



Machine Learning Model for Inverting Convective Boundary Layer Height with Implicit Physical Constraints and Its Multi-Site Applicability

Yufei Chu¹, Guo Lin^{2,3}, Min Deng⁴, Lulin Xue⁵, Weiwei Li⁵, Hyeyum Hailey Shin⁵, Jun A. Zhang^{2,3},
5 Hanqing Guo⁶, and Zhien Wang¹

¹School of Marine and Atmospheric Sciences, Stony Brook University, Stony Brook, 11790, USA

²NOAA /AOML/Hurricane Research Division, Miami, 33149, USA.

³Cooperative Institute for Marine and Atmospheric Studies, University of Miami, Miami, 33149, USA

⁴Environmental Science and Technologies Department, Brookhaven National Laboratory, Upton,
10 11793, USA

⁵National Science Foundation National Center for Atmospheric Research, Boulder, 80307, USA

⁶Department of Electrical and Computer Engineering, University of Hawaii at Manoa, Honolulu, 96822,
USA

15 Correspondence to: Zhien Wang (Zhien.wang@stonybrook.edu)

Abstract. Accurate estimation of convective boundary layer height (CBLH) is vital for weather, climate, and air quality modeling. Machine learning (ML) shows promise in CBLH prediction, but input parameter selection often lacks physical grounding, limiting generalizability. This study introduces a novel ML framework for CBLH inversion, integrating thermodynamic constraints and the diurnal CBLH cycle as an implicit physical guide. Boundary layer growth is modeled as driven by surface heat fluxes and atmospheric heat absorption, using the diurnal cycle as input and output. TPOT and AutoKeras are employed to select optimal models, validated against Doppler lidar-derived CBLH data, achieving an R^2 of 0.84 across untrained years. Comparisons of eddy covariance (ECOR) and energy balance Bowen ratio (EBBR) flux measurements show consistent predictions (R^2 difference ~ 0.011 , MAE ~ 0.002 km). Models trained on ARM SGP C1 site ECOR data and tested at E37 and E39 yield R^2 values of 0.787 and 0.806, respectively, demonstrating adaptability. Training with all sites' data enhances C1 ECOR and EBBR performance over C1-only training: ECOR (R^2 : 0.851 vs. 0.845; MAE: 0.198 km vs. 0.207 km), EBBR (R^2 : 0.837 vs. 0.834; MAE: 0.203 km vs. 0.205 km). The interquartile range (IQR) error bars for predicted CBLH are consistently narrower than those for DL-derived CBLH, reflecting lower variability in predicted CBLH compared to DL-derived CBLH, which is influenced by additional factors. Transferability across ARM Southern Great Plains sites and seasonal performance during summer (JJA) confirm the model's robustness, offering a scalable approach for improving boundary layer parameterization in atmospheric models.



1 Introduction

35 The convective boundary layer (CBL) is a critical component of the Earth's atmosphere, governing the exchange of heat, moisture, and momentum between the surface and the free troposphere (Stull, 1988; Garratt, 1994). Accurate estimation of the convective boundary layer height (CBLH) is essential for understanding atmospheric processes, including turbulence, pollutant dispersion, and cloud formation (Stull, 1988; Seibert et al., 2000). In numerical weather prediction (NWP) and climate models, CBLH
40 serves as a key parameter for parameterizing turbulent mixing and convective processes, directly impacting forecast accuracy and climate projections (Grenier, 2001; Holtslag et al., 2013; Baklanov et al., 2014). Errors in CBLH estimation can lead to significant biases in surface temperature, humidity, and air quality predictions (Vogelezang and Holtslag, 1996; Hu et al., 2010). Consequently, improving CBLH
45 predictions has been a priority in atmospheric science, with numerous studies emphasizing its role in model performance and data assimilation (Helmis, 2012; Cohen et al., 2015; Wulfmeyer et al., 2016; Brown et al., 2008; Barlow et al., 2020; Chu et al., 2022a).

50 While current observational techniques have greatly contributed to the determination of the CBLH, each method still presents inherent limitations related to resolution, sensitivity, or applicability under different atmospheric conditions. Radiosondes provide direct measurements of temperature and humidity profiles but suffer from low temporal resolution, typically limited to twice to fourth-daily launches (Seidel et al., 2010; Liu and Liang, 2010; Lin, 2024). Meteorological towers measure near-surface variables but are constrained by their height, rarely capturing the full CBL (Bianco et al., 2011; Emeis et al., 2009). Weather radars offer vertical profiles but lack the resolution to resolve fine-scale CBL structures (Heo et
55 al., 2003; Compton et al., 2013). Aerosol lidars, while effective for detecting entrainment zones, are often confounded by residual layers, leading to ambiguous CBLH estimates (Hennemuth and Lammert, 2006; Sawyer and Li, 2013; Schween et al., 2014; Luo, 2014). Doppler lidars provide high-resolution velocity and backscatter data, enabling precise CBLH retrievals, but their algorithms vary widely (Tucker et al., 2009; Barlow et al., 2011; Chu et al., 2020). Each method employs different inversion algorithms—such
60 as gradient-based, variance-based, or wavelet techniques—each with inherent uncertainties depending on atmospheric conditions and data quality (Cohn and Angevine, 2000; Hägeli et al., 2000; Lammert and Bösenberg, 2006; Compton et al., 2013; Chu et al., 2022b).

65 Recent advances in machine learning (ML) have revolutionized CBLH prediction by leveraging large datasets to model complex atmospheric relationships. Early ML approaches used simple regression models to estimate CBLH from radiosonde data (Krishnamurthy et al., 2020; Madonna et al., 2021). Subsequent studies adopted random forests and neural networks, incorporating inputs from aerosol lidars, Doppler lidars, and reanalysis datasets (Liu et al., 2022; Krishnamurthy et al., 2021; Peng et al., 2023; Wei et al., 2025; Zhang et al., 2025). For instance, random forest models have been applied to lidar-derived backscatter profiles (Du et al., 2023; Chu et al., 2025a), while deep neural networks have integrated reanalysis data for regional CBLH predictions (Ayazpour et al., 2023; Su et al., 2024). Despite these advances, most ML models select input parameters empirically, lacking physical constraints, which limits their generalizability across diverse sites (de Arruda Moreira et al., 2022; Su et al., 2024; Chu et al., 2025a; Macatangay et al., 2025; Stapleton et al., 2025). Few studies have explored physically



75 constrained ML frameworks or evaluated model performance across multiple stations, highlighting a critical gap in the literature (Krishnamurthy et al., 2021; Su et al., 2024; Wei et al., 2025; Stapleton et al., 2025). The above discussion and existing literature clearly demonstrate that numerous ML algorithms are currently available. Comparing each algorithm to identify the optimal one is highly time-consuming.

80 Building on these insights, this study introduces an Auto-ML framework that automatically selects the optimal ML algorithm for CBLH inversion, utilizing Doppler lidar-derived CBLH data and thermodynamically constrained input parameters. We incorporate physical principles by assuming CBL growth is driven by thermodynamic equilibrium, with surface heat flux and atmospheric heat absorption as primary drivers. This approach ensures robust predictions across varying atmospheric conditions and

85 sites. To assess the model's transferability, we evaluate its performance at four sub-sites (C1, E32, E37, and E39) within the Atmospheric Radiation Measurement (ARM) Southern Great Plains (SGP) supersite. These locations were selected due to their comprehensive observations of surface heat flux (SHF), latent heat flux (LHF), and lower tropospheric stability (LTS). These sites provide a diverse testbed for validating the model's generalizability and its potential to enhance CBLH predictions in atmospheric

90 models.

This paper is organized as follows: Section 2 describes the data sources and ML methodology, including the implicit physical constraints. Section 3 presents the model results, encompassing performance metrics, site-to-site comparisons, and contrasts across different seasons and ML approaches. Section 4 discusses

95 the findings, their implications for atmospheric model, and future research directions.

2 Data and methods

2.1 Site description

This study utilizes data from the ARM SGP facility, a premier research site established by the U.S. Department of Energy to investigate land-atmosphere interactions in a continental mid-latitude environment (Mather et al., 2016). Located in Oklahoma, USA, the SGP spans a diverse agricultural landscape, making it ideal for studying CBL dynamics under varying meteorological conditions (Mather and Voyles, 2013). We focus on four SGP sites: the central facility (C1) and three extended facilities (E32, E37, E39), selected for their comprehensive measurements of surface fluxes and atmospheric profiles. The latitude and longitude coordinates of the four sites are shown in Table 1 (Wulfmeyer et al., 2018).

100 The C1 site, located near Lamont, Oklahoma, serves as the primary hub, hosting a suite of instruments including radiosondes, a Doppler lidar, and an Atmospheric Emitted Radiance Interferometer (AERI). The extended sites—E37, and E39—are equipped with Eddy Correlation (ECOR) systems for surface flux measurements. Additionally, the nearby E14 site (Lamont, Oklahoma; 36.605°N, 97.485°W, 315 m elevation), co-located with C1, which we attribute to C1 for consistency. The E32, E39, and E13 (near

105 C1) sites employ Energy Balance Bowen Ratio (EBBR) technology to measure heat flux.

110



Table 1. Instruments and datasets used in this study.

Sites	Latitude (°, N)	Longitude (°, W)	Altitude (m)	AERI data stream	DL data stream	ECOR data stream	EBBR data stream	Other
C1	36.6073	97.4876	314	sgpaerioel1tu rnC1.c1.	sgpdllfptc1 .b1	sgp30qcecorE14.s1.	sgp30baebbrE13.c1.	
E32	36.8193	97.8198	328	sgpaerioel1tu rnE32.c1.	Sgpdlfpte3 2.b1	/	sgp30baebbrE32.c1.	E14's ECOR or E13's EBBR used as C1
E37	36.3104	97.9274	379	sgpaerioel1tu rnE37.c1.	Sgpdlfpte3 7.b1	sgp30qcecorE37.s1.	/	
E39	36.3735	97.0691	279	sgpaerioel1tu rnE39.c1.	Sgpdlfpte3 9.b1	sgp30qcecorE39.s1.	sgp30baebbrE39.c1.	

115

The distances between the ARM SGP sites are as follows: C1 to E32 is approximately 40 km, C1 to E37 is approximately 77km, C1 to E39 is approximately 41 km, E32 to E37 is approximately 57km, E32 to E39 is approximately 67 km, and E37 to E39 is approximately 77 km. These distances ensure a range of spatial variability in surface and atmospheric conditions, enabling robust evaluation of the model's transferability across sites (Turner et al., 2016). The C1 site's Doppler lidar (DL) provides high-resolution vertical velocity and backscatter data, while radiosondes offer 4-th-daily temperature and humidity profiles. The AERI at four sites measures downwelling infrared radiance to derive atmospheric stability metrics.

120

2.2 Data and preprocessing

125

The dataset comprises multiple variables critical for CBLH estimation, sourced from the ARM SGP sites over the period 2016–2019. The DL used are Halo Photonics Stream Line models (1.5 μ m wavelength), with the C1 site featuring an upgraded Stream Line XR+ model for enhanced signal-to-noise ratio (SNR). These lidars provide a vertical resolution of 30 m and a temporal resolution of 1-3s, ensuring detailed vertical velocity profiles (Manninen et al., 2019). The CBLH is calculated using Chu et al. (2022a)'s algorithm on ARM DL data, utilizing wavelet analysis to account for turbulence eddy size and gravity wave effects, and applying dynamic thresholds to estimate CBLH from 2-D vertical velocity variance. LTS is derived from AERI observations at C1, calculated as the potential temperature difference between 700 hPa and 1000 hPa ($LTS = \theta_{700} - \theta_{1000}$), validated against radiosonde data (Feltz et al., 2003; Wood et al., 2006). Surface fluxes, including SHF and LHF, are obtained from ECOR systems at C1 (via E14), E37, and E39, and from the EBBR system at C1 (via E13), E32, and E39. The ECOR systems use eddy covariance techniques to measure turbulent fluxes, while the EBBR system estimates fluxes via the Bowen ratio method, incorporating net radiation, soil heat flux, and temperature-humidity gradients (Cook, 2005; Cook, 2011).

130

135

140

Previous studies have shown significant flux discrepancies between ECOR and EBBR beams obtained through different detection techniques, making them non-interchangeable for direct use (Tang et al., 2019; Chu et al., 2022b). Data preprocessing involves quality control to remove outliers and missing values, following ARM's standard protocols (e.g., flagging data with unrealistic values or low SNR).



2.3 Machine learning methods

145 Machine learning (ML) algorithms have emerged as powerful tools in atmospheric science, enabling the analysis of complex, non-linear relationships within large datasets to improve predictions of phenomena such as CBLH (Schultz et al., 2021). ML methods excel at identifying patterns in atmospheric data, enhancing applications like weather forecasting, air quality modeling, and boundary layer parameterization by integrating diverse data sources, including ground-based observations and reanalysis 150 products (Reichstein et al., 2019). Two prominent ML approaches for regression tasks are decision tree-based methods and neural networks, each offering distinct advantages for atmospheric applications (Bauer et al., 2015; de Burgh-Day et al., 2023).

155 Decision tree-based methods partition data into hierarchical decision nodes, creating a flowchart-like structure to predict outcomes based on input features. Advanced ensemble techniques, such as random forests and gradient boosting, combine multiple trees to improve accuracy and robustness, making them well-suited for tasks like CBLH estimation (Breiman, 2001; Chen and Guestrin, 2016). Neural networks, conversely, consist of interconnected layers of nodes that learn intricate patterns through backpropagation, 160 excelling in capturing non-linear dynamics in atmospheric datasets, such as turbulence or stability gradients (Goodfellow et al., 2016). Automated ML frameworks streamline model development by optimizing architectures and hyperparameters. Prior studies have compared frameworks like AutoKeras (Zhong et al., 2024), which automates neural network design, and the Tree-based Pipeline Optimization Tool (TPOT), which focuses on tree-based models, finding comparable performance in atmospheric 165 applications (Jin et al., 2019; Olson et al., 2016). Considering computational efficiency and the adaptability of algorithms to diverse datasets, this study does not simultaneously compare the results of various machine learning methods. Instead, it focuses on comparing the outcomes of TPOT and AutoKeras after their automated selection of optimal models. Specifically, we adopt the TPOT Regressor from the TPOT library, which automates the construction and optimization of tree-based pipelines, minimizing manual tuning efforts while preserving high predictive accuracy (Olson et al., 2016). 170 Similarly, we employ AutoKeras, a neural architecture search framework that leverages Bayesian optimization and network morphism to automatically design and fine-tune deep learning models, enabling efficient and adaptive model selection for complex datasets (Jin et al., 2019; Liang et al., 2024). This research employs the automated machine learning frameworks TPOT (version 0.12.2) and AutoKeras (version 1.0.20), integrated within a Python 3.9 environment. Development was executed using the 175 spyder-kernels package (version 2.4.4), ensuring robust and reproducible computational workflows. To facilitate reader understanding, the table 2 below provides an overview of commonly supported algorithms in TPOT and AutoKeras, along with their descriptions, advantages, applicable scenarios, disadvantages, and limitations, supported by relevant references.

180



Table 2. Commonly supported algorithms in TPOT and AutoKeras.

Framework	Algorithm	Description	Advantages	Limitations	Reference
TPOT	Decision Tree Regressor	A tree-based model that splits data into branches based on feature thresholds to predict outcomes.	Simple, interpretable; suitable for small-to-medium datasets with clear feature relationships.	Prone to overfitting; struggles with high-dimensional or noisy data.	Breiman et al. (1984)
	Random Forest Regressor	An ensemble of decision trees that aggregates predictions to reduce variance and improve accuracy.	Robust to overfitting, handles non-linear relationships well; ideal for medium-sized datasets.	Computationally intensive; less interpretable due to ensemble nature.	Breiman (2001)
	Gradient Boosting Regressor (XGBoost)	Boosts weak decision trees sequentially, optimizing a loss function using gradient descent.	High predictive accuracy, handles missing data well; effective for structured/tabular data.	Requires careful hyperparameter tuning; can be slow to train on large datasets.	Chen and Guestrin (2016)
	Support Vector Regressor (SVR)	Finds a hyperplane that best fits the data within a margin, using kernel tricks for non-linearity.	Effective for small datasets with non-linear patterns; robust to outliers.	Scales poorly with large datasets; sensitive to kernel choice and parameter tuning.	Drucker et al. (1997)
	Linear Regression	Models the relationship between features and target using a linear equation.	Simple, interpretable, fast to train; suitable for datasets with linear relationships.	Assumes linearity and independence of features; performs poorly with non-linear or noisy data.	Seber and Lee (2003)
AutoKeras	Convolutional Neural Network (CNN)	A deep learning model with convolutional layers to extract spatial features from data.	Excels in image or spatial data processing; automatically learns hierarchical features.	Requires large datasets and computational resources; less interpretable.	LeCun et al. (1998)
	Recurrent Neural Network (RNN)	A deep learning model designed for sequential data, with loops to retain memory of past inputs.	Suitable for time-series or sequential data; captures temporal dependencies.	Prone to vanishing gradients; computationally expensive for long sequences.	Hochreiter and Schmidhuber (1997)
	Transformer	A model using self-attention mechanisms to process sequential data, often for time-series tasks.	Efficient for long sequences, captures global dependencies; ideal for complex time-series data.	High computational cost; requires large datasets for effective training.	Vaswani et al. (2017)
	Multilayer Perceptron (MLP)	A fully connected neural network with multiple layers to model non-linear relationships.	Versatile for tabular data; can approximate complex functions with sufficient depth.	Prone to overfitting on small datasets; requires careful tuning of architecture and parameters.	Hornik et al. (1989)
	Residual Network (ResNet)	A deep CNN with residual connections to mitigate vanishing gradients in deep architectures.	Enables training of very deep networks; effective for complex pattern recognition tasks.	High computational cost; may overfit on small datasets without proper regularization.	He et al. (2016)

2.4 Evaluation metrics

To assess the performance of the ML model, we use a suite of standard regression metrics, providing a comprehensive evaluation of predictive accuracy and error characteristics. Let y_i represent the observed CBLH, \hat{y}_i the predicted CBLH, and n the number of observations. The following metrics are used:

Coefficient of Determination (R^2): Measures the proportion of variance in the observed CBLH explained by the model, calculated as:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}, \quad (1)$$



195 where $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$ is the mean of the observed values (Krause et al., 2005; Legates and McCabe, 1999).
The coefficient of determination (R^2) quantifies the model's explanatory power by measuring the proportion of variance in the observed CBLH that is accounted for by the model predictions. As shown in Equation (1), R^2 compares the sum of squared residuals (SSR) against the total variance in the observations (SST). When R^2 approaches 1, it indicates that the model explains nearly all variability in
200 the observed CBLH ($SSR \approx 0$), representing a near-perfect fit where predicted values (\hat{y}_i) closely match observations (y_i). Conversely, an R^2 near 0 suggests the model performs no better than simply using the mean (\bar{y}) as a predictor. While useful for goodness-of-fit assessment, we need to notice that high R^2 values alone do not guarantee model validity, as they may be artificially inflated by overfitting or insensitive to systematic biases in the predictions.

205

Mean Absolute Error (MAE): Quantifies the average absolute difference between predicted and observed CBLH, given by:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| , \quad (2)$$

210 MAE provides a straightforward measure of prediction error in the same units as CBLH (Willmott and Matsuura, 2005). MAE measures the average absolute difference between predicted and true values. It is used to monitor the model's performance during training and validation but does not directly influence the optimization of the model parameters (i.e., it does not affect the minimization of the loss function). MAE provides an intuitive metric to reflect the average magnitude of prediction errors.

215 **Mean Squared Error (MSE):** Computes the average squared difference between predictions and observations, defined as:

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 , \quad (3)$$

220 The quadratic nature of Mean Squared Error (MSE) inherently penalizes larger prediction errors more severely (Hyndman & Koehler, 2006). AutoKeras optimizes model parameters by directly minimizing this loss function to enhance predictive accuracy. In contrast, TPOT utilizes Negative MSE ($-MSE$), a transformed version where the sign inversion converts the minimization problem into a maximization framework. This approach maintains consistency with scikit-learn's convention where higher scores denote superior models, allowing TPOT's automated search to effectively identify configurations that maximize $-MSE$ (thereby minimizing the actual MSE).

225

230 This study primarily focuses on comparing the CBLH across different sites, utilizing the R^2 and MAE as robust metrics to evaluate the performance of machine learning models. These metrics provide a comprehensive assessment of model accuracy and predictive reliability, with R^2 quantifying the proportion of variance in CBLH that is explained by the model, and MAE offering a direct measure of the average prediction error in physical units. These metrics collectively ensure a thorough evaluation of the model's accuracy, bias, and robustness across the ARM SGP sites.



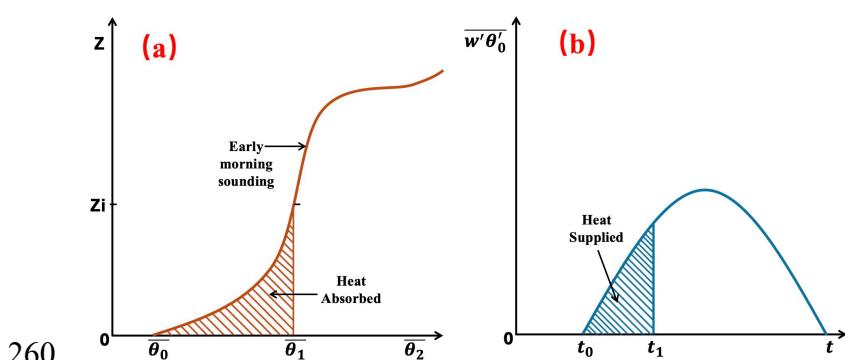
2.5 Implicit physical constraints

2.5.1 Parameter selection based on thermodynamic equilibrium constraints

235 Traditional machine learning methods for predicting CBLH typically employ Principal Component Analysis (PCA) or random combinatorial approaches to select input parameters (Liu et al., 2022). While these methods can achieve good predictive performance at specific sites, they lack a physical basis, resulting in poor transferability across different sites and limiting their applicability in physical space. To address this issue, this study proposes an innovative approach by incorporating the physical foundation
240 of thermodynamic equilibrium to optimize parameter selection, thereby developing a CBLH prediction model that is transferable across sites. As noted by Stull (1988) in his classic textbook, although the development of the convective boundary layer is influenced by multiple factors, thermodynamic equilibrium is the primary driver of its evolution. Figure 1a illustrates the heat required for CBL growth, while Fig.1b depicts the surface heat flux emitted following solar radiation absorption, which together
245 determine the CBLH. Additionally, this study builds on the theoretical framework of thermodynamic equilibrium to quantify the dynamic evolution of the CBLH. Specifically, the relationship is described by the following integral expression:

$$\int_{t=0}^{t_1} \overline{w' \theta'_s(t)} dt = \int_{\theta=\theta_0}^{\theta_1} Z(\theta) d\theta , \quad (4)$$

250 where the left-hand side represents the cumulative contribution of surface heat flux over time, and the right-hand side describes the amount of heat required to produce the observed temperature profile variation with height within the boundary layer. This heat conservation relationship jointly determines the dynamic evolution of CBLH. This formula provides a solid physical foundation for the model, ensuring that parameter selection not only enhances predictive capability but also maintains physical consistency across different sites (e.g., C1, E32, E37, E39), thereby overcoming the transferability
255 limitations of traditional machine learning approaches. By integrating the thermodynamics information of the sites, the model can effectively capture the variations in thermodynamic processes across different sites, thus improving both prediction accuracy and generalizability.



260 Figure 1. Graphical approach to estimate Convective Boundary Layer Height (CBLH) thermodynamically by equating heat supplied with heat absorbed; Zi is CBLH. (Stull, 1988)



Based on the thermodynamic constraints outlined, traditional physical constraint methods can incorporate the thermodynamic constraints from Equation 4 into the loss function of machine learning models. This
265 approach adds to the MSE the difference between the cumulative energy provided by the heat flux and the energy required to account for the atmospheric temperature difference, multiplied by an adjustment coefficient to optimize the model. However, the selection of this coefficient significantly impacts the results, posing challenges to the method's applicability. From the perspective of boundary layer physics, Equation 4 provides a robust physical constraint during the development from sunrise to the top of the
270 CBL. Nevertheless, after reaching the boundary layer top, the entrainment process in the entrainment zone must also be considered, where the application of Equation 4 may impose inappropriate constraints. The dissipation process, on the other hand, involves distinct physical mechanisms. Additionally, factors such as moisture, wind speed, wind direction, and cloud cover introduce complex, nonlinear effects, making the direct application of physical constraints highly challenging.

275 This study proposes an innovative approach that combines heat flux with LTS to achieve implicit physical constraints. Initially, the cumulative heat flux is decomposed into SHF and LHF, which reflect moisture characteristics. Furthermore, instantaneous SHF and LHF can account for the influences of clouds and wind. Simultaneously, LTS is calculated as an instantaneous value for each hour, dynamically capturing
280 temperature difference variations. To address diurnal and seasonal variations in CBLH, the model incorporates sunrise and sunset times along with their corresponding timestamps, defining a normalized temporal parameter,

$$\text{SUNPERCENT} = (\text{TIME} - \text{SUNRISE}) / (\text{SUNSET} - \text{SUNRISE}) , \quad (5)$$

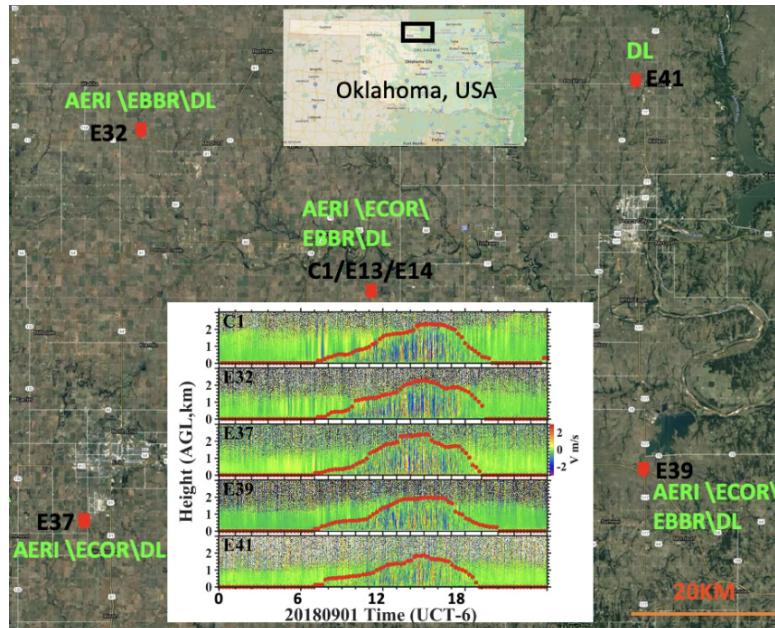
285 which represents the proportion of the current time relative to the daylight duration. In summary, this study employs physically driven variables for parameter selection—specifically surface heat flux and LTS as core inputs—to ensure the model's physical consistency and transferability. The heat flux is further broken down into physical components, including the cumulative sensible heat flux (C_SHF) and
290 latent heat flux (C_LHF) since sunrise, as well as the instantaneous sensible heat flux (I_SHF) and latent heat flux (I_LHF) within a one-hour window, while LTS is taken as an hourly instantaneous value. This parameterization effectively captures diurnal variations in solar radiation, enriching the model with more comprehensive physical information.

295 2.5.2 Integrated Diurnal Evolution of the CBL

The current literature on predicting CBLH using machine learning predominantly focuses on discrete, moment-to-moment predictions, often overlooking the integrated diurnal evolution of the CBL as a unified process. For instance, Chu et al. (2025a) employed ML to estimate CBLH over the Southern Great Plains, but this approach centered on individual time steps, neglecting the full diurnal cycle. However, as
300 shown in Fig. 2, the diurnal variation of CBLH across five ARM sites reveals distinct site-specific patterns. The CBLH at each moment evolves continuously from the preceding moment, establishing a dynamic and interconnected developmental trajectory. Treating these moments in isolation disrupts this continuity,



305 failing to capture the underlying evolutionary dynamics. Specifically, the peak CBLH values vary across the sites, and the morning growth and evening decay phases exhibit notable differences, highlighting the critical role of temporal dependencies in boundary layer evolution. While some studies incorporate the CBLH of the previous moment—or CBLH derived from alternative methods, such as sensible heat flux or parcel methods—as an input variable, this approach often overemphasizes the influence of prior CBLH values, thereby overshadowing the contributions of other meteorological drivers. For example, Su et al. (2020) demonstrated that machine learning models relying heavily on CBLH derived from sensible heat 310 flux and parcel methods tend to exhibit excessive dependence on temporal autocorrelation, which diminishes the model's sensitivity to key meteorological factors such as heat flux and atmospheric stability. Consequently, these methods are limited in their ability to comprehensively predict the diurnal variation of CBLH, constraining the scope of their investigations.



315

Figure 2. Relative Locations of ARM SGP Sites C1, E32, E37, E39, and E41. Top inset: Geographic position of SGP sites in Oklahoma; Bottom inset: Diurnal variation of CBLH observed by Doppler lidar across the five sites on 01 September 2018.

To address these shortcomings, this study adopts the CBLH across the entire diurnal cycle as the training 320 target for the machine learning model, treating the boundary layer evolution as a continuous and interconnected process. This holistic approach enables the model to comprehensively capture the dynamic evolution of the boundary layer, from the gradual rise of the CBLH after sunrise, through its midday peak accompanied by oscillations, to the rapid decay observed after sunset. By integrating the complete developmental trajectory, the model not only better represents the interconnected dynamics of the CBL 325 but also accounts for the complex interplay of meteorological drivers that govern its evolution. For instance, the morning growth phase is heavily influenced by surface heating and turbulent mixing, while the midday peak often reflects a balance between entrainment processes at the boundary layer top and surface-driven convection. The evening decay phase, on the other hand, is modulated by radiative cooling



and the cessation of surface heat fluxes, which vary significantly across different sites due to local land
330 surface characteristics and atmospheric conditions. To enhance the model's predictive capability, we incorporate time-dependent variables that reflect the diurnal cycle's progression. This approach mitigates the overreliance on prior CBLH values by ensuring that the model learns the underlying physical relationships between CBLH and its meteorological drivers, rather than simply exploiting temporal autocorrelation. As a result, the model is expected to improve the accuracy of CBLH predictions across
335 the diurnal cycle, offering a more comprehensive understanding of boundary layer dynamics.

2.6 Auto-ML model for convective boundary layer height

We prepared the relevant input parameters and employed the following methodology to enable the machine learning approach to uncover the complex physical mechanisms underlying the physical
340 parameters. After the compilation environment was set up and the data was prepared, the specific model application process (Fig. 3) is as follows:

1) Data Collection and Pre-processing. Collect and pre-process the necessary input data for CBLH prediction. Based on the content of Section 2.1 and 2.2, we prepare the data for each timestamp of the
345 day, including C_SHF, C_LHF, I_SHF, I_LHF, LTS, TIME, SUNRISE, SUNSET, SUNPERCENT, and CBLH. The SUNRISE and SUNSET represent the sunrise and sunset times calculated based on the latitude and longitude coordinates of the site. To simplify the dataset, we aggregate the data from 06:00 to 21:00 (UTC-6), covering a 15-hour period, as a single daily dataset. The CBLH for the entire day is designated as the target variable for output, while the other parameters serve as input variables. We
350 randomly split all the data into 70% for training and 30% for testing by date. The subfigure in Fig. 3 that depicts the ARM site, included in the Data Collection section, is adapted from Wulfmeyer et al. (2018).

2) Use AutoML to find the Best Train Model. Using the training dataset, we employ TPOT and AutoKeras to derive their respective optimal algorithms or hyperparameters. By comparing the R^2 and
355 Mean MAE metrics, we select the model that performs best in both MAE and R^2 as the optimal model, which is then designated as the candidate best model for further evaluation and application.

3) Use the Best Model with training Data for Training. The best model is trained on the training dataset and saved for later use, ensuring optimal performance for future applications.
360

4) One-Day CBLH Prediction. Use the trained model to predict CBLH for a single day.

5) Check if All Days are Processed. If not all days in the test dataset have been processed, repeat step 4 for the next day.
365

6) Save Data. When all days' data have been processed, save the predicted CBLH data for further analysis.

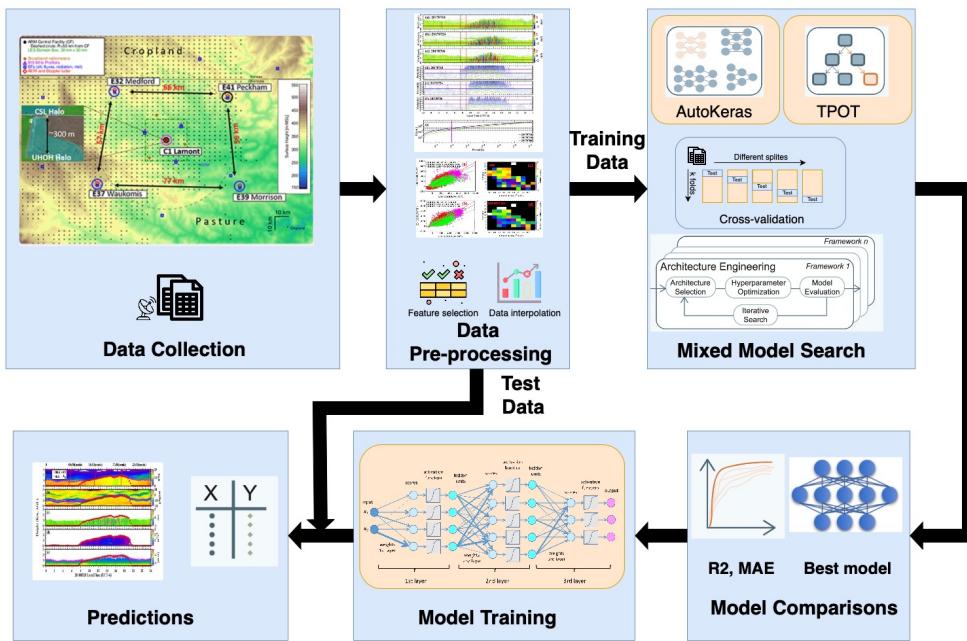


Figure 3. The Auto-ML workflow of Convective Boundary Layer Height (CBLH)

370 Figure 3's flowchart outlines the algorithm proposed in this study (termed the Auto-ML algorithm). Its
371 core principles are: (1) utilizing thermodynamical variables as input parameters with implicit physical
372 constraints; (2) incorporating the complete CBL development cycle as unified input, with corresponding
373 CBLH as output; and (3) employing TPOT and AutoKeras models to automatically select the optimal
374 machine learning algorithm. This approach enables the model to capture the entire CBL development
375 process, enhancing prediction accuracy and representation of CBL dynamics.

3 Results

To validate the effectiveness of the Auto-ML framework, we first conducted tests using data from the C1 site spanning 2016 to 2019, presenting the results for ECOR and EBBR heat flux, respectively. Subsequently, the algorithm was applied to model data transfer from one site to other sites. Next, we 380 compared the performance of the optimal TPOT and AutoKeras algorithms for summer (JJA) and further evaluated the advantages and limitations of different methods for computing SHAP (Shapley Additive exPlanations) values. Furthermore, we analyzed the variations in Auto-ML's relative importance across seasons. Finally, we compared the performance of models trained on multi-site data and tested on site-specific data.



385 3.1 Application of the Auto-ML to the ARM SGP C1 Site

3.1.1 Results from ECOR flux

The Auto-ML algorithm demonstrates robust performance in predicting CBLH when using ECOR flux dataset. The scatter predicted CBLH in Fig.4a demonstrates a strong linear correlation ($R^2 = 0.845$) between predicted and observed CBLH across the annual dataset, suggesting that the Auto ML model effectively captures the general trends of CBL evolution (Fig.4a). However, the MAE of 0.207 km highlights a non-negligible systematic bias, potentially linked to specific meteorological conditions or seasonal variations not fully resolved by the model. Notably, the density of points clusters around the 1:1 line in the lower CBLH range (0–1.5 km), while deviations increase slightly at higher CBLH values (>2 km), possibly indicating reduced model sensitivity to extreme events (e.g., intense convective days).

395

The diurnal variability between predicted and observed CBLH also shows good agreement (Fig.4b). The predicted CBLH closely tracks observed values during the morning development phase (07:30–13:30 UTC-6), with overlapping interquartile ranges (IQR error bars), reflecting reliable performance during periods of rapid boundary layer growth driven by surface heating and turbulent mixing. However, a 400 significant divergence emerges in the afternoon (15:30–17:30 UTC-6), where the predicted mean CBLH underestimates observations by ~100 m. This discrepancy coincides with the typical peak phase of the CBL, characterized by weakening turbulence, entrainment processes at the CBL top, and increasing influence of subsidence or advection.

405

The afternoon underestimation may stem from the algorithm's limited ability to resolve complex interactions during the CBL peak phase. During midday, solar radiation maximizes surface heat flux, driving vigorous turbulent eddies that homogenize the CBL, making CBLH prediction relatively straightforward. By late afternoon, surface heating diminishes, turbulence decays, and the entrainment zone at the CBL top becomes dynamically significant. Entrainment of free-tropospheric air into the CBL 410 can temporarily elevate the observed CBLH, a process that may not be easily captured in the Auto ML model. Additionally, the advection of air masses with different thermodynamic properties (e.g., moisture or temperature gradients) could introduce spatial heterogeneity, further challenging the algorithm's generalizability during transitional periods.

415

Moreover, the model's training data might underrepresent late-afternoon scenarios, where PBL dynamics are influenced by mesoscale phenomena (e.g., cloud cover, or topographic effects). For instance, enhanced subsidence or cloud shading at 15:30–17:30 (UTC-6) could suppress turbulent mixing, leading to a shallower predicted CBLH compared to observations.

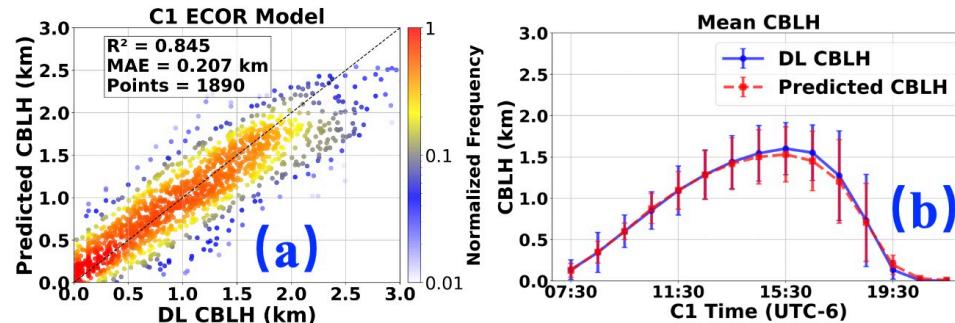


Figure 4. Results of Auto ML model with ECOR dataset for predicting CBLH: (a) Comparison of all test data, (b) Diurnal variation average with IQR (interquartile range) error bars.

3.1.2 Results from EBBR flux

The Auto-ML algorithm exhibits robust performance in predicting CBLH using the EBBR flux dataset. As shown in Fig. 5a, the scatter plot of predicted CBLH demonstrates a strong linear correlation ($R^2 = 0.834$) between predicted and observed CBLH across the annual dataset, indicating that the Auto-ML model effectively captures the overall trends of CBL evolution. However, the MAE of 0.205 km, comparable to that obtained with the ECOR dataset, suggests that both flux datasets yield similar predictive accuracy. The diurnal variability between predicted and observed CBLH also shows good agreement (Fig. 5b). During the morning development phase (07:30–11:30 UTC-6), the predicted CBLH closely aligns with observed values. However, a non-notable divergence emerges in the afternoon (around 15:30 UTC-6), where the predicted CBLH is lower than the observed values, with the deviation reducing from 100 m (as seen with ECOR) to approximately 50 m.

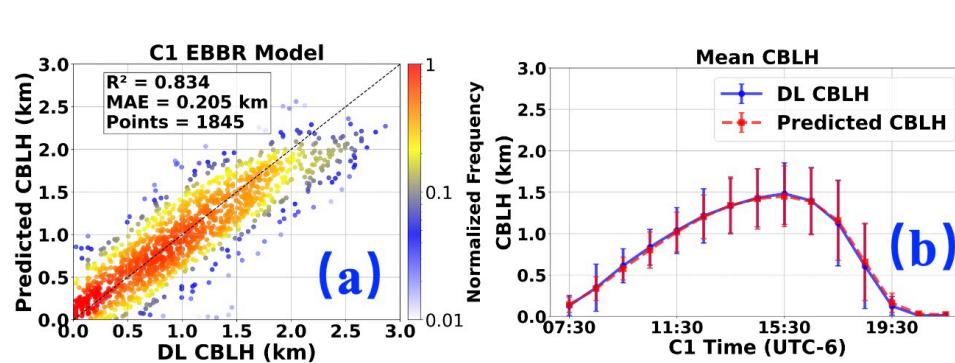


Figure 5. Results of Auto ML model with EBBR dataset for predicting CBLH: (a) Comparison of all test data, (b) Diurnal variation average with IQR (interquartile range) error bars.

By comparing Fig. 4 and Fig. 5, it is evident that, with the same Auto-ML model, the CBLH predictions using ECOR flux data slightly outperform those using EBBR flux data, though the difference is minor (R^2 : 0.845 vs 0.834; MAE: 0.207 km vs 0.205 km). This suggests that heat flux measurements from different instruments exhibit discrepancies, but these discrepancies follow a systematic, inherent pattern.



445 The Auto-ML model effectively mitigates these inherent differences, achieving comparable CBLH predictions, thereby demonstrating its robustness and adaptability.

450 A comparison of Fig.4 and Fig.5 reveals that the discrepancies between the predicted values and the DL observations are primarily evident at the PBL top (~15:30 UTC-6) and during the dissipation phase (~19:30 UTC-6). Specifically, at the CBL top (~15:30 UTC-6), the predicted values are generally lower than the DL observations, whereas during the dissipation phase (~19:30 UTC-6), the predicted values tend to exceed the observed values.

3.2 Effectiveness of the Auto-ML Across Multiple Sites

455 The Auto-ML algorithm demonstrates notable adaptability beyond the C1 site, highlighting its potential for broader application across multiple observation sites. To evaluate this, we tested an Auto-ML model trained on ECOR heat flux data at the C1 site (E14 site) for its performance at the E37 and E39 sites. Similarly, an Auto-ML model trained on EBBR heat flux data at the C1 site (E13 site) was assessed for its performance at the E32 and E39 sites.

3.2.1 Apply C1 site ECOR model to E37 and E39 Sites

460 The Auto-ML model trained at the C1 site with ECOR data exhibits strong performance at the E37 and E39 sites, achieving R^2 values of 0.787 and 0.806, and MAE values of 0.219 km and 0.208 km, respectively (Fig. 6a and 6c). However, as observed in Fig. 6b and 6d, the model's performance varies across different time periods at these sites. Specifically, Fig.6b shows that at the E37 site, the model predictions align well with observations during the CBLH dissipation phase (15:30–21:30 UTC-6).
465 However, during the initial development phase (07:30–14:30 UTC-6), a significant discrepancy is observed, with predicted values consistently higher than the DL observations. Notably, no similar discrepancy is evident in Figure 4b, suggesting that the heat flux differences between the C1 and E37 sites are more pronounced during the initial phase, while the differences diminish after reaching the boundary layer top.

470 Similarly, Figure 6d indicates that at the E39 site, the model predictions closely match observations during the CBLH dissipation phase (15:30–21:30 UTC-6). However, discrepancies are observed during the initial phase (09:30–11:30 UTC-6) and the top phase (14:30–16:30 UTC-6). During the initial phase, the predicted values are higher than the DL observations, while at the top phase, the predicted values are lower. In comparison, Figure 4b shows a similar discrepancy at the top phase but not during the initial phase. This suggests that the heat flux differences between the C1 and E39 sites are more significant during the initial phase, with the differences decreasing after reaching the boundary layer top.

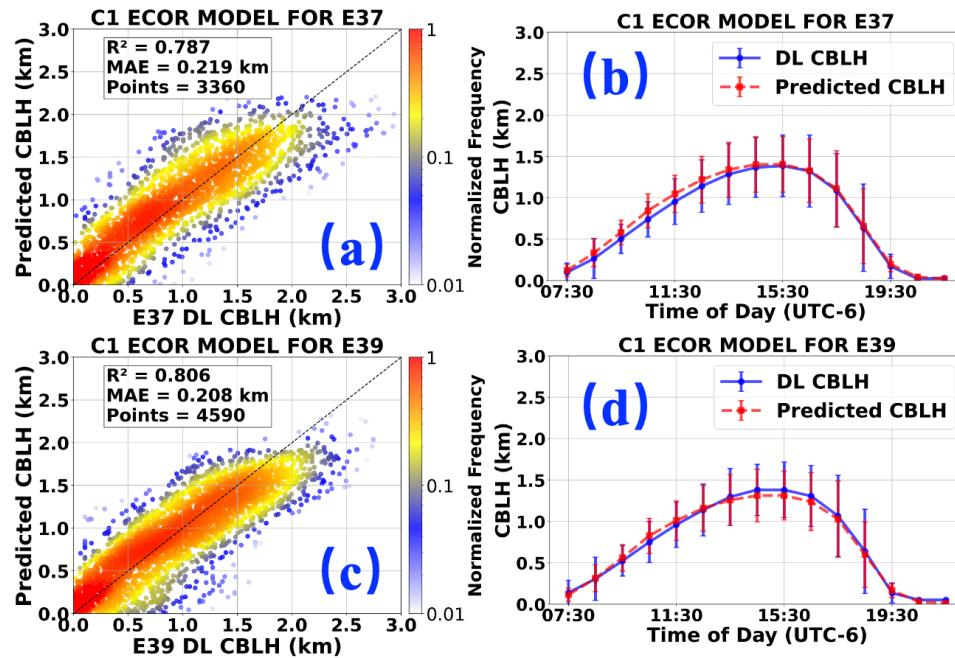


Figure 6. C1 ECOR model for E37 and E39 sites; (a) and (c) represent the R^2 and MAE for E37 and E39 respectively; (b) and (d) show the MEAN CBLH with IQR error bars for E37 and E39 respectively.

The above analysis indicates that while the Auto-ML model trained at the C1 site performs well at the E37 and E39 sites, its performance varies across different time periods, reflecting both similarities and differences in behavior at these sites. This highlights the spatial variability between sites. Furthermore, the differences between the C1 and E39 sites are smaller than those between the C1 and E37 sites, which aligns with their relative distances (41 km vs. 77 km).

3.2.2 Apply C1 site EBBR model to E32 and E39 Sites

The Auto-ML model trained at the C1 site using EBBR data performs well at the E39 site, achieving an R^2 of 0.795 and an MAE of 0.211 km (Fig. 7a), but shows a slight decline in performance at the E32 site, with an R^2 of 0.719 and an MAE of 0.269 km (Fig. 7c). However, Figure 7d reveals discrepancies in the predicted values near the CBL top, which closely resemble the performance of the C1 site with EBBR data (Fig. 5b). In contrast, Figure 6b indicates significant deviations at the E32 site during both the initial and top phases of CBLH (07:30–17:30 UTC-6), where the predicted values exceed the DL observations by up to ~200 m. Notably, no such pronounced discrepancies are observed in Fig. 5b, suggesting that the heat flux differences between the C1 and E32 sites are substantial during the initial phase and persist even after reaching the boundary layer top.

The above discussion highlights that while the Auto-ML model trained at the C1 site performs effectively at the E37 and E39 sites, it exhibits more pronounced differences at the E32 site. The variability between the C1 and E39 sites is smaller than that between the C1 and E32 sites, despite their comparable distances (41 km vs. 40 km).

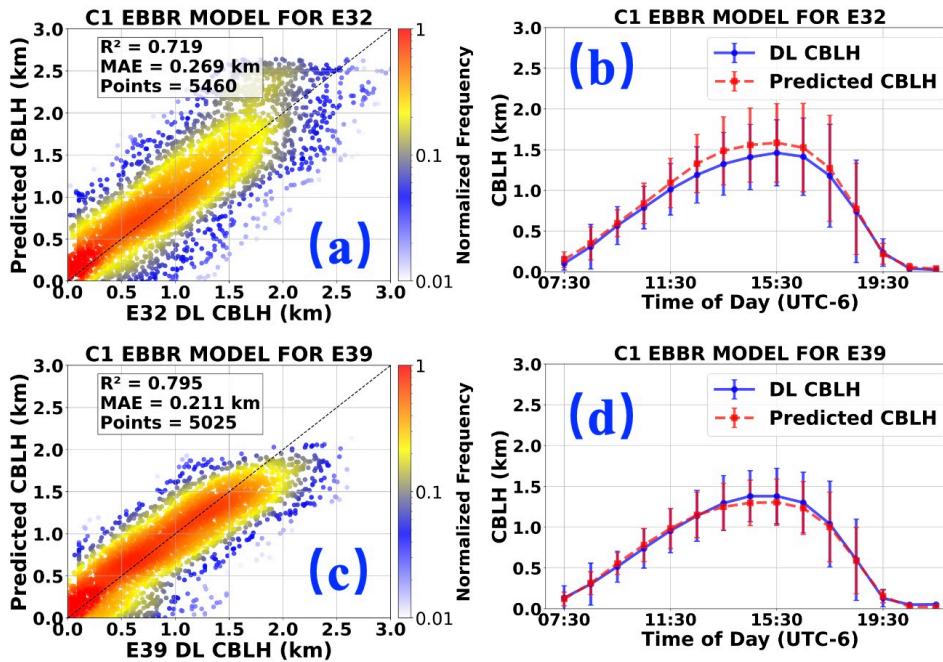


Figure 7. C1 EBBR model for E32 and E39 sites; (a) and (c) represent the R^2 and MAE for E32 and E39 respectively; (b) and (d) show the MEAN CBLH with IQR error bars for E32 and E39 respectively.

The Auto-ML algorithm demonstrates significant adaptability beyond the C1 site, underscoring its potential for wider application across multiple observational stations (Figs. 6–7). The Auto-ML model trained at the C1 site performs effectively at the E37 and E39 sites, while models trained at the E37 and E39 sites also exhibit robust performance at the C1 site, achieving R^2 values of approximately 0.80–0.85 and MAE values ranging from 0.19 to 0.23 km (figures not shown in this study).

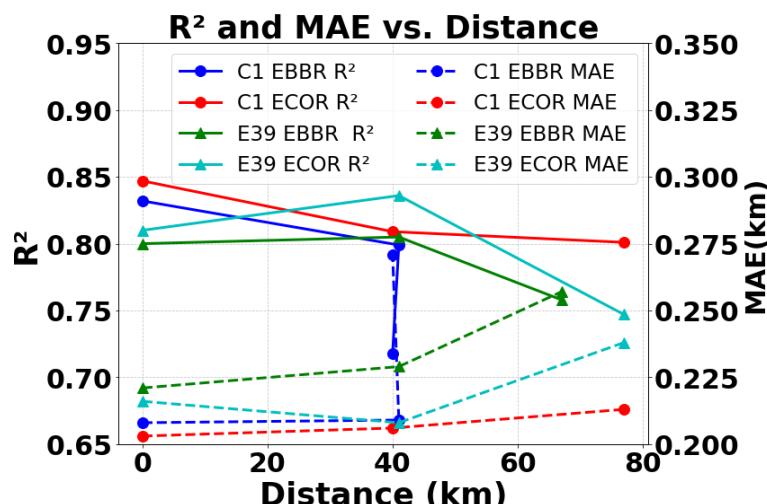
3.3 The relationship between Auto-ML performance and the spatial separation between sites

Section 3.2 demonstrates the cross-site applicability of the Auto-ML algorithm. To further investigate the relationship between Auto-ML performance and the spatial separation between sites, we selected data from the C1 site (E14 and E13 sites with ECOR and EBBR data, respectively) and the E39 site, training separate models using both datasets. These models were then applied to other sites with similar heat flux characteristics to assess the correlation between performance and inter-site distance.

The predictive performance of the Auto-ML algorithm exhibits a clear negative correlation with the spatial separation between observational sites (Fig. 8). For instance, as shown by the solid red and dashed lines in Fig. 8, the model trained on ECOR data from the C1 site performs robustly at its primary site ($R^2 = 0.847$, $MAE = 0.203 \text{ km}$), but its accuracy decreases at the E39 site 41 km away ($R^2 = 0.809$, $MAE = 0.206 \text{ km}$) and further declines at the E37 site 77 km away ($R^2 = 0.801$, $MAE = 0.213 \text{ km}$). Similarly, as depicted by the solid blue and dashed lines in Fig. 8, the model trained on ECOR data from the E39 site achieves strong performance at its primary site ($R^2 = 0.810$, $MAE = 0.216 \text{ km}$), but its accuracy higher at the C1 site 41 km away ($R^2 = 0.836$, $MAE = 0.208 \text{ km}$) and decreases further at the E37 site 77 km away ($R^2 = 0.747$, $MAE = 0.238 \text{ km}$).



525 Despite this overall trend of performance decline with distance, notable irregularities are observed. For example, the model trained on EBBR data from the C1 site performs best at its primary site ($R^2 = 0.834$, MAE = 0.205 km), with a slight decrease in accuracy at the E39 site 41 km to the southeast ($R^2 = 0.795$, MAE = 0.2011 km), but a more significant decline at the E32 site 40 km to the northwest ($R^2 = 0.719$, MAE = 0.269 km). Likewise, the model trained on EBBR data from the E39 site excels at its primary site
530 ($R^2 = 0.800$, MAE = 0.221 km), maintains comparable performance at the C1 site 41 km to the southeast ($R^2 = 0.805$, MAE = 0.229 km), but shows a substantial drop at the E32 site 67 km to the northwest ($R^2 = 0.758$, MAE = 0.257 km).



535 **Figure 8: Relationship between Auto-ML effectiveness and distance.**

It is noteworthy that the model trained on C1 EBBR data exhibits a sharp decline in R^2 to 0.71 at the E32 site, located 40 km away, while maintaining a robust R^2 of approximately 0.8 at the E39 site, 41 km from C1. One potential reason is that ARM's ECOR systems are typically surrounded by winter wheat fields or farmland, whereas EBBR systems are primarily deployed in pastures. The performance drop at E32
540 may stem from vegetation differences and measurement principles. E32's pasture dominated EBBR data, prone to overestimating latent heat flux, contrasts with the winter wheat fields around C1 and E39, likely measured by ECOR, which directly captures turbulent fluxes. These discrepancies in surface heat flux inputs challenge the model's generalization, particularly at E32, where site-specific factors like soil moisture or EBBR measurement errors near sunrise/sunset may further degrade performance.

545 This analysis reveals that the predictive performance of the Auto-ML algorithm exhibits a clear negative correlation with spatial separation between sites, accompanied by spatial heterogeneity. These irregularities align with theoretical expectations: local factors such as terrain variations (e.g., changes in elevation or surface roughness), land use differences (e.g., urban vs. rural settings), and microclimate effects (e.g., humidity or temperature gradients) disrupt the coherence of CBL dynamics with increasing distance. These site-specific perturbations limit the algorithm's generalizability across diverse regions.



3.4 Comparison of Performances of two ML methods for summer at the C1 site.

To compare algorithmic performance, this study evaluates two high-performing machine learning models: an ExtraTreesRegressor from TPOT and a neural network from AutoKeras. Both models are trained on the same dataset. The ExtraTreesRegressor is selected in this section because of its inherent resistance to overfitting. This robustness is achieved through randomized feature selection and split point selection. As a result, the model performs well on high-dimensional data and noisy datasets and shows strong resilience to outliers. However, it is not suitable for small-sample datasets. The best-performing model selected by AutoKeras is a neural network with 10,836 parameters, implemented using the Functional API. It comprises an input layer for 9-dimensional features, preprocessing layers for multi-category encoding and normalization (19 non-trainable parameters), two dense hidden layers with 256 and 32 units respectively (ReLU activation, 10,817 trainable parameters), and a regression output layer. The architecture leverages AutoKeras's automated feature engineering through integrated preprocessing, while its two-layer structure maintains moderate complexity. The parameter distribution (256 to 32 units) indicates a progressive reduction in feature dimensionality, supporting effective feature extraction for the regression task.

3.4.1 SHAP Computation Methods: Tree-Based vs. Gradient Approaches.

To compare the relative importance of features between the two methods, SHAP values for the ExtraTreesRegressor are computed directly using the TreeExplainer, which leverages the tree structure (split paths and leaf node values). In contrast, the AutoKeras neural network employs the GradientExplainer, a gradient-based method, to estimate SHAP values. Their summer (JJA) performance, shown in Fig. 9, reveals similar R^2 and MAE values: ExtraTreesRegressor (0.855, 0.221 km) versus neural networks (0.840, 0.245 km), as depicted in Fig. 9a (ExtraTreesRegressor) and Fig. 9e (neural networks). These consistent metrics highlight the robustness of both approaches in capturing CBLH. However, despite their similarity in overall performance, the two models diverge significantly (SHAP method) in their assessment of feature importance. In Fig. 9c, ExtraTreesRegressor assigns a notably higher importance to LTS (~0.23); attributes nearly equal importance to I_SHF, I_LHF, TIME, and SUNPERCENT each hovering around 0.15, indicating a clear prioritization of LTS in its decision-making process. In contrast, the neural network, as shown in Fig. 9g, assigns a notably higher importance to I_LHF (~0.28); attributes nearly equal importance to C_SHF, and I_SHF, each hovering around 0.2. The different machine learning models can achieve comparable accuracy by using varied nonlinear combinations of predictors. In such scenarios, the physical interpretation of these models becomes challenging or may lack sufficient reliability.

Figures 9b and 9d show that the diurnal variations predicted by the AutoKeras neural network and the TPOT ExtraTreesRegressor are generally comparable. However, Figure 9f reveals that the neural network predicts lower CBLH values than the ExtraTreesRegressor (Figure 9b) between 07:30–11:30 and around 19:30. Despite this, the neural network exhibits inferior performance, with lower R^2 and higher MAE compared to the ExtraTreesRegressor.



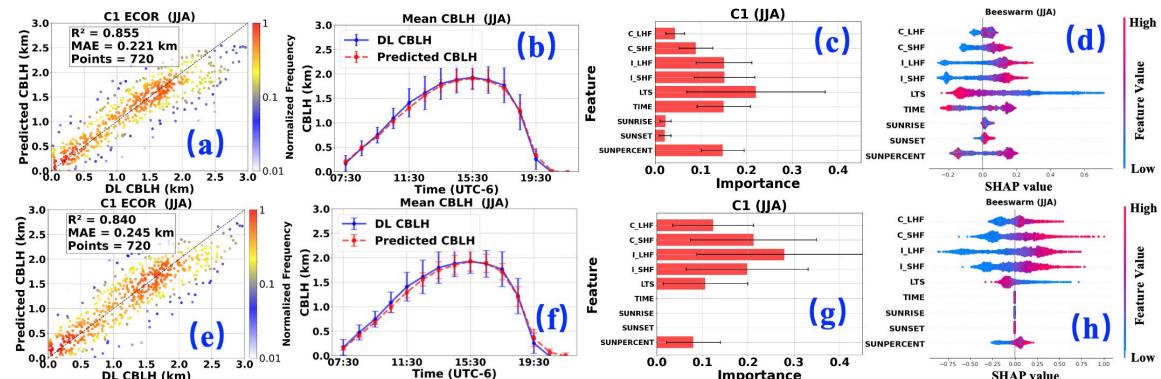
595

A comparison of Figures 9g–h (neural network) with Figures 9c–d (ExtraTreesRegressor) highlights distinct differences in feature contributions. For the neural network, the SHAP values and relative importance of TIME, SUNRISE, and SUNSET are zero, whereas SUNPERCENT retains a non-zero SHAP value. This suggests that the neural network effectively captures the information encoded in Equation 5, prioritizing SUNPERCENT as the primary contributor to CBLH predictions. In contrast, the ExtraTreesRegressor assigns reduced but non-zero relative importance to SUNRISE and SUNSET, indicating a broader distribution of feature contributions.

600

These differences likely stem from the distinct SHAP explainers used for each model. The ExtraTreesRegressor employs the TreeExplainer, which leverages the tree structure (split paths and leaf node values) to compute SHAP values directly, without requiring a background dataset. Conversely, the neural network uses the GradientExplainer, a local explanation method that relies on a background dataset (100 samples in this study) and computes SHAP values based on gradients near specific input points. 605 When the local gradient for features such as TIME, SUNRISE, and SUNSET approaches zero, this reflects their negligible impact on the model's local decision boundary, resulting in corresponding SHAP values of zero. This explains the neural network's tendency to assign zero importance to these features, while the ExtraTreesRegressor's global approach captures their residual contributions.

605



610

Figure 9. Performance comparison of two machine learning frameworks during summer (JJA) (a-d) ExtraTreesRegressor: (a) R^2 and MAE, (b) diurnal variations, (c) relative influence of input variables and (d) SHAP values of input variables. (e-h) Neural network: (e) R^2 and MAE, (f) diurnal variations, (g) relative influence of input variables and (h) SHAP values of input variables.

3.4.2 Comparative Analysis of SHAP Value Estimation Methods for AutoKeras Neural Networks

615

To validate the reliability of SHAP values and assess differences across computation methods, we compare the results of alternative SHAP explainers with those shown in Fig. 9g–h. The GradientExplainer, used for the AutoKeras neural network, approximates SHAP values by computing gradients of input features relative to model outputs, relying on a background dataset (100 samples in this study) to estimate feature contributions. The choice of background dataset can influence results, as GradientExplainer assumes local differentiability and quantifies feature importance based on gradient information. 620 Consequently, features with near-zero gradients (e.g., those with minimal local influence near the background dataset) may yield zero SHAP values.



To mitigate this limitation, two additional explainers were employed: (1) KernelExplainer, a model-agnostic method that estimates SHAP values through sampling and weighted regression, suitable for any model. By sampling the global feature space, it captures non-local or nonlinear contributions, potentially yielding non-zero SHAP values even when local gradients are zero. However, KernelExplainer still requires a background dataset. (2) ExactExplainer, which does not require an explicit background dataset but uses a masking strategy, typically `shap.maskers.Independent`, to implicitly define the background distribution based on the data itself. By precisely computing Shapley values for all feature combinations, ExactExplainer provides the theoretically most accurate SHAP estimates, though it is computationally intensive.

Table 3. Performance Comparison of Various SHAP Value Calculation Methods.

	SHAP Values		
	GradientExplainer	KernelExplainer	ExactExplainer
C_LHF	0.12	0.09	0.09
C_SHF	0.21	0.22	0.21
I_LHF	0.28	0.18	0.18
I_SHF	0.2	0.13	0.13
LTS	0.11	0.15	0.16
TIME	0	0.13	0.13
SUNRISE	0	0.02	0.02
SUNSET	0	0.01	0.01
SUNPERCENT	0.08	0.07	0.07

635

Table 3 summarizes the performance of different SHAP explainers. The GradientExplainer assigns zero SHAP values to input features with low influence, resulting in larger errors, but it offers high computational efficiency and requires fewer resources. In contrast, the ExactExplainer provides more reliable results but incurs high computational complexity, making it resource-intensive. For scenarios with limited computational resources and a need for high-accuracy SHAP values, the KernelExplainer is recommended as a balanced alternative. Notably, for features with lower relative importance, such as TIME, SUNRISE, and SUNSET, the ExactExplainer and KernelExplainer yield nearly identical results, with minor differences (approximately 0.01) observed for C_SHF and LTS.

645

Based on these findings, the SHAP value computation strategy in this study is as follows: For the TPOT-selected model (ExtraTreesRegressor), the TreeExplainer is used, leveraging its efficiency for tree-based models. For the AutoKeras neural network, the ExactExplainer is employed to compute SHAP values, as real-time computation is not required, prioritizing accuracy over speed.

650

3.5 Seasonal comparative analysis of Auto-ML's performance

3.5.1 Seasonal comparative analysis of Auto-ML's comprehensive performance

The performance of Auto-ML varies across different sites and seasons. As shown in Figure 10, after training with multi-year ECOR data at C1 sites, the model's performance is evaluated across four seasons. Figures 10a1–8a4 illustrate that autumn (SON) achieves the highest overall R^2 (0.860) but ranks second



in MAE (0.178 km). Winter (DJF) exhibits the lowest MAE (0.173 km) but the poorest R^2 (0.736). Summer (JJA) records the highest overall CBLH with strong performance (R^2 : 0.855, MAE: 0.221 km). Spring (MAM) yields an R^2 of 0.768 and an MAE of 0.239 km. Given the higher CBLH in summer and lower CBLH in winter, Auto-ML performs best overall in autumn.

660 3.5.2 Season-wise comparison of hourly averaged AutoML performance

665 However, when considering diurnal variations across seasons, summer (JJA) appears to perform best. As shown in Figures 10b1–9b4, predictions for spring (MAM) and autumn (SON) near the CBL top phase are approximately 0.1 km lower than DL observations. In winter (DJF), due to lower overall CBLH, predictions are about 0.05 km below observations. In contrast, summer (JJA) shows no significant discrepancy near the CBL top (12:30–14:30), with predictions only slightly lower (~0.05 km) around 11:30. Potential reasons include: (1) a larger number of summer data points, leading Auto-ML to prioritize summer performance, and (2) distinct entrainment processes in summer compared to other seasons.

670 The entrainment process at the top of the atmospheric boundary layer exhibits a dual influence on boundary layer development. When warm, dry air is entrained into the boundary layer, it enhances turbulent mixing and promotes vertical growth (Angevine et al., 1998). Conversely, if a strong inversion layer exists aloft, entrainment can suppress convection by dissipating turbulent kinetic energy and reducing upward heat flux (Lenschow et al., 2012). The entrainment rate ω_e , defined as the volume flux of air drawn from the free atmosphere into the mixing layer per unit time at the boundary layer top, follows 675 a modified form of the classical entrainment parameterization (Lilly, 1968; Deardorff, 1979; Stull, 1988; Sullivan et al., 1998):

$$\omega_e = A \frac{\omega_*^3}{h \Delta \theta_v} \quad (6)$$

680 where, the entrainment efficiency coefficient A is around 0.2 (Beare et al., 2006; Cuxart et al., 2006); $\beta \approx 0.5$ is an empirical coefficient representing stratification suppression (Pino et al., 2003); h is CBLH, ω_* is convective velocity scale (Deardorff, 1979); $\Delta \theta_v$ is virtual potential temperature jump (virtual potential temperature at the bottom of the inversion layer - average virtual potential temperature of the mixing layer); θ_{v0} is the surface virtual potential temperature. Equate 6 indicates that deeper boundary layers (larger h) exhibit weaker entrainment due to diminished turbulent energy (Beare et al., 2008). Additionally, when cloud fraction exceeds ~60%, the boundary layer growth rate declines by over 50%, 685 as cloud shading suppresses surface-driven turbulence (Zhang et al., 2020).

690 In humid summer conditions, high specific humidity (>18–20 g/kg) further inhibits boundary layer growth through multiple pathways: (a) increased cloudiness reduces surface solar heating (Luo et al., 2024), (b) precipitation depletes convective available potential energy (Hohenegger et al., 2013), and (c) evaporative cooling enhances stability (Zhang et al., 2003). Observational studies confirm that tropical moist boundary layers are 30–40% shallower than their arid counterparts (von Engeln & Teixeira, 2013), highlighting moisture's threshold-like suppression effect.

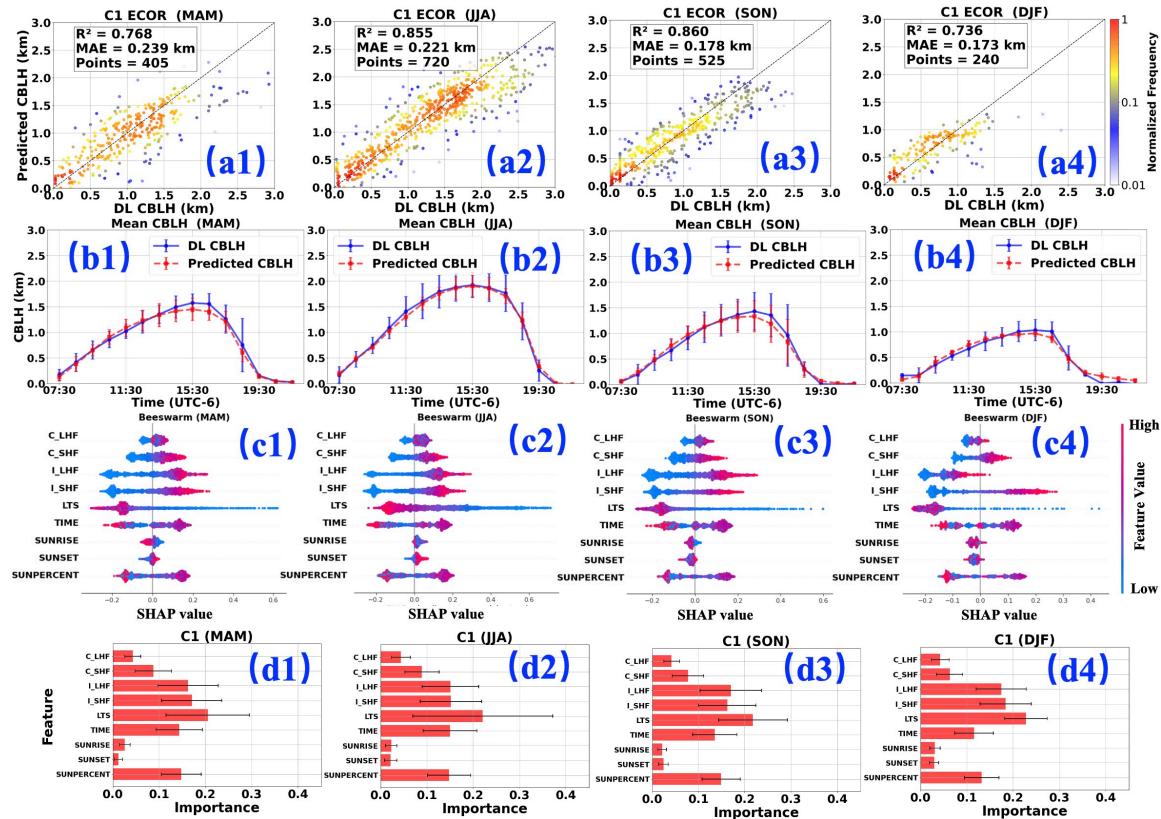


Figure 10. Seasonal performance of the C1 ECOR model. The four columns represent the four seasons: spring (MAM), summer (JJA), autumn (SON), and winter (DJF); (a) R^2 and MAE metrics; (b) Diurnal evolution of mean CBLH; (c) Beeswarm (SHAP values); (d) SHAP-derived feature importance.

Further analysis of Fig. 10b1–10b4 reveals that the interquartile range (IQR) error bars for predicted CBLH are consistently narrower than those for DL-derived CBLH. This is particularly evident in summer (JJA, Fig. 10b2), where, for instance, at 11:30, the IQR of DL CBLH is approximately four times that of the predicted values. This indicates that, despite the mean values of DL CBLH and predicted CBLH being similar (nearly overlapping at 11:30 with an error <0.01 km), the boundary layer development phase is governed by heat flux and LTS, which determine the attainable CBLH height. However, DL CBLH exhibits greater variability due to influences from additional meteorological factors, such as wind and low-level jets (LLJ). In contrast, the predicted CBLH, derived solely from thermodynamic parameters, lacks information on these factors, resulting in a more constrained range. A similar trend is observed in winter (DJF) with pronounced consistency. In spring and autumn, while a comparable pattern exists, the differences between predicted and observed values are smaller, suggesting lower variability (or complexity) in meteorological conditions compared to summer.

3.5.3 SHAP dependence between Auto-ML input parameters and CBLH

Analysis of Fig. 10c1–c4 reveals the relationships between various variables and CBLH. Initially, heat flux components (C_SHF, C_LHF, I_SHF, I_LHF) exhibit a predominantly positive correlation with

715 CBLH, indicating that stronger heat flux corresponds to greater CBLH. Conversely, LTS shows a negative correlation with CBLH, suggesting that higher LTS values are associated with reduced CBLH. Time and SUNPERCENT display a biphasic relationship with CBLH: a positive correlation is observed during the

720 pre-peak phase, where CBLH increases with time, while a negative correlation emerges post-peak as CBLH decreases with time. This behavior is consistent with CBL dynamics. Additionally, SUNRISE exhibits a weak negative correlation with CBLH, implying that later sunrise times correspond to lower CBLH, whereas earlier sunrise times are linked to higher CBLH. Similarly, later sunset times are associated with higher CBLH, and earlier sunset times with lower CBLH. These patterns align well with the established development processes of the atmospheric boundary layer. Since SUNPERCENT integrates the effects of time, sunrise, and sunset, its relationship with CBLH closely mirrors that of time.

725 While the general relationships between input parameters and CBLH are outlined above, seasonal variations are notable. For instance, LTS consistently exhibits a negative correlation with CBLH in spring

730 (MAM), autumn (SON), and winter (DJF). However, in summer, certain data points show a positive correlation, suggesting that under specific complex meteorological conditions in summer, factors beyond LTS dominate CBLH development, further highlighting the complexity of summer CBL dynamics. Additionally, I_LHF displays a positive correlation with CBLH across spring, summer, and autumn, but a negative correlation in winter, despite most corresponding SHAP values exceeding -0.1, creating a stark contrast with the other seasons. A similar, though less pronounced, trend is observed with C_SHF in winter. This phenomenon may be attributed to the dominance of northerly monsoon winds in winter, exacerbating cold, dry conditions. An increase in I_LHF could reduce I_SHF, potentially suppressing turbulence generation. In winter, both SUNRISE and SUNSET exhibit a negative correlation with CBLH, indicating that the overall CBLH remains low during this season. The specific mechanisms underlying

735 these phenomena require further investigation.

3.5.4 Relative importance of input parameters in Auto-ML

740 The relative importance of different parameters across the four seasons exhibits a consistent pattern. LTS ranks highest (0.2–0.25), as it determines the energy required for CBLH growth. Next are the instantaneous heat flux components (I_SHF: 0.15–0.18 and I_LHF: 0.15–0.18), indicating that current

745 heat flux plays a critical role in sustaining CBLH. Following these are TIME (0.12–0.15) and SUNPERCENT (0.12–0.15), which collectively govern the diurnal variation in boundary layer development; a positive correlation is observed before CBLH peaks (typically when SUNPERCENT is around 0.5), transitioning to a negative correlation post-peak. Subsequently, C_SHF (0.06–0.09) exceeds C_LHF (~0.05) in influence. The least impactful factors are SUNRISE (~0.02) and SUNSET (0.01–0.02).

750 These relative importance values are consistent with the discussion in Section 3.5.3.

However, minor seasonal variations in the relative importance of these parameters are observed. For instance, in summer, LTS reaches a relative importance of 0.22, but its error bar extends to 0.3, suggesting that while LTS temporarily dominates CBLH development, its influence is significantly modulated by other conditions, underscoring the complexity of summer boundary layer dynamics. In contrast, during



winter, LTS maintains a relative importance of approximately 0.23 with an error bar of only 0.1, indicating its dominant and stable role in governing winter boundary layer development.

3.6 Comparative analysis of multi-site training with site-specific testing

Previous analyses employed Auto-ML models trained separately for individual sites. To investigate potential performance improvements, we conducted experiments using combined multi-site training followed by site-specific testing while controlling for cross-heat-flux interference. Two independent test groups were evaluated: the ECOR cluster (C1, E37, E39) and the EBBR cluster (C1, E32, E39), with comparative results presented in Table 4. Figure 10 further illustrates the CBLH diurnal variation patterns to enhance temporal resolution analysis.

Table 4. The results of multi-site training with site-specific testing

	ECOR			EBBR		
	C1	E37	E39	C1	E32	E39
R2	0.851	0.832	0.806	0.837	0.778	0.82
MAE (km)	0.198	0.19	0.211	0.203	0.249	0.205
C_LHF	0.04	0.03	0.03	0.07	0.06	0.07
C_SHF	0.11	0.11	0.11	0.03	0.04	0.04
I_LHF	0.09	0.09	0.09	0.15	0.14	0.15
I_SHF	0.17	0.18	0.18	0.14	0.13	0.15
LTS	0.2	0.19	0.18	0.2	0.24	0.18
TIME	0.16	0.16	0.16	0.18	0.17	0.17
SUNRISE	0.04	0.05	0.05	0.03	0.03	0.04
SUNSET	0.03	0.03	0.03	0.03	0.03	0.03
SUNPERCENT	0.16	0.16	0.17	0.17	0.16	0.17

To evaluate the performance of the ECOR and EBBR models, we trained both using 70% of the data from three sites combined and tested them on the remaining 30% of site-specific data. The results show that ECOR outperforms EBBR overall, with an average R^2 of 0.830 and MAE of 0.200 across the three sites, compared to EBBR's average R^2 of 0.812 and MAE of 0.219. However, at site E39, EBBR achieves a better R^2 (0.820) and MAE (0.205) than ECOR's R^2 (0.806) and MAE (0.211). The overall performance surpassed that of training and testing solely with C1 site data: ECOR (R^2 : 0.851 vs. 0.845; MAE: 0.198 km vs. 0.207 km), EBBR (R^2 : 0.837 vs. 0.834; MAE: 0.203 km vs. 0.205 km).

In terms of input parameter importance, ECOR exhibits minimal variation, with all parameters contributing approximately 0.01, except for LTS at 0.02. For EBBR, the LTS contribution at site E32 (0.24) is 0.06 higher than at E39 (0.18). Despite E32's lower overall performance (R^2 : 0.778, MAE: 0.249) compared to E39 (R^2 : 0.820, MAE: 0.205), E32's predictions are more accurate near the convective boundary layer top (~15:30), closely aligning with observations (mean values nearly overlap). In contrast, E39's predictions at the same time are approximately 0.1 km lower than observed. This suggests that E32's emphasis on LTS enhances prediction accuracy near the CBL top (Fig.10), where LTS is a critical factor.

780

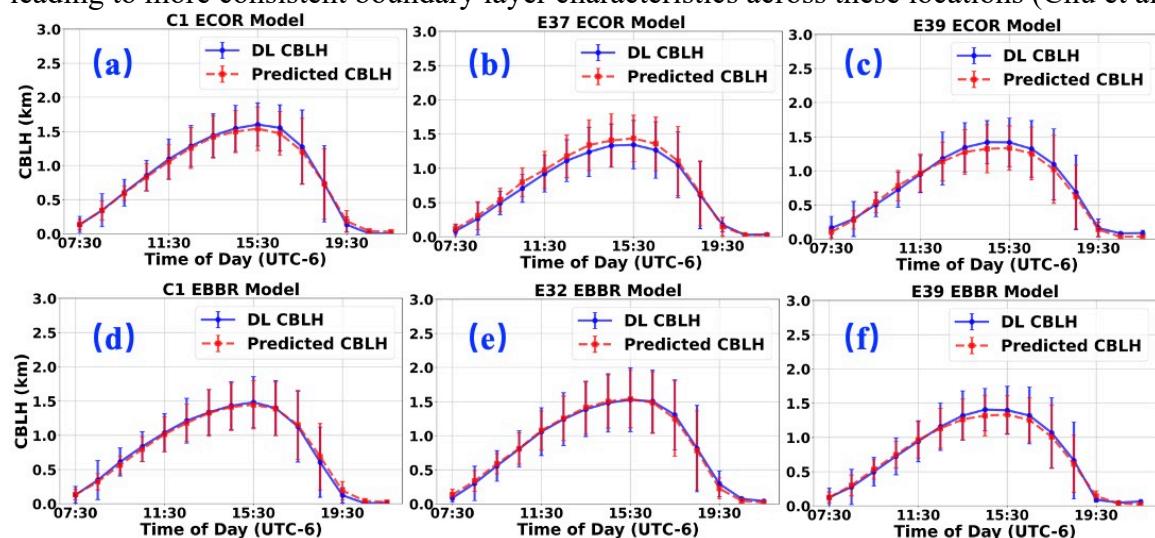
The primary influencing factors for ECOR and EBBR were similar (in table 4), with the largest difference observed in LTS (E32: 0.24, E39: 0.18). Additionally, the allocation of SHF and LHF at E39 differed



significantly between ECOR and EBBR, with notable disparities in C_LHF (0.07 vs. 0.03) and C_SHF (0.03 vs. 0.04). However, other factors, such as LTS and SUNPERCENT, showed near-identical patterns.
 785 These findings highlight discrepancies in heat flux measurements between ECOR and EBBR. Such differences typically require data assimilation in traditional PBL schemes but can be mitigated through machine learning's nonlinear combinations, yielding comparable CBLH estimates. This approach could also facilitate future heat flux data assimilation.

790 The lower overall R^2 and MAE at E32, combined with its distinct LTS contribution, indicate that local meteorological conditions at E32 differ from those at the other sites. According to Tang (2019), E32 is surrounded primarily by pasture, unlike the seasonal crops and grasslands near C1 and E39, consistent with findings in Section 3.3. Including E32's data in training improves its performance (R^2 : 0.778, MAE: 0.249) compared to Section 3.3 (R^2 : 0.718, MAE: 0.271), highlighting the benefit of site-specific data in
 795 model training.

800 As shown in the diurnal variations across different sites (Fig. 11), noticeable discrepancies remain among them. For the ECOR measurements, the predicted CBL top heights are lower than the observations at the C1 and E39 sites, whereas at the E37 site, the predicted values exceed the observed ones. This suggests
 805 that the entrainment process near the CBL top at E37 differs from that at C1 and E39. In contrast, the EBBR-based results show that predictions at C1 and E32 closely match the observations near the CBL top throughout the day, while at E39, the model tends to underestimate the CBL top height. Overall, the discrepancy in EBBR-based estimates is smaller than that of ECOR, particularly near the CBL top. One possible explanation is that the E39, C1, and E32 sites lie along a southeast wind trajectory, potentially leading to more consistent boundary layer characteristics across these locations (Chu et al., 2025a).



810 Figure 11. Diurnal variation comparison of multi-site trained models evaluated through site-specific testing. Results are grouped by heat flux measurement system: (Top) ECOR sites (a) C1 (E14), (b) E37, (c) E39; (Bottom) EBBR sites (a) C1 (E13), (b) E32, (c) E39.



4 Summary and conclusions

This study develops an Auto-ML framework for predicting CBLH, guided by thermodynamic physical constraints and the implicit diurnal cycle of CBLH. By leveraging TPOT and AutoKeras to automatically select optimal models, the approach bypasses manual comparisons of machine learning algorithms, enhancing efficiency and reproducibility. The resulting Auto-ML models, validated against Doppler lidar CBLH measurements, demonstrate robust performance with an overall R^2 of 0.84. Comparisons between ECOR and EBBR techniques for measuring surface heat and energy fluxes reveal consistent predictions, with an R^2 difference of approximately 0.011 and a MAE of 0.002 km. The models exhibit strong adaptability across multiple sites. When trained on ECOR data from the C1 site and applied to E37 and E39 sites within the ARM SGP network, the models achieve R^2 values of 0.787 and 0.806, respectively. Models trained on combined C1 and E39 data and tested on other sites show a gradual decline in R^2 and MAE with increasing distance yet maintain high predictive accuracy. These results underscore the transferability of the Auto-ML framework, highlighting its potential for integration with traditional numerical weather prediction models.

To enhance model interpretability, we conducted a comprehensive analysis of SHAP values using different explainers. The ExactExplainer was selected for its high accuracy in computing SHAP values for the AutoKeras neural network, despite its computational cost, as real-time computation was not required. In contrast, the TreeExplainer was employed for the TPOT-selected ExtraTreesRegressor, leveraging its efficiency for tree-based models. Seasonal performance at the C1 site was further evaluated, with a targeted comparison of model predictions during summer (JJA), revealing consistent feature importance patterns (e.g., dominance of LTS).

The study compared the performance of the C1 ECOR site across four seasons, revealing that autumn (SON) exhibited the best performance (R^2 : 0.860, MAE: 0.178 km). This may be attributed to fewer clouds in autumn, reducing the influence of CBL top entrainment on CBLH development. Subsequently, the model was trained using data from multiple sites and tested individually. The overall performance surpassed that of training and testing solely with C1 site data: ECOR (R^2 : 0.851 vs. 0.845; MAE: 0.198 km vs. 0.207 km), EBBR (R^2 : 0.837 vs. 0.834; MAE: 0.203 km vs. 0.205 km). ECOR sites (C1, E32, E39) generally outperformed EBBR sites (C1, E37, E39) on average, though EBBR at E39 outperformed ECOR. The primary influencing factors for ECOR and EBBR were similar, with the largest difference observed in LTS (E32: 0.24, E39: 0.18). Additionally, the allocation of SHF and LHF at E39 differed significantly between ECOR and EBBR, with notable disparities in C_LHF (0.07 vs. 0.03) and C_SHF (0.03 vs. 0.04). However, other factors, such as LTS and SUNPERCENT, showed near-identical patterns. These findings highlight discrepancies in heat flux measurements between ECOR and EBBR. Such differences typically require data assimilation in traditional PBL schemes but can be mitigated through machine learning's nonlinear combinations, yielding comparable CBLH estimates. This approach could also facilitate future heat flux data assimilation. This implicit physically constrained Auto-ML approach significantly improves the accuracy and generalizability of CBLH predictions across diverse sites and seasons. By providing a scalable framework for boundary layer parameterization, it offers valuable



insights for refining atmospheric models and advancing the integration of machine learning in operational weather forecasting.

It should be noted that, although this study consistently refers to Auto-ML as "predicting" the CBLH, in
855 the context of PBL schemes, it is more accurately described as "diagnosing" CBLH, given that the model
uses a full day of data as both input and output. To enhance the model's applicability, it is critical to align
it with the conventions of traditional PBL schemes by incorporating the CBLH output from the previous
time step as an input for predicting the subsequent CBLH. Preliminary application of the model at the C1
site produces results consistent with those reported in this study ($R^2 = 0.822$; $MAE = 0.199$ km). The next
860 step involves further optimization to meet additional requirements: extracting parameters from the CCPP-
SCM PBL framework to predict the CBLH by Auto-ML, and then feeding this output back into the PBL
framework to forecast the CBLH at the subsequent time step.

The Auto-ML PBL model has broad applications due to its accuracy and efficiency. It can support air
865 quality forecasting by better predicting pollutant dispersion within the PBL, which is crucial for urban
and industrial areas (Garratt, 1992; Stensrud, 2007). Additionally, its lightweight design makes it ideal
for integration into local data-driven weather forecasting systems, providing accurate CBLH inputs to
support low-altitude economic activities (Ben et al., 2024). The Auto-ML driven scalability further
870 enables its use in data assimilation, integrating diverse observations for improved model initialization
(Arcucci et al., 2021; Arcomano et al., 2023). As observational networks like ARM expand, this model
offers a versatile tool for global atmospheric research.

The lightweight Auto-ML PBL model exhibits limitations in predicting peak MLH values near the PBL
875 top, primarily due to two interrelated factors. First, its reliance on lidar-derived training data introduces
uncertainties at higher altitudes, where a reduced SNR obscures sharp inversion layers. Second, although the AutoML framework captures energy balance constraints to some extent, it does not fully
represent critical physical processes such as entrainment dynamics and turbulence-driven mixing at the
880 top of the CBL. These processes are especially important during the peak development phase of the
boundary layer. At this stage, shear-driven turbulence and buoyancy fluxes play a dominant role in
promoting vertical mixing and facilitating the entrainment of free-atmosphere air into the CBL. Without
885 explicitly incorporating these mechanisms, the model may underrepresent key drivers of boundary layer
growth, particularly under conditions of strong surface heating or elevated wind shear. This limitation
highlights the need for physically informed hybrid models that can integrate data-driven approaches with
boundary-layer process understanding. Unlike traditional schemes that parameterize these through TKE
890 budgets or non-local mixing, the Auto-ML model lacks such dynamic constraints, reducing its sensitivity
to abrupt inversion layer changes or synoptic-scale forcings (e.g., advective momentum fluxes) (Stevens,
2002; Cuxart et al., 2006; Fernando, 2010; Shin et al., 2021;). The IQR error bars for predicted CBLH
are consistently narrower than those for DL-derived CBLH across all seasons, reflecting lower variability
in predicted CBLH (based on thermodynamic parameters) compared to DL-derived CBLH, which is
influenced by additional factors such as wind and low-level jets. Despite these limitations, a follow-up
work is underway to develop an AI/ML-based emulator for parameterizing PBL, which will be applied
into real-case simulations to assess its performance against conventional PBL schemes. Furthermore,



future enhancements may incorporate high-resolution observational data (e.g., uncrewed aircraft systems or airborne LiDAR) to directly sample the entrainment zone, thereby improving physical consistency.
895 Additionally, integrating other boundary-layer parameters (e.g., turbulent dissipation rate; Chu et al., 2025b) could further refine the model, enabling predictions of additional PBL variables for a more comprehensive parameterization scheme.

Code and Data Availability

900 ARM data are available at <https://doi.org/10.2172/1253897> (ARM, 2016). The algorithm for convective boundary layer height determination is also accessible at <https://doi.org/10.1364/OE.451728> (Chu et al., 2022a).

Author Contributions

905 YC and HG conceptualized the study. YC, LX, GL, and HG curated the data. YC conducted the formal analysis. YC, LX, and HG developed the methodology. YC and HG implemented the software. ZW provided supervision. YC performed validation. YC, GL, and LX wrote the original draft. YC, GL, LX, MD, H.H.S., JZ, HG, and ZW contributed to writing, review, and editing.

Competing Interests

910 The contact author has declared that none of the authors has any competing interests.

Disclaimer

No disclaimer is applicable.

Acknowledgements

915 This research was supported by the DOE Atmospheric System Research (ASR) program under grant DE-SC0020171 and the National Science Foundation (NSF) under grant AGS-1917693. Jun Zhang acknowledges support from NOAA under grant NA22OAR4050669D, the Office of Naval Research (ONR) MURI under grant N00014-24-1-2554, and the National Science Foundation (NSF) under awards 2228299 and 2211308.



Financial Support

920 No disclaimer is applicable.

References

Angevine, W. M., Grimsdell, A. W., Hartten, L. M., and Delany, A. C.: The Flatland boundary layer experiments, *Bull. Am. Meteorol. Soc.*, 79, 419–432, doi:10.1175/1520-0477(1998)079<0419:TFBLE>2.0.CO;2, 1998.

925 Arcomano, T., Szunyogh, I., Wikner, A., Pathak, J., Hunt, B. R., Ott, E.: A hybrid atmospheric model incorporating machine learning can capture dynamical processes not captured by its physics-based component, *Geophys. Res. Lett.*, 50, e2022GL102649, doi:10.1029/2022GL102649, 2023.

930 Arcucci, R., Zhu, J., Hu, S., Guo, Y. K.: Deep data assimilation: integrating deep learning with data assimilation, *Appl. Sci.*, 11, 1114, doi:10.3390/app11031114, 2021.

935 Ayazpour, Z., Tao, S., Li, D., Scarino, A. J., Kuehn, R. E., and Sun, K.: Estimates of the spatially complete, observational-data-driven planetary boundary layer height over the contiguous United States, *Atmos. Meas. Tech.*, 16, 563–580, doi: 10.5194/amt-16-563-2023, 2023.

940 Baklanov, A., Schlünzen, K. H., Suppan, P., Baldasano, J. M., Brunner, D., Aksoyoglu, S., Carmichael, G., Dourous, J., Flemming, J., Forkel, R., ... & Lohmann, U.: Online coupled regional meteorology chemistry models in Europe: current status and prospects, *Atmos. Chem. Phys.*, 14, 317–398, doi: 10.3929/ethz-b-000077229, 2013.

Barlow, J. F., Dunbar, T. M., Nemitz, E. G., Wood, C. R., Gallagher, M. W., Davies, F., O'Connor, E., and Harrison, R. M.: Boundary layer dynamics over London, UK, as observed using Doppler lidar during REPARTEE-II, *Atmos. Chem. Phys.*, 11, 2111–2125, doi:10.5194/acp-11-2111-2011, 2011.

945 Barlow, J. F., Halios, C. H., Lane, S. E., and Wood, C. R.: Observations of urban boundary layer structure during a strong urban heat island event, *Environ. Fluid Mech.*, 15, 373–398, doi:10.1007/s10652-014-9335-6, 2015.

950 Beare, R. J., MacVean, M. K., Holtslag, A. A. M., Cuxart, J., Esau, I., Golaz, J.-C., Jimenez, M. A., Khairoutdinov, M., Kosovic, B., Lewellen, D., Lund, T. S., Lundquist, J. K., McCabe, A., Moene, A. F., Noh, Y., Raasch, S., and Sullivan, P.: An intercomparison of large-eddy simulations of the stable boundary layer, *Bound.-Lay. Meteorol.*, 118, 247–272, doi:10.1007/s10546-005-3780-3, 2006.

955 Beare, R. J.: The role of shear in the morning transition boundary layer, *Bound.-Layer Meteorol.*, 129, 395–410, doi:10.1007/s10546-008-9324-8, 2008.



960 Ben Bouallègue, Z., Clare, M. C. A., Magnusson, L., and Co-authors: The rise of data-driven weather forecasting: a first statistical assessment of machine learning-based weather forecasts in an operational-like context, *Bull. Am. Meteorol. Soc.*, 105, E864–E883, doi:10.1175/BAMS-D-23-0162.1, 2024.

965 Bianco, L., Djalalova, I. V., King, C. W., and Wilczak, J. M.: Diurnal evolution and annual variability of boundary-layer height and its correlation to other meteorological variables in California's Central Valley, *Boundary-Layer Meteorol.*, 141, 491–514, doi:10.1007/s10546-011-9637-2, 2011.

970 Breiman, L., Friedman, J. H., Olshen, R. A., and Stone, C. J.: *Classification and regression trees*, Chapman and Hall/CRC, doi:10.1201/9781315139470, 1984.

975 Breiman, L.: Random forests, *Mach. Learn.*, 45, 5–32, doi:10.1023/A:1010933404324, 2001.

980 Brown, A. R., Beare, R. J., Edwards, J. M., Lock, A. P., Keogh, S. J., Milton, S. F., and Walters, D. N.: Upgrades to the boundary-layer scheme in the Met Office numerical weather prediction model, *Boundary-Layer Meteorol.*, 128, 117–132, doi:10.1007/s10546-008-9275-0, 2008.

985 975 Chen, T. and Guestrin, C.: XGBoost: A scalable tree boosting system, in: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794, doi:10.1145/2939672.2939785, 2016.

990 Chu, Y., Liu, D., Wang, Z., Wu, D., Deng, Q., Li, L., Zhuang, P., and Wang, Y.: Basic principle and technical progress of Doppler wind lidar, *Chinese J. Quantum Electron.*, 37, 580–600, doi:10.3969/j.issn.1007-5461.2020.05.008, 2020.

995 Chu, Y., Wang, Z., Xue, L., Li, Y., Wang, Q., and Zhang, Y.: Characterizing warm atmospheric boundary layer over land by combining Raman and Doppler lidar measurements, *Opt. Express*, 30, 11892–11911, doi:10.1364/OE.30.011892, 2022a.

990 Chu, Y., Wang, Z., Deng, M., Lin, G., and Zhang, Y.: The spatial and temporal variability of the planetary boundary layer at the ARM SGP supersite, in: *AGU Fall Meeting Abstracts*, Vol. 2022, A35K-1601, 2022b.

995 Chu, Y., Lin, G., Deng, M., Wang, Z., Zhang, Y., and Li, Y.: Characterizing seasonal variation of the atmospheric mixing layer height using machine learning approaches, *Remote Sens.*, 17(8), 1399, doi:10.3390/rs17081399, 2025a.

995 Chu, Y., Lin, G., Deng, M., and Wang, Z.: Characteristics of Eddy Dissipation Rates in Atmosphere Boundary Layer Using Doppler Lidar, *Remote Sens.*, 17(9), 1652, doi:10.3390/rs17091652, 2025b.



000 Cohen, A. E., Cavallo, S. M., Coniglio, M. C., and Brooks, H. E.: A review of planetary boundary layer parameterization schemes and their sensitivity in simulating southeastern U.S. cold season severe weather environments, *Weather Forecast.*, 30, 591–612, doi:10.1175/WAF-D-14-00105.1, 2015.

005 Cohn, S. A. and Angevine, W. M.: Boundary layer height and entrainment zone thickness measured by lidars and wind-profiling radars, *J. Appl. Meteorol.*, 39, 1233–1247, doi:10.1175/1520-0450(2000)039<1233:BLHAEZ>2.0.CO;2, 2000.

010 Compton, J. C., Delgado, R., Berkoff, T. A., and Hoff, R. M.: Determination of planetary boundary layer height on short spatial and temporal scales: A demonstration of the covariance wavelet transform in ground-based wind profiler and lidar measurements, *J. Atmos. Ocean. Tech.*, 30, 1566–1575, doi:10.1175/JTECH-D-12-00116.1, 2013.

015 Cook, D. R.: Energy Balance Bowen Ratio (EBBR) handbook, ARM Tech. Rep., DOE/SC-ARM/TR-037, 2005.

Cook, D. R.: Eddy Correlation Flux Measurement System (ECOR) handbook, ARM Tech. Rep., DOE/SC-ARM/TR-052, 2011.

020 Compton, J. C., Delgado, R., Berkoff, T. A., and Hoff, R. M.: Determination of planetary boundary layer height on short spatial and temporal scales: A demonstration of the covariance wavelet transform in ground-based wind profiler and lidar measurements, *J. Atmos. Oceanic Technol.*, 30, 1566–1575, doi:10.1175/JTECH-D-12-00116.1, 2013.

025 Cuxart, J., Holtslag, A. A. M., Beare, R. J., Bazile, E., Beljaars, A., Cheng, A., Conangla, L., Ek, M., Freedman, F., Hamdi, R., Kerstein, A., Kitagawa, H., Lenderink, G., Lewellen, D., Mailhot, J., Mauritsen, T., Perov, V., Schayes, G., Steeneveld, G.-J., Svensson, G., Taylor, P., Weng, W., Wunsch, S., and Xu, K. M.: Single-column model intercomparison for a stable PBL, *Bound.-Layer Meteorol.*, 118, 273–303, doi:10.1007/s10546-005-3780-1, 2006.

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

035 de Burgh-Day, C. O. and Leeuwenburg, T.: Machine learning for numerical weather and climate modelling: a review, *Geosci. Model Dev.*, 16, 6433–6477, doi:10.5194/gmd-16-6433-2023, 2023.

040 Deardorff, J. W.: Prediction of convective mixed-layer entrainment for realistic capping inversion structure, *J. Atmos. Sci.*, 36, 424–436, doi:10.1175/1520-0469(1979)036<0424:POCMLE>2.0.CO;2, 1979.



Drucker, H., Burges, C. J. C., Kaufman, L., Smola, A. J., and Vapnik, V.: Support vector regression machines, in: Advances in Neural Information Processing Systems, 9, 155–161, <https://papers.nips.cc/paper/1238-support-vector-regression-machines>, 1997.

045 Du, L., Pan, Y. N., and Wang, W.: Random sample fitting method to determine the planetary boundary layer height using satellite-based lidar backscatter profiles, *Remote Sens.*, 12, 4006, doi:10.3390/rs12234006, 2020.

050 Emeis, S., Schäfer, K., and Münkel, C.: Observation of the structure of the urban boundary layer with different ceilometers and validation by RASS data, *Meteorol. Z.*, 18(2), 149–164 doi: 10.1127/0941-2948/2009/0365, 2009.

055 Feltz, W. F., Smith, W. L., Howell, H. B., Knuteson, R. O., Woolf, H., Revercomb, H. E.: Near-continuous profiling of temperature, moisture, and atmospheric stability using the Atmospheric Emitted Radiance Interferometer (AERI), *J. Appl. Meteorol.*, 42, 584–597, doi:10.1175/1520-0450(2003)042<0584:NPOTMA>2.0.CO;2, 2003.

060 Fernando, H. J. S.: Fluid dynamics of urban atmospheres in complex terrain, *Annu. Rev. Fluid Mech.*, 42, 365–389, doi:10.1146/annurev-fluid-121108-145459, 2010.

065 Garratt, J. R.: The atmospheric boundary layer, *Earth-Sci. Rev.*, 37, 89–134, doi:10.1016/0012-8252(94)90026-4, 1994.

070 Grenier, H. and Bretherton, C. S.: A moist PBL parameterization for large-scale models and its application to subtropical cloud-topped marine boundary layers, *Mon. Weather Rev.*, 129, 357–377, doi:10.1175/1520-0493(2001)129<0357:AMPPFL>2.0.CO;2, 2001.

075 Hägeli, P., Steyn, D. G., and Strawbridge, K. B.: Spatial and temporal variability of mixed-layer depth and entrainment zone thickness, *Boundary-Layer Meteorol.*, 97, 47–71, doi:10.1023/A:1002790114980, 2000.

080 He, K., Zhang, X., Ren, S., and Sun, J.: Deep residual learning for image recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 770–778, doi:10.1109/CVPR.2016.90, 2016.

075 Helmis, C. G., Sgouros, G., Tombrou, M., Schäfer, K., Münkel, C., Bossioli, E., and Dandou, A.: A comparative study and evaluation of mixing-height estimation based on sodar-RASS, ceilometer data and numerical model simulations, *Boundary-Layer Meteorol.*, 145, 507–526, doi:10.1007/s10546-012-9743-4, 2012.



Hennemuth, B. and Lammert, A.: Determination of the atmospheric boundary layer height from radiosonde and lidar backscatter, *Boundary-Layer Meteorol.*, 120, 181–200, doi:10.1007/s10546-005-9035-8, 2006.

085 Heo, B.-H., Jacoby-Koaly, S., Kim, K.-E., Campistron, B., Benech, B., and Jung, E.-S.: Use of the Doppler spectral width to improve the estimation of the convective boundary layer height from UHF wind profiler observations, *J. Atmos. Oceanic Technol.*, 20, 408–424, doi:10.1175/1520-0426(2003)020<0408:UOTDSW>2.0.CO;2, 2003.

090 Hochreiter, S. and Schmidhuber, J.: Long short-term memory, *Neural Comput.*, 9, 1735–1780, doi:10.1162/neco.1997.9.8.1735, 1997.

Hohenegger, C. and Stevens, B.: Preconditioning deep convection with cumulus congestus, *J. Atmos. Sci.*, 70, 448–464, doi:10.1175/JAS-D-12-089.1, 2013.

095 Holtslag, A. A. M., Svensson, G., Baas, P., Basu, S., Beare, B., Beljaars, A. C. M., ... and van de Wiel, B. J. H.: Stable atmospheric boundary layers and diurnal cycles: challenges for weather and climate models, *Bull. Amer. Meteor. Soc.*, 94, 1691–1706, doi: 0.1175/BAMS-D-11-00187.1, 2013.

100 Hornik, K., Stinchcombe, M., and White, H.: Multilayer feedforward networks are universal approximators, *Neural Netw.*, 2, 359–366, doi:10.1016/0893-6080(89)90020-8, 1989.

105 Hu, X.-M., Nielsen-Gammon, J. W., and Zhang, F.: Evaluation of three planetary boundary layer schemes in the WRF model, *J. Appl. Meteorol. Climatol.*, 49, 1831–1844, doi:10.1175/2010JAMC2432.1, 2010.

Jin, H., Song, Q., and Hu, X.: Auto-Keras: An Efficient Neural Architecture Search System, in: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, Anchorage, AK, USA, 4–8 August 2019, 1946–1956, doi:10.1145/3292500.3330648, 2019.

110 Krause, P., Boyle, D. P., and Bäse, F.: Comparison of different efficiency criteria for hydrological model assessment, *Adv. Geosci.*, 5, 89–97, doi:10.5194/adgeo-5-89-2005, 2005.

115 Krishnamurthy, R., Newsom, R., Berg, L., Xiao, H., Ma, P., and Turner, D.: On the estimation of boundary layer heights: a machine learning approach, *Atmos. Meas. Tech. Discuss.*, 2020, 1–34, doi:10.5194/amt-2020-217, 2020.

Krishnamurthy, R., Newsom, R. K., Chand, D., and Shaw, W. J.: Boundary layer climatology at ARM Southern Great Plains, PNNL-30832, Pacific Northwest National Laboratory, Richland, WA, USA, doi:10.2172/1778833, 2021.

120



125 Lammert, A. and Bösenberg, J.: Determination of the convective boundary-layer height with laser remote sensing, *Boundary-Layer Meteorol.*, 119, 159–185, doi:10.1007/s10546-005-9020-2, 2006.

130 125 LeCun, Y., Bottou, L., Bengio, Y., and Haffner, P.: Gradient-based learning applied to document recognition, *Proc. IEEE*, 86, 2278–2324, doi:10.1109/5.726791, 1998.

135 130 Legates, D. R. and McCabe Jr., G. J.: Evaluating the use of “goodness-of-fit” measures in hydrologic and hydroclimatic model validation, *Water Resour. Res.*, 35, 233–241, doi:10.1029/1998WR900018, 1999.

140 135 Lenschow, D. H., Lothon, M., Mayor, S. D., Sullivan, P. P., and Canut, G.: A comparison of higher-order vertical velocity moments in the convective boundary layer from lidar with in situ measurements and large-eddy simulation, *Bound.-Layer Meteorol.*, 143, 107–123, doi:10.1007/s10546-011-9615-3, 2012.

145 140 Liang, S., Fei, Z., Wu, J., and Co-authors: Tree-based pipeline optimization-based automated-machine learning model for performance prediction of materials and structures: case studies and UI design, *Struct. Control Health Monit.*, 2024, 1485739, doi:10.1155/2024/1485739, 2024.

150 145 Lilly, D. K.: Models of cloud-topped mixed layers under a strong inversion, *Quart. J. Roy. Meteor. Soc.*, 94, 292–309, doi:10.1002/qj.49709440106, 1968.

155 150 Lin, G., Butterworth, B. J., Ondov, B. D., Bal, M., Paleri, S., Ma, J., Desai, A. R., and Mineau, J. M.: Airborne measurements of scale-dependent latent heat flux impacted by water vapor and vertical velocity over heterogeneous land surfaces during the CHEESEHEAD19 campaign, *J. Geophys. Res. Atmos.*, 129, e2023JD039586, doi:10.1029/2023JD039586, 2024.

155 150 Liu, S. and Liang, X.-Z.: Observed diurnal cycle climatology of planetary boundary layer height, *J. Clim.*, 23, 5790–5809, doi:10.1175/2010JCLI3552.1, 2010.

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

155 Luo, T., Yuan, R., and Wang, Z.: Lidar-based remote sensing of atmospheric boundary layer height over land and ocean, *Atmos. Meas. Tech.*, 7, 173–182, doi:10.5194/amt-7-173-2014, 2014.



160 Luo, Z., Song, X., Yin, J., Bu, Z., Chen, Y., Yu, Y., and Zhang, Z.: Comparison and verification of
coherent Doppler wind lidar and radiosonde data in the Beijing urban area, *Adv. Atmos. Sci.*, 41, 2203–
2214, doi:10.1007/s00376-024-3240-9, 2024.

165 Macatangay, R., Sonkaew, T., Bran, S. H., Thongsame, W., Supasri, T., Panya, M., and Haug, A.:
Decoupling the PBL Height, the Mixing Layer Height, and the Aerosol Layer Top in LiDAR
Measurements over Chiang Mai, Northern Thailand, *EGUsphere*, 2025, 1–14, doi:10.5194/egusphere-
2025-630, 2025.

170 Madonna, F., Summa, D., Di Girolamo, P., Marra, F., Wang, Y., and Rosoldi, M.: Assessment of trends
and uncertainties in the atmospheric boundary layer height estimated using radiosounding observations
over Europe, *Atmosphere*, 12, 301, doi:10.3390/atmos12030301, 2021.

175 Manninen, A. J., O'Connor, E. J., Vakkari, V., Petäjä, T.: A novel post-processing algorithm for Halo
Doppler lidars, *Atmos. Meas. Tech.*, 12, 435–444, doi:10.5194/amt-12-435-2019, 2019.

180 Olson, R. S., Bartley, N., Urbanowicz, R. J., and Moore, J. H.: Evaluation of a Tree-based Pipeline
Optimization Tool for Automating Data Science, in: *Proceedings of the Genetic and Evolutionary
Computation Conference 2016*, Denver, CO, USA, 20–24 July 2016, 485–492,
doi:10.1145/2908812.2908918, 2016.

185 Mather, J. H. and Voyles, J. W.: The ARM Climate Research Facility: a review of structure and
capabilities, *Meteorol. Monogr.*, 57, 1–24, doi:10.1175/AMSMONOGRAPHSD-15-0003.1, 2016.

Peng, K., Xin, J., Zhu, X., Wang, X., Cao, X., Ma, Y., and Wang, Z.: Machine learning model to
accurately estimate the planetary boundary layer height of Beijing urban area with ERA5 data, *Atmos.
Res.*, 293, 106925, doi:10.1016/j.atmosres.2023.106925, 2023.

190 Pino, D., Vilà-Guerau de Arellano, J., and Duynkerke, P. G.: The contribution of shear to the evolution
of a convective boundary layer, *J. Atmos. Sci.*, 60, 1913–1926, doi:10.1175/1520-
0469(2003)060<1913:TCOSTT>2.0.CO;2, 2003.

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

195 Schween, J. H., Hirsikko, A., Löhnert, U., and Crewell, S.: Mixing-layer height retrieval with
ceilometer and Doppler lidar: from case studies to long-term assessment, *Atmos. Meas. Tech.*, 7, 3685–
3704, doi:10.5194/amt-7-3685-2014, 2014.

200 Seber, G. A. F. and Lee, A. J.: *Linear regression analysis*, 2nd Edn., Wiley, doi:10.1002/0471725315,
2003.



Seibert, P., Beyrich, F., Gryning, S.-E., Joffre, S., Rasmussen, A., and Tercier, P.: Review and intercomparison of operational methods for the determination of the mixing height, *Atmos. Environ.*, 34, 1001–1027, doi:10.1016/S1352-2310(99)00449-0, 2000.

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

210 Shin, H. H., Xue, L., Li, W., Angevine, W. M., and Zhang, M. H.: Large-scale forcing impact on the development of shallow convective clouds revealed from LASSO large-eddy simulations, *J. Geophys. Res. Atmos.*, 126, e2021JD035208, doi:10.1029/2021JD035208, 2021.

215 Stapleton, A., Dias-Júnior, C. Q., von Randow, C., Farias D’Oliveira, F. A., Pöhlker, C., de Araújo, A. C., Sá, L. D., Santana, R. A., and Eichelmann, E.: Intercomparison of machine learning models to determine the planetary boundary layer height over Central Amazonia, *J. Geophys. Res. Atmos.*, 130, e2024JD042488, doi:10.1029/2024JD042488, 2025.

220 Stensrud, D. J.: *Parameterization Schemes: Keys to Understanding NWP Models*, Cambridge Univ. Press, Cambridge, UK, 2007.

225 Stevens, B.: Entrainment in stratocumulus-topped mixed layers, *Q. J. Roy. Meteorol. Soc.*, 128, 2663–2690, doi:10.1256/qj.01.202, 2002.

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

235 Sullivan, P. P., McWilliams, J. C., and Moeng, C.-H.: A grid nesting method for large-eddy simulation of planetary boundary-layer flows, *Bound.-Layer Meteorol.*, 84, 1–38, doi:10.1023/A:1000990124594, 1998.

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

240 Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I.: Attention is all you need, in: *Advances in Neural Information Processing Systems*, 30, 5998–6008, <https://papers.nips.cc/paper/7181-attention-is-all-you-need>, 2017.



245 Vogezelang, D. H. P. and Holtslag, A. A. M.: Evaluation and model impacts of alternative boundary-layer height formulations, *Boundary-Layer Meteorol.*, 81, 245–269, doi:10.1007/BF02430331, 1996.

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

255 Wei, Y., Sun, Y., Ma, Y., Tan, Y., Ren, X., Peng, K., and Xin, J.: Deviations of Boundary Layer Height and Meteorological Parameters Between Ground-Based Remote Sensing and ERA5 over the Complex Terrain of the Mongolian Plateau, *Remote Sens.*, 17, 393, doi:10.3390/rs17030393, 2025.

260 Wood, R., Bretherton, C. S.: On the relationship between stratiform low cloud cover and lower-tropospheric stability, *J. Clim.*, 19, 6425–6432, doi:10.1175/JCLI3988.1, 2006.

265 Wulfmeyer, V. and Turner, D.: Land-Atmosphere Feedback Experiment (LAFE) Science Plan, DOE Office of Science Atmospheric Radiation Measurement (ARM) Program, United States, DOE/SC-ARM-16-038, 2016.

270 Wulfmeyer, V. and Turner, D.: Land-Atmosphere Feedback Experiment Field Campaign Report, DOE Office of Science Atmospheric Radiation Measurement (ARM) Program, United States, DOE/SC-ARM-18-007, 2018.

275 Zhang, D., Comstock, J., Sivaraman, C., Mo, K., Krishnamurthy, R., Tian, J., and Roldán-Henao, N.: Best Estimate of the Planetary Boundary Layer Height from Multiple Remote Sensing Measurements, *EGUsphere*, 2025, 1–36, doi:10.5194/egusphere-2024-3959, 2025.

Zhang, G. J.: Roles of tropospheric and boundary layer forcing in the diurnal cycle of convection in the US southern great plains, *Geophys. Res. Lett.*, 30, 2281, doi:10.1029/2003GL018554, 2003.

280 Zhang, J., Zheng, Y., Li, Z., Xia, X., and Chen, H.: A 17-year climatology of temperature inversions above clouds over the ARM SGP site: The roles of cloud radiative effects, *Atmos. Res.*, 237, 104810, doi:10.1016/j.atmosres.2019.104810, 2020.

Zhong, X., Zhao, L., Ren, P., and Co-authors: A physics-guided automated machine learning approach for obtaining surface radiometric temperatures on sunny days based on UAV-derived images, *Comput. Environ. Urban Syst.*, 113, 102175, doi:10.1016/j.compenvurbsys.2024.102175, 2024.