A Novel Machine Learning Algorithm for Planetary Boundary Layer Height Estimation Using AERI Measurement Data

Jin Ye, Lei Liu, Qi Wang, Shuai Hu, and Shulei Li

***Abstract*— Accurately determining the height of the planetary boundary layer (PBL) is important since it can affect the climate, weather, and air quality. Ground-based infrared hyperspectral remote sensing is an effective way to obtain this parameter. Compared with radiosonde measurements, its temporal resolution is much higher. In this study, a method to retrieve the PBL height (PBLH) from the ground-based infrared hyperspectral radiance data is proposed based on machine learning. In this method, the channels that are sensitive to temperature and humidity profiles are selected as the feature vectors, and the PBLHs derived from radiosonde are taken as the true values. The support vector machine (SVM) is applied to train and test the data set, and the parameters are optimized in the process. The data set collected at the Atmospheric Radiation Measurement (ARM) program Southern Great Plains (SGP) from 2012 to 2015 is analyzed. The instruments used in this letter include Atmospheric Emitted Radiance Interferometer (AERI), Vaisala CL31 ceilometer, and radiosonde. It shows that the root mean square error (RMSE) between the PBLHs calculated by the proposed method using AERI data and those from radiosonde data can be within 370 m, and the square correlation coefficient (SCC) is greater than 0.7. Compared with the PBLHs derived from the ceilometer, it can be found that the new method is more stable and less affected by clouds.**

***Index Terms*— Atmospheric emitted radiance interferometer (AERI), diurnal cycles, planetary boundary layer height (PBLH), seasonal cycles, support vector machine (SVM).**

I. Introduction

THE planetary boundary layer (PBL) is the lowest layer of the troposphere, which is a physically mixed layer due to the effects of shear-induced turbulence and convective overturning near the Earth's surface [1]. The PBL height (PBLH) is an important parameter for characterizing many atmospheric processes, including the dispersion of air pollutants and the formation of clouds [2]. PBLH is usually inferred from radiosonde measurements [3]; however, its temporal resolution is too sparse to detect the evolution of the diurnal structure. Therefore, to monitor PBL in a more continuous way, several remote sensing methods have been proposed to estimate PBLH by using the wind profiler, lidar, and ceilometer. The wind profiler can be used to detect the PBLH based on the turbulence structure data [4], but it is restricted to shallow boundary layers due to their limited range. Using aerosol as a tracer, the PBLH can be inferred from the aerosol vertical distribution measured by a lidar or a ceilometer [5]. However, the measurement under cloudy conditions needs to be further determined.

Some thermodynamic profile remote sensing instruments, such as atmospheric emitted radiance interferometer (AERI) [6] and microwave radiometer (MWR) [7], are also used to determine PBLHs from the inverted vertically thermodynamic variables [8]. One of the main limitations of these methods is their low vertical profile resolution. Moreover, the profile is particularly oversmoothed at higher altitudes so that the important features within the retrieved profiles are missing, such as vertical gradients of temperature and water vapor that used to do PBLH estimation [9]. Another challenge is how to reconcile PBLH from aerosol as a tracer and that from thermodynamic profiles. They often do not match up in their representation of the PBLH. The purpose of this letter is to demonstrate the potential of the machine learning approach for a fast, robust, accurate, and automated PBLH estimation approach to overcome the above problems. We do not carry out the inversion of thermodynamic profiles but directly use the AERI radiance (AERIRAD) data to estimate PBLH. Radiance data of the selected bands that are sensitive to temperature and humidity profiles are extracted as the input variables. The PBLHs derived from the radiosonde are taken as the "truths." The regression relationship between AERIRAD data and "truths" is established by using the support vector machine (SVM).

This letter is organized as follows. Section II is a brief summary of all the instruments and the data set used in this work. Section III formulates the PBLH retrieved algorithm using AERIRAD data. Section IV compares the PBLHs retrieved from AERI with those from radiosonde and Vaisala CL31 ceilometer (VCEIL). Finally, the conclusions are presented in Section V.

II. Establishment of Data Set

The instruments used in this study include AERI, radiosonde, and ceilometer located at the Atmospheric Radiation Measurement (ARM) Southern Great Plains (SGP) site. The data set covers the period from January 2010 to December 2015. There are many methods to calculate PBLHs by using radiosonde. In this letter, we use Liu and Liang method based on the vertical potential temperature gradient to calculate the PBLHs [10], [11] as "truths."

AERI is a ground-based spectrometer that measures the downwelling infrared radiance from 3.3 to 19 μm (3020–520 cm^{-1}) at 0.5-cm^{-1} resolution. When using AERIRAD to retrieve PBLHs, radiances of 555 channels

Manuscript received January 6, 2021; revised March 15, 2021; accepted April 6, 2021. Date of publication April 26, 2021; date of current version December 15, 2021. This work was supported by the National Natural Science Foundation of China under Grant 41875025 and Grant 41575024. *(Corresponding author: Lei Liu.)*

Jin Ye, Lei Liu, Shuai Hu, and Shulei Li are with the College of Meteorology and Oceanography, National University of Defense Technology, Changsha 410073, China (e-mail: jin_ye2020@foxmail.com; liuleidll@gmail.com; hushuai2012@nudt.edu.cn; lishulei@nudt.edu.cn).

Qi Wang is with the Unit No. 91922 of PLA, Sanya 572099, China (e-mail: qiqiqiwang@outlook.com).

Digital Object Identifier 10.1109/LGRS.2021.3073048

TABLE I
NUMBER OF DATA SETS FROM 2010 TO 2015 AT THE ARM SGP SITE

Particular year	Number of sample set groups		
	Cloudy	Cloud free	Total
2010	429	916	1345
2011	462	911	1373
2012	181	573	754
2013	183	383	566
2014	443	799	1242
2015	528	913	1441

that are sensitive to thermodynamic profiles are selected as the input parameters [12]. The sensitive channels of water vapor are 538–588 and 1250–1350 cm^{-1}, and the sensitive channels of temperature are 612–618, 624–660, 674–713, and 2223–2260 cm^{-1}. The observational time of radiosonde and AERI is different, so it is necessary to do time-matching before establishing a database. Table I shows the number of time-matching between AERI and total available radiosondes of the whole day. We use VCEIL and Total Sky Image (TSI) to judge cloudy or cloud-free at the corresponding time. Data from 2010 and 2011 are used as training samples, and the rest are testing samples. PBLHs derived from thermodynamic profiles given by the value-added product of AERI (AERIPROF) [13], [14] are also calculated, which are based on the maximum potential temperature gradient [15]. PBLHs from ceilpbl (the value-added product of ceilometer) [16] are used to do the comparison with those from AERI.

III. INVERSION OF PBLH BASED ON SVM

Using AERIRAD to estimate PBLHs can be regarded as a regression problem. Section III-A presents the channel selection process. Section III-B describes the support vector regression method. Section III-C summarizes the proposed algorithm for PBLH estimation.

A. Channel Selection

First, the radiance data of 538–588, 612–618, 624–660, and 674–713 cm^{-1} are divided by 180, and 1250–1350 cm^{-1} is divided by 80. The radiance data of 1250–1350 and 2223–2260 cm^{-1} are divided by 5. This normalization process makes the radiance data of each channel in the same order of magnitude.

Then, the minimal-redundancy-maximal-relevance (mRMR) method is used to screen the characteristic quantities [17] by considering the correlation between features and categories and the redundancy between features. Finally, the most important 50 channels are selected, which can be found in Table II.

B. Support Vector Machine

We consider a training set of $(x_i, y_i)|_{i=1}^{k}, i = 1, \ldots, k$, where x_i and y_i represent the input and output and k is the dimension of the training set. The form of SVM estimation is taken as

$$f(x) = w \cdot \Phi(x) + b \tag{1}$$

where w is a weighting matrix and b is a bias term. Φ denotes a nonlinear transformation to a higher dimensional feature space. Equation (1) can be converted to minimize the regression error as [18]

$$\min \frac{1}{2}||w||^2 + C\sum_{i=1}^{k}\left(\zeta_i + \zeta_i^*\right) \tag{2}$$

$$\text{s.t. } |y_i - w \cdot \Phi(x) + b| \le \varepsilon + \xi_i, \quad \xi_i,\ \zeta_i^* > 0 \tag{3}$$

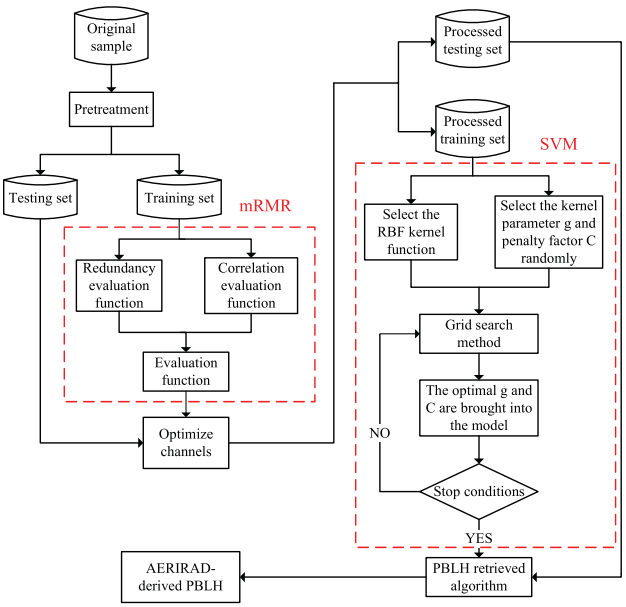

Fig. 1. Proposed algorithm for PBLH estimation.

where ε is the permissible error and C denotes the penalty factor. Every vector outside ε-tube is captured in slack variables ζ_i and ζ_i^*. The radial basis function (RBF) kernel is applied as

$$K(x_i, x) = \exp(-g||x_i - x||^2) \tag{4}$$

where g is a parameter inversely proportional to the width of the Gaussian kernel.

C. Proposed Algorithm for PBLH Estimation

Fig. 1 shows the flowchart of the proposed algorithm for PBLHs estimation. The main processes are as follows.

1) Create the training and testing samples. Find the matching data between AERI and radiosonde to establish the database.
2) Select the best channels based on the mRMR method. Calculate the maximum correlation and minimum redundancy between channels.
3) Search for the best parameters of the SVM algorithm. The selected channels are brought into the SVM algorithm, and the best kernel parameter g and penalty factor C are selected by grid search method [19] to verify the testing set. The range of C and g is 2^{-8}–2^{8}, and the search step is $2^{0.8}$.
4) Use the optimal model to estimate the PBLHs of all testing samples. The accuracy of the model is measured by the root mean square error (RMSE) and the square correlation coefficient (SCC). RMSE is the deviation between the retrieved results and the "truths"; SCC is the degree of the fitting.

Table III shows the retrieved results with different channel numbers selected by the mRMR method. Considering the computational complexity and the accuracy of the model, the optimal 50 channels are selected, as shown in Table II. The corresponding C and g are 9.1896 and 5.278.

IV. EVALUATION OF PBLH INVERSION ALGORITHM

The data set from 2012 to 2015 is used to evaluate the PBLHs retrieved algorithm in this section. The comparison

TABLE II
50 OPTIMAL CHANNELS SELECTED BY MRMR

Serial number	Wavenumber (cm^{-1})	Serial number	Wavenumber (cm^{-1})	Serial number	Wavenumber (cm^{-1})	Serial number	Wavenumber (cm^{-1})	Serial number	Wavenumber (cm^{-1})
1	1349.05	**11**	2247.29	**21**	2240.06	**31**	2236.68	**41**	2232.34
2	1250.21	**12**	540.00	**22**	2253.56	**32**	541.93	**42**	1252.14
3	561.22	**13**	1253.10	**23**	1250.69	**33**	1255.99	**43**	538.56
4	2251.63	**14**	2241.50	**24**	1305.65	**34**	2243.43	**44**	1288.30
5	1260.81	**15**	1312.40	**25**	559.29	**35**	2225.59	**45**	2251.15
6	2234.27	**16**	558.33	**26**	2229.45	**36**	2250.18	**46**	1276.24
7	1278.17	**17**	2228.00	**27**	1251.65	**37**	1268.05	**47**	2237.16
8	2259.34	**18**	2255.48	**28**	2257.41	**38**	558.81	**48**	565.56
9	1287.33	**19**	1255.51	**29**	1272.39	**39**	2249.70	**49**	1339.89
10	2226.56	**20**	1259.85	**30**	1338.44	**40**	1319.15	**50**	1269.98

TABLE III
ESTIMATION RESULTS WITH DIFFERENT CHANNEL NUMBERS

Channel numbers	RMSE (m)	SCC
3	363.15	0.6391
5	321.26	0.7140
10	256.94	0.8162
20	247.87	0.8290
30	246.37	0.8293
50	227.73	0.8574
200	225.92	0.8581
555	213.16	0.8636

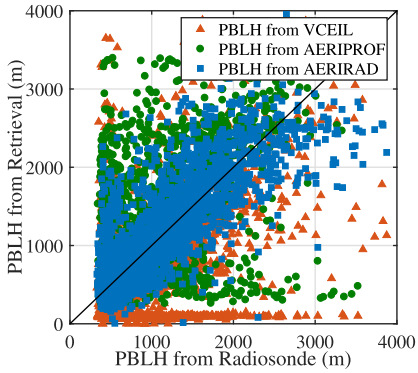

Fig. 2. Scatter plots of PBLHs derived from radiosonde, AERIRAD, AERIPROF, and VCEIL of the whole day from 2012 to 2015 at the ARM SGP site.

of the PBLHs obtained by AERIRAD, AERIPROF, VCEIL, and radiosonde is shown in Fig. 2. When the PBLH is below 2500 m, the difference between AERIRAD- and radiosonde-derived PBLHs is relatively small. However, the difference gradually increases when the PBLHs are above 3000 m; the maximum difference can reach more than 1000 m. It may be partly because AERIRAD is sensitive to the near-surface layer. The RMSE and SCC of PBLHs derived from AERIRAD, AERIPROF, and VCEIL are shown in Table IV. AERIPROF-derived PBLHs are based on the thermodynamic profiles inversed from radiance on selected bands. The error of temperature and humidity profile inversion will increase the PBLH estimation error of AERIPROF. However, using AERIRAD to retrieve PBLHs directly can avoid this error. This is the reason why the accuracy of AERIRAD-derived PBLHs is higher than AERIPROF-derived PBLHs. What is more, AERIPROF can give temperature and humidity information every 50 m below 1000 m and every 100 m between 1000 and 2000 m, the vertical resolution is low, and it decreases with height. This is also the reason why the accuracy of PBLHs' inversion from AERIPROF is lower.

TABLE IV
RMSE AND SCC OF PBLHS DERIVED FROM AERIRAD, VCEIL, AND AERIPROF OF THE WHOLE DAY FROM 2012 TO 2015

	AERIRAD	AERIPROF	VCEIL
RMSE (m)	364.8	641.1	548.1
SCC	0.7041	0.4644	0.3145

VCEIL detects the PBLHs using aerosol as a tracer; the aerosol-derived PBLHs may not be consistent with the PBLHs derived from thermodynamic profiles. Therefore, VCEIL- and radiosonde-derived PBLHs may have a certain deviation. AERIRAD uses the radiance of selected bands to retrieve PBLHs, which is consistent with the thermodynamic profile method. Hence, compared with VCEIL-derived PBLHs, the SCC of AERIRAD-derived PBLHs (0.7041) is significantly larger, and the RMSE (364.8 m) is smaller. In addition, VCEIL is more easily affected by the clouds, which also affects the accuracy of PBLHs derived from VCEIL. To take an insightful look at this phenomenon, Section IV-A introduces the influence of cloudy on the PBLHs of two instruments. Sections IV-B and IV-C compare the accuracy of two instruments in different time periods. Section IV-D analyzes the edge cases.

A. Evaluation When a Cloud Is in the Field of View

Table V shows the RMSE and SCC of AERIRAD- and VCEIL-derived PBLHs under cloudy and cloud-free conditions. The agreement of PBLH detection from AERIRAD is better in the cloud-free case than cloudy. The SCC of the AERIRAD-derived PBLHs (0.7219) under cloud-free conditions is significantly higher. Clouds can attenuate the downward infrared hyperspectral radiance. Sawyer and Li [20] have proved that AERI-derived PBLHs based on the thermodynamic profiles are unreliable in cloudy conditions. Therefore, the cloud may also increase the retrieval error of AERIRAD. Based on this phenomenon, we add cloudy samples in the training set to reduce the influence of clouds on the AERIRAD-derived PBLHs. The proportion of cloudy and cloud-free samples in the training set is given in Table I.

Compared to radiosonde-derived PBLHs, AERIRAD-derived PBLHs are more accurate than VCEIL-derived PBLHs no matter in cloudy cases or cloud-free cases. Moreover, the PBLHs derived from VCEIL are more seriously affected by clouds. When it is cloudy, the SCC of the VCEIL-derived PBLHs is small. Fig. 3 shows an example in cloudy conditions on December 26, 2014. After 1700 Coordinated Universal Time (UTC), there were low and thick clouds, which had a great influence on the PBLHs' inversion. In this case, it is

TABLE V

COMPARISON OF THE PBLHS DERIVED FROM AERIRAD AND VCEIL UNDER CLOUDY OR CLOUD-FREE CONDITIONS

	RMSE (m)		SCC	
	AERI	VCEIL	AERI	VCEIL
Cloudy	324.6	550.9	0.5166	0.1424
Cloud-free	377.9	551.2	0.7219	0.3759

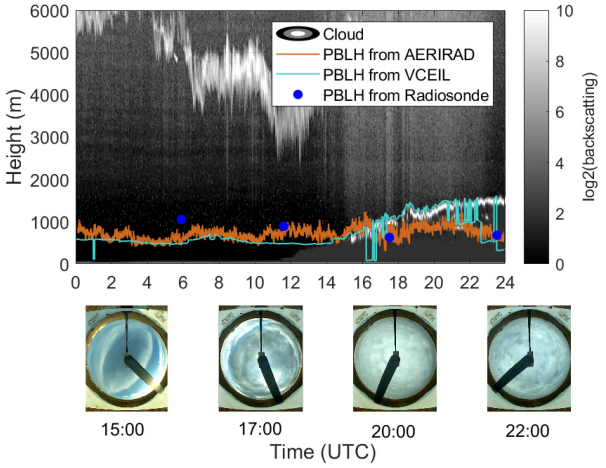

Fig. 3. Comparison of PBLHs derived from radiosonde, AERIRAD, and VCEIL under the cloudy condition on December 26, 2014, at the ARM SGP site.

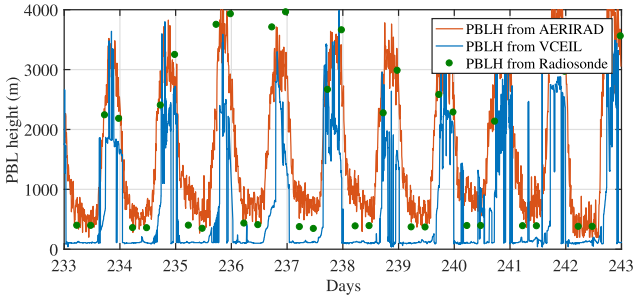

Fig. 4. PBLHs derived from radiosonde, AERIRAD, and VCEIL during a ten-day period from August 21 to 30, 2013, at the ARM SGP site.

easy to misjudge the cloud base height as the PBLH by using VCEIL.

B. Diurnal Cycle

Fig. 4 shows the results of ten-day continuous detection of the PBLHs from August 21 to 30, 2013. Boundary layers undergo strong diurnal cycling at the ARM SGP site [20], and the PBLHs reach the minimum in the early morning and the maximum in the afternoon. PBL can be classified into three regimes: the convective boundary layer (CBL), the stable boundary layer (SBL), and the residual layer (RL) [10]. As shown in Table VI, it can be seen that CBL (91.92%) is dominant in the daytime, and SBL (75.02%) is dominant in the nighttime of the testing set. Compared to radiosonde-derived PBLHs, the errors of both the AERIRAD- and VCEIL-derived PBLHs are relatively larger in the nighttime. The SCC of AERIRAD-derived PBLHs in the daytime is higher than that in the nighttime, and the RMSE is relatively smaller. This may due to the relative stability of SBL at night, which leads to the insufficient thermodynamic distribution to determine the PBLHs. Therefore, turbulent kinetic energy (TKE) profiles are

TABLE VI

COMPARISON OF THE PBLHS DERIVED FROM AERIRAD AND VCEIL DURING DAYTIME AND NIGHTTIME

	PBL regimes (%)			RMSE (m)		SCC	
	CBL	SBL	RL	AERI	VCEIL	AERI	VCEIL
Daytime	91.92	5.57	2.51	425.1	629.5	0.5610	0.2781
Nighttime	24.98	75.02	0	251.2	479.9	0.1224	0.0246

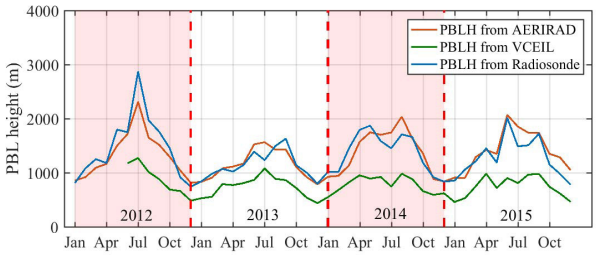

Fig. 5. Monthly mean values of PBLHs derived from radiosonde, AERIRAD, and VCEIL during daytime from 2012 to 2015 at the ARM SGP site.

used to calculate PBLHs [21]. However, AERIRAD is not sensitive to TKE as TKE does not affect the downwelling infrared radiance. The AERIRAD is more suitable for determining the PBLH in the daytime than at nighttime. Furthermore, the PBL is well-mixed in the daytime; VCEIL-derived PBLHs are also better in the daytime.

In addition, compared with VCEIL-derived PBLHs, AERIRAD-derived PBLHs have better accuracy. The SCC of the AERIRAD-derived PBLHs is relatively higher, and the RMSE is small no matter in the daytime or nighttime. It may be due to the inconsistency between aerosol-derived PBLHs and the PBLHs derived from thermodynamic profiles.

C. Seasonal Cycle

Due to the inaccuracy of AERI and VCEIL in the measurement of PBLHs in the nighttime, this section is limited to analyze the PBLHs during daytime. Fig. 5 shows monthly averages (AVGs) of PBLHs measured by three instruments at the ARM SGP site. Compared to the radiosonde-derived PBLHs, the AERIRAD-derived PBLHs are more consistent than the VCEIL-derived PBLHs on the seasonal scale. The AVGs are given in Table VII. The VCEIL-derived PBLHs are about 500 m lower than the radiosonde-derived PBLHs in spring and autumn, 700 m lower in summer, and 400 m lower in winter. It reveals that we can reconcile the PBLHs from aerosol and thermodynamic profiles in a simple bias correction on the seasonal time scale. Table VII also shows that the SCC of the AERIRAD-derived PBLHs in summer and autumn is higher than that in spring and winter. It may be due to the varieties of cloudy conditions in different seasons. As shown in Table VIII, there are more cloudy scenes (about 40%) in the spring and winter of the testing set.

D. Edge Cases

The AERIRAD- and radiosonde-derived PBLHs are compared in different height regions during daytime in this section. The mean and standard deviation of the difference between these two PBLHs as a function of height are shown in Fig. 6(a). The region from 500 to 2500 m occupies the majority in both training and testing set [see Fig. 6(b)]. PBLHs derived from AERIRAD and radiosonde are consistent in this region also. The error of AERIRAD-derived PBLHs is larger

TABLE VII

COMPARISON OF THE PBLHS DERIVED FROM THREE INSTRUMENTS DURING DAYTIME IN FOUR SEASONS

	RMSE (m)		SCC		AVG (m)		
	AERI	VCEIL	AERI	VCEIL	AERI	VCEIL	RS
Spring	479.0	572.4	0.3152	0.1858	1302.8	831.8	1369.2
Summer	428.9	776.3	0.4667	0.1995	1778.0	968.3	1705.0
Autumn	380.7	611.5	0.5108	0.2203	1295.2	734.8	1292.2
Winter	267.0	447.8	0.2654	0.1130	889.9	531.3	901.2

TABLE VIII

PROPORTION OF CLOUDY SAMPLES DURING DAYTIME IN FOUR SEASONS

	Spring	Summer	Autumn	Winter
Proportion (%)	41.25	25.22	21.86	46.23

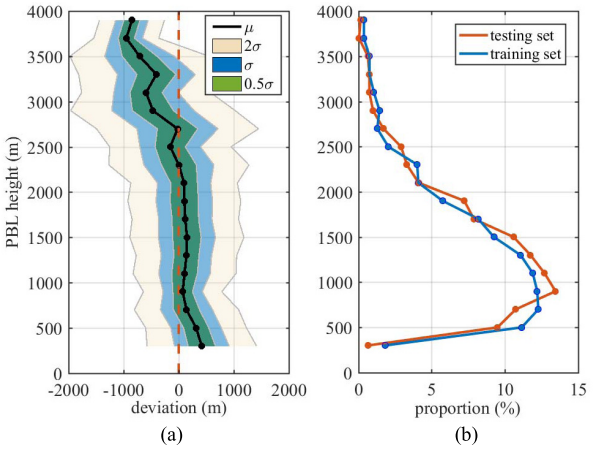

Fig. 6. (a) Mean (μ) and standard deviation (σ) of the difference between AERIRAD- and radiosonde-derived PBLHs in different height regions of the testing set. (b) Proportion of PBLHs in different height regions of training and testing set during the daytime.

when the PBLHs are above 3000 m or below 500 m. It may be caused by the small number of such edge cases in the training set. There are only 46 groups (3.3%) with the PBLHs above 3000 m and 94 groups (6.7%) below 500 m. In addition, AERI is less sensitive to the temperature and humidity profiles above 3 km [13], which means low sensitivity for high PBLHs.

V. CONCLUSION

This letter proposes a machine learning approach for PBLHs estimation using AERIRAD. The results of the proposed approach were compared with those from AERIPROF, VCEIL, and radiosonde to determine its performance. The experimental results demonstrated that the AERIRAD-based method had a better response to the sharp change of the PBLHs in the daytime with high accuracy and was less affected by clouds. What is more, it could better reflect the diurnal and seasonal cycles of PBLHs. The major advantage of the proposed method is that PBLHs can be retrieved directly from AERIRAD avoiding the thermodynamic profiles inversion. We can get fast, robust, accurate, and high temporal resolution PBLHs by using this approach. Insufficiently, we only used the Liu and Liang [11] method of radiosonde as the "truth." Our future work includes testing the influence of the "truth" values obtained by different algorithms on the inversion results. We also intend to use deep-learning techniques to avoid the process of channel selection and will verify the applicability of this algorithm in different sites.

REFERENCES

[1] S. Emeis, K. Schäfer, and C. Münkel, "Surface-based remote sensing of the mixing-layer height a review," *Meteorologische Zeitschrift*, vol. 17, no. 5, pp. 621–630, Oct. 2008.

[2] W. Zhang *et al.*, "Planetary boundary layer height from CALIOP compared to radiosonde over China," *Atmos. Chem. Phys.*, vol. 16, no. 15, pp. 9951–9963, Aug. 2016.

[3] T. Su *et al.*, "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.*, vol. 122, no. 7, pp. 3929–3943, Apr. 2017.

[4] A. Molod, H. Salmun, and M. Dempsey, "Estimating planetary boundary layer heights from NOAA profiler network wind profiler data," *J. Atmos. Ocean. Technol.*, vol. 32, no. 9, pp. 1545–1561, Sep. 2015.

[5] G. Tsaknakis *et al.*, "Inter-comparison of lidar and ceilometer retrievals for aerosol and planetary boundary layer profiling over Athens, Greece," *Atmos. Meas. Techn.*, vol. 4, no. 6, pp. 1261–1273, Jun. 2011.

[6] W. G. Blumberg, T. J. Wagner, D. D. Turner, and J. Correia, "Quantifying the accuracy and uncertainty of diurnal thermodynamic profiles and convection indices derived from the atmospheric emitted radiance interferometer," *J. Appl. Meteorol. Climatol.*, vol. 56, no. 10, pp. 2747–2766, Oct. 2017.

[7] D. Cimini, T. J. Hewison, L. Martin, J. Güldner, C. Gaffard, and F. S. Marzano, "Temperature and humidity profile retrievals from ground-based microwave radiometers during TUC," *Meteorologische Zeitschrift*, vol. 15, no. 1, pp. 45–56, Feb. 2006.

[8] M. Wiegner *et al.*, "Mixing layer height over Munich, Germany: Variability and comparisons of different methodologies," *J. Geophys. Res.*, vol. 111, no. 13, pp. 1–17, 2006.

[9] U. Saeed, F. Rocadenbosch, and S. Crewell, "Adaptive estimation of the stable boundary layer height using combined lidar and microwave radiometer observations," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 12, pp. 6895–6906, Dec. 2016.

[10] *Planetary Boundary Layer Height Value Added Product: Radiosonde Retrievals*, U.S. Dept. Energy, Washington, DC, USA, 2013, pp. 1–36.

[11] S. Liu and X.-Z. Liang, "Observed diurnal cycle climatology of planetary boundary layer height," *J. Climate*, vol. 23, no. 21, pp. 5790–5809, Nov. 2010.

[12] U. Löhnert, D. D. Turner, and S. Crewell, "Ground-based temperature and humidity profiling using spectral infrared and microwave observations. Part I: Simulated retrieval performance in clear-sky conditions," *J. Appl. Meteorol. Climatol.*, vol. 48, no. 5, pp. 1017–1032, May 2009.

[13] *Retrieving Temperature and Moisture Profiles From AERI Radiance Observations: AERIPROF Value-Added Product Technical Description Revision 1*, U.S. Dept. Energy, Washington, DC, USA, 2007.

[14] W. F. Feltz, W. L. Smith, H. B. Howell, R. O. Knuteson, H. Woolf, and H. E. Revercomb, "Near-continuous profiling of temperature, moisture, and atmospheric stability using the atmospheric emitted radiance interferometer (AERI)," *J. Appl. Meteorol.*, vol. 42, no. 5, pp. 584–597, May 2003.

[15] P. Seibert, F. Beyrich, S. E. Gryning, S. Joffre, A. Rasmussen, and P. Tercier, "Review and intercomparison of operational methods for the determination of the mixing height," *Atmos. Environ.*, vol. 34, no. 7, pp. 1001–1027, 2000.

[16] N. Eresmaa, A. Karppinen, S. M. Joffre, J. Räsänen, and H. Talvitie, "Mixing height determination by ceilometer," *Atmos. Chem. Phys.*, vol. 6, no. 6, pp. 1485–1493, May 2006.

[17] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 8, pp. 1226–1238, Aug. 2005.

[18] A. G. Abo-Khalil and D.-C. Lee, "MPPT control of wind generation systems based on estimated wind speed using SVR," *IEEE Trans. Ind. Electron.*, vol. 55, no. 3, pp. 1489–1490, Mar. 2008.

[19] X. Wang, F. Zhang, H.-T. Kung, V. C. Johnson, and A. Latif, "Extracting soil salinization information with a fractional-order filtering algorithm and grid-search support vector machine (GS-SVM) model," *Int. J. Remote Sens.*, vol. 41, no. 3, pp. 953–973, Feb. 2020.

[20] V. Sawyer and Z. Li, "Detection, variations and intercomparison of the planetary boundary layer depth from radiosonde, lidar and infrared spectrometer," *Atmos. Environ.*, vol. 79, pp. 518–528, Nov. 2013.

[21] A. C. Fitch, J. K. Lundquist, and J. B. Olson, "Mesoscale influences of wind farms throughout a diurnal cycle," *Monthly Weather Rev.*, vol. 141, no. 7, pp. 2173–2198, Jul. 2013.