https://doi.org/10.5194/egusphere-2024-376
Preprint. Discussion started: 13 February 2024

Deep-Learning-derived Planetary Boundary Layer Height from Conventional Meteorological Measurements

Tianning Su[1*], Yunyan Zhang[1]

[1]Lawrence Livermore National Laboratory, Livermore, CA, USA

[*]*Correspondence to*: su10@llnl.gov

https://doi.org/10.5194/egusphere-2024-376
Preprint. Discussion started: 13 February 2024

Abstract. The planetary boundary layer (PBL) height (PBLH) is an important parameter for various meteorological and climate studies. This study presents a multi-structure deep neural network (DNN) model, designed to estimate PBLH by integrating morning temperature profiles with surface meteorological observations. The DNN model is developed by leveraging a rich dataset of PBLH derived from long-standing radiosonde records and augmented with high-resolution micro-pulse lidar and Doppler lidar observations. We access the performance of the DNN with an ensemble of ten members, each featuring distinct hidden layer structures, which collectively yield a robust 27-year PBLH dataset over the Southern Great Plains from 1994 to 2020. The influence of various meteorological factors on PBLH is rigorously analyzed through the importance test. Moreover, the DNN model's accuracy is evaluated against radiosonde observations and juxtaposed with conventional remote sensing methodologies, including Doppler lidar, ceilometer, Raman lidar, and Micro-pulse lidar. The DNN model exhibits reliable performance across diverse conditions and demonstrates lower biases relative to remote sensing methods. In addition, the DNN model, originally trained over a plain region, demonstrates remarkable adaptability when applied to the heterogeneous terrains and climates encountered during the GoAmazon (Tropical Rainforest) and CACTI (Middle Latitude Mountain) campaigns. These findings demonstrate the effectiveness of deep learning models in estimating PBLH, enhancing our understanding of boundary layer dynamics with implications for enhancing the representation of PBL in weather forecasting and climate modeling.

https://doi.org/10.5194/egusphere-2024-376
Preprint. Discussion started: 13 February 2024

1 Introduction

The Planetary Boundary Layer (PBL) is the atmosphere's lowest part, where the Earth's surface directly influences meteorological variables, impacting the climate system (Garratt, 1994; Kaimal and Finnigan, 1994). The PBL height (PBLH) is a meteorological factor that strongly influences surface-atmosphere exchanges of heat, moisture, and energy (Stull, 1988; Caughey, 1984; Holtslag and Nieuwstadt, 1986; Mahrt, 1999; Helbig et al., 2021; Guo et al., 2024; Beamesderfer et al., 2022). In addition, PBLH it is a crucial variable for monitoring and simulating surface pollutant behaviors since it determines the volume available for near-surface pollutant dispersion (Li et al., 2017; Su et al., 2022a; Tucker et al., 2009). Due to its impacts on the development of convective systems, PBLH is also a key parameter in numerical weather forecasts and climate models (Menut et al., 1999; Park et al., 2001; 2023; Emanuel, 1994; Guo et al., 2017; Lilly, 1968; Matsui et al., 2004).

Radiosonde (SONDE) remains the standard method for estimating PBLH, yet it is hampered by limitations in temporal frequency, restricting its ability to comprehensively capture the whole diurnal cycle of PBL development (Stull, 1988; Seidel et al. 2010; Guo et al. 2021; Liu and Liang, 2010). To overcome these challenges, there has been an increasing dependence on remote sensing techniques, especially lidar systems. These techniques capture atmospheric vertical information (e.g., aerosols, temperature, humidity, and wind) at high temporal and vertical resolutions, leading to remote sensing-based retrievals of PBLH (Menut et al., 1999; Kotthaus et al., 2023; Sawyer and Li, 2013). The remote sensing systems, including Doppler lidar (Barlow et

https://doi.org/10.5194/egusphere-2024-376
Preprint. Discussion started: 13 February 2024

al. 2011), ceilometer (Zhang et al. 2022), Raman lidar (Summa et al. 2013), and Micro-pulse lidar (Melfi et al., 1985), utilize laser-based technology to track PBLH diurnal evolutions, helping us understand the PBL dynamics (Cohn and Angevine, 2000; Davis et al., 2000).

Despite advancements in remote sensing for PBLH estimation, challenges persist in aligning the results from various remote sensing instruments with those from SONDE measurements (Zhang et al. 2022; Su et al. 2020; Chu et al., 2019). Specifically, interpreting aerosol, turbulence, and moisture profiles derived from remote sensing techniques to determine PBLH bears inherent limitations due to the unstable signal-to-noise ratio (Su et al., 2017; Kotthaus et al., 2023; Krishnamurthy et al., 2021). This issue is compounded by the differing measurement methodologies employed by various remote sensing tools, leading to notable uncertainties when comparing their PBLH estimates to those obtained from standard SONDE retrievals (Zhang et al. 2022; Sawyer and Li, 2013).

As machine learning (ML) has shown potential in atmospheric science (McGovern et al., 2017; Gagne et al., 2019; Vassallo et al., 2020; Cadeddu et al., 2009; Molero et al. 2022), this technique presents a promising tool for refining the estimation of PBLH to resolve the inherent complexity and variability of PBL. For example, several studies use ML to identify PBL heights using thermodynamic profiles or backscatter profiles from Lidar or Atmospheric Emitted Radiance Interferometer (AERI), highlighting the ML's superiority over conventional techniques under different scenarios (Sleeman et al. 2020; Krishnamurthy et al., 2021; Rieutord et al. 2021; Liu et al. 2022; Ye et al. 2021).

https://doi.org/10.5194/egusphere-2024-376
Preprint. Discussion started: 13 February 2024

Moreover, Li et al. (2023) used an ML algorithm that considers the vertical distribution of aerosols to find the PBLH under complex atmospheric conditions.

While existing ML methodologies have marked progress in estimating PBLH, these studies mainly focus on refining retrievals from remote sensing data, particularly lidar-based technologies. Thus, there is an inherent limitation to the applicability due to a reliance on specific remote sensing instruments. To address this issue, we aim to leverage and integrate the comprehensive field observations (i.e., radiosonde and remote sensing techniques) to develop a deep learning model for direct PBLH estimation from conventional meteorological data. This strategy circumvents the limitations of relying on particular remote sensing technologies. Furthermore, our model employs a multi-structure deep neural network (DNN), diverging from traditional ML methods like random forest, to enhance its adaptivity for PBLH estimations. This multi-structure DNN approach offers great potential for wide applications under various meteorological conditions, as well as a stable performance for both trained and untrained periods. This underscores the versatility of DNN as a tool for PBLH estimation, which can be utilized under different scenarios and locations.

By focusing on the interaction between surface meteorology and the PBL, this study introduces a DNN-based method to estimate the daytime evolution of PBLH from morning temperature profiles and surface meteorology. We evaluate the model's performance using extensive datasets over the Southern Great Plains (SGP) for a period spanning 27 years (1994-2020) and includes comparisons with PBLH estimations obtained from Doppler lidar, ceilometer, Raman lidar, and micro-pulse lidar.

https://doi.org/10.5194/egusphere-2024-376
Preprint. Discussion started: 13 February 2024

Furthermore, we explore the generalizability of the model to different geographic regions and climates, as tested during the field campaigns, e.g., Green Ocean Amazon (GoAmazon) and Cloud, Aerosol, and Complex Terrain Interactions (CACTI).

2 Data and instruments

2.1 ARM Sites

The Atmospheric Radiation Measurement (ARM) program, funded by the U.S. Department of Energy, has been employed at the Southern Great Plains (SGP) site in Oklahoma (36.607°N, 97.488°W) for several decades. This study use comprehensive field observations at the SGP site during 1994 to 2020. In addition to the SGP site, this study utilizes data from the ARM GoAmazon (3.213°S, 60.598°W) and ARM CACTI (32.126°S, 64.728°W) field campaigns to carry out indenpendant tests for the deep learning model. Specifically, the GoAmazon campaign is located in the amazon tropical forests and provides rich field observations data during 2014-2015. Meanwhile, the CACTI central site, at an elevation of 1141 meters within the Sierras de Córdoba Mountain range in north-central Argentina, offers the observations during the 2018-2019 period. Utilizing these comprehensive ARM datasets, our study includes thermodynamic profiles derived from radiosondes, data from the Active Remote Sensing of Clouds (Clothiaux et al. 2000, 2001; Kollias et al. 2020), in-situ surface flux measurements, and standard meteorological observations at the surface, as documented by Cook (2018) and Xie et al. (2010).

https://doi.org/10.5194/egusphere-2024-376
Preprint. Discussion started: 13 February 2024

SONDE measurements at the ARM sites launch routinely several times a day and provide detailed information into the thermodynamic conditions of the atmosphere. The technical details of the ARM SONDE data are documented in Holdridge et al. (2011). Moreover, we use the surface meteorological parameters at the standard meteorological station. In-situ measurements at 2 meters above ground level provide data on temperature, relative humidity, and vapor pressure. Moreover, this study obtain the surface sensible and latent heat fluxes the surface instruments (Wesely et al., 1995). In SGP, we use the best-estimate surface fluxes in the Bulk Aerodynamic Energy Balance Bowen Ratio (BAEBBR) product, which is derived from the measurements by Energy Balance Bowen Ratio (EBBR). Due to the availability, we utilize the surface fluxes from Quality Controlled Eddy CORrelation (QCECOR) datasets from CACTI and GoAmazon sites (Tang et al. 2019).

2.2 Existing PBLH datasets over the ARM sites

For analyzing PBLH, we have utilized a variety of datasets to get a full picture of PBLH derived from different instruments. These datasets are developed by using different methodologies and instruments and jointly offer a detailed information of PBLH under various meteorological conditions. Among these datasets, SONDE- and ceilometer-derived PBLH are available for all three sites, other datasets are only available over the SGP. The technique details for these datasets can be found in the corresponding publications or technique reports.

(1) SONDE-derived PBLH by Liu and Liang (2010):

https://doi.org/10.5194/egusphere-2024-376
Preprint. Discussion started: 13 February 2024

PBLHs are retrieved using a method developed by Liu and Liang (2010), based on potential temperature gradients from SONDE. We focus on daytime data during 05:00–18:00 Local Time (LT), with a resampled vertical resolution of 5-hPa. The SONDE dataset is available at DOI: https://doi.org/10.5439/1595321.

(2) Doppler Lidar-derived PBLH by Sivaraman and Zhang (2021):

Doppler lidar PBLH estimates are derived using a vertical velocity variance method during 2010-2019 (Tucker et al., 2009; Lareau et al., 2018; Sivaraman and Zhang 2021). The dataset is available at DOI: https://doi.org/10.5439/1726254.

(3) Combined MPL and SONDE PBLH by Su et al. (2020):

We utilize a PBLH dataset that merges lidar and SONDE measurements during 1998-2023, ensuring vertical coherence and temporal continuity (Su et al. 2020). An additional method for handling cloudy conditions is detailed in Su et al. (2022b). The dataset is available at DOI: https://doi.org/10.5439/2007149.

(4) Ceilometer-derived PBLH by Zhang et al. (2022):

The Vaisala CL31 ceilometer, with a 7.7 km vertical range, provides detailed backscatter profiles used for PBLH estimation via gradient methods during 2011-2023 (Zhang et al. 2022). Enhanced algorithms ensure robust estimations under all weather conditions. The dataset is available at DOI: https://doi.org/10.5439/1095593.

(5) MPL-derived PBLH by Sawyer and Li (2013):

Micropulse lidar (MPL) is utilized for its high temporal resolution to retrieve PBLH during 2009-2020. MPL-derived PBLH, validated against SONDE and infrared spectrometer (AERI) data, improves understanding of boundary-layer processes

https://doi.org/10.5194/egusphere-2024-376
Preprint. Discussion started: 13 February 2024

(Sawyer and Li. 2013). The dataset is available at DOI: https://doi.org/10.5439/1637942.

(6) Combined Raman Lidar and AERI PBLH by Ferrare (2012):

PBLH is calculated using merged potential temperature profiles from Raman lidar and AERI, with criteria established for the SGP site. PBL heights are computed hourly for 2009-2011. The dataset is available at DOI: https://doi.org/10.5439/1169501.

In the datasets, (1-3) serve as the foundation for training. Concurrently, considering radiosonde as the benchmark standard, we utilized dataset (1) for validating PBLH retrievals obtained from various sources. Meanwhile, datasets (4-6) are used for the intercomparisons between PBLH derived from DNN and remote sensing techniques.

3 Deep Learning Model to Estimate PBLH

3.1 The Multi-Structure Deep Learning Model

Our deep learning model for estimating PBLH leverages the robustness of ensemble learning using a multi-structure DNN (Sze et al. 2017; Schmidhuber, 2015; Nielsen, 2015; Pang et al. 2020). This model used the TensorFlow Package, developed by Google (Abadi et al., 2016; https://www.tensorflow.org/). By employing an array of varied network architectures, we capitalize on the unique strengths of each structure to synthesize a more accurate and reliable estimation of PBLH. Figure 1 outlines the DNN's comprehensive design, beginning with the input layer that ingests a suite of morning meteorological features. We first present a preliminary run for the model to obtain the importance of each input feature. Then, these inputs undergo a filtration

https://doi.org/10.5194/egusphere-2024-376
Preprint. Discussion started: 13 February 2024

process based on their importance (Date and Kikuchi, 2018; Altmann et al. 2010), ensuring that only the impactful data guide the model (detailed in Section 3.3). Subsequently, the filtered inputs traverse through an ensemble of ten structures with distinct hidden layers. Each structure here represents an ensemble member and contributes to the prediction of PBLH in its unique way (Ganaie et al. 2022). The ensemble employs a three-layer base structure [52, 28, 16] for neural networks, from which ten unique configurations are derived by applying random perturbations to the default settings of the base structure. These different structures for ensembles 1-10 are presented in Table 1.

At the final stage, the model use the PBLH esimations from different ensembles to get a mean value as the final PBLH retrieval. This process allows the model to leverage the different results of all structures and enhance the generalizability of results. In the DNN model, neuron biases in the output and hidden layers are important for the network's architecture (Battaglia et al. 2018). These biases serve as fine-tuning parameters to adjust the activation thresholds of neurons in different layers and further refine the model's predictive capabilities. Neuron biases are initialized with small random values at the start of the training process and then iteratively adjusted according to the network weights during the training. Normalization is a preprocessing technique that often leads to improvements in model training by scaling the input features and target values to a standard range (Raju et al. 2020). The normalization process was applied to each input data to ensure that they have a zero mean and a standard deviation

https://doi.org/10.5194/egusphere-2024-376
Preprint. Discussion started: 13 February 2024

of one, as well as the target data. This standardization scales the different input data to a similar range, and thus, contributes a more stable and efficient training process.

The hidden layers of the DNN model incorporate L2 regularization to curtail overfitting, while batch normalization aids in stabilizing learning. Moreover, a dropout rate of 0.2 helps the model to generalize better by reducing reliance on any specific neurons during training. We choose the Adam optimizer and mean squared error as the loss function, which aligns with one of the best practices for regression models (Zhang. 2018). The mean absolute error is selected as a metric to evaluate the model's accuracy during the training. We incorporate the early stopping and learning rate reduction callbacks in in the model's training for regularization and fine-tuning (Liu et al. 2019). Such measures ensure optimal performance by terminating training at the right juncture and avoid the overfitting in the final results.

3.2 Training the DNN Model

The training of the DNN model was conducted using a PBLH dataset enriched by SONDE and lidar measurements during 1994 to 2016 over the SGP. Table 2 presents the distribution of dataset samples under different local time, which were important for both the training and validation processes of the DNN model. The primary dataset (i.e., PBLH derived from SONDE measurements) is listed in the first column and are available routinely for 5, 11, and 17 LT. The training dataset was augmented with the combined MPL-SONDE PBLH dataset (Su et al. 2020) and the Doppler Lidar-derived PBLH (Sivaraman and Zhang, 2021) to address the gaps where SONDE measurements

https://doi.org/10.5194/egusphere-2024-376
Preprint. Discussion started: 13 February 2024

were not available. In instances where radiosonde data are unavailable, the lidar datasets are used for training, contingent upon their agreement with radiosonde measurements within a margin of 0.2 km over a 3-hour window.

For the purpose of training the DNN model, 70% of the hourly data from both SONDE and the lidar combined dataset were randomly selected. The remaining 30% dataset, comprises the portion of SONDE measurements set aside for validation purposes, including a separate subset from the years 2017 to 2020 to test the model's predictive capabilities on independent data. This training and validation scheme ensures that the DNN model is not only well-trained but also thoroughly evaluated, reinforcing its reliability in accurately estimating PBLH. As morning SONDE data constitute the primary input and boundary conditions for the model, the validation of PBLH retrievals is consequently confined to the 08:00 to 18:00 LT.

3.3 Feature Importance Score

In the DNN model, we quantified the significance of each input parameter using the permutation importance technique, which is a widely-used method for the deep learning (Date and Kikuchi, 2018; Altmann et al. 2010). Initially, we carry out a test run to determine a baseline performance by calculating the mean absolute error (MAE) on the validation set. Then, each feature within this set was then individually shuffled, severing its correlation with the target PBLH, and the MAE was recalculated. Compared to the baseline performance, the increase in MAE from this shuffled state indicates the feature's predictive value: the greater the increase, the more significant the feature. We

https://doi.org/10.5194/egusphere-2024-376
Preprint. Discussion started: 13 February 2024

repeat this shuffling and evaluation for 15 times, each with a unique random seed to ensure statistical robustness. Furthermore, we calculated the average MAE increase across these iterations as the importance score. Therefore, we derived a composite importance metric for feature groups to represent their significance.

Figure 2 presents the importance scores to demonstrate each primary feature's relative influence on the model's performance. Prominently, features such as the boundary layer height derived from parcel methods (BLH_{Parcel}), morning potential temperature profiles (θ), and surface relative humidity are identified as pivotal, with their substantial impact on the accuracy of PBLH estimation being highlighted. BLH_{Parcel} is defined as the height where the morning potential temperature first exceeds the current surface potential temperature by more than 1.5 K (Holzworth, 1964; Chu et al., 2019). Complementing this, Table 3 offers an exhaustive breakdown of importance scores for all considered input features within the deep learning model. In refining the model, features contributing a negligible or negative effect on performance (i.e., importance scores less than zero) are excluded. As a result, this selection criterion has led to the inclusion of 58 out of the original 64 features. This process ensures we only use inputs with a proven positive influence in the DNN model.

4 Evaluation of Deep Learning Model

4.1 Comparative analysis of biases among different datasets

A critical component of evaluating our deep learning model's efficacy is analyzing the biases of individual ensemble members and their collective output. Figure 3 offers

https://doi.org/10.5194/egusphere-2024-376
Preprint. Discussion started: 13 February 2024

a visual assessment of the mean absolute error (MAE), root mean square error (RMSE), and correlation coefficient (R) for each ensemble member, alongside a comparison with the ensemble mean (average of all individual ensemble members). The plotted data points reveal the variation in performance across different model architectures, while the ensemble mean, represented by the horizontal dashed lines, indicates the collective accuracy of the ensemble approach. The structures of different hidden layer configurations are listed in the Table 1.

This methodological consolidation results in a more reliable and accurate PBLH estimation, leveraging the strengths and mitigating the weaknesses of individual models. By integrating multiple neural network configurations, we revealed that an ensemble prediction that consistently outperforms the individual models. This strategy can improve the MAE by up to 4.4%, rendering the model less dependent on any specific structural configuration.

An in-depth comparative analysis of biases among various PBLH estimation methods is essential for validating the reliability and accuracy of the DNN developed in this study. Figure 4 illustrates the MAE trends for several methods over a multi-year span, with the SONDE-derived PBLH serving as the benchmark for ground truth. The analysis reveals the performance of different methodologies: the DNN approach, doppler lidar, ceilometer, MPL, and Raman lidar. Significantly, the DNN model, depicted in black, maintains a consistent MAE trend throughout the trained period (1994-2016) as well as the subsequent untrained period (2017-2020), demonstrating robust predictive stability. In contrast, the remote sensing-based methods show a

https://doi.org/10.5194/egusphere-2024-376
Preprint. Discussion started: 13 February 2024

reduction in bias from 2010 to 2022, possibly due to the improvement of remote sensing data quality. The discrepancy in PBLH estimates between the DNN and SONDE remains consistently lower than those observed with conventional remote sensing techniques.

Figure 5 provides a detailed evaluation of the DNN model in comparison to ceilometer and doppler lidar-derived PBLH, as these two methods have demonstrated the high quality with more than nine years of datasets. Figure 5a-b contrast the PBLH predictions from the DNN model for both the trained period (1994-2016) and untrained periods (2017-2020), respectively, showcasing strong correlations and low MAEs, indicative of the model's robust training and generalization capabilities. Figure 5c-d further this examination with ceilometer and Doppler lidar comparisons, respectively. Overall, Doppler lidar exhibits a closer alignment with SONDE-derived PBLH than the ceilometer. However, the MAE from Doppler lidar-based estimates is still approximately 48% higher than those derived from the DNN model. The correlation coefficient for the DNN-derived PBLH estimates has seen a substantial improvement, rising from the 0.5-0.6 range typically observed with remote sensing-based PBLH methods to exceed 0.8 when compared to SONDE-derived PBLH measurements. This comparative analysis not only confirms the DNN model's accuracy but also offers insights into the relative performance of various contemporary PBLH estimation methodologies.

https://doi.org/10.5194/egusphere-2024-376
Preprint. Discussion started: 13 February 2024

4.2 Performances of PBLH retrievals under different conditions

The performance of PBLH retrievals under varying atmospheric conditions is a crucial aspect of model evaluation. In Figure 6, the seasonal diurnal cycles of PBLH estimated by different methods are presented, offering information into the diurnal and seasonal evolution of PBL. As PBLH demonstrates notable variations for different seasons and local time with large differences between summer and winter, the DNN and Doppler lidar estimates show good agreement and closely track the variations observed in SONDE data. Meanwhile, the ceilometer presents an underestimation of PBLH, especially for the summer afternoon, indicating the potential bias of ceilometer derived PBLH under a convective environment.

Figure 7 illustrates the diurnal variation in the model's performance by comparing the correlation coefficient, RMSE, and MAE against SONDE-derived PBLH as the reference. The bar graphs for each local time hour offer a comparison of the RMSE and MAE, as well as the correlation, showcasing the model's precision and consistency relative to remote sensing methods (i.e., ceilometer and Doppler lidar). The ceilometer-derived PBLH exhibits the greatest variations during different hours, particularly around noon, suggesting a time-dependent bias in its measurements. Conversely, both the DNN and Doppler lidar-derived PBLH demonstrate stable performance in term of MAE and RMSE throughout the day. Regarding the correlation, remote sensing methods like ceilometer and Doppler lidar exhibit a lower correlation with SONDE-derived PBLH, especially in the early hours (8-9 LT) with a value of 0.1-0.3, indicating potential limitations in their reliability during these times. On the other hand, the DNN

https://doi.org/10.5194/egusphere-2024-376
Preprint. Discussion started: 13 February 2024

model shows a relatively good correlation with SONDE retrievals (above 0.6 under different hours). This comparison shows the efficacy of DNN in tracking the diurnal cycle of PBLH.

Continuing our assessment of the DNN model, we analyze the DNN model's monthly performance in estimating PBLH, as shown in Figure 8. The analysis compares MAE, RMSE, and correlation coefficients for each month to assess the model's precision and dependability. The summer months (June-July-August) exhibit higher biases, with MAE values for the DNN, ceilometer, and Doppler lidar at 0.3 km, 0.56 km, and 0.45 km, respectively. In contrast, the winter months (December-January-February) show reduced biases, with MAE values of 0.2 km for the DNN, 0.27 km for the ceilometer, and 0.24 km for the Doppler lidar. Specifically, the DNN model shows a much lower bias during the summer season. Compared to the remote sensing-based retrievals, the DNN-derived PBLH shows a much better agreement with SONDE-derived PBLH, increasing from 0.3-0.6 to approximately 0.8 in term of correlation coefficients.

Figure 9 presents the biases of PBLH retrievals under clear-sky and low cloud conditions. We calculated biases as the absolute deviation from the mean PBLH for each condition, focusing particularly on the differences between low cloud (maximum cloud fraction between 0-4 km exceeding 1%) and clear-sky (total cloud fraction below 1%) scenarios. The violin plots in this figure illustrate the data distribution of biases for each method to demonstrate their variability. For the DNN model and ceilometer, the relative biases between clear and cloudy conditions are comparable and the difference

https://doi.org/10.5194/egusphere-2024-376
Preprint. Discussion started: 13 February 2024

is less than than 1%. This suggests a consistent performance across these atmospheric states. However, the Doppler lidar exhibits a larger disparity, showing a 5.5% bias under cloudy conditions compared to clear skies. Moreover, the spread of biases (shaded areas and error bars) is notably wider for both the ceilometer and Doppler lidar. This indicates large variability in their performance. For all three methods, the mean biases are notably higher than the median values. Such differences indicate that the mean values are notably influenced by outliers under both clear-sky and cloudy conditions.

The analyses presented in this section illustrate the effectiveness of the DNN model in capturing the PBLH variations across different local times, seasons, and atmospheric conditions. Compared to the traditional remote sensing methods, the DNN model exhibits relatively good accuracy in aligning with SONDE-derived PBLH, indicating its capability and stable performance under different scenarios.

4.3 Testing the DNN Model's Adaptability

The DNN model relies on the incorporation of morning temperature profiles as inputs, such as detailed in Table 3. This dependency prompts the question of how to proceed the DNN model in the absence of SONDE data at specific locations. As a solution, we suggest employing morning temperature profiles from the European Centre for Medium-Range Weather Forecasts' fifth-generation global reanalysis (ERA-5, Hersbach et al., 2020) dataset when radiosonde data is not available to maintain the model's operational integrity under sounding-data-constrained conditions. As one of the

https://doi.org/10.5194/egusphere-2024-376
Preprint. Discussion started: 13 February 2024

most advanced reanalysis data, the ERA-5 is generated by the Integrated Forecasting System coupled with a data assimilation system, and offer the meteorological data at a spatial resolution of 0.25°- 0.25°.

Figure 10 assess the performance of DNN produced by multi-sources field observations in estimating the PBLH by using morning temperature profiles from ERA-5 (5 LT) and observed surface meteorological data. The temperature profiles in ERA-5 have a vertical resolution of 25-hPa in the lower atmosphere and are interpolated into different levels described in Table 3. By utilizing ERA-5 morning profiles, the model demonstrates similar performance to those results achieved with radiosonde inputs, as evidenced by comparing Figure 10a and Figure 5. Moreover, this alternative approach also shows enhanced accuracy over the native PBLH model outputs from ERA-5, increasing the correlation coefficient from 0.74 to 0.86 and reducing the MAE from 0.3 km to 0.25 km. In addition, it is important to acknowledge that the PBLH represented in ERA-5 is indicative of a grid-average value, approximately 25 km in scale, and therefore inherently differs from site-specific data.

These findings highlight the alternative DNN model's robustness, offering a reliable substitute for radiosonde data by leveraging reanalysis data with similar performance. This demonstrates the DNN model's adaptability and potential as a practical tool for PBLH estimation across various meteorological sites, especially in regions or periods where radiosonde data may be lacking.

We further test the adaptability and generalizability of the DNN model, by applying across different climatic and geographic regions. To this end, we extended our model

https://doi.org/10.5194/egusphere-2024-376
Preprint. Discussion started: 13 February 2024

evaluation to include SONDE and surface meteorological data from the GoAmazon (Tropical Rainforest) and CACTI (Middle Latitude Mountain) field campaigns. Seasonality is accounted for as an input variable in the DNN model, with months in the Southern Hemisphere adjusted to reflect their Northern Hemisphere seasonal counterparts (e.g., July inputs are treated as January). The normalization process (Section 3.1) was reapplied for the CACTI campaign data to adjust for notable pressure level variations, ensuring input standardization with zero mean and unit variance.

Figure 11 presents the model's performance, in comparison to SONDE observations for both GoAmazon and CACTI campaigns. The DNN model demonstrates commendable adaptability, maintaining a strong correlation (0.86-0.88) with SONDE measurements (Figure 11a-b). Further comparison is provided, which assess the performance of ceilometer derived PBLH against SONDE for the same campaigns. When assessing the performance of the ceilometer-derived PBLH against SONDE for the same campaigns, the DNN model exhibited both stronger correlations and smaller biases, as shown in Figure 11b-d.

Nevertheless, the analysis highlighted the presence of systematic biases, with relatively larger MAE at the GoAmazon and CACTI sites compared to the SGP site. Figure 12 underscores this by presenting a comparative analysis of PBLH means and standard deviations across the three ARM sites. The early morning measurements during 05-07 LT are excluded. The results, derived from SONDE, the DNN model, ceilometer, and Doppler lidar data, reveal average differences in PBLH means relative to SONDE. These differences suggest an overestimation (+15%) and underestimation

https://doi.org/10.5194/egusphere-2024-376
Preprint. Discussion started: 13 February 2024

(-23%) by the DNN model for the GoAmazon and CACTI sites, respectively, compared to the more consistent PBLH values at the SGP site.

The evident systematic deviations when applying the SGP-trained DNN model to the diverse environments of GoAmazon and CACTI underscore the challenges in generalizing the model to regions with significantly different meteorological backgrounds. These findings point to the potential of DNN models for PBLH estimation while also highlighting the necessity for region-specific model adjustments.

5 Summary

This study has developed a Multi-Structure DNN model for estimating PBLH using conventional meteorological data. The DNN model is developed by leveraging a long-term dataset of PBLH derived from radiosonde data and augmented with high-resolution MPL and Doppler lidar observations. This model produced an PBLH dataset over the SGP with robust accuracy, consistently yielding lower bias values across various conditions and datasets. Utilizing conventional meteorological data, this method generates a 27-year dataset over the SGP, encompassing periods with limited remote sensing data availability. In situations where morning radiosonde data is unavailable, ERA-5 data can be effectively employed to initiate the model, offering a practical alternative.

An important aspect of this research involved comparing DNN models with diverse remote sensing instruments. Although these instruments offer high temporal and vertical resolution, discrepancies in PBLH estimation remain. Our DNN model,

https://doi.org/10.5194/egusphere-2024-376
Preprint. Discussion started: 13 February 2024

leveraging a broad range of input features refined by their importance, constructs a representation of PBL evolutions, frequently demonstrating a closer agreement with SONDE-derived PBLH. In the absence of remote sensing data, the DNN model can produce high-quality PBLH results from the conventional meteorology data.

The study has shown the DNN model's ability to synthesize complex patterns from meteorological data, reflecting the versatility of machine learning in simulating the boundary layer processes. Its application to varied geographic terrains and climates during the GoAmazon and CACTI campaigns has further validated its adaptability, demonstrating a high correlation between DNN-derived PBLH and SONDE-derived PBLH. Nonetheless, systematic biases in regions outside the SGP highlight the influence of regional factors in PBLH estimation and suggest the need for region-specific refinements to the model.

In summary, this research introduces a machine learning framework for PBLH estimation that is able to generate high-quality PBLH using meteorological data, independent of remote sensing instruments. This methodology, alongside the datasets derived from the deep learning model, is beneficial in advancing our understanding of PBL daytime development including thermodynamics and dynamics. It also has implications for improved representation of the PBL processes in weather forecasting and climate models, particularly by offering the potential to diagnose PBL in models through the integration of modeled meteorological data as input. Future efforts will be directed towards refining this model to ensure its wide applicability over a global scale.

https://doi.org/10.5194/egusphere-2024-376
Preprint. Discussion started: 13 February 2024

These developments aim to effectively tackle the challenges of systematic biases and regional variability in PBLH estimation.

Data Availability. ARM radiosonde data, surface fluxes, and cloud masks are available at https://adc.arm.gov/discovery/#/results/instrument_class_code::armbe. The datasets of planetary boundary layer height used in this study can be downloaded from https://adc.arm.gov/discovery/#/results/instrument_class_code::pblht. The DNN-derived PBLH datasets over the SGP, CACTI, and GoAmazon are available at Zenodo (https://zenodo.org/records/10633811) and will be uploaded to ARM data archive as a product with detailed information upon acceptance.

Author contributions. TS conceptualized this study and carried out the analysis. TS and YZ interpreted the data and wrote the manuscript. YZ supervised the project.

Competing interests. The contact author has declared that neither they nor their co-authors have any competing interests.

Acknowledgements: We acknowledge the provision of radiosonde, lidar data, surface meteorological data, and cloud products by the U.S. Department of Energy's ARM program. Work at LLNL is performed under the auspices of the U.S. DOE by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344. This research used resources of the National Energy Research Scientific Computing Center

https://doi.org/10.5194/egusphere-2024-376
Preprint. Discussion started: 13 February 2024

(NERSC), a U.S. Department of Energy Office of Science User Facility located at Lawrence Berkeley National Laboratory, operated under Contract No. DE-AC02-05CH11231.

Financial support. This work is supported by the DOE Atmospheric System Research (ASR) Science Focus Area (SFA) THREAD project.

References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G.S., Davis, A., Dean, J., Devin, M. and Ghemawat, S., 2016. Tensorflow: Large-scale machine learning on heterogeneous distributed systems. arXiv preprint arXiv:1603.04467.

Altmann, A., Toloşi, L., Sander, O. and Lengauer, T., 2010. Permutation importance: a corrected feature importance measure. Bioinformatics, 26(10), pp.1340-1347.

Barlow, J.F., Dunbar, T.M., Nemitz, E.G., Wood, C.R., Gallagher, M.W., Davies, F., O'Connor, E. and Harrison, R.M., 2011. Boundary layer dynamics over London, UK, as observed using Doppler lidar during REPARTEE-II. *Atmospheric Chemistry and Physics*, *11*(5), pp.2111-2125.

Battaglia, P.W., Hamrick, J.B., Bapst, V., Sanchez-Gonzalez, A., Zambaldi, V., Malinowski, M., Tacchetti, A., Raposo, D., Santoro, A., Faulkner, R. and Gulcehre, C., 2018. Relational inductive biases, deep learning, and graph networks. arXiv preprint arXiv:1806.01261.

Beamesderfer, E.R., Buechner, C., Faiola, C., Helbig, M., Sanchez-Mejia, Z.M., Yáñez-Serrano, A.M., Zhang, Y. and Richardson, A.D., 2022. Advancing cross-disciplinary understanding of land-atmosphere interactions. Journal of Geophysical Research: Biogeosciences, 127(2), p.e2021JG006707.

Cadeddu, M. P., Turner, D. D., and Liljegren, J. C., 2009. A neural network for real-time retrievals of PWV and LWP from Arctic millimeter-wave ground-based observations, IEEE T. Geosci. Remote, 47, 1887–1900.

Caughey, S.J. (1984). Observed characteristics of the atmospheric boundary layer. In Atmospheric turbulence and air pollution modelling (pp. 107-158). Springer, Dordrecht.

Chu, Y., Li, J., Li, C., Tan, W., Su, T. and Li, J. (2019). Seasonal and diurnal variability of planetary boundary layer height in Beijing: Intercomparison between MPL and WRF results. Atmospheric research, 227, pp.1-13.

https://doi.org/10.5194/egusphere-2024-376
Preprint. Discussion started: 13 February 2024

Chu, Y., Li, J., Li, C., Tan, W., Su, T. and Li, J., 2019. Seasonal and diurnal variability of planetary boundary layer height in Beijing: Intercomparison between MPL and WRF results. Atmospheric research, 227, pp.1-13.

Clothiaux, E. E., Ackerman, T. P., Mace, G. G., Moran, K. P., Marchand, R. T., Miller, M. A., and Martner, B. E. (2000): Objective determination of cloud heights and radar reflectivities using a combination of active remote sensors at the ARM CART sites. J. Appl. Meteorol., 39(5), 645–665.

Clothiaux, E.E., Miller, M.A., Perez, R.C., Turner, D.D., Moran, K.P., Martner, B.E., Ackerman, T.P., Mace, G.G., Marchand, R.T., Widener, K.B. and Rodriguez, D.J. (2001). The ARM millimeter wave cloud radars (MMCRs) and the active remote sensing of clouds (ARSCL) value added product (VAP) (No. DOE/SC-ARM/VAP-002.1). DOE Office of Science Atmospheric Radiation Measurement (ARM) Program (United States).

Cohn, S. A. and Angevine, W. M. (2000). Boundary layer height and entrainment zone thickness measured by lidars and wind-profiling radars, J. Appl. Meteorol., 39, 1233–1247.

Cook, D.R. (2018). Energy balance bowen ratio station (EBBR) instrument handbook (No. DOE/SC-ARM/TR-037). DOE Office of Science Atmospheric Radiation Measurement (ARM) Program (United States).

Date, Y. and Kikuchi, J., 2018. Application of a deep neural network to metabolomics studies and its performance in determining important variables. Analytical chemistry, 90(3), pp.1805-1810.

Davis, K.J., Gamage, N., Hagelberg, C.R., Kiemle, C., Lenschow, D.H., & Sullivan, P.P. (2000). An objective method for deriving atmospheric structure from airborne lidar observations. Journal of Atmospheric and Oceanic Technology, 17, 1455-1468

Emanuel, K.A. (1994). Atmospheric convection.: Oxford University Press on Demand

Ferrare, Richard. Raman lidar/AERI PBL Height Product. United States: N. p., 2012. Web. doi:10.5439/1169501.

Gagne II, D. J., Haupt, S. E., Nychka, D. W., and Thompson, G.: Interpretable deep learning for spatial analysis of severe hailstorms, Mon. Weather Rev., 147, 2827–2845, (2019).

Ganaie, M.A., Hu, M., Malik, A.K., Tanveer, M. and Suganthan, P.N., 2022. Ensemble deep learning: A review. Engineering Applications of Artificial Intelligence, 115, p.105151.

Garratt, J. R. (1994): The atmospheric boundary layer. Earth-Sci. Rev., 37(1–2), 89–134.

Guo, J., Su, T., Li, Z., Miao, Y., Li, J., Liu, H., Xu, H., Cribb, M. and Zhai, P. (2017). Declining frequency of summertime local-scale precipitation over eastern China from 1970 to 2010 and its potential link to aerosols. Geophysical Research Letters, 44(11), pp.5700-5708.

Guo, J., Zhang, J., Shao, J., Chen, T., Bai, K., Sun, Y., Li, N., Wu, J., Li, R., Li, J. and Guo, Q., 2024. A merged continental planetary boundary layer height dataset

https://doi.org/10.5194/egusphere-2024-376
Preprint. Discussion started: 13 February 2024

based on high-resolution radiosonde measurements, ERA5 reanalysis, and GLDAS. Earth System Science Data, 16(1), pp.1-14.

Guo, J., Zhang, J., Yang, K., Liao, H., Zhang, S., Huang, K., Lv, Y., Shao, J., Yu, T., Tong, B. and Li, J., 2021. Investigation of near-global daytime boundary layer height using high-resolution radiosondes: first results and comparison with ERA5, MERRA-2, JRA-55, and NCEP-2 reanalyses. Atmospheric Chemistry and Physics, 21(22), pp.17079-17097.

Helbig, M., Gerken, T., Beamesderfer, E.R., Baldocchi, D.D., Banerjee, T., Biraud, S.C., Brown, W.O., Brunsell, N.A., Burakowski, E.A., Burns, S.P. and Butterworth, B.J., 2021. Integrating continuous atmospheric boundary layer and tower-based flux measurements to advance understanding of land-atmosphere interactions. Agricultural and Forest Meteorology, 307, p.108509.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D. and Simmons, A., 2020. The ERA5 global reanalysis. Quarterly Journal of the Royal Meteorological Society, 146(730), pp.1999-2049.

Holdridge, D., Ritsche, M., Prell, J., and Coulter, R. (2011): Balloon-borne sounding system (SONDE) handbook, https://www.arm.gov/capabilities/instruments/sonde.

Holtslag, A.A. and Nieuwstadt, F.T. (1986). Scaling the atmospheric boundary layer. Boundary-Layer Meteorology, 36(1-2), pp.201-209.

Holzworth, G. C. (1964). Estimates of mean maximum mixing depths in the contiguous United States. Monthly Weather Review, 92(5), 235–242. https://doi.org/10.1175/1520-0493(1964)092<0235:EOMMMD>2.3.CO;2

Kaimal, J.C. and Finnigan, J.J. (1994). Atmospheric boundary layer flows: their structure and measurement. Oxford university press.

Kollias, P., Bharadwaj, N., Clothiaux, E.E., Lamer, K., Oue, M., Hardin, J., Isom, B., Lindenmaier, I., Matthews, A., Luke, E.P. and Giangrande, S.E. (2020). The ARM radar network: At the leading edge of cloud and precipitation observations. Bulletin of the American Meteorological Society, 101(5), pp.E588-E607.

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. and Mona, L., 2023. Atmospheric boundary layer height from ground-based remote sensing: a review of capabilities and limitations. Atmospheric Measurement Techniques, 16(2), pp.433-479.

Krishnamurthy, R., Newsom, R.K., Berg, L.K., Xiao, H., Ma, P.L. and Turner, D.D., 2021. On the estimation of boundary layer heights: a machine learning approach. Atmospheric Measurement Techniques, 14(6), pp.4403-4424.

Lareau, N.P., Zhang, Y. and Klein, S.A., 2018. Observed boundary layer controls on shallow cumulus at the ARM Southern Great Plains site. Journal of the Atmospheric Sciences, 75(7), pp.2235-2255.

Li, H., Liu, B., Ma, X., Jin, S., Wang, W., Fan, R., Ma, Y., Wei, R. and Gong, W., 2023. Estimation of Planetary Boundary Layer Height from Lidar by Combining Gradient Method and Machine Learning Algorithms. IEEE Transactions on Geoscience and Remote Sensing.

https://doi.org/10.5194/egusphere-2024-376
Preprint. Discussion started: 13 February 2024

Li, Z., Guo, J., Ding, A., Liao, H., Liu, J., Sun, Y., Wang, T., Xue, H., Zhang, H. and Zhu, B., 2017. Aerosol and boundary-layer interactions and impact on air quality. National Science Review, 4(6), pp.810-833.

Lilly, D.K. (1968). Models of Cloud-Topped Mixed Layers under a Strong Inversion. Quarterly Journal of the Royal Meteorological Society, 94, 292-&

Liu, F., Page, A., Strode, S. A., Yoshida, Y., Choi, S., Zheng, B., Lamsal, L. N., Li, C., Krotkov, N. A., Eskes, H., and Veefkind, P. (2020). Abrupt decline in tropospheric nitrogen dioxide over China after the outbreak of COVID-19. Science Advances, 6(28), eabc2992

Liu, L., Jiang, H., He, P., Chen, W., Liu, X., Gao, J. and Han, J., 2019. On the variance of the adaptive learning rate and beyond. arXiv preprint arXiv:1908.03265.Garratt, J.R. (1994). The atmospheric boundary layer. Earth-Science Reviews, 37(1-2), pp.89-134.

Liu, S. and Liang, X.Z., 2010. Observed diurnal cycle climatology of planetary boundary layer height. Journal of Climate, 23(21), pp.5790-5809.

Liu, Z., Chang, J., Li, H., Chen, S. and Dai, T., 2022. Estimating boundary layer height from lidar data under complex atmospheric conditions using machine learning. Remote Sensing, 14(2), p.418.

Liu, Z., Chang, J., Li, H., Chen, S. and Dai, T., 2022. Estimating boundary layer height from lidar data under complex atmospheric conditions using machine learning. Remote Sensing, 14(2), p.418.

Mahrt, L. (1999). Stratified atmospheric boundary layers. Boundary-Layer Meteorology, 90(3), pp.375-396.

Matsui, T., Masunaga, H., Pielke, R.A., & Tao, W.K. (2004). Impact of aerosols and atmospheric thermodynamics on cloud properties within the climate system. Geophysical Research Letters, 31

McGovern, A., Elmore, K. L., Gagne, D. J., Haupt, S. E., Karstens, C. D., Lagerquist, R., Smith, T., and Williams, J. K.: Using artificial intelligence to improve real-time decision-making for high-impact weather, B. Am. Meteorol. Soc., 98, 2073–2090, (2017).

Melfi, S.H., Spinhirne, J.D., Chou, S.H. and Palm, S.P. (1985). Lidar observations of vertically organized convection in the planetary boundary layer over the ocean. Journal of climate and applied meteorology, 24(8), pp.806-821.

Menut, L., Flamant, C., Pelon, J., & Flamant, P.H. (1999). Urban boundary-layer height determination from lidar measurements over the Paris area. Applied Optics, 38, 945-954

Molero, F., Barragán, R. and Artíñano, B., 2022. Estimation of the atmospheric boundary layer height by means of machine learning techniques using ground-level meteorological data. Atmospheric Research, 279, p.106401.

Nielsen, M.A., 2015. Neural networks and deep learning (Vol. 25, pp. 15-24). San Francisco, CA, USA: Determination press.

Pang, B., Nijkamp, E. and Wu, Y.N., 2020. Deep learning with tensorflow: A review. Journal of Educational and Behavioral Statistics, 45(2), pp.227-248.

Park, O.H., Seo, S.J., & Lee, S.H. (2001). Laboratory simulation of vertical plume dispersion within a convective boundary layer - Research note. Boundary-Layer Meteorology, 99, 159-169

Raju, V.G., Lakshmi, K.P., Jain, V.M., Kalidindi, A. and Padma, V., 2020, August. Study the influence of normalization/transformation process on the accuracy of supervised classification. In 2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT) (pp. 729-735). IEEE.

Rieutord, T., Aubert, S. and Machado, T., 2021. Deriving boundary layer height from aerosol lidar using machine learning: KABL and ADABL algorithms. Atmospheric Measurement Techniques, 14(6), pp.4335-4353.

Sawyer, V. and Li, Z., 2013. Detection, variations and intercomparison of the planetary boundary layer depth from radiosonde, lidar and infrared spectrometer. Atmospheric environment, 79, pp.518-528.

Schmidhuber, J., 2015. Deep learning in neural networks: An overview. Neural networks, 61, pp.85-117.

Seidel, D.J., Ao, C.O. and Li, K., 2010. Estimating climatological planetary boundary layer heights from radiosonde observations: Comparison of methods and uncertainty analysis. Journal of Geophysical Research: Atmospheres, 115(D16).

Sivaraman, Chitra, and Zhang, Damao. Planetary Boundary Layer Height derived from Doppler Lidar (DL) data. United States: N. p., 2021. Web. doi:10.5439/1726254.

Sleeman, J., Halem, M., Yang, Z., Caicedo, V., Demoz, B. and Delgado, R., 2020, September. A deep machine learning approach for lidar based boundary layer height detection. In IGARSS 2020-2020 IEEE international geoscience and remote sensing symposium (pp. 3676-3679). IEEE.

Stull, R.B. (1988). An Introduction to Boundary Layer Meteorology. Dordrecht: Springer Netherlands

Su, T., Li, J., Li, C., Xiang, P., Lau, A.K.H., Guo, J., Yang, D. and Miao, Y., 2017. An intercomparison of long-term planetary boundary layer heights retrieved from CALIPSO, ground-based lidar, and radiosonde measurements over Hong Kong. Journal of Geophysical Research: Atmospheres, 122(7), pp.3929-3943.

Su, T., Li, Z. and Zheng, Y. (2023). Cloud-Surface Coupling Alters the Morning Transition From Stable to Unstable Boundary Layer. Geophysical Research Letters, 50(5), p.e2022GL102256.

Su, T., Li, Z., and Kahn, R. (2020): 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.

Su, T., Li, Z., Zheng, Y., Wu, T., Wu, H. and Guo, J., (2022a). Aerosol-boundary layer interaction modulated entrainment process. NPJ Climate and Atmospheric Science, 5(1), p.64.

Su, T., Zheng, Y. and Li, Z., (2022b). Methodology to determine the coupling of continental clouds with surface and boundary layer height under cloudy conditions from lidar and meteorological data. Atmospheric Chemistry and Physics, 22(2), pp.1453-1466.

https://doi.org/10.5194/egusphere-2024-376
Preprint. Discussion started: 13 February 2024

Summa, D., Di Girolamo, P., Stelitano, D. and Cacciani, M., 2013. Characterization of the planetary boundary layer height and structure by Raman lidar: comparison of different approaches. Atmospheric Measurement Techniques, 6(12), pp.3515-3525.

Sze, V., Chen, Y.H., Yang, T.J. and Emer, J.S., 2017. Efficient processing of deep neural networks: A tutorial and survey. Proceedings of the IEEE, 105(12), pp.2295-2329.

Tang, S., Xie, S., Zhang, M., Tang, Q., Zhang, Y., Klein, S. A., Cook, D. R., and Sullivan, R. C. (2019): Differences in eddy-correlation and energy-balance surface turbulent heat flux measurements and their impacts on the large-scale forcing fields at the ARM SGP site. J. Geophy. Res. Atmos., 124, 3301–3318, doi.org/10.1029/2018JD029689.

Tucker, S.C., Brewer, W.A., Banta, R.M., Senff, C.J., Sandberg, S.P., Law, D.C., Weickmann, A.M., & Hardesty, R.M. (2009). Doppler Lidar Estimation of Mixing Height Using Turbulence, Shear, and Aerosol Profiles. Journal of Atmospheric and Oceanic Technology, 26, 673-688

Vassallo, D., Krishnamurthy, R., and Fernando, H. J. S.: Decreasing wind speed extrapolation error via domain-specific feature extraction and selection, Wind Energ. Sci., 5, 959–975, https://doi.org/10.5194/wes-5-959-2020, 2020.

Wesely, M. L., Cook, D. R., and Coulter, R. L. (1995): Surface heat flux data from energy balance Bowen ratio systems (No. ANL/ER/CP-84065; CONF-9503104-2). Argonne National Lab., IL (United States).

Xie, S., McCoy, R. B., Klein, S. A., Cederwall, R. T., Wiscombe, W. J., Jensen, M. P., Johnson, K. L., Clothiaux, E. E., Gaustad, K. L., Long, C. N., and Mather, J. H. (2010): Clouds and more: ARM climate modeling best estimate data: a new data product for climate studies. Bull. Amer. Meteorol. Soc., 91(1), 13–20.

Ye, J., Liu, L., Wang, Q., Hu, S. and Li, S., 2021. A novel machine learning algorithm for planetary boundary layer height estimation using AERI measurement data. IEEE Geoscience and Remote Sensing Letters, 19, pp.1-5.

Zhang, D., Comstock, J. and Morris, V., (2022). Comparison of planetary boundary layer height from ceilometer with ARM radiosonde data. Atmospheric Measurement Techniques, 15(16), pp.4735-4749.

Zhang, Z., (2018), June. Improved adam optimizer for deep neural networks. In 2018 IEEE/ACM 26th international symposium on quality of service (IWQoS) (pp. 1-2). IEEE.

https://doi.org/10.5194/egusphere-2024-376
Preprint. Discussion started: 13 February 2024

Table list:

Table 1. This table lists the varying structures of hidden layers used by each ensemble member for PBLH estimation. Each configuration is expressed as an array, with the number of elements indicating the number of layers and each value specifying the number of neurons activated in the corresponding layer. For instance, a structure denoted as [52, 28, 16] comprises three hidden layers containing 52, 28, and 16 neurons, respectively.

Ensemble Member	Different Structures in Hidden Layer	Ensemble Member	Different Structures in Hidden Layer
Member 1	[52, 28, 16]	Member 6	[57, 44, 19]
Member 2	[61, 43, 20]	Member 7	[55, 43, 19]
Member 3	[59, 45, 19]	Member 8	[57, 43, 15]
Member 4	[60, 45, 23]	Member 9	[59, 41, 20, 10]
Member 5	[57, 45, 23]	Member 10	[57, 43, 18, 9]

https://doi.org/10.5194/egusphere-2024-376
Preprint. Discussion started: 13 February 2024

Table 2. Distribution of Dataset Samples for deep learning neural network (DNN) Training and Validation. This table details the sample data in different local time used for the development and validation of DNN to estimate planetary boundary layer height (PBLH). The first column lists the available PBLH derived from radiosonde (SONDE, Liu and Liang. 2010) during various local hours from 1994 to 2016. The second column supplements the dataset with a combined MPL and SONDE approach (Su et al. 2020) and Doppler Lidar-derived PBLH (Sivaraman and Zhang, 2021) used in the absence of SONDE measurements. Seventy percent of the combined dataset from the first and second columns was randomly selected for the model's training. The third column provides the number of SONDE measurements available for validation purposes. Since morning SONDE serves as the input and boundary condition.

Local Time (h)	SONDE	Supplement Lidar Dataset	SONDE for Validation
5	7163	0	0
6	22	1181	0
7	3	1186	0
8	1225	2541	453
9	16	2629	8
10	9	2732	3
11	6513	13	3307
12	26	2797	9
13	14	2694	47
14	2131	2334	728
15	28	2555	9
16	3	2730	1
17	6503	2	3348

https://doi.org/10.5194/egusphere-2024-376
Preprint. Discussion started: 13 February 2024

Table 3. Feature Importance in the Deep Learning Model. This table presents the importance scores of each input feature used in the deep learning model to estimate the planetary boundary layer height. The features include local time, month, relative humidity, U and V wind components, surface pressure, precipitation, temperature, lifting condensation level (LCL), boundary layer height derived from sensible heat and parcel methods (Sensible Heat BLH and Parcel Method BLH), sensible and latent heat, and profiles of potential temperature (θ) at different heights. The importance scores quantify the relative contribution of each feature to the model's predictive accuracy.

Feature	Importance	Feature	Importance
Local Time	0.001553446	θ 0.45km	0.002378
Month	0.01447574	θ 0.5km	0.002168
RH (i-1)	0.006151263	θ 0.55km	0.002156
RH (i)	0.065531985	θ 0.6km	0.00223
U Wind (i-1)	0.001555849	θ 0.65km	0.001738
U Wind (i)	0.008374529	θ 0.7km	0.001382
V Wind (i-1)	0.010233951	θ 0.75km	0.001251
V Wind (i)	0.009699108	θ 0.8km	0.001533
Surface Pressure (i-1)	0.000757657	θ 0.85km	0.001889
Surface Pressure (i)	0.004098737	θ 0.9km	0.001667
Rain Rate (i-1)	0.000313072	θ 0.95km	0.001062
Rain Rate (i)	0.000442731	θ 1km	0.000533
Temperature (i-1)	0.004147774	θ 1.1km	0.000657
Temperature (i)	0.005575494	θ 1.2km	0.000172
LCL (i-1)	0.001331462	θ 1.3km	-8.3E-05
LCL (i)	0.011779424	θ 1.4km	-0.00047
Sensible Heat BLH (i-1)	0.004322382	θ 1.5km	-8.1E-05
Sensible Heat BLH (i)	0.01068823	θ 1.6km	0.000436
Parcel Method BLH (i-1)	0.035470469	θ 1.7km	0.000855
Parcel Method BLH (i)	0.089339075	θ 1.8km	0.000374
Sensible Heat (i-1)	0.00440638	θ 1.9km	0.000542
Sensible Heat (i)	0.00138861	θ 2km	0.00044
Latent Heat (i-1)	0.005000932	θ 2.2km	-0.00044
Latent Heat (i)	0.006878718	θ 2.4km	-0.00088
θ 0.05km	0.054674179	θ 2.6km	-0.00072
θ 0.1km	0.004824675	θ 2.8km	0.000325
θ 0.15km	0.000101218	θ 3km	0.001006
θ 0.2km	0.000781841	θ 3.2km	0.000577
θ 0.25km	0.001795084	θ 3.4km	0.000799
θ 0.3km	0.002307328	θ 3.6km	0.00064
θ 0.35km	0.003030368	θ 3.8km	0.000747
θ 0.4km	0.00309969	θ 4km	0.004221

https://doi.org/10.5194/egusphere-2024-376
Preprint. Discussion started: 13 February 2024

Figures

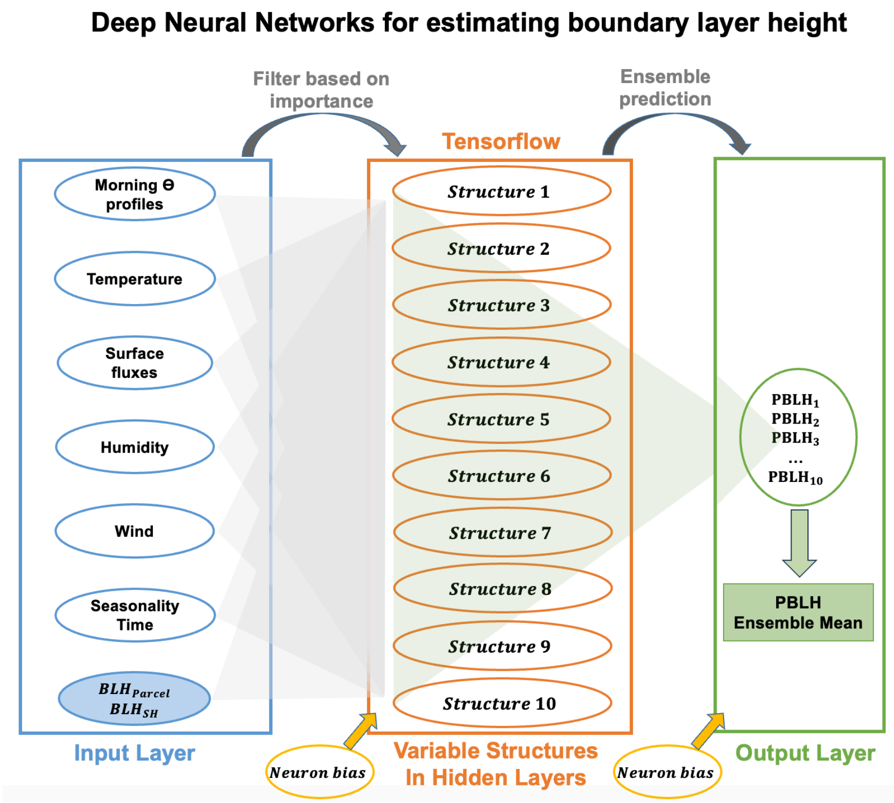


Figure 1. Schematic of the multi-structure deep neural networks (DNN) used for estimating the planetary boundary layer height (PBLH). Input features, including morning potential temperature profiles, temperature, wind, humidity, surface fluxes, seasonality, and time, are filtered based on importance and fed into the network. The system comprises ten distinct hidden layer structures, each processing the inputs to model PBLH. The outputs from these structures are then synthesized to determine the final PBLH value, leveraging the diverse representations of atmospheric dynamics captured by each neural network configuration. Neuron biases are applied at the output and hidden layers to fine-tune the model's performance.

https://doi.org/10.5194/egusphere-2024-376
Preprint. Discussion started: 13 February 2024

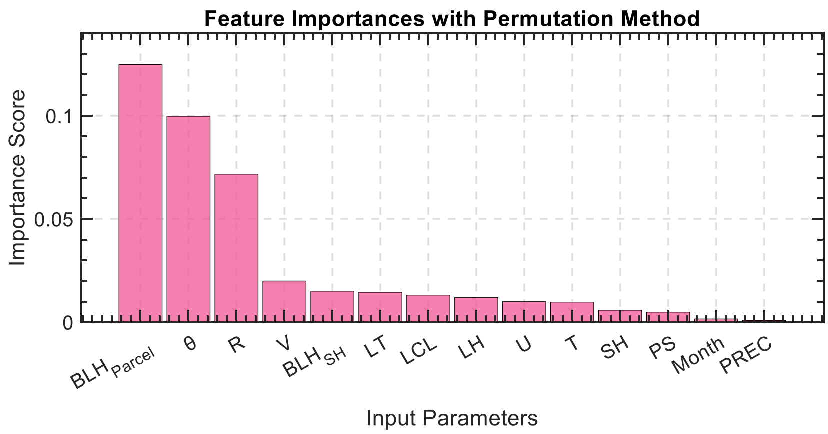


Figure 2. Feature importance with permutation method in the deep learning model. This table presents the importance scores of each input feature used in the deep learning model to estimate the PBLH. The features include local time (LT), month, relative humidity (R), surface U and V wind components, pressure at the surface (PS), precipitation (PREC), surface temperature (T), sensible and latent heat (SH and LH), surface-derived lifting condensation level (LCL), boundary layer height derived from sensible heat and parcel methods (BLH_{Parcel} and BLH_{SH}), and profiles of potential temperature (θ). The importance scores quantify the relative contribution of each feature to the model's predictive accuracy.

https://doi.org/10.5194/egusphere-2024-376
Preprint. Discussion started: 13 February 2024

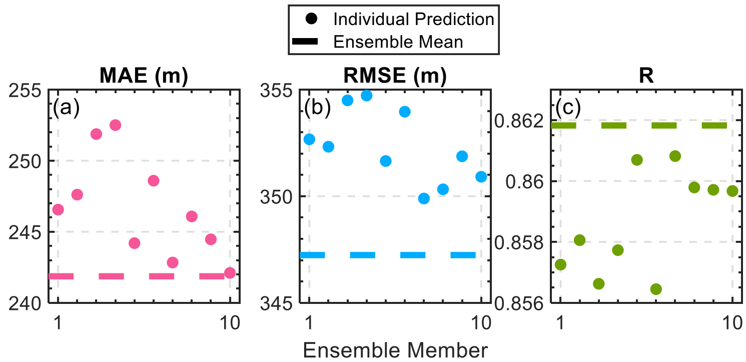
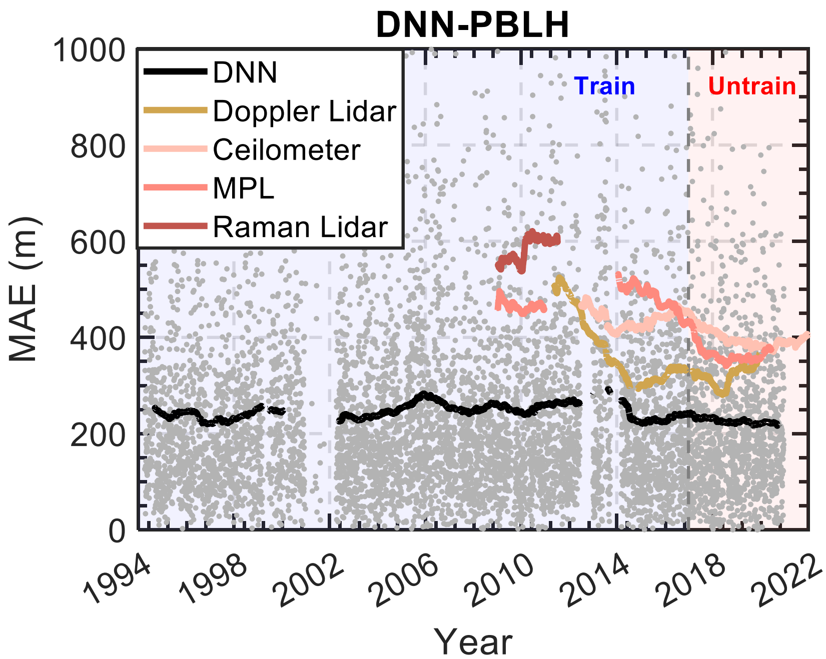


Figure 3: Performance metrics of individual ensemble members and the ensemble mean in estimating planetary boundary layer height (PBLH). Panel (a) displays the mean absolute error (MAE), panel (b) the root mean square error (RMSE), and panel (c) the correlation coefficient (R) for each of the ten ensemble members (represented by dots) and the ensemble mean (indicated by the horizontal dash line). The ensemble approach demonstrates improved accuracy and reliability in PBLH estimation as evidenced by the aggregation of individual model predictions into a robust ensemble mean.

https://doi.org/10.5194/egusphere-2024-376
Preprint. Discussion started: 13 February 2024



822

823 **Figure 4:** Comparative analysis of the mean absolute error (MAE) in PBLH estimation
824 using different methodologies. PBLH derived from SONDE is considered as the ground
825 truth. The DNN approach is shown in black, doppler lidar (Sivaraman and Zhang. 2021)
826 in yellow, ceilometer (Zhang et al. 2022) in pink, micro-pulse lidar (MPL, Sawyer and
827 Li. 2013) in light red, and Raman lidar (Ferrare. 2012) in dark red. DNN model is
828 trained during 1994-2016. Individual MAE values for DNN are represented by gray
829 dots, while the solid lines denote the smoothed MAE for each method with a 2-year
830 smooth window.

https://doi.org/10.5194/egusphere-2024-376
Preprint. Discussion started: 13 February 2024

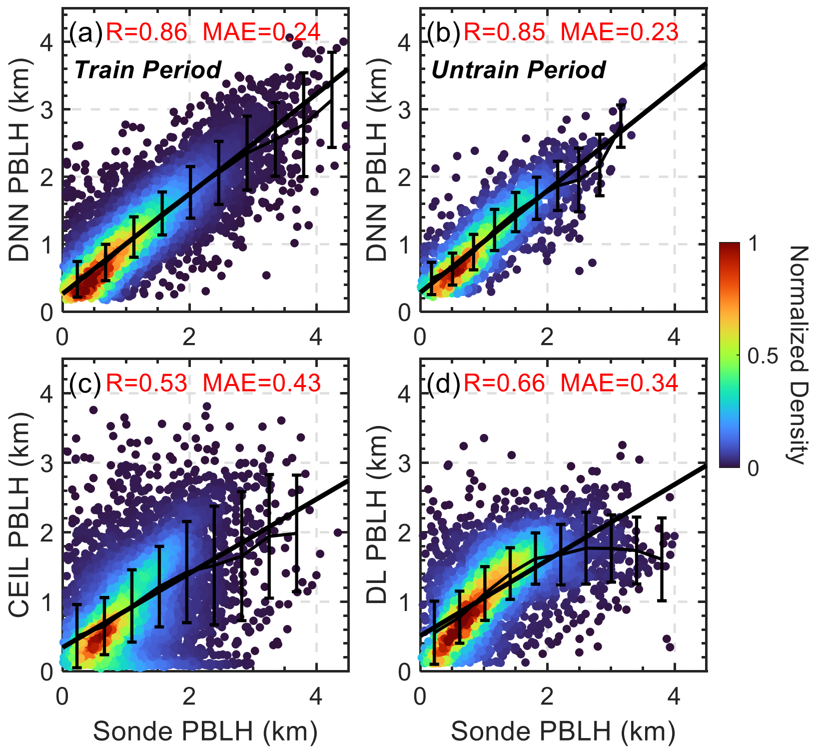


Figure 5: Scatter plots comparing observed radiosonde (SONDE) PBLH with estimates from the machine learning model and lidar observations. Panels (a) and (b) show the PBLH estimated by the deep neural network (DNN) during the trained period (1994-2016) and the untrained period (2017-2020), respectively, with corresponding correlation coefficients (R) and mean absolute errors (MAE). Panels (c) and (d) display comparisons of Sonde PBLH with ceilometer (CEIL) and doppler lidar (DL) derived PBLH, respectively. The color gradient indicates the normalized density of data points, while the solid black line represents the line of best fit and error bars indicates the mean and standard deviations for each bin.

https://doi.org/10.5194/egusphere-2024-376
Preprint. Discussion started: 13 February 2024

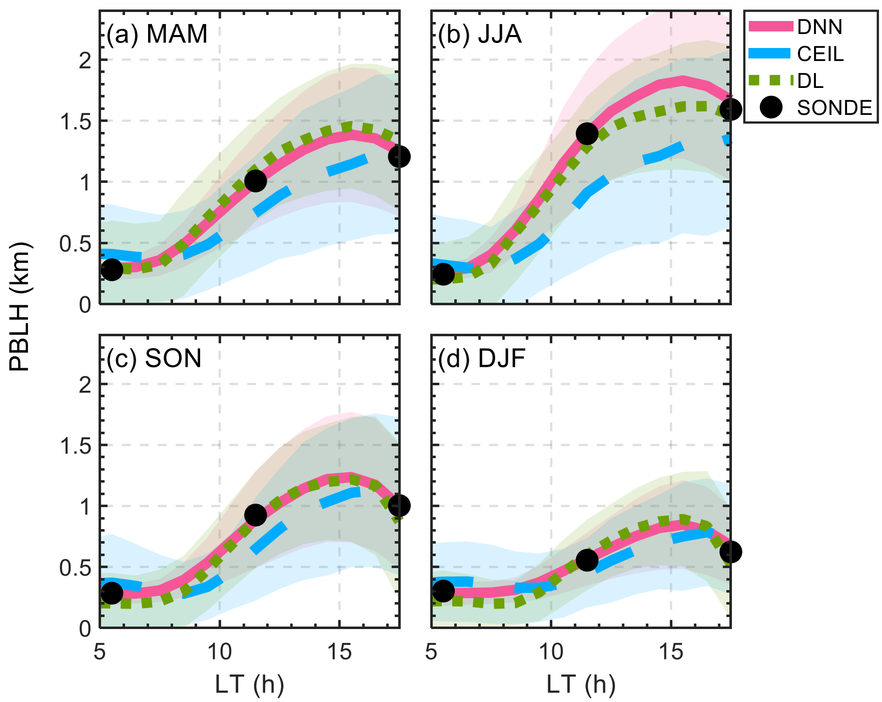


Figure 6: Seasonal-averaged daytime evolution of planetary boundary layer height (PBLH) derived from various methods. The panels represent the mean PBLH values throughout the day for different seasons: (a) March-April-May (MAM), (b) June-July-August (JJA), (c) September-October-November (SON), and (d) December-January-February (DJF). The PBLH values estimated by the deep neural network (DNN) are shown in red, ceilometer (CEIL) estimates in blue, Doppler lidar (DL) in green, and observed radiosonde (SONDE) data in black. Shaded areas around the lines indicate the standard deviations within each method.

https://doi.org/10.5194/egusphere-2024-376
Preprint. Discussion started: 13 February 2024

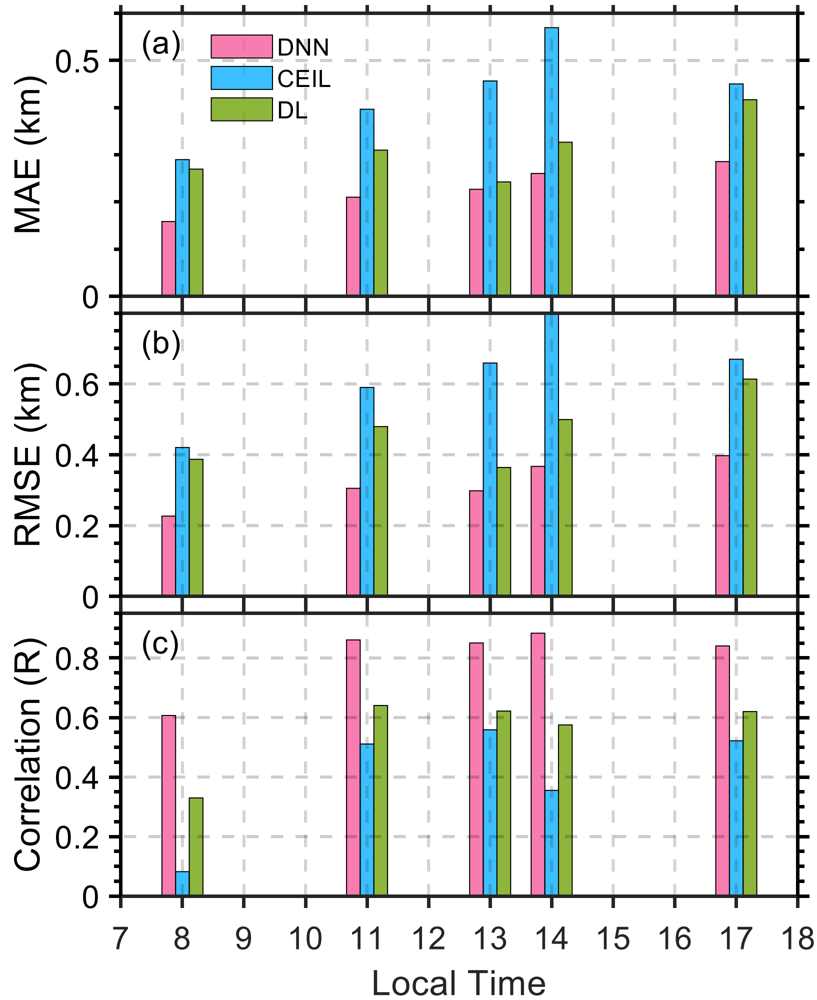


Figure 7: Diurnal variations in the performance metrics for estimating PBLH using different datasets. (a) Shows the correlation coefficient (R), (b) represents the root mean square error (RMSE), and (c) depicts the mean absolute error (MAE) at various local times throughout the day. The deep learning neural network (DNN) estimates are in blue, ceilometer (CEIL) derived estimates are in pink, and doppler lidar (DL) estimates are in green. Note that these biases metrics are calculated using SONDE PBLH as the standard. The availability of SONDE data for different hours is detailed in Table 2.

https://doi.org/10.5194/egusphere-2024-376
Preprint. Discussion started: 13 February 2024

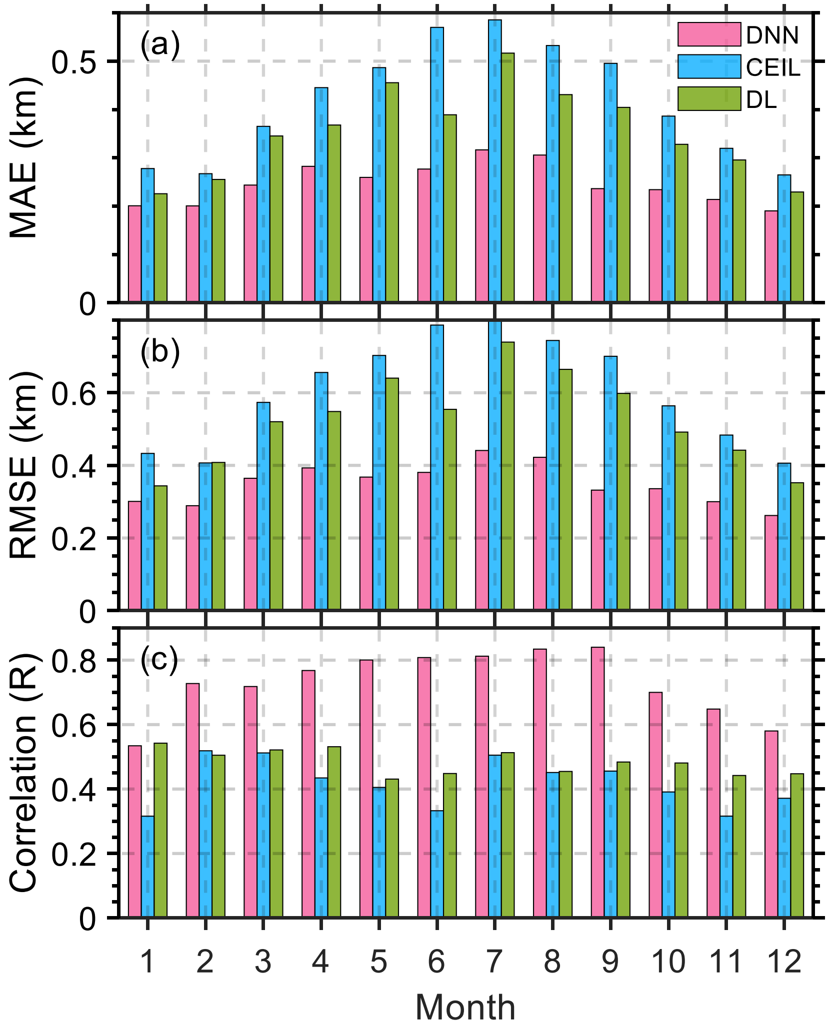


Figure 8: Similar to Figure 7, but for MAE, RMSE, and R for different month.

https://doi.org/10.5194/egusphere-2024-376
Preprint. Discussion started: 13 February 2024

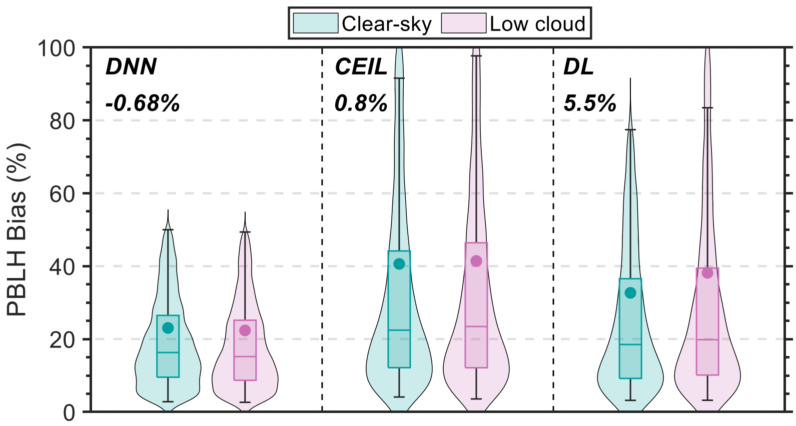


Figure 9: Comparative analysis of PBLH estimation bias under clear-sky and low cloud conditions for various methods. Bias percentages are computed as the absolute bias normalized by the mean PBLH for each condition, with the number above each method indicating the difference in bias between low cloud and clear-sky scenarios. The boxplots detail the 10th, 25th, 50th, 75th, and 90th percentiles, while shaded areas in violin plots illustrate the distribution of dataset biases. The dots indicate the mean value for each condition.

https://doi.org/10.5194/egusphere-2024-376
Preprint. Discussion started: 13 February 2024

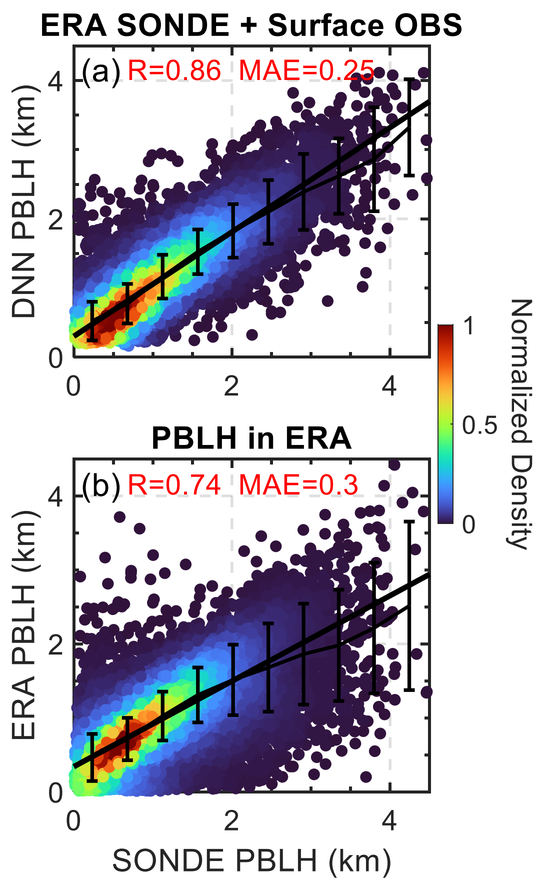
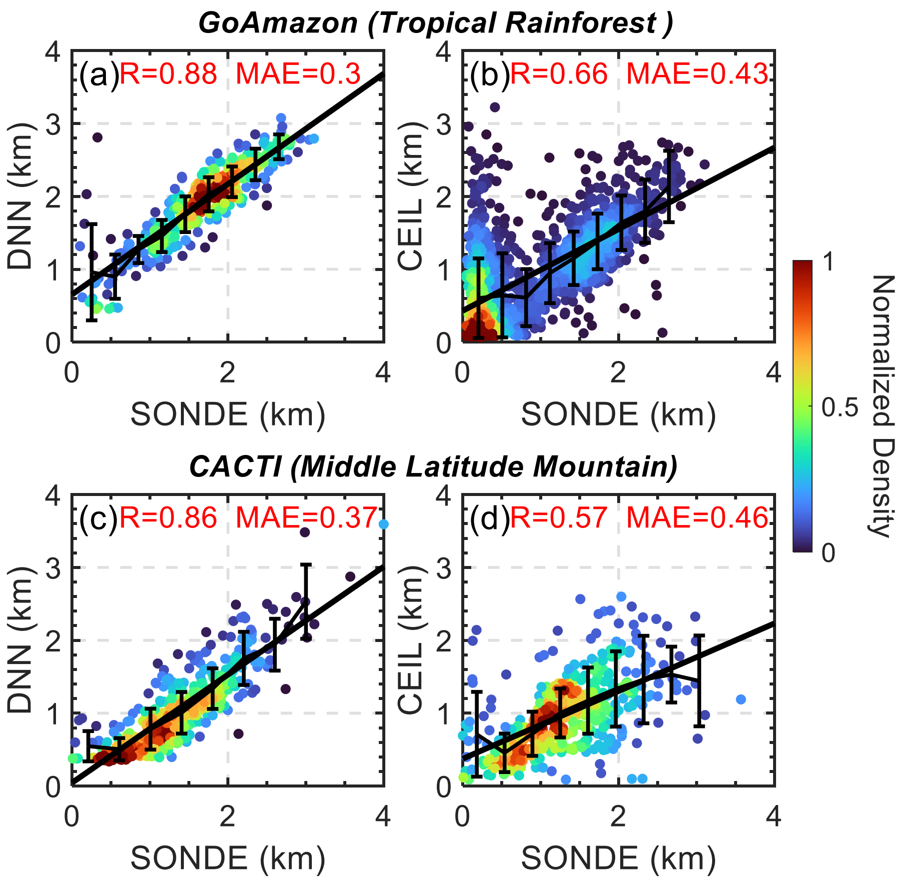


Figure 10: Scatter plots comparing SONDE PBLH with estimates from the DNN and ERA-5. (a) The comparison between observed SONDE PBLH and estimates from the DNN model, which utilizes morning temperature profiles (5 LT) from ERA-5 (ERA Profile) and observed surface meteorological data (surface OBS) as inputs. (b) The correlation comparison observed SONDE PBLH and PBLH model outputs from the ERA-5 datasets. The color gradient in both panels represents the normalized density of data points, while the solid black line indicates the linear regression, and the error bars denote the mean and standard deviations for each bin.

https://doi.org/10.5194/egusphere-2024-376
Preprint. Discussion started: 13 February 2024



879

880 **Figure 11:** Validation of the DNN trained over the SGP for the GoAmazon (Tropical
881 Rainforest) and CACTI (Middle Latitude Mountain) field campaigns. Panels (a) and (c)
882 illustrate the correlation (R) and mean absolute error (MAE) between DNN predictions
883 and SONDE observations for GoAmazon and CACTI, respectively. Panels (b) and (d)
884 show the performance of ceilometer (CEIL) derived PBLH compared to SONDE for
885 the same campaigns. The color gradient indicates the normalized density of data points,
886 while the solid black line represents the line of best fit and error bars indicates the mean
887 and standard deviations for each bin.

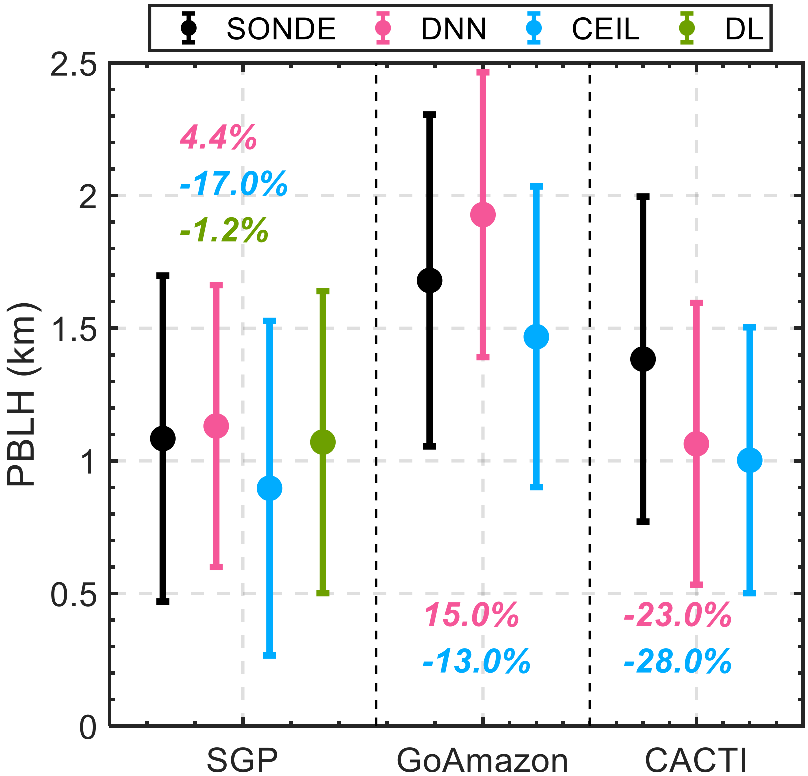


Figure 12: Comparative PBLH mean (dots) and standard deviations (error bars) across ARM sites (SGP, GoAmazon, and CACTI). The datasets are derived from radiosonde (SONDE, in black), the DNN model (in pink), ceilometer (CEIL, in blue), and Doppler lidar (DL, in green), respectively. Noted the DL-derived PBLH is only available at the SGP. The percentages in various colors denote the differences in PBLH means derived from the DNN, CEIL, and DL methods relative to SONDE observations. To mitigate sampling bias, these mean values and standard deviations are computed exclusively for intervals where all instruments have concurrently available data.