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

Submission to *Atmospheric Chemistry and Physics*

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

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, which can estimate PBLH by integrating the morning temperature profiles and 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 processes with implications for enhancing the representation of PBL in weather forecasting and climate modeling.

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 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., 2024a; Tucker et al., 2009; Wang et al. 2020). Due to its impact on cloud evolution and the development of convective systems, PBLH is also a key parameter in numerical weather forecasts and climate projections (Deardorff, 1970; Kaimal et al. 1976; Menut et al., 1999; Park et al., 2001; Emanuel, 1994; Guo et al., 2017, 2019; 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 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;

65   Wang et al., 2023). The remote sensing systems, including Doppler lidar (Barlow et al.

66   2011), ceilometer (Zhang et al. 2022), Raman lidar (Summa et al. 2013), and Micro-

67   pulse lidar (Melfi et al., 1985), utilize laser-based technology to track PBLH diurnal

68   evolutions, helping us understand the PBL evolutions (Cohn and Angevine, 2000;

69   Davis et al., 2000). In addition, wind profilers can estimate the PBLH using algorithms

70   that analyze the signal-to-noise ratio from wind profiler data (Molod et al. 2015; Solanki

71   et al. 2022; Liu et al. 2019; Salmun et al. 2023; Bianco and Wilczak 2002; Bianco et al.

72   2008; Tao et al. 2021).

73   However, the advancement in remote sensing for the estimation of PBLH

74   challenges is still posing in bridging the results obtained by different remote sensing

75   instruments with those obtained from the SONDE measurements (Zhang et al. 2022;

76   Chu et al., 2019). Specifically, interpreting aerosol, turbulence, and moisture profiles

77   derived from remote sensing techniques to determine PBLH bears inherent limitations

78   due to the unstable signal-to-noise ratio (Kotthaus et al., 2023; Krishnamurthy et al.,

79   2021). This issue is compounded by the different measurement methodologies and

80   definitions employed by various remote sensing tools, leading to uncertainties when

81   comparing their PBLH estimates to the retreivals derived from SONDE measurements

82   (Zhang et al. 2022; Sawyer and Li, 2013).

83   As machine learning (ML) has shown potential in atmospheric science (McGovern

84   et al., 2017; Gagne et al., 2019; Su et al. 2020a; Vassallo et al., 2020; Cadeddu et al.,

85   2009; Molero et al. 2022), this technique presents a promising tool for refining the

86   estimation of PBLH to resolve the inherent complexity and variability of PBL. For

87 example, several studies use ML to identify PBLH using thermodynamic profiles

88 Atmospheric Emitted Radiance Interferometer (AERI) or using backscatter profiles

89 from lidar, highlighting the ML's superiority over conventional techniques under

90 different scenarios (Sleeman et al. 2020; Rieutord et al. 2021; Liu et al. 2022; Ye et al.

91 2021). For example, Li et al. (2023) applied an ML algorithm for retrieving PBLH

92 under complex atmospheric conditions with account of the vertical distribution of

93 aerosols. Krishnamurthy et al. (2021) incorporated a random forest model, along with

94 machine learning, to use Doppler lidar data for the extraction of PBLH with better

95 results compared to the results retrieved by traditional methods.

96   While existing ML methodologies have made great progress in estimating PBLH,

97 these studies mainly focus on refining retrievals from remote sensing data, particularly

98 lidar-based technologies. Thus, there is an inherent limitation to the applicability due to

99 a reliance on specific remote sensing instruments. To address this issue, we aim to

100 leverage and integrate the comprehensive field observations (i.e., radiosonde and

101 remote sensing techniques) to develop a deep learning model for direct PBLH

102 estimation from conventional meteorological data. This strategy circumvents the

103 limitations of relying on particular remote sensing technologies. Furthermore, our

104 model employs an advanced deep neural network (DNN) approach (Sze et al. 2017;

105 Schmidhuber, 2015; Nielsen, 2015; Pang et al. 2020), diverging from traditional ML

106 methods like random forest. This deep learning model utilizes ensemble techniques,

107 constructing arrays of various structures and using their average for the final estimation.

108 This approach method provides particular advantages in the context of complex and

109 nonlinear processes (Ganaie et al. 2022; Mohammed and Kora. 2023). Ensemble DNN

110 with multi-structure designs shows very strong flexibility and robustness, so it

111 relatively performs better and has high stability across a wide range of conditions (Xue

112 et al. 2020; Dong et al. 2020). This facilitates the adaptability of DNN as a tool for

113 PBLH estimation, which can be utilized under different scenarios and locations.

114 By focusing on the interaction between surface meteorology and the PBL, this study

115 introduces a DNN-based method to estimate the daytime evolution of PBLH from

116 morning temperature profiles and surface meteorology. We evaluate the model's

117 performance using extensive datasets over the Southern Great Plains (SGP) for a period

118 spanning 27 years (1994-2020) and includes comparisons with PBLH estimations

119 obtained from Doppler lidar, ceilometer, Raman lidar, and micro-pulse lidar.

120 Furthermore, we explore the generalizability of the model to different geographic

121 regions and climates, as tested during the field campaigns, e.g., Green Ocean Amazon

122 (GoAmazon) and Cloud, Aerosol, and Complex Terrain Interactions (CACTI).

123

124 **2 Data and instruments**

125 **2.1 ARM Sites**

126 The Atmospheric Radiation Measurement (ARM) program, funded by the U.S.

127 Department of Energy, has been employed at the Southern Great Plains (SGP) site in

128 Oklahoma (36.607°N, 97.488°W), situated 314 meters above mean sea level. This study

129 use comprehensive field observations at the SGP site during 1994 to 2020. In addition

130 to the SGP site, this study utilizes data from the ARM GoAmazon (3.213°S, 60.598°W)

131   and ARM CACTI (32.126°S, 64.728°W) field campaigns to carry out indenpendant

132   tests for the deep learning model. Specificly, the GoAmazon campaign is located in the

133   amazon tropical forests and provides rich field observations data during 2014-2015

134   (Martin et al. 2016). Meanwhile, the CACTI central site, at an elevation of 1141 meters

135   within the Sierras de Córdoba Mountain range in north-central Argentina, offers the

136   observations during the 2018-2019 period (Varble et al. 2021). Utilizing these

137   comprehensive ARM datasets, our study includes thermodynamic profiles derived from

138   radiosondes, data from the Active Remote Sensing of Clouds (ARSCL, Clothiaux et al.

139   2000, 2001; Kollias et al. 2020), in-situ surface flux measurements, and standard

140   meteorological observations at the surface, as documented by Cook (2018) and Xie et

141   al. (2010).

142   SONDE measurements at the ARM sites launch routinely several times a day and

143   provide detailed information into the thermodynamic conditions of the atmosphere. The

144   technical details of the ARM SONDE data are documented in Holdridge et al. (2011).

145   Moreover, we use the surface meteorological parameters at the standard meteorological

146   station. In-situ measurements at 2 meters above ground level provide data on

147   temperature, relative humidity, and vapor pressure. Moreover, this study obtain the

148   surface sensible and latent heat fluxes from the surface instruments (Wesely et al.,

149   1995). In SGP, we use the best-estimate surface fluxes in the Bulk Aerodynamic Energy

150   Balance Bowen Ratio (BAEBBR) product, which is derived from the measurements by

151   Energy Balance Bowen Ratio (EBBR). Due to the availability, we utilize the surface

152   fluxes from Quality Controlled Eddy CORrelation (QCECOR) datasets from CACTI

153   and GoAmazon sites (Tang et al. 2019).

154

155   **2.2 Existing PBLH datasets over the ARM sites**

156   For analyzing PBLH, we have utilized a variety of datasets to get a full picture of

157   PBLH derived from different instruments. These datasets are developed by using

158   different methodologies and instruments and jointly offer a detailed information of

159   PBLH under various meteorological conditions. Among these datasets, SONDE- and

160   ceilometer-derived PBLH are available for all three sites, other datasets are only

161   available over the SGP. The technique details for these datasets can be found in the

162   corresponding publications or technique reports.

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

164   PBLHs are retrieved using a method developed by Liu and Liang (2010), based on

165   potential temperature gradients from SONDE. We focus on daytime data during 05:00–

166   18:00 Local Time (LT), with a resampled vertical resolution of 5-hPa. The SONDE

167   dataset is available at DOI: https://doi.org/10.5439/1595321.

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

169   Doppler lidar PBLH estimates are derived using a vertical velocity variance method

170   during 2010-2019 (Tucker et al., 2009; Lareau et al., 2018; Sivaraman and Zhang 2021).

171   The dataset is available at DOI: https://doi.org/10.5439/1726254.

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

173   We utilize a PBLH dataset that merges lidar and SONDE measurements during

174   1998-2023, ensuring vertical coherence and temporal continuity (Su et al. 2020b). An

175   additional method for handling cloudy conditions is detailed in Su et al. (2022). The

176   dataset is available at DOI: https://doi.org/10.5439/2007149.

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

178   The Vaisala CL31 ceilometer, with a 7.7 km vertical range, provides detailed

179   backscatter profiles used for PBLH estimation via gradient methods during 2011-2023

180   (Zhang et al. 2022). Enhanced algorithms ensure robust estimations under all weather

181   conditions. The dataset is available at DOI: https://doi.org/10.5439/1095593.

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

183   Micropulse lidar (MPL) is utilized for its high temporal resolution to retrieve PBLH

184   during 2009-2020. MPL-derived PBLH, validated against SONDE and AERI data,

185   improves understanding of boundary-layer processes (Sawyer and Li. 2013). The

186   dataset is available at DOI: https://doi.org/10.5439/1637942.

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

188   PBLH is calculated using merged potential temperature profiles from Raman lidar

189   and AERI, with criteria established for the SGP site. PBL heights are computed hourly

190   for 2009-2011. The dataset is available at DOI: https://doi.org/10.5439/1169501.

191   In the datasets, (1-3) serve as the foundation for training. Concurrently, considering

192   radiosonde as the benchmark standard, we utilized dataset (1) for validating PBLH

193   retrievals obtained from various sources. Meanwhile, datasets (4-6) are used for the

194   intercomparisons between PBLH derived from DNN and remote sensing techniques.

195

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. The DNN model derives the PBLH from surface meteorological parameters. We also incorporate boundary layer heights derived from sensible heat and parcel methods (BLH_{Parcel} and BLH_{SH}) as inputs. Specifically, BLH_{Parcel} is calculated based on the morning profile of potential temperature (Holzworth. 1964), while BLH_{SH} is determined using the surface temperature combined with surface sensible heat, following the methodologies (Stull, 1988; Su et al. 2023). We first present a preliminary run for the model to obtain the importance of each input feature. Then, these inputs undergo a filtration 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

217 PBLH in its unique way (Ganaie et al. 2022). The ensemble employs a three-layer base

218 structure [52, 28, 16] for neural networks, from which ten unique configurations are

219 derived by applying random perturbations to the default settings of the base structure.

220 These different structures for ensembles 1-10 are presented in Table 1.

221 At the final stage, the model use the PBLH esimations from different ensembles to

222 get a mean value as the final PBLH retrieval. This process allows the model to leverage

223 the different results of all structures and enhance the generalizability of results. In the

224 DNN model, neuron biases in the output and hidden layers are important for the

225 network's architecture (Battaglia et al. 2018). These biases serve as fine-tuning

226 parameters to adjust the activation thresholds of neurons in different layers and further

227 refine the model's predictive capabilities. Neuron biases are initialized with small

228 random values at the start of the training process and then iteratively adjusted according

229 to the network weights during the training. Normalization is a preprocessing technique

230 that often leads to improvements in model training by scaling the input features and

231 target values to a standard range (Raju et al. 2020). The normalization process was

232 applied to each input data to ensure that they have a zero mean and a standard deviation

233 of one, as well as the target data. This standardization scales the different input data to

234 a similar range, and thus, contributes a more stable and efficient training process.

235 The hidden layers of the DNN model incorporate L2 regularization to curtail

236 overfitting, while batch normalization aids in stabilizing learning. Moreover, a dropout

237 rate of 0.2 helps the model to generalize better by reducing reliance on any specific

238 neurons during training. We choose the Adam optimizer and mean squared error as the

239 loss function, which aligns with one of the best practices for regression models (Zhang.

240 2018). The mean absolute error is selected as a metric to evaluate the model's accuracy

241 during the training. We incorporate the early stopping and learning rate reduction

242 callbacks in in the model's training for regularization and fine-tuning (Liu et al. 2019).

243 Such measures ensure optimal performance by terminating training at the right juncture

244 and avoid the overfitting in the final results.

245

246 **3.2 Training the DNN Model**

247 The training of the DNN model was conducted using a PBLH dataset enriched by

248 SONDE and lidar measurements during 1994 to 2016 over the SGP. Table 2 presents

249 the distribution of dataset samples under different local time, which were important for

250 both the training and validation processes of the DNN model. The primary dataset (i.e.,

251 PBLH derived from SONDE measurements) is listed in the first column and are

252 available routinely for 5, 11, and 17 LT. The training dataset was augmented with the

253 combined MPL-SONDE PBLH dataset (Su et al. 2020b) and the Doppler Lidar-derived

254 PBLH (Sivaraman and Zhang, 2021) to address the gaps where SONDE measurements

255 were not available. In instances where radiosonde data are unavailable, the lidar datasets

256 are used for training, contingent upon their agreement with radiosonde measurements

257 within a margin of 0.2 km over a 3-hour window. Specifically, out of the total

258 comparisons during the study period, 40.2% of the lidar measurements do not agree

259 within the 0.2 km threshold with the SONDE results. The cases with relatively larger

260 inconsistencies stem from various factors, including instrumental errors, rainy

261 conditions, stable PBL conditions, differing definitions, and lidar signal attenuation, as

262 discussed in previous studies (Su et al., 2020b; Kotthaus et al., 2023). These cases were

263 excluded from the DNN model training to maintain the quality of the process.

264 For the purpose of training the DNN model, 70% of the hourly data from both

265 SONDE and the lidar combined dataset were randomly selected. The remaining 30%

266 dataset, comprises the portion of SONDE measurements set aside for validation

267 purposes, including a separate subset from the years 2017 to 2020 to test the model's

268 predictive capabilities on independent data. This training and validation scheme ensures

269 that the DNN model is not only well-trained but also thoroughly evaluated, reinforcing

270 its reliability in accurately estimating PBLH. As morning SONDE data constitute the

271 primary input and boundary conditions for the model, the validation of PBLH retrievals

272 is consequently confined to the 08:00 to 18:00 LT.

273

274 **3.3 Feature Importance Score**

275 In the DNN model, we quantified the significance of each input parameter using the

276 permutation importance technique, which is a widely-used method for the deep learning

277 (Date and Kikuchi, 2018; Altmann et al. 2010; Breiman, L., 2001). Initially, we carry

278 out a test run to determine a baseline performance by calculating the mean absolute

279 error (MAE) on the validation set. Then, each feature within this set was then

280 individually shuffled, severing its correlation with the target PBLH, and the MAE was

281 recalculated. Compared to the baseline performance, the increase in MAE from this

282 shuffled state indicates the feature's predictive value: the greater the increase, the more

283   significant the feature. We repeat this shuffling and evaluation for 15 times, each with

284   a unique random seed to ensure statistical robustness. Furthermore, we calculated the

285   average MAE increase across these iterations as the importance score. These scores are

286   expressed as percentages, with each feature's importance score normalized to sum to

287   100%. Each score quantitatively represents how much the shuffling of a feature

288   increases the MAE, indicating the relative significance of that feature in the model's

289   predictive accuracy and facilitating a straightforward comparison of the influence of

290   each feature within the model. Therefore, we derived a composite importance metric

291   for feature groups to represent their significance as the cumulative sum of reltaed inputs.

292   Figure 2 presents the importance scores to demonstrate the relative influence of

293   different feature groups on the model's performance. Prominently, features such as the

294   BLH_{Parcel}, morning potential temperature profiles (θ profile), and surface relative

295   humidity are identified as most important three features, with their substantial impact

296   on the accuracy of PBLH estimation being highlighted. BLH_{Parcel} is defined as the

297   height where the morning potential temperature first exceeds the current surface

298   potential temperature by more than 1.5 K (Holzworth, 1964; Chu et al., 2019). Among

299   these features, BLH_{Parcel} captures the response of the PBL to surface heating, which

300   can drastically affect local convection and thus serves as one of the key parameters in

301   the DNN model. Incorporating this parameter and its association with PBL

302   development better simulates diurnal variations of PBLH in the DNN model.

303   Meanwhile, the morning θ profile represents the vertical stratification of

304   thermodynamics and is essential for understanding stability and mixing processes

305   within the PBL. Thus, θ profile serves as the initial boundary condition for the PBLH

306   estimation with a significant importance score. Surface relative humidity also emerges

307   as a key influencer, affecting the model's performance significantly. Humidity levels

308   influence the condensation and evaporation processes within the PBL, which are

309   important in determining its vertical extent layer and structure. Fair-weather and dry

310   conditions are typically associated with a more turbulent and higher PBL. Conversely,

311   high surface humidity often contributes to the formation of boundary layer clouds,

312   which introduces complex interactions with PBL thermodynamics.

313   In this analysis, each feature, such as θ profile, comprises several different inputs,

314   and the relative importance scores presented in Figure 2 are calculated as the cumulative

315   sum of these inputs. Complementing this, Table 3 offers an exhaustive breakdown of

316   importance scores for all considered input features within the deep learning model. In

317   refining the model, features contributing a negligible or negative effect on performance

318   (i.e., importance scores less than zero) are excluded. As a result, this selection criterion

319   has led to the inclusion of 58 out of the original 64 features. This process ensures we

320   only use inputs with a proven positive influence in the DNN model.

321

322   **4   Evaluation of Deep Learning Model**

323   **4.1 Comparative analysis of biases among different datasets**

324   A critical component of evaluating our deep learning model's efficacy is analyzing

325   the biases of individual ensemble members and their collective output. Figure 3 offers

326   a visual assessment of the mean absolute error (MAE), root mean square error (RMSE),

327   and correlation coefficient (R) for each ensemble member, alongside a comparison with

328   the ensemble mean (average of all individual ensemble members). The plotted data

329   points reveal the variation in performance across different model architectures, while

330   the ensemble mean, represented by the horizontal dashed lines, indicates the collective

331   accuracy of the ensemble approach. The structures of different hidden layer

332   configurations are listed in the Table 1.

333   This methodological consolidation results in a more reliable and accurate PBLH

334   estimation, leveraging the strengths and mitigating the weaknesses of individual

335   models. By integrating multiple neural network configurations, we revealed that an

336   ensemble prediction that consistently outperforms the individual models. This strategy

337   can improve the MAE by up to 4.4%, rendering the model less dependent on any

338   specific structural configuration.

339   An in-depth comparative analysis of biases among various PBLH estimation

340   methods is essential for validating the reliability and accuracy of the DNN developed

341   in this study. Figure 4 illustrates the MAE trends for several methods over a multi-year

342   span, with the SONDE-derived PBLH serving as the benchmark for ground truth. The

343   analysis reveals the performance of different methodologies: the DNN approach,

344   doppler lidar, ceilometer, MPL, and Raman lidar. Significantly, the DNN model,

345   depicted in black, maintains a consistent MAE trend throughout the trained period

346   (1994-2016) as well as the subsequent untrained period (2017-2020), demonstrating

347   robust predictive stability. In contrast, the remote sensing-based methods show a

348   reduction in bias from 2010 to 2022, possibly due to the improvement of remote sensing

349 data quality. The discrepancy in PBLH estimates between the DNN and SONDE

350 remains consistently lower than those observed with conventional remote sensing

351 techniques.

352   Figure 5 provides a detailed evaluation of the DNN model in comparison to

353 ceilometer and doppler lidar-derived PBLH, as these two methods have demonstrated

354 the high quality with more than nine years of datasets. Figure 5a-b contrast the PBLH

355 predictions from the DNN model for both the trained period (1994-2016) and untrained

356 periods (2017-2020), respectively, showcasing strong correlations and low MAEs,

357 indicative of the model's robust training and generalization capabilities. Figure 5c-d

358 further this examination with ceilometer and Doppler lidar comparisons, respectively.

359 Overall, Doppler lidar exhibits a closer alignment with SONDE-derived PBLH than the

360 ceilometer. However, the MAE from Doppler lidar-based estimates is still

361 approximately 48% higher than those derived from the DNN model. The correlation

362 coefficient for the DNN-derived PBLH estimates has seen a substantial improvement,

363 rising from the 0.5-0.6 range typically observed with remote sensing-based PBLH

364 methods to exceed 0.8 when compared to SONDE-derived PBLH measurements. This

365 comparative analysis not only confirms the DNN model's accuracy but also offers

366 insights into the relative performance of various contemporary PBLH estimation

367 methodologies.

368

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

391   model shows a relatively good correlation with SONDE retrievals (above 0.6 under

392   different hours). This comparison shows the efficacy of DNN in tracking the diurnal

393   cycle of PBLH.

394   Continuing our assessment of the DNN model, we analyze the DNN model's

395   monthly performance in estimating PBLH, as shown in Figure 8. The analysis compares

396   MAE, RMSE, and correlation coefficients for each month to assess the model's

397   precision and dependability. The summer months (June-July-August) exhibit higher

398   biases, with MAE values for the DNN, ceilometer, and Doppler lidar at 0.3 km, 0.56

399   km, and 0.45 km, respectively. In contrast, the winter months (December-January-

400   February) show reduced biases, with MAE values of 0.2 km for the DNN, 0.27 km for

401   the ceilometer, and 0.24 km for the Doppler lidar. Specifically, the DNN model shows

402   a much lower bias during the summer season. Compared to the remote sensing-based

403   retrievals, the DNN-derived PBLH shows a much better agreement with SONDE-

404   derived PBLH, increasing from 0.3-0.6 to approximately 0.8 in term of correlation

405   coefficients.

406   Figure 9 presents the biases of PBLH retrievals under clear-sky and low cloud

407   conditions. We calculated biases as the absolute deviation from the mean PBLH for

408   each condition, focusing particularly on the differences between low cloud (maximum

409   cloud fraction between 0-4 km exceeding 1%) and clear-sky (total cloud fraction below

410   1%) scenarios. The threshold of 1% for cloud fraction is also used to identify cloud base

411   height (CBH) in the European Centre for Medium-Range Weather Forecasts' fifth-

412   generation global reanalysis (ERA-5, Hersbach et al., 2023). The violin plots in this

413   figure illustrate the data distribution of biases for each method to demonstrate their

414   variability. For the DNN model and ceilometer, the relative biases between clear and

415   cloudy conditions are comparable and the difference is less than than 1%. This suggests

416   a consistent performance across these atmospheric states. However, the Doppler lidar

417   exhibits a larger disparity, showing a 5.5% bias under cloudy conditions compared to

418   clear skies. Moreover, the spread of biases (shaded areas and error bars) is notably wider

419   for both the ceilometer and Doppler lidar. This indicates large variability in their

420   performance. For all three methods, the mean biases are notably higher than the median

421   values. Such differences indicate that the mean values are notably influenced by outliers

422   under both clear-sky and cloudy conditions.

423   The evolution of the PBLH under shallow cumulus conditions offers insights into

424   the interactions between clouds, PBL, and land surface (Zhang and Klein, 2010, 2013).

425   Figure 10 demonstrates the variations of PBLH measurements from different methods

426   during conditions typical of shallow cumulus clouds. Shallow cumulus clouds were

427   identified following Su et al. (2024b). Specifically, these coupled clouds form post-

428   sunrise; and the sky must not be overcast, characterized by a cloud fraction less than

429   90%. This selection criterion ensures that the observed cloud formations are primarily

430   driven by surface heating and local convection. The DNN model closely matches the

431   SONDE-derived PBLH and the CBH from ARSCL. This alignment underscores the

432   physical validity of the DNN approach, confirming its capability to replicate traditional

433   measurement techniques to a good extend of accuracy. Meanwhile, Doppler lidar-

434   derived PBLH retrievals also show high consistency with SONDE measurements,

435 whereas ceilometer-derived PBLH generally underestimates values under shallow

436 cumulus conditions.

437 Figure 10 also demonstrates the general relationship between the development of

438 shallow cumulus clouds and the PBL, which are driven by local convection and

439 turbulence. The formation of these cumulus clouds is linked to rising thermals and an

440 increase in surface heat fluxes, essential for driving vertical mixing within the sub-cloud

441 layer. This relationship is evidenced by the increased occurrence of cumulus clouds

442 along with an increase in DNN-derived PBLH from morning to late afternoon.

443 Specifically, during periods with a high frequency of shallow cumulus, the DNN-

444 derived PBLH often surpasses the CBH. This indicates that rising air parcels extend

445 beyond the condensation level, facilitating the formation and development of coupled

446 cumulus clouds.

447 In this context, these analyses confirm the physical consistency of DNN-derived

448 PBLH with traditional measurement techniques and highlight its physically reasonable

449 variations during cloudy conditions. The results presented in this section illustrate the

450 effectiveness of the DNN model in capturing the PBLH variations across different local

451 times, seasons, and cloudy conditions. Compared to the traditional remote sensing

452 methods, the DNN model exhibits relatively good accuracy in aligning with SONDE-

453 derived PBLH, indicating its capability and stable performance under different

454 scenarios.

455

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 ERA-5 (Hersbach et al., 2020) dataset when radiosonde data is not available to maintain the model's operational integrity for the conditions without SONDE data. As one of the 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 11 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 11a 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.

478   These findings highlight the alternative DNN model's robustness, offering a reliable

479   substitute for radiosonde data by leveraging reanalysis data with similar performance.

480   This demonstrates the DNN model's adaptability and potential as a practical tool for

481   PBLH estimation across various meteorological sites, especially in regions or periods

482   where radiosonde data may be lacking.

483   We further test the adaptability and generalizability of the DNN model, by applying

484   across different climatic and geographic regions. To this end, we extended our model

485   evaluation to include SONDE and surface meteorological data from the GoAmazon

486   (Tropical Rainforest) and CACTI (Middle Latitude Mountain) field campaigns.

487   Seasonality is accounted for as an input variable in the DNN model, with months in the

488   Southern Hemisphere adjusted to reflect their Northern Hemisphere seasonal

489   counterparts (e.g., July inputs are treated as January). The normalization process

490   (Section 3.1) was reapplied for the CACTI campaign data to adjust for notable pressure

491   level variations, ensuring input standardization with zero mean and unit variance.

492   Figure 12 presents the model's performance, in comparison to SONDE observations

493   for both GoAmazon and CACTI campaigns. The DNN model demonstrates

494   commendable adaptability, maintaining a strong correlation (0.86-0.88) with SONDE

495   measurements (Figure 12a-b). Further comparison is provided, which assess the

496   performance of ceilometer derived PBLH against SONDE for the same campaigns.

497   When assessing the performance of the ceilometer-derived PBLH against SONDE for

498   the same campaigns, the DNN model exhibited both stronger correlations and smaller

499   biases, as shown in Figure 12b-d.

500   Nevertheless, the analysis highlighted the presence of systematic biases, with

501   relatively larger MAE at the GoAmazon and CACTI sites compared to the SGP site.

502   Figure 13 underscores this by presenting a comparative analysis of PBLH means and

503   standard deviations across the three ARM sites. The early morning measurements

504   during 05-07 LT are excluded. The results, derived from SONDE, the DNN model,

505   ceilometer, and Doppler lidar data, reveal average differences in PBLH means relative

506   to SONDE. These differences suggest an overestimation (+15%) and underestimation

507   (-23%) by the DNN model for the GoAmazon and CACTI sites, respectively, compared

508   to the more consistent PBLH values at the SGP site.

509   The evident systematic deviations when applying the SGP-trained DNN model to

510   the diverse environments of GoAmazon and CACTI underscore the challenges in

511   generalizing the model to regions with significantly different meteorological

512   backgrounds. These findings point to the potential of DNN models for PBLH estimation

513   while also highlighting the necessity for region-specific model adjustments.

514

515   **5 Summary**

516   This study has developed a Multi-Structure DNN model for estimating PBLH using

517   conventional meteorological data. The DNN model is developed by leveraging a long-

518   term dataset of PBLH derived from radiosonde data and augmented with high-

519   resolution MPL and Doppler lidar observations. This model produced an PBLH dataset

520   over the SGP with robust accuracy, consistently yielding lower bias values across

521   various conditions and datasets. Utilizing conventional meteorological data, this

522   method generates a 27-year dataset over the SGP, encompassing periods with limited

523   remote sensing data availability. In situations where morning radiosonde data is

524   unavailable, ERA-5 data can be effectively employed to initiate the model, offering a

525   practical alternative.

526   An important aspect of this research involved comparing DNN models with diverse

527   remote sensing instruments. Although these instruments offer high temporal and

528   vertical resolution, discrepancies in PBLH estimation remain. Our DNN model,

529   leveraging a broad range of input features refined by their importance, constructs a

530   representation of PBL evolutions, frequently demonstrating a closer agreement with

531   SONDE-derived PBLH. In the absence of remote sensing data, the DNN model can

532   produce high-quality PBLH results from the conventional meteorology data.

533   The study has shown the DNN model's ability to synthesize complex patterns from

534   meteorological data, reflecting the versatility of machine learning in simulating the

535   boundary layer processes. Its application to varied geographic terrains and climates

536   during the GoAmazon and CACTI campaigns has further validated its adaptability,

537   demonstrating a high correlation between DNN-derived PBLH and SONDE-derived

538   PBLH. Nonetheless, systematic biases in regions outside the SGP highlight the

539   influence of regional factors in PBLH estimation and suggest the need for region-

540   specific refinements to the model.

541   In summary, this research introduces a machine learning framework for PBLH

542   estimation that is able to generate high-quality PBLH using meteorological data,

543   independent of remote sensing instruments. This methodology, alongside the datasets

544   derived from the deep learning model, is beneficial in advancing our understanding of

545   PBL daytime development including thermodynamics and dynamics. It also has

546   implications for improved representation of the PBL processes in weather forecasting

547   and climate models, particularly by offering the potential to diagnose PBL in models

548   through the integration of modeled meteorological data as input. Future efforts will be

549   directed towards refining this model to ensure its wide applicability over a global scale.

550   These developments aim to effectively tackle the challenges of systematic biases and

551   regional variability in PBLH estimation.

552

553   **Data Availability.** ARM radiosonde data, surface fluxes, and cloud masks are available

554   at https://doi.org/10.5439/1333748 (ARM User Facility. 1994). The datasets of

555   planetary boundary layer height used in this study can be downloaded from

556   https://adc.arm.gov/discovery/#/results/instrument_class_code::pblht (last access: 7

557   January 2024; ARM User Facility, 2024). Climate Data Store offers the ERA-5

558   reanalysis data (https://doi.org/10.24381/cds.adbb2d47, Hersbach et al., 2023). The

559   DNN-derived PBLH datasets over the SGP, CACTI, and GoAmazon are available at

560   Zenodo (https://zenodo.org/records/10633811, Su. 2024) and will be uploaded to ARM

561   data archive as a product with detailed information upon acceptance.

562

563   **Author contributions.** TS conceptualized this study and carried out the analysis. TS

564   and YZ interpreted the data and wrote the manuscript. YZ supervised the project.

565

566   **Competing interests.** The contact author has declared that neither they nor their co-

567   authors have any competing interests.

568

569   **Acknowledgements:** We acknowledge the provision of radiosonde, lidar data, surface

570   meteorological data, and cloud products by the U.S. Department of Energy's ARM

571   program. Work at LLNL is performed under the auspices of the U.S. DOE by Lawrence

572   Livermore National Laboratory under Contract DE-AC52-07NA27344. This research

573   used resources of the National Energy Research Scientific Computing Center

574   (NERSC), a U.S. Department of Energy Office of Science User Facility located at

575   Lawrence Berkeley National Laboratory, operated under Contract No. DE-AC02-

576   05CH11231.

577

578   **Financial support.** This work is supported by the DOE Office of Science Atmospheric

579   System Research (ASR) program Science Focus Area (SFA) project, Tying in High

580   Resolution E3SM with ARM Data (THREAD).

581

582   **References**
583   Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G.S.,
584   Davis, A., Dean, J., Devin, M. and Ghemawat, S., 2016. Tensorflow: Large-scale
585   machine learning on heterogeneous distributed systems. arXiv preprint
586   arXiv:1603.04467.
587   Altmann, A., Toloşi, L., Sander, O. and Lengauer, T., 2010. Permutation importance:
588   a corrected feature importance measure. Bioinformatics, 26(10), pp.1340-1347.
589   ARM User Facility. (1994). ARM best estimate data products (ARMBEATM).
590   Southern Great Plains (SGP) central facility, Lamont, OK (C1). Compiled by C.
591   Xiao and X. Shaocheng [Dataset]. ARM Data Center.
592   https://doi.org/10.5439/1333748

593 ARM User Facility. (2024). PLANETARY BOUNDARY LAYER HEIGHT.
594   Southern Great Plains (SGP) central facility, Lamont, OK (C1). [Dataset]. ARM
595   Data Center. https://adc.arm.gov/discovery/#/results/instrument_class_code::pblht
596   (last access: 7 January 2024)

597 Atmospheric Radiation Measurement (ARM) user facility. 2015. Planetary Boundary
598   Layer Height (PBLHTSONDE1MCFARL). 2024-04-16 to 2024-04-19, ARM
599   Mobile Facility (ACX) Off the Coast of California - NOAA Ship Ronald H.
600   Brown; AMF2 (M1). Compiled by D. Zhang and D. Zhang. ARM Data Center.
601   Data set accessed 2024-03-05 at http://dx.doi.org/10.5439/1991783.

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

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

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

614 Bianco, L., and J. M. Wilczak, 2002: Convective boundary layer depth: Improved
615   measurements by Doppler radar wind profiler using fuzzy logic methods. J.
616   Atmos. Oceanic Technol., 19, 1745–1758, https://doi.org/10.1175/1520-
617   0426(2002)019,1745: CBLDIM.2.0.CO;2.

618 Bianco, L., Wilczak, J. M., & White, A. B. (2008). Convective boundary layer depth
619   estimation from wind profilers: Statistical comparison between an automated
620   algorithm and expert estimations. Journal of Atmospheric and Oceanic
621   Technology, 25(8), 1397-1413.

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

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

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

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

633 Clothiaux, E. E., Ackerman, T. P., Mace, G. G., Moran, K. P., Marchand, R. T.,
634   Miller, M. A., and Martner, B. E. (2000): Objective determination of cloud heights
635   and radar reflectivities using a combination of active remote sensors at the ARM
636   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

Deardorff, J. W. (1970). Convective velocity and temperature scales for the unstable planetary boundary layer and for Rayleigh convection. J. atmos. Sci, 27(8), 1211-1213.

Dong, X., Yu, Z., Cao, W., Shi, Y., & Ma, Q. (2020). A survey on ensemble learning. Frontiers of Computer Science, 14, 241-258.

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., Chen, D., Wang, J., Li, Z., Lv, Y., ... & Zhai, P. (2019). Declining summertime local-scale precipitation frequency over China and the United States, 1981–2012: The disparate roles of aerosols. Geophysical Research Letters, 46(22), 13281-13289.

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.

680 Guo, J., Zhang, J., Shao, J., Chen, T., Bai, K., Sun, Y., Li, N., Wu, J., Li, R., Li, J. and
681   Guo, Q., 2024. A merged continental planetary boundary layer height dataset
682   based on high-resolution radiosonde measurements, ERA5 reanalysis, and
683   GLDAS. Earth System Science Data, 16(1), pp.1-14.
684 Guo, J., Zhang, J., Yang, K., Liao, H., Zhang, S., Huang, K., Lv, Y., Shao, J., Yu, T.,
685   Tong, B. and Li, J., 2021. Investigation of near-global daytime boundary layer
686   height using high-resolution radiosondes: first results and comparison with ERA5,
687   MERRA-2, JRA-55, and NCEP-2 reanalyses. Atmospheric Chemistry and
688   Physics, 21(22), pp.17079-17097.
689 Helbig, M., Gerken, T., Beamesderfer, E.R., Baldocchi, D.D., Banerjee, T., Biraud,
690   S.C., Brown, W.O., Brunsell, N.A., Burakowski, E.A., Burns, S.P. and
691   Butterworth, B.J., 2021. Integrating continuous atmospheric boundary layer and
692   tower-based flux measurements to advance understanding of land-atmosphere
693   interactions. Agricultural and Forest Meteorology, 307, p.108509.
694 Hersbach, H., Bell, B., Berrisford, P., Biavati, G., Horányi, A., Muñoz Sabater, J.,
695   Nicolas, J., Peubey, C., Radu, R., Rozum, I., Schepers, D., Simmons, A., Soci, C.,
696   Dee, D., Thépaut, J-N. (2023): ERA5 hourly data on single levels from 1940 to
697   present. Copernicus Climate Change Service (C3S) Climate Data Store (CDS),
698   DOI: https://doi.org/10.24381/cds.adbb2d47
699 Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J.,
700   Nicolas, J., Peubey, C., Radu, R., Schepers, D. and Simmons, A., 2020. The
701   ERA5 global reanalysis. Quarterly Journal of the Royal Meteorological Society,
702   146(730), pp.1999-2049.
703 Holdridge, D., Ritsche, M., Prell, J., and Coulter, R. (2011): Balloon-borne sounding
704   system (SONDE) handbook, https://www.arm.gov/capabilities/instruments/sonde.
705 Holtslag, A.A. and Nieuwstadt, F.T. (1986). Scaling the atmospheric boundary layer.
706   Boundary-Layer Meteorology, 36(1-2), pp.201-209.
707 Holzworth, G. C. (1964). Estimates of mean maximum mixing depths in the
708   contiguous United States. Monthly Weather Review, 92(5), 235–242.
709   https://doi.org/10.1175/1520-0493(1964)092<0235:EOMMMD>2.3.CO;2
710 Kaimal, J.C. and Finnigan, J.J. (1994). Atmospheric boundary layer flows: their
711   structure and measurement. Oxford university press.
712 Kaimal, J.C., Wyngaard, J.C., Haugen, D.A., Coté, O.R., Izumi, Y., Caughey, S.J. and
713   Readings, C.J., 1976. Turbulence structure in the convective boundary layer.
714   Journal of Atmospheric Sciences, 33(11), pp.2152-2169.
715 Kollias, P., Bharadwaj, N., Clothiaux, E.E., Lamer, K., Oue, M., Hardin, J., Isom, B.,
716   Lindenmaier, I., Matthews, A., Luke, E.P. and Giangrande, S.E. (2020). The ARM
717   radar network: At the leading edge of cloud and precipitation observations.
718   Bulletin of the American Meteorological Society, 101(5), pp.E588-E607.
719 Kotthaus, S., Bravo-Aranda, J.A., Collaud Coen, M., Guerrero-Rascado, J.L., Costa,
720   M.J., Cimini, D., O'Connor, E.J., Hervo, M., Alados-Arboledas, L., Jiménez-
721   Portaz, M. and Mona, L., 2023. Atmospheric boundary layer height from ground-
722   based remote sensing: a review of capabilities and limitations. Atmospheric
723   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.

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, B., Ma, Y., Guo, J., Gong, W., Zhang, Y., Mao, F., ... & Shi, Y. (2019). Boundary layer heights as derived from ground-based Radar wind profiler in Beijing. IEEE Transactions on Geoscience and Remote Sensing, 57(10), 8095-8104.

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.

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

Martin, S. T., Artaxo, P., Machado, L. A. T., Manzi, A. O., Souza, R. A. F. D., Schumacher, C., ... & Wendisch, M. (2016). Introduction: observations and modeling of the Green Ocean Amazon (GoAmazon2014/5). Atmospheric Chemistry and Physics, 16(8), 4785-4797.

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-

767   time decision-making for high-impact weather, B. Am. Meteorol. Soc., 98, 2073–
768   2090, (2017).

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

772   Menut, L., Flamant, C., Pelon, J., & Flamant, P.H. (1999). Urban boundary-layer
773   height determination from lidar measurements over the Paris area. Applied Optics,
774   38, 945-954

775   Mohammed, A., & Kora, R. (2023). A comprehensive review on ensemble deep
776   learning: Opportunities and challenges. Journal of King Saud University-
777   Computer and Information Sciences, 35(2), 757-774.

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

781   Molod, A., Salmun, H., and Dempsey, M., 2015: Estimating Planetary Boundary
782   Layer Heights from NOAA Profiler Network Wind Profiler Data, J. Atmos.
783   Ocean. Tech., 32, 1545–1561, https://doi.org/10.1175/JTECH-D-14-00155.1.

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

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

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

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

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

798   Salmun, H., Josephs, H., & Molod, A. (2023). GRWP-PBLH: Global Radar Wind
799   Profiler Planetary Boundary Layer Height Data. Bulletin of the American
800   Meteorological Society, 104(5), E1044-E1057.

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

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

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

809   Sivaraman, Chitra, and Zhang, Damao. Planetary Boundary Layer Height derived
810   from Doppler Lidar (DL) data. United States: N. p., 2021. Web.
811   doi:10.5439/1726254.
812   Sleeman, J., Halem, M., Yang, Z., Caicedo, V., Demoz, B. and Delgado, R., 2020,
813   September. A deep machine learning approach for lidar based boundary layer
814   height detection. In IGARSS 2020-2020 IEEE international geoscience and
815   remote sensing symposium (pp. 3676-3679). IEEE.
816   Solanki, R., Guo, J., Lv, Y., Zhang, J., Wu, J., Tong, B., & Li, J. (2022). Elucidating
817   the atmospheric boundary layer turbulence by combining UHF radar wind profiler
818   and radiosonde measurements over urban area of Beijing. Urban Climate, 43,
819   101151.
820   Stull, R.B. (1988). An Introduction to Boundary Layer Meteorology. Dordrecht:
821   Springer Netherlands
822   Su, T. (2024). PBLH using DNN method [Dataset]. Zenodo.
823   https://zenodo.org/records/10633811
824   Su, T., Laszlo, I., Li, Z., Wei, J., & Kalluri, S. (2020a). Refining aerosol optical depth
825   retrievals over land by constructing the relationship of spectral surface
826   reflectances through deep learning: Application to Himawari-8. Remote Sensing
827   of Environment, 251, 112093.
828   Su, T., Li, Z. and Zheng, Y. (2023). Cloud-Surface Coupling Alters the Morning
829   Transition From Stable to Unstable Boundary Layer. Geophysical Research
830   Letters, 50(5), p.e2022GL102256.
831   Su, T., Li, Z., and Kahn, R. (2020b): A new method to retrieve the diurnal variability
832   of planetary boundary layer height from lidar under different thermodynamic
833   stability conditions. Remote Sens. Environ., 237, 111519.
834   Su, T., Li, Z., Roldán, N., Luan, Q., Yu, F.: Constraining Effects of Aerosol-Cloud
835   Interaction by Accounting for Coupling between Cloud and Land Surface, Science
836   Advances, 2024a (Accepted).
837   Su, T., Li, Z., Zhang, Y., Zheng, Y., & Zhang, H.: Observation and Reanalysis
838   Derived Relationships Between Cloud and Land Surface Fluxes Across Cumulus
839   and Stratiform Coupling Over the Southern Great Plains. Geophysical Research
840   Letters, 51(8). https://doi.org/10.1029/2023GL108090, 2024b.
841   Su, T., Zheng, Y. and Li, Z., (2022). Methodology to determine the coupling of
842   continental clouds with surface and boundary layer height under cloudy conditions
843   from lidar and meteorological data. Atmospheric Chemistry and Physics, 22(2),
844   pp.1453-1466.
845   Summa, D., Di Girolamo, P., Stelitano, D. and Cacciani, M., 2013. Characterization
846   of the planetary boundary layer height and structure by Raman lidar: comparison
847   of different approaches. Atmospheric Measurement Techniques, 6(12), pp.3515-
848   3525.
849   Sze, V., Chen, Y.H., Yang, T.J. and Emer, J.S., 2017. Efficient processing of deep
850   neural networks: A tutorial and survey. Proceedings of the IEEE, 105(12),
851   pp.2295-2329.

852   Tang, S., Xie, S., Zhang, M., Tang, Q., Zhang, Y., Klein, S. A., Cook, D. R., and
853   Sullivan, R. C. (2019): Differences in eddy-correlation and energy-balance surface
854   turbulent heat flux measurements and their impacts on the large-scale forcing
855   fields at the ARM SGP site. J. Geophy. Res. Atmos., 124, 3301–3318,
856   doi.org/10.1029/2018JD029689.
857   Tao, C., Y. Zhang, Q. Tang, H. Ma, V. P. Ghate, S. Tang, S. Xie, and J. A.
858   Santanello, 2021: Land–Atmosphere Coupling at the U.S. Southern Great Plains:
859   A Comparison on Local Convective Regimes between ARM Observations,
860   Reanalysis, and Climate Model Simulations. J. Hydrometeor., 22, 463–481,
861   https://doi.org/10.1175/JHM-D-20-0078.1.
862   Tucker, S.C., Brewer, W.A., Banta, R.M., Senff, C.J., Sandberg, S.P., Law, D.C.,
863   Weickmann, A.M., & Hardesty, R.M. (2009). Doppler Lidar Estimation of Mixing
864   Height Using Turbulence, Shear, and Aerosol Profiles. Journal of Atmospheric
865   and Oceanic Technology, 26, 673-688
866   Varble, A. C., Nesbitt, S. W., Salio, P., Hardin, J. C., Bharadwaj, N., Borque, P., ... &
867   Zipser, E. J. (2021). Utilizing a storm-generating hotspot to study convective
868   cloud transitions: The CACTI experiment. Bulletin of the American
869   Meteorological Society, 102(8), E1597-E1620.
870   Vassallo, D., Krishnamurthy, R., and Fernando, H. J. S.: Decreasing wind speed
871   extrapolation error via domain-specific feature extraction and selection, Wind
872   Energ. Sci., 5, 959–975, https://doi.org/10.5194/wes-5-959-2020, 2020.
873   Wang, J., Su, H., Wei, C., Zheng, G., Wang, J., Su, T., Li, C., Liu, C., Pleim, J.E., Li,
874   Z. and Ding, A., 2023. Black-carbon-induced regime transition of boundary layer
875   development strongly amplifies severe haze. One Earth, 6(6), pp.751-759.
876   Wang, Y., Zheng, X., Dong, X., Xi, B., Wu, P., Logan, T., & Yung, Y. L. (2020).
877   Impacts of long-range transport of aerosols on marine-boundary-layer clouds in
878   the eastern North Atlantic. Atmospheric Chemistry and Physics, 20(23), 14741-
879   14755.
880   Wesely, M. L., Cook, D. R., and Coulter, R. L. (1995): Surface heat flux data from
881   energy balance Bowen ratio systems (No. ANL/ER/CP-84065; CONF-9503104-
882   2). Argonne National Lab., IL (United States).
883   Xie, S., McCoy, R. B., Klein, S. A., Cederwall, R. T., Wiscombe, W. J., Jensen, M.
884   P., Johnson, K. L., Clothiaux, E. E., Gaustad, K. L., Long, C. N., and Mather, J.
885   H. (2010): Clouds and more: ARM climate modeling best estimate data: a new
886   data product for climate studies. Bull. Amer. Meteorol. Soc., 91(1), 13–20.
887   Xue, W., Dai, X., & Liu, L. (2020). Remote sensing scene classification based on
888   multi-structure deep features fusion. IEEE Access, 8, 28746-28755.
889   Ye, J., Liu, L., Wang, Q., Hu, S. and Li, S., 2021. A novel machine learning algorithm
890   for planetary boundary layer height estimation using AERI measurement data.
891   IEEE Geoscience and Remote Sensing Letters, 19, pp.1-5.
892   Zhang, D., Comstock, J. and Morris, V., (2022). Comparison of planetary boundary
893   layer height from ceilometer with ARM radiosonde data. Atmospheric
894   Measurement Techniques, 15(16), pp.4735-4749.

895   Zhang, Y., & Klein, S. A. (2010). Mechanisms affecting the transition from shallow
896   to deep convection over land: Inferences from observations of the diurnal cycle
897   collected at the ARM Southern Great Plains site. Journal of the Atmospheric
898   Sciences, 67(9), 2943–2959. https://doi.org/10.1175/2010jas3366.1
899   Zhang, Y., & Klein, S. A. (2013). Factors controlling the vertical extent of fair-
900   weather shallow cumulus clouds over land: Investigation of diurnal-cycle
901   observations collected at the ARM Southern Great Plains site. Journal of the
902   Atmospheric Sciences, 70(4), 1297–1315. https://doi.org/10.1175/jas-d-12-0131.1
903   Zhang, Z., (2018), June. Improved adam optimizer for deep neural networks. In 2018
904   IEEE/ACM 26th international symposium on quality of service (IWQoS) (pp. 1-
905   2). IEEE.

906 **Table list:**

907 **Table 1.** This table lists the varying structures of hidden layers used by each ensemble

908 member for PBLH estimation. Each configuration is expressed as an array, with the

909 number of elements indicating the number of layers and each value specifying the

910 number of neurons activated in the corresponding layer. For instance, a structure

911 denoted as [52, 28, 16] comprises three hidden layers containing 52, 28, and 16 neurons,

912 respectively.

913

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]

914

915

916

917

918

919

920

921

922

923

924

925

926

927 **Table 2.** Distribution of Dataset Samples for deep learning neural network (DNN)
928 Training and Validation. This table details the sample data in different local time used
929 for the development and validation of DNN to estimate planetary boundary layer height
930 (PBLH). The first column lists the available PBLH derived from radiosonde (SONDE,
931 Liu and Liang. 2010) during various local hours from 1994 to 2016. The second column
932 supplements the dataset with a combined MPL and SONDE approach (Su et al. 2020b)
933 and Doppler Lidar-derived PBLH (Sivaraman and Zhang, 2021) used in the absence of
934 SONDE measurements. Seventy percent of the combined dataset from the first and
935 second columns was randomly selected for the model's training. The third column
936 provides the number of SONDE measurements available for validation purposes. Since
937 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

Table 3. The relative 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 are expressed as percentages, indicating each feature's relative contribution to the model's predictive accuracy, normalized to sum to 100%.

Feature	Importance (%)	Feature	Importance (%)
Local Time	0.385238096	θ 0.45km	0.589744268
Month	3.589829217	θ 0.5km	0.537731259
RH (i-1)	1.525447612	θ 0.55km	0.534610382
RH (i)	16.25123402	θ 0.6km	0.552997086
U Wind (i-1)	0.385834048	θ 0.65km	0.431060615
U Wind (i)	2.076794013	θ 0.7km	0.342764903
V Wind (i-1)	2.537910928	θ 0.75km	0.310147803
V Wind (i)	2.405275378	θ 0.8km	0.380120894
Surface Pressure (i-1)	0.187890954	θ 0.85km	0.468503984
Surface Pressure (i)	1.016443163	θ 0.9km	0.413498983
Rain Rate (i-1)	0.077638613	θ 0.95km	0.263411835
Rain Rate (i)	0.10979265	θ 1km	0.132168034
Temperature (i-1)	1.028603672	θ 1.1km	0.163035362
Temperature (i)	1.382663171	θ 1.2km	0.042643843
LCL (i-1)	0.330188472	θ 1.3km	-0.020619871
LCL (i)	2.92117154	θ 1.4km	-0.117425464
Sensible Heat BLH (i-1)	1.071904572	θ 1.5km	-0.020003889
Sensible Heat BLH (i)	2.650567178	θ 1.6km	0.10811159
Parcel Method BLH (i-1)	8.796298485	θ 1.7km	0.211953821
Parcel Method BLH (i)	22.15513884	θ 1.8km	0.092761568
Sensible Heat (i-1)	1.09273529	θ 1.9km	0.134436502
Sensible Heat (i)	0.344360459	θ 2km	0.109195516
Latent Heat (i-1)	1.240177933	θ 2.2km	-0.10805866
Latent Heat (i)	1.705848738	θ 2.4km	-0.217483536
θ 0.05km	13.55861389	θ 2.6km	-0.178324068
θ 0.1km	1.19646809	θ 2.8km	0.08071272
θ 0.15km	0.025100917	θ 3km	0.249503653
θ 0.2km	0.193888217	θ 3.2km	0.143137953
θ 0.25km	0.445161715	θ 3.4km	0.19819078
θ 0.3km	0.572192811	θ 3.6km	0.158828504
θ 0.35km	0.751498918	θ 3.8km	0.185359544
θ 0.4km	0.768690105	θ 4km	1.046682377

946 **Figures**

947

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.

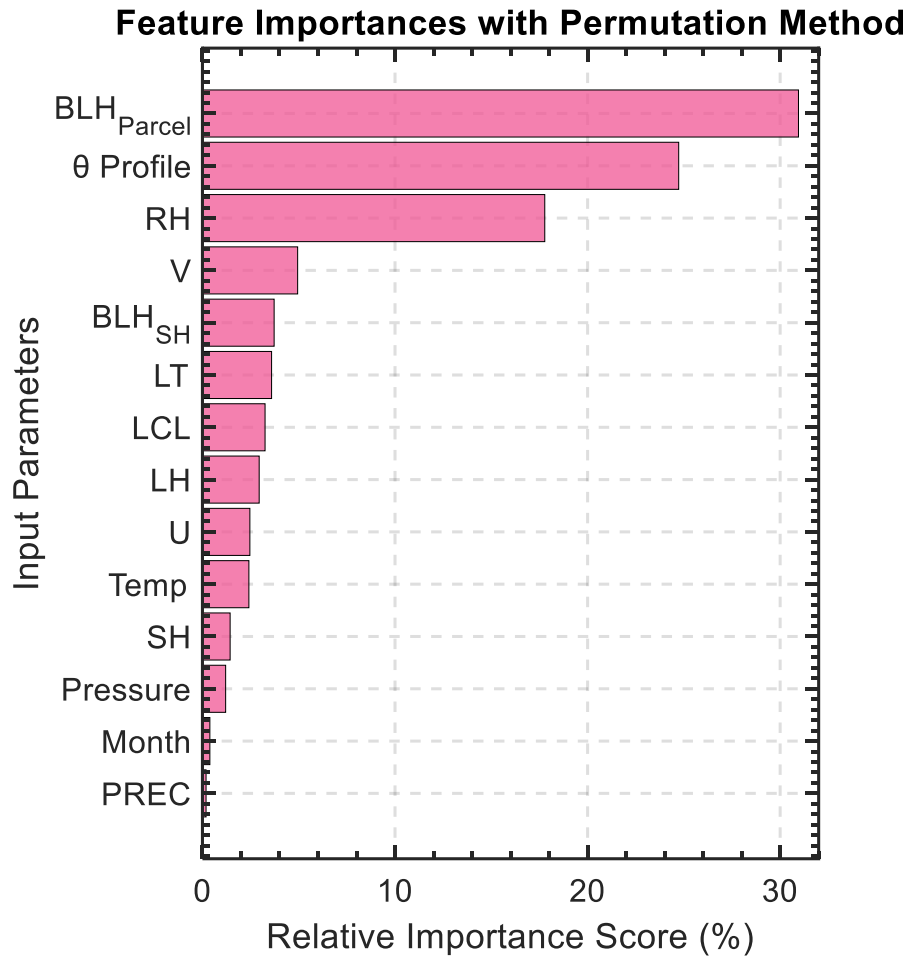



Figure 2. Feature importance with permutation method in the deep learning model. This table presents the importance scores of each input feature used in the deep learning model to estimate the PBLH. The features include local time (LT), month, relative humidity (RH), surface U and V wind components, pressure at the surface (Pressure), precipitation (PREC), surface temperature (Temp), 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 morning profiles of potential temperature (θ Profile). The importance scores are presented as percentages, representing each feature's relative contribution to the model's predictive accuracy, normalized to sum to 100%.

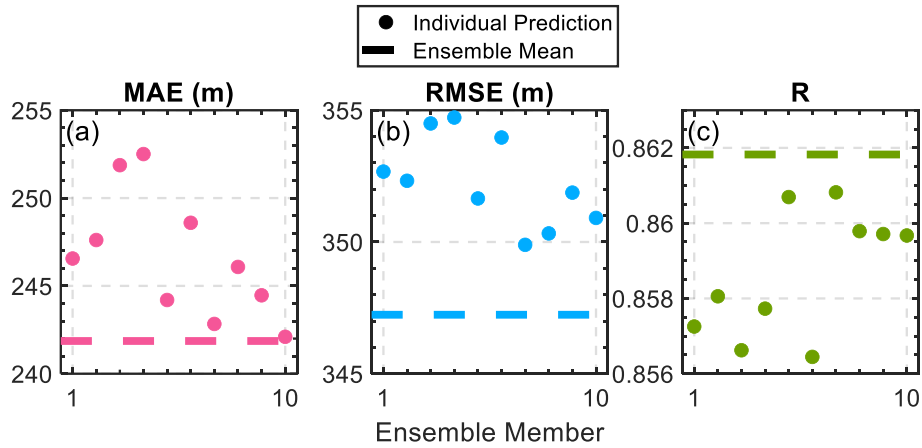



Figure 3: Performance metrics of individual ensemble members and the ensemble mean in estimating planetary boundary layer height (PBLH). Panel (a) displays the mean absolute error (MAE), panel (b) the root mean square error (RMSE), and panel (c) the correlation coefficient (R) for each of the ten ensemble 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.

979

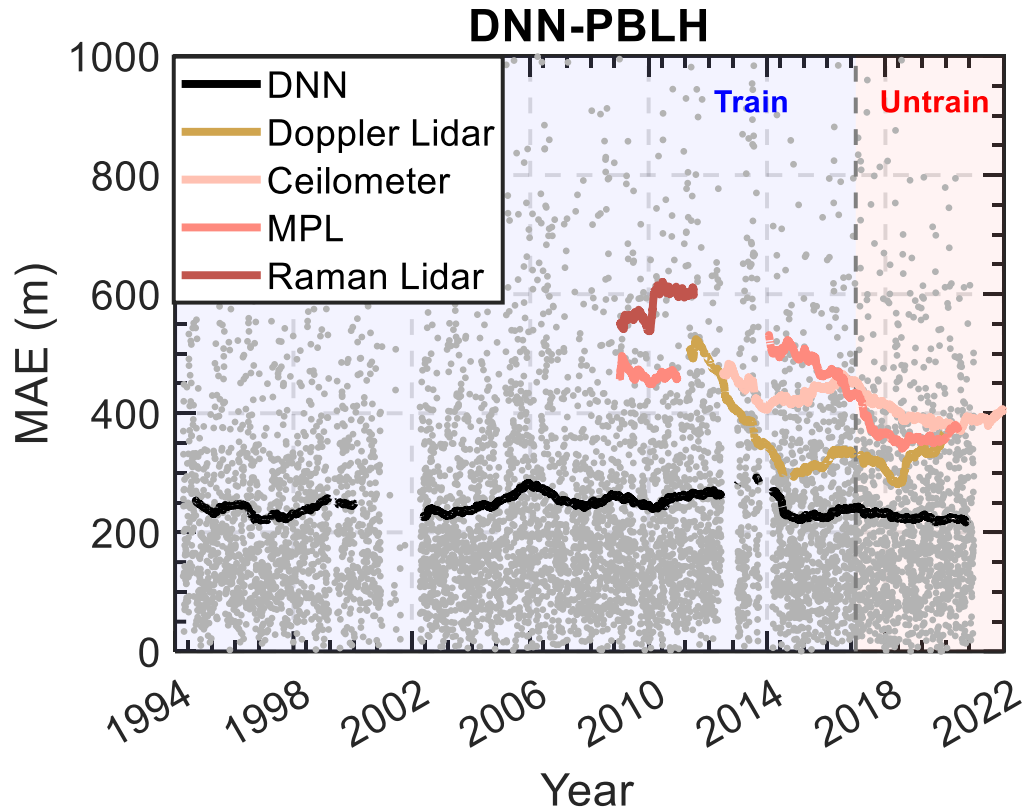



Figure 4: Comparative analysis of the mean absolute error (MAE) in PBLH estimation

using different methodologies. PBLH derived from SONDE is considered as the ground

truth. The DNN approach is shown in black, doppler lidar (Sivaraman and Zhang. 2021)

in yellow, ceilometer (Zhang et al. 2022) in pink, micro-pulse lidar (MPL, Sawyer and

Li. 2013) in light red, and Raman lidar (Ferrare. 2012) in dark red. DNN model is

trained during 1994-2016. Individual MAE values for DNN are represented by gray

dots, while the solid lines denote the smoothed MAE for each method with a 2-year

smooth window.

988

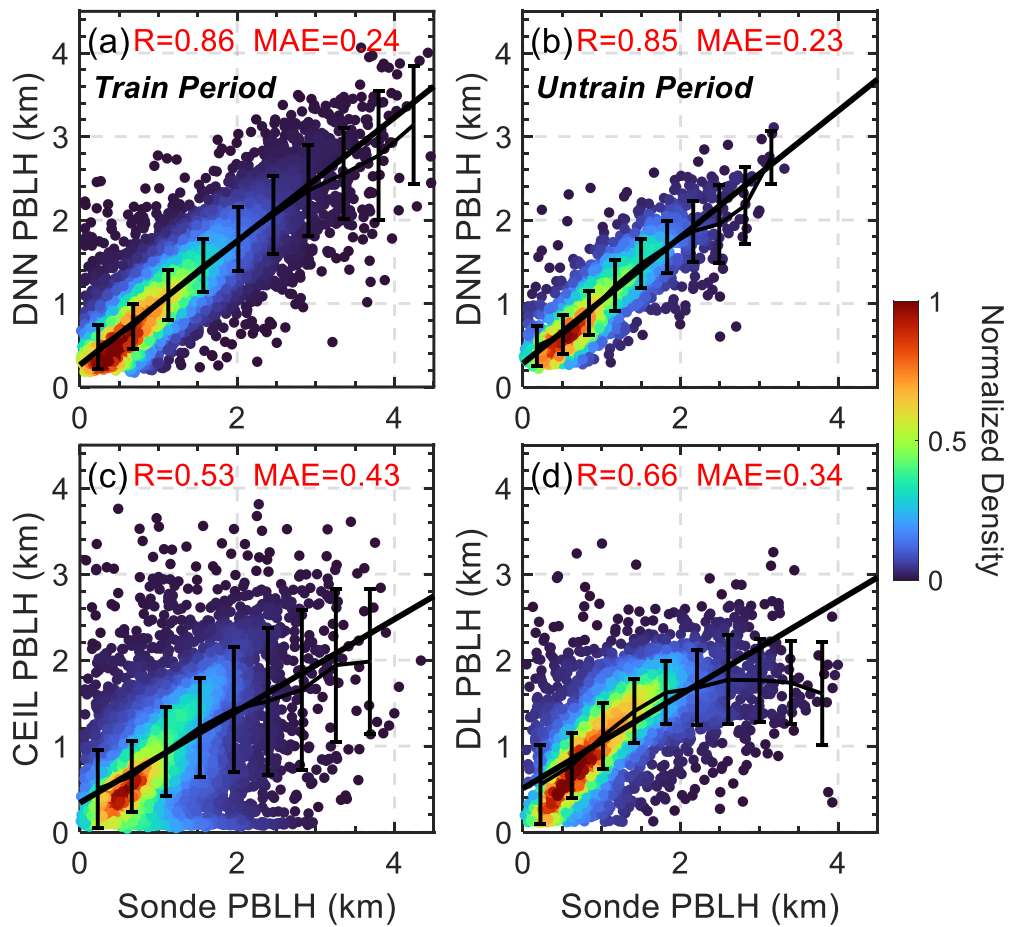



Figure 5: Scatter plots comparing observed radiosonde (SONDE) PBLH with estimates

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.

998

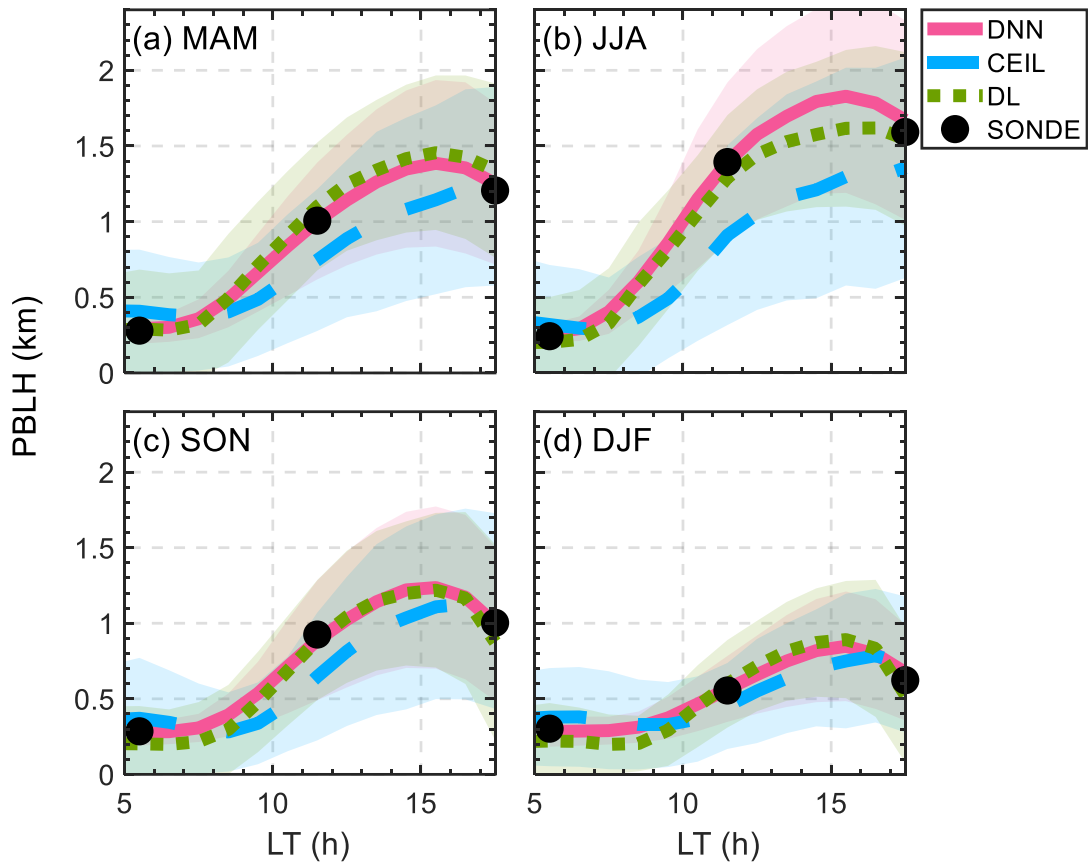



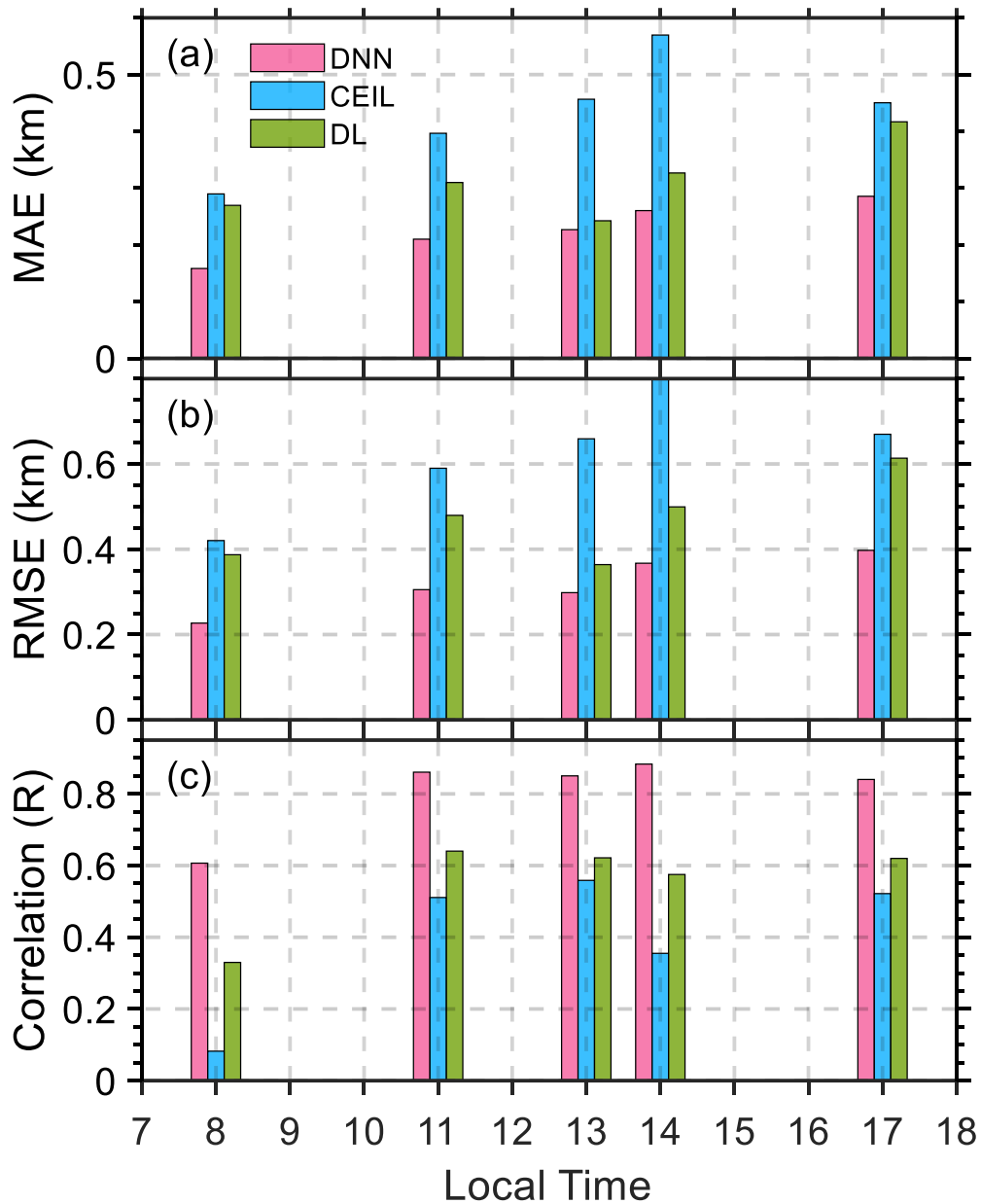
Figure 6: Seasonal-averaged daytime evolution of planetary boundary layer height (PBLH) derived from various methods. The panels represent the mean PBLH values throughout the day for different seasons: (a) March-April-May (MAM), (b) June-July-August (JJA), (c) September-October-November (SON), and (d) 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.

1007

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

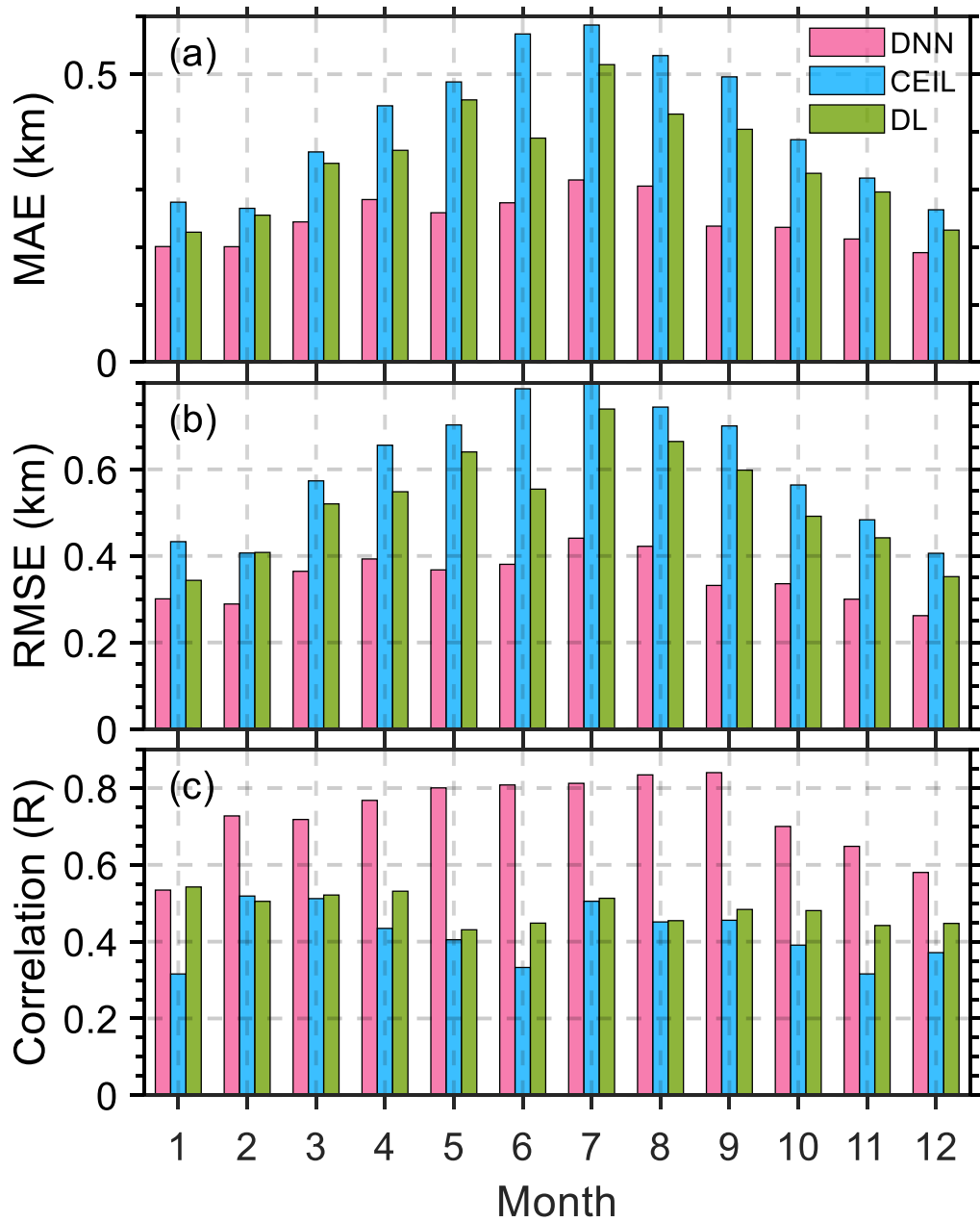
standard. The availability of SONDE data for different hours is detailed in Table 2.

1015

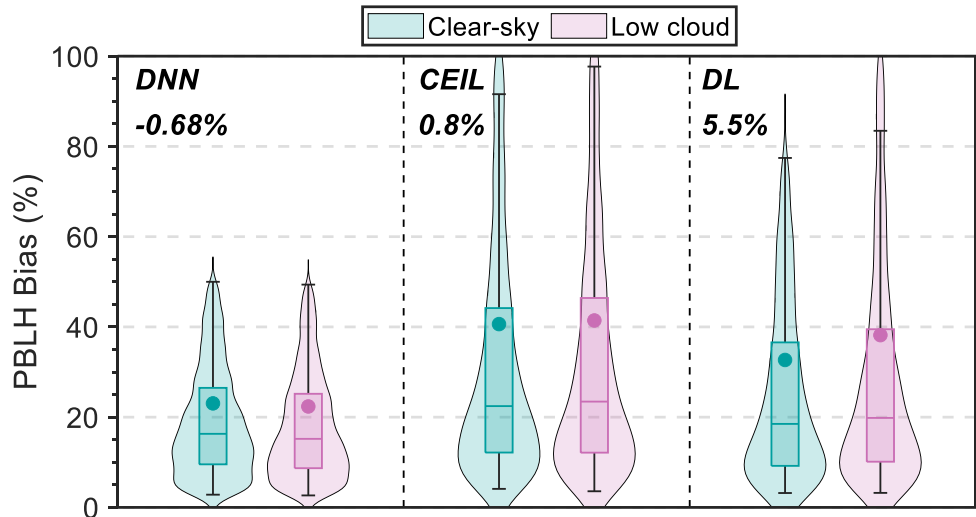
Figure 8: Similar to Figure 7, but for MAE, RMSE, and R for different month.

1017

Figure 9: Comparative analysis of PBLH estimation bias under clear-sky and low cloud

1018

conditions for various methods. Bias percentages are computed as the absolute bias

1019

normalized by the mean PBLH for each condition, with the number above each method

1020

indicating the difference in bias between low cloud and clear-sky scenarios. The

1021

boxplots detail the 10th, 25th, 50th, 75th, and 90th percentiles, while shaded areas in

1022

violin plots illustrate the distribution of dataset biases. The dots indicate the mean value

1023

for each condition.

1024

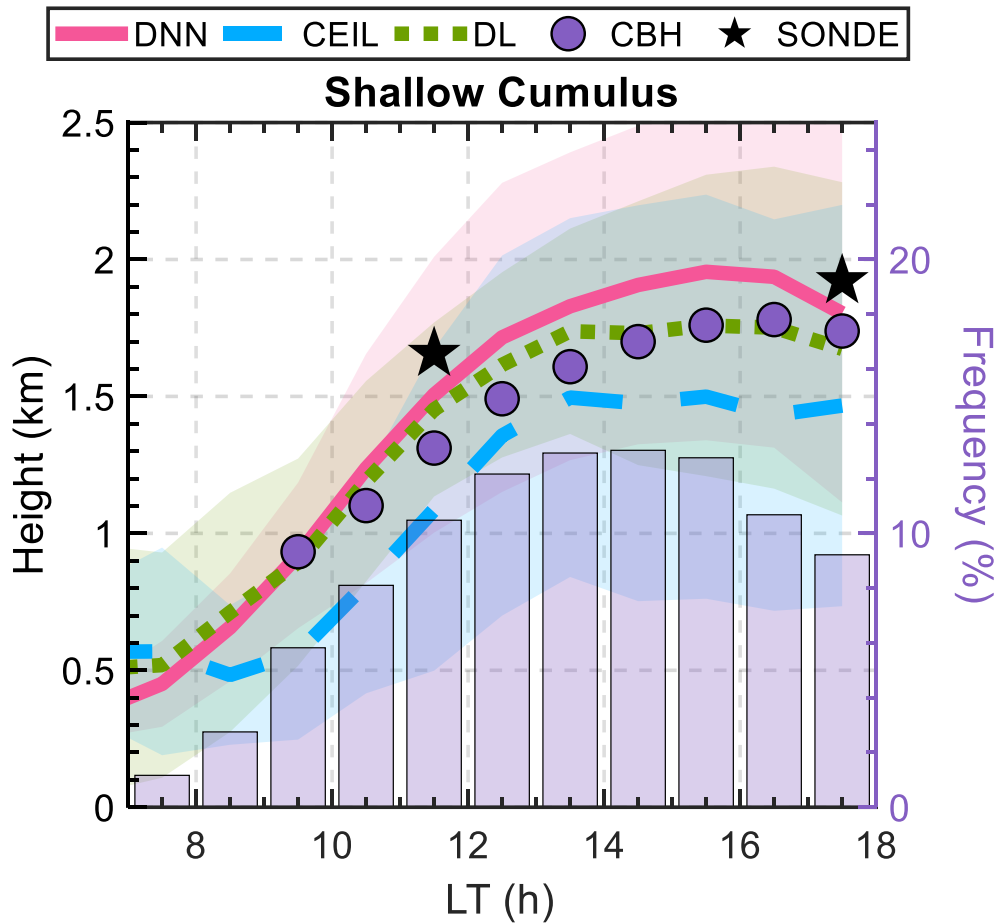
1025

1026

Figure 10: Daytime evolution of planetary boundary layer height (PBLH) derived from

various methods under the shallow cumulus condition. PBLH values estimated by the

deep neural network (DNN) are shown in red, ceilometer (CEIL) estimates in blue,

Doppler lidar (DL) in green. Observed radiosonde (SONDE) data are represented by

black stars. Purple bars show the relative frequency of shallow cumulus occurrences

throughout the day, while purple dots mark the corresponding cloud-base heights

(CBH). Shaded areas around each line reflect the standard deviations for each method.

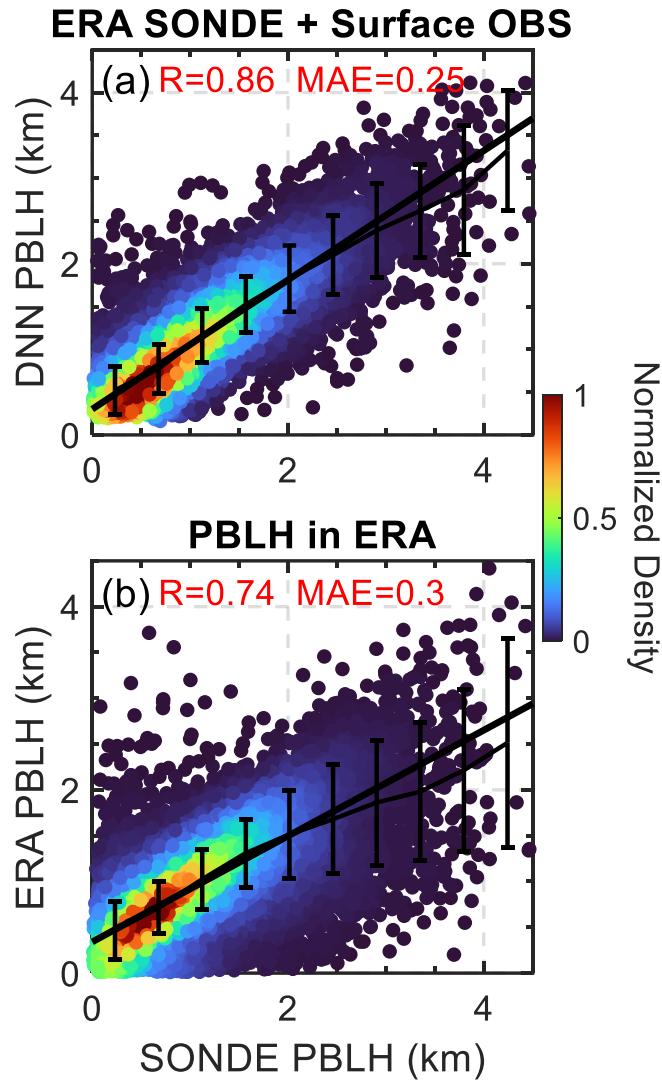
1034

1035

Figure 11: 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.

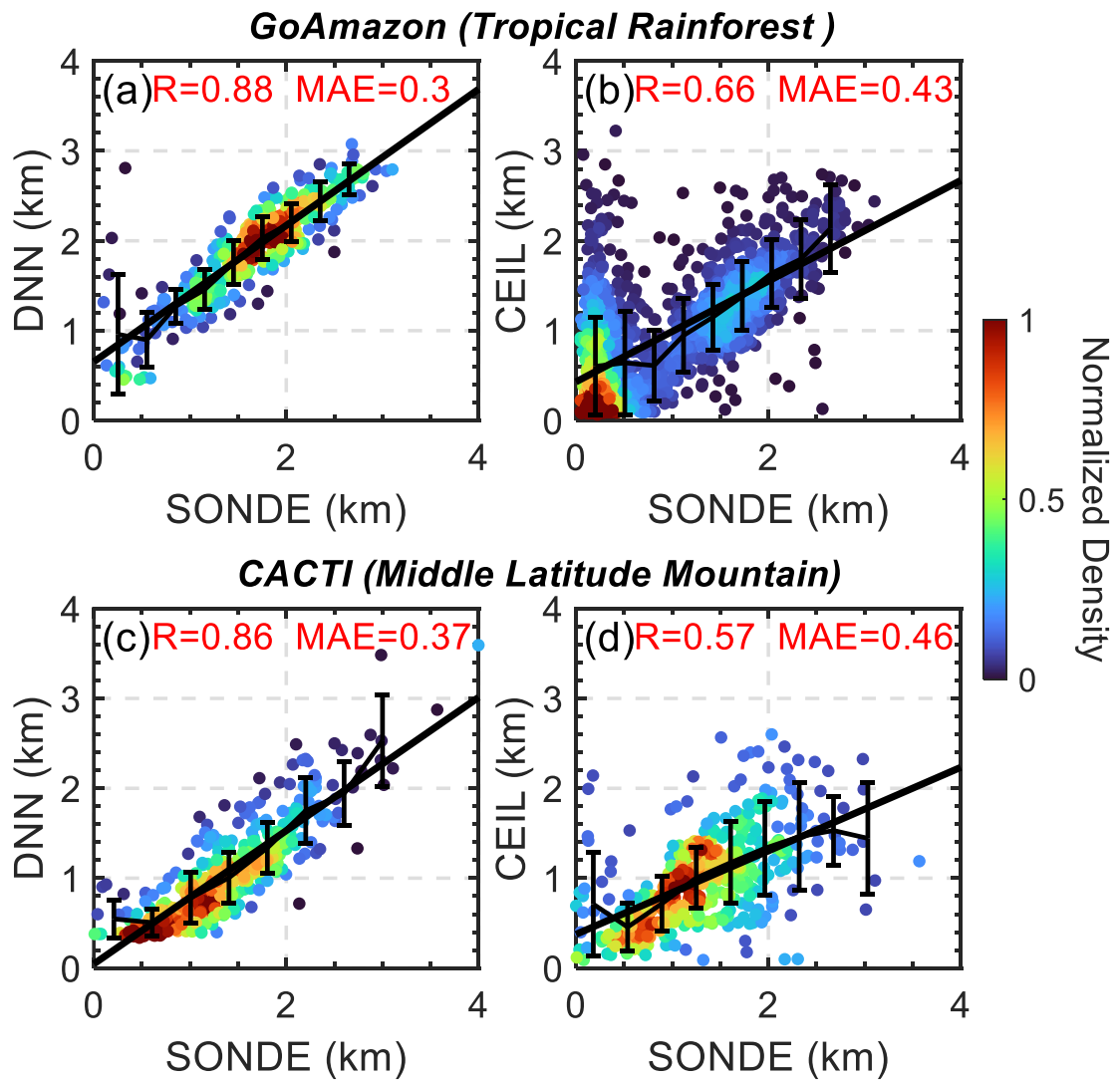



Figure 12: Validation of the DNN trained over the SGP for the GoAmazon (Tropical

Rainforest) and CACTI (Middle Latitude Mountain) field campaigns. Panels (a) and (c)

illustrate the correlation (R) and mean absolute error (MAE) between DNN predictions

and SONDE observations for GoAmazon and CACTI, respectively. Panels (b) and (d)

show the performance of ceilometer (CEIL) derived PBLH compared to SONDE for

the same campaigns. 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

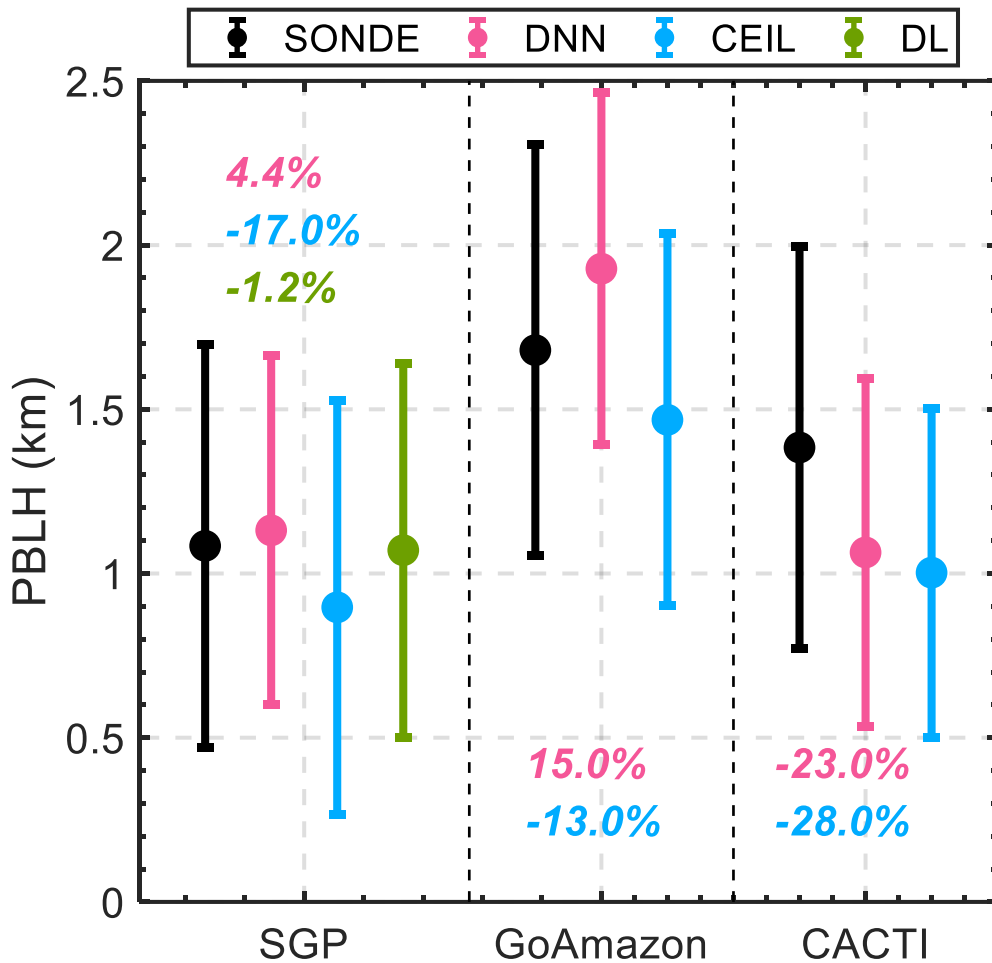
and standard deviations for each bin.

1053

Figure 13: 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.