Responds to Anonymous Referee #1:

General comments:

The article presented a new approach combining gradient method and cluster analysis to distinguish multi-layers (i.e., the cloud layer, the elevated aerosol layer, and the noise layer) and therefore retrieving NBLH based on lidar data. More information about such layers can also be obtained by the K-mean cluster analysis. However, the writing of the article needs to be further improved. And some doubts about your work are as follows:

Response:

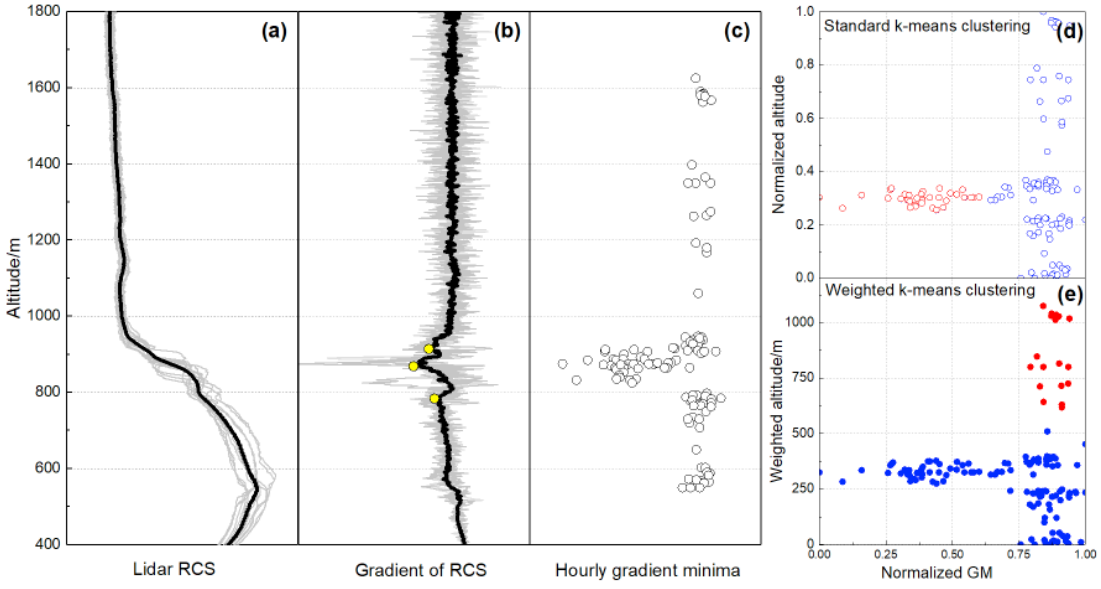
Thanks a lot for your reviews on our manuscript entitled "A novel Mie lidar gradient cluster analysis method of nocturnal boundary layer detection during air pollution episodes (ID: amt-2020-167). We have revised the manuscript according to your suggestion, the language has been polished by Elsevier Language Editing Services and mentioned references have been added. The details are shown as follows.

Specific comments:

1.Figure 2 should be described clearly. Is the red solid line the lidar signal profile averaged every 1h in figure 2(a)? And I'm confused about the weighted altitudes in figure 2(b), is hw equals h (the real height) minus hmin? If yes, the maximum of hw is obviously lower than 1000 m, why a point exist higher than 1000 m in your figure?

Response:

1) The description of the figure 2 has been added at P6.line 138-142.



'Figure 2. The theoretical schematic of the weighted-k means clustering. (a)The real profile of a lidar RCS(light gray line) and the hour averaged RCS (black line). (b) The gradient of RCS (light gray line), the hour averaged gradient RCS (black line), and the three minima in the profile (yellow points). (c) The distribution of the gradient minima within an hour. (d-e) The results obtained by standard k-means and weighted k-means clustering, where two clusters are differentiated, as shown by red and blue hollow and solid points, respectively.'

2) Yes, the weight in k-means clustering equals 1075 ($G = h_{max} - h_{min}$). We modified the scale of the y-axis range, and check that there are several points located at around 1600 m, which indicated the weighted k-means points larger than 1000 m.

Thank you for your suggestion. The figure 2 has been changed.

2. Line145: "a dataset of three gradient minima of RCS". Do your mean three gradient minima of RCS at every 50 s within 1h are chosen to have a k-means cluster analysis?

Response:

Yes, every profile of the RCS gradient is used to seek the three minima. Then, all the minima within an hour are used as the dataset of k-means classification.

The contents has been added to the article in P6 line 144.

'a dataset of three minima of RCS gradient within an hour works as the dataset of weighted k-means classification.'

3 From table 3, the altitude of NBL is always lower than that of EALs, Cloud, and Noise layers, so is there a simple top limiter works?

Response:

No, not exactly, the method contains height restriction on both upper limiter and lower limiter.

In this algorithm, it contain the top limiter conception in the first weighted k-means analysis, because the location of cloud layer and noise is above the NBL. Previous studies (Dang et al., 2019b; Li et al., 2017a) have successfully evaluate the works of the top-limiter.

However, in the second weighted k-means cluster processing. In order to classify the elevated aerosol layers (EALs) and NBL, we use the distance between two aerosol layers and the threshold of the backscatter coefficient as a sign to identifying the EALs and NBL. Here is an example of the height restriction on 18 Dec 2016 on 2：00-3：00 LST (Figure R1-1).

Through the first k-means clustering (Figure R1-1(a)), the noise is identified above 912.5 m. Next, in the second clustering analysis (Figure R1(b)), the upper groups are not meet the criteria for the EALs. After checking the standard

deviation of the normalized gradient method value between the centriod on each clusters ($S_{Ci'} > S_{Cj'}$), we found that the NBL is the cluster of Ci'. The lower and top limiter are shown in the Figure R1(b).

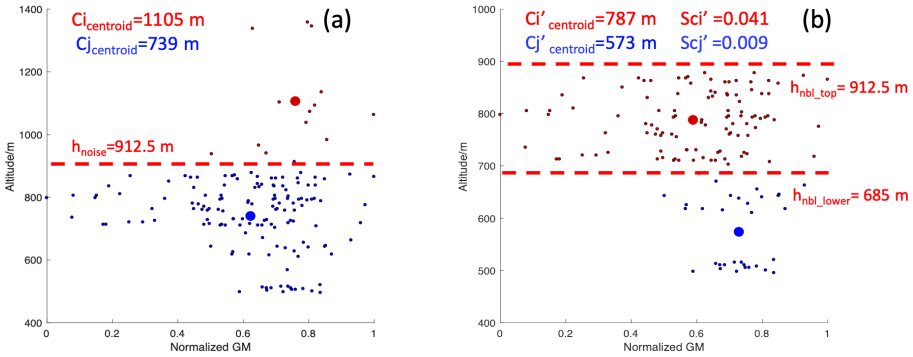


Figure R1-1. The two weighted k-means clustering on 18 Dec 2016 between 2：00-3：00 LST.

(a) The first weighted k-means clustering. The results are shown by red and blue hollow and solid points, and their the centroids are represented by larger points of the same colour. The red dash line is shown the height of the limitation of noise. (b) The second weighted k-means clustering. The $S_{Ci'}$ and $S_{Cj'}$ represents the standard deviation of the normalized gradient values in each cluster. The two dash red line is the height restriction of the cluster.

4 The word "starfield" appears many times in the article. Do you mean "stratified"? Please confirm it.

Response:

Thank you for your suggestion. Yes, the word has been changed as stratified.(Line 151 & 170& 274).

5. From figure 7-2 (c), the NBLH between 21:00 to 22:00 LST is about 640 m (hcjcenter). However, from Figure 7-1, the NBLH of that time period is much higher, why?

Response:

In our algorithm, we have defined two constraints to identifying noise as in figure 3 shown. The first is that the noise signal distribution is not clearly stratified ($D_{intra} < D_{sig}$), and the second is that the noise is located at a higher height($h_{ci} > h_{cj}$) and the average standard deviation of the points in noise cluster is smaller than the NBL($S_{ci} < S_{cj}$). At 21:00-22:00 on April 6, 2017, the distribution of the upper cluster is not meet the requirement of the standard deviation (S_{ci}=0.016, S_{cj}=0.033). Therefore, the NBL are in the cluster of the upper layer (cluster in blue).

The content has been changed in P13 Line 274-277.

6. In your CA-GM algorithm, the cluster number is set as two in prior, that is, except NBL, assume that there is only one layer exist above NBL. So what if two or more layers (EALs, cloud layer, or noise layer) exist above NBLH? Besides, I'm

concerned that if there is no EALs, cloud layer, or noise layer, does the cluster method affect the NBLH retrieval? Are the NBLHs from the CA-GM similar to that from the GM?

Response:

As for the effect two or three more layers, the following results are testing with the real signal.

1) As the example of the 00：00-01:00 Jan 6, 2017. This is a typical multiple layer structure of noise, cloud and NBL.

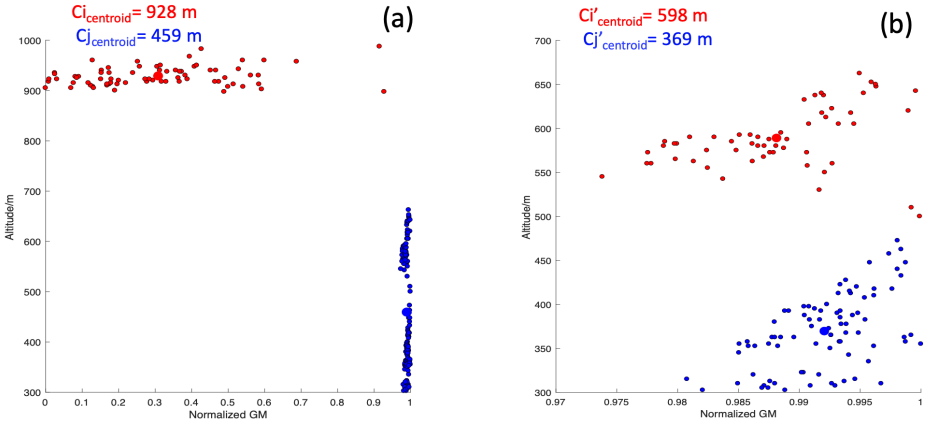


Figure R1-2.The example of the multi-layer structure of noise, cloud and NBL.
(a) the first weighted k-means clustering (b) the second weighted k-means clustering.

As a result of the first k-means clustering, seldom noise is located above the cloud layer (Figure R1-2(a)). We set the cluster as two in prior, there are two groups which indicated the cloud and the possible NBL. According to the criteria to distinguish the cloud layer, the CL and the noise are removed in the upper cluster. Then, we use the second weighted k-means clustering to further identify the NBL. Due to the standard deviation of the GM value in the upper layer is bigger than the lower cluster ($S_{Ci'} > S_{Cj'}$), the final NBL location is at the cluster whose centroid is 598 m.

2) As the example of the 19:00 Dec 19, 2016. There is a scenario of the structure of noise, EALs and NBL.

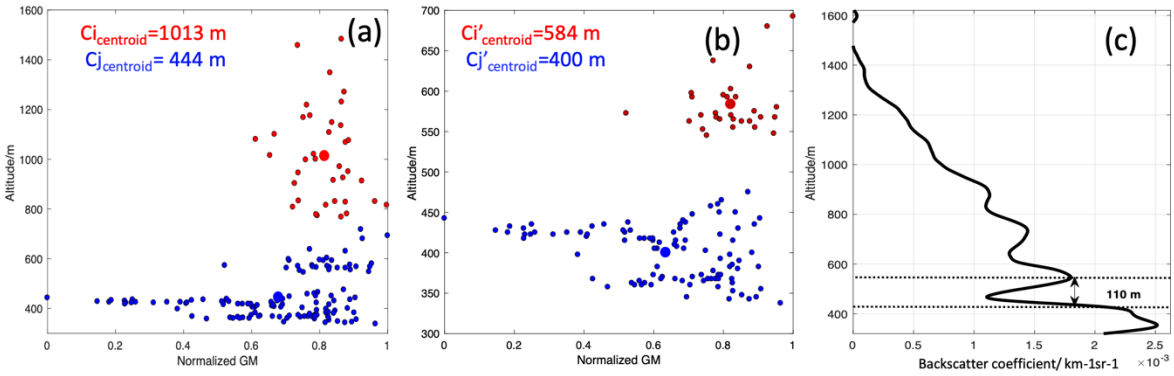


Figure R1-3.The example of the structure of noise, EALs and NBL.

(a) the first weighted k-means clustering. (b) the second weighted k-means clustering.

(c) the backscatter coefficient of the lidar. The distance of two layers are larger than D threshold=100 m.

As a result of the first k-means clustering, the noise is located at the cluster in red (Figure R1-3(a)). We set the cluster as two in prior, there are two groups which indicated the noise and the possible NBL. According to the criteria to distinguish the noise, noise is removed in the upper cluster. Then, we use the second weighted k-means clustering to further identify the NBL. Due to the backscatter coefficient is excess $1.786 \times 10 - 3km^{-1}sr^{-1}$. The cluster in red (Figure R1-3(b)) are to be defined as the EALs. Therefore, the NBL is located at the lower cluster (in blue).

3) As for the situation of the structure of noise, CLs, EALs and NBL.

There is no real case can be present in this experiment. It is relatively rare in the experiment of this scene.

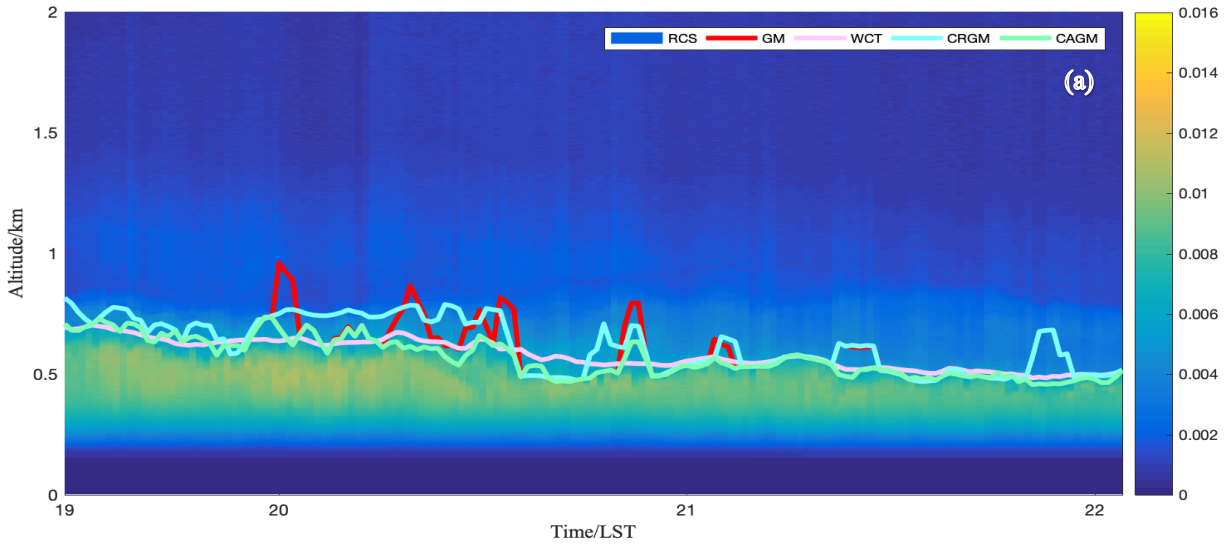
According to the performance of the algorithm, the three possible minima will be located as the CL, EALs ,and the NBL. Seldom will locate as noise affection. If the noise exists and locate over the cloud, the structure can be solved as Figure R1-2. The cloud with noise will be removed by the upper limiter. the four-layer structure has been transformed into a three-layer structure as the Figure R1-3.

As for the complex condition for four or more layers exist, it may cause a certain degree of misjudgement. The algorithm cloud be further developed in seeking for optimal k of the k-means clustering to suitable for more complex condition accurately.

Thank you for your idea.

4) As the situation in the clear condition, on 29 Dec 2016.

If there are no that mentioned three layers , in order to use the CA-GM, it still needs to find the minima in each profile. One point will be layered in obvious NBL, the other two will be removed as the noise cluster. As for the dataset of an hour k-means cluster, the extreme value will be move from the GM. As the shown in Figure R1-4 (a).



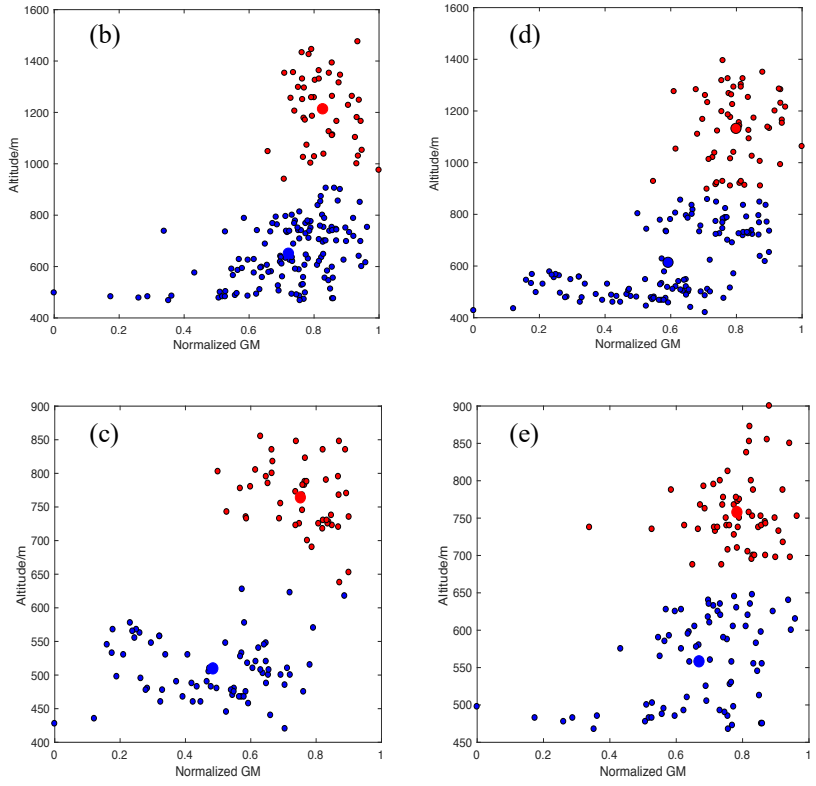


Figure R1-4.The example of the no clear structure of the lidar signal. (a). the time-height cross section of the range-corrected signal (RCS) with four NBLH retrieved method on 29 Dec, 2016. (b-c)The first and second weighted k-means clustering in the 20:00-21:00 LST. (d-e) The first and second weighted k-means clustering in the 21:00-22:00 LST.

7. Lines 40-53: the authors have clarified some BLH retrieval methods; however, more previous evaluation works should be cited here, for example: McGrath-Spangler et al., 2012, Li et al., 2017.

[1] McGrath-Spangler, E. L., and A. S. Denning (2012), Estimates of North American summertime planetary boundary layer depths derived from space-borne lidar, J. Geophys. Res., 117, D15101, doi: 10.1029/2012JD017615.

[2] H Li, Y Yang*, X-M Hu, Z Huang, G Wang, B Zhang, T Zhang (2017), Evaluation of retrieval methods of daytime convective boundary layer height based on Lidar data, J. Geophy. Res. Atmos., 122, doi: 10.1002/2016JD025620.

Meanwhile, there are some studies have worked to detect cloud or aerosol layers based on lidar data, like Winker et al., 1994, Wang et al., 2001, Li et al., 2017, Dang et al., 2019, should also be cited here, and explain why your work is needed compared to the others'.

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

[4] Wang, Z.; Sassen, K. Cloud type and macrophysical property retrieval using multiple remote sensors. J. of Appl. Meteorol., 2001, 40, 1665–1683.

[5] H Li, Y Yang*, X-M Hu, Z Huang, G Wang, B Zhang. Application of Convective Condensation Level Limiter in Convective Boundary Layer Height Retrieval Based on Lidar Data. Atmosphere, 2017, 8, 79, doi: 10.3390/atmos8040079.

[6] Dang, R., Yang, Y., Li, H., Hu, X.-M., Wang, Z., Huang, Z., Zhou, T. and Zhang, T.: Atmosphere Boundary Layer Height (ABLH) Determination under Multiple-Layer Conditions Using Micro-Pulse Lidar, Remote Sensing, 11(3), 263, doi:10.3390/rs11030263, 2019.

Response:

Thanks for your suggestion. The following reference has been added.

The mentioned reference [1] and [2] had been add in P2.lines 31-32.

'Multiple approaches have been developed to determine the ABLH based on various observations, including radiosounding, remote sensing, and parameterisation from laboratory experiments (Li et al., 2017b; McGrath-Spangler and Denning, 2012; Nakoudi et al., 2019; Su et al., 2020a).'

The mentioned reference [3-6] has been add in P2.lines 55-63.

'The retrieval of BLHs under cloudy conditions is quite challenging. Some researchers have used the threshold of the attenuated scattering ratio (Campbell et al., 2008; Winker and Vaughan, 1994), the ratio of peaks to the base of the range-

corrected signal (RCS) (Wang and Sassen, 2001) to locate cloud tops and bases, while others have employed the objective upper limit of the convective condensation level (CCL)(Li et al., 2017a), as well as the analysis of signal continuity and the classification of whether the cloud caps the ABLH or is decoupled from the ABL (Dang et al., 2019b). The height restriction has significant advantages in removing the influence of clouds. Elevated aerosol layers (EALs) are characteristically similar to the aerosol trapped in ABL, using the threshold of lidar backscatter coefficient can distinguish them (Dubovik et al., 2002; Hänel et al., 2012; Peng et al., 2017). More instrument and multi-wavelength lidar systems are combined to obtain more accurate results to identified the EALs (Liu et al., 2019; Ortega et al., 2016).'

Minor revision:

1.Line 23: Change "continues" to "continuous".

2.Line 30: Change "on observation" to "based on various observations".

Response:

The word had been changed in P1.lines 23.

The word had been changed in P2.lines 30.

3.Line 34: coefficient between what?

Response:

The correlation coefficient between lidar retrieval algorithm and radiosonde is lower under stable conditions due to a complex aerosol structure that increases the difficulty of NBLH retrieval.

The sentence have rephrased P2.lines 33-35.

'The stable condition shows further agreement between lidar and radiosonde than the unstable condition because of the complex aerosol structure that complicates NBLH retrieval (Emeis and Schäfer, 2006; Martucci et al., 2007; Sawyer and Li, 2013). '

4.Lines 51-52: The sentence is difficult to understand.

Response:

The sentence have been revised in P2.lines 49-54.

'Some graph theory methods, such as the extended Kalman filter (Banks et al., 2014), Pathfinder and PathfinderTURB (de Bruine et al., 2017; Poltera et al., 2017), k-means clustering (Liu et al., 2018; Toledo et al., 2014), and The STRAT-2D algorithm (Haeffelin et al., 2012) have been proposed to yield promising results via an automated method that reduces the incorrect detection of ABLH. However, these techniques strongly depend on the vertical distribution of particle layers (aerosols and clouds) and are unsuitable for use under complicated multilayer conditions (Granados-Muñoz et al., 2012). '

5.Line 55: The fluctuation of NBLH, such statement is not completed. Line 63: Delete "in the experiment".

Response:

Thank you for your suggestion.

The word have been changed in P2.lines 65.

'Digressing from these previous efforts to estimate the ABLH, we herein present a new approach—cluster analysis of the gradient method (CA-GM)—to overcome the multilayer structure and remove the noise fluctuation of NBLH with raw data resolution.'

And the word has been delete.

6.Line 87: the value of turbulence? Such statement is incorrect.

Response:

Thank you for your suggestion. The word has been changed as turbulence intensity in P4 Line104-105.

'The assumption of the NBL at which the aerosol concentration and turbulence intensity are significantly higher in the NBL than in the free atmosphere (FA)(Dang et al., 2019a; Wang et al., 2020).'

7.Line 127: Change "the noise from the GM" to "the NBLH from the GM".

Response:

The word has been changed in P6 line 150.

8.Line 148: Please explain Dsig here.

Response:

The word have been changed in P7 line 172-173.

'Dsig is the empirical value to distinguish noise layer for verified starfield. '

9.Line 281: Change "influence " to "influencing".

Response:

The word have been changed in P17 line 311.

Thank you so much for your reviewing! We deeply appreciate your recognition of our research work.

Reference

Banks, R. F., Tiana-Alsina, J., María Baldasano, J. and Rocadenbosch, F.: Retrieval of boundary layer height from lidar using extended Kalman filter approach, classic methods, and backtrajectory cluster analysis, edited by A. Comerón, E. I. Kassianov, K. Schäfer, R. H. Picard, K. Stein, and J. D. Gonglewski, p. 92420F, Amsterdam, Netherlands., 2014.

de Bruine, M., Apituley, A., Donovan, D. P., Klein Baltink, H. and de Haij, M. J.: Pathfinder: applying graph theory to consistent tracking of daytime mixed layer height with backscatter lidar, Atmospheric Measurement Techniques, 10(5), 1893–1909, doi:10.5194/amt-10-1893-2017, 2017.

Campbell, J. R., Sassen, K. and Welton, E. J.: Elevated Cloud and Aerosol Layer Retrievals from Micropulse Lidar Signal Profiles, Journal of Atmospheric and Oceanic Technology, 25(5), 685–700, doi:10.1175/2007JTECHA1034.1, 2008.

Dang, R., Yang, Y., Hu, X.-M., Wang, Z. and Zhang, S.: A Review of Techniques for Diagnosing the Atmospheric Boundary Layer Height (ABLH) Using Aerosol Lidar Data, Remote Sensing, 11(13), 1590, doi:10.3390/rs11131590, 2019a.

Dang, R., Yang, Y., Li, H., Hu, X.-M., Wang, Z., Huang, Z., Zhou, T. and Zhang, T.: Atmosphere Boundary Layer Height (ABLH) Determination under Multiple-Layer Conditions Using Micro-Pulse Lidar, Remote Sensing, 11(3), 263, doi:10.3390/rs11030263, 2019b.

Emeis, S. and Schäfer, K.: Remote Sensing Methods to Investigate Boundary-layer Structures relevant to Air Pollution in Cities, Boundary-Layer Meteorology, 121(2), 377–385, doi:10.1007/s10546-006-9068-2, 2006.

Granados-Muñoz, M. J., Navas-Guzmán, F., Bravo-Aranda, J. A., Guerrero-Rascado, J. L., Lyamani, H., Fernández-Gálvez, J. and Alados-Arboledas, L.: Automatic determination of the planetary boundary layer height using lidar: One-year analysis over southeastern Spain: DETERMINATION OF THE PBL HEIGHT, Journal of Geophysical Research: Atmospheres, 117(D18), n/a-n/a, doi:10.1029/2012JD017524, 2012.

Haeffelin, M., Angelini, F., Morille, Y., Martucci, G., Frey, S., Gobbi, G. P., Lolli, S., O'Dowd, C. D., Sauvage, L., Xueref-Rémy, I., Wastine, B. and Feist, D. G.: Evaluation of Mixing-Height Retrievals from Automatic Profiling Lidars and Ceilometers in View of Future Integrated Networks in Europe, Boundary-Layer Meteorology, 143(1), 49–75, doi:10.1007/s10546-011-9643-z, 2012.

Li, H., Yang, Y., Hu, X.-M., Huang, Z., Wang, G. and Zhang, B.: Application of Convective Condensation Level Limiter in Convective Boundary Layer Height Retrieval Based on Lidar Data, Atmosphere, 8(12), 79, doi:10.3390/atmos8040079, 2017a.

Li, H., Yang, Y., Hu, X.-M., Huang, Z., Wang, G., Zhang, B. and Zhang, T.: Evaluation of retrieval methods of daytime convective boundary layer height based on lidar data: MEASUREMENT OF BOUNDARY LAYER HEIGHT, Journal of Geophysical Research: Atmospheres, 122(8), 4578–4593, doi:10.1002/2016JD025620, 2017b.

Liu, B., Ma, Y., Liu, J., Gong, W., Wang, W. and Zhang, M.: Graphics algorithm for deriving atmospheric boundary layer heights from CALIPSO data, Atmospheric Measurement Techniques, 11(9), 5075–5085, doi:10.5194/amt-11-5075-2018, 2018.

Liu, B., Ma, Y., Gong, W., Zhang, M. and Yang, J.: Improved two-wavelength Lidar algorithm for retrieving atmospheric boundary layer height, Journal of Quantitative Spectroscopy and Radiative Transfer, 224, 55–61, doi:10.1016/j.jqsrt.2018.11.003, 2019.

Martucci, G., Matthey, R., Mitev, V. and Richner, H.: Comparison between backscatter lidar and radiosonde measurements of the diurnal and nocturnal stratification in the lower troposphere, Journal of Atmospheric and Oceanic Technology, 24(7), 1231–1244, 2007.

McGrath-Spangler, E. L. and Denning, A. S.: Estimates of North American summertime planetary boundary layer depths derived from space-borne lidar: PBL DEPTH ESTIMATES FROM CALIPSO LIDAR, Journal of Geophysical Research: Atmospheres, 117(D15), n/a-n/a, doi:10.1029/2012JD017615, 2012.

Nakoudi, K., Giannakaki, E., Dandou, A., Tombrou, M. and Komppula, M.: Planetary boundary layer height by means of lidar and numerical simulations over New Delhi, India., Atmospheric Measurement Techniques, 12(5), 2019.

Poltera, Y., Martucci, G., Collaud Coen, M., Hervo, M., Emmenegger, L., Henne, S., Brunner, D. and Haefele, A.: PathfinderTURB: an automatic boundary layer algorithm. Development, validation and application to study the impact on in situ measurements at the Jungfraujoch, Atmospheric Chemistry and Physics, 17(16), 10051–10070, doi:10.5194/acp-17-10051-2017, 2017.

Sawyer, V. and Li, Z.: Detection, variations and intercomparison of the planetary boundary layer depth from radiosonde, lidar and infrared spectrometer, Atmospheric Environment, 79, 518–528, doi:10.1016/j.atmosenv.2013.07.019, 2013.

Su, T., Li, Z. and Kahn, R.: A new method to retrieve the diurnal variability of planetary boundary layer height from lidar under different thermodynamic stability conditions, Remote Sensing of Environment, 237, 111519, doi:10.1016/j.rse.2019.111519, 2020.

Toledo, D., Córdoba-Jabonero, C. and Gil-Ojeda, M.: Cluster Analysis: A New Approach Applied to Lidar Measurements for Atmospheric Boundary Layer Height Estimation, Journal of Atmospheric and Oceanic Technology, 31(2), 422–436, doi:10.1175/JTECH-D-12-00253.1, 2014.

Wang, H., Li, Z., Lv, Y., Zhang, Y., Xu, H., Guo, J. and Goloub, P.: Determination and climatology of the diurnal cycle of the atmospheric mixing layer height over Beijing 2013–2018: lidar measurements and implications for air pollution, Atmospheric Chemistry and Physics, 20(14), 8839–8854, doi:10.5194/acp-20-8839-2020, 2020.

Wang, Z. and Sassen, K.: Cloud Type and Macrophysical Property Retrieval Using Multiple Remote Sensors, Journal of Applied Meteorology, 40(10), 1665–1682, doi:10.1175/1520-0450(2001)040<1665:CTAMPR>2.0.CO;2, 2001.

Winker, D. M. and Vaughan, M. A.: Vertical distribution of clouds over Hampton, Virginia observed by lidar under the ECLIPS and FIRE ETO programs, Atmospheric Research, 34(1–4), 117–133, doi:10.1016/0169-8095(94)90084-1, 1994.