https://doi.org/10.5194/egusphere-2026-1778
Preprint. Discussion started: 3 June 2026

A merged hurricane boundary layer height dataset in the western Atlantic based on dropsonde measurements and ERA5 reanalysis

Xiangyu Pi[1], Yuanjie Zhang[1], Xiaoze Xu[2], Yubin Li[1]

[1]Key Laboratory for Aerosol-Cloud-Precipitation of China Meteorological Administration, School of Atmospheric Physics, Nanjing University of Information Science & Technology, Nanjing, China;
[2]Shanghai Academy of Artificial Intelligence for Science, Shanghai, & 200232, China.

***Correspondence to*:** yuanjiez@nuist.edu.cn (Yuanjie Zhang)

Abstract. The hurricane boundary layer (HBL) critically regulates exchanges of heat, moisture, and momentum between the ocean and atmosphere. Accurate estimation of HBL height (HBLH) is essential for understanding hurricane dynamics. Dropsonde observations provide high accuracy but have limited spatiotemporal coverage, whereas ERA5 reanalysis offers continuous coverage with systematic biases. This study presents a merged HBLH dataset for 75 hurricanes over the western Atlantic during 2002–2024, generated by integrating 4438 dropsonde profiles, ERA5 reanalysis, and IBTrACS hurricane records through a Random Forest machine learning framework. Nineteen input variables representing thermodynamic, dynamic, and hurricane-specific parameters were used to predict the HBLH bias between dropsondes and ERA5. The corrected dataset retains the original ERA5 resolution (0.25° × 0.25°, 1-hourly) and significantly reduces systematic errors relative to dropsonde observations. Validation shows a correlation coefficient of 0.93 with dropsonde-derived HBLH, and reductions in MAE from 544 m to 159 m and RMSE from 661 m to 246 m. The merged HBLH reproduces the radial and asymmetric structure within hurricane domains more accurately than ERA5, while providing continuous temporal and spatial coverage suitable for further analysis of HBL dynamics under hurricane conditions. The dataset is publicly available at https://doi.org/10.5281/zenodo.17196964.

1 Introduction

Hurricanes (also called typhoon in the western Pacific Ocean) are multi-hazard events accompanied by extreme weather phenomena such as strong winds, heavy precipitation, and storm surges (Enderami et al., 2022). In recent decades, hurricanes have emerged as one of the most devastating natural disasters due to global warming and rising sea surface temperatures (Li et al., 2023), inflicting enormous economic losses and extensive damage to infrastructure worldwide, and causing hundreds of deaths each year (Klotzbach et al., 2018; Mendelsohn et al., 2012). Enhancing the accuracy of hurricane intensity and track predictions is crucial for providing earlier warnings of hurricane-related disasters and implementing effective mitigation strategies. As a result, studying hurricanes and their structural characteristics has consistently been a key area of focus (Lin et al., 2022).

https://doi.org/10.5194/egusphere-2026-1778
Preprint. Discussion started: 3 June 2026

The hurricane boundary layer (HBL) critically regulates heat, moisture, and momentum exchanges that drive storm intensification (Smith and Montgomery, 2010), making its height (HBLH) a key factor in hurricane dynamics (Nolan et al., 2009; Zhu et al., 2014). Numerical simulations demonstrate that HBLH parameterization schemes significantly influence hurricane intensity and track forecasts (Kepert, 2012; Smith and Thomsen, 2010). Current approaches to defining HBLH are generally classified as dynamic (e.g., inflow layer height, height of maximum tangential wind) or thermodynamic (e.g., maximum potential temperature gradient) (Zeng et al., 2004; Smith et al., 2009; Zhang et al., 2011, 2013; Ren et al., 2019, 2020b). However, single-variable definitions often fail under extreme hurricane conditions due to blurred profile signatures (Zhang et al., 2014). Thus, the bulk Richardson number (Rib) method, a hybrid approach that integrates both thermal and dynamic parameters (Vogelezang & Holtslag, 1996; Holtslag & Boville, 1993), has become a widely adopted alternative (Seidel et al., 2012). The Rib method has been effectively applied in observational and numerical studies of tropical cyclones, including the diagnosis of HBLH from dropsonde and aircraft measurements (Zhang et al., 2011; Liao et al., 2017), offering a physically comprehensive criterion that remains applicable under complex hurricane environments. More recently, Wang et al. (2023) further improved the performance of the Ri_b method in estimating HBLH by incorporating the sea surface friction term based on an air-sea drag relationship between sea surface friction velocity and 10-m wind speed (Andreas et al., 2012; Edson et al., 2013; Hsu et al., 2019).

Under extreme conditions like hurricanes, traditional radiosonde or airborne measurements are difficult to conduct. Consequently, GPS dropsonde observations providing high-resolution profiles of wind, temperature and humidity have become essential for studying hurricane structure (Wang et al., 2015; Liu et al., 2015). Such data enable reliable retrieval of HBLH using the Rib method (Zhang et al., 2011) and have been widely applied to analyze vertical wind shear, thermodynamic structure, convective intensity, and boundary layer dynamics under hurricane conditions (Zhang et al., 2013; Zhu et al., 2013; Liao et al., 2017; Wei et al., 2025). However, for individual hurricanes, the limited spatiotemporal coverage of dropsonde data hinders a comprehensive depiction of HBL structure and evolution. Although the composite analysis combining multiple hurricanes is often used to improve the spatial coverage (Zhang et al., 2011; Zhang and Uhlhorn, 2012), this approach tends to smooth detailed structural features (Zhang et al., 2013) and cannot capture the temporal evolution within single hurricanes. Compared with dropsonde observations, ERA5 reanalysis offers higher spatiotemporal continuity and thus can serve as an effective alternative or complementary dataset for hurricane research (Hersbach et al., 2020; Yu, 2020; Farjami et al., 2020; Xiong et al.,2022). While ERA5 generally performs well in representing boundary layer structures under conventional weather conditions (Guo et al., 2021), numerous studies have reported systematic biases in ERA5 boundary layer height (BLH) product relative to site observations (e.g., Zhang et al., 2020; Guo et al., 2024; Xi et al., 2024). Under hurricane conditions, specifically, ERA5 tends to overestimate lower HBLHs in the eye and outer regions but notably underestimates higher HBLHs near the radius of maximum wind (RMW), compared to dropsonde observations (Wang et al., 2023).

In summary, dropsonde observations provide high accuracy in describing the HBLH but suffer from limited spatiotemporal coverage, whereas ERA5 reanalysis offers continuous and extensive coverage but contains systematic biases. Therefore,

https://doi.org/10.5194/egusphere-2026-1778
Preprint. Discussion started: 3 June 2026

integrating the strengths of both datasets to develop a merged HBLH dataset is feasible and meaningful, which constitutes the main objective of this study. In recent year, machine learning techniques have emerged as a powerful approach for data fusion, by enabling the integration of strengths from multiple datasets (Cai et al., 2020; J. Liu et al., 2021; de Arruda Moreira et al., 2022; Jiang and Zhao, 2022; Molero et al., 2022). For instance, Guo et al. (2024) developed a global BLH dataset by training a random forest model to predict biases between radiosonde and ERA5 BLH using ERA5 and GLDAS inputs, achieving improved accuracy and resolution. Similarly, Wei et al. (2025) applied an XGBoost model with dropsonde data to correct ERA5 temperature and humidity in tropical cyclones, reducing errors by approximately 60%. These results demonstrate the strong potential of machine learning for correcting and merging HBLHs based on dropsonde and ERA5 data. This study integrates dropsonde data from 72 hurricanes during 2002–2023, along with ERA5 reanalysis and best-track data, to investigate HBL characteristics under tropical cyclone conditions. A Rib method that accounts for sea surface friction velocity is employed to estimate the HBLH. The relationships between HBLH bias and a suite of meteorological factors are first examined, based on which a Random Forest (RF) model is developed to predict and correct the bias. The proposed framework ultimately produces an integrated, hourly HBLH dataset for the western Atlantic hurricanes during 2002-2024, covering a radial extent within 4 RMW. This dataset provides robust data support for analyzing HBL structure and offers valuable resources for tracking and investigating HBL structural features and hurricane evolution.

2 Data and methods

2.1. Dropsonde Data

The dropsonde data utilized in this study were obtained from the Hurricane Research Division (HRD) of NOAA (ftp://ftp.aoml.noaa.gov/hrd/pub/data/dropsonde), which were collected during hurricane reconnaissance missions conducted over the western Atlantic from 2002 to 2024. The data format we utilized is Full-Resolution Data (FRD) files, which contain observations recorded at a frequency of 4 Hz (0.25s). Prior to 2008, the sampling frequency was 2 Hz (0.5 s). The vertical resolution ranges from 5 to 15 m depending on the descent rate (Wang et al., 2015). These data underwent rigorous quality control before being released (Ren et al., 2019), and provide meteorological elements such as wind speed, temperature, pressure, and relative humidity (Wang et al., 2015). The precision of the measurements is ±1.0 hPa for pressure, ±0.2°C for temperature, ±5% for relative humidity, and ±0.5 m/s for horizontal wind speed (Franklin et al. 2003).

To ensure that the dropsonde data properly capture the boundary layer structure under hurricane conditions, we selected only dropsonde observations released within four times the RMW from the hurricane center, when the maximum sustained wind speed (MSW) exceeded 64 kt (Zhang et al., 2011, Wang et al., 2023), corresponding to hurricane strength according to the Saffir–Simpson Hurricane Wind Scale. To ensure that the dropsonde profiles adequately captured the HBL, we required the highest observation point of the dropsonde profile to exceed 1000 m and the lowest point to be below 200 m. In addition, profiles with continuous missing data exceeding 200 m in the vertical were excluded. Based on these criteria, 4438

https://doi.org/10.5194/egusphere-2026-1778
Preprint. Discussion started: 3 June 2026

dropsonde profiles from 75 hurricanes spanning 2002 to 2024 were selected, with their locations indicated by blue dots in Fig. 1. The numbers of selected dropsonde observations for each hurricane are also listed in Table 1. Because dropsonde observations are characterized by uneven vertical sampling due to variations in descent rate and flight altitude, the selected profiles were interpolated linearly onto uniform vertical levels at 10 m intervals. The 10 m interval choice is reasonable given that the vertical resolution of most dropsonde observations is approximately 6-7 m (Zhang et al., 2011). The interpolation was performed from the sea surface up to a maximum altitude of 5000 m. It should be noted that for the bias analysis and machine-learning model training described below, only the selected dropsonde observations from 2002–2023 were used. The 2024 dropsonde data were retained as an independent dataset for validation.

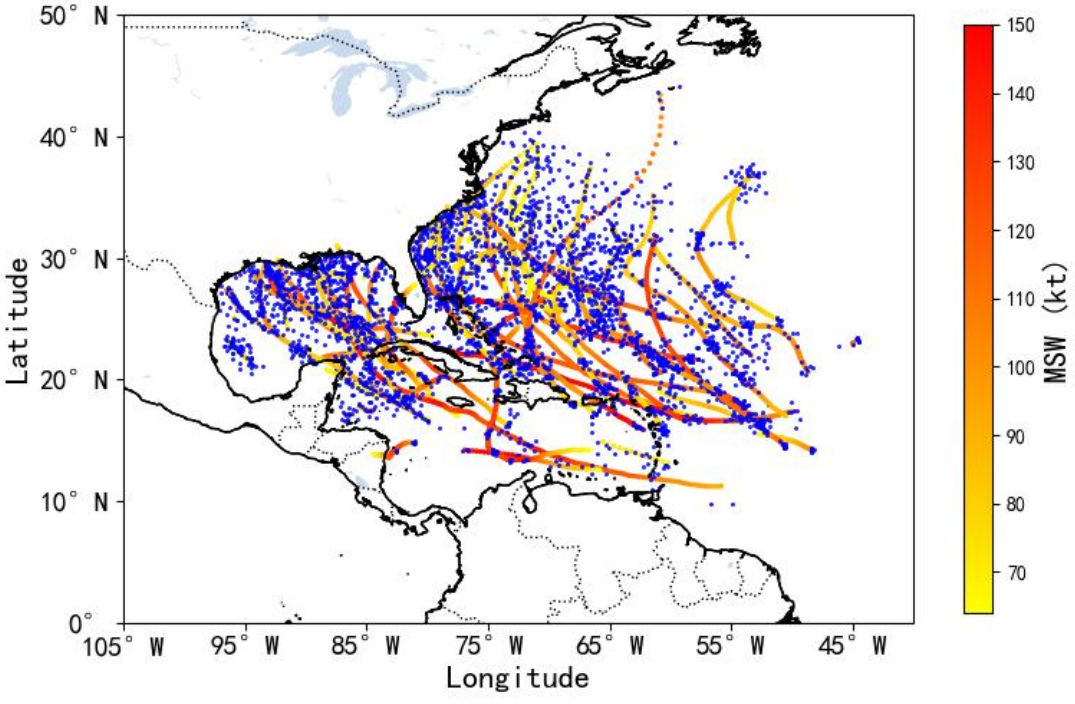


Figure 1. Distribution of dropsonde observations (blue dots) and hurricane tracks (warm-colored dots) during the period of 2002-2024 over the western Atlantic. The color of tracks indicates the hurricane MSW.

2.2 IBTrACS Dataset

The best-track dataset of hurricanes is sourced from the International Best Track Archive for Climate Stewardship (IBTrACS; Knapp et al., 2018; Gahtan et al., 2024), with a temporal resolution of 3 hours. The dataset contains hurricane information such as the MSW in kt, the RMW in nautical miles (nmi) and the hurricane center location (latitude and longitude). These

https://doi.org/10.5194/egusphere-2026-1778
Preprint. Discussion started: 3 June 2026

best-track data were first linearly interpolated in time to an hourly temporal resolution, in order to match dropsonde observations within four times the RMW from the hurricane center at the same hour. Besides, the absolute distance (r), relative distance (r/RMW) and azimuth (θ) of each dropsonde relative to the hurricane center were also calculated from the location information of dropsondes and hurricane centers. The absolute distance refers to the actual distance between the dropsonde and the hurricane center, while the relative distance represents the distance normalized by the RMW. The azimuth was determined following the method in Wang et al., (2023). Thus, a total of 5 parameters were obtained from IBTrACS data to represent the hurricane intensity and the relative location of the dropsonde (Table 2).

Table 1. Hurricane information and numbers of selected dropsondes during 2002-2024.

ID	Hurricane name	Dropsonde number	ID	Hurricane name	Dropsonde number
1	2002-ISIDORE	30	39	2016-MATTHEW	126
2	2002-LILI	26	40	2017-HARVEY	55
3	2003-FABIAN	29	41	2017-IRMA	106
4	2003-ISABEL	108	42	2017-JOSE	17
5	2004-CHARLEY	3	43	2017-MARIA	113
6	2004-FRANCES	60	44	2017-NATE	28
7	2004-IVAN	72	45	2018-FLORENCE	72
8	2004-JEANNE	22	46	2018-MICHAEL	89
9	2005-DENNIS	42	47	2019-DORIAN	290
10	2005-EMILY	8	48	2019-HUMBERTO	34
11	2005-OPHELIA	15	49	2019-JERRY	16
12	2005-RITA	95	50	2019-LORENZO	10
13	2005-WILMA	12	51	2020-DELTA	131
14	2006-HELENE	35	52	2020-ETA	16
15	2007-FELIX	13	53	2020-HANNA	5
16	2008-DOLLY	10	54	2020-ISAIAS	71
17	2008-GUSTAV	45	55	2020-LAURA	123
18	2008-IKE	77	56	2020-PAULETTE	27
19	2008-KYLE	7	57	2020-SALLY	50
20	2008-PALOMA	14	58	2020-TEDDY	118
21	2009-BILL	42	59	2020-ZETA	62
22	2010-EARL	68	60	2021-ELSA	16
23	2010-TOMAS	5	61	2021-GRACE	20
24	2011-IRENE	87	62	2021-HENRI	43
25	2011-KATIA	32	63	2021-LARRY	96
26	2011-RINA	10	64	2021-SAM	107
27	2012-ISAAC	66	65	2022-ERAL	123
28	2012-SANDY	148	66	2022-FIONA	162
29	2013-INGRID	54	67	2022-LAN	108
30	2014-ARTHUR	61	68	2022-LISA	7
31	2014-BERTHA	29	69	2023-IDALIA	47
32	2014-CRISTOBAL	57	70	2023-LEE	390
33	2014-EDOUARD	76	71	2023-NIGEL	33
34	2015-DANNY	11	72	2023-TAMMY	55
35	2015-JOAQUIN	53	73	2024-Beryl	41
36	2015-KATE	1	74	2024-Debby	5
37	2016-EARL	10	75	2024-Ernesto	82
38	2016-HERMINE	11			

https://doi.org/10.5194/egusphere-2026-1778
Preprint. Discussion started: 3 June 2026

Table 2. Summary of input parameters of machine learning algorithms and their correlation coefficients with the bias between ERA5 and dropsonde HBLHs.

	Parameter	Acronym (Unit)	Data source
1	Surface Pressure	SP (hPa)	ERA5
2	Skin Temperature	SKT (K)	ERA5
3	Latent heat flux	LE (w/m^2)	ERA5
4	Sensible heat flux	H (w/m^2)	ERA5
5	Friction Velocity	u_* (m/s)	ERA5
6	2m Temperature	T_{2m} (K)	ERA5
7	2-m Dew Temperature	D_{2m} (K)	ERA5
8	10-m Wind Speed	WS_{10m} (m/s)	ERA5
9	Lower Tropospheric Stability	LTS (K)	ERA5
10	Boundary layer Specific Humidity	q_{BL} (g/kg)	ERA5
11	Boundary layer Temperature	T_{BL} (K)	ERA5
12	Boundary layer zonal wind	u_{BL} (m/s)	ERA5
13	Boundary layer meridional wind	v_{BL} (m/s)	ERA5
14	Boundary Layer Height	BLH (m)	ERA5
15	Maximum Sustained Wind	MSW (kt)	IBTrACS
16	Radius of Maximum Wind	RMW (nmi)	IBTrACS
17	Absolute distance	r (nmi)	IBTrACS
18	Relative distance	r/RMW	IBTrACS
19	Azimuth	θ (°)	IBTrACS

2.3 ERA5 Reanalysis Data

ERA5 is one of the most widely used reanalysis datasets, with updated parameterization schemes, improved data assimilation systems, and enhanced spatial and temporal resolutions compared to previous versions of ECMWF reanalysis (e.g., Xi et al., 2024; Guo et al., 2021; Hersbach et al., 2020). It provides a temporal resolution of one hour and a spatial resolution of 0.25° × 0.25°. The ERA5 BLHs were extracted for grids located within four times the RMW under hurricane conditions, according to the best-track data. Correspondingly, the sea surface and near-surface parameters, including surface pressure (*SP*), skin temperature (*SKT*), 2-m dew point temperature (D_{2m}), 2-m temperature (T_{2m}), surface latent heat flux (*LE*), surface sensible heat flux (*H*), friction velocity (u_*) and 10-m wind speed (WS_{10m}), were also taken from ERA5 single-level data. In addition, the boundary layer parameters were derived from the average values of ERA5 data at four pressure levels: 1000 hPa, 925 hPa, 850 hPa, and 700 hPa, including wind components (u_{BL} and v_{BL}), specific humidity (q_{BL}), temperature (T_{BL}) and the lower tropospheric stability (*LTS*), to characterize HBL dynamics and thermodynamics. Here, the LST is defined as the potential temperature difference between 700 hPa and 1000 hPa, following Guo et al. (2016). In total, 14 parameters were extracted from ERA5 data (Table 2).

https://doi.org/10.5194/egusphere-2026-1778
Preprint. Discussion started: 3 June 2026

3 Methodology

3.1 Determination of HBLH using the bulk Richardson number method

The Rib method is widely used to compute the BLH due to its applicability in various atmospheric stable conditions (Seidel et al., 2012). The Ri_b quantifies the turbulence and thermodynamic stability, and is defined as the ratio of contributions of buoyancy and shear forces to turbulence (Vogelezang and Holtslag, 1996), expressed as:

$$Ri_b = \frac{(g/\theta_{vs})(\theta_{vz}-\theta_{vs})(z-z_s)}{(u_z-u_s)^2+(v_z-v_s)^2+bu_*^2}, \quad (1)$$

where θ_{vz} and θ_{vs} are the virtual potential temperatures at height z and z_s, respectively, and g/θ_{vs} is the buoyancy parameter. u and v are wind components. z_s is the height of the lower boundary for the BL. bu_*^2 represents the turbulence production due to surface friction, which is particularly non-negligible under near-neutral conditions, with its empirical parameter b typically set to 100 (Vogelezang and Holtslag, 1996; Seibert, 2000). Recent studies also highlighted the importance of the surface friction term, especially under hurricane conditions (Zhang et al., 2020; Wang et al., 2023). The $\mathrm{Ri_b}$ method assumes the BLH is the height at which the Ri_b first reaches a threshold value, known as the critical Ri_b (Ri_bc). The most commonly used Ri_bc values are 0.25 and 0.5 (e.g., Holtslag and Boville, 1993; Hong, 2010). According to the sensitivity analysis of the Ri_b method under hurricane conditions in Wang et al. (2023), the parameters setting of Ri_bc=0.25, b=100 and z_s=80 m are most suitable for calculating HBLHs from dropsonde observations, which are also adopted in this study. It is worth noting that the BLH in ERA5 is also computed using the Rib method with Ri_bc=0.25 but without accounting for surface friction (i.e., b=0) (Seidel et al., 2012).

Owing to the lack of sea surface friction velocity measurements corresponding to the dropsonde observations, u_* was estimated from the 10-m wind speed using an air-sea drag relationship for aerodynamic rough flow (Edson et al., 2013; Hsu et al., 2019), as follows:

$$u_* = 0.062U_{10} - 0.28, \quad (2)$$

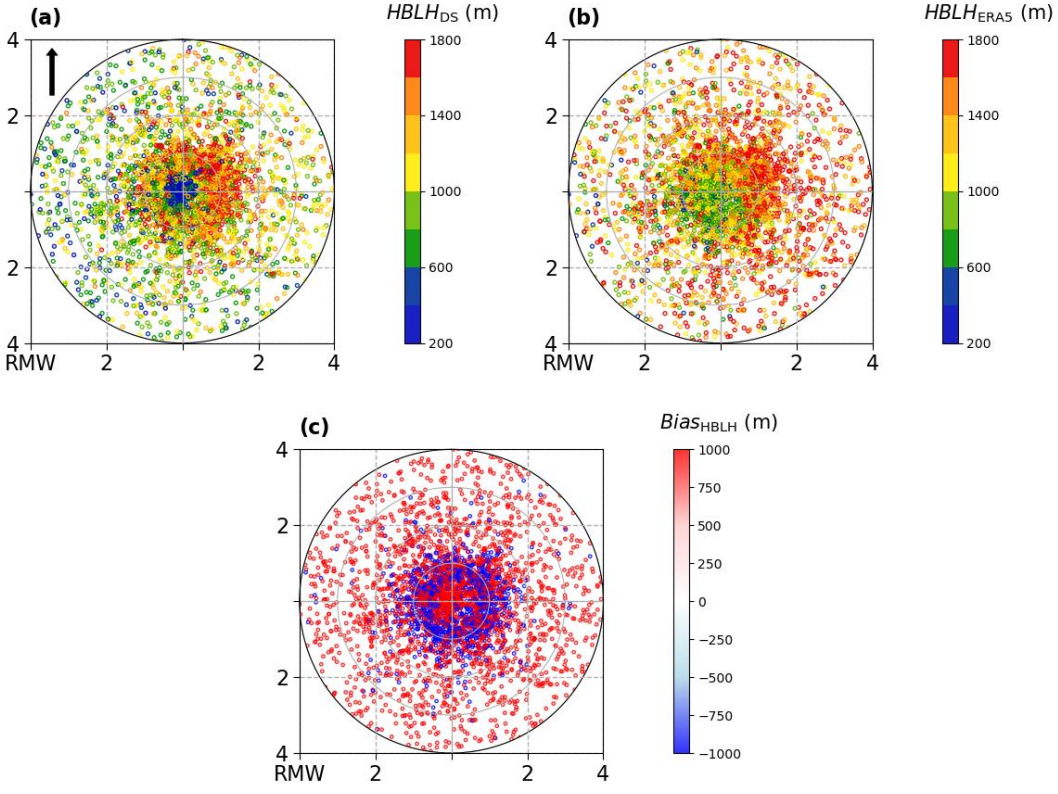
where U_{10} is the 10-m wind speed. Eq. (2) is valid for U_{10} values in the range of 9–70 m/s (Hsu et al., 2019). With the Rib method, HBLHs calculated with u_* derived from Eq. (2) are close to those with $u_*$from the near-neutral logarithmic law, but exhibit clear advantages over the latter in certain cases (Wang et al., 2023). Furthermore, the relationship in Eq. (2) is valid for a wide range of U_{10} (9-70 m/s, Hsu et al., 2019), making it applicable to high-wind conditions associated with hurricanes.

3.2 HBLH biases between ERA5 and dropsonde observations

Figure 2 compares HBLHs from dropsonde composites during 2002-2023 with the corresponding ERA5 BLHs, with the radial distance and azimuth of each dropsonde normalized by the corresponding RMW and storm motion direction, respectively. Hereafter, the dropsonde-derived HBLH, the ERA5 BLH are referred to as $HBLH_\mathrm{DS}$, and $HBLH_\mathrm{ERA5}$, respectively. The $HBLH_\mathrm{DS}$ exhibits pronounced spatial variability, characterized by lower values in the eye region (0–

https://doi.org/10.5194/egusphere-2026-1778
Preprint. Discussion started: 3 June 2026

165 0.5RMW) and higher values in high-wind zone (0.5–1.5RMW), along with a clear left–low-right-high asymmetry, which are consistent with previous observational findings (Zhang et al 2011, 2013; Wang et al., 2023). Comparatively, the $HBLH_{\mathrm{ERA5}}$ shows a broadly similar spatial pattern, including the asymmetric feature, but with substantially weaker spatial variability. Overall, ERA5 tends to overestimate HBLHs across most regions, particularly within the eye region, while underestimate HBLHs in the high-wind zone, compared to dropsonde observations. These results indicate that although ERA5 data can
170 reasonably capture the spatial distribution characteristics of HBLHs, confirming its applicability under hurricane conditions (Farjami et al., 2020; Xiong et al., 2022), it still exhibits non-negligible biases that necessitate correction. Furthermore, the HBLH biases are strongly associated with the dropsonde position relative to the hurricane center, showing predominantly negative values in the high-wind zone and positive values in the eye and other regions (Fig. 2c). This suggests that the dropsonde location-related variables, such as the r, r/RMW and θ, are likely important determinants of the HBLH bias.



175

Figure 2. 2D distribution of the selected dropsonde observations. The radial distance to the hurricane center is normalized by the RMW and the azimuth is relative to the hurricane motion direction (black arrow). Scatter colors indicate (a) HBLHs estimated from dropsonde profiles, (b) corresponding ERA5 HBLHs, and (c) ERA5 HBLH biases compared to dropsonde observations ($Bias_{\mathrm{HBLH}}$, ERA5 minus dropsonde)

180 The systematic overestimation of ERA5 BLH has been well documented (e.g., Zhang et al., 2020; Wang et al., 2023; Xi et al., 2024). Previous studies have shown that HBLH biases are closely linked to ERA5 inaccuracies in near-surface wind

https://doi.org/10.5194/egusphere-2026-1778
Preprint. Discussion started: 3 June 2026

speed (Graham et al., 2019; Bian et al.,2021; Wang et al., 2023), suggesting that the HBLH bias is influenced by hurricane intensity (MSW and RMW) and by dynamical conditions at the sea surface and in the lower troposphere. In addition, the supply of heat and moisture from the sea surface is a fundamental driver of hurricane and HBL development (Wei et al., 2025), implying that thermodynamic conditions may also shape the magnitude of the HBLH bias. ERA5 consistently underestimates the LTS, which likely also contributes to the overestimation of $HBLH_{ERA5}$ (Wang et al., 2023), since reduced LTS generally facilitates stronger PBL development (Slingo, 1987). Taken together, these results indicate that the HBLH bias maybe jointly influenced by thermodynamic and dynamic parameters at the sea surface, in the near-surface layer, and within the HBL, as well as the hurricane own intensity.

To further examine the relationships of the HBLH bias with these meteorological factors and hurricane characteristics, Fig. 3 presents variations of HBLH biases with each parameter. These parameters were broadly grouped into three categories: thermodynamic factors (Fig. 3a-i), dynamic factors (Fig. 3j-l), and hurricane-specific factors (Fig. 3m-o). All samples were first divided into ten bins with equal sample size. The binned correlation coefficients, as shown in figures, were then calculated using the mean HBLH bias and the mean parameter value within each bin. The results show that, among the thermodynamic factors, q_{BL}, T_{BL}, LTS, and SKT exhibit significantly negative correlations with the HBLH bias, with correlation coefficients ranging from –0.67 to –0.91. H and D_{2m} also display weak negative correlations, whereas SP and LE are positively associated with the bias. For the dynamical factors, u_* demonstrates a pronounced positive correlation with the HBLH bias, underscoring the critical role of surface friction in governing HBL development (e.g., Wang et al., 2023). The wind speed at 10 m (WS_{10m}) and lower tropospheric levels (WS_{BL}) show modest positive correlations. Regarding the hurricane-specific factors, MSW is significantly negatively correlated with the HBLH bias, with the correlation coefficient as -0.92, while the dropsonde azimuth, θ, exhibits a weak positive correlation.

https://doi.org/10.5194/egusphere-2026-1778
Preprint. Discussion started: 3 June 2026

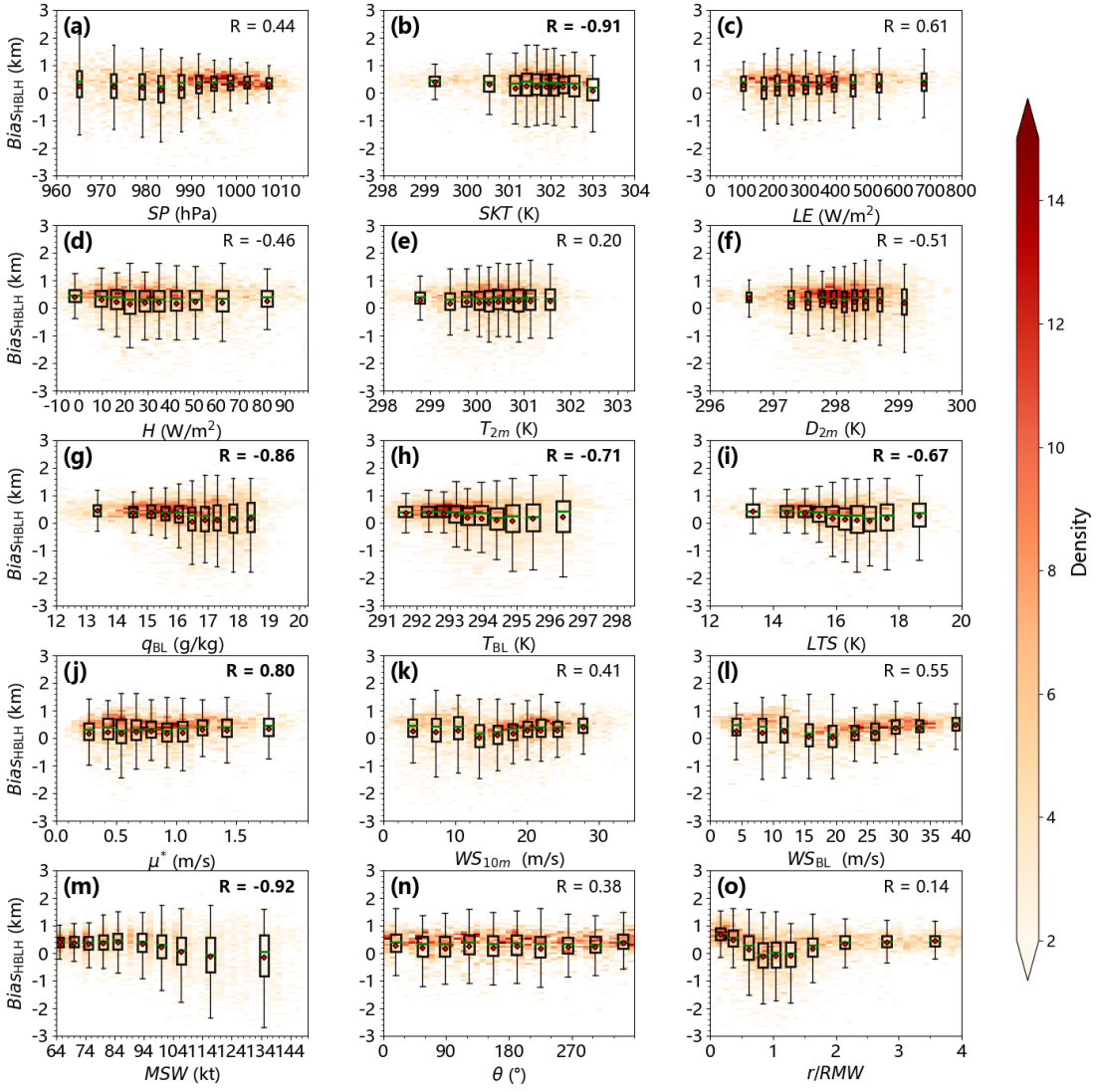


Figure 3. The 2D histogram distribution of the HBLH bias with respect to each feature variable. The boxplots are based on ten equal-sized intervals of data, where the red dots represent the mean of the box, and the green short lines indicate the median within the box. The correlation coefficient, calculated based on the median of the boxplot, is labeled in the upper-right corner, with bold text indicating significance ($p < 0.05$) based on the significance test.

It is worth noting that the correlation analysis above only captures the linear relationships between the HBLH bias and individual variables. For several variables, such as WS_{10m} and r/RMW, the HBLH bias exhibits clear fluctuations as their values increase. Although these variables show weak or insignificant linear correlations with the HBLH bias, their nonlinear relationships should not be overlooked. In addition, the large variability of HBLH bias within each bin, as indicated by the error bars, introduces substantial uncertainties that limit the representativeness of the binned correlation coefficients. Considering that machine learning is well suited for identifying nonlinear relationship and handling uncertainties that

https://doi.org/10.5194/egusphere-2026-1778
Preprint. Discussion started: 3 June 2026

conventional statistics struggle with, all variables are included as feature predictors in the subsequent model training, even if some of them display weak linear correlations.

To account for the potential influence of wind direction on the HBLH bias, the boundary-layer wind speed, WS_{BL}, was decomposed into its zonal and meridional components (u_{BL} and v_{BL}) and treated as two separate predictors. For the factor *r/RMW*, both the numerator and denominator were also included as independent predictors. In addition, the relatively high $HBLH_{\mathrm{ERA5}}$ in the high-wind region tends to produce negative HBLH biases, whereas the lower $HBLH_{\mathrm{ERA5}}$ in other areas, particularly within the eye region, generally results in positive biases (Fig. 2). Therefore, $HBLH_{\mathrm{ERA5}}$ itself (denoted as BLH) was also incorporated as one of the predictors. In total, 19 variables were selected as model predictive features, as listed in Table 2.

Unlike previous studies that rely on multisource datasets (e.g., Guo et al., 2024), all meteorological parameters used in this study are derived exclusively from ERA5. This choice is motivated by three considerations: 1) ERA5 provides a complete set of variables required for this study with higher temporal resolution; 2) developing the model using a single, consistent data source facilitates long-term and continuous bias correction in future applications; 3) the systematic biases inherent in ERA5 variables and their relationships with the HBLH bias are more stable under a unified numerical model and data assimilation system, which is expected to improve the efficiency of model training and learning.

3.3 Development of a predictive model

To correct the ERA5 BLH product and generate a temporally and spatially continuous HBLH dataset under hurricane conditions, the HBLH bias ($HBLH_{\mathrm{ERA5}} - HBLH_{\mathrm{DS}}$) was used as the prediction target for the machine learning model. By explicitly modeling this bias, the correction framework aims to bridge the systematic discrepancies between reanalysis data and dropsonde observations while preserving the large-scale consistency of ERA5. Once the HBLH bias is obtained from the prediction model, a bias-corrected HBLH dataset, referred to as the merged HBLH product ($HBLH_{\mathrm{merged}}$), can be constructed by removing the predicted bias from the $HBLH_{\mathrm{ERA5}}$, which can be formulated as:

$$HBLH_{\mathrm{merged}} = HBLH_{\mathrm{ERA5}} - Bias_{\mathrm{HBLH}}\,, \tag{3}$$

Several machine learning algorithms, including Decision Tree model, Gradient Boosting decision trees, Support Vector machines and Random Forest, were evaluated for their suitability in predicting HBLH biases. Informed by these comparisons and guided by previous studies (Guo et al., 2022; Xi et al., 2024), the RF model was selected due to its superior overall performance and robustness for this task. The RF framework was therefore adopted to predict the HBLH bias ($Bias_{\mathrm{HBLH}}$), expressed as:

$$Bias_{\mathrm{HBLH}} = f_{\mathrm{RF}}\,(\boldsymbol{X};\,\varepsilon)\,, \tag{4}$$

where X denotes the set of input predictors, including thermodynamic, dynamic, and hurricane-specific parameters (Table 2), and ε represents the optimized hyperparameters of the RF model.

https://doi.org/10.5194/egusphere-2026-1778
Preprint. Discussion started: 3 June 2026

The dataset that contains the $Bias_{\mathrm{HBLH}}$ and the input array from 2002-2023 is randomly divided into training and testing subsets using a 7:3 ratio. Within the training subset, a 10-fold cross-validation strategy was applied to assess the generalization capability of the model, while the remaining 30% of the data were reserved as an independent test set for evaluating the final model performance. The RF model was implemented using the scikit-learn library (version 1.0.2). The number of trees was set to 651, and other parameters were kept at their default values. Hyperparameter optimization was performed using Bayesian black-box optimization, which efficiently explores the parameter space while reducing the risk of overfitting. The final hyperparameter configurations adopted for each HBLH interval are summarized in Table 3.

Table 3. Random Forest Parameter Settings.

Hyper-Parameters	Description	Final Values
Max_depth	Maximum depth of each decision tree	100
Max_features	Fraction of features considered at each split	0.7
Min_samples_leaf	Minimum number of samples per leaf node	2

The relative importance of individual predictors in the RF model is illustrated in Fig. 4. It reveals that hurricane-related parameters, especially MSW and RMW, are the most influential predictors of the HBLH Bias. Notably, the $HBLH_{\mathrm{ERA5}}$ itself contributes approximately 14% to the model, highlighting its role in anchoring the correction to the original reanalysis data. Overall, hurricane-related parameters contribute the largest share to the model, followed by dynamic parameters, while thermodynamic parameters exert the weakest influence. Dynamic predictors generally have a greater impact than thermodynamic ones, consistent with Wang et al. (2023), who found that the HBL development is more strongly controlled by dynamical processes than by thermodynamic conditions.

To further assess the robustness of predictor selection and evaluate the necessity of individual variables, a series of feature ablation experiments was conducted (results not shown). The results indicate that the model trained with all predictors achieves the best predictive performance. Model skill is primarily dominated by high-importance predictors, while removing low-importance predictors leads to only a slight degradation in performance. This demonstrates the overall stability of the RF-based correction framework. In this study, the full-predictor model is adopted to obtain the most accurate corrected data.

https://doi.org/10.5194/egusphere-2026-1778
Preprint. Discussion started: 3 June 2026

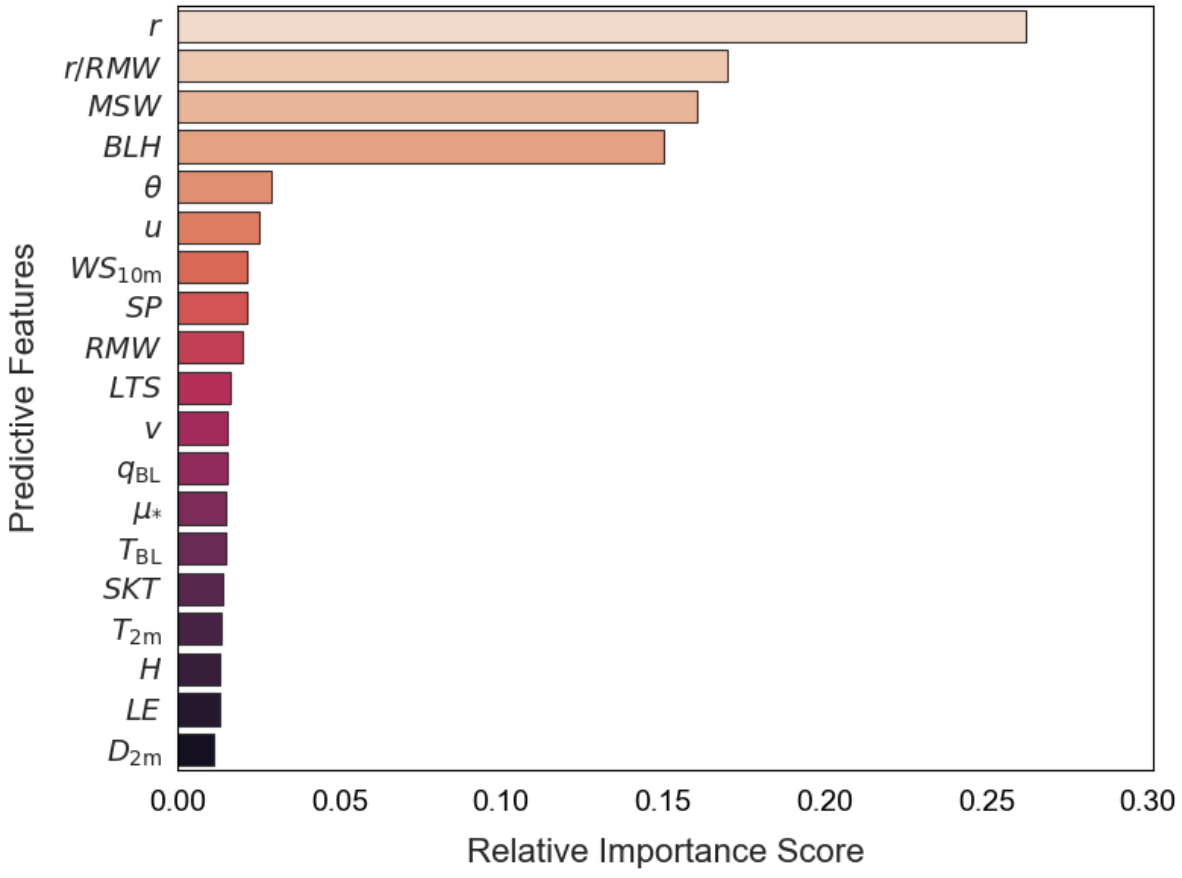


Figure 4. Feature importance of the Random Forest model.

Table 4 summarizes the performance metrics of the RF model on both the training and testing datasets, including the coefficient of determination (R^2), root mean squared error (RMSE), and mean absolute error (MAE). While the model achieves strong performance on the training set (R^2=0.94, RMSE=148 m, MAE=94 m), a moderate reduction in performance is observed on the testing set (R^2=0.74, RMSE= 305 m, MAE=209 m), suggesting a certain degree of overfitting. Such behavior can be primarily attributed to the characteristics of both the dataset and the target variable. First, the available training dataset remains relatively small compared with the substantial variability and structural complexity of HBLs (Table 1), which inherently constrains the model generalization capability. Under these conditions, some divergence between training and testing performance is expected. Second, as illustrated in Fig. 2c, the systematic bias in HBLH is considerable, indicating a relatively high level of intrinsic uncertainty associated with the target variable itself. This inherent uncertainty limits the achievable prediction accuracy and further contributes to the observed performance gap.

Despite these factors, the overall performance of the model remains robust and practically meaningful. The RF model retains substantial explanatory power and provides a clear improvement over the uncorrected ERA5 HBLH product. Moreover, the achieved error level is comparable to that reported in related studies on BLH retrieval over land (Guo et al., 2024). Therefore,

https://doi.org/10.5194/egusphere-2026-1778
Preprint. Discussion started: 3 June 2026

considering the data limitations and the intrinsic uncertainty of the problem, the proposed model demonstrates satisfactory reliability and is well suited for constructing the corrected HBLH dataset.

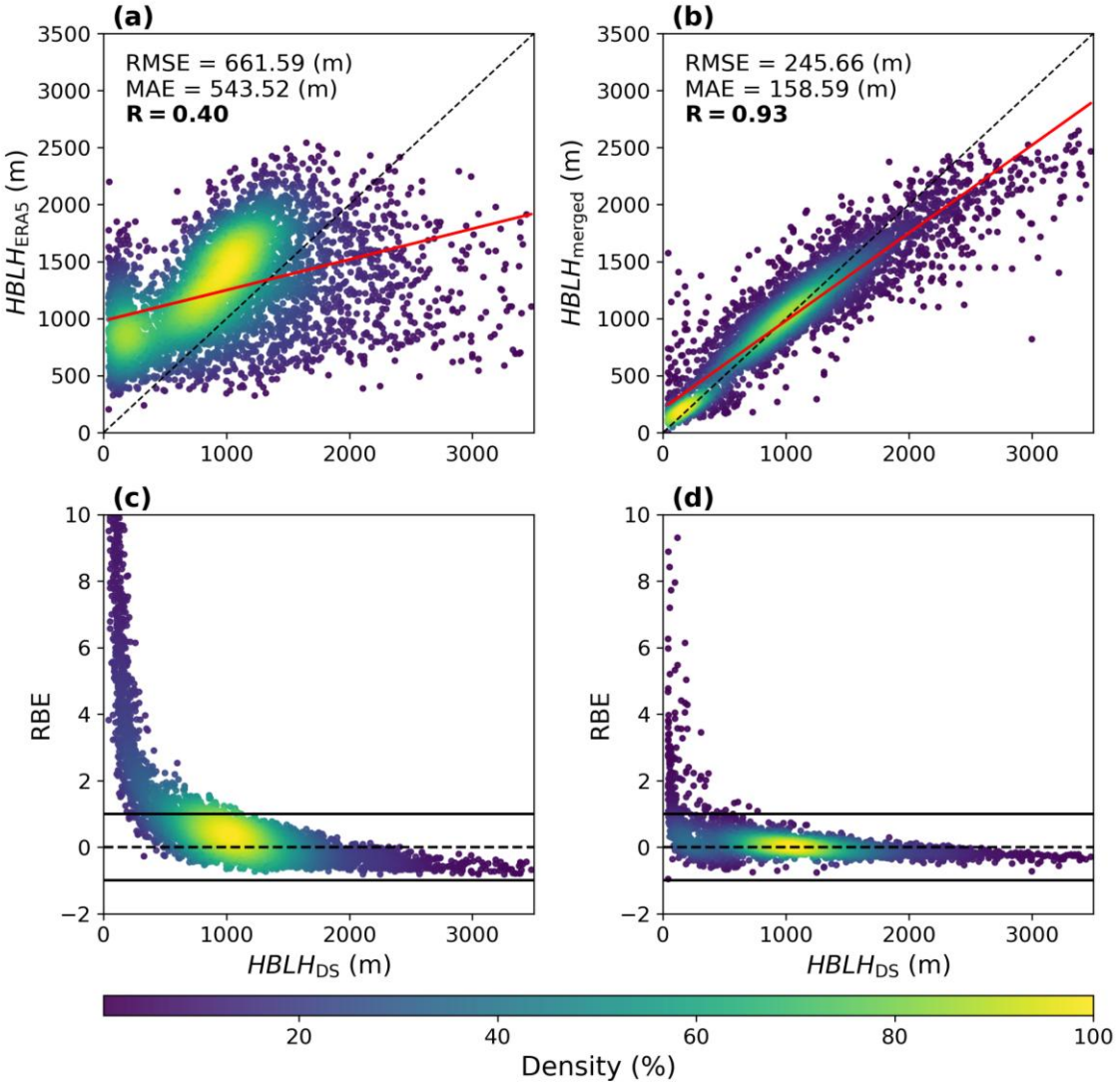
Table 4. Evaluation of Random Forest Model with coefficient of determination (R^2); root mean squared error (RMSE); mean absolute error (MAE).

Data Set	R^2	RMSE	MAE
Train	0.94	147.71	94.42
Test	0.74	305.17	208.60

4 Evaluation of the merged HBLH dataset

After applying the RF-based correction and constructing the merged HBLH dataset, this section systematically evaluates its performance through comparing the biases of $HBLH_{ERA5}$ before and after correction relative to dropsonde observations. Figure 5 presents comparisons of $HBLH_{DS}$ from 2002 to 2023 with the corresponding original $HBLH_{ERA5}$ and $HBLH_{merged}$. The results show that, before correction, ERA5 generally overestimates HBLH, while exhibiting a pronounced underestimation at higher HBLH values, particularly when HBLH exceeds 2000 m (Fig. 5a-b). Statistically, the correlation coefficient (R) between ERA5 and dropsonde observations is only 0.40, accompanied by MAE of 544 m and a RMSE as large as 661 m. In contrast, after correction, biases of $HBLH_{merged}$ relative to dropsonde observations are substantially reduced, demonstrating a high level of consistency. The R increases to 0.93, while the MAE and RMSE decrease to 159 m and 246 m, respectively. In terms of the relative bias error (RBE) as shown in Fig. 5c-d, ERA5 shows pronounced order-of-magnitude mismatches for HBLHs lower than 500 m, with RBEs exceeding 10, indicating a failure to represent low HBLs, especially in the eye region. Comparatively, the $HBLH_{merged}$ shows a much narrower RBE distribution centered near zero, with extreme deviations largely eliminated and most values confined to the range of −1 to 1. This substantial improvement confirms the effectiveness of the RF-based bias prediction model in correcting $HBLH_{ERA5}$ and highlights its reliability for constructing a merged HBLH dataset.

https://doi.org/10.5194/egusphere-2026-1778
Preprint. Discussion started: 3 June 2026



300 **Figure 5.** Comparison of HBLH before and after correction against dropsonde observations. Panels (a) and (b) show scatter density plots of $HBLH_{ERA5}$ and the $HBLH_{merged}$ against $HBLH_{DS}$. Panels (c) and (d) present the corresponding relative biases, defined as ($HBLH_{ERA5}$-$HBLH_{DS}$)/$HBLH_{DS}$ and ($HBLH_{merged}$ -$HBLH_{DS}$)/$HBLH_{DS}$. The color gradient in all panels represents the normalized density of data points.

Given the asymmetric distribution and radial dependence of HBLHs as shown in Fig. 2, Fig. 6 further compares the radial variations of quadrant-mean $HBLH_{DS}$ with those of the corresponding $HBLH_{ERA5}$ and $HBLH_{merged}$, based on a composite
305 analysis within four times the RMW. Dropsonde observations show that HBLH is lowest in the eye region (0-0.5 RMW), peaks in the high-wind region (0.5-1.5 RMW), and then gradually decreases with increasing radial distance. In contrast, $HBLH_{ERA5}$ exhibits a much weaker radial variation, characterized by a gradual increase with radius, indicating its inability to capture the observed radial structure of HBLH. Correspondingly, its biases show an opposite radial pattern to the observed

https://doi.org/10.5194/egusphere-2026-1778
Preprint. Discussion started: 3 June 2026

HBLH, with pronounced positive biases in the eye and outer regions and negative biases in the high-wind region. After correction, $HBLH_{merged}$ shows good agreement with observations and successfully reproduces the radial variation of HBLH. Specifically, it closely matches observations in the LB quadrant, with only a slight underestimation persisting near the high-wind region in the right quadrants. A comparison of biases before and after correction indicates that the RF-based correction primarily reduces the positive HBLH biases in the eye and outer regions.

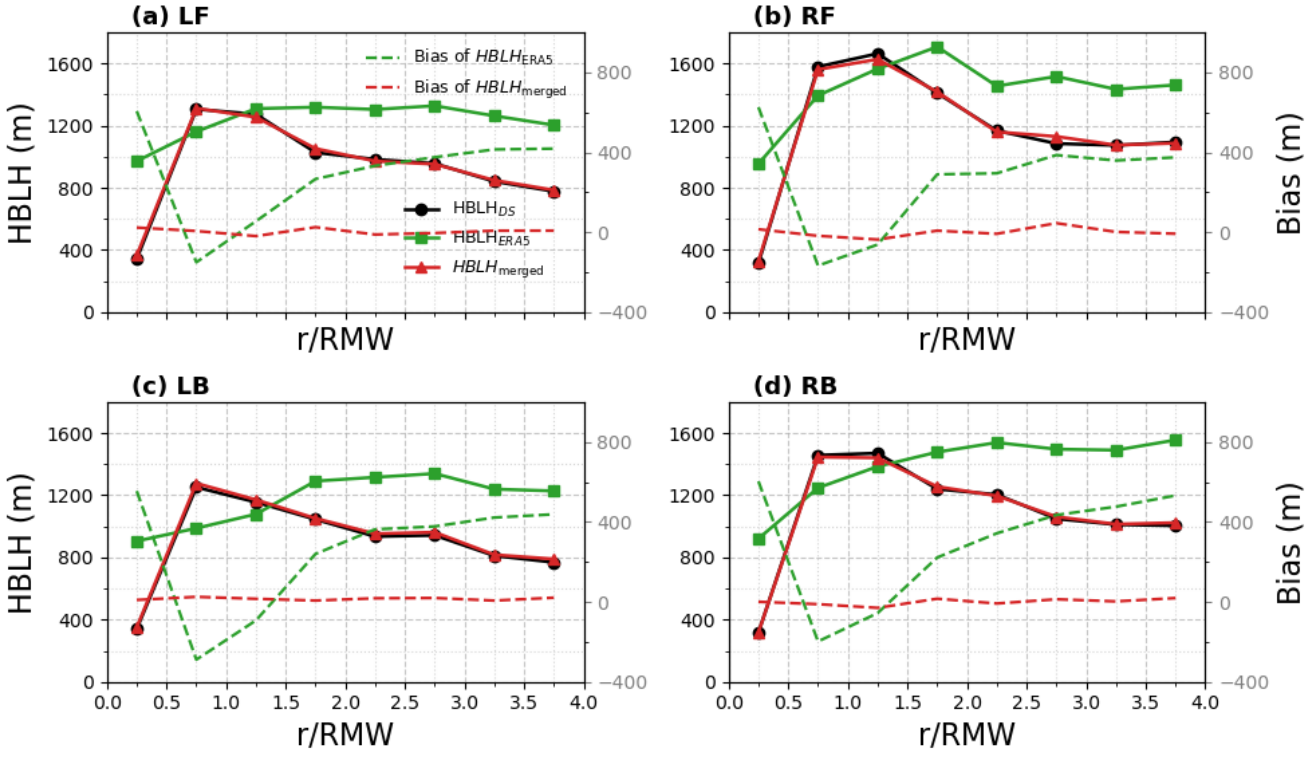
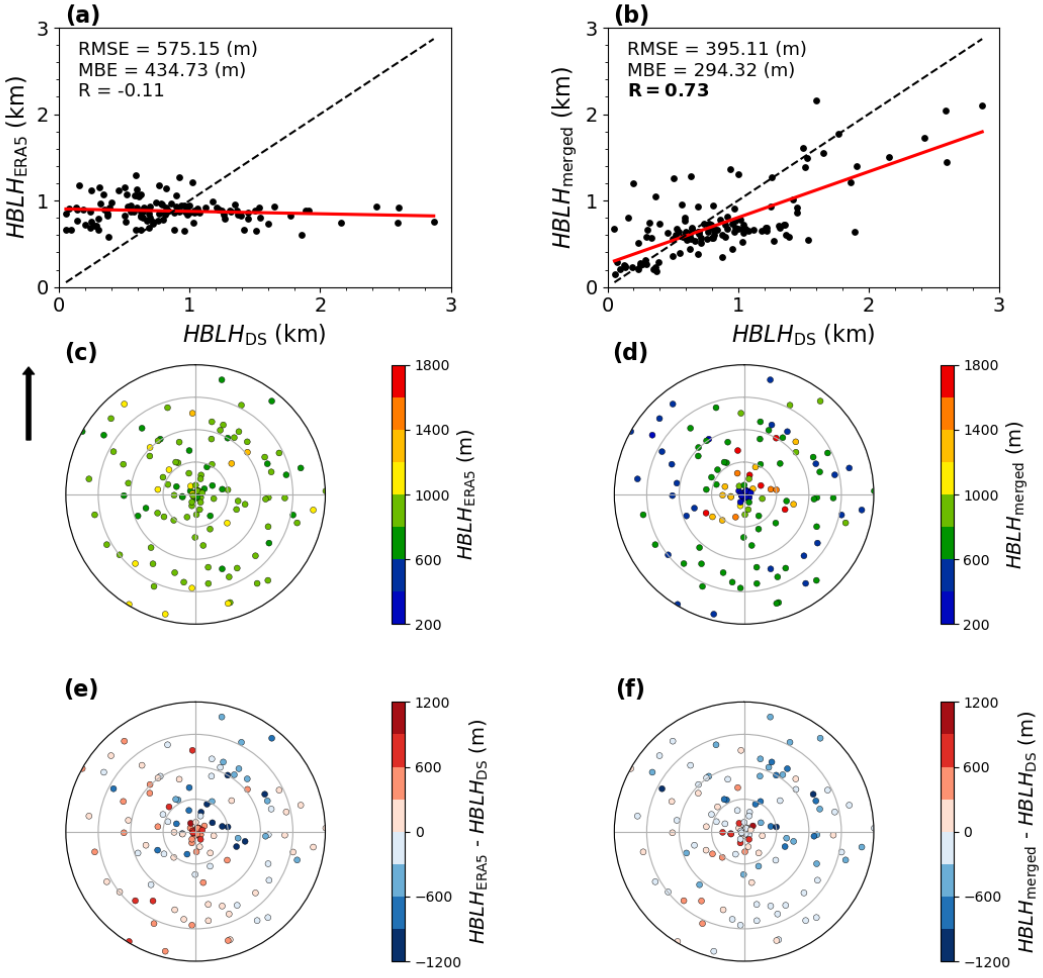


Figure 6. Radial variations HBLHDS, HBLHERA5 and HBLHmerged averaged in the (a) left forward (LF), (b) right forward (RF), (c) left backward (LB), and (d) right backward (RB) quadrant, as well as their biases. The radial distance and quadrants are defined based on the composite analysis shown in Fig. 2.

To further evaluate the generalization capability of the correction model, Fig. 7 presents a comparative assessment of HBLHs before and after correction based on 128 dropsonde observations collected in 2024. Notably, the 2024 dropsonde dataset was not involved in either model training or testing, ensuring its independence and enabling an objective validation of model performance. The post-correction results demonstrate marked improvements in performance. The RMSE and MAE decrease substantially (from 574 m to 395 m and from 435 m to 294 m, respectively), while the correlation coefficient increases from a negative value to 0.73, significant at the $p < 0.01$ level (Figure 8a-b). Despite remaining residual errors, the model substantially reduces systematic biases in ERA5, notably correcting the overestimation at low HBLH and the underestimation at high HBLH.

https://doi.org/10.5194/egusphere-2026-1778
Preprint. Discussion started: 3 June 2026

From the composite spatial distribution, the corrected HBLH more accurately represents the radial structure of the hurricane, exhibiting a clearer dependence on distance from the storm center. Relative to the pre-correction results, it shows reduced HBLH within the eye zone and enhanced HBLH near the high-wind regions (Fig. 8c-d). Overall, the corrected spatial distribution more closely resembles observational evidence (Wang et al., 2023), with the corresponding biases generally converging toward zero (Fig. 8e–f). Notably, in the RF quadrant, both the pre- and post-correction HBLH datasets exhibit a pronounced underestimation relative to observations, indicating limited model improvement in this region (Fig. 8e–f). This deficiency is likely related to the strong dynamical asymmetry of hurricanes. Previous studies have shown that wind speeds are typically enhanced in the RF quadrant (Olfateh et al., 2017; Wang et al., 2023), where intensified turbulence and dynamical variability increase the complexity and nonlinearity of the HBL, thereby posing greater challenges for accurate representation.



https://doi.org/10.5194/egusphere-2026-1778
Preprint. Discussion started: 3 June 2026

Figure 7. Panels (a) and (b) show the scatter distributions of dropsonde profiles meeting the criteria for the entire year of 2024, compared with ERA5 and the RF model. Relevant statistical metrics are displayed in the upper-left corner (total number of dropsondes, N=128). Panels (c) – (f) are similar to Fig.2 but include only the data in 2024.

Through the application of the correction model, an hourly HBLH dataset was constructed for 75 hurricanes over the western Atlantic from 2002 to 2024. The dataset covers the region within 4 RMW from the hurricane center, with a spatial resolution of 0.25°×0.25°. Figure 8 compares spatial distributions merged HBLHs and original ERA5 HBLHs and their evolution during Hurricane Nigel (2023) as an example. The period shown spans from 15 UTC 18 September to 00 UTC 22 September, 2023, including different intensity stages of the hurricane development, during which the MSW retained above hurricane intensity (64 kt). The panels are shown at 3-hourly intervals, consistent with the original time steps of the best-track data.

Compared with the original ERA5 product, the merged dataset provides a more physically consistent depiction of HBLH distribution within the hurricane domain. The most prominent difference lies in the representation of the radial structure. $HBLH_{\text{merged}}$ captures a more realistic radial pattern, with HBLH increasing outward from the eye, reaching a maximum typically around 1-2 RMW, and then decreasing farther outward. In contrast, ERA5 tends to produce a much smoother radial gradient and fails to reproduce the sharp increase in HBLH near the eye region, which results in generally overestimated HBLH values within the eye. The merged dataset also performs better in tracking the hurricane center and preserving the inner-core structure. Beginning at approximately 21 UTC on 20 September, ERA5 no longer accurately captures the storm-center position, as its low-HBLH core becomes noticeably displaced toward the LB quadrant**.** By contrast, $HBLH_{\text{merged}}$ consistently maintains a distinct low-value center within the hurricane eye throughout the period, indicating a more accurate representation of the storm inner core. In addition, the merged HBLH more clearly captures the evolution of both its magnitude and spatial distribution in response to changes in hurricane intensity, whereas the ERA5 HBLH and its spatial pattern vary much more slowly and remain nearly unchanged in some periods.

Overall, these results demonstrate that the merged HBLH dataset, constrained by dropsonde observations, provides a more realistic and observation-consistent characterization of the spatial distribution and temporal evolution of hurricane boundary layer height than the original ERA5 product.

https://doi.org/10.5194/egusphere-2026-1778
Preprint. Discussion started: 3 June 2026

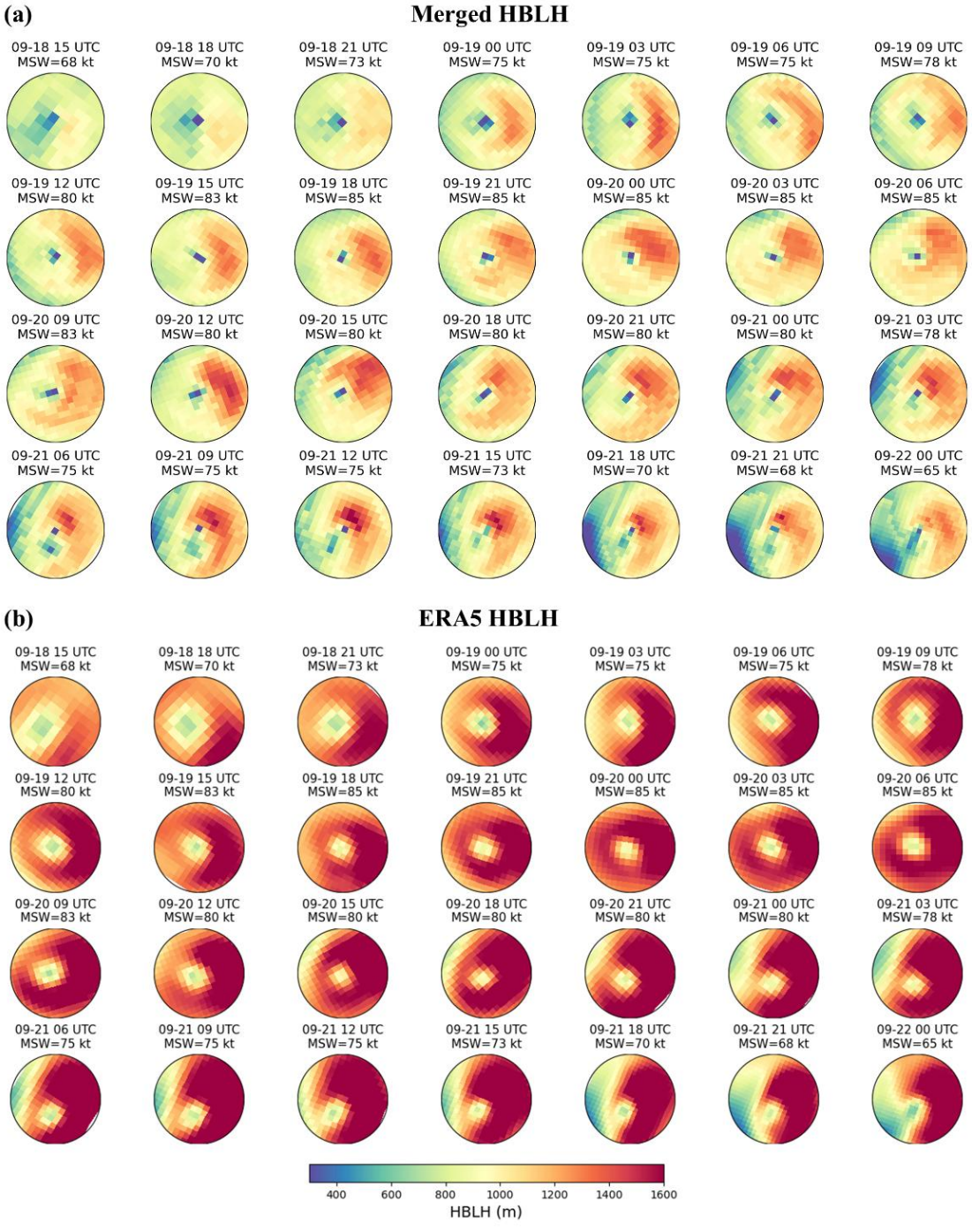


Figure 8. Spatial distributions of (a) merged HBLHs and (b) original ERA5 HBLHs and their temporal evolution within four times the RMW during development of Hurricane Nigel (2023) from 15 UTC 18 September to 00 UTC 22 September. Each panel is rotated so that the storm forward motion is oriented upward.

https://doi.org/10.5194/egusphere-2026-1778
Preprint. Discussion started: 3 June 2026

5 Conclusions and summary

This study develops a merged HBLH dataset for the western Atlantic by integrating dropsonde observations, ERA5 reanalysis, and IBTrACS data using a random forest model. The dataset provides hourly HBLH fields at 0.25° × 0.25° resolution for the period 2002 – 2024 within four times the RMW under hurricane conditions.

Results show that ERA5 exhibits systematic biases in HBLH under hurricane conditions relative to dropsonde-derived values. This bias shows clear spatial dependence, with pronounced overestimation in the eye region and outer storm areas, while underestimation occurs at higher observed HBLH values within the high-wind region. These discrepancies highlight limitations of ERA5 in representing the complex boundary layer structure of intense tropical cyclones.

Using 19 predictors that describe thermodynamic, dynamic, and hurricane-related conditions, the Random Forest model effectively predicts the bias between ERA5 and dropsonde-derived HBLH. The merged HBLH dataset ($HBLH_{\mathrm{merged}}$) retains ERA5 original spatiotemporal resolution (0.25° × 0.25°, hourly) and effectively reduces systematic errors relative to dropsonde observations. Compared with $HBLH_{\mathrm{DS}}$ over the period 2002-2023, $HBLH_{\mathrm{merged}}$ improves the correlation coefficient from 0.40 to 0.93, while reducing the RMSE from 661 m to 246 m and the MAE from 544 m to 159 m. Validation against the independent 2024 dropsonde dataset further highlights the advantage of the post-correction HBLH data, with the correlation coefficient increasing from a negative value to 0.73, demonstrating strong generalization of the correction model.

Beyond the statistical improvement, the merged dataset better reproduces the observed radial structure of HBLH than the original ERA5 product. It effectively reduces the overestimation at low HBLHs and the underestimation at high HBLHs, thereby yielding a more realistic representation of lower HBLHs in the eye region, peak HBLHs in the high-wind region, and the subsequent decrease in HBLH with increasing distance from the storm center. A case study of Hurricane Nigel (2023) further shows that the merged dataset provides a more consistent depiction of the spatial distribution and temporal evolution of HBLH within the hurricane domain.

Despite these improvements, the performance of the correction framework is still limited by the spatial distribution and sample size of available dropsonde observations. Future work should further expand the observational basis by incorporating multi-source profile observations during hurricanes, such as satellite-retrieved profiles, and by extending the method to other regions prone to tropical cyclones, such as the western North Pacific. This would increase the sample size and allow a more comprehensive assessment of the applicability of the method across different regions. Overall, the merged dataset provides an improved and observation-constrained HBLH product that can support future studies of HBL structure and evolution.

https://doi.org/10.5194/egusphere-2026-1778
Preprint. Discussion started: 3 June 2026

Code and data availability. Dropsonde data are publicly available from the NOAA Atlantic Oceanographic and Meteorological Laboratory (AOML) at https://www.aoml.noaa.gov/ftp/hrd/data/dropsonde/. ERA5 reanalysis data are available from the Copernicus Climate Data Store (ECMWF, 2024) at https://cds.climate.copernicus.eu/. IBTrACS data can be accessed at https://www.ncei.noaa.gov/products/international-best-track-archive. The processed dataset and codes used in this study is publicly available at https://doi.org/10.5281/zenodo.17196964.

Author contributions. Xiangyu Pi: Conceptualization, Software, Writing – original draft, Writing – review & editing, Visualization. Yuanjie Zhang: Conceptualization, Methodology, Resources, Writing – review & editing, Supervision, Project administration, Funding acquisition. Yubin Li: Resources, Writing – review & editing, Supervision, Investigation. Xiaoze Xu: Methodology, Writing – review & editing.

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

Acknowledgements. We are thankful for the support from the National Natural Science Foundation of China (Grants 42075070 and 41705004). We acknowledge the NOAA/AOML/Hurricane Research Division in Miami, FL (USA) for providing the dropsonde datasets. The numerical calculations in this paper have been done on the supercomputing system in the Supercomputing Center of Nanjing University of Information Science & Technology.

References

Andreas, E. L., Mahrt, L., and Vickers, D.: A new drag relation for aerodynamically rough flow over the ocean, J. Atmos. Sci., 69, 2521–2537, https://doi.org/10.1175/JAS-D-11-0312.1, 2012.

Bian, G.-F., Nie, G.-Z., and Qiu, X.: How well is outer tropical cyclone size represented in the ERA5 reanalysis dataset?, Atmos. Res., 249, 105339, https://doi.org/10.1016/j.atmosres.2020.105339, 2021.

Cai, J., Xu, K., Zhu, Y., Hu, F., and Li, L.: Prediction and analysis of net ecosystem carbon exchange based on gradient boosting regression and random forest, Appl. Energy, 262, 114566, https://doi.org/10.1016/j.apenergy.2020.114566, 2020.

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

Edson, J. B., Jampana, V., Weller, R. A., Bigorre, S. P., Plueddemann, A. J., Fairall, C. W., Miller, S. D., Mahrt, L., Vickers, D., and Hersbach, H.: On the exchange of momentum over the open ocean, J. Phys. Oceanogr., 43, 1589–1610, https://doi.org/10.1175/JPO-D-12-0173.1, 2013.

https://doi.org/10.5194/egusphere-2026-1778
Preprint. Discussion started: 3 June 2026

Enderami, S. A., Mazumder, R. K., and Sutley, E. J.: Framework for incorporating community social vulnerability in the assessment of hurricane-induced wind risk to residential buildings, https://doi.org/10.48550/arXiv.2207.14206, 28 July 2022.

Franklin, J. L., Black, M. L., and Valde, K.: GPS dropwindsonde wind profiles in hurricanes and their operational implications, Weather Forecasting, 18, 32–44, https://doi.org/10.1175/1520-0434(2003)018%253C0032:GDWPIH%253E2.0.CO;2, 2003.

Farjami, H. and Hesari, A. R. E.: Assessment of sea surface wind field pattern over the caspian sea using EOF analysis, Reg. Stud. Mar. Sci., 35, 101254, https://doi.org/10.1016/j.rsma.2020.101254, 2020.

Graham, R. M., Hudson, S. R., and Maturilli, M.: Improved performance of ERA5 in arctic gateway relative to four global atmospheric reanalyses, Geophys. Res. Lett., 46, 6138–6147, https://doi.org/10.1029/2019GL082781, 2019.

Guo, J., Zhang, J., Yang, K., Liao, H., Zhang, S., Huang, K., Lv, Y., Shao, J., Yu, T., Tong, B., Li, J., Su, T., Yim, S. H. L., Stoffelen, A., Zhai, P., and Xu, X.: Investigation of near-global daytime boundary layer height using high-resolution radiosondes: first results and comparison with ERA5, MERRA-2, JRA-55, and NCEP-2 reanalyses, Atmos. Chem. Phys., 21, 17079–17097, https://doi.org/10.5194/acp-21-17079-2021, 2021.

Guo, J., Zhang, J., Shao, J., Chen, T., Bai, K., Sun, Y., Li, N., Wu, J., Li, R., Li, J., Guo, Q., Cohen, J. B., Zhai, P., Xu, X., and Hu, F.: A merged continental planetary boundary layer height dataset based on high-resolution radiosonde measurements, ERA5 reanalysis, and GLDAS, Earth Syst. Sci. Data, 16, 1–14, https://doi.org/10.5194/essd-16-1-2024, 2024.

Gahtan, J., K. R. Knapp, C. J. Schreck, H. J. Diamond, J. P. Kossin, M. C. Kruk.: International Best Track Archive for Climate Stewardship (IBTrACS) Project, Version 4r01. [NA]. NOAA National Centers for Environmental Information. doi:10.25921/82ty-9e16, accessed 6 October 2024, 2024

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., De Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., and Thépaut, J.-N.: The ERA5 global reanalysis, Q. J. R. Meteorolog. Soc., 146, 1999–2049, https://doi.org/10.1002/qj.3803, 2020.

Hock, T. F. and Franklin, J. L.: The NCAR GPS Dropwindsonde, Bull. Am. Meteorol. Soc., 80, 407–420, https://doi.org/10.1175/1520-0477(1999)080%253C0407:TNGD%253E2.0.CO;2, 1999.

Holtslag, A. a. M. and Boville, B. A.: Local Versus Nonlocal Boundary-Layer Diffusion in a Global Climate Model, J. Clim., 6, 1825–1842, https://doi.org/10.1175/1520-0442(1993)006%253C1825:LVNBLD%253E2.0.CO;2, 1993.

Hong, S.-Y.: A new stable boundary-layer mixing scheme and its impact on the simulated East Asian summer monsoon. Q. J. R. Meteorol. Soc. 136, 1481-1496. https://doi.org/10.1002/qj.665, 2010.

https://doi.org/10.5194/egusphere-2026-1778
Preprint. Discussion started: 3 June 2026

Hsu, S. A., Shen, H., and He, Y.: Characterizing overwater roughness reynolds number during hurricanes, Meteorol. Atmos. Phys., 131, 279–285, https://doi.org/10.1007/s00703-017-0569-y, 2019.

Jiang, R. and Zhao, K.: Using machine learning method on calculation of boundary layer height, Neural Comput. Appl., 34, 2597–2609, https://doi.org/10.1007/s00521-021-05865-3, 2022.

Kepert, J. D.: Choosing a boundary layer parameterization for tropical cyclone modeling, Mon. Weather Rev., 140, 1427–1445, https://doi.org/10.1175/MWR-D-11-00217.1, 2012.

Klotzbach, P. J., Bowen, S. G., Pielke, R., and Bell, M.: Continental U.S. Hurricane Landfall Frequency and Associated Damage: Observations and Future Risks, Bull. Am. Meteorol. Soc., 99, 1359–1376, https://doi.org/10.1175/BAMS-D-17-0184.1, 2018.

Knapp, K. R., M. C. Kruk, D. H. Levinson, H. J. Diamond, and C. J. Neumann.: The International Best Track Archive for Climate Stewardship (IBTrACS): Unifying tropical cyclone best track data. Bulletin of the American Meteorological Society, 91, 363-376. doi:10.1175/2009BAMS2755.1, 2010

Li, S., Wang, Y., Huang, H., Huang, L., and Chen, Y.: Study on typhoon disaster assessment by mining data from social media based on artificial neural network, Nat. Hazard., 116, 2069–2089, https://doi.org/10.1007/s11069-022-05754-5, 2023.

Liao, F., Deng, H., Gao, Z., and Chan, P.: The research on boundary layer evolution characteristics of typhoon usagi based on observations by wind profilers, Acta Oceanolog. Sin., 36, 39–44, https://doi.org/10.1007/s13131-017-1109-9, 2017.

Lin, J., Qian, T., and Klotzbach, P.: Tropical cyclones, Atmos. Ocean, 60, 360–398, https://doi.org/10.1080/07055900.2022.2086849, 2022.

Liu, J., Zuo, Y., Wang, N., Yuan, F., Zhu, X., Zhang, L., Zhang, J., Sun, Y., Guo, Z., Guo, Y., Song, X., Song, C., and Xu, X.: Comparative Analysis of Two Machine Learning Algorithms in Predicting Site-Level Net Ecosystem Exchange in Major Biomes, Remote Sens., 13, 2242, https://doi.org/10.3390/rs13122242, 2021.

Liu, L., Han, Y., Xia, Y., Guo, Q., Gao, W., and Guo, J.: Investigation of atmospheric dynamic and thermodynamic structures of typhoon Sinlaku (2020) from high-resolution dropsonde and two-way rawinsonde measurements, Remote Sens., 14, 2704, https://doi.org/10.3390/rs14112704, 2022.

Mendelsohn, R., Emanuel, K., Chonabayashi, S., and Bakkensen, L.: The impact of climate change on global tropical cyclone damage, Nat. Clim. Change, 2, 205–209, https://doi.org/10.1038/nclimate1357, 2012.

Molero, F., Barragán, R., and Artíñano, B.: Estimation of the atmospheric boundary layer height by means of machine learning techniques using ground-level meteorological data, Atmospheric Research, 279, 106401, https://doi.org/10.1016/j.atmosres.2022.106401, 2022.

Nolan, D. S., Zhang, J. A., and Stern, D. P.: Evaluation of planetary boundary layer parameterizations in tropical cyclones by comparison of In Situ observations and high-resolution simulations of hurricane Isabel (2003). Part I: initialization, maximum winds, and the outer-core boundary layer, Mon. Weather Rev., 137, 3651–3674, https://doi.org/10.1175/2009MWR2785.1, 2009.

https://doi.org/10.5194/egusphere-2026-1778
Preprint. Discussion started: 3 June 2026

Olfateh, M., Callaghan, D. P., Nielsen, P., and Baldock, T. E.: Tropical cyclone wind field asymmetry—development and evaluation of a new parametric model, Journal of Geophysical Research: Oceans, 122, 458–469, https://doi.org/10.1002/2016JC012237, 2017.

Ren, Y., Zhang, J. A., Guimond, S. R., and Wang, X.: Hurricane boundary layer height relative to storm motion from GPS dropsonde composites, Atmosphere, 10, 339, https://doi.org/10.3390/atmos10060339, 2019.

Ren, Y., Zhang, J. A., Vigh, J. L., Zhu, P., Liu, H., Wang, X., and Wadler, J. B.: An observational study of the symmetric boundary layer structure and tropical cyclone intensity, Atmosphere, 11, 158, https://doi.org/10.3390/atmos11020158, 2020.

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

Seidel, D. J., Zhang, Y., Beljaars, A., Golaz, J.-C., Jacobson, A. R., and Medeiros, B.: Climatology of the planetary boundary layer over the continental united states and Europe, J. Geophys. Res.: Atmos., 117, https://doi.org/10.1029/2012JD018143, 2012.

Slingo, J. M.: The development and verification of a cloud prediction scheme for the ecmwf model, Q. J. R. Meteorolog. Soc., 113, 899–927, https://doi.org/10.1002/qj.49711347710, 1987.

Smith, R. K. and Montgomery, M. T.: Hurricane boundary-layer theory, Q. J. R. Meteorolog. Soc., 136, 1665–1670, https://doi.org/10.1002/qj.679, 2010.

Smith, R. K. and Thomsen, G. L.: Dependence of tropical-cyclone intensification on the boundary-layer representation in a numerical model, Q. J. R. Meteorolog. Soc., 136, 1671–1685, https://doi.org/10.1002/qj.687, 2010.

Smith, R. K., Montgomery, M. T., and Van Sang, N.: Tropical cyclone spin-up revisited, Quarterly Journal of the Royal Meteorological Society, 135, 1321–1335, https://doi.org/10.1002/qj.428, 2009.

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

Wang, J., Young, K., Hock, T., Lauritsen, D., Behringer, D., Black, M., Black, P. G., Franklin, J., Halverson, J., Molinari, J., Nguyen, L., Reale, T., Smith, J., Sun, B., Wang, Q., and Zhang, J. A.: A long-term, high-quality, high-vertical-resolution GPS dropsonde dataset for hurricane and other studies, Bull. Am. Meteorol. Soc., 96, 961–973, https://doi.org/10.1175/BAMS-D-13-00203.1, 2015.

Wang, X., Zhang, Y., Li, Y., and Gao, Z.: Hurricane boundary layer height distribution from dropsonde composites based on the bulk richardson number method, Atmos. Res., 292, 106853, https://doi.org/10.1016/j.atmosres.2023.106853, 2023.

Wei, C., Zhao, X., Liu, Y., Yang, P., Zhou, Z., and Chen, Y.: Bias analysis and correction of ERA5 reanalysis in the context of tropical cyclones, J. Geophys. Res.: Atmos., 130, e2024JD042737, https://doi.org/10.1029/2024JD042737, 2025.

Xiong, J., Yu, F., Fu, C., Dong, J., and Liu, Q.: Evaluation and improvement of the ERA5 wind field in typhoon storm surge simulations, Appl. Ocean Res., 118, 103000, https://doi.org/10.1016/j.apor.2021.103000, 2022.

https://doi.org/10.5194/egusphere-2026-1778
Preprint. Discussion started: 3 June 2026

Xi, X., Yang, Q., Liu, C., Shupe, M. D., Han, B., Peng, S., Zhou, S., and Chen, D.: Evaluation of the planetary boundary layer height from ERA5 reanalysis with MOSAiC observations over the Arctic Ocean, J. Geophys. Res.: Atmos., 129, e2024JD040779, https://doi.org/10.1029/2024JD040779, 2024.

Yu, Fujiang, et al.: Modern Storm Surge Prediction Technology and Its Application (in Chinese). Science Press, Beijing, pp. 130–134, 2020

Zeng, X., Brunke, M. A., Zhou, M., Fairall, C., Bond, N. A., and Lenschow, D. H.: Marine atmospheric boundary layer height over the eastern pacific: data analysis and model evaluation, J. Clim., 17, 4159–4170, https://doi.org/10.1175/JCLI3190.1, 2004.

Zhang, J. A., Rogers, R. F., Nolan, D. S., and Marks, F. D.: On the characteristic height scales of the hurricane boundary layer, Mon. Weather Rev., 139, 2523–2535, https://doi.org/10.1175/MWR-D-10-05017.1, 2011.

Zhang, J. A., Rogers, R. F., Reasor, P. D., Uhlhorn, E. W., and Marks, F. D.: Asymmetric hurricane boundary layer structure from dropsonde composites in relation to the environmental vertical wind shear, Mon. Weather Rev., 141, 3968–3984, https://doi.org/10.1175/MWR-D-12-00335.1, 2013.

Zhang, Y., Gao, Z., Li, D., Li, Y., Zhang, N., Zhao, X., and Chen, J.: On the computation of planetary boundary-layer height using the bulk richardson number method, Geosci. Model Dev., 7, 2599–2611, https://doi.org/10.5194/gmd-7-2599-2014, 2014.

Zhang, Y., Sun, K., Gao, Z., Pan, Z., Shook, M. A., and Li, D.: Diurnal climatology of planetary boundary layer height over the contiguous united states derived from AMDAR and reanalysis data, J. Geophys. Res.: Atmos., 125, e2020JD032803, https://doi.org/10.1029/2020JD032803, 2020.

Zhu, P., Menelaou, K., and Zhu, Z.: Impact of subgrid-scale vertical turbulent mixing on eyewall asymmetric structures and mesovortices of hurricanes, Q. J. R. Meteorolog. Soc., 140, 416–438, https://doi.org/10.1002/qj.2147, 2014.