using Machine Learning. *Remote Sens.* 2022, *14*, 418. https:// doi.org/10.3390/rs14020418

Academic Editors: Patricia K. Quinn, Maria João Costa, Oleg Dubovik and Jean-Christophe Raut

Received: 7 December 2021 Accepted: 12 January 2022 Published: 17 January 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). monitor the evolving process of ABLH which is necessary for a range of meteorological applications. LiDAR is an active remote sensing tool with high spatial and temporal resolution, long detection range, and strong anti-interference ability, which can continuously and automatically measure the ABLH. It is assumed that the distribution of aerosols in the ABL is almost uniform and the concentration of aerosols in the ABL and FT, known as the entrainment zone (EZ). The center of the EZ is the ABLH at which the aerosol concentration decreases sharply [17,18]. There are two main kinds of methods to estimate ABLH based on LiDAR: (1) Methods to finding an abrupt change based on vertical distribution information regarding the aerosol concentration reflected by LiDAR backscatter signals, such as the threshold method [19], gradient method (GM), [20] and wavelet covariance transform method (WM) [21]; (2) methods based on statistical analysis that seek the maximum variance of the LiDAR backscatter signal [22]. However, these methods operate in cloud-free meteorological conditions, and will misjudge the ABLH from cloud interference.

Machine learning (ML)—which extracts features from the training data—is a powerful tool for classification and regression problems [23,24]. Moreover, ML has been widely used in target recognition [25], computer vision [26], and other fields [27–29], and has achieved remarkable results. In this paper, we regard the detection of ABLH as a cluster problem and explore how the appropriate algorithm can be used to solve this problem. Toledo proposed a cluster analysis (CA) method to calculate the ABLH by taking advantage of both the vertical distribution of aerosols and the variation in concentration [30]. The study indicated that cluster analysis is a reliable method for studying the ABLH. Toledo tested the robustness and capability of six numerical methods usually used to determine ABLH from LiDAR data under different atmospheric conditions [18]. The results showed that all methods are in agreement with radiosonde measurements under sea-land breeze and non-dusty conditions. However, the presence of a residual layer (RL) affected the use of these methods and CA performed best under dusty conditions. Caicedo assessed three aerosol LiDAR-retrieved algorithms for ABLH detection and found that there is good agreement between radiosonde and aerosol-derived ABLH in cloud-free conditions [17]. The presence of clouds creates difficulties for estimating ABLH for all methods. Rieutord described two machine learning methods that derive ABLH from LiDAR data, which are based on K-means and AdaBoost [31]. The case study indicated that both algorithms performed well. Nevertheless, K-means is sensitive to the clustering initial value, and AdaBoost is constrained by its training data. Moreover, the meteorological conditions of low-level clouds (cloud base height below 3 km) were discarded.

As is known, clouds cover about 60% of the globe [32]. Estimating ABLH under cloudy conditions is a necessary but challenging task. Krishnamurthy [33] used multiple instruments to define the ABLH under clouds conditions. However, these instruments are not always available for field measurements. Dang [34] proposed a top limit technique to eliminate interference from the cloud layer. However, in that study, the threshold -2 gradient used for cloud location determination was not applicable to other observation sites. Li [14] determined the ABLH based on the WCT method with whale optimization algorithm and the top limit. The empirical threshold -2 was also used as a reference for the detection of boundary layer height information. Zhong [35] estimated the ABLH under multi-layer conditions by using maximum limited height initialization and range restriction. Nevertheless, the parameter was determined by experience which is limited to the research area and is unsuitable for other areas. The applicability of these ABLH estimation methods is limited due to the uncertainty of threshold selection. Therefore, it is necessary to use methods other than threshold to estimate the ABLH. To improve the reliability and convenience of ABLH estimation under clouds or residual layer conditions, we propose a new algorithm, known as Mahalanobis transform K-near-means (MKnm), in which only LiDAR data are used. In this study, the observation datasets are constructed based on the characteristics of the categories. Mahalanobis transform is used to partition the data due to the correlation and different magnitudes of each feature in the observation data, and the interference from outliers is reduced with the data that near the previous centers to update the next centers. The most important factor is the LiDAR backscatter signal gradient which is used as a reference to select the initial clustering value. To verify the effectiveness of the proposed MKnm, micropulse LiDAR data—under cloud or residual layer aerosol conditions—from the Southern Great Plains site C1 were used for testing, and the ABLH defined by MKnm, K-means, VM, and HM were compared with reference ABLH calculated from radiosonde data.

The rest of this article is organized as follows. Section 2 introduces the materials used in our study. ABLH retrieval using existing methods and proposed MKnm are described in Section 3. The results of our study, including comparisons of the methods with radiosonde, are presented in Section 4. In Section 5, we discuss the proposed method. Section 6 concludes the study.

2. Materials

Ground-based micro pulse LiDAR (MPL) and radiosonde data used in this work were collected from the central facility (36.607322° N, 97.487643° W) near Lamont in northcentral Oklahoma at the Southern Great Plains (SGP) site of the Atmospheric Radiation Measurement (ARM) program [33,36,37]. The dataset spanned 2 years (2003 and 2004). The instruments described below refer mainly to works during this period.

2.1. Micro-Pulse LiDAR (MPL)

The micro-pulse LiDAR (Version 2.0) instrument installed at the SGP site is an eye-safe instrument that can be operated autonomously. It emits a laser pulse at a wavelength of 523 nm. Moreover, with pulse-repetition frequencies of 2.5 KHz, the maximum detection range is up to 60 km. The raw MPL data have a vertical resolution of 30 m and temporal resolution of 30 s. After system dead-time correction, background subtraction, range correction, overlap, after-pulsing, and energy-level normalization, the resulting signal is referred to as normalized relative backscatter (NRB):

$$NRB(r) = \frac{\left\{ [n(r) \times D[n(r)]] - n_{ap}(r) - n_b \right\} r^2}{O_c(r)E} = C\beta(r)T(r)^2$$
(1)

where n(r) represents the measured signal return in photoelectron counts per second at range r. D[n(r)] is the dead time; $n_{ap}(r)$ is the after-pulsing; n_b is background; $O_c(r)$ is the overlap factor; *C* represents the calibration constant of a dimensional system; *E* is the transmitted laser pulse energy; β is the backscatter coefficient; *T* is the atmospheric transmittance; NRB is the value-added data product (VAP) of ARM that is used for detecting clouds and aerosols, and the vertical resolution is 90 m. In this work, the interval thresholding technique is used to reduce noise interference [29]. NRB data below 4.37 km are used, and data from rain and fog meteorological conditions are discarded. In this study, we focus mainly on demonstrating the ability to detect the ABLH under cloud or residual layer aerosol conditions by the proposed method.

2.2. Radiosonde (RS)

Vaisala RS90 radiosonde was equipped at the SGP site that is usually launched four times daily—05:30, 11:30, 17:30, and 23:30 UTC [33]. It provides vertical variation information of air pressure, temperature, relative humidity. The most widely used approaches to determine the ABLH involve finding the maximal positive gradient of the vertical potential temperature profile or the minimum negative gradient value of the vertical relative humidity profile. Three different algorithms were used to estimate ABLH from RS data in the ARM ABLH VAP: (1) the Heffter method, (2) bulk Richardson number method [38,39], and (3) Liu and Liang method [33]. The Liu and Liang method determines ABLH of the convective, stable, and neutral boundary layer from radiosonde soundings profiles collected from 14 major field campaigns around the world [10]. It is a robust method and can produce realistic ABLH that can be verified by several thousand additional soundings; its perfor-

mance is also validated by substantial literature [14,33]. In particular, Krishnamurthy [33] compared these three methods using 1785 cases with radiosonde data and daytime clear or shallow cumulus conditions. From comparisons results, they found that the Liu and Liang technique resulted in the best overall agreement with the Tucker method, and used as a reference to calibrate the RF model in their study. The observation site in our study is same as Krishnamurthy's. Therefore, we also use Liu and Liang method as a reference to evaluate the proposed method. In order to match the ABLH detected by LiDAR data and RS data, the LiDAR data are the average within 10 min of launching radiosonde.

3. Methods

Figure 1 shows a typical example of NRB under cloudy conditions at SGP. Figure 1a illustrates the vertical distribution profiles of normalized relative LiDAR backscatter (NRB) signals observed at 23:31 UTC, 05 February 2003. Figure 1b,c shows the gradient and relative increase in NRB (RE = (NRB ($z + \Delta z$) – NRB (z))/NRB (z)), respectively. In Figure 1a, NRB almost decreases with altitude until approximately 2.2 km above ground level (AGL) where a rapid increase occurs and a sharp decline at approximately 2.4 km AGL, which correspond to the strongest positive and negative gradient pairs shown in Figure 1b. In addition, In Figure 1c, a relative increase in the NRB—larger than 0.55—is clearly shown between 2.2 km and 2.4 km AGL, indicating that one cloud layer is located between 2.2 km to 2.5 km, which is also confirmed by the cloud data from the ARM SGP site [36]. In Figure 1a, the LiDAR signal above the cloud top almost decays to zero due to cloud attenuation. The ABLH determined by RS is 1.082 km.



Figure 1. Vertical distributions of normalized relative LiDAR backscatter signal on typical cloud conditions: (**a**) one cloud layer is observed at 23:31UTC, 05 February 2003. Horizontal line with red color depicts the position of ABLH by RS; (**b**) NRB gradient; (**c**) relative increase in NRB.

3.1. Gradient Method (GM)

As mentioned above, the aerosol concentration in the boundary layer is significantly higher than that in the free atmosphere. There is a marked change in the LiDAR signal when it enters the free atmosphere from the boundary layer. Therefore, the boundary layer height is defined as the position where the LiDAR signal decays fastest or the height at which the minimum gradient is obtained, and is defined as:

$$ABLH_{GM} = \operatorname{argmin}_{r} \frac{d(NRB(r))}{dr}$$
(2)

Or

$$ABLH_{GLM} = \operatorname{argmin}_{r} \frac{d(\ln NRB(r))}{dr}$$
(3)

where NRB(r) represents the normalized relative backscatter at the height *r*. Here, we use Equation (2) to compute ABLH_{GM}. GM is simple and convenient but is easily disturbed by noise and the aerosol layer structure.

5 of 18

3.2. Wavelet Covariance Transform Method (WM)

Wavelet covariance transformation is a method that uses the Haar step function to detect the LiDAR signal step changes. The wavelet covariance function is defined as:

$$W_{NRB}(a,b) = \frac{1}{a} \int_{r_l}^{r_u} NRB(r)h\left(\frac{r-b}{a}\right) dr$$
(4)

where r_l and r_u are the lower and upper boundaries of the LiDAR profile; *a* and *b* are the dilation and translation parameters of the Haar step function. Here, $a = 2n\Delta r$, n is a positive integer, Δr denotes the vertical resolution of LiDAR, and b is set to 90 m step size from 0.3 km to 4.37 km considering the vertical resolution of NRB.

The Haar wavelet function is expressed as:

$$h\left(\frac{r-b}{a}\right) = \begin{cases} +1: & b-\frac{a}{2} \le r \le b\\ -1: & b \le r \le b+\frac{a}{2}\\ 0: & elsewhere \end{cases}$$
(5)

The ABLH obtained by WM is to find the maximum $W_{NRB}(a,b)$ and is shown as:

$$ABLH_{WM} = \operatorname{argmax}_{r}(W_{NRB}(a, b))$$
(6)

where WM shows good performance in many studies and is often used as the manufacturer's method. However, parameters *a* and *b* need to be appropriately selected.

3.3. K-Means Method

K-means is a classical unsupervised learning algorithm that partitions objects according to their distance from k cluster centers [40]. It is commonly used in many applications since it is simple and computationally efficient. The K-means implementation process to identify ABLH is as follows:

Step 1—Build dataset $X \in \mathbb{R}^{N \times F}$. N is the number of data points; F is the dimension of data that each dimension represents a feature of clusters. Here, F =3 –altitude r, normalized relative backscatter NRB(r), and the absolute value of the relative change in NRB(r) expressed as $|\Delta NRB(r)|$.

Step 2—Normalize the data making each dimension contribute equally to the partition.

Step 3—Choose the number of clusters k, and select the initial cluster center C_1, \ldots, C_k at random locations inside the dataset. Here, the number of clusters k is same as that determined by MKnm.

Step 4—Cluster the data X_i according to Euclidean distance:

$$X_i \in j, \qquad j = \operatorname{argmin}_i d(X_i, C_j)$$
(7)

where j = 1, ..., k, indicates the cluster; $d(X_i, C_j)$ is the Euclidean distance between data X_i and cluster center C_i .

Step 5—Update cluster center using the average of the data in the cluster, defined as:

$$C_j = \frac{1}{N_j} \sum_{X_i \in j} X_i \tag{8}$$

where N_i is the number of data points in cluster *j*.

Steps 4 and 5 are repeated until the maximum number of iterations is reached or the cluster centers stop moving. In this study, the ABLH determine by K-means is located at the category boundary for the first decrease in cluster strength from bottom to top.

3.4. MKnm Method

K-means has been used to detect ABLH from LiDAR data. However, it is sensitive to the number of clusters and initial cluster centers. Outliers can also affect cluster centers that keep them away from real values [31,40]. Moreover, the Euclidean distance measurement—often used in K-means—does not take into account the relationship between object features [41,42]. Therefore, MKnm is proposed to solve these problems.

3.4.1. Algorithm Description

MKnm estimates ABLH by cluster analysis from only the LiDAR signal. One key work of this method is to select features of the clusters. There are at least three clusters to be identified, such as ABL, FT, and cloud (for Figure 1). The algorithm is based on the fact that the intensity of the LiDAR backscatter signal generally decreases with altitude, however, increases sharply when interacting with clouds or uplifting aerosols and decreases rapidly at the upper part of them due to scattering and attenuation [34,35]. Therefore, the LiDAR backscatter signal changes significantly when it encounters clouds, as shown in Figure 1. In this study, we choose the feature of clusters as altitude r, normalized relative backscatter NRB(r), and the absolute value of the relative change in NRB(r) expressed as $|\Delta NRB(r)|$. Aerosols with adjacent positions, similar concentrations, and varying intensities are classified into one category. This supposes that the ABLH will be of higher accuracy under improved clustering.

In order to accurately partition data, the relationship of each feature should be considered. Two pieces of data that are closely related may not belong to the same cluster, and the magnitude of each feature can also affect the clustering results. MKnm takes the Mahalanobis transform to the raw dataset in order to eliminate effects arising from correlation and different magnitudes from each feature. These processes are expressed as:

$$Z = Y \begin{bmatrix} \frac{1}{\sigma_1} & & \\ & \frac{1}{\sigma_2} & \\ & & \frac{1}{\sigma_3} \end{bmatrix} = X U^T \begin{bmatrix} \frac{1}{\sigma_1} & & \\ & \frac{1}{\sigma_2} & \\ & & \frac{1}{\sigma_3} \end{bmatrix}$$
(9)

where *Z* is the new dataset obtained by taking Mahalanobis transform to raw dataset *X*; *Y* is the dataset after projecting *X* onto U^T , which is similar to principal component analysis (PCA); σ_i is the standard deviation of the ith feature of *Y*; U is the orthogonal matrix of the eigenvectors of Σ , expressed as $\Sigma = UQU^T$; Σ is the covariance matrix of *X*.

In the next step, MKnm uses K-near-means to partition dataset Z into k clusters by calculating the Euclidean distance between data and cluster centers. It should appropriately select the number of clusters k and the initial cluster centers C. The aim of this algorithm is to estimate ABLH which means the main focus is to identify ABL and FT. Therefore, a low number of clusters will not extract ABL and FT from NRB. However, a greater number of clusters will only further classify the ABL or FT and increase the cost, which has a minimal effect on ABLH estimation. As mentioned previously, the ABLH will be of higher accuracy under improved clustering. Figure 2 illustrates the clustering under varying numbers of clusters with random initial centers.

From Figure 2, it can be observed that the aerosol layer is further classified with an increase in the number of clusters k. It divides into cloud layer and non-cloud layer at k = 2 and divides into aerosol layer 1, aerosol layer 2, clouds layer, and the attenuation layer above the cloud k = 4. In our study, in order to clustering the aerosol properly, the NRB gradient is chosen as the observed object. The optimal number of clusters is determined by the number of same direction intervals on the gradient of NRB after the process using the interval thresholding technique. It is known that NRB almost decreases with altitude under ideal atmospheric conditions in which the number of same direction intervals is one. It can be identified as ABLH by dividing aerosols into two clusters. Under cloud atmospheric conditions, the NRB gradient is negative until it interacts with clouds or the elevated aerosol layer; it is positive at the lower part and negative at the upper part, corresponding to an enhancement and attenuation of backscatter signals. There are three same direction intervals of the NRB gradient that are needed to divide into three or four clusters to identify ABLH that includes ABL, FT, and cloud or elevated aerosol layer, and the attenuation layer

above the cloud may need to be selected. Therefore, the optimal number of clusters is the number of same direction intervals of the NRB gradient or the addition of one. The change in the NRB gradient from Figure 1 is shown in Figure 3.



Figure 2. The clustering under a different number of clusters with random initial center: (a) k = 2; (b) k = 3; (c) k = 4; (d) k = 5.



Figure 3. The gradient of NRB processed by interval thresholding technique.

In Figure 3, there are three same direction intervals. Thus, the optimal number of clusters is three or four. Next, we analyze the effect of the initial center in the cluster analysis. Figure 4 illustrates the clustering under different initial centers with a given

number of clusters. From the figure, it can be seen that the result of the cluster will be affected by the initial center. We can obtain different ABLH with different initial centers from the same number of clusters. Therefore, the initial center of clusters needs to be selected carefully. In our study, the maximum of the same direction intervals is selected as the initial centers, and is then fine-tuned.



Figure 4. The clustering under different initial center with four clusters.(**a**) C = [Z(4,:); Z(16,:); Z(24,:); Z(37,:)]; (**b**) C = [Z(4,:); Z(24,:); Z(25,:); Z(37,:)]. Z(n,:) represents the nth data in dataset Z.

Since the cluster center is the mean of the data in the cluster during the update process, the outliers will remove it from the real value. In this study, the data point that nears the cluster center is chosen to update it, defined as:

$$C_{j}^{(n+1)} = \frac{1}{N_{j}^{(n)}} \sum_{Z_{i}^{(n)} \in k_{i}^{(n)}, d(Z_{i}^{(n)}, C_{j}^{(n)}) \le 2H^{(n)}} Z_{i}^{(n)}$$
(10)

$$H^{(n)} = \frac{1}{N_j^{(n)}} \sum_{Z_i^{(n)} \in k_j^{(n)}} d(Z_i^{(n)}, C_j^{(n)})$$
(11)

where $C_j^{(n+1)}$ is the center of the *j*th cluster at time n + 1, $k_j^{(n)}$ is the *j*th cluster at time n, $N_j^{(n)}$ is the number of the *j*th cluster at time n, $Z_i^{(n)}$ is the data point *i* of the *j*th cluster at time n, $H^{(n)}$ is the mean Euclidean distance between data, and cluster centers at time n.

3.4.2. Flowchart of ABLH Estimated by MKnm

Figure 5 is the flowchart of ABLH estimated under complex atmospheric conditions based on LiDAR data by MKnm. The gradient and relative increases in NRB need to be calculated first to construct dataset X. Indeed, the intensity of the LiDAR backscatter signal usually declines with altitude. However, it increases rapidly when the cloud or elevated aerosol layer is encountered and decreases sharply at the upper boundary of each layer. Therefore, zero-crossing of the derivative occurs at the boundary of the cloud or elevated aerosol layer. Campbell also indicates that the relative increase in NRB will be at least 0.55 when a cloud exists [36]. Interval thresholding—which considers the same direction interval as the whole to implement threshold operations—shows better performance for removing noise interference. The NRB gradient that deals with interval thresholding can reflect the distribution of the cloud or elevated aerosol layer in vertical space. Then, the optimal number of clusters K can be determined by calculating the performance metrics. We select the maximum in the same direction interval from the NRB gradient as the initial cluster center.

Considering the correlation of elements contained in the data, a Mahalanobis transform is implemented for the dataset. In addition, data that near the cluster center are used to update the center to reduce interference from outliers. Performance metrics are calculated with the given initial cluster center, and fine-tuning is implemented until 10 times. Finally, the dataset is cluster analyzed using the optimal cluster number and initial cluster center. In order to estimate the ABLH, the average intensity of each cluster needs to be calculated. The ABLH is located at the category boundary for the first decrease in cluster strength from bottom to top.



Figure 5. Flowchart of ABLH estimation by MKnm from LiDAR data.

3.4.3. Performance Metrics

In this study, three indexes were used: (1) the sum of the squared errors (SSE); (2) the mean of the silhouette coefficient (MSC) [43]; and (3) Davies–Bouldin indices (DBI) [44], defined as:

$$SSE = \sum_{i=1}^{k} \sum_{Z \in C_i} d(Z, C_i)^2$$
(12)

where Z is the data in cluster *i*; *k* is the number of clusters; C_i is the cluster center of cluster *i*.

$$MSC = \frac{1}{N} \sum_{i=1}^{N} \frac{b_i - a_i}{\max\{a_i, b_i\}}$$
(13)

where *N* is the number of data points; a_i is the mean distance between data point Z_i and other data points with the same cluster. b_i is the minimum mean distance between data point Z_i and other data points with a different cluster and is defined as:

$$a_{i} = \frac{1}{N_{C_{i}} - 1} \sum_{Z_{i}, Z_{j} \in C_{i}, j \neq i} d(Z_{i}, Z_{j})$$
(14)

$$b_{i} = \min_{k \neq i} \frac{1}{N_{C_{k}}} \sum_{Z_{j} \in C_{k}} d(Z_{i}, Z_{j})$$
(15)

where N_{Ci} is the number in the cluster i; a_i represents the similarity between data point Z_i and other data points in the cluster. b_i represents the dissimilarity between data point Z_i and other data points in other clusters. MSC locates in the range of [-1, 1]. In general, a bigger MSC indicates a better clustering result, however, does not always work since it is affected by a_i and b_i . Hence, in our study, the initial centers of clustering are selected based on SSE and MSC. Davies–Bouldin indices also need to be used when selecting the optimal number of clusters, defined as:

$$DBI = \frac{1}{k} \sum_{i=1}^{k} \max_{i \neq j} \left[\frac{\mu_i + \mu_j}{d(C_i, C_j)} \right]$$
(16)

where *k* is the number of clusters, μ_i is the mean distance between data and their cluster center in cluster *i*, and $d(C_i, C_j)$ is the distance between cluster center C_i and cluster center C_j .

4. Results

To assess the validity of the proposed Mknm, this section shows the ABLH obtained different methods and analyzes the ABLH diurnal cycles under cloudy conditions using different methods. Furthermore, ABLHs retrieved by the LiDAR method are compared with radiosonde methods under residual layer or cloudy conditions.

4.1. ABLH Estimated by Different Methods

Figure 6 shows the ABLH obtained by four methods based on LiDAR data from Figure 1 defined as ABLH-GM, ABLH-WM, ABLH-Kmeans, and ABLH-Mknm. The reference ABLH defined by RS is 1.082 km. Notably, the best result is provided by Mknm, which has the lowest deviate (0.1622 km) and relative deviation (0.1499). The ABLH estimated by GM, WM, and K-means are close to the top of the cloud with a large deviate and relative deviation. The ABLH obtained by GM is highest in all methods which is farthest from the reference ABLH.



Figure 6. The ABLH obtained by four methods under cloudy conditions.

4.2. ABLH Diurnal Cycles under Cloudy Conditions

Figure 7 is the NRB time-height distribution, and the ABLH retrieved by different methods are shown using different symbols. The diurnal cycles in ABLH are detected by four methods with data measured at the ARM SGP site C1 on 2 October 2003. The NCRB are averaged every 10 min, and the convective boundary layer (CBL) in the daytime (from 11:05–23:45 UTC) is analyzed. Radiosonde launches at 11:30, 17:30, and 23:30, and ABLH determined by RS is shown in Figure 7 as the black diamond. As can be seen from Figure 7, a layer or more clouds appear in the sky throughout the day. ABLHs—calculated by GM and WM—are almost located at the top of the cloud where the strongest change in aerosol concentration occurs. However, ABLHs detected by Kmeans and Mknm are in good agreement with the change in ABLHs, and Mknm can precisely identify ABLHs. From 11:05–14:00, the boundary layer is clearly visible at lower altitudes, and there is a thick high cloud layer above 2.1 km AGL which is shown as a white region. The NCRB is completely attenuated above the cloud top at this time. GM and WM capture the upper edge of the cloud as ABLH with a large error. Kmeans and Mknm are able to correctly identify ABLHs, and Mknm obtains the best results that correctly and accurately capture the change process of ABLHs. Between 17:05 and 19:05 UTC, more than one cloud layer appears in the sky. GM and WM detect the cloud layer, which occurs as the strongest change in aerosol concentration as ABLH. Kmeans and Mknm capture the location as a marked change in aerosol concentration under clouds as ABLH. From 19:25 to 21:05 UTC, the ABLH may couple with the cloud layer due to no sharp change in NRB beside them; the ABLH is located at the upper edge of the cloud estimated by all four methods. It is clear that the ABLH determined by MKnm and Kmeans match very well with that obtained by radiosonde at 11:30 and 23:30 UTC, and MKnm performance best. GM and WM overestimate ABLH as estimated from radiosonde. At 17:30 UTC, there is a low cloud. The ABLH estimated by radiosonde method is the base of the cloud layer where the temperature profile changes rapidly. Considering the continuity of ABLH in two adjacent moments, the ABLH determined by MKnm is selected [34].



Figure 7. Resulting ABLH for 2 October 2003 with 10 min averages for all methods. Radiosondeestimated ABLHs are shown as the black diamond.

4.3. Comparisons of ABLH Retrieval between LiDAR and Radiosonde Methods under Clouds or RL Conditions

In order to further evaluate and quantify the performance of the proposed method, the ABLH detected by four LiDAR-based methods in the presence of RL and cloud layers are compared with ABLH identified by radiosonde. The mean absolute error (MAE) and Pearson correlation coefficient (R) are used as scores to assess the quality of methods. They are defined as:

$$MAE = \frac{1}{N} \sum_{n=1}^{N} \left| \hat{Z}(n) - Z(n)_{RS} \right|$$
(17)

$$R = \frac{\operatorname{cov}(\hat{Z}, Z_{RS})}{\sigma(\hat{Z})\sigma(Z_{RS})}$$
(18)

where *N* is the number of data points, \hat{Z} is the ABLH detected by LiDAR-based methods, and Z_{RS} is the ABLH determined by radiosonde. cov(•) denotes the covariance, and $\sigma(\bullet)$ denotes the standard deviation. A lower MAE and higher R are better since this indicates that ABLHs detected by LiDAR-based methods are closer to references. Moreover, the mean deviation D—which is the difference between the mean of ABLH detected by LiDAR-based methods and radiosonde—was also used to assess the methods.

The residual layer (RL) occurs in the morning or evening and is disconnected from the surface aerosol layer. It has the same properties as the cloud layer when interacting with the LiDAR beam that affects the defined ABLH. Figure 8 shows the comparisons between ABLH results determined by LiDAR-based methods and radiosonde in cloud or RL scenarios at the ARM SGP site from January 2003 to May 2004. Previous studies have indicated that the ABLH derived from LiDAR-based methods is affected by distinct weather patterns and LiDAR-blind areas. Therefore, meteorological conditions including rain, snow, or fog were not considered, and the ABLH below 300 m, determined by radiosonde, was also discarded. This study selected 56 well-defined cases for comparison. Figure 8a compares the ABLH estimated by GM and radiosonde. The ABLH determined by GM is much higher than that obtained from radiosonde due to the presence of clouds or RL, with a correlation coefficient of 0.23, a mean absolute error of 1698 m, and a mean deviation of 1690 m. The comparison results of ABLH-determined by WM and by radiosonde-are shown in Figure 8b, with a correlation coefficient of 0.24, a mean absolute error of 1639 m, and a mean deviation of 1624 m. From Figure 8a,b, it can be seen that ABLH determined by GM and WM both have large errors with ABLH determined by radiosonde. Given the number of clusters, the ABLH determined by cluster analysis (K-means and MKnm) is in close agreement with those by radiosonde as shown in Figure 8c,d. The correlation coefficient between K-means and radiosonde derived ABLH is 0.77, the mean absolute error is 220 m, and the mean deviation is 86 m. The proposed MKnm method obtained the best result, which is consistent with radiosonde, with a correlation coefficient of 0.95, a mean absolute error of 87 m, and a mean deviation of -8 m. Table 1 shows the comparison results of ABLH-determined by LiDAR-based methods and by radiosonde. The results indicate that MKnm performs well under cloud or RL conditions.

Table 1. Correlation coefficients (R), mean absolute error (MAE), mean deviation (D) between ABLH determined by LiDAR-based methods and radiosonde.

Method	R	MAE(m)	D(m)
GM	0.23	1698	1690
WM	0.24	1639	1634
K-means	0.77	220	86
MKnm	0.95	87	-8



Figure 8. Comparisons between ABLH results determined by LiDAR-based methods and radiosonde for 56 cases in cloud or RL conditions. (**a**) GM and RS; (**b**) WM and RS; (**c**) Kmeans and RS; (**d**) MKnm and RS. The red dotted line is the expected 1:1 reference line. The correlation coefficient, mean absolute error and mean deviation are represented by R, MAE and D, respectively.

Figure 9 shows a comparison of the ABLH determined by four LiDAR-based methods. From the figure, it can be seen that ABLHs—determined by GM and WM—have the closest correlation, with a correlation coefficient of 0.99. The result is consistent with those described previously since they are methods that find the maximum change in the LiDAR backscatter signal. The cluster analysis methods—K-means and MKnm—are closely related, with a correlation coefficient of 0.8. However, it also shows that the ABLH—defined by MKnm—varies slightly from that defined by K-means. Compared with MKnm, K-means with random initial center has a larger error. Table 2 shows the comparison results of ABLH—determined by these four LiDAR-based methods. Overall, in most cases, the ABLHs estimated by K-means and MKnm are quite different from those by GM and WM in the presence of clouds or RL.

Table 2. Correlation coefficients (R) of ABLH determined by LiDAR-based methods.

	GM	WM	K-means	MKnm
GM	1	0.99	0.24	0.19
WM	0.99	1	0.25	0.21
K-means	0.24	0.25	1	0.80
MKnm	0.19	0.21	0.80	1



Figure 9. Intercomparison of the ABLH determined by LiDAR-based methods for 56 cases in cloud or RL conditions. (a) WM and GM; (b) Kmeans and GM; (c) MKnm and GM; (d) Kmeans and WM; (e) MKnm and WM; (f) MKnm and Kmeans. The red dotted line is the expected 1:1 reference line.

5. Discussion

5.1. Evaluation Index to Quantify the Quality of Clustering

SSE is the sum of the square errors between observations and their cluster center and decreases with an increase in the number of clusters. The magnitude of decline is initially large and gradually decreases. Researchers often use the elbow method to select the optimal number of clusters. Nevertheless, it is easy to fall into local optimum, and the selection of elbow points has a certain uncertainty in which errors are easily produced. In this study, we used SSE as one evaluation index to select the initial cluster centers; because, in the case of the same cluster number K, SSE can better evaluate the clustering effect, which is conducive to selecting better initial centers. MSC combines the cohesion of observations within clusters and the separation of observations between clusters. In terms of good clustering performance, observations of the same cluster should be as similar as

15 of 18

possible and vice versa. Thus, a larger MSC is better. The DBI index tests the similarity of observations among clusters, and better clustering results lead to significant differences among clusters, that is, a smaller DBI is better. MSCand DBI are effect evaluation indexes that quantify the quality of clustering which is conducive to selecting the optimal clustering category K and the initial center C.

5.2. Define the ABLH after Clustering

According to the aerosol vertical structure distribution characteristics, the ABL is in the lowest layer of the troposphere. Therefore, after clustering, the bottom category is generally ABL, and the upward category adjacent to its potential boundary layer, free atmosphere, cloud, or RL, can be judged according to the aerosol intensity (average) information from these categories. If the intensity of the upward category does not differ greatly from that of the bottom category but is weaker, the upward category is the boundary layer. If the intensity differs greatly from the known boundary layer category and is weaker, the category is free atmosphere. If the intensity differs greatly from the known boundary layer category, the category is residual layer aerosol or cloud. Based on this, the boundary layer height can be obtained.

5.3. Estimation of ABLH above the Cloud

The ABLH is mostly below the cloud, but in rare cases, it may be coupled to or appear above the cloud [34]. In this scenario, it is challenging to define ABLH based on ground-based LiDAR only, and presently, there is no suitable method. In many cases, due to cloud attenuation, even for high-power systems, the LiDAR signal completely decays before reaching the cloud top, and the LiDAR pulse may be unable to penetrate the cloud to obtain the backscattered LiDAR signal generated by aerosols above the cloud top [45]. In this case, the LiDAR echo signal on the cloud top will attenuate into background noise, therefore, making it difficult to detect the ABLH on and above the cloud top. As the LiDAR backscattering signal rapidly decays at the upper boundary of the cloud, the gradient-based boundary layer height detection method (gradient method and wavelet method) determines the upper boundary of the cloud as the boundary layer height, and the clustering analysis method also determines the upper boundary of the cloud as the boundary layer height. However, with the weak attenuation of clouds, the LiDAR pulse can penetrate the clouds and obtain aerosol information above the cloud. The methods based on the clustering analysis have great advantages, e.g., they can estimate the ABLH locating at that area. however, the gradient-based method will also define the ABLH as the cloud top.

6. Conclusions

The estimation of ABLH based on the LiDAR backscatter signal is significantly influenced by complex atmospheric structures such as clouds and uplifting aerosol layers. This paper proposed a new method known as MKnm to define ABLH under cloud or RL conditions. MKnm is a clustering method that aims to overcome the disadvantage of classic clustering methods as K-means. Firstly, it estimates the optimal number of clustering. The NRB gradient reflects the varying aerosol concentration intensity. We can obtain the overall distribution of aerosol concentration intensity in vertical space by implementing interval thresholding. Thus, the optimal number of clustering is related to the number of same direction intervals on the NRB gradient. Secondly, it selects initial centers of clustering. The experiment proved that the ABLH can be affected by initial centers of clustering. In this paper, the interval extreme value was used as the candidate center, and the initial center was determined by fine-tuning based on the evaluation index. Thirdly, Mahalanobis distance was used to partition observations by considering the correlation and different magnitudes of each feature, and the near observations were used to update the centers by eliminating interference from outliers. MKnm, K-means, GM, and WM were compared with RS by taking the LiDAR data under cloud or RL conditions from the ARM SGP C1 site. The diurnal cycles of ABLH from cloudy weather were detected by using four methods. MKnm and K-means can trace the changing trend of ABLH, however, GM and WM misidentified the cloud top or residual layer as ABLH. MKnm outperformed the K-means which sometimes had a large error. The ABLH estimated by MKnm, K-means, GM, and WM were compared with that defined by radiosonde based on LiDAR data (from January 2003 to May 2004) under cloud or RL conditions. Experimental results showed that GM and WM were similar in estimating ABLH, and the ABLH obtained by both was less related to the ABLH defined by radiosonde; the value was only 0.23–0.24, and the mean absolute error and mean deviation were larger (above 1600 m). However, the ABLH obtained by the clustering analysis method is in close agreement with the ABLH defined by radiosonde—the mean absolute error and mean deviation are small. The proposed MKnm performed best in these methods, with a correlation coefficient of 0.95, a mean absolute error of 87 m, and a mean deviation of -8 m. Overall, MKnm can—with high accuracy—estimate the presence of ABLH in complex weather conditions, and, from the results of this study, the performance of K-means can be improved.

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

Funding: This research was funded by the National Natural Science Foundation of China, grant number 61875089, 61905116, 61605082, 62175114; the Priority Academic Program Development of Jiangsu Higher Education Institutions, China, grant number 1081080015001; the Top-notch Academic Programs Project of Jiangsu Higher Education Institutions, China, grant number 1181081501003; the Natural Science Foundation of Jiangsu Province, China, grant number BK20190771; the Natural Science Foundation of the Jiangsu Higher Education Institutions of China, grant number 16KJB510020; the "Qing Lan" project of Jiangsu Universities, China; the Key Scientific Research Project of Taizhou Polytechnic College, China, grant number TZYKYZD-19-5.

Data Availability Statement: The LiDAR and radiosonde data used in this article are publicly available on the ARM web page (https://adc.arm.gov/discovery/#/ (accessed on 02 October 2021)). The meteorological data at the Southern Great Plains site C1 can be available from the publicly web page (https://en.tutiempo.net/climate/oklahoma-state.html (accessed on 01 December 2021), http://weather.uwyo.edu/ (accessed on 01 December 2021)).

Acknowledgments: The authors would like to acknowledge the provision of radiosonde and MPL by the U.S. Department of Energy's ARM program.

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

Abbreviations

Atmospheric boundary layer	ABL
Atmospheric boundary layer height	ABLH
Radiosonde	RS
Free Troposphere	FT
Entrainment zone	ΕZ
Gradient method	GM
Wavelet covariance transform method	WM
Machine learning	ML
Cluster analysis	CA
Residual layer	RL
Mahalanobis transform K-near-means	MKnm
Micro-pulse LiDAR	MPL

Southern Great Plains	SGP
Atmospheric Radiation Measurement	ARM
Normalized relative backscatter	NRB
Value-added data product	VAP
Sum of the squared errors	SSE
Mean of the silhouette coefficient	MSC
Davies-Bouldin indices	DBI
Mean absolute error	MAE

References

- Helbig, M.; Gerken, T.; Beamesderfer, E.R.; Baldocchi, D.D.; Banerjee, T.; Biraud, S.C.; Brown, W.O.J.; Brunsell, N.A.; Burakowski, E.A.; Burns, S.P.; et al. Integrating continuous atmospheric boundary layer and tower-based flux measurements to advance understanding of land-atmosphere interactions. *Agric. For. Meteorol.* 2021, 307, 108509. [CrossRef]
- Su, T.; Li, J.; Li, C.; Xiang, P.; Lau, A.K.H.; Guo, J.; Yang, D.; Miao, Y. An intercomparison of long-term planetary boundary layer heights retrieved from CALIPSO, ground-based lidar, and radiosonde measurements over Hong Kong. J. Geophys. Res. 2017, 122, 3929–3943. [CrossRef]
- Kim, M.H.; Yeo, H.; Park, S.; Park, D.H.; Omar, A.; Nishizawa, T.; Shimizu, A.; Kim, S.W. Assessing CALIOP-derived planetary boundary layer height using ground-based lidar. *Remote Sens.* 2021, 13, 1496. [CrossRef]
- 4. Feudo, T.L.; Calidonna, C.R.; Avolio, E.; Sempreviva, A.M. Study of the Vertical Structure of the Coastal Boundary Layer Integrating Surface Measurements and Ground-Based Remote Sensing. *Sensors* **2020**, *13*, 6516. [CrossRef]
- 5. Zhang, Y.; Chen, S.; Chen, S.; Chen, H.; Guo, P. A novel lidar gradient cluster analysis method of nocturnal boundary layer detection during air pollution episodes. *Atmos. Meas. Tech.* **2020**, *13*, 6675–6689. [CrossRef]
- Wang, D.; Szczepanik, D.; Stachlewska, I.S. Interrelations between surface, boundary layer, and columnar aerosol properties derived in summer and early autumn over a continental urban site in Warsaw, Poland. *Atmos. Chem. Phys.* 2019, 19, 13097–13128. [CrossRef]
- 7. Fang, Z.; Yang, H.; Cao, Y.; Xing, K.; Liu, D.; Zhao, M.; Xie, C. Study of persistent pollution in hefei during winter revealed by ground-based LiDAR and the CALIPSO satellite. *Sustainability* **2021**, *13*, 875. [CrossRef]
- 8. Bain, C.L.; Parker, D.J.; Taylor, C.M.; Kergoat, L.; Guichard, F. Observations of the nocturnal boundary layer associated with the West African monsoon. *Mon. Weather Rev.* **2010**, *138*, 3142–3156. [CrossRef]
- Moreira, G.d.A.; Guerrero-Rascado, J.L.; Bravo-Aranda, J.A.; Foyo-Moreno, I.; Cazorla, A.; Alados, I.; Lyamani, H.; Landulfo, E.; Alados-Arboledas, L. Study of the planetary boundary layer height in an urban environment using a combination of microwave radiometer and ceilometer. *Atmos. Res.* 2020, 240, 104932. [CrossRef]
- 10. Liu, S.; Liang, X.Z. Observed diurnal cycle climatology of planetary boundary layer height. J. Clim. 2010, 23, 5790–5809. [CrossRef]
- 11. Guo, J.; Li, Y.; Cohen, J.B.; Li, J.; Chen, D.; Xu, H.; Liu, L.; Yin, J.; Hu, K.; Zhai, P. Shift in the Temporal Trend of Boundary Layer Height in China Using Long-Term (1979–2016) Radiosonde Data. *Geophys. Res. Lett.* **2019**, *46*, 6080–6089. [CrossRef]
- 12. Seidel, D.J.; Ao, C.O.; Li, K. Estimating climatological planetary boundary layer heights from radiosonde observations: Comparison of methods and uncertainty analysis. *J. Geophys. Res. Atmos.* **2010**, *115*, 16113. [CrossRef]
- 13. Korhonen, K.; Giannakaki, E.; Mielonen, T.; Pfüller, A.; Laakso, L.; Vakkari, V.; Baars, H.; Engelmann, R.; Beukes, J.P.; Van Zyl, P.G.; et al. Atmospheric boundary layer top height in South Africa: Measurements with lidar and radiosonde compared to three atmospheric models. *Atmos. Chem. Phys.* **2014**, *14*, 4263–4278. [CrossRef]
- 14. Li, H.; Chang, J.; Liu, Z.; Zhang, L.; Dai, T.; Chen, S. An improved method for automatic determination of the planetary boundary layer height based on lidar data. *J. Quant. Spectrosc. Radiat. Transf.* **2020**, 257, 107382. [CrossRef]
- 15. Du, L.; Pan, Y.; Wang, W. Random sample fitting method to determine the planetary boundary layer height using satellite-based lidar backscatter profiles. *Remote Sens.* **2020**, *12*, 4006. [CrossRef]
- 16. Dang, R.; Yang, Y.; Hu, X.M.; Wang, Z.; Zhang, S. A review of techniques for diagnosing the atmospheric boundary layer height (ABLH) using aerosol lidar data. *Remote Sens.* **2019**, *11*, 1590. [CrossRef]
- 17. Caicedo, V.; Rappenglück, B.; Lefer, B.; Morris, G.; Toledo, D.; Delgado, R. Comparison of aerosol lidar retrieval methods for boundary layer height detection using ceilometer aerosol backscatter data. *Atmos. Meas. Tech.* **2017**, *10*, 1609–1622. [CrossRef]
- Toledo, D.; Córdoba-Jabonero, C.; Adame, J.A.; De La Morena, B.; Gil-Ojeda, M. Estimation of the atmospheric boundary layer height during different atmospheric conditions: A comparison on reliability of several methods applied to lidar measurements. *Int. J. Remote Sens.* 2017, *38*, 3203–3218. [CrossRef]
- 19. Melfi, S.H.; Spinhirne, J.D.; Chou, S.H.; Palm, S.P. Lidar observations of vertically organized convection in the planetary boundary layer over the ocean. *J. Clim. Appl. Meteorol.* **1985**, *24*, 806–821. [CrossRef]
- Hayden, K.L.; Anlauf, K.G.; Hoff, R.M.; Strapp, J.W.; Bottenheim, J.W.; Wiebe, H.A.; Froude, F.A.; Martin, J.B.; Steyn, D.G.; McKendry, I.G. The vertical chemical and meteorological structure of the boundary layer in the Lower Fraser Valley during Pacific '93. *Atmos. Environ.* 1997, *31*, 2089–2105. [CrossRef]
- 21. Compton, J.C.; Delgado, R.; Berkoff, T.A.; Hoff, R.M. Determination of planetary boundary layer height on short spatial and temporal scales: A demonstration of the covariance wavelet transform in ground-based wind profiler and lidar measurements. *J. Atmos. Ocean. Technol.* **2013**, *30*, 1566–1575. [CrossRef]

- 22. Menut, L.; Flamant, C.; Pelon, J.; Flamant, P.H. Urban boundary-layer height determination from lidar measurements over the Paris area. *Appl. Opt.* **1999**, *38*, 945. [CrossRef]
- Kang, J.; Ullah, Z.; Gwak, J. Mri-based brain tumor classification using ensemble of deep features and machine learning classifiers. Sensors 2021, 21, 2222. [CrossRef]
- Al-Sahaf, H.; Bi, Y.; Chen, Q.; Lensen, A.; Mei, Y.; Sun, Y.; Tran, B.; Xue, B.; Zhang, M. A survey on evolutionary machine learning. J. R. Soc. N. Z. 2019, 49, 205–228. [CrossRef]
- 25. Li, Y.; Yu, F.; Cai, Q.; Qian, M.; Liu, P.; Guo, J.; Yan, H.; Yuan, K.; Yu, J. Design of Target Recognition System Based on Machine Learning Hardware Accelerator. *Wirel. Pers. Commun.* **2018**, *102*, 1557–1571. [CrossRef]
- Zhang, F.; Li, W.; Zhang, Y.; Feng, Z. Data Driven Feature Selection for Machine Learning Algorithms in Computer Vision. *IEEE Internet Things J.* 2018, *5*, 4262–4272. [CrossRef]
- 27. Maxwell, A.E.; Warner, T.A.; Fang, F. Implementation of machine-learning classification in remote sensing: An applied review. *Int. J. Remote Sens.* 2018, 39, 2784–2817. [CrossRef]
- Yu, W.; Liu, Y.; Ma, Z.; Bi, J. Improving satellite-based PM2.5 estimates in China using Gaussian processes modeling in a Bayesian hierarchical setting. *Sci. Rep.* 2017, 7, 7048. [CrossRef]
- 29. Liu, Z.; Chang, J.; Li, H.; Zhang, L.; Chen, S. Signal Denoising Method Combined with Variational Mode Decomposition, Machine Learning Online Optimization and the Interval Thresholding Technique. *IEEE Access* **2020**, *8*, 223482–223494. [CrossRef]
- Toledo, D.; Córdoba-Jabonero, C.; Gil-Ojeda, M. Cluster analysis: A new approach applied to lidar measurements for atmospheric boundary layer height estimation. J. Atmos. Ocean. Technol. 2014, 31, 422–436. [CrossRef]
- 31. Rieutord, T.; Aubert, S.; MacHado, T. Deriving boundary layer height from aerosol lidar using machine learning: KABL and ADABL algorithms. *Atmos. Meas. Tech.* **2021**, *14*, 4335–4353. [CrossRef]
- 32. Rossow, W.B.; Schiffer, R.A. Advances in Understanding Clouds from ISCCP. *Bull. Am. Meteorol. Soc.* **1999**, *80*, 2261–2287. [CrossRef]
- Krishnamurthy, R.; Newsom, R.K.; Berg, L.K.; Xiao, H.; Ma, P.L.; Turner, D.D. On the estimation of boundary layer heights: A machine learning approach. *Atmos. Meas. Tech.* 2021, 14, 4403–4424. [CrossRef]
- Dang, R.; Yang, Y.; Li, H.; Hu, X.M.; Wang, Z.; Huang, Z.; Zhou, T.; Zhang, T. Atmosphere boundary layer height (ABLH) determination under multiple-layer conditions using micro-pulse lidar. *Remote Sens.* 2019, 11, 263. [CrossRef]
- 35. Zhong, T.; Wang, N.; Shen, X.; Xiao, D.; Xiang, Z.; Liu, D. Determination of planetary boundary layer height with lidar signals using maximum limited height initialization and range restriction (MLHI-RR). *Remote Sens.* **2020**, *12*, 2272. [CrossRef]
- Campbell, J.R.; Hlavka, D.L.; Welton, E.J.; Flynn, C.J.; Turner, D.D.; Spinhirne, J.D.; Stanley, S.; Hwang, I.H. Full-time, eye-safe cloud and aerosol lidar observation at atmospheric radiation measurement program sites: Instruments and data processing. *J. Atmos. Ocean. Technol.* 2002, 19, 431–442. [CrossRef]
- 37. Welton, E.J.; Campbell, J.R.; Spinhirne, J.D.; Scott III, V.S. Global monitoring of clouds and aerosols using a network of micropulse lidar systems. *Lidar Remote Sens. Ind. Environ. Monit.* **2001**, *4153*, 151. [CrossRef]
- Li, H.; Liu, B.; Ma, X.; Jin, S.; Ma, Y.; Zhao, Y.; Gong, W. Evaluation of retrieval methods for planetary boundary layer height based on radiosonde data. *Atmos. Meas. Tech.* 2021, 14, 5977–5986. [CrossRef]
- 39. Min, J.S.; Park, M.S.; Chae, J.H.; Kang, M. Integrated System for Atmospheric Boundary Layer Height Estimation (ISABLE) using a ceilometer and microwave radiometer. *Atmos. Meas. Tech.* 2020, *13*, 6965–6987. [CrossRef]
- Qi, J.; Yu, Y.; Wang, L.; Liu, J.; Wang, Y. An effective and efficient hierarchical K-means clustering algorithm. Int. J. Distrib. Sens. Netw. 2017, 13, 8. [CrossRef]
- 41. Xiang, S.; Nie, F.; Zhang, C. Learning a Mahalanobis distance metric for data clustering and classification. *Pattern Recognit.* 2008, 41, 3600–3612. [CrossRef]
- 42. Omara, I.; Hagag, A.; Ma, G.; El-Samie, F.E.A.; Song, E. A novel approach for ear recognition: Learning Mahalanobis distance features from deep CNNs. *Mach. Vis. Appl.* **2021**, *32*, 38. [CrossRef]
- Lovmar, L.; Ahlford, A.; Jonsson, M.; Syvänen, A.C. Silhouette scores for assessment of SNP genotype clusters. BMC Genom. 2005, 6, 35. [CrossRef]
- 44. Davies, D.L.; Bouldin, D.W. A Cluster Separation Measure. *IEEE Trans. Pattern Anal. Mach. Intell.* **1979**, *PAMI-1*, 224–227. [CrossRef]
- Pal, S.R.; Steinbrecht, W.; Carswell, A.I. Automated method for lidar determination of cloud-base height and vertical extent. *Appl. Opt.* **1992**, *31*, 1488. [CrossRef] [PubMed]