https://doi.org/10.5194/egusphere-2024-3959
Preprint. Discussion started: 14 February 2025

Best Estimate of the Planetary Boundary Layer Height from Multiple Remote Sensing Measurements

Damao Zhang[1], Jennifer Comstock[1], Chitra Sivaraman[1], Kefei Mo[1], Raghavendra Krishnamurthy[1], Jingjing Tian[1], Tianning Su[2], Zhanqing Li[3], Natalia Roldán-Henao[3]

[1]Pacific Northwest National Laboratory, Richland, Washington, USA
[2] Lawrence Livermore National Laboratory, Livermore, CA, USA
[3]Department of Atmospheric and Oceanic Sciences, University of Maryland, College Park, College Park, MD, USA

Correspondence to: Damao Zhang (damao.zhang@pnnl.gov)

Abstract. Remote sensing measurements have been widely used to estimate the planetary boundary layer height (PBLHT). Each remote sensing approach offers unique strengths and faces different limitations. In this study, we use machine learning (ML) methods to produce a best-estimate PBLHT (PBLHT-BE-ML) by integrating four PBLHT estimates derived from remote sensing measurements at the Department of Energy (DOE) Atmospheric Radiation Measurement (ARM) Southern Great Plains (SGP) observatory. Three ML models—Random Forest (RF) Classifier, RF Regressor, and Light Gradient-Boosting Machine (LightGBM)—were trained on a dataset from 2017 to 2023 that included radiosonde, various remote sensing PBLHT estimates, and atmospheric meteorological conditions. Evaluations indicated that PBLHT-BE-ML from all three models improved alignment with PBLHT derived from Radiosonde data (PBLHT-Sonde), with LightGBM demonstrating the highest accuracy under both stable and unstable boundary layer conditions. Feature analysis revealed that the most influential input features at the SGP site were the PBLHT estimates derived from, a) potential temperature profiles retrieved using Raman Lidar (RL) and Atmospheric Emitted Radiance Interferometer (AERI) measurements (PBLHT-THERMO), b) vertical velocity variance profiles Doppler Lidar (PBLHT-DL), and c) aerosol backscatter profiles from Micropulse Lidar (PBLHT-MPL). The trained models were then used to predict PBLHT-BE-ML at a temporal resolution of 10 minutes, effectively capturing the diurnal evolution of PBLHT and its significant seasonal variations, with the largest diurnal variation observed over summer at the SGP site. We applied these trained models to data from the ARM Eastern Pacific Cloud Aerosol Precipitation Experiment (EPCAPE) field campaign (EPC), where the PBLHT-BE-ML, particularly with the LightGBM model, demonstrated improved accuracy against PBLHT-Sonde. Analyses of model performance at both the SGP and EPC sites suggest that expanding the training dataset to include various surface types, such as ocean and ice-covered areas, could further enhance ML model performance for PBLHT estimation across varied geographic regions.

https://doi.org/10.5194/egusphere-2024-3959
Preprint. Discussion started: 14 February 2025

1 Introduction

The planetary boundary layer (PBL) refers to the lowest part of the Earth's atmosphere that directly interacts with the Earth's surface (Stull, 1988). This layer responds to surface forcing within an hour or less and closely follows the diurnal cycle of surface heating and cooling over land (Deardorff et al. 1974; Xi et al., 2022). Within the PBL, turbulent motion drives significant exchanges of heat, mass, moisture, and momentum between the surface and the free troposphere. These exchanges significantly influence atmospheric processes, including aerosol mixing and transport, cloud formation and evolution, aerosol-cloud interactions (Painemal et al. 2017; Su et al., 2024), and precipitation formation, which strongly affect human activities (Teixeira et al., 2021). The vertical depth of PBL is represented by the planetary boundary layer height (PBLHT), which corresponds to an important parameter in atmospheric process studies and numerical model simulations (Zhang et al., 2020). The PBLHT is often used to characterize PBL structures and as a key factor for estimating flux exchanges between the surface and the atmosphere. Although there are several well-accepted definitions of the PBL as described by LeMone et al. (2019), accurate estimates of the PBLHT remain challenging.

Due to the convective nature of the PBL, vertical gradients of thermodynamic parameters including temperature (or potential temperature) and water vapor (or relative humidity), as well as trace gases and aerosols are commonly used to estimate the PBLHT. For well-mixed convective PBL, the PBL top is characterized by a positive gradient of potential temperature, and negative gradients of water vapor, trace gases, and aerosols. Radiosonde data, offering high vertical-resolution measurements of temperature and moisture profiles, are widely used for estimating PBLHT (Liu and Liang 2010; Seidel et al., 2010). However, radiosonde data suffers from poor temporal resolutions. Most radiosonde stations launch a sounding system only twice daily, making it challenging to study and understand the temporal evolution of the PBL based on radiosonde data. Model forecasts and reanalysis data have also been used to estimate global PBLHT climatology (Von Engeln and Teixeira 2013). However, the uncertainty and coarse spatiotemporal resolution of these data could prevent reliable estimates of PBLHT.

The use of continuous remote sensing observations provides high temporal resolution of PBLHT estimates. These observations include sodar (Contini et al., 2008), radar wind profilers (Salmun et al., 2023), aerosol lidars (Dang et al., 2019; Su et al., 2020), Doppler lidar (DL) (Tucker et al., 2009; Krishnamurthy et al., 2021), water vapor and/or temperature lidars and radiometers (Turner et al., 2014), as well as Global Navigation Satellite System Radio Occultation (GNSS-RO) (Nelson et al., 2021). Although obtaining the vertical distribution of turbulence parameters remains challenging, these observations provide valuable data on the thermodynamic properties (e.g., water vapor and/ or temperature lidars and radiometers), dynamic properties (e.g., Sodar, radar wind profilers, DL, GNSS-RO), and distribution of tracer substance (e.g., aerosol lidars) of PBL, all of which can be used to estimate PBLHT. Kotthaus et al. (2023) present a comprehensive review of the capabilities and limitations of PBLHT estimates from ground-based remote sensing observations. When remote sensing instruments are deployed on various platforms, such as ground stations or spaceborne satellites, they can provide PBLHT estimates for fixed

https://doi.org/10.5194/egusphere-2024-3959
Preprint. Discussion started: 14 February 2025

locations as well as on a global scale (Kalmus et al., 2022; Luo et al., 2016; Roldán-Henao et al. 2024 a, b; Salmun et al., 2023; Su et al., 2020).

It is important to note that different observational techniques capture varying characteristics of the PBL, and as a result, PBLHT estimates from different observations may differ significantly. Each observation has its own advantages and limitations. Radiosonde data offer high accuracy in situ measurements of atmospheric temperature and water vapor profiles, making PBLHT estimates from radiosondes more reliable. For example, the Vaisala RS92 radiosonde thermodynamic sensor can measure pressure, temperature, and relative humidity with accuracies of 0.5 hPa, 0.2 °C, and 2%, respectively (Holdridge 2020). As a result, radiosonde data are often used to evaluate PBLHT estimates from remote sensing methods (Su et al., 2020) or serve as the 'ground truth' for training machine learning models (Krishnamurthy et al., 2021). Aerosol lidars and ceilometers (CEIL) use aerosol as a tracer for PBL observations (Zhang et al., 2022). These instruments are affordable, portable, and reliable even under harsh weather conditions. However, PBLHT estimates from aerosol lidars suffer from the impacts of the aerosol residual layer during nighttime and early morning, and from transported aerosol layers. DL directly employs the measured vertical velocity variance profile to estimate PBLHT. Studies show that the PBLHT estimated from DL compares well with that from radiosonde data, especially during the growing and decaying periods of the PBL evolution (Tucker et al., 2009). However, the DL approach is sensitive to the choice of velocity variance threshold and can vary as a function of site. In addition, the approach has difficulty providing reliable PBLHT estimates under stable PBL conditions (Krishnamurthy et al., 2021). PBLHT estimates from thermodynamic profiles retrieved using water vapor/temperature lidars or high spectral radiometer measurements can employ the same methods for PBLHT estimates from radiosonde data, which works under clear sky. However, the retrieved thermodynamic profiles sometimes suffer from significant retrieval uncertainties, which hinder the accurate estimate of PBLHT. Although no single method or measurement consistently provides the most reliable PBLHT estimates across all PBL schemes and environmental conditions, combining PBLHT estimates from multiple remote sensing techniques can lead to more accurate estimates under various conditions (Kotthaus et al., 2023).

Machine learning (ML) approaches offer powerful tools for integrating information from various observations, identifying patterns, making predictions, and gaining deeper insights from complex datasets. Even when the relationships between variables are not fully understood, ML methods enable the extraction of valuable knowledge from diverse data sources. ML techniques have been applied to improve PBLHT estimates. For example, Krishnamurthy et al. (2021) used a Random Forest (RF) model to enhance PBLHT estimates from DL measurements by combining PBLHT estimates from the Tucker method with environmental factors such as surface sensible and latent heat fluxes, wind speed and direction, surface upward longwave and shortwave radiation, soil moisture and temperature, the Monin-Obukhov length, and cloud base height. Their results demonstrated significant improvements in PBLHT estimates compared to those derived solely from the Tucker method, validated against radiosonde data. Su and Zhang (2024) introduced a deep learning framework trained on extensive remote sensing and radiosonde data, leveraging conventional meteorological measurements to produce robust and reliable PBLHT

https://doi.org/10.5194/egusphere-2024-3959
Preprint. Discussion started: 14 February 2025

estimates across diverse environmental conditions. Several studies have explored various ML models, such as gradient-boosting regression trees, K-means clustering, AdaBoost, and deep neural networks, have shown enhanced PBLHT estimates from various remote sensing measurements (Rieutord et al., 2021; de Arruda Moreira et al., 2022; Liu et al., 2022).

In this study, we apply ML methods to enhance PBLHT estimates by integrating multiple remote sensing-derived PBLHTs. Specifically, we utilize data from the Micropulse Lidar (PBLHT-MPL), Ceilometer (PBLHT-CEIL), Doppler Lidar (PBLHT-DL), and thermodynamic profiles retrieved using Raman Lidar (RL) and Atmospheric Emitted Radiance Interferometer (AERI) measurements (PBLHT-THERMO). We hypothesize that PBLHTs derived from observations of various PBL characteristics, e.g., thermodynamic effects (PBLHT-THERMO), dynamic effects (PBLHT-DL), and tracer particles (PBLHT-MPL and PBLHT-CEIL), complement each other. Together, they can provide valuable information on the true PBLHT under various PBL schemes and environmental conditions. We use radiosonde-derived PBLHT (PBLHT-Sonde, more details in section 2.2) to train ML models, aiming to predict the best estimate of PBLHT (PBLHT-BE) by combining the four approaches with ancillary environmental data. We use multiple years of advanced remote sensing and radiosonde data collected at the Department of Energy (DOE) Atmospheric Radiation Measurement (ARM) Southern Great Plains (SGP) atmospheric observatory to train ML models and test their predictions. In addition, we use remote sensing and radiosonde data at a recent ARM field campaign to test whether the ML method works at different locations.

The paper is organized as follows: Section 2 presents an overview of the ARM SGP site, its instruments and observations, PBLHT estimation approaches from radiosonde data and various remote sensing observations; the ML method, training and testing, feature importance analysis are presented in Section 3; finally, Section 4 presents the summary and conclusions.

2 ARM Observation and PBLHT Estimates

2.1 the ARM SGP Site Observations

The DOE ARM user facility deploys advanced remote sensing and in situ instruments at climatically critical locations. The SGP atmospheric observatory central facility, located in north-central Oklahoma (36° 36′ 26″ N, 97° 29′ 15″ W), is the world's most extensive climate research facility. Surrounded by cattle pasture and wheat fields, the SGP central facility is equipped with a wide range of advanced instrument clusters, providing critical observations for research on atmospheric process-level studies of aerosol, cloud, precipitation, land-atmospheric interactions, etc. The advanced instruments include various radars, lidars, radiometers, aerosol/gas sampling instruments, and cloud sample instruments. A detailed description of these instruments can be found on ARM's SGP instrument webpage (https://www.arm.gov/capabilities/observatories/sgp) and in Mather and Voyles (2013). Key ground-based instruments, data streams, and their measurements used for PBLHT estimates are listed in Table 1. The SGP has been collecting data since 1992 and has accumulated extensive long-term datasets. This study utilizes the data collected between 2017 and 2023 for PBLHT analysis as well as training and testing ML models.

https://doi.org/10.5194/egusphere-2024-3959
Preprint. Discussion started: 14 February 2025

Table 1. Key Ground-based instruments and measurements at the SGP site used for PBLHT estimates.

Instrument	ARM data stream	Measured or derived quantities	References
Balloon-Borne Sounding System (SONDE)	pblhtsonde1mcfarl.c1	PBLHTs from radiosonde (PBLHT-sonde), PBL regimes (PBL regime, details in section 2.2)	Sivaraman et al. (2013), Liu and Liang (2010)
Micropulse Lidar (MPL)	pblhtmpl1sawyerli.c1	PBLHT from MPL (PBLHT-MPL)	Sawyer and Li (2013)
Ceilometer (CEIL)	ceilpblht.a0	PBLHT from CEIL (PBLHT-CEIL)	Münkel et al. (2007)
	ceil.b1	Cloud base height (CBH)	Münkel and Roininen (2010)
Doppler Lidar (DL)	pblhtdl.c1	PBLHT from DL (PBLHT-DL)	Tucker et al. (2009)
	dlprofwstats4news.c1	Vertical velocity variance profiles	Newsome et al. (2019)
Raman Lidar (RL)	pblhtrl1zhang	PBLHT from thermodynamic profile (PBLHT-THERMO)	Heffter et al. (1980)
	rlproftemp2news10m.c0	Temperature profiles	Newsome et al. (2013)
Atmospheric Emitted Radiance Interferometer (AERI)	tropoe.c1	Temperature profiles	Turner and Löhnert (2021)
Eddy Correlation Flux Measurement System (ECOR) and Carbon Dioxide Flux Measurement Systems (CO2FLUX)	30co2flx25m.b1	Sensible heat flux, latent heat flux, total kinetic energy (TKE), friction velocity (u^*), Monin-Obukhov length (MO-Len), surface wind speed (SWS) and direction (SWD)	Tang et al., (2019)
Solar and Infrared Radiation Station (SIRS)	qcrad1long.c2	Upward longwave (LRAD) and shortwave radiation (SRAD)	Long and Shi (2006)
Surface Meteorology Systems (MET)	met.b1	Air temperature (T), Relative humidity (RH)	Kyrouac and Tuftedal (2024)

https://doi.org/10.5194/egusphere-2024-3959
Preprint. Discussion started: 14 February 2025

2.2 PBLHT-SONDE Value Added Product

The balloon-borne sounding system (SONDE) is typically launched four times a day at approximately 5:30 (0:30), 11:30 (6:30), 17:30 (12:30), and 23:30 (18:30) Universal Time Coordinate (UTC) (Local Time, LT) at the SGP site. During intensive observing periods (IOPs), SONDE launches occur more frequently, often up to eight times per day. SONDE measures vertical profiles of the atmospheric thermodynamic state, including atmospheric pressure, temperature, moisture, and wind speed and direction, with a temporal resolution of 1 second. This corresponds to vertical resolutions ranging from several meters to over 10 meters. The measured temperature can reach an accuracy of 0.2 °C.

To estimate PBLHT, the ARM PBLHT-SONDE Value Added Product (VAP) applies three commonly used methods: the Heffter (1980), Liu and Liang (2010), and bulk Richardson number approaches (Seibert et al., 2000). Details of the three methods are described in the corresponding references and by Sivaraman et al. (2013). In brief, the Heffter method (PBLHT-Heffter) determines PBLHT as the height of the base of the lowest inversion layer in the potential temperature profile, requiring a temperature difference of at least 2 K between the base and the top of the inversion layer.

The Liu-Liang method (PBLHT-Liuliang) first classifies the PBL into three regimes: convective boundary layer (CBL), stable boundary layer (SBL), and neutral residual layer (NRL). This classification is based on the potential temperature (θ) difference between the fifth and second level of sounding data ($\theta_5 - \theta_2$), compared to a stability threshold. CBL is identified when $\theta_5 - \theta_2 < -\delta_s$, SBL when $\theta_5 - \theta_2 > +\delta_s$, and NRL when $-\delta_s \leq \theta_5 - \theta_2 \leq +\delta_s$. The PBL scheme is a critical factor, as PBLHT estimates exhibit significantly different characteristics across different PBL regimes. As noted by Zhang et al. (2022), the CBL and NRL regimes can be combined and referred as the unstable PBL condition, in contrast to the SBL regime, which is the stable PBL condition. In this study, we follow the same approach to separate our analysis into unstable and stable PBL conditions. The Liu-Liang method then applies specific algorithms tailored to each regime for estimating PBLHT. For the unstable PBL, it identifies PBLHT as the level k at which $\theta_k - \theta_2 > \delta_u$, where δ_u is another stability threshold. The thresholds for different surface types are provided in Liu and Liang (2010). For stable PBL, PBLHT is determined as the top of the stable layer above the surface or the height of the low-level jet (LLJ nose), whichever is lower.

https://doi.org/10.5194/egusphere-2024-3959
Preprint. Discussion started: 14 February 2025

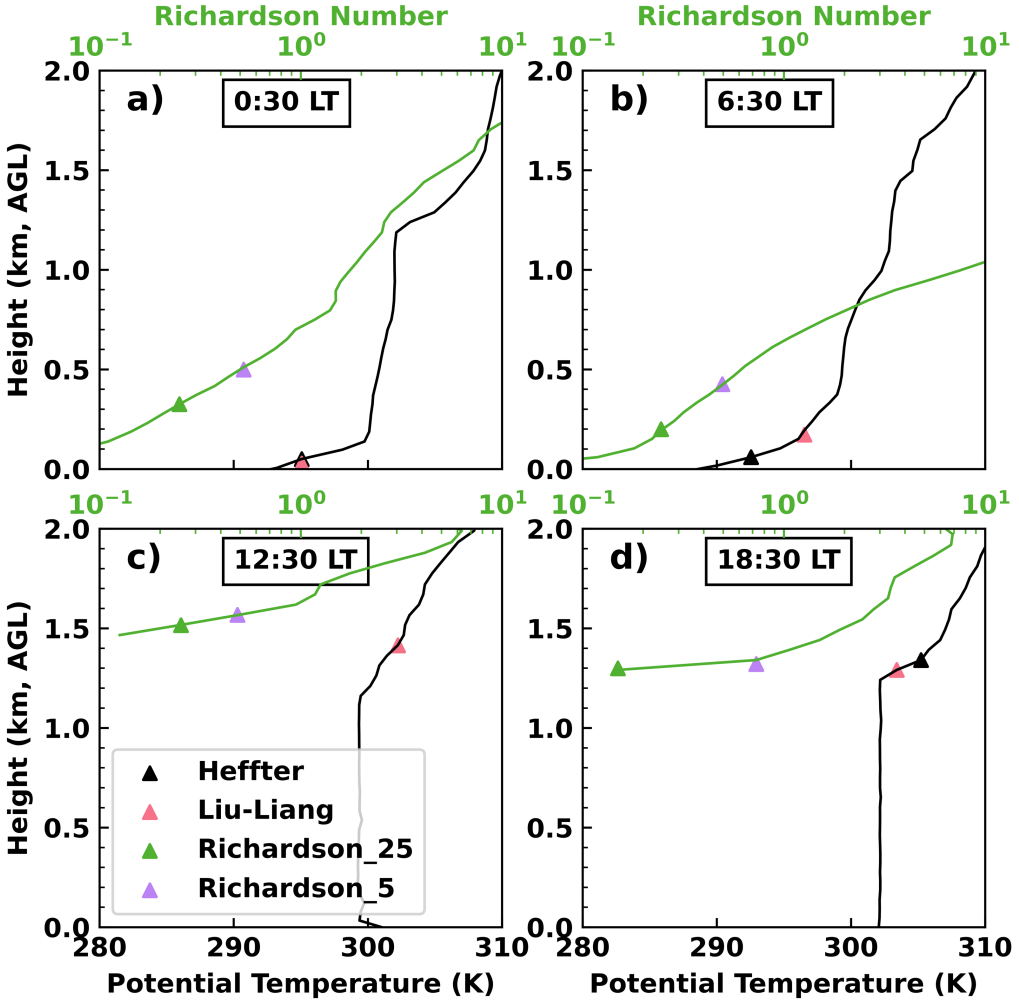


Figure 1. PBLHT estimates from the PBLHT-SONDE VAP during the four SONDE launches on May 8, 2017. Since PBLHTs from Heffter and Liu-Liang methods primarily rely on potential temperature profiles, their PBLHT estimates are plotted alongside these profiles. PBLHTs from the bulk Richardson method are plotted on the corresponding *Ri* profiles. Richardson_25 refers to PBLHT estimate using a Ri_c threshold of 0.25; while Richardson_5 for a Ri_c threshold of 0.5. AGL stands for Above Ground Level, and all altitudes in this study are in terms of AGL.

The bulk Richardson method (PBLHT-Richardson) determines PBLHT as the altitude where the bulk Richardson number (*Ri*) first exceeds a critical value Ri_c. The PBLHT-SONDE VAP provides PBLHT estimates using two Ri_c thresholds, 0.25 and 0.5, based on previous studies (Holtslag et al., 1990; Seibert et al., 2000). An example of PBLHT estimates from the PBLHT-SONDE VAP during the four SONDE launches on May 8, 2017, is shown in Figure 1. PBLHTs are low during nighttime and early morning (0:30 and 6:30 LT), grow to ~1.5 km at noon (12:30 LT) and decay slightly in the late afternoon (18:30 LT).

https://doi.org/10.5194/egusphere-2024-3959
Preprint. Discussion started: 14 February 2025

Although PBLHT estimates from sonde data are generally considered more reliable, there is no definitive ground truth to determine which method performs better. Different methods and algorithms, even when using the same measurements, can yield significantly different PBLHT estimates (Seidel et al., 2010; Roldán-Henao et al., 2024). This is evident in Figures 1a and 1b, which show substantial discrepancies in PBLHT estimates from the PBLHT-SONDE VAP during nighttime (0:30 LT) and early morning (6:30 LT). Statistical comparisons of PBLHT estimates from the three approaches also reveal large discrepancies, particularly under stable PBL conditions (Figures S1 and S2). As a general rule, the median of PBLHT estimates from the three approaches is recommended (Smith and Carlin 2024). Since the two PBLHT estimates from the bulk Richardson method use the same approach but with different Ri_c values, only the estimate with Ri_c value of 0.25 is used when calculating the median PBLHT across the three approaches (Seibert et al., 2000).

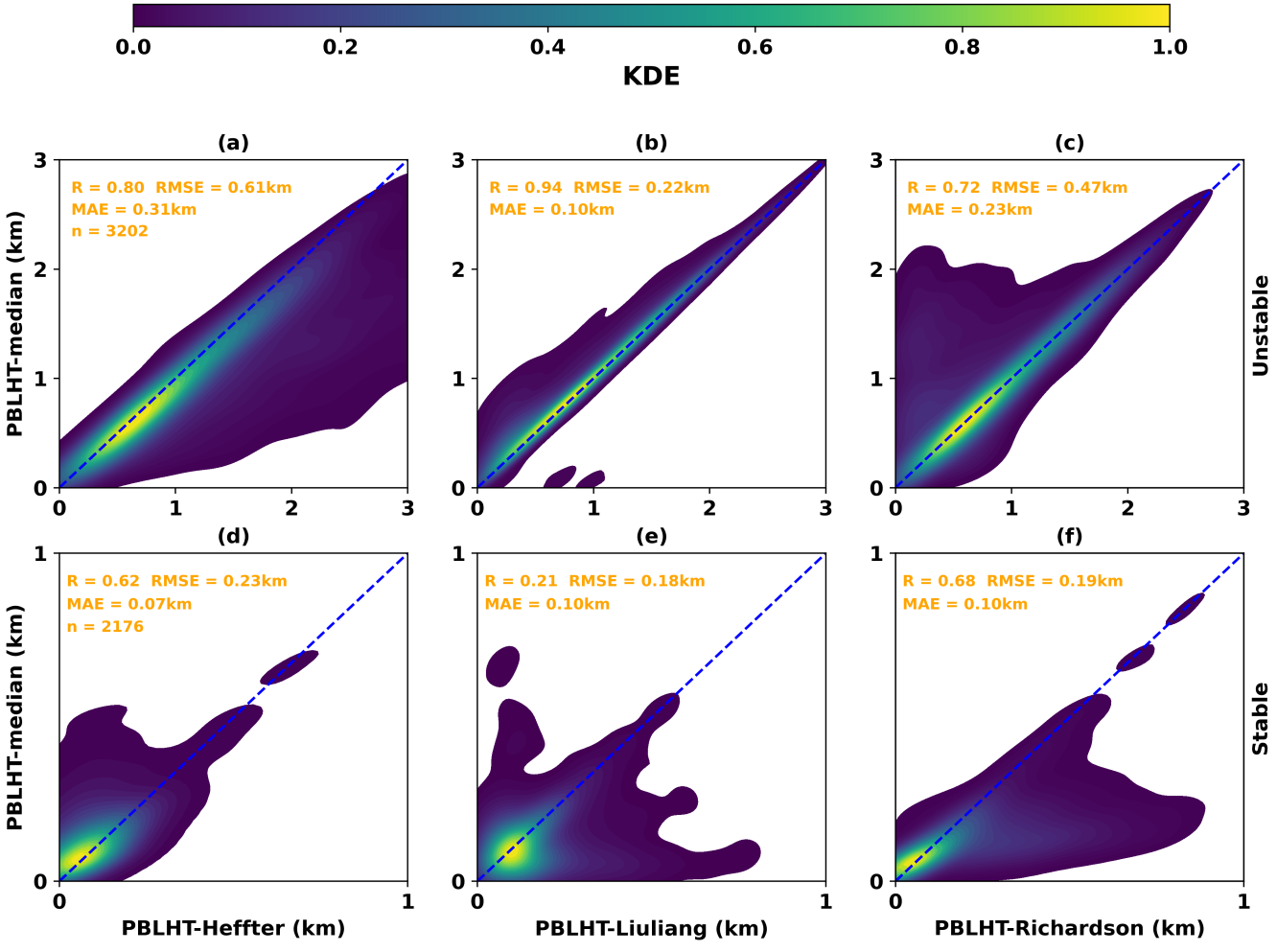


Figure 2. Comparisons of the three PBLHT estimates from the PBLHT-SONDE VAP with their median (PBLHT-median) using kernel distribution estimate (KDE) under unstable (a-c) and stable (d-f) PBL conditions. R is the correlation coefficient. RMSE is the root mean square error. MAE is the mean absolute error. n is the number of samples.

https://doi.org/10.5194/egusphere-2024-3959
Preprint. Discussion started: 14 February 2025

This study uses three parameters to evaluate how well two datasets are compared: the linear correlation coefficient (R), the Root Mean Square Error (RMSE), and the Mean Absolute Error (MAE). Figure 2 shows this evaluation and compares the three PBLHT estimates from the PBLHT-SONDE VAP with their median (PBLHT-median). Under unstable PBL conditions, the PBLHT-Liuliang estimate tends to align closely with the PBLHT-median, which shows the highest R, the lowest RMSE and MAE values. When PBLHT exceeds 2 km, PBLHT-Heffter is generally larger than the PBLHT-median, whereas PBLHT-Richardson is typically smaller than the median when PBLHT is below 1 km. On the other hand, under stable PBL conditions, the PBLHT-median is more likely to resemble either the PBLHT-Heffter or PBLHT-Richardson estimates. These patterns are consistent with previous studies. For example, research using aerosol lidars and Doppler lidar (DL) suggests that PBLHT estimates from lidar are more in line with the Liu-Liang method during daytime at the SGP site (Sawyer and Li, 2013; Su et al., 2020, Krishnamurthy et al., 2021). Additionally, Lewis (2016) found that the Heffter method produces reasonable PBLHT values based on careful inspection of temperature and humidity profiles during the Marine ARM GPCI Investigation of Clouds (MAGIC) field campaign. Seibert et al. (2000) noted that the bulk Richardson number method offers better PBLHT estimates under SBL conditions. In this study, we use the PBLHT-median as the 'ground truth' to train and test the ML models. For simplicity, in the following sections, we refer to the PBLHT-median as PBLHT-Sonde.

2.2 PBLHTs from Remote Sensing Measurements

In this section, we describe four PBLHT estimates derived from remote sensing observations: PBLHT-MPL, PBLHT-CEIL, PBLHT-DL, and PBLHT-THERMO.

2.2.1 PBLHT-MPL

PBLHT-MPL is estimated using range and overlap-corrected MPL backscatter profiles through the wavelet covariance method, as described by Sawyer and Li (2013). The ARM MPL system provides aerosol backscatter intensity and depolarization ratio profile measurements at a wavelength of 532, with a vertical resolution of 15 meters and a temporal resolution of 10 seconds. The raw data are post-processed to subtract background noise and apply deadtime, afterpulse, and overlap corrections (Campbell et al., 2002). Figure 3a presents an example of MPL backscatter profiles from May 8, 2017, at the SGP site. Residual aerosol layers are clearly visible up to approximately 2 km during the nighttime and early morning hours until 6:00 LT, with elevated aerosol layers above 2 km at the start of the day, which can lead to significant errors in lidar-based PBL retrievals (Su et al., 2020). After 11:00 LT, strong aerosol backscatter gradients at the top of the PBL are evident.

The Sawyer and Li's approach uses the Haar wavelet covariance transform on the lidar backscatter intensity profile to identify the strongest negative gradient with height, which is then taken as the PBL top (Brooks 2003). This wavelet covariance transform is an advanced edge detection technique well-suited for vertically resolved active remote sensing data. The method relies only on lidar backscatter intensity profiles and is independent of absolute lidar calibration, making it adaptable for

https://doi.org/10.5194/egusphere-2024-3959
Preprint. Discussion started: 14 February 2025

different MPL systems across various locations. A key parameter in this approach is the dilation factor, *a*, of the Haar wavelet, which usually ranges between 100 m and 1 km. A small value of *a* may result in noisy PBLHT estimates due to spurious gradients, while large values increase the minimal detectable PBLHT. The minimal detectable PBLHT is half of *a.* Following Sawyer and Li (2013), we set the dilation factor, *a* a, to 1 km, corresponding to a minimal detectable PBLHT of 0.5 km above the ground level (AGL). Figure 3e shows that before 11:00 LT, PBLHT-MPL captures the top of the residual layer as the PBLHT, while after 11:00 LT, it accurately tracks the PBLHT of convective PBL.

2.2.2 PBLHT-CEIL

In this study, we use ceilometer PBLHT estimates directly provided by the vendor built-in software, "BL-View", which provides real-time monitoring of boundary layer structures and identifies up to three potential boundary layer heights (Münkel and Räsänen, 2010). ARM ceilometers utilize the Vaisala CL31 model, which has a maximum vertical range of 7.7 km (Münkel and Räsänen, 2004). The CL31 provides total attenuated backscatter coefficient profiles at a wavelength of 910 nm, with a vertical resolution of 10 meters and a temporal resolution of 2 seconds. Figure 3b shows the time-height plot of CEIL backscatter profiles. Due to longer wavelength and larger field of view, CEIL backscatter profiles tend to be noisier compared to those from MPL. Despite this, aerosol layer structures are still discernible within the PBL.

The CL31 ceilometer estimates PBLHT using the gradient method, which identifies local gradient minima in the range- and overlap-corrected backscatter coefficient profiles. First, the data are averaged to a 16-second temporal resolution to improve aerosol signal reliability. Then, local gradient minima are detected using a 30-minute temporal and 360-meter vertical sliding window average. Furthermore, the enhanced gradient method applies a cloud and precipitation filter during the averaging process to suppress false layer identification. This allows for robust PBLHT estimates under various weather conditions (Münkel and Roininen, 2010). For the three potential boundary layer heights, the BL-View algorithm assigns a quality index (ranging from 1 to 3) to each boundary layer height candidate. A higher quality index is assigned to stronger gradients, greater distances between minima, and when no clouds are detected near the boundary layer. We select the boundary layer height candidate with the highest quality index as the ceilometer-estimated PBLHT, consistent with previous studies (Zhang et al., 2022).

As illustrated in Figure 3e on the example day, the PBLHT-MPL consistently identifies the top of the residual layer as PBLHT before 11:00 LT. In contrast, the PBLHT-CEIL sometimes detects lower PBLHTs that are closer to PBLHT-Sonde. After 11:00 LT, when the convective PBL develops, PBLHT-CEIL typically captures PBLHT accurately, although occasional underestimations still occur. Since ceilometers are cost-effective, portable, and reliable, they are widely deployed at various ground-based atmospheric observatories. As a result, numerous studies have developed methods to improve PBLHT estimates from ceilometers (Caicedo et al., 2020; de Arruda Moreira et al., 2022). In this study, we directly use PBLHT-CEIL from the

https://doi.org/10.5194/egusphere-2024-3959
Preprint. Discussion started: 14 February 2025

Vaisala BL-View software, as it is readily available at ARM observatories, quality-controlled, and widely adopted by the research community.

2.2.3 PBLHT-DL

ARM DL provides range-resolved measurements of attenuated backscatter intensity at approximately 1.5 μm and radial velocity profiles. Most ARM DLs have full upper-hemispheric scanning capabilities and typically operate in a vertical staring mode, offering vertical velocity measurements with a resolution of 30 meters and a temporal resolution of 1 second or less. Additionally, they perform plan-position-indicator (PPI) scans to capture three-dimensional turbulent flows (Newsom and Krishnamurthy, 2022). The DLPROF-WSTATS VAP (https://arm.gov/capabilities/science-data-products/vaps/dlprof-wstats) uses vertical velocity data to calculate 30-minute statistics of vertical velocity variance (σ^2_w), skewness, and kurtosis. The vertical velocity profiles are oversampled by a factor of three to produce these statistical quantities at a 10-minute temporal resolution. Figure 3c shows the time-height plot of σ^2_w. During nighttime the stable atmospheric stratification created by radiative cooling leads to negative buoyancy production, effectively suppressing turbulence. σ^2_w stays low values in nighttime and early morning. While during daytime, enhanced convection increases surface-based mixing and increases atmospheric turbulence. As a result, σ^2_w starts to increase after 8:00 LT.

According to Tucker et al. (2009), the depth of the convective boundary layer can be estimated by identifying the height where σ^2_w falls below a specified threshold, such as 0.04 m^2s^{-2}, which is the approach used in this study to derive PBLHT-DL from DLPROF-WSTATS. Figure 3e illustrates that PBLHT-DL reports lower PBLHT values before 8:00 LT, which aligns closely with PBLHT-Sonde. However, during this period, PBLHT-DL consistently reports the lowest levels, indicating its limitations in providing valid PBLHT estimates under stable PBL conditions. Between 8:00 LT and 12:00 LT, PBLHT-DL effectively captures PBL growth, offering valuable PBLHT estimates during this critical period. After 12:00 LT, PBLHT-DL continues to provide reasonable PBLHT estimates, closely matching PBLHT-Sonde.

2.2.4 PBLHT-THERMO

At night, the presence of residual layers can create strong water vapor and aerosol gradients, making it difficult to derive PBLHT from water vapor or aerosol lidar backscatter profiles. In such conditions, PBLHT estimates based on thermodynamic profiles, such as potential temperature, are more reliable (Seidel et al., 2010). Both ARM RL and AERI provide height- and time-resolved temperature profiles that can be used for this purpose. ARM RL systems measure elastically backscattered light from aerosols at a wavelength of 355 nm, as well as inelastically scattered light from atmospheric molecules at specifically tuned channels. The ARM RLPROF-TEMP VAP (https://www.arm.gov/capabilities/science-data-products/vaps/rlproftemp) uses measurements from RL's rotational Raman channels to retrieve temperature profiles (Newsom et al., 2013). These temperature profiles have a vertical resolution of 60 meters and a temporal resolution of 10 minutes (Newsom et al., 2018). The uncertainty in the RL's temperature retrievals is calculated using standard error analysis, and by default, temperature

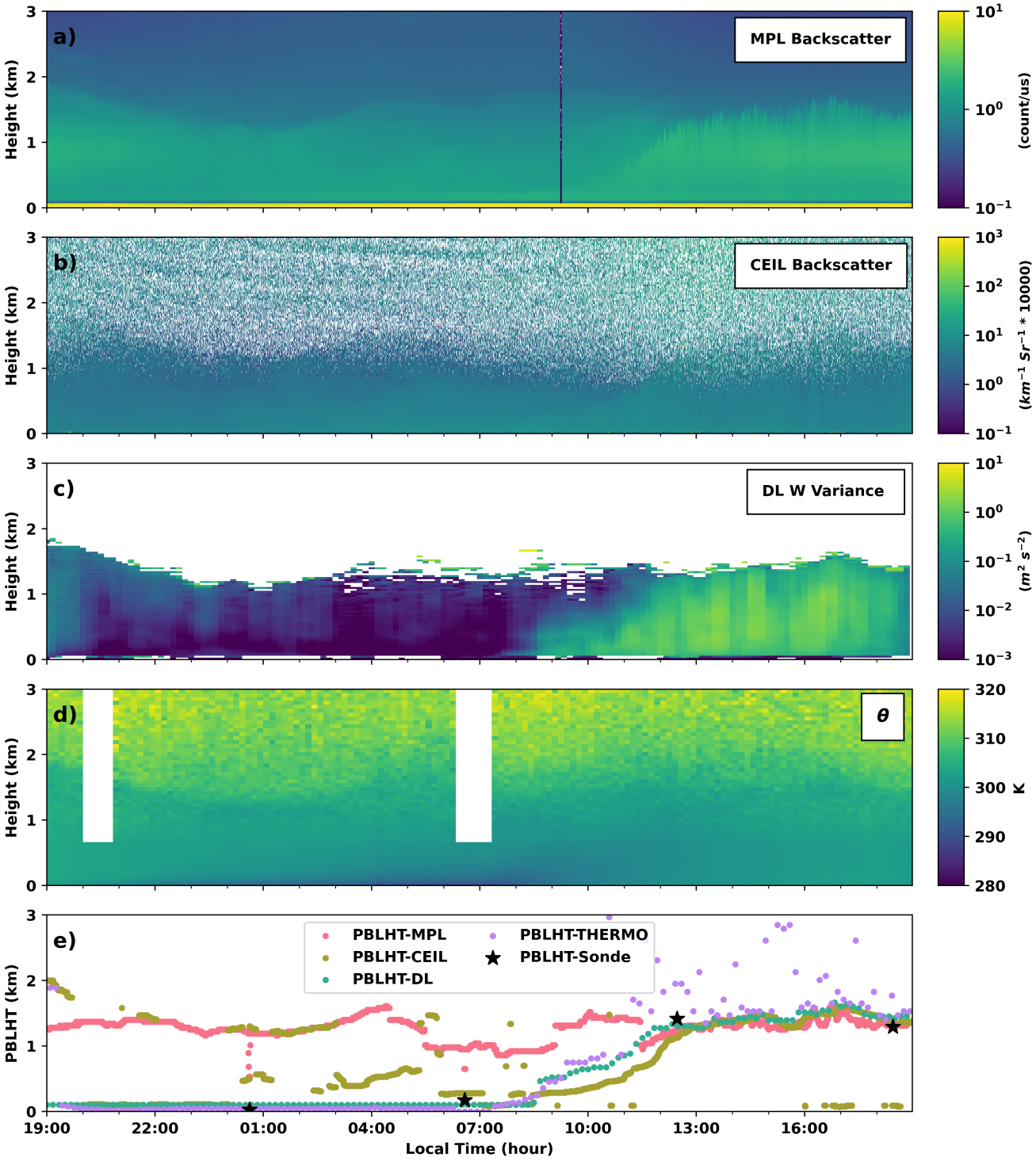
https://doi.org/10.5194/egusphere-2024-3959
Preprint. Discussion started: 14 February 2025

retrievals with relative uncertainties greater than 0.05 are excluded. It is important to note that lidar signals and temperature retrievals can be affected by incomplete overlap between the outgoing laser beam and the receiver's field of view. As a result, overlap corrections are typically applied at low altitudes, where retrievals have higher uncertainties.

The ARM AERI retrieves boundary layer temperature profiles, which can also be retrieved from the measured infrared (IR) radiance spectrum. The ARM AERI instrument measures absolute IR spectral radiance from 3000 to 520 wavenumbers (cm^{-1}) with a spectral resolution of 1 cm^{-1}, capturing data every 20 seconds. The ARM TROPospheric Optimal Estimation (TROPoe) VAP (https://www.arm.gov/capabilities/science-data-products/vaps/tropoe) provides retrievals of lower tropospheric temperature and water vapor mixing ratio profiles from AERI IR spectral radiance measurements using a physical-iterative retrieval approach, as described by Turner and Löhnert (2014, 2021). TROPoe retrievals are produced at a 5-minute temporal resolution, with vertical resolution starting at approximately 25 meters near the surface, decreasing to 800 meters by 3 km in altitude. The output file also includes the one-sigma uncertainty for each retrieved variable, derived from the error covariance matrix within the optimal estimation framework. The mean bias errors relative to radiosonde profiles are less than 0.2 K for temperatures at heights below 2 km under clear-sky conditions (Turner and Löhnert, 2014).

To address the impact of overlap correction on RL temperature retrievals and leverage the higher vertical resolution of TROPoe near the surface, we combine the RLPROF-TEMP and TROPoe temperature profiles. Specifically, TROPoe temperatures are used for altitudes below 700 meters. At altitudes between 700 and 1100 meters, RLPROF-TEMP temperatures are linearly scaled using the ratio of the mean TROPoe temperature to the mean RLPROF-TEMP temperature within this layer. RLPROF-TEMP temperatures are applied for altitudes above 1100 meters. Potential temperature (θ) profiles are then calculated using the combined temperature profiles, along with pressure profiles from the Interpolated Sonde (INTERPSONDE) VAP (https://www.arm.gov/capabilities/science-data-products/vaps/interpsonde). Figure 3d displays the time-height plots of the θ profiles. PBLHT-THERMO is derived from these θ profiles using the Heffter method, the same approach employed in the PBLHT-SONDE VAP. As shown in Figure 3e, PBLHT-THERMO reports lower PBLHT values before 8:00 LT. Notably, PBLHT-THERMO indicates that the PBL begins to grow around 7:30 LT, while PBLHT-DL shows noticeable PBL growth after approximately 8:30 LT. This discrepancy may arise because thermodynamic measurements are more effective for estimating PBLHT than dynamic measurements in shallow, stable PBL conditions. After 12:00 LT, during convective PBL, PBLHT-THERMO exhibits more scattered estimates, likely due to greater uncertainties in temperature retrievals and smaller potential temperature gradients at higher altitudes.

https://doi.org/10.5194/egusphere-2024-3959
Preprint. Discussion started: 14 February 2025



https://doi.org/10.5194/egusphere-2024-3959
Preprint. Discussion started: 14 February 2025

Figure 3. Remote sensing observations and retrievals on May 8, 2017at the ARM SGP site. Panels from top to bottom are: a) time-height plot of MPL backscatter profiles; b) ceilometer total attenuated backscatter coefficient profiles; c) DL derived vertical velocity variance (W) profiles; d) potential temperature (θ) profiles from RL and TROPospheric Optimal Estimation (TROPoe) retrievals; e) PBLHT estimates from remote sensing observations and the PBLHT-SONDE VAP.

2.3 Statistical Evaluations of PBLHT Estimates from Remote Sensing Measurements Against PBLHT-Sonde

With multiple years of concurrent MPL, CEIL, DL, RL, TROPoe, and radiosonde data at the SGP site, we can statistically evaluate the performance of each PBLHT estimate from remote sensing measurements with PBLHT-Sonde. Figure 4 shows comparisons of the four PBLHT estimates from remote sensing measurements with PBLHT-Sonde under unstable and stable PBL conditions using 7 years of data between 2017 and 2023 at the SGP site.

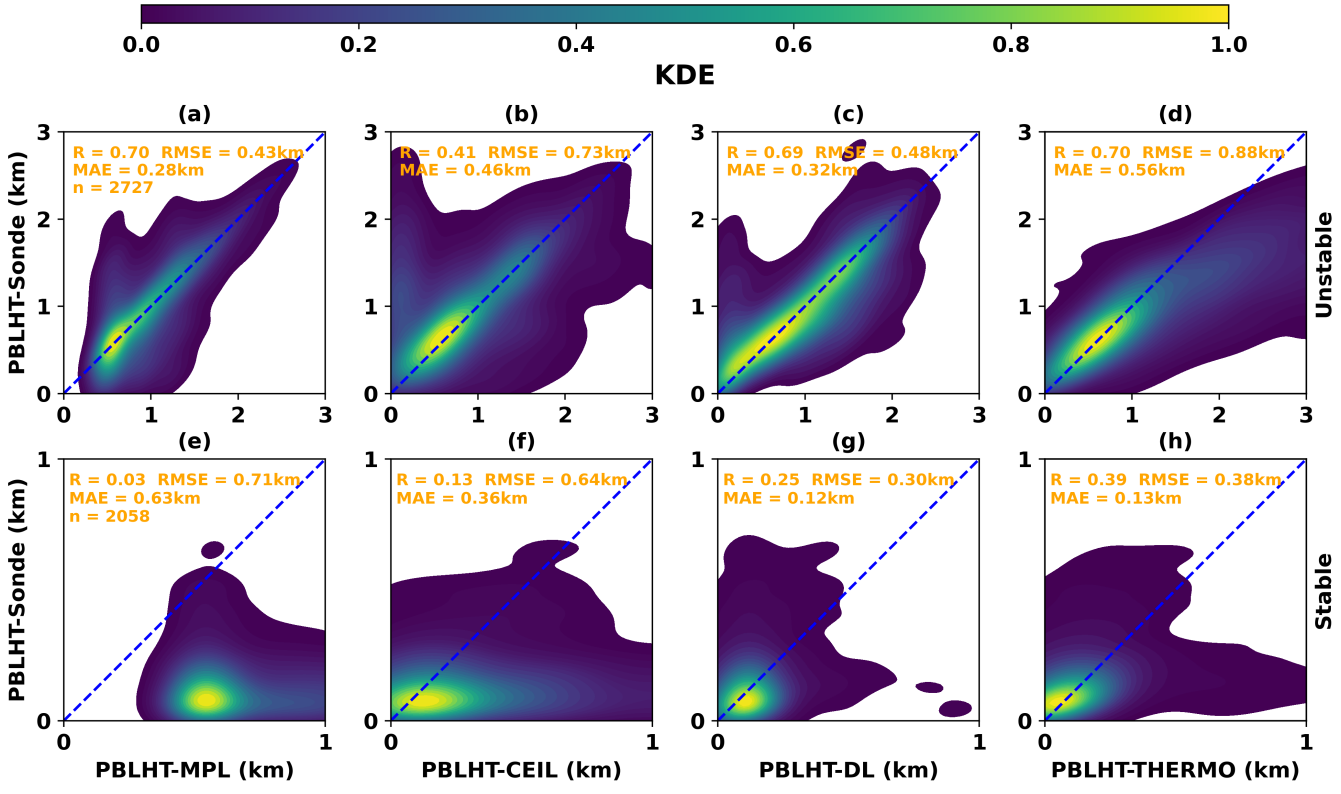


Figure 4. Comparisons of the four PBLHT estimates from remote sensing measurements with PBLHT-Sonde under unstable (a-d) and stable (e-h) PBL conditions. Note that different axis ranges are used for unstable and stable PBL conditions.

Under unstable PBL conditions, both PBLHT-MPL and PBLHT-DL show better performance, with relatively higher R, and lower RMSE and MAE values. PBLHT-MPL exhibits a smaller spread around PBLHT-Sonde but is unable to detect PBLHT values below 0.5 km. In contrast, during unstable conditions PBLHT-CEIL performs poorly, with the lowest R, and relatively higher RMSE and MAE values, due to a larger spread around PBLHT-Sonde. PBLHT-DL, while showing a good correlation with PBLHT-Sonde across all PBLHT ranges, has a larger spread compared to PBLHT-MPL for PBLHTs larger than 0.5 km.

https://doi.org/10.5194/egusphere-2024-3959
Preprint. Discussion started: 14 February 2025

PBLHT-THERMO performs well for PBLHT values below 1 km but significantly overestimates PBLHT for values above 1 km, leading to the highest RMSE and MAE values among the four PBLHT estimates.

The performance of the four PBLHT estimates from remote sensing measurements deteriorates significantly under stable PBL conditions (Krishnamurthy et al., 2021; Su et al., 2020; Zhang et al., 2022), with much lower R values, as shown in Figure 4e-h. The uncertainty in estimating PBLHTs from observations during stable atmospheric conditions has been documented in several studies (Su et al. 2020; Zhang et al. 2022). Reduced accuracy during nighttime is primarily attributed to the formation of SBL, which generates smaller thermal gradients and weaker turbulence near the surface. These features are difficult to detect with remote sensing instruments due to weak or noisy returns (in terms of backscatter or scattering). In addition, PBLHTs can be shallow during nighttime conditions, wherein the PBLHT is below the lowest range-gate for some of these remote sensing devices making it hard to detect. Among the four methods, during stable conditions, PBLHT-DL and PBLHT-THERMO perform better than PBLHT-MPL and PBLHT-CEIL. PBLHT-MPL shows poor performance, with the lowest R, and highest RMSE and MAE values, primarily due to its inability to detect PBLHT values below 0.5 km, while most PBLHT values under stable PBL conditions at SGP are below 0.5 km (Krishnamurthy et al., 2021). PBLHT-CEIL continues to exhibit a large spread around PBLHT-Sonde. PBLHT-DL has the lowest RMSE and MAE values, while PBLHT-THERMO achieves the highest R-value.

The above analysis reveals that no single remote sensing approach for estimating PBLHT consistently outperforms the others. Each method has unique strengths and limitations under different PBL conditions, primarily due to instrument constraints or retrieval uncertainties, as discussed in Section 1. However, these approaches are complementary, and integrating them can provide improved and more continuous PBLHT estimates.

3 Best Estimate PBLHT (PBLHT-BE)

3.1 PBLHT-BE from Remote Sensing Measurements (PBLHT-BE-Lidar)

To evaluate the hypothesis that combining PBLHT estimates from different remote sensing approaches can enhance PBLHT accuracy, we designed an idealized experiment. Among the four PBLHT estimates from remote sensing measurements, the closest to the 'ground truth'—PBLHT-Sonde in this study—can be selected as the PBLHT-BE. Since all remote sensing methods include lidar measurements, this PBLHT-BE is referred to as PBLHT-BE-Lidar. We manually searched for PBLHT-BE-Lidar from the four remote sensing estimated PBLHTs by comparing them against PBLHT-Sonde at each radiosonde launch time. Figure 5 compares PBLHT-BE-Lidar with PBLHT-Sonde, showing that, under both unstable and stable conditions, the R is significantly improved to 0.97 and 0.85, the RMSE is greatly reduced to 0.15 and 0.08 km, and the MAE is reduced to 0.10 and 0.04 km, respectively. The data distributions of PBLHT-BE-Lidar and PBLHT-Sonde largely fall within a narrower range relative to the 1:1 line. This suggests that carefully selecting the best estimate from multiple remote sensing measurements can significantly improve PBLHT accuracy compared to relying on a single remote sensing method. Please note

https://doi.org/10.5194/egusphere-2024-3959
Preprint. Discussion started: 14 February 2025

that this experiment provides PBLHT-BE-Lidar estimates only when PBLHT-Sonde data is available and cannot be applied to other time periods.

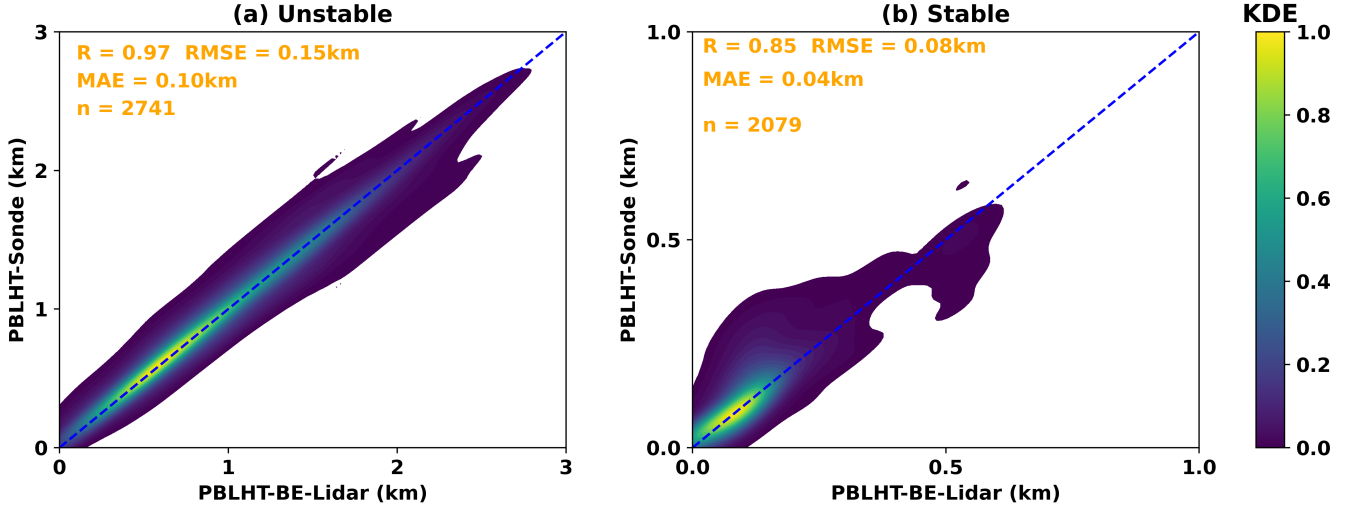


Figure 5. Similar as Figure 4., but for comparisons of PBLHT-BE-Lidar and PBLHT-Sonde.

Understanding when or under what conditions a specific PBLHT estimate from remote sensing measurements best aligns with PBLHT-Sonde and is selected as the PBLHT-BE-Lidar is valuable. Figure 6 presents histograms of the probability distributions of PBLHT estimates from remote sensing measurements selected as the PBLHT-BE-Lidar at different local times (Figure 6a) and PBL regimes (Figure 6b). The results show that during nighttime and early morning, before 10:00 LT, or under stable PBL conditions, PBLHT-THERMO is more frequently chosen as the PBLHT-BE-Lidar, followed by PBLHT-DL. After 12:00 LT, or under unstable PBL conditions, PBLHT-DL and PBLHT-MPL are more likely to be selected as the PBLHT-BE-Lidar. These findings indicate that PBLHT-THERMO is more accurate during nighttime, early morning, or stable PBL conditions, whereas PBLHT-MPL and PBLHT-DL perform better in the afternoon or under unstable conditions, consistent with the patterns shown in Figure 4.

https://doi.org/10.5194/egusphere-2024-3959
Preprint. Discussion started: 14 February 2025

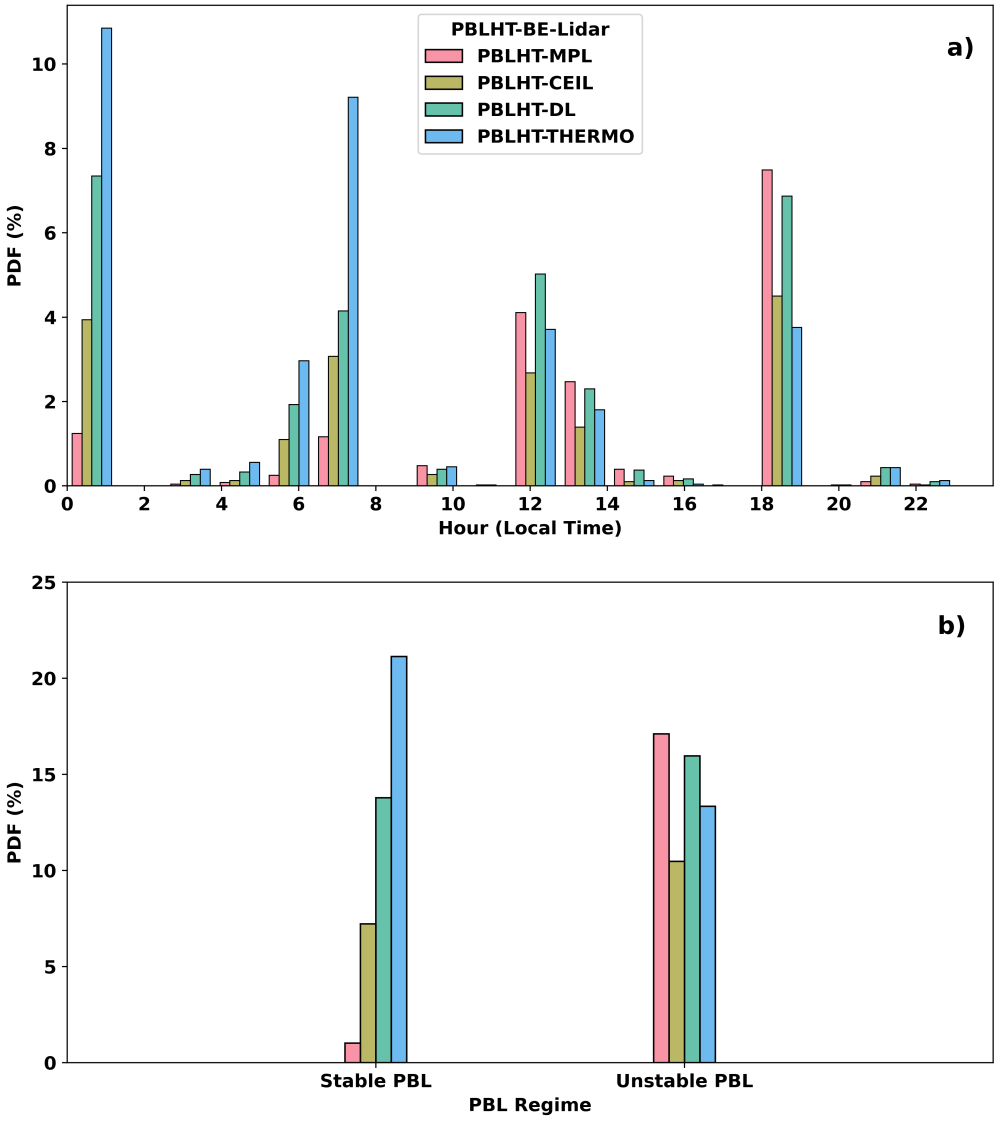


Figure 6. Histogram plots of probability distribution functions (PDFs) of PBLHT estimates from remote sensing measurements that are selected as the PBLHT-BE-Lidar at different LTs (a) and PBL regimes (b).

3.2 PBLHT-BE Using ML (PBLHT-BE-ML)

The discussion above highlights the need for an automated approach to derive more accurate PBLHT estimates from multiple remote sensing measurements when PBLHT-Sonde data are unavailable. This would allow for improved continuous PBLHT estimates, facilitating the study of PBL evolution and the evaluation of model simulations. To achieve this, we tested three machine learning models—RF classifier, RF regressor, and LightGBM—for predicting PBLHT. These models utilized four PBLHT estimates derived from remote sensing measurements along with environmental variables listed in Table 1.

https://doi.org/10.5194/egusphere-2024-3959
Preprint. Discussion started: 14 February 2025

3.2.1 Data Preparation and ML Models

Remote sensing, radiosonde, and surface measurements, as summarized in Table 1 and collected between 2017 and 2023, are used to train and test ML models. Because radiosondes are routinely launched every 6 hours, we find the closest remote sensing and surface data to radiosonde launch time to create the concurrent observation dataset. For ARM data products as listed in Table 1 and PBLHT estimates from remote sensing measurements, quality control (QC) flags are added for each variable. These QC flags are used to filter erroneous data or bad retrievals. Given that PBLHT estimates from the four lidar measurements are key input data, we also remove cases when a whole day of PBLHT estimates is missing from any lidar measurements. In total, 4,785 PBLHT-Sonde estimates from radiosonde measurements were used for the training process. Following Krishnamurthy et al. (2021), we refer to the variables listed in Table 1 as input features. Missing data, often caused by instrument malfunctions or failures, are not uncommon in datasets that integrate a wide range of observational data streams. To address this, we apply the Pandas Python library forward- and backward-fill methods to handle missing values. The forward-fill method replaces missing data with the last known value, while the backward-fill method uses the next valid value moving backwards. To minimize mismatches caused by large data gaps, forward-fill and backward-fill methods were applied only within one hour of missing data. Finally, we substitute the missing value with the average of the forward- and backward-filled values. Data standardization is commonly applied to enhance machine learning model performance (Sujon et al., 2024). However, random forest (RF) models and LightGBM are not sensitive to standardization or scaling (Breiman, 2001). As a result, we use the input features directly, without standardizing, for the training process.

The Scikit-Learn library for Python (Pedregosa et al., 2011) was used in this study for model hyperparameter tuning, training, and evaluation of machine learning algorithms. Hyperparameters are settings that govern the learning process of machine learning algorithms, influencing aspects such as model complexity, learning rate, number of layers, and regularization. Optimal hyperparameter selection is essential to maximize model performance. This process is automated using Scikit-Learn's GridSearchCV, which conducts cross-validated hyperparameter tuning to determine the optimal model configuration. Hyperparameters optimized in this study were the number trees in the forest (n_estimators), the maximum number of features to consider when looking for the best split (max_features), the maximum number of splits each tree can take (max_depth), and the maximum number of leaf nodes a single decision tree within the forest can have (max_leaf_nodes). Model training and testing data are split at random using Scikit-Learn's train_test_split function. 75% of the data are used for training and the rest for testing the model.

https://doi.org/10.5194/egusphere-2024-3959
Preprint. Discussion started: 14 February 2025

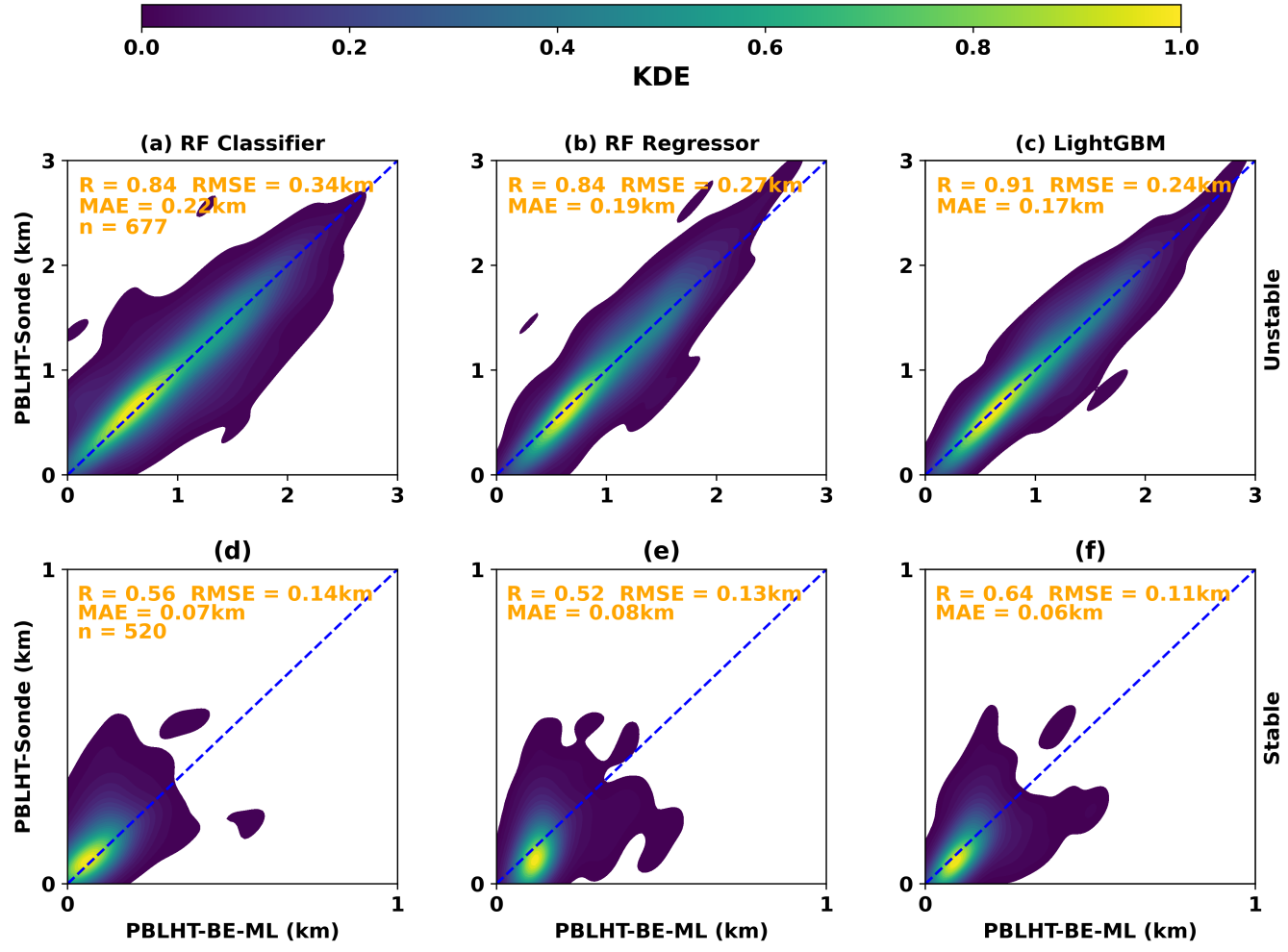


Figure 7. Similar as Figure 4, but for comparisons of predicted PBLHT-BE-ML from ML methods with PBLHT-Sonde.

We first train a RF classifier model to find the PBLHT-BE (PBLHT-BE-ML) as the PBLHT estimate that is the closest to PBLHT-Sonde from the four lidar-based PBLHTs. The RF classifier is an ensemble machine learning model that builds multiple decision trees during training and merges their results to improve classification accuracy and reduce overfitting (Breiman 2001). This ensemble ML approach helps improve the robustness and accuracy of predictions by reducing variance and leveraging the collective output of individual trees. Vertically resolved profiles of atmospheric state variables (e.g., temperature, water vapor, and vertical velocity) and tracers (e.g., aerosol particles) from remote sensing measurements can be used to reliably detect atmospheric features such as temperature and water vapor inversion layers, as well as aerosol layers. However, identifying which of these features is most closely related to PBLHT remains a challenge. The RF classifier model offers an effective solution to address this challenge. It has been widely used in many fields due to its high accuracy, flexibility, and ability to handle complex datasets. Figures 7a and 7d present evaluations of the predicted PBLHT-BE-ML from the RF classifier model compared with PBLHT-Sonde under both unstable and stable PBL conditions. When compared to the PBLHT

https://doi.org/10.5194/egusphere-2024-3959
Preprint. Discussion started: 14 February 2025

estimates derived from individual remote sensing measurements shown in Figure 4, the predicted PBLHT-BE-ML, which uses multi-remote sensing estimates, demonstrates a significantly higher correlation (R) and lower root mean square error (RMSE) against PBLHT-Sonde under both regimes. This indicates that the ML approach provides more accurate PBLHT estimates from combined remote sensing data than estimates based on single remote sensing measurements.

However, the R (RMSE) value between the predicted PBLHT-BE-ML from the trained RF Classifier model and PBLHT-Sonde is lower (larger) than that of the idealized PBLHT-BE-Lidar shown in Figure 5. To address this, we tested other ML models, including the RF Regressor and the Light Gradient-Boosting Machine (LightGBM) Regressor models. Similar to the RF Classifier model, the RF Regressor is an ensemble learning model that uses averaging to enhance predictive accuracy and control overfitting. The RF Regressor extends the principles of the RF Classifier to regression tasks by fitting multiple decision tree regressors on various sub-samples of the dataset to predict continuous outcomes (Breiman 2001). LightGBM is a widely used decision tree-based ML model known for its speed, efficiency, and high performance (Ke et al., 2017). Unlike other boosting algorithms, LightGBM constructs decision tree leaf-wise instead of level-wise, enabling it to achieve lower loss and higher accuracy. Figure 7b, 7c, 7e, and 7f show the evaluations of predicted PBLHT-BE-ML from the RF Regressor and LightGBM models compared with PBLHT-Sonde under both unstable and stable PBL conditions. Overall, the RF Regressor model performs similarly to the RF Classifier, with the predicted PBLHT-BE-ML from the RF Regressor showing a narrower data distribution relative to the 1:1 line and a lower RMSE compared to the RF Classifier under unstable PBL conditions (Figure 7b), indicating a slight performance advantage. However, the LightGBM Regressor model outperforms both RF models, demonstrating the highest R and lowest RMSE values in evaluations against PBLHT-Sonde for both unstable and stable PBL conditions (Figure 7c and 7f). The LightGBM model's predicted PBLHT-BE-ML shows substantially better correlations with PBLHT-Sonde than the RF models and approaches the ideal PBLHT-BE depicted in Figure 5.

Understanding whether the significant improvement in predicted PBLHT-BE-ML is due to the use of ML methods or the combination of various PBLHT estimates is important. Figures S3 and S4 show evaluations of predicted PBLHT using RF and LightGBM regressors based on individual remote sensing PBLHT estimates. The results indicate that ML models applied to individual remote sensing PBLHT estimates can also lead to substantial improvements in PBLHT prediction. However, there are several advantages to using multiple remote sensing data sources with ML models: (1) it can further enhance predicted PBLHT accuracy, as evidenced by comparing Figures 7 with Figures S3 and S4; (2) it allows for more reliable PBLHT predictions during periods outside of routine radiosonde launch times, as will be discussed in section 3.2.3; and (3) it enables consistent and reliable PBLHT predictions across different geographic regions, as will be shown in section 3.2.4.

3.2.2 Feature Importance Analysis

Input features for the ML models, along with their abbreviations and source ARM datastreams, are detailed in Table 1. We did not include a local time parameter in the input features because the parameter changes with seasons and may have different

https://doi.org/10.5194/egusphere-2024-3959
Preprint. Discussion started: 14 February 2025

relations with PBLHT at different locations, which could bring issues when using the ML prediction at other geographic locations. The quality and quantity of input features can significantly affect the model's accuracy and efficiency. Therefore, understanding the importance of input features is crucial for optimizing ML model performance. Scikit-Learn's random forest models have built-in feature importance metrics that help identify which input features contribute most to the model's predictive power. This feature importance is calculated based on the average reduction in impurity each time a feature is used to split a node across all trees in the forest. Features causing substantial decreases in impurity are assigned higher importance scores and deemed more significant. These scores are normalized so that they sum to 1. The LightGBM model offers two main types of feature importance: "Split" and "Gain." In this study, we use the Gain importance, which measures the improvement in the model's accuracy when a specific feature is used for splitting.

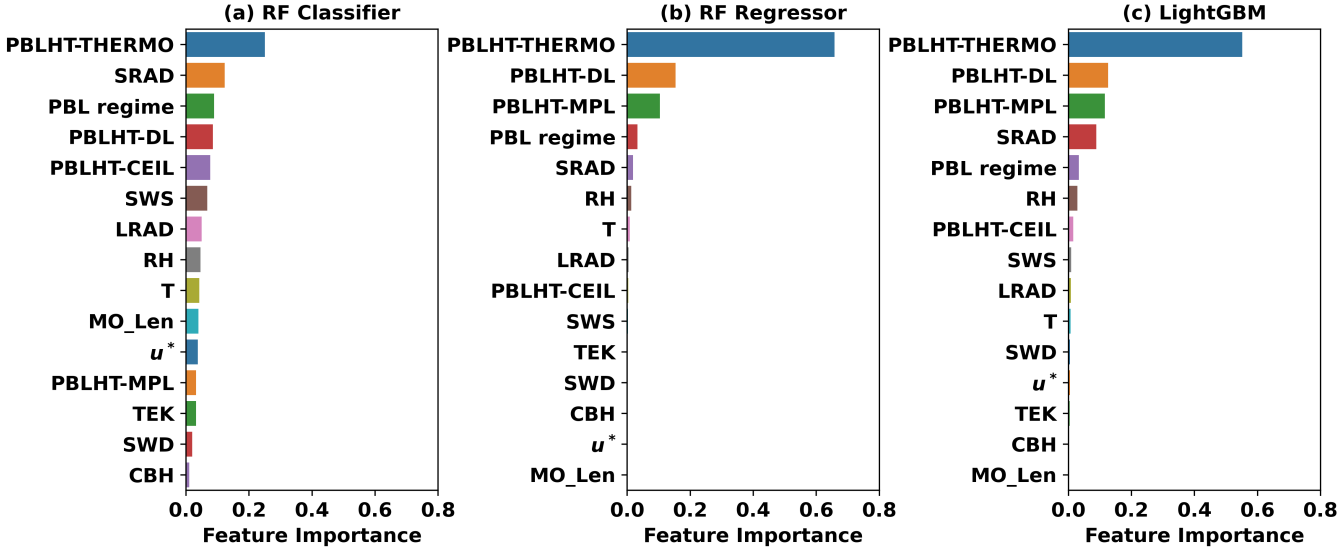


Figure 8. Feature importance for the three ML models at the training stage: a) RF Classifier; b) RF Regressor; c) LightGBM. Feature importance scores are normalized so that they sum up to 1.

Figure 8 shows the feature importance for the three ML models under all conditions. PBLHT-THERMO emerges as the most significant feature across all three models, likely because it compares best with PBLHT-Sonde under stable PBL conditions and during nighttime or early morning, as demonstrated in Figures 5 and 6. In the RF Classifier, each input feature contributes to the model's decision-making, as the prediction involves selecting the most appropriate PBLHT estimate from the four remote sensing-based estimates. Each input feature can influence this selection to varying degrees under different environmental conditions. In contrast, PBLHT-THERMO, PBLHT-DL, and PBLHT-MPL are the top three most important features for the RF and LightGBM Regressor models. This is because these models aim to predict PBLHT by interpolating or extrapolating from the input features to produce an estimate that aligns closely with PBLHT-Sonde. Consequently, other atmospheric or surface parameters play a less role in these models. PBLHT-CEIL, which uses aerosols as tracers similar to PBLHT-MPL, has a smaller impact because CEIL is less sensitive to aerosols and typically has a lower signal-to-noise ratio compared to MPL.

https://doi.org/10.5194/egusphere-2024-3959
Preprint. Discussion started: 14 February 2025

As a result, PBLHT-CEIL estimates are generally less accurate and play a minor role in the RF and LightGBM Regressor models. Additionally, comparisons of feature importance between daytime and nighttime reveal that PBLHT-DL and PBLHT-MPL are the most influential features during daytime, whereas PBLHT-THERMO dominates during nighttime for RF Regressor and LightGBM model predictions (Figure S5).

3.2.3 Applying PBLHT-BE ML to Continuous Remote Sensing Measurements

Once the ML models were trained and tested using PBLHT-Sonde data, they were applied to predict high-temporal-resolution PBLHT-BE-ML using PBLHT estimates from various remote sensing measurements. Since PBLHT-THERMO has the coarsest temporal resolution (10 minutes) among the four remote sensing PBLHT estimates, its time dimension was used as the reference for the predicted PBLHT-BE-ML. All other input features were aligned to this time dimension using the 'nearest' data mapping principle. Figure 9 presents an example of PBLHT estimates from individual remote sensing measurements and ML model predictions for July 12, 2019, at the ARM SGP site, with PBLHT-Sonde estimates included for comparison. On this day, radiosondes were launched eight times due to an intensive observation period (IOP) for testing a tethered balloon system at SGP. The PBLHT estimates from individual remote sensing measurements exhibit similar issues to those seen in the May 8, 2017, case shown in Figure 3. The predicted PBLHT-BE-ML from the three ML models aligns well with all eight PBLHT-Sonde estimates and displays a smooth, complete diurnal evolution of the PBLHT, as shown in Figure 9b. The RF Classifier model, however, shows abrupt jumps during the PBL growth period around 20:00 and 09:00 LT, which is probably caused the transition of PBL regimes and/or quick changes in SRAD. In contrast, the RF Regressor and LightGBM Regressor models demonstrate smooth PBL growth during these periods.

https://doi.org/10.5194/egusphere-2024-3959
Preprint. Discussion started: 14 February 2025

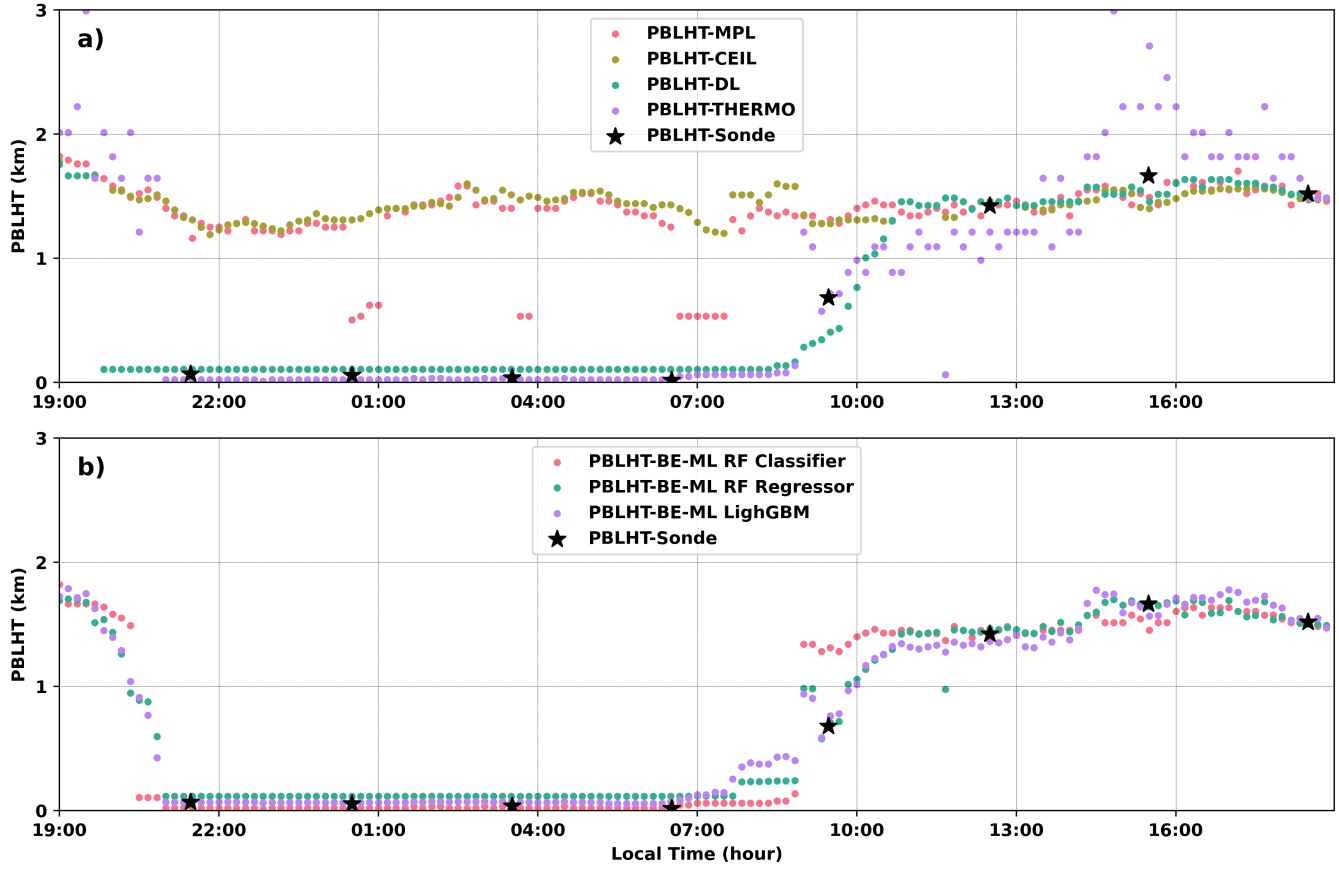


Figure 9. PBLHT estimates on July 12, 2019, at the ARM SGP site: a) PBLHT estimates from individual remote sensing measurements and PBLHT-Sonde; b) PLBHT-BE-ML prediction from the three ML models and PBLHT-Sonde.

High-temporal-resolution PBLHT estimates are crucial for studying the daily diurnal evolution of PBLHT as well as its seasonal variations. Figure 10 shows box and whisker plots of PBLHT diurnal cycles and their seasonal variations from the three ML model predictions and from PBLHT-Sonde at the ARM SGP observatory using data between 2017 and 2023. As expected, the comparison between PBLHT-BE-ML and PBLHT-Sonde are very well because 75% of the data were used to train the ML models and the validation using the remaining 25% data show good testing results as presented in section 3.2.1. Noticeable differences between PBLHT-BE-ML and PBLHT-Sonde occur at about 14:00 and 15:00 LT in Fall as shown in Figure 10 c, which is probably caused by small samples from a couple of IOPs. Clear PBLHT diurnal evolutions at all seasons are observed from all PBLHT estimation methods at the ARM SGP observatory, revealing a typical PBLHT diurnal evolution over midlatitude land surface. The PBLHT remains shallow throughout the nighttime. It begins to grow around 9:00 a.m. local time, reaches its peak in the late afternoon, and start to decay. Summer exhibits the highest convective PBLHT during the day and the lowest nighttime PBLHT compared to other seasons, likely due to the strongest daytime shortwave surface heating

https://doi.org/10.5194/egusphere-2024-3959
Preprint. Discussion started: 14 February 2025

and the most intense nighttime longwave radiative cooling in this season. As a comparison, PBLHT diurnal cycles and their seasonal variations from PBLHT-MPL, PBLHT-CEIL, PBLHT-DL, and PBLHT-THERMOD are shown in Figure S6.

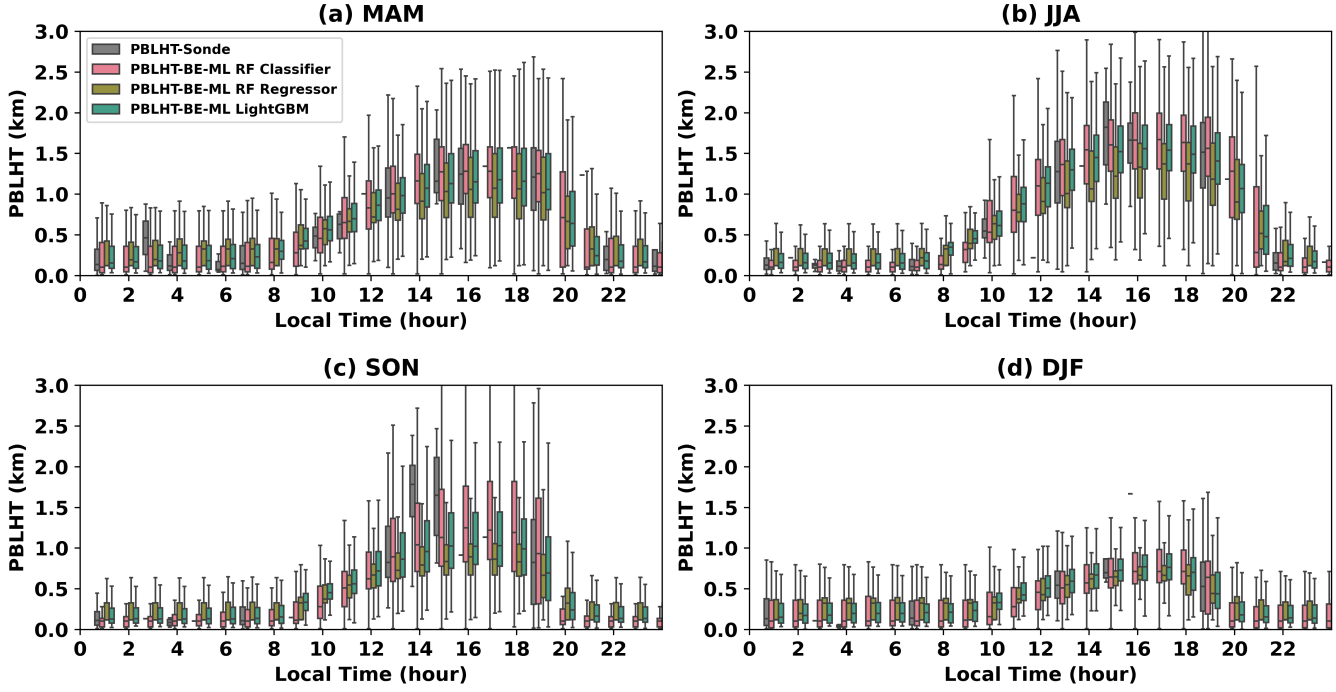


Figure 10. PBLHT diurnal cycles and their seasonal variations from the three ML model predictions and from PBLHT-Sonde at the ARM SGP observatory. MAM (March-April-May) represents the Spring season, JJA (Jun-July-August) for Summer, SON (September-October-November) for Fall, and DJF (December-January-February) for Winter. Horizontal bars, boxes and whiskers represent the median, interquartile range and range of the data.

The predicted PBLHT-BE-ML from the three ML models generally align well; however, they show notable differences under afternoon convective PBL conditions between 14:00 and 18:00 local time during summer and fall. The RF Classifier predicts the highest PBLHT, while the RF Regressor predicts the lowest. Unfortunately, we could not directly evaluate which model performs better due to the lack of routine radiosonde launches during this period. This highlights a general challenge for ML-based methods: how well do they perform when the target conditions differ from the training conditions? This period is characterized by strong turbulence within the PBL. Under such conditions, PBLHT estimates from aerosol lidars and Doppler lidars (DL) are considered reliable (Kotthaus et al., 2023) and show smaller relative standard deviations among different methods, as seen in Figure S2b. Therefore, we can use lidar-derived PBLHT estimates to evaluate the ML model predictions. Similar to the method we used for different PBLHT estimates from radiosonde data, we calculate the median of PBLHT estimates from PBLHT-MPL, PBLHT-CEIL, and PBLHT-DL, referring to it as PBLHT-Lidar. PBLHT-THERMO is excluded because it clearly overestimates PBLHT under strong convective conditions due to large uncertainties in temperature retrievals

https://doi.org/10.5194/egusphere-2024-3959
Preprint. Discussion started: 14 February 2025

from RL measurements, as shown in Figure 4d. We assume that PBLHT-Lidar can be regarded as the "true" PBLHT under strong convective conditions and used to evaluate predicted PBLHT-BE-ML.

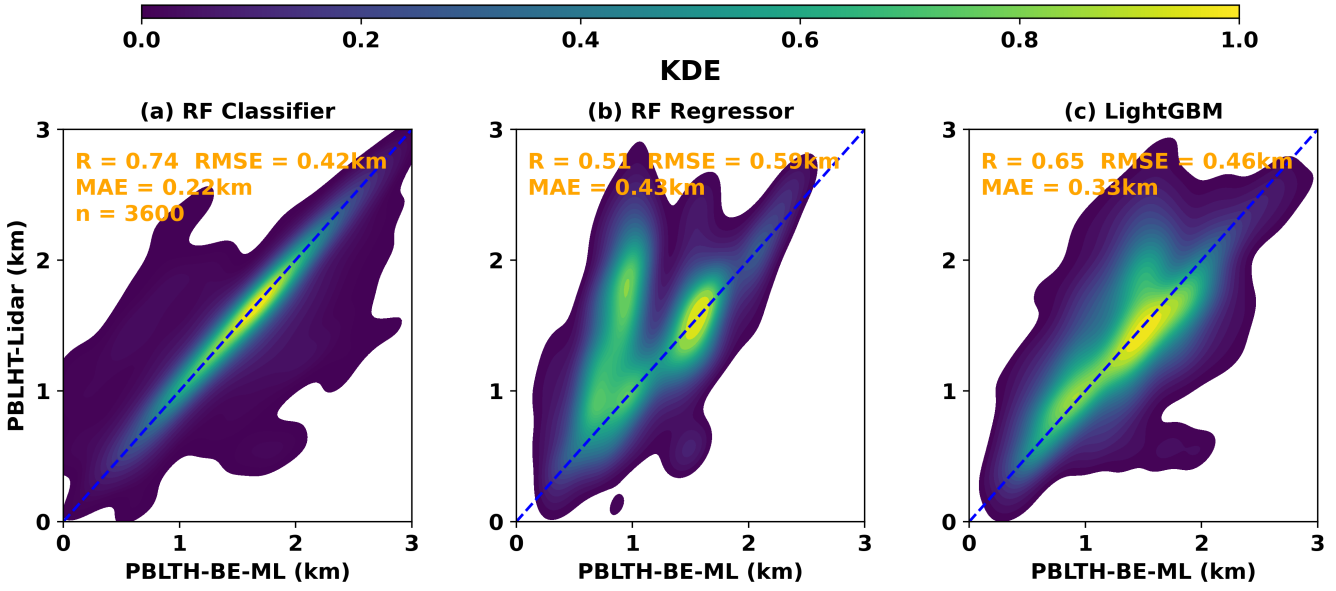


Figure 11. Similar as Figure 4, except for comparisons of predicted PBLHT-BE-ML against PBLHT-Lidar from: a) RF Classifier; b) RF Regressor; c) LightGBM under afternoon convective PBL conditions between 14:00 and 18:00 local time during summer and fall at the ARM SGP site.

Figure 11 compares the predicted PBLHT-BE-ML from the three ML models with PBLHT-Lidar. As expected, the agreement between predicted PBLHT-BE-ML and PBLHT-Lidar is not as strong as during the radiosonde launch periods illustrated in Figure 7. Among the three ML models, the RF Classifier's predicted PBLHT-BE-ML shows the best agreement with PBLHT-Lidar, with the highest R and lowest RMSE. The RF Classifier's predictions closely align with the PBLHT-Lidar along the 1:1 line, whereas the RF Regressor and LightGBM predictions tend to be lower than both PBLHT-Lidar and the RF Classifier predictions, consistent with Figures 10b and 10c. Additionally, the RF Classifier's predicted PBLHT-BE-ML outperforms the best comparison between individual lidar PBLHT estimates and PBLHT-Lidar, as shown in Figure S7. To conclude, the performance of predicted PBLHT-BE-ML, particularly from regression-based ML models, may decrease when applied to periods that differ from the training data periods. More training data under afternoon convective PBL conditions are needed to improve ML model predictions. However, the classifier-based model remains effective in identifying the candidate PBLHT as the PBLHT-BE, providing enhanced predictions during afternoon convective PBL conditions.

3.2.4 Evaluation of Predicted PBLHT-BE-ML at a Different ARM Site

To further evaluate the performance of the ML models in predicting PBLHT-BE, we compared the predicted PBLHT-BE-ML from the three ML models with PBLHT-Sonde data at a different ARM Mobile Facility (AMF) observatory. The ARM Eastern

https://doi.org/10.5194/egusphere-2024-3959
Preprint. Discussion started: 14 February 2025

Pacific Cloud Aerosol Precipitation Experiment (EPCAPE) field campaign deployed an AMF at the Scripps Memorial Pier in La Jolla, California (the EPC site) from February 15, 2023, to February 14, 2024 (https://www.arm.gov/research/campaigns/amf2023epcape). This field campaign focuses on characterizing the extent, radiative properties, aerosol interactions, and precipitation characteristics of stratocumulus clouds in the Eastern Pacific at a coastal location, making PBLHT a critical parameter for understanding these processes. The EPC site is dominated by low-level marine stratocumulus clouds year-round, representing markedly different meteorological conditions compared to the SGP site. The deployment includes most instruments listed in Table 1, except for the RL. Consequently, PBLHT-MPL, PBLHT-CEIL, and PBLHT-DL for EPCAPE are derived in the same manner as at the SGP site. However, PBLHT-THERMO is derived solely using TROPoe data.

Turbulence over a warm ocean surface is generally weaker and exhibits less diurnal variability than over land, making it more challenging to obtain reliable PBLHT estimates in marine environments. Figure 12 presents comparisons of PBLHT estimates from various remote sensing measurements against PBLHT-Sonde at the EPC site. These comparisons are noticeably weaker than those observed at the SGP site (Figure 4), as indicated by significantly lower correlation coefficients (R) at the EPC site. Additionally, PBLHT values under unstable PBL conditions at the EPC site are markedly lower than those at the SGP site, likely due to the limited thermal energy available for convection over the ocean. Among the different remote sensing PBLHT estimates, PBLHT-DL shows the strongest agreement with PBLHT-Sonde data under both unstable and stable PBL conditions, suggesting that PBLHT-DL may represent a reliable estimate for marine environments like the EPC site. PBLHT-THERMO performs best under stable PBL conditions at the SGP site; however, it consistently overestimates PBLHT compared to PBLHT-Sonde at the EPC site. This discrepancy is likely due to higher uncertainties in potential temperature profile retrievals from TROPoe and the less distinct gradient in the potential temperature profile under stable (cloudy) PBL conditions at the EPC site.

https://doi.org/10.5194/egusphere-2024-3959
Preprint. Discussion started: 14 February 2025

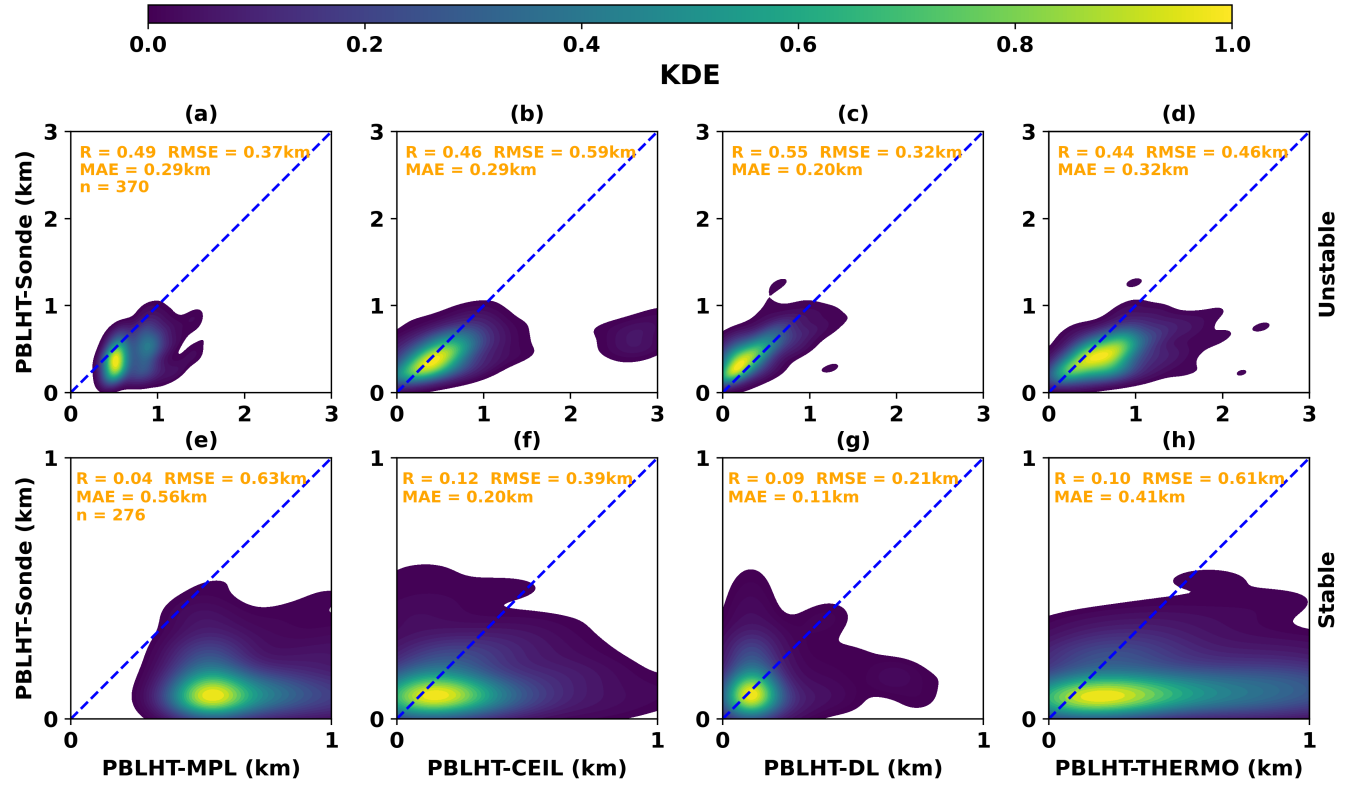


Figure 12. Similar as Figure 4, except for the EPC site.

We applied the three trained ML models directly to EPC data to evaluate their performance against PBLHT-Sonde. Figure 12 presents evaluations of the predicted PBLHT-BE-ML from the three ML models compared to PBLHT-Sonde under both unstable and stable PBL conditions. Compared to the PBLHT estimates from individual remote sensing measurements shown in Figure 12, the predicted PBLHT-BE-ML from all three ML models shows improvement under both unstable and stable PBL conditions. Although the RF Classifier and Regressor models do not show an increase in R compared to PBLHT-DL under unstable conditions, they do demonstrate a notable reduction in RMSE and MAE. Additionally, both models show improvements in R, RMSE, and MAE under stable conditions. Among the three ML models, LightGBM exhibits the best performance against PBLHT-Sonde, with significantly higher R, and lower RMSE and MAE than both the RF Classifier and Regressor, as well as the individual PBLHT estimates from remote sensing measurements under both unstable and stable PBL conditions. However, it should be noted that the performance of the ML models at the EPC site is not as strong as at the SGP site. Expanding the training dataset to include data from different surface types, such as warm ocean surfaces, ice, and snow-covered areas, could further enhance the ML model's performance across diverse locations.

https://doi.org/10.5194/egusphere-2024-3959
Preprint. Discussion started: 14 February 2025

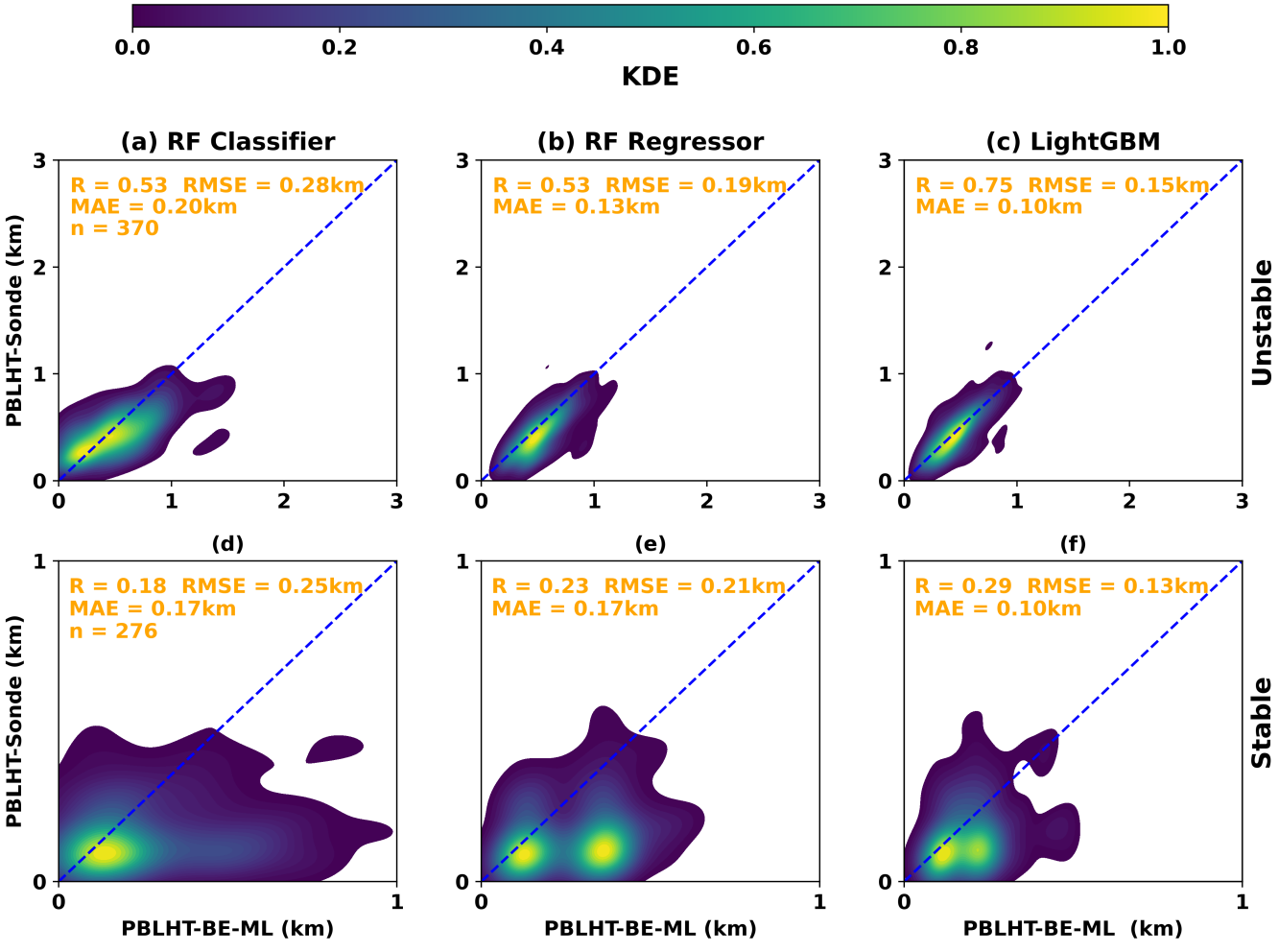


Figure 13. Similar as Figure 7, except for the EPC site.

4 Summary and Discussions

The planetary boundary layer height (PBLHT) is commonly determined using radiosonde data and remote sensing measurements. PBLHT estimates from radiosondes are generally considered more reliable and are commonly used to validate estimates from remote sensing. The Department of Energy (DOE) Atmospheric Radiation Measurement (ARM) program provides PBLHT estimates from radiosonde data through its PBLHT-SONDE Value-Added Product (VAP), which includes three methods: 1) the Heffter method, 2) the Liu-Liang method, and 3) the bulk Richardson method with critical thresholds of 0.25 and 0.5. However, radiosonde data suffers from low temporal resolution and is subject to sampling error. In contrast, lidar and radiometer remote sensing instruments offer high-temporal-resolution, continuous PBLHT estimates. ARM provides various PBLHT estimates from remote sensing measurements, including: 1) PBLHT-MPL, derived from the wavelet covariance of the Micro-pulse Lidar (MPL) backscatter profile (Sawyer and Li, 2013); 2) PBLHT-CEIL, based on the

https://doi.org/10.5194/egusphere-2024-3959
Preprint. Discussion started: 14 February 2025

VAISALA CL31's enhanced gradient method; 3) PBLHT-DL, based on vertical velocity variance measured by Doppler Lidar, using a threshold of 0.04 m^2/s^2 (Tucker et al., 2016); and 4) PBLHT-THERMO, derived from Raman Lidar (RL) and Atmospheric Emitted Radiance Interferometer (AERI) temperature and water vapor profiles using the Heffter method. Each remote sensing approach has its own strengths and limitations. To achieve reliable PBLHT estimates throughout the day and under varying boundary layer conditions, we trained machine learning (ML) models on the PBLHT-SONDE VAP to produce the best-estimate PBLHT (PBLHT-BE) from multiple remote sensing measurements at the ARM SGP site.

Comparisons of the three PBLHT estimates from the PBLHT-SONDE VAP reveal substantial differences across various PBL conditions. To address this variability, we use their median as the 'ground truth' for training and testing the ML models, referring to it as PBLHT-Sonde. Evaluations of PBLHT estimates from individual remote sensing methods at the ARM Southern Great Plains (SGP) site indicate that both PBLHT-MPL and PBLHT-DL perform well under unstable conditions, with higher correlation coefficients (R) and lower Root Mean Square Error (RMSE). PBLHT-THERMO performs accurately for PBLHT values below 1.5 km but significantly overestimates values above 1.5 km. Under stable PBL conditions, the accuracy of PBLHT estimates from remote sensing measurements by all methods decreases substantially but varies among different methods of mixed performance at different time of a day, with PBLHT-DL and PBLHT-THERMO showing slightly better performance than PBLHT-MPL and PBLHT-CEIL especially at nighttime and early morning.

We combine the four PBLHT estimates from remote sensing measurements to identify the PBLHT estimate (referred to as PBLHT-BE-Lidar) that best aligns with PBLHT-Sonde. In an ideal scenario, where the PBLHT estimate closest to PBLHT-Sonde is accurately selected as PBLHT-BE-Lidar, the comparison with PBLHT-Sonde improves significantly. During the nighttime, early morning, or under stable PBL conditions, PBLHT-THERMO is more frequently selected as the PBLHT-BE-Lidar. In the afternoon or under unstable PBL conditions, PBLHT-DL and PBLHT-MPL are more commonly chosen as the PBLHT-BE-Lidar at the SGP site.

Automated approaches using ML methods were tested to derive PBLHT-BE (PBLHT-BE-ML). Remote sensing, radiosonde, and surface measurements spanning 2017 to 2023 were utilized for training and testing the ML models. A total of 4,785 PBLHT-Sonde estimates from radiosonde measurements were included in the training (75%) and testing (25%) processes. We tested three ML models: a Random Forest (RF) Classifier, an RF Regressor, and LightGBM. All three models demonstrated improved alignment of PBLHT-BE-ML with PBLHT-Sonde, yielding higher R, and lower RMSE and MAE values compared to PBLHT estimates from individual remote sensing measurements. LightGBM, in particular, demonstrated the best performance against PBLHT-Sonde for both unstable and stable PBL conditions. Feature analysis for these models revealed that PBLHT-THERMO is the most significant feature across all three, with PBLHT-DL and PBLHT-MPL also ranking as important features for the RF and LightGBM Regressor models.

https://doi.org/10.5194/egusphere-2024-3959
Preprint. Discussion started: 14 February 2025

The trained ML models were then applied to various lidar remote sensing measurements to predict high-temporal-resolution PBLHT-BE-ML. An example from an intensive observation period (IOP) day shows that the predicted PBLHT-BE-ML from all three models aligns well with the eight PBLHT-Sonde estimates, capturing a smooth, complete diurnal evolution of PBLHT. Seasonal analysis of the diurnal evolution reveals that summer has the largest diurnal PBLHT variation, while winter shows the smallest. The predicted PBLHT-BE-ML from the three ML models are generally consistent, except for noticeable differences under afternoon convective PBL conditions between 14:00 and 18:00 local time in summer and fall. Due to a lack of routine radiosonde launches during this period, we used the median of PBLHT-MPL, PBLHT-CEIL, and PBLHT-DL as the "true" PBLHT, referred to as 'PBLHT-Lidar,' for evaluating ML model predictions. The results indicate that the RF Classifier predictions align closely with PBLHT-Lidar along the 1:1 line, while the RF Regressor and LightGBM predictions tend to be slightly lower than both PBLHT-Lidar and RF Classifier predictions, suggesting that additional training data under afternoon convective PBL conditions could enhance ML model accuracy.

To further assess model performance, we applied the trained ML models to remote sensing measurements from the ARM Eastern Pacific Cloud Aerosol Precipitation Experiment (EPCAPE) field campaign at the EPC site. Due to weaker turbulence over the warm ocean surface, obtaining reliable PBLHT estimates at the EPC site is more challenging. PBLHT estimates from individual remote sensing measurements show significantly lower correlation (R) at the EPC site than at the SGP site. However, the predicted PBLHT-BE-ML from all three ML models demonstrates improvement under both unstable and stable PBL conditions, with LightGBM showing the best agreement with PBLHT-Sonde. Nonetheless, model performance at the EPC site is not as robust as at the SGP site, highlighting the need to expand the training dataset to include data from diverse surface types, such as ocean, ice, and snow-covered surfaces.

https://doi.org/10.5194/egusphere-2024-3959
Preprint. Discussion started: 14 February 2025

Data availability

Remote sensing, surface, and radiosonde measurements, the PBLHT-SONDE VAP, and PBLHT-CEIL data from the ARM SGP central facility and from the EPCAPE field campaign used in this study can be directly downloaded from the ARM data discovery website: https://www.archive.arm.gov/discovery/. PBLHT-MPL, PBLHT-DL, and PBLHT-THERMO at the SGP central facility are also available at from the ARM data discovery website. These PBLHT estimates at the EPC site and model predicted PBLHT data are available upon request and will be available from the ARM data discover in near future.

Author contributions

Conceptualization, D.Z. and J.C; methodology, D.Z.; software, C.S. and K.M.; validation, D.Z.; formal analysis, D.Z., J.C.; investigation, D.Z.; resources, D.Z.; data curation, D.Z., C.S.; writing—original draft preparation, D.Z.; writing—review and editing, all co-authors; visualization, D.Z.; supervision, D.Z.; project administration, D.Z.; funding acquisition, J.C.. All authors have read and agreed to the published version of the manuscript.

Competing interests

The authors declare that they have no conflict of interest.

Acknowledgements

Data were obtained from the ARM user facility, a U.S. DOE Office of Science user facility managed by the Biological and Environmental Research (BER) program. This research was supported by the DOE ARM program.

https://doi.org/10.5194/egusphere-2024-3959
Preprint. Discussion started: 14 February 2025

References

Breiman, L.: Random Forests, Mach. Learn., 45, 5–32, https://doi.org/10.1023/A:1010933404324, 2001.

Brooks, I. M.: Finding Boundary Layer Top: Application of a Wavelet Covariance Transform to Lidar Backscatter Profiles, J. Atmos. Oceanic Technol., 20, 1092–1105, https://doi.org/10.1175/1520-0426(2003)020<1092:FBLTAO>2.0.CO;2, 2003.

Caicedo, V., Delgado, R., Sakai, R., Knepp, T., Williams, D., Cavender, K., Lefer, B., and Szykman, J.: An automated common algorithm for planetary boundary layer retrievals using aerosol lidars in support of the u.S. epa photochemical assessment monitoring stations program. J. Atmos. Oceanic Technol., *37*(10), 1847–1864. https://doi.org/10.1175/JTECH-D-20-0050.1, 2020.

Campbell, J. R., Hlavka, D. L., Welton, E. J., Flynn, C. J., Turner, D. D., Spinhirne, J. D., Scott, V. S., and Hwang, I. H.: Full-time, Eye-Safe Cloud and Aerosol Lidar Observation at Atmospheric Radiation Measurement Program Sites: Instrument and Data Processing, J. Atmos. Ocean. Tech., 19, 431–442, https://doi.org/10.1175/1520-0426(2002)019<0431:FTESCA>2.0.CO;2, 2002.

Contini, D., Cava, D., Martano, P., Donateo, A., and Grasso, F.: Boundary layer height estimation by sodar and sonic anemometer measurements. IOP Conference Series: Earth and Environmental Science, IOP Publishing, 012034, 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 Sens, 11, 1590, https://doi.org/10.3390/rs11131590, 2019.

Deardorff, J. W.: Three-dimensional numerical study of the height and mean structure of a heated planetary boundary layer. Boundary-Layer Meteorology, 7, 81-106, 1974.

de Arruda Moreira, G., Sánchez-Hernández, G., Guerrero-Rascado, J. L., Cazorla, A., and Alados-Arboledas, L.: Estimating the urban atmospheric boundary layer height from remote sensing applying machine learning techniques, Atmos. Res., 266, 105962, https://doi.org/10.1016/j.atmosres.2021.105962, 2022.

Heffter J. L.: Transport Layer Depth Calculations, Second Joint Conference on Applications of Air Pollution Meteorology, New Orleans, Louisiana, 1980.

Holdridge, D.: Balloon-Borne Sounding System (SONDE) Instrument Handbook, Atmospheric Radiation Measurement, U.S. Department of Energy Office of Science, DOE/SC-ARM/TR-029, https://doi.org/10.2172/1020712, 2020.

Holtslag, A. A. M., De Bruijn, E. I. F., and Pan, H.: A High Resolution Air Mass Transformation Model for Short-Range Weather Forecasting, Mon. Weather Rev., 118, 1561–1575, https://doi.org/10.1175/1520-0493(1990)118<1561:AHRAMT>2.0.CO;2, 1990.

Kalmus, P., Ao, C. O., Wang, K. N., Manzi, M. P., and Teixeira, J.: A high-resolution planetary boundary layer height seasonal climatology from GNSS radio occultations, Remote Sens. Environ., 276, 113037, https://doi.org/10.1016/j.rse.2022.113037, 2022.

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q. and Liu, T.Y.: Lightgbm: A highly efficient gradient boosting decision tree. Advances in neural information processing systems, 30, 2017.

Kotthaus, S., Bravo-Aranda, J. A., Collaud Coen, M., Guerrero-Rascado, J. L., Costa, M. J., Cimini, D., O'Connor, E. J., Hervo, M., Alados-Arboledas, L., Jiménez-Portaz, M., Mona, L., Ruffieux, D., Illingworth, A., and Haeffelin, M.: Atmospheric boundary layer height from ground-based remote sensing: a review of capabilities and limitations, Atmos. Meas. Tech., 16, 433–479, https://doi.org/10.5194/amt-16-433-2023, 2023.

Krishnamurthy, R., Newsom, R. K., Berg, L. K., Xiao, H., Ma, P.-L., and Turner, D. D.: On the estimation of boundary layer heights: a machine learning approach, Atmos. Meas. Tech., 14, 4403–4424, https://doi.org/10.5194/amt-14-4403-2021, 2021.

Kyrouac, J, and M Tuftedal.: Surface Meteorological System (MET) Instrument Handbook. U.S. Department of Energy, Atmospheric Radiation Measurement user facility, Richland, Washington. DOE/SCARM-TR-086, https://doi.org/10.2172/1007926, 2024.

LeMone, M. A., and Coauthors: 100 Years of Progress in Boundary Layer Meteorology, Meteor. Monogr., 59 9.1–9.85. doi: https://doi.org/10.1175/AMSMONOGRAPHS-D-18-0013.1, 2018.

Lewis, E. R.: Marine ARM GPCI Investigation of Clouds (MAGIC) field campaign report, Ed. by Robert Stafford, DOE ARM Climate Research Facility. DOE/SC-ARM-16-057. doi: https://doi.org/10.2172/1343577, 2016.

Liu, S., and Liang, X.: Observed Diurnal Cycle Climatology of Planetary Boundary Layer Height, J. Climate, **23**, 5790–5809, https://doi.org/10.1175/2010JCLI3552.1, 2010.

Liu, Z., Chang, J., Li, H., Chen, S., and Dai, T.: Estimating boundary layer height from lidar data under complex atmospheric conditions using machine learning, Remote Sens., 14, 418, https://doi.org/10.3390/rs14020418, 2022.

Long, C. N. and Shi, Y.: The QCRad Value Added Product: Surface Radiation Measurement Quality Control Testing, Including Climatology Configurable Limits, Office of Biological and Environmental Research, U.S. Department of Energy, 69 pp., 2006.

Luo, T., Wang, Z., Zhang, D., and Chen, B.: Marine boundary layer structure as observed by A-train satellites, Atmos. Chem. Phys., 16, 5891–5903, https://doi.org/10.5194/acp-16-5891-2016, 2016.

Münkel, C., and Räsänen, J.: New optical concept for commercial lidar ceilometers scanning the boundary layer, Proc. SPIE 5571, Remote Sensing of Clouds and the Atmosphere IX, https://doi.org/10.1117/12.565540, 2004.

https://doi.org/10.5194/egusphere-2024-3959
Preprint. Discussion started: 14 February 2025

Münkel, C. and Roininen, R.: Automatic monitoring of boundary layer structures with ceilometers, Vaisala News 184, 2010.

Münkel, C., Eresmaa, N., Räsänen, J., and Karppinen, A.: Retrieval of mixing height and dust concentration with lidar ceilometer, Bound.-Lay. Meteorol. 124, 117–128, https://doi.org/10.1007/s10546-006-9103-3, 2007.

Münkel, C., Roininen, R.: Automatic Monitoring of Boundary Layer Structures with Ceilometer. vol. 184 Vaisala News., 2010.

Nelson, K. J., Xie, F., Ao, C. O., and Oyola-Merced, M. I.: Diurnal Variation of the Planetary Boundary Layer Height Observed from GNSS Radio Occultation and Radiosonde Soundings over the Southern Great Plains, J. Atmos. Ocean. Tech., 38, 2081–2093, https://doi.org/10.1175/jtech-d-20-0196.1, 2021.

Newsom, R. K., Turner, D. D., and Goldsmith, J. E. M.: Long-Term Evaluation of Temperature Profiles Measured by an Operational Raman Lidar, J. Atmos. Ocean. Tech., 30, 1616–1634, https://doi.org/10.1175/JTECH-D-12-00138.1, 2013.

Newsom, R. and Sivaraman, C.: Raman Lidar Water Vapor Mixing Ratio and Temperature Value-Added Products, Tech. rep., DOE Office of Science Atmospheric Radiation Measurement (ARM) Program (United States), https://doi.org/10.2172/1489497, 2018.

Newsom, R. K., Sivaraman, C., Shippert, T. R., and Riihimaki, L. D.: Doppler Lidar Vertical Velocity Statistics Value-Added Product (No. DOE/SC-ARM/TR-149), DOE Office of Science Atmospheric Radiation Measurement (ARM) Program (United States), Richland, Washington, USA, https://doi.org/10.2172/1238068, 2019.

Newsom, R., and Krishnamurthy, R.: Doppler lidar (DL) instrument handbook, U.S. Department of Energy, Office of Science, DOE/SC-ARM-TR-101, https://doi.org/10.2172/1034640, 2022.

Painemal, D., Chiu, J.-Y. C., Minnis, P., Yost, C., Zhou, X., Cadeddu, M., Eloranta, E., Lewis, E. R., Ferrare, R., and Kollias, P.: Aerosol and cloud microphysics covariability in the northeast Pacific boundary layer estimated with ship-based and satellite remote sensing observations, J. Geophys. Res.-Atmos., 122, 2403-2418, https://doi.org/10.1002/2016JD025771, 2017.

Pedregosa, F., Michel, V., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Vanderplas, J., Cournapeau, D., Pedregosa, F., Varoquaux, G., Gramfort, A., Thirion, B., Grisel, O., Dubourg, V., Passos, A., Brucher, M., Perrot, M., and Duchesnay, É. (2011). Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research, 12, 2825–2830. http://scikit-learn.sourceforge.net.

Rieutord, T., Aubert, S., and Machado, T.: Deriving boundary layer height from aerosol lidar using machine learning: KABL and ADABL algorithms, Atmos. Meas. Tech., 14, 4335–4353, https://doi.org/10.5194/amt-14-4335-2021, 2021.

https://doi.org/10.5194/egusphere-2024-3959
Preprint. Discussion started: 14 February 2025

Roldán-Henao, N., Su, T., and Li, Z.: Refining planetary boundary layer height retrievals from micropulse-lidar at multiple ARM sites around the world, J. Geophys. Res.-Atmos., 129, e2023JD040207, https://doi.org/10.1029/2023JD040207, 2024a.

Roldán-Henao, N., Yorks, J. E., Su, T., Selmer, P. A, and Li, Z.: Statistically Resolved Planetary Boundary Layer Height Diurnal Variability Using Spaceborne Lidar Data. Remote Sensing. 16(17):3252. https://doi.org/10.3390/rs16173252. 2024b.

Salmun, H., Josephs, H., and Molod, A.: GRWP-PBLH: Global Radar Wind Profiler Planetary Boundary Layer Height Data, B. Am. Meteorol. Soc., 104, E1044–E1057, https://doi.org/10.1175/BAMS-D-22-0002.1., 2023.

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

Seibert, P., Beyrich, F., Gryning, S. E., Joffre, S., Rasmussen, A., and Tercier, P.: Review and intercomparison of operational methods for the determination of the mixing height, Atmos. Environ., 34, 1001–1027, https://doi.org/10.1016/S1352-2310(99)00349-0, 2000.

Seidel, D. J., Ao, C. O., and Li, K.: Estimating climatological planetary boundary layer heights from radiosonde observations: Comparison of methods and uncertainty analysis, J. Geophys. Res.-Atmos., 115, D16113, https://doi.org/10.1029/2009JD013680, 2010.

Sivaraman, C., McFarlane, S., Chapman, E., Jensen, M., Toto, T., Liu, S., and Fischer, M.: Planetary boundary layer (PBL) height value added product (VAP): Radiosonde retrievals, U.S. Department of Energy Rep. DOE/SC-ARM-TR-132, 36 pp., https://doi.org/10.2172/1808688, 2013.

Smith, E. N. and Carlin, J. T.: A multi-instrument fuzzy logic boundary-layer-top detection algorithm, Atmos. Meas. Tech., 17, 4087–4107, https://doi.org/10.5194/amt-17-4087-2024, 2024.

Stull, R. B.: An Introduction to Boundary Layer Meteorology, Kluwer Academic, 1988.

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 Sens. Environ., 237, 111519, https://doi.org/10.1016/j.rse.2019.111519, 2020.

Su, T., Li, Z., Roldán, N., Luan, Q., and Yu, F.: Constraining Effects of Aerosol-Cloud Interaction by Accounting for Coupling between Cloud and Land Surface, Sci. Adv., 10, eadl5044, https://doi.org/10.1126/sciadv.adl504, 2024.

Su, T. and Zhang, Y.: Deep-learning-derived planetary boundary layer height from conventional meteorological measurements, Atmos. Chem. Phys., 24, 6477–6493, https://doi.org/10.5194/acp-24-6477-2024, 2024.

https://doi.org/10.5194/egusphere-2024-3959
Preprint. Discussion started: 14 February 2025

Sujon, K. M., Hassan, R. B., Towshi, Z. T., Othman, M. A., Samad, M. A., and Choi, K.: When to Use Standardization and Normalization: Empirical Evidence From Machine Learning Models and XAI, *IEEE Access*, vol. 12, pp. 135300-135314, doi: 10.1109/ACCESS.2024.3462434, 2024.

Tang, S., Xie, S., Zhang, Y., and Cook, D. R.: The QCECOR Value-Added Product: Quality-Controlled Eddy Correlation Flux Measurements (No. DOE/SC-ARM-TR-223), DOE Office of Science Atmospheric Radiation Measurement (ARM) Program (United States), Livermore, California, USA, https://doi.org/10.2172/1557426, 2019.

Teixeira, J., Piepmeier, J. R., Nehrir, A. R., Ao, C. O., Chen, S. S., Clayson, C. A., Fridlind, A. M., Lebsock, M., McCarty, W., Salmun, H., Santanello, J. A., Turner, D. D., Wang, Z., and Zeng, X.: Toward a global planetary boundary layer observing system: the NASA PBL incubation study team report, NASA PBL Incubation Study Team, 134 pp., available at: https://science.nasa.gov/science-red/s3fs-public/atoms/files/NASAPBLIncubationFinalReport.pdf, last access: 21 September, 2024.

Tucker, S. C., Senff, C. J., Weickmann, A. M., Brewer, W. A., Banta, R. M., Sandberg, S. P., Law, D. C., and Hardesty, R. M.: Doppler lidar estimation of mixing height using turbulence, shear, and aerosol profiles, J. Atmos. Ocean. Tech., 26, 673–688, https://doi.org/10.1175/2008JTECHA1157.1, 2009.

Turner, D. and Löhnert, U.: Information content and uncertainties in thermodynamic profiles and liquid cloud properties retrieved from the ground-based Atmospheric Emitted Radiance Interferometer (AERI), J. Appl. Meteor. Climatol., 53, 752–771, https://doi.org/10.1175/JAMC-D-13-0126.1, 2014.

Turner, D. D., Wulfmeyer, V., Berg, L. K., and Schween, J. H.: Water vapor turbulence profiles in stationary continental convective mixed layers, J. Geophys. Res., 119, 1–15, https://doi.org/10.1002/2014JD022202, 2014.

Turner, D. D. and Löhnert, U.: Ground-based temperature and humidity profiling: combining active and passive remote sensors, Atmos. Meas. Tech., 14, 3033–3048, https://doi.org/10.5194/amt-14-3033-2021, 2021.

von Engeln, A. and Teixeira, J.: A planetary boundary layer height climatology derived from ECMWF reanalysis data, J. Climate, 26, 6575–6590, https://doi.org/10.1175/JCLI-D-12-00385.1, 2013.

Xi, X., Zhang, Y., Gao, Z., Yang, Y., Zhou, S., Duan, Z., and Yin, J.: Diurnal climatology of correlations between the planetary boundary layer height and surface meteorological factors over the contiguous United States. International Journal of Climatology, 42(10), 5092–5110. https://doi.org/10.1002/joc.7521, 2022.

Zhang, D., Comstock, J., and Morris, V.: Comparison of planetary boundary layer height from ceilometer with ARM radiosonde data, Atmos. Meas. Tech., 15, 4735–4749, https://doi.org/10.5194/amt-15-4735-2022, 2022.