

Article

A Cluster Analysis Approach for Nocturnal Atmospheric Boundary Layer Height Estimation from Multi-Wavelength Lidar

Zhongmin Zhu ^{1,†}, Hui Li ^{2,†}, Xiangyang Zhou ^{1,*}, Shumin Fan ³, Wenfa Xu ¹ and Wei Gong ²¹ College of Information Science and Engineering, Wuchang Shouyi University, Wuhan 430064, China² School of Electronic Information, Wuhan University, Wuhan 430072, China³ School of Information Science and Engineering, Dalian Polytechnic University, Dalian 116034, China

* Correspondence: zhoxiangyang@wsyu.edu.cn

† These authors contributed equally to this work.

Abstract: The atmospheric boundary layer provides useful information about the accumulation and diffusion of pollutants. As a fast method, remote sensing techniques are used to retrieve the atmospheric boundary layer height (ABLH). Atmospheric detection lidar has been widely applied for retrieving the ABLH by providing information on the vertical distribution of aerosols. However, these previous algorithms that rely on gradient change are susceptible to residual layers. Contrary to the use of gradient change to retrieve ABLH, in this paper, we propose using a cluster analysis approach through multifunction lidar remote sensing techniques due to its increasing availability. The clustering algorithm for multi-wavelength lidar data can be divided into two parts: characteristic signal selection and selection of the classifier. First, since the separability of each type of signal is different, careful selection of the input characteristic signal is important. We propose using Fourier transform for all the observed signals; the most suitable characteristic signal can be determined based on the dispersion degree of the signal in the frequency domain. Then, the performances of four common classifiers (K-means method, Gaussian mixture model, hierarchical cluster method (HCM), and density-based spatial clustering of applications with noise) are evaluated by comparing with the radiosonde measurements from June 2015 to June 2016. The results show that the performance of the HCM classifier is the best under all states ($R^2 = 0.84$ and RMSE = 0.18 km). The findings obtained here offer insight into ABLH remote sensing technology.

Keywords: atmospheric boundary layer; multi-wavelength lidar; cluster analysis



Citation: Zhu, Z.; Li, H.; Zhou, X.; Fan, S.; Xu, W.; Gong, W. A Cluster Analysis Approach for Nocturnal Atmospheric Boundary Layer Height Estimation from Multi-Wavelength Lidar. *Atmosphere* **2023**, *14*, 847. <https://doi.org/10.3390/atmos14050847>

Academic Editor: Philippe Thunis

Received: 3 April 2023

Revised: 7 May 2023

Accepted: 8 May 2023

Published: 9 May 2023



Copyright: © 2023 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/>).

1. Introduction

The atmospheric boundary layer (ABL) is commonly defined as the layer of the troposphere that is directly impacted by the Earth's surface and its features [1]. It is characterized by a rapid response to surface forcings, typically on a timescale of an hour or less [2]. The boundary layer thickness is quite variable in time and space, ranging from hundreds of meters to a few kilometers [3]. The mixing atmospheric boundary layer height (ABLH) is an important parameter to quantify the evolution of the ABL [4,5]. Moreover, the ABLH is directly related to the accumulation and diffusion of pollutants [6,7]. Therefore, accurate ABLH retrieval is crucial for understanding the vertical extent of turbulent mixing, vertical diffusion, and convective transport within the ABL.

Various profiling methods have been employed to estimate ABLHs, including radiosondes (RSs), lidar, wind profile radar, and ceilometers [8–14]. Among these, RSs are widely used in ABL research since they can determine dynamic processes in the atmosphere, in particular atmospheric stability, which drives the diurnal variation of the ABL. RSs can measure meteorological parameters directly, and thus, can determine thermally or mechanically driven ABLs [15,16]. However, the time resolution of a radiosonde is too low, i.e., two or three times a day (some special periods), to investigate the temporal variation of

the ABL [17]. In contrast, lidar systems can provide atmospheric vertical information with high temporal and spatial resolution [18,19]. A lidar system is an active remote sensing instrument that can measure the vertical distribution of atmospheric aerosols. Atmospheric aerosol vertical profiles are widely used to monitor the nocturnal stable layer height, internal aerosol layers, and the nighttime residual layer height. The ABLH can be inverted based on the aerosol vertical profiles [20–22]. For long-term observation of ABL structure using lidar, a reliable algorithm is required to manipulate the large datasets acquired.

The principal algorithms used to determine the ABLH from lidar systems include the gradient method, ideal profile fitting method, wavelet covariance transform (WCT) method, and maximum variance technique [23–26], among others. Specifically, as one of the earliest algorithms applied to the lidar system, the gradient method determines the ABLH by searching for the local maximum gradient from the vertical aerosol profile [23]; however, it is vulnerable to background noise. The ideal profile fitting method is designed to retrieve the well-mixed ABLs, but it is susceptible to the effect of complex aerosol layers [24]. The WCT method performs well when processing complex cases because the operator can select an appropriate base function and set an appropriate threshold [25]. The commonality of these algorithms is based on the vertical distribution of aerosol concentration to identify the ABLH. However, the accuracies of the ABLHs retrieved by this commonality are affected by the residual layer height. In particular, under weak convection conditions during the nighttime, the vertical structure of the atmosphere can be divided into the stable ABL, the residual layer, and the free atmosphere; thus, accurate determination of the ABLH using these lidar algorithms is difficult [27,28].

To overcome the effect of the residual layer height (RLH), many novel lidar algorithms have been developed to determine the ABLH, which have provided more insight into ABL remote sensing [29–31]. For instance, Pal et al. [30] and Bruine et al. [31] proposed a multi-parameter combination method that combined lidar data and meteorological data (e.g., temperature, humidity, and stability index) to estimate ABLH. This method had good stability and practicability; however, it depended on the corresponding meteorological parameters and could not rely on lidar data for independent retrieval. Due to the development of lidar system hardware technology, a multifunction lidar system can provide the backscatter coefficients (BCs) of multiple wavelengths as well as the color ratio (CR) and depolarization ratio (DR) information of aerosol particles. Based on a variety of lidar signals, Toledo et al. and Liu et al. proposed a cluster analysis approach that could be applied to lidar data to retrieve the ABLH [28,32]. Its main process comprised two steps: first, the characteristic signals representing the size scale (i.e., CR), shape (i.e., DR), or scattering ability (i.e., BCs) of the atmospheric particles are input; then, the vertical structure of the atmosphere is divided into the ABL category and the free atmosphere category according to the similar characteristics of atmospheric particles in the ABL. This algorithm avoids the effects of residual layers by combining multiple lidar signals; however, it should be noted that the selection of the characteristic signal and the selection of the classifier are very important for the accuracy of the ABLH [6]. As the separability of each type of signal is different, careful selection of the input characteristic signal is crucial. Moreover, the performance of the classifier is targeted, and therefore, it is necessary to select the most appropriate classifier for the feature signal classification.

This study evaluates the selection of the input characteristic signal and the performance of several different classifiers using a cluster analysis approach to obtain the ABLH. First, taking the dual-wavelength polarization lidar as an example, the BCs, CR, and DR are used to study the selection of the input characteristic signal. Then, the performances of the K-means method (KM), Gaussian mixture model (GMM), hierarchical cluster method (HCM), and density-based spatial clustering of applications with noise (DBSCAN) classifiers are compared under different ABL conditions. This comprehensive analysis will help researchers to make decisions when deducing the ABLH using multi-wavelength lidar data. The remainder of this work proceeds as follows: Section 2 describes the datasets used. The methodology is detailed in Section 3, followed by a comprehensive analysis of the

algorithm's performance and discussions in Sections 4 and 5, respectively. In Section 6, the key findings are summarized.

2. Study Area and Data

2.1. Lidar Data

To comprehensively test the performance of the cluster analysis approach, lidar observation data and a corresponding reference value are required. The experimental data in this study were collected using a two-wavelength polarization lidar system. The ground-based lidar system is located at Wuhan City Hubei Province, China [6,28]. The laser emitting system is a Nd:YAG (Quantel CFR) laser, which emits laser wavelengths of 355 nm and 532 nm. The aerosol BCs at 355 nm and 532 nm, and the DR information at 532 nm can be detected by this lidar system. The BC_{355} and BC_{532} are calculated by using the Fernald method. The lidar ratios at 532 nm and 355 nm, i.e., the ratio of the extinction to the backscattering coefficient, are assumed to be a constant over the Wuhan site (50 sr). The aerosol CR is defined as the ratio of the BCs at 532 nm to 355 nm, and the DR is defined as the ratio of the BC at the 532 nm perpendicular polarized channel to the 532 nm parallel polarized channel. Moreover, the temporal and vertical resolutions of the lidar observation data are 1 s and 7.5 m, respectively. The overlap of the lidar system is 150 m. Further instrumental details on the parameters can be found in previous studies [33,34]. In total, 132 days of experimental data were collected from June 2015 to June 2016, and the lidar data were averaged into hourly profiles. In addition, the observation data for cloud and dust were removed based on an extinction coefficient $> 2 \text{ km}^{-1}$ and the magnitude of the depolarization ratio [35]. The lidar data were matched with the RS data. After the screening and matching process, a total of 52 hourly nocturnal profiles remained.

2.2. RS Data

To evaluate the performance of different classifiers, the retrieved ABLHs and RS-determined ABLHs were compared for each classifier. The RS data used in this study were derived from launches at Wuhan, at 20:00 local time (LT) [36]. The RS data were obtained from the Bureau of Meteorology (Wuhan site) located at $30.37^\circ \text{ N } 114.08^\circ \text{ E}$, which is 20 km northwest of the lidar site. The Richardson number method was used to calculate the ABLH. The lowest level at which the interpolated Ri crosses the critical value of 0.25 is judged as the ABLH [17]. Note that if the Richardson number method failed to detect an ABLH, the ABLH was labeled as "not identified". In addition, it is worth noting that clouds and dust are also forms of aerosols, and cloud or dust boundaries will be misclassified as ABLH. After removing these cases, a total of 52 sets of matching data were obtained.

3. Methodology

The cluster analysis approach based on lidar data was first proposed by Toledo et al. [32]. Subsequently, Liu et al. [6] applied the approach to multi-wavelength lidar data to retrieve ABLH. The precondition of the clustering algorithm is that the lidar signal can characterize the difference in the characteristics of particles inside and outside the ABL. The clustering algorithm classifies the particles in the vertical direction to obtain the atmospheric aerosol category within the ABL and the free atmosphere category above the ABL; then, it defines the junction of the category as the ABLH. The principle of the cluster analysis approach is mainly based on the fact that aerosols in the ABL are distinctly different than those above the ABL. The ABLH is retrieved based on the difference between the aerosol load in the ABL and above the ABL. A case study is presented in Figure 1. The vertical structure of the atmosphere comprises an ABL and a free atmosphere. The free atmosphere has a small amount of aerosol particles, but the ABL contains a large amount of aerosol particles (Figure 1a). Liu et al. [34] pointed out that the BCs (532 nm and 355 nm), DR, and color ratio typically vary between the ABL and above-ABL air. For each observation, the profiles of the atmosphere BC and the CR can be obtained; then, it is possible to form a set of two-dimensional (2D) characteristic signals (Figure 1b). As the CR and BC of atmospheric

aerosols and clean particles are different, the atmospheric particles at different heights are concentrated in different regions of the 2D characteristic signal distribution. Therefore, the sample points in the ABL and the sample points in the free atmosphere can be distinguished by clustering (Figure 1c). Then, their junction can be identified as the ABLH. The ABLH result retrieved by lidar is also compared with RS measurements. Figure 1d shows the ABLH result at 20:00 LT.

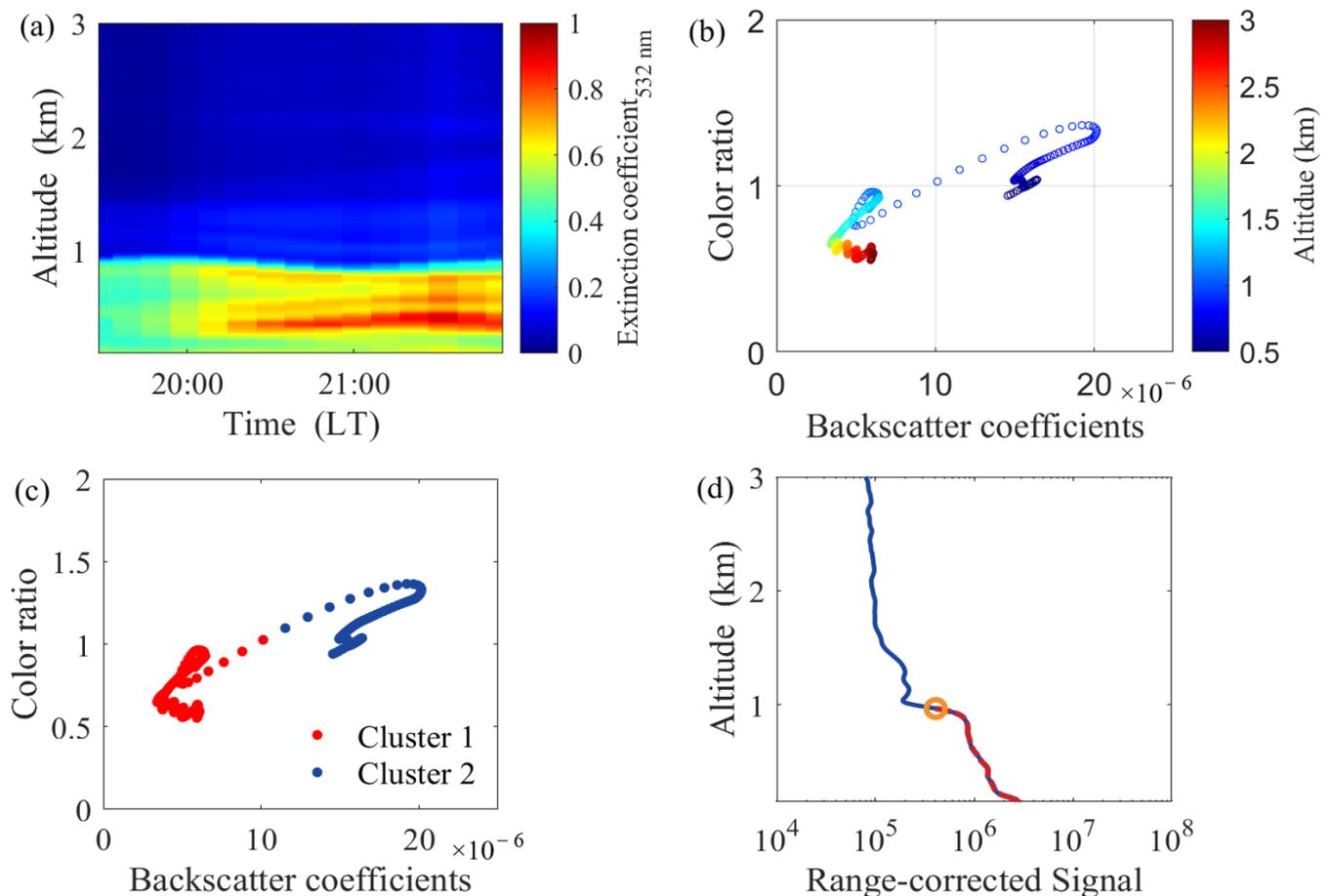


Figure 1. Case study of cluster analysis approach on 14 October 2015: (a) Time-height cross-section of extinction coefficients; (b) scatter plot of the color ratio vs. the backscatter coefficients (BCs) at 532 nm; (c) classification result; (d) the ABLH result at 20:00 (LT).

When using the clustering algorithm, there are two points to note: the selection of the input characteristic signal and the selection of the classifier. The flow chart of the cluster analysis approach is shown in Figure 2. First, regarding characteristic signal selection, the dual-wavelength polarized lidar system, for example, can provide BCs of 355 nm and 532 nm, as well as the aerosol CR and DR. The BC represents the backscatter intensity of particles, the CR represents the size of the particles, and the DR indicates the shape of the particles. Therefore, these four types of signals can form six sets of input signals: type-1 (CR and BC_{532}), type-2 (CR and BC_{355}), type-3 (CR and DR), type-4 (DR and BC_{532}), type-5 (DR and BC_{355}), and type-6 (BC_{355} and BC_{532}). As seen in Figure 3, the 2D distributions of each set of signals are different. This indicates that the separability of each signal type is also different. Therefore, careful selection of the input characteristic signal is very important.

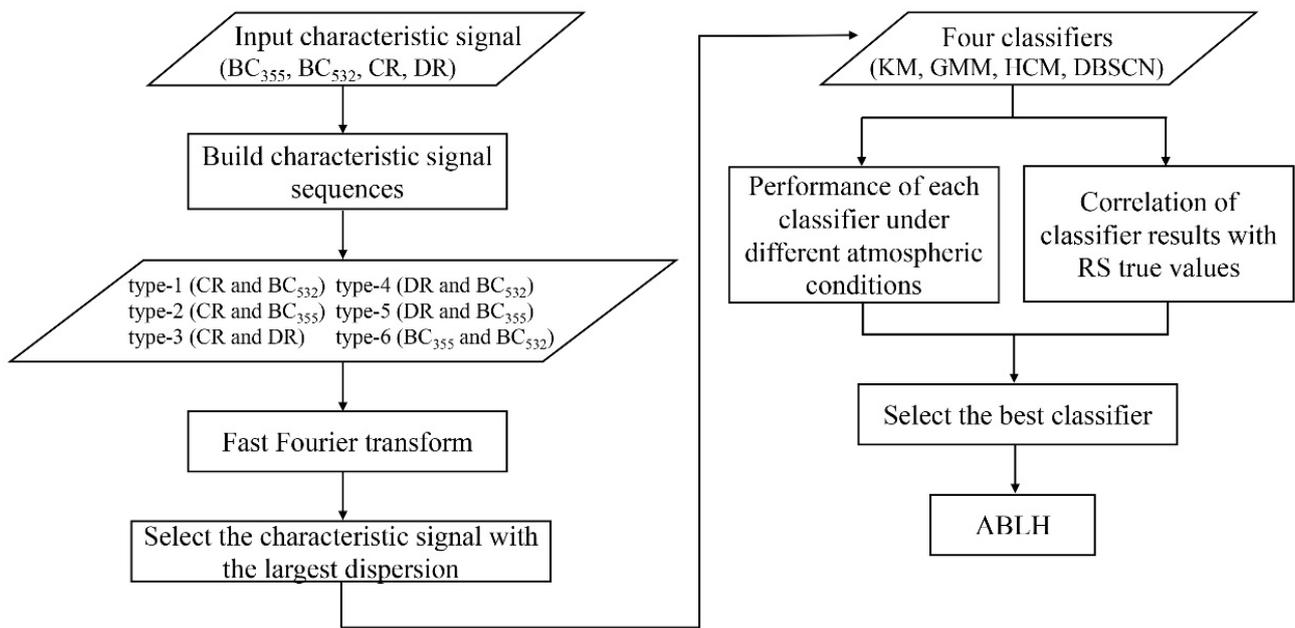


Figure 2. Flow chart of the cluster analysis approach.

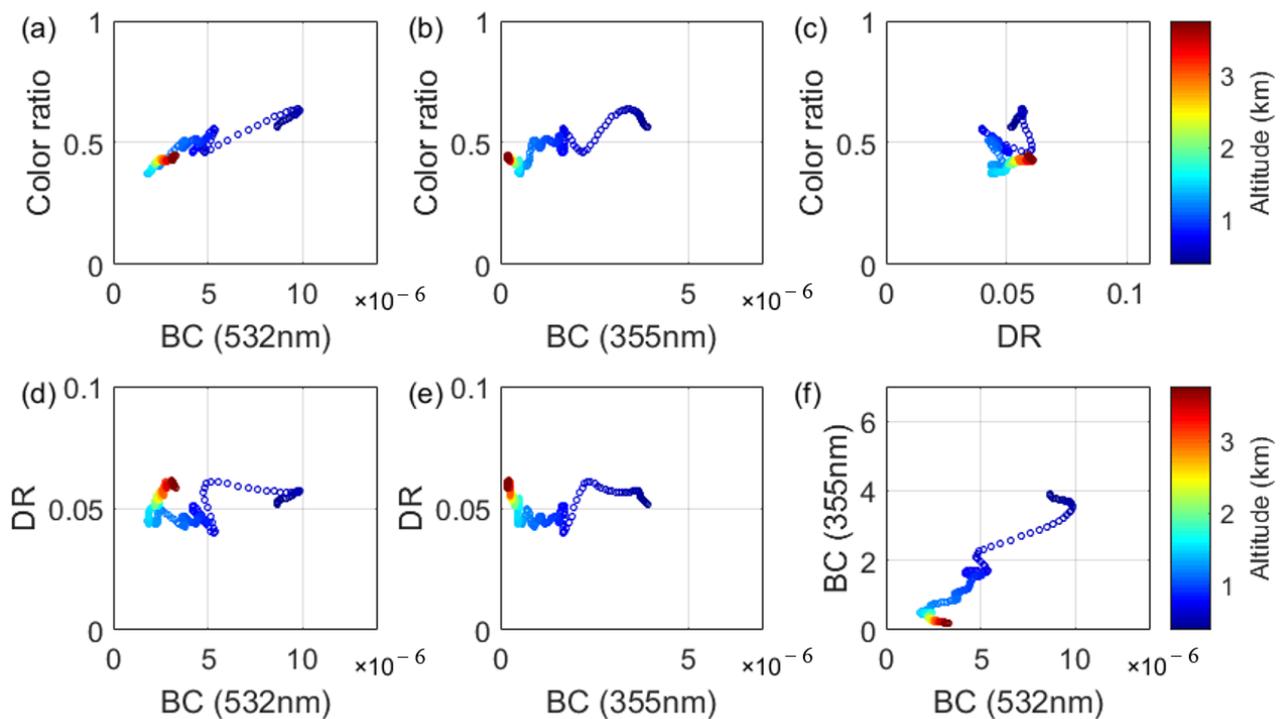


Figure 3. Lidar observation data at 20:00 (LT) on 7 December 2015, two-dimensional distribution of: (a) Type-1 (CR and BC532); (b) type-2 (CR and BC355); (c) type-3 (CR and DR); (d) type-4 (DR and BC532); (e) type-5 (DR and BC355); (f) type-6 (BC355 and BC532) signals.

Another point is the selection of the classifier. Common classifiers mainly include the KM, GMM, HCM, and DBSCAN [37–41]. The principle of the KM classifier is to classify the sample sequence by the distance between sample points; the adjacent sample points are divided into one category. The effect of the KM classifier is that the sample points in the category are close, while those between the categories are far away [34,38]. The GMM classifier refers to a method based on a probabilistic model. According to the probability

distribution of each sample point, the sample points with the same probability distribution are grouped into one category [38]. The HCM classifier is a connected algorithm that assumes each sample point is a category at the beginning, and then, the same category is sought based on linkage, after which the required categories are finally formed [41]. The DBSCAN classifier is a density-based clustering method that is designed to solve the clustering of irregular shapes. It defines a cluster as the largest set of points connected by density, and it is able to divide a region with a sufficiently high density into clusters. This method also works well for noisy data. However, it is very sensitive to the setting of the search radius and density parameters [38]. The KM and GMM classifiers solve the clustering problem of a simple distribution; however, when the data distribution is complicated, the classification results of the KM and GMM classifiers are much lower than those of HCM and DBSCAN classifiers. As the performance of the classifier is targeted, it is necessary to select the most appropriate classifier for the feature signal classification. In general, when using the clustering method to invert the ABLH classifier, it is necessary to select the appropriate characteristic signal. At the same time, the performance of the classifier should be evaluated. We discuss how to select the characteristic signal and compare the performance of the classifiers in Section 4.

4. Results

4.1. Determining the Characteristic Signal

In this section, we investigate the distribution of the six signal types to determine which type is best suited as the characteristic signal. From the physical principle, the difference in the characteristics of aerosol particles in the ABL and molecules outside the ABL will affect the dispersion of the characteristic signals. In other words, the more discrete the feature signal is, the easier it is for the clustering algorithm to classify it. The selection criterion for the characteristic signal is that the input 2D signal facilitates subsequent sample classification. More specifically, for each type of signal, the more discrete the 2D signal, the better the sample classification [6,28]. Therefore, we introduce fast Fourier transform (FFT) to calculate the degree of discreteness of feature signals. After FFT, the number of the step points of the signal in the frequency domain represents the degree of dispersion of the signal. The more step points, the more discrete the feature signal is, making it easier for the clustering algorithm to classify it. FFT can be performed on each type of signal, and the dispersion degree of each type can be estimated by calculating the number of step points in the frequency domain for each signal type. Finally, the 2D signal with the largest dispersion is selected as the characteristic signal.

Figure 4 shows the frequency distributions of the six signal types after Fourier transform at five different times, and the statistical average number of singularities for each type. The six types of signals (types 1–6) are represented by the gray, blue, green, black, orange, and red colors, respectively. After the Fourier transform, the distribution of each signal type in the frequency domain can be obtained. The sample points represent the step points of 2D signals, and the number of sample points indicates the degree of the signal dispersion. According to this distribution, when the number of step points in the frequency domain is more, the corresponding signal types are more discrete in the time domain, which is more favorable for classification. From the case studies (Figure 4a–e), the number of step points in the type-1 (gray points), type-4 (black points), and type-6 (red points) signals are more than those in the other types. This indicates that the type-1, type-4, and type-6 signals are more suitable as the characteristic signal. To determine which signal type is best suited as the characteristic signal, we perform a Fourier transform on all the signals, and count the average number of step points for each signal type. The average number of step points for the six types of signals are 2.2, 1.6, 1, 2.2, 1.6, and 3.2, respectively (Figure 4f). The type-6 signal has the most step points, indicating that it is the most discrete signal. This result shows that the type-6 signal is most suitable as the characteristic signal.

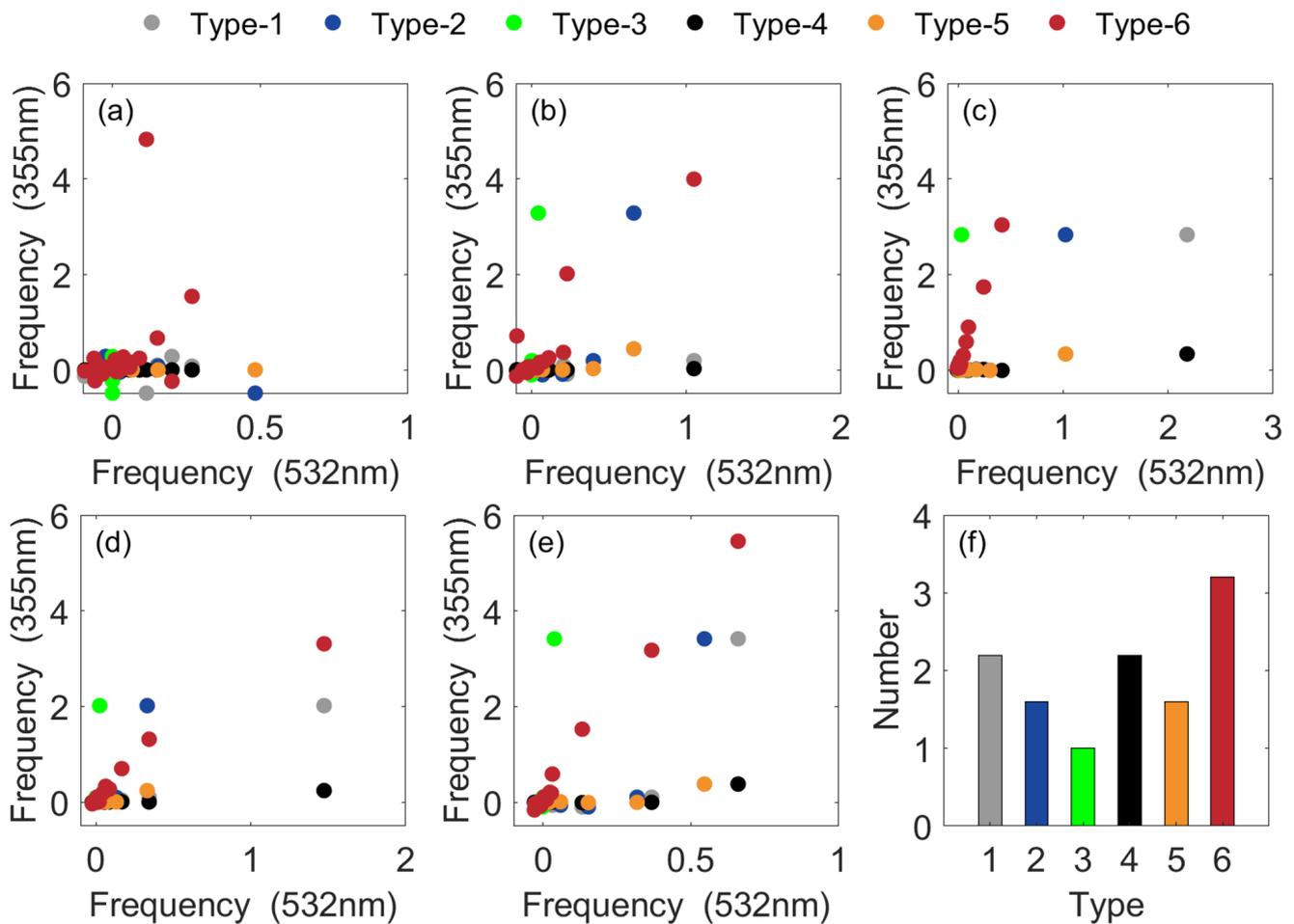


Figure 4. Frequency domain distributions of the six types of signals at 20:00 LT on: (a) 10 September 2015; (b) 14 October 2015; (c) 25 November 2015; (d) 7 December 2015; (e) 25 February 2016; and (f) the average number of step points for the six types of signals. The gray, blue, green, black, orange, and red colors represent the type 1–6 signals, respectively.

4.2. Case Study

In this section, after determining the input characteristic signal, three different states of atmospheric stability are selected to illustrate the performance of different classifiers under different atmospheric conditions. Then, the classification results of the different classifiers are calculated and compared with the RS-determined ABLH to evaluate the performance of the classifier.

Figure 5 shows the clustering results for the different classifiers on 14 October 2015. The ABLH is marked by the orange line, which is approximately 1 km above the ground. Under this state, the ABLH is a clear boundary between the ABL and the free atmosphere (Figure 5a), because the aerosol loading in the ABL is much larger than that above ABL. Consequently, the BCs at 532 nm or 355 nm wavelength below 1 km are greater than those above 1 km. The temperature profile observed by the RS at 20:00 LT is shown in Figure 5b. It confirms that the ABLH is approximately 1 km above the ground. The clustering results for the KM, GMM, HCM, and DBSCAN classifiers can be seen in Figure 5c–f. It can be found that the ABLH results from different methods are similar to each other. These results indicate that the performances of the four classifiers are similar under this state. Li et al. [42] also pointed out that the retrieved ABLH results of various methods were similar when the aerosols were uniformly distributed below the boundary layer.

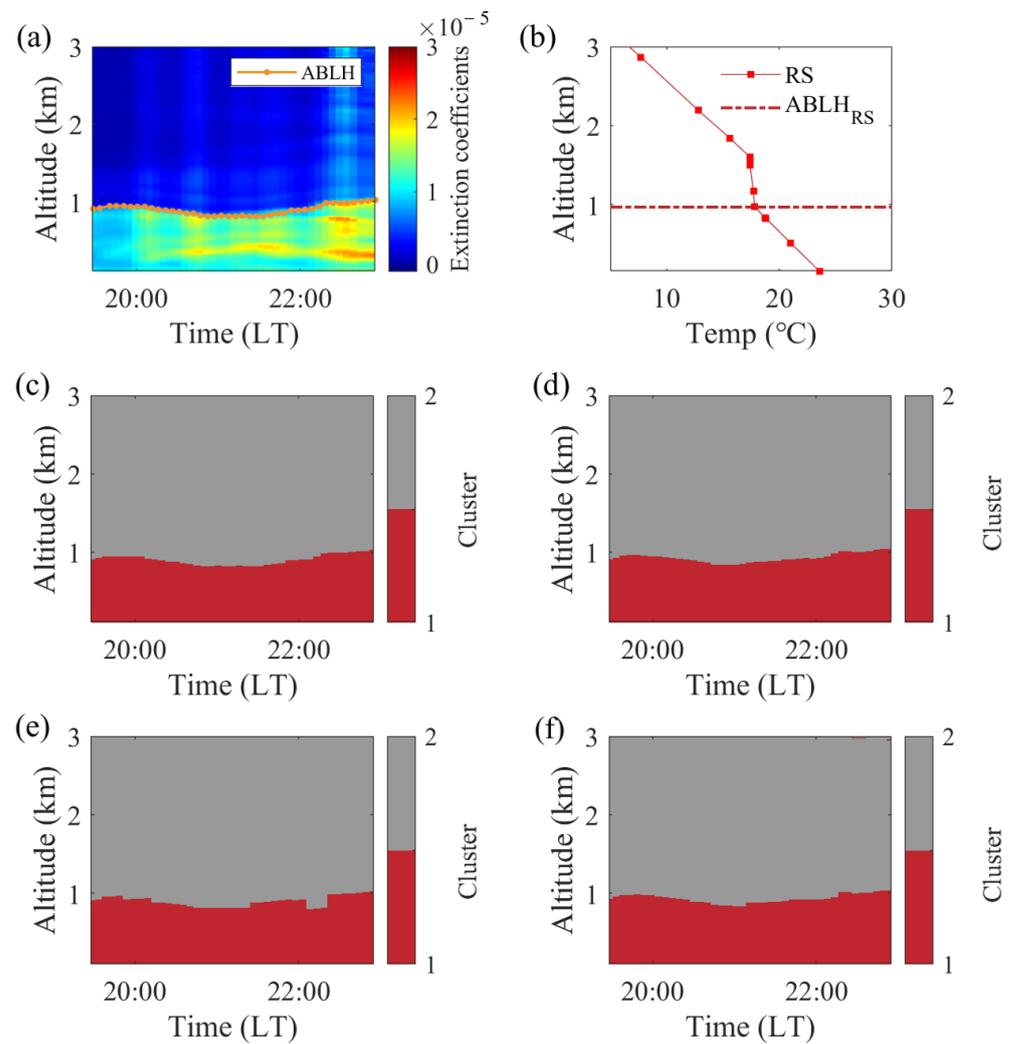


Figure 5. Case study on 14 October 2015: (a) Time-height cross-section of extinction coefficient; (b) temperature profile from RS observation at 20:00 LT. Clustering results for the: (c) KM; (d) GMM; (e) HCM; (f) DBSCAN classifiers.

The clustering results for the different classifiers on 23 October 2015 are shown in Figure 6. The black and orange lines represent ABLH and RLH, respectively. The atmospheric vertical structure is divided into the stable boundary layer, residual layer, and free atmosphere under this condition. In Figure 6a, a large number of aerosols are concentrated below 0.4 km, forming a stable boundary layer structure. Moreover, some thin aerosol layers are suspended at 0.4–1 km. This implies that the thin aerosol layer suspended in the ABL would affect the retrieval of the stable ABLH. The RS observation indicates that the ABLH is approximately 0.4 km at 20:00 LT (Figure 6b). Due to the relatively stable surface heat transfer at nighttime, the ABL has not evolved, and the ABLH should remain at 0.4 km during the nighttime. According to the distribution of the characteristic signal, the aerosols in the residual layer cause a large step change in the characteristic signal. This change may affect the accuracy of the classification results. As seen in Figure 6c,d, the KM and GMM classifiers are affected by the residual layer, and they misjudge the top of the residual layer as the stable ABLH. In contrast, the HCM and DBSCAN classifiers can overcome the effect of the residual layer, and can identify the stable ABLH. These results indicate that the performance of the HCM and DBSCAN classifiers are better than those of the KM and GMM classifiers under the residual layer state.

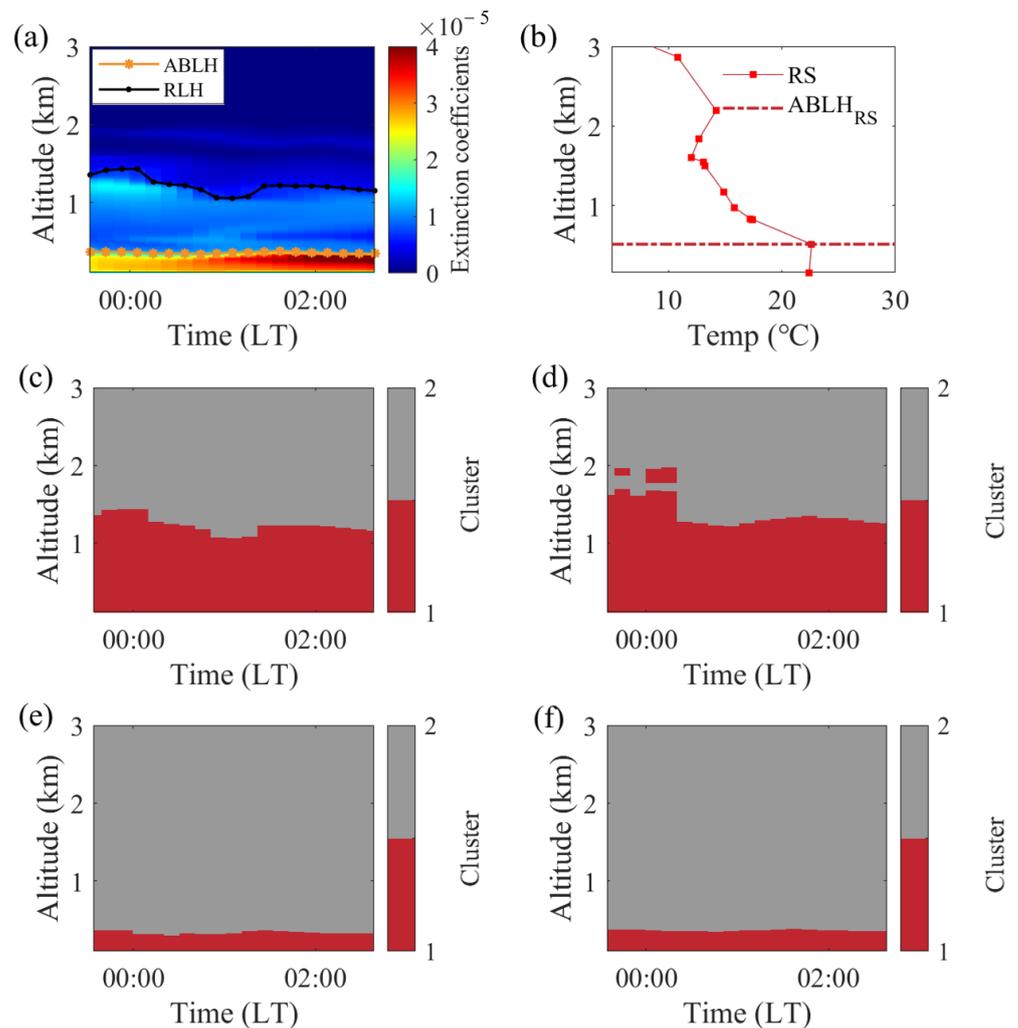


Figure 6. Case study on 23 October 2015: (a) Time-height cross-section of extinction coefficient; (b) temperature profile from RS observation at 20:00 LT. Clustering results for the: (c) KM; (d) GMM; (e) HCM; (f) DBSCAN classifiers.

Figure 7 shows the clustering results for the different classifiers on 2 November 2015. The black and orange lines represent ABLH and RLH, respectively. Evolution of the ABL can be observed under this state. The variations of surface temperature and surface net radiation are shown in Figure S1. The sunset time is 17:30 LT at this case. The surface temperature gradually decreases from 18:00 to 20:00 LT. During this period, the boundary layer remains at a height of approximately 1 km above the ground. The RS observation at 20:00 LT also indicates that the ABLH is approximately 1 km (Figure 7b). This result is consistent with the lidar observation. After 20:00 LT, the cooling of the surface temperature is completed, forming a nocturnal stable boundary layer. Due to the weakening of atmospheric convection, some aerosols cannot mix into the stable boundary layer. These aerosols are suspended above the ABL to form a residual layer [27,35]. The RLH remain to be at 1 km above the ground, which the ABLH is approximately 0.3 km above the ground. Similarly, the clustering results indicate that the performances of the four classifiers are similar before sunset (Figure 7c–f). In contrast, the performances of the HCM and DBSCAN classifiers are better than those of the KM and GMM classifiers after 20:00 LT. The KM and GMM classifiers wrongly judge RLH as BLH from 20:00 to 03:00 LT. Moreover, the ABLH result of the HCM classifier has a better change trend than that of the DBSCAN classifier.

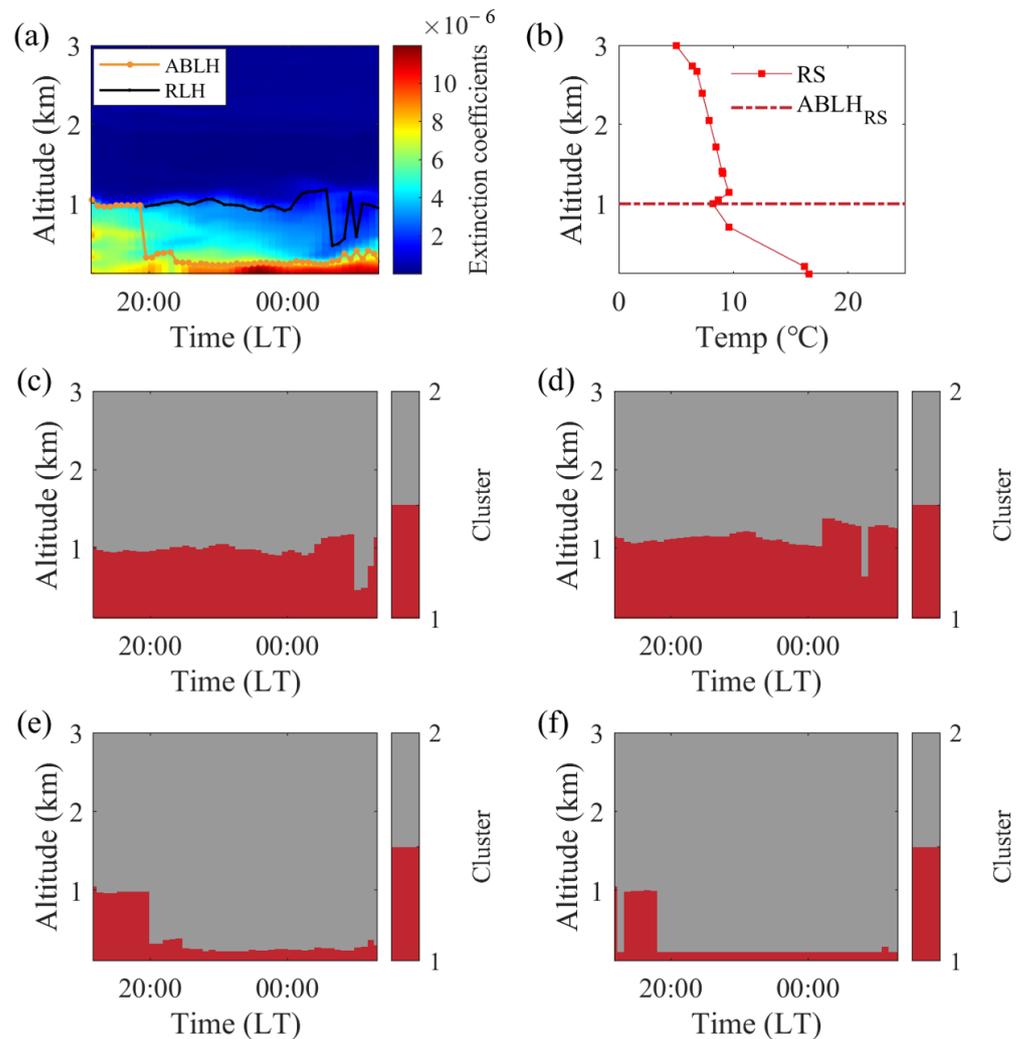


Figure 7. Case study on 2 November 2015: (a) Time-height cross-section of extinction coefficient; (b) temperature profile from RS observation at 20:00 LT. Clustering results for the: (c) KM; (d) GMM; (e) HCM; (f) DBSCAN classifiers.

4.3. Classifier Performance Evaluation

To test the performance of the four classifiers, the ABLHs retrieved by the different classifiers and the RS measurements are compared. Note that the total number of samples is 52 after the screening and matching process. Figure 8 shows the comparison of the results obtained using the different classifiers and that obtained from the RS measurements. Crosses, asterisks, triangles, and circle dots represent sample points at March–April–May (MAM), June–July–August (JJA), September–October–November (SON), and December–January–February (DJF), respectively. An asterisk indicates that the R passed the statistical significance difference test ($p < 0.05$). The correlation coefficients between the RS-determined ABLHs and the results of the four classifiers (KM, GMM, HCM, and DBSCAN) are 0.12, 0.14, 0.84, and 0.75, respectively. For K-means and GMM classifiers, some of ABLH results are higher than the ABLH results by RS. It is affected by the residual layers. The RLH is misjudged as ABLH by the K-means and GMM classifiers, leading to high results. This is because the classifiers based on distance or probability distribution divide the residual layer and the stable layer into one category when the particle characteristics in the different layers are similar. This leads to overestimation of the ABLH results. For the HCM and DBSCAN classifiers, the correlation coefficients are high, and the correlation coefficient of the HCM classifier is larger than that of the DBSCAN classifier. In addition, it is worth noting that the number of samples in the DBSCAN classifier is less. The reason is

that the DBSCAN classifier is too sensitive to the settings of the search radius and density parameters [43–48]. It is impossible to obtain the results for all observation data when using fixed initial parameters. The ABLH cannot be identified when the DBSCAN classifier cannot classify the characteristic signal. The results show that the ABLH retrieved using the HCM classifier is more consistent with that from RS measurements.

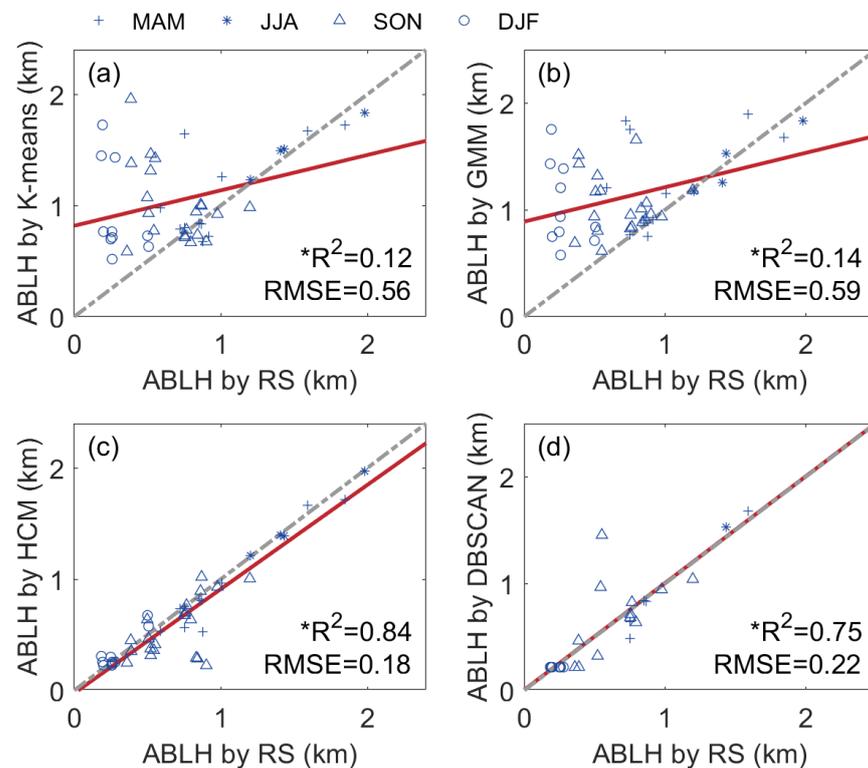


Figure 8. Correlation plots between the ABLHs retrieved by using (a) K-means, (b) GMM, (c) HCM, (d) DBSCAN and the RS-determined ABLHs. The grey and red lines represent the 1:1 reference lines and the regression lines, respectively. The crosses, asterisks, triangles, and circle dots represent sample points at spring (MAM), summer (JJA), autumn (SON), and winter (DJF), respectively. Each point was calculated from the hourly average profile. A star indicates the data with significant trends (p value < 0.05).

5. Discussion

In general, lidar-retrieved ABLH is actually to find the gradient change at the top of the ABL [13,28]. The principle of the cluster analysis approach is to select characteristic signals that can characterize the gradient changes at the top of the ABL, and then, retrieve the ABLH through a classifier. Here, first, we propose a method for selecting the input characteristic signal. The multi-wavelength lidar can provide six sets of input signals. By performing fast Fourier transform on each signal type, the number of step points in the frequency domain of each type is calculated to determine which signal is most suitable as the input characteristic signal. We perform Fourier transform on all the signals and count the average number of step points for each signal type. The results indicate that the type-6 signal composed of BC₃₅₅ and BC₅₃₂ is most suitable as the input characteristic signal. This is because the aerosols are mainly concentrated within the ABL. Backscatter signals can effectively characterize the differences in aerosol scattering intensities inside and outside the ABL. In contrast, the CR represents particle size and the DR represents particle shape, which have no obvious distribution pattern in the vertical direction, and are not recommended as characteristic signal. Next, we compare the performances of the four classifiers; the results indicate that the HCM classifier is the most appropriate classifier for the characteristic signal classification, because the HCM classifier performs a breakpoint

detection according to the linkage between sample points. This ensures that once a step point is detected in the characteristic signal, a category result is given. Under this condition, as long as there is a difference between the residual layer and the stable layer, the stable layer can be accurately separated. Therefore, the HCM classifier is recommended for investigating ABL variation at nighttime as it can effectively solve the effects of the residual layer. In addition, the HCM classifier is not suitable for cloud and dust days. The dust and cloud layers have strong scattering characteristics, and HCM classifiers may misjudge the tops of the cloud (dust) layer as ABLH. Some studies apply the height limitation or graph theory to avoid the effect of the cloud (dust) layer [6,31]. However, when the top of the ABL is mixed with the cloud (dust) layer, the ABLH retrieval based on lidar data needs further research.

6. Conclusions

In this study, we investigate, in detail, the application of a cluster analysis approach to ABL research. The cluster analysis approach is applied to 52 days of lidar measurements in Wuhan for the period from June 2015 to June 2016. The process of ABLH retrieval using a clustering algorithm can be divided into two parts: selection of the input characteristic signal and selection of the classifier. First, regarding selection of the characteristic signal, the dual-wavelength polarization lidar, for example, performs Fourier transform on all the observed signals. According to the dispersion degree of the signal in the frequency domain, the best characteristic signal can be determined, which comprises BCs at 355 nm and 532 nm. Similarly, other researchers can use this method to filter characteristic signals when a multifunction lidar system is used to obtain the ABLH. Then, for the selection of the classifier, the performances of the four typical classifiers are compared under different atmospheric states. The results indicate that the HCM classifier is more suitable under the well-mixed and residual layer conditions. Therefore, the HCM classifier is recommended for classifying characteristic signals when clustering algorithms are used for nocturnal ABL research.

In this study, we attempt to utilize a cluster analysis approach to estimate nocturnal ABLH from multi-wavelength lidar data. The advantage of this method is that it can rely on the lidar data to stably invert the ABLH, while reducing the effects of the residual layer. A limitation of this method is that it does not work well when those differences in aerosol characteristics between the ABL and above-ABL air are not present. Meanwhile, it is useless under special situations such as low cloud and dust cases. Because the clouds and dust are also forms of aerosols, cloud or dust boundaries will be misclassified as ABLH. In addition, due to the number limitations of observational data, we were unable to evaluate the performance of the algorithm during the day and different seasons. It will be further studied in the future. This study offers insight into ABL remote sensing technology.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/atmos14050847/s1>. Figure S1: The variations of surface temperature and surface net radiation from ERA5 at 2 November 2015. Sample dataset, demo code and readme file.

Author Contributions: Conceptualization, Z.Z. and H.L.; methodology, X.Z.; software, H.L.; validation, Z.Z., S.F., W.X. and W.G.; formal analysis, Z.Z.; resources, X.Z. and W.G.; data curation, H.L.; writing—original draft preparation, Z.Z. and H.L.; writing—review and editing, X.Z.; visualization, Z.Z. and H.L.; supervision, S.F.; project administration, W.X. and W.G.; funding acquisition, Z.Z.; investigation, H.L., X.Z., S.F., W.X. and W.G. 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 under grant 42071353.

Data Availability Statement: The lidar data at Wuhan station given in this paper are available upon request via email: zhongmin.zhu@whu.edu.cn.

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

References

1. Stull, R.B. *An Introduction to Boundary Layer Meteorology*; Springer: Dordrecht, The Netherlands, 1988.
2. Seibert, P.; Beyrich, F.; Gryning, S.E.; Joffre, S.; Rasmussen, A.; Tercier, P. Review and intercomparison of operational methods for the determination of the mixing height. *Atmos. Environ.* **2000**, *34*, 1001–1027. [[CrossRef](#)]
3. Yang, T.; Wang, Z.; Zhang, W.; Gbaguidi, A.; Sugimoto, N.; Wang, X.; Matsui, I.; Sun, Y. Technical note: Boundary layer height determination from lidar for improving air pollution episode modeling: Development of new algorithm and evaluation. *Atmos. Chem. Phys.* **2017**, *17*, 6125–6225. [[CrossRef](#)]
4. Ji, X.; Liu, C.; Xie, Z.; Hu, Q.; Dong, Y.; Fan, G.; Zhang, T.; Xing, C.; Wang, Z.; Javed, Z.; et al. Comparison of mixing layer height inversion algorithms using lidar and a pollution case study in Baoding, China. *J. Environ. Sci.* **2018**, *79*, 81–90. [[CrossRef](#)]
5. Liu, J.; Huang, J.; Chen, B.; Zhou, T.; Yan, H.; Jin, H.; Huang, Z.; Zhang, B. Comparisons of PBL heights derived from CALIPSO and ECMWF reanalysis data over China. *J. Quant. Spectrosc. Radiat. Transf.* **2015**, *153*, 102–112. [[CrossRef](#)]
6. Liu, B.; Ma, Y.; Gong, W.; Zhang, M.; Yang, J. Improved Two-wavelength Lidar algorithm for Retrieving Atmospheric Boundary Layer Height. *J. Quant. Spectrosc. Radiat. Transf.* **2019**, *224*, 55–61. [[CrossRef](#)]
7. Li, H.; Yang, Y.; Hu, X.; Huang, Z.; Wang, G.; Zhang, B.; Zhang, T. Evaluation of retrieval methods of daytime convective boundary layer height based on lidar data. *J. Geophys. Res. Atmos.* **2017**, *122*, 4578–4593. [[CrossRef](#)]
8. Davis, K.J.; Gamage, N.; Hagelberg, C.R.; Kiemle, C.; Lenschow, D.H.; Sullivan, P.P. An objective method for deriving atmospheric structure from airborne lidar observations. *J. Atmos. Ocean. Technol.* **2000**, *17*, 1455–1468.
9. Hägeli, P.; Steyn, D.G.; Strawbridge, K.B. Strawbridge. Spatial and temporal variability of mixed layer depth and entrainment zone thickness. *Bound. Layer Meteorol.* **2000**, *97*, 47–71. [[CrossRef](#)]
10. Bianco, L.; Wilczak, J.M. Convective Boundary Layer Depth: Improved Measurement by Doppler Radar Wind Profiler Using Fuzzy Logic Methods. *J. Atmos. Ocean. Technol.* **2002**, *19*, 1745–1758. [[CrossRef](#)]
11. Campbell, J.R.; Hlavka, D.L.; Welton, E.J.; Flynn, C.J.; Turner, D.D.; Spinhirne, J.D. 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](#)]
12. Emeis, S.; Schafer, K.; Munkel, C. Surface-based remote sensing of the mixing layer height a review. *Meteorol. Z.* **2008**, *17*, 621–630. [[CrossRef](#)]
13. Su, T.; Li, J.; Li, C.; Xiang, P.; Lau AK, H.; Guo, J.; Yang, Y.; 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. Atmos.* **2017**, *122*, 3929–3943. [[CrossRef](#)]
14. Gregori, D.A.M.; Luis, G.R.J.; Ntonio, J.B.A.A.; Antonio, J.B.O.; Pablo, O.A.; Roberto, R.; Esteban AB, V.; Landulfo, E.; Lucas, A.A. Study of the planetary boundary layer by microwave radiometer, elastic lidar and Doppler lidar estimations in Southern Iberian Peninsula. *Atmos. Res.* **2018**, *213*, 185–195.
15. Liu, S.; Liang, X. Observed diurnal cycle climatology of planetary boundary layer height. *J. Clim.* **2010**, *23*, 5790–5809. [[CrossRef](#)]
16. 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*, D16113. [[CrossRef](#)]
17. Guo, J.; Miao, Y.; Zhang, Y.; Liu, H.; Li, Z.; Zhang, W.; He, J.; Luo, M.; Yan, Y.; Bian, L.; et al. The climatology of planetary boundary layer height in China derived from radiosonde and reanalysis data. *Atmos. Chem. Phys.* **2016**, *16*, 13309–13319. [[CrossRef](#)]
18. Huang, Z.; Huang, J.; Bi, J.; Wang, G.; Wang, W.; Fu, Q.; Li, Z.; Tsay, S.; Shi, J. Dust aerosol vertical structure measurements using three mpl lidars during 2008 China-U.S. joint dust field experiment. *J. Geophys. Res. Atmos.* **2010**, *115*, 1307–1314.
19. Liu, S.; He, W.; Liu, H.; Chen, H. Retrieval of Atmospheric Boundary Layer Height from Ground-based Microwave Radiometer Measurements. *J. Appl. Meteorol. Sci.* **2015**, *26*, 626–635.
20. Granados-Muñoz, M.J.; Navas-Guzmán, F.; Bravo-Aranda, J.A.; Guerrero-Rascado, J.L.; Lyamani, H.; Fernández-Gálvez, J.; Alados-Arboledas, L. Automatic determination of the planetary boundary layer height using lidar: One-year analysis over southeastern Spain. *J. Geophys. Res. Atmos.* **2012**, *117*, D18208. [[CrossRef](#)]
21. Luo, T.; Yuan, R.; Wang, Z. Lidar-based remote sensing of atmospheric boundary layer height over land and ocean. *Atmos. Meas. Tech.* **2014**, *7*, 173–182. [[CrossRef](#)]
22. Toledo, D.; Córdoba-Jabonero, C.; Adame, J.A.; Benito DL, M.; 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](#)]
23. Emeis, S.; Jahn, C.; Munkel, C.; Munsterer, C.; Schafer, K. Multiple atmospheric layering and mixing-layer height in the Inn valley observed by remote sensing. *Meteorol. Z.* **2007**, *16*, 415–424. [[CrossRef](#)]
24. Steyn, D.G.; Baldi, M.; Hoff, R.M. The detection of mixed layer depth and entrainment zone thickness from lidar backscatter profiles. *J. Atmos. Ocean. Technol.* **1999**, *16*, 953–959. [[CrossRef](#)]
25. Brooks, I.M. Finding boundary layer top: Application of a wavelet covariance transform to lidar backscatter profiles. *J. Atmos. Ocean. Technol.* **2003**, *20*, 1092–1105. [[CrossRef](#)]
26. Jordan, N.S.; Hoff, R.M.; Bacmeister, J.T. Validation of Goddard Earth Observing System-version 5 MERRA planetary boundary layer heights using CALIPSO. *J. Geophys. Res.* **2010**, *115*, D24218. [[CrossRef](#)]
27. Wang, Z.; Cao, X.; Zhang, L.; Notholt, J.; Zhou, B.; Liu, R. Lidar measurement of planetary boundary layer height and comparison with microwave profiling radiometer observation. *Atmos. Meas. Tech.* **2012**, *5*, 1965–1972. [[CrossRef](#)]

28. Liu, B.; Ma, Y.; Gong, W.; Yang, J.; Zhang, M. Two-wavelength Lidar inversion algorithm for determining planetary boundary layer height. *J. Quant. Spectrosc. Radiat. Transf.* **2018**, *206*, 117–124. [[CrossRef](#)]
29. Lange, D.; Tiana-Alsina, J.; Saeed, U.; Tomás, S.; Rocadenbosch, F. Atmospheric Boundary Layer Height Monitoring Using a Kalman Filter and Backscatter Lidar Returns. *IEEE Trans. Geosci. Remote Sens.* **2013**, *52*, 4717–4728. [[CrossRef](#)]
30. Pal, S.; Haeffelin, M.; Batchvarova, E. Exploring a geophysical process-based attribution technique for the determination of the atmospheric boundary layer depth using aerosol lidar and near-surface meteorological measurements. *J. Geophys. Res. Atmos.* **2013**, *118*, 9277–9295. [[CrossRef](#)]
31. Bruine, M.D.; Apituley, A.; Donovan, D.; Baltink, H.K.; Haij, M.D. Pathfinder: Applying graph theory for consistent tracking of daytime mixed layer height with backscatter lidar. *Atmos. Meas. Tech.* **2017**, *10*, 1893–1909. [[CrossRef](#)]
32. 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](#)]
33. Yang, S.; Yang, J.; Shi, S.; Song, S.; Luo, Y.; Du, L. The rising impact of urbanization-caused CO₂ emissions on terrestrial vegetation, Ecological Indicators. *Ecol. Indic.* **2023**, *148*, 110079. [[CrossRef](#)]
34. Liu, B.; Ma, Y.; Gong, W.; Zhang, M. Observations of aerosol color ratio and depolarization ratio over wuhan. *Atmos. Pollut. Res.* **2017**, *8*, 1113–1122. [[CrossRef](#)]
35. Liu, B.; Ma, Y.; Shi, Y.; Jin, S.; Jin, Y.; Gong, W. The characteristics and sources of the aerosols within the nocturnal residual layer over Wuhan, China. *Atmos. Res* **2020**, *241*, 104959. [[CrossRef](#)]
36. Li, H.; Liu, B.; Ma, X.; Ma, Y.; Jin, S.; Fan, R.; Wang, W.; Fang, J.; Zhao, Y.; Gong, W. The Influence of Temperature Inversion on the Vertical Distribution of Aerosols. *Remote Sens.* **2022**, *14*, 4428. [[CrossRef](#)]
37. Jain, A.K. Data clustering: 50 years beyond K-means. *Pattern Recognit. Lett.* **2010**, *31*, 651–666. [[CrossRef](#)]
38. Birant, D.; Kut, A. ST-DBSCAN: An algorithm for clustering spatial-temporal data. *Data Knowl. Eng.* **2007**, *60*, 208–221. [[CrossRef](#)]
39. Biernacki, C.; Celeux, G.; Govaert, G. Assessing a mixture model for clustering with the integrated completed likelihood. *IEEE Trans. Pattern Anal. Mach. Intell.* **2000**, *22*, 719–725. [[CrossRef](#)]
40. Hartigan, J.A.; Wong, M.A. Algorithm AS 136: A k-means clustering algorithm. *J. R. Stat. Society. Ser. C Appl. Stat.* **1979**, *28*, 100–108. [[CrossRef](#)]
41. Langfelder, P.; Zhang, B.; Horvath, S. Defining clusters from a hierarchical cluster tree: The Dynamic Tree Cut package for R. *Bioinformatics* **2007**, *24*, 719–720. [[CrossRef](#)]
42. Li, H.; Shi, R.; Jin, S.; Wang, W.; Fan, R.; Zhang, Y.; Liu, B.; Zhao, P.; Gong, W.; Zhao, Y. Study of Persistent Haze Pollution in Winter over Jinan (China) Based on Ground-Based and Satellite Observations. *Remote Sens.* **2021**, *13*, 4862. [[CrossRef](#)]
43. 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](#)]
44. Liu, B.; Ma, X.; Ma, Y.; Li, H.; Jin, S.; Fan, R.; Gong, W. The relationship between atmospheric boundary layer and temperature inversion layer and their aerosol capture capabilities. *Atmos. Res.* **2022**, *271*, 106121. [[CrossRef](#)]
45. Xu, W.; Wang, W.; Wang, N.; Chen, B. New Algorithm for Himawari-8 Aerosol Optical Depth Retrieval by Integrating Regional PM_{2.5} Concentrations. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 3155503.
46. Shi, T.; Han, G.; Ma, X.; Mao, H.; Chen, C.; Han, Z.; Gong, W. Quantifying factory-scale CO₂/CH₄ emission based on mobile measurements and EMISSION-PARTITION model: Cases in China. *Environ. Res. Lett.* **2023**, *18*, 034028. [[CrossRef](#)]
47. Liu, B.; Ma, X.; Guo, J.; Li, H.; Jin, S.; Ma, Y.; Gong, W. Estimating hub-height wind speed based on a machine learning algorithm: Implications for wind energy assessment. *Atmos. Chem. Phys.* **2023**, *23*, 3181–3193. [[CrossRef](#)]
48. Chen, B.; Tan, J.; Wang, W.; Dai, W.; Ao, M.; Chen, C. Tomographic Reconstruction of Water Vapor Density Fields from the Integration of GNSS Observations and Fengyun-4A Products. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 4100712. [[CrossRef](#)]

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