



# Bias Assessment and Random Forest-Based Correction of Temperature Observations from the Level-Drift Phase of Ascent-Drift-Descent Radiosonde System

Xiaojuan YAO<sup>1</sup>, Qiyun GUO<sup>2,3,4</sup>, Xin SUN<sup>1</sup>, Jincheng WANG<sup>5,6,7</sup>, Yanxia JI<sup>1</sup>, Linchun LIU<sup>1</sup>, Dan WANG<sup>5,6</sup>, Feng ZHU<sup>1</sup>, and Ke LIU<sup>1</sup>

<sup>1</sup>Inner Mongolia Meteorological Observatory, Hohhot, 010051, China

<sup>2</sup>Meteorological Observation Centre of China Meteorological Administration, Beijing, 100081, China

<sup>3</sup>State Key Laboratory of Environment Characteristics and Effects for Near-space, Beijing, 100081, China

<sup>4</sup>Engineering Technology Research Center for Meteorological Observation of CMA, Beijing, 100081, China

10 <sup>5</sup>CMA Earth System Modeling and Prediction Centre (CEMC), Beijing 100081, China

<sup>6</sup>State Key Laboratory of Severe Weather Meteorological Science and Technology (LaSW), Beijing, 100081, China

<sup>7</sup>College of Atmospheric Sciences, Lanzhou University, Lanzhou, 730000, China

**Correspondence:** Qiyun GUO (feng\_cloudy@163.com); Xin SUN (sunx\_beht@163.com)

**Abstract.** To facilitate the quantitative application of Ascent-Drift-Descent Radiosonde System (ADDRS) temperature data from the level-drift phase in numerical weather prediction (NWP) models, this study performs a comprehensive quality assessment of the temperature observations using the fifth generation European Centre for Medium-Range Weather Forecasts reanalysis (ERA5) temperature field data as a reference. The results indicate that, due to solar radiative heating, the temperature biases between the level-drift observations and the ERA5 are positive during daytime and tend to increase with the solar elevation angle (SEA). During nighttime, temperature biases are slightly negative. Statistical analysis of temperature biases reveals that the biweight mean values are 3.48 K (daytime) and -0.91 K (nighttime), with corresponding biweight standard deviations of 5.3 K and 2.3 K. Thus, the daytime biweight standard deviation meets the World Meteorological Organization (WMO) "threshold" target, while its nighttime counterpart meets the WMO breakthrough target. Furthermore, a random forest-based temperature bias correction model is developed. After correction, the biweight mean values of temperature biases decrease to 0.52 K (daytime) and -0.23 K (nighttime), and the biweight standard deviations decrease to 2.03 K (daytime) and 1.54 K (nighttime). Both evaluation metrics meets the WMO breakthrough target, with the nighttime metrics notably exceeding it. The probability distribution function of the corrected biases aligns more closely with a normal distribution, demonstrating the effectiveness of the random forest-based model in correcting the biases of level-drift temperature observations. This research establishes a critical foundation for the future application of ADDRS data assimilation in NWP models.

## 30 1 Introduction

Sounding data, serving as benchmark observations for atmospheric observation networks, can provide real three-dimensional atmospheric information (such as temperature, pressure, humidity, and wind) from the troposphere to the lower stratosphere,



and thus constitute an essential source of atmospheric data (Faccani et al., 2009; He et al., 2009). With the continuous advancements in sounding technology, China has progressively upgraded from the Type 59 radiosonde system to the L-band sounding system. This upgrade has enabled second-level detection with a high temporal resolution of 1.2 s and a vertical resolution of 8 m, providing denser vertical atmospheric profiles (Hao et al., 2014). However, the current sounding frequency in China remains twice daily, which is still insufficient to meet the monitoring and forecasting demands for meso- and small-scale hazardous weather systems (Chen et al., 2011). Additionally, increasing the frequency of conventional soundings poses challenges such as high operational costs and difficulties in sustaining long-term operations (Guo et al., 2018; Zhou et al., 2023).

To address these challenges and extend the scope of existing sounding observations, the Meteorological Observation Center of the China Meteorological Administration (CMA) has innovatively developed a next-generation sounding observation system. Through a single balloon launch, the system achieves a three-phase observation strategy: ascent, level-drift and descent phases (Cao et al., 2019; Cao et al., 2026; Liu et al., 2022; Zhang 2022). This technology extends the effective observation duration to over 6 hours. Without substantial cost increases, it provides two vertical tropospheric profiles (ascent and descent phases) and continuous horizontal observations at stratospheric altitudes for up to 4 hours. This approach effectively achieves spatio-temporal observation densification and a doubling of observation counts, while balancing the costs of densifying the upper-level observation network (Wang et al., 2023; Yang et al., 2023). Moreover, to ensure more accurate positioning and more stable data acquisition, this system integrates navigation satellite positioning technology and is referred to as the Ascent-Drift-Descent Radiosonde System (ADDRS) (Wang et al., 2021).

Ensuring high data quality is essential for the application of ADDRS observation data. Wang et al. (2020) developed a quality control scheme for ADDRS observation data and evaluated its performance using the Global/Regional Assimilation and PrEdiction System – four-dimensional variational data assimilation. Their results indicated that the wind, temperature and humidity fields in the ascent and descent phases meet the World Meteorological Organization (WMO) breakthrough target. Specifically, the temperature field within the troposphere and the pressure field throughout the entire observation period achieved the ideal targets, and the wind field and the nighttime temperature field were suitable for direct data assimilation, indicating strong application potential of ADDRS data. Zhang et al. (2021) evaluated the quality of ADDRS data using the National Centers for Environmental Prediction Final Operational Global Analysis (NCEP-FNL) data. They found that the temperature and pressure deviations for the ascent and descent phases were within  $\pm 2^{\circ}\text{C}$  and  $\pm 4$  hPa, respectively. Assimilating the descent-phase data increased the equitable threat scores for downpour, torrential rain and heavy rain by 100%, 9.2% and 29%, respectively. Current research consistently supports the reliability and usability of ADDRS data in the ascent and descent phases. Compared with traditional sounding methods, a notable advantage of ADDRS is its capability for prolonged, direct and continuous observations of the stratosphere, effectively filling gaps in upper-atmospheric data. Thus, ADDRS holds great potential for application. Regarding the quality assessment of observation



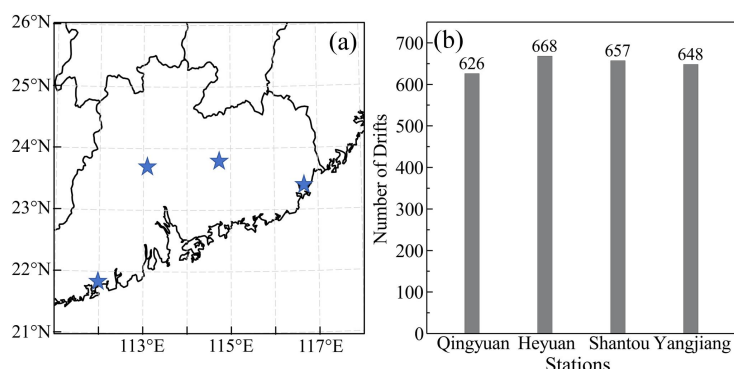
65 data in the level-drift phase, Wang et al. (2020) found that the wind and nighttime temperature fields in the level-drift phase met the breakthrough targets, and the daytime temperature field was close to the “threshold” target. Therefore, daytime temperature, pressure and humidity require bias correction prior to data assimilation. Similarly, Rong et al. (2023) found that, due to solar radiative heating, the nighttime temperature biases of the level-drift observations relative to the global navigation satellite system (GNSS) radio occultation data met the breakthrough target, while the daytime temperature biases  
70 approached the “threshold” target. Additionally, Qian (2019) also noted that the biases between the daytime temperature in the level-drift phase and the NCEP-FNL data were notably larger. These persistent quality issues in level-drift observations are key reasons why such data have not yet been applied in data assimilation, thereby hindering further application.

The aforementioned findings were primarily derived from the ADDRS network trials conducted by the CMA Meteorological Observation Center in the middle and lower reaches of the Yangtze River Basin. These studies identified large daytime  
75 temperature biases during the level-drift phase but did not implement further bias correction. To address these shortcomings, this study focuses on two scientific objectives: (1) to examine and evaluate the causes of temperature errors in the level-drift phase of ADDRS, and (2) to correct temperature biases in the level-drift phase to meet the requirements for data assimilation in numerical weather prediction (NWP) models. Consequently, the fifth-generation European Centre for Medium-Range Weather Forecasts reanalysis (ERA5) data are used in this study as a reference to evaluate the quality of temperature  
80 observations in the level-drift phase of ADDRS. By statistically analyzing the characteristics and variations of temperature biases and standard deviations in the level-drift phase, we investigate the underlying error mechanisms and establish a preliminary bias correction model based on a random forest algorithm. This work provides a foundation for future assimilation and application of ADDRS observations in NWP models.

## **2 Data and preliminary quality control**

### **85 2.1 Data description**

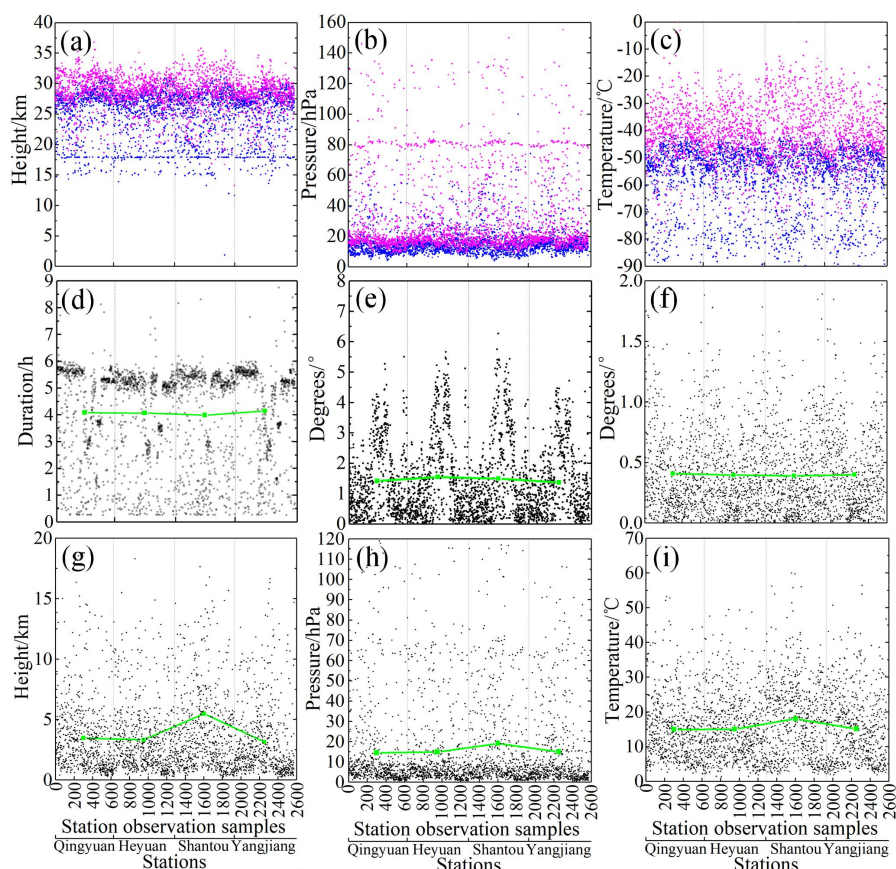
From July 2022 to October 2023, the CMA Meteorological Observation Center conducted pilot ADDRS observations at four stations in Guangdong Province (Fig. 1a): Qingyuan, Heyuan, Shantou, and Yangjiang stations. The ADDRS balloons were launched twice daily, with a time interval of about half an hour apart from the operational sounding schedule. The temporal resolution of the observation data is 2 s. Due to the limited number of launches in July 2022 and September–October 2023  
90 (fewer than 10 per month per station), we primarily utilize the data from August 2022 to August 2023 in this research, which exhibit better temporal continuity. Figure 1b presents the statistics of successful level-drift events for the four stations, and all the stations had over 600 cases.



**Figure 1: (a) Distribution of ADDRS observation stations and (b) statistics of successful level-drift cases at each station.**

## 95 2.2 Distribution characteristics and preliminary quality control of observation variables in the level-drift phase of ADDRS

The level-drift data used in this study have undergone quality control based on the scheme developed by Wang et al. (2020), which assigns quality flags to temperature, geopotential height, pressure and horizontal wind (0 for "correct", 1 for "suspect", 2 for "erroneous", and 9 for "unchecked"). For the 13-month level-drift observation data at the four stations, samples with a quality flag of 0 are selected for a comprehensive evaluation of observations. Figure 2 presents the scatter distributions of the minimum and maximum values of the observed geopotential height (Fig. 2a), pressure (Fig. 2b) and temperature (Fig. 2c) during each level-drift phase at each station. The results indicate that the variation range of each variable at each station is relatively concentrated during the level-drift phase, with geopotential height, pressure and temperature primarily clustered within 26–31 km, 8–20 hPa and  $-55^{\circ}\text{C}$  to  $-40^{\circ}\text{C}$ , respectively. Furthermore, by calculating the difference between the maximum and minimum values of each variable during each level-drift phase, we obtain the observation duration of each level-drift phase (Fig. 2d), the drifts in the longitudinal, latitudinal and geopotential height directions (Figs. 2e–g), pressure difference (Fig. 2h) and temperature difference (Fig. 2i). Evidently, the level-drift durations for most stations exceeded 5 hours, with high scatter overlap indicating good observation stability. The average level-drift durations were all above 4 hours, satisfying the technical requirement that a balloon level-drift duration should be  $\geq 4$  hours. The mean longitudinal and latitudinal drifts were around  $1.5^{\circ}$  and  $0.4^{\circ}$ , respectively. Notably, for samples where level-drift durations were less than 4 hours with a high degree of scatter point clustering, the longitudinal drift was also relatively larger, further suggesting that the observation data with level-drift durations shorter than 4 hours are less stable. For Qingyuan, Heyuan and Yangjiang stations, the averages of the drift in the geopotential height direction, pressure difference and temperature difference were approximately 3 km, 15 hPa and  $15^{\circ}\text{C}$ , respectively. However, at Shantou station, the average drift in the geopotential height direction was 5.5 km, and the averages of the pressure difference (19 hPa) and temperature difference ( $18^{\circ}\text{C}$ ) were also slightly larger compared with the other stations, indicating relatively lower data stability at this location.



**Figure 2:** Scatter distributions of the minimum (blue) and maximum (pink) values (vertical axis) of various variables across sample size intervals at the four stations during the level-drift phase of ADDR from August 2022 to August 2023: (a) geopotential height, (b) pressure, (c) temperature and (d) level-drift duration, (e) longitudinal drift, (f) latitudinal drift, (g) drift along the geopotential height direction, (h) pressure difference and (i) temperature difference. The polylines with squares represent the average values at each station.

Moreover, for the geopotential height of about 18 km and the pressure of approximately 80 hPa, their sample distributions exhibit relatively horizontal linear patterns. Focusing on the samples with the minimum geopotential height of <19 km in the level-drift phase (Fig. 3a), we find that the level-drift durations for such samples were generally less than 4 hours and mostly below 2 hours, and the maximum pressure clustered around 80 hPa. Given the overall evaluation, these observation data clearly suffer from quality issues. Therefore, a preliminary quality control procedure was applied to exclude the level-drift observations with the minimum geopotential height of <19 km.

Figure 3b presents the statistics of level-drift cases across the four stations following the preliminary quality control. Figure 4 displays the scatter distributions of various variables after the preliminary quality control. The results indicate that after quality control, the average level-drift duration at each station increases to about 4.5 hours. Most observations with level-drift durations of <2 hours are removed, but some samples with level-drift durations of <4 hours remain. In addition, the average values of the drifts along the longitudinal, latitudinal and geopotential height directions, as well as mean values



of the pressure difference and temperature difference, markedly decrease, especially the average pressure difference, which drops to around 8 hPa.

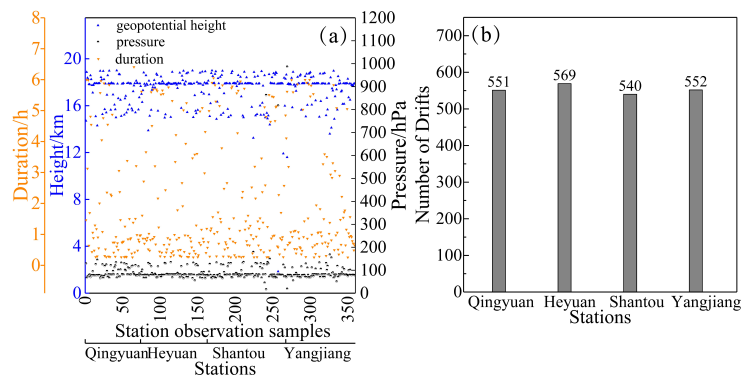


Figure 3: (a) Scatter distributions of the minimum geopotential height below 19 km (blue), the level-drift duration (orange) and the maximum pressure (black) during the level-drift phase across sample size intervals, and (b) number of level-drift occurrences at each station after the preliminary quality control.

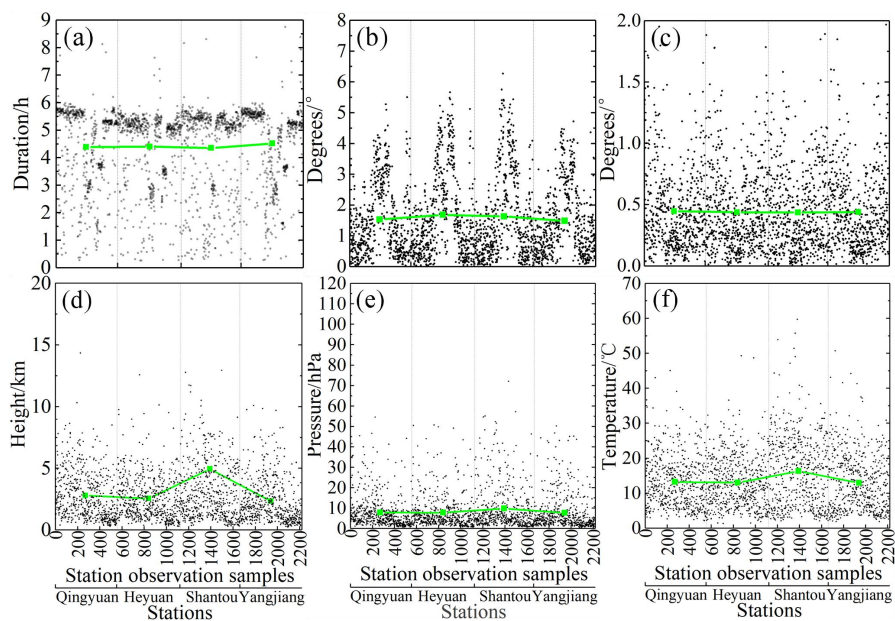


Figure 4: Same as Fig. 2, but for the (a) level-drift duration, (b) longitudinal drift, (c) latitudinal drift, (d) drift along the geopotential height direction, (e) pressure difference and (f) temperature difference after the preliminary quality control.

3 Quality assessment of level-drift temperature observations from ADDR

3.1 Overall comparison between the ADDR temperature observations and the ERA5 data

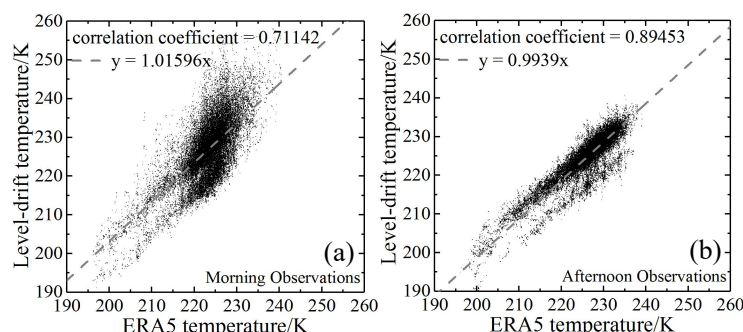
Rong et al. (2023) evaluated the quality of temperature observations in the level-drift phase of ADDR using GNSS radio occultation data, while performing cross-comparisons with ERA5 reanalysis data. Their results demonstrated high consistency between the ERA5 temperature and radio occultation dry temperature, confirming that ERA5 reanalysis data



exhibit high accuracy and can serve as a reference for the quality assessment of other observation data. Therefore, we use hourly ERA5 reanalysis temperature data in this study to analyze the temperature biases in the level-drift phase of ADDRS.

150 Prior to the comparison analysis, the ERA5 data are spatially interpolated to match the latitude, longitude and geopotential height of the level-drift trajectory. Temporally, considering the 2 s resolution of the level-drift observations, a time-weighted interpolation is performed using the two adjacent ERA5 time steps, with weights assigned according to their temporal distance from each radiosonde observation, thereby achieving spatiotemporal collocation.

Figure 5 illustrates the overall comparison between the temperature data from the level-drift phase of ADDRS and the ERA5 temperature data. According to the launch times of the ADDRS balloons, the observations are categorized into morning and afternoon sessions. The results show that for the morning observation period, the correlation coefficient of the temperature variable between the two datasets is 0.71, and the slope of the linear fit is slightly greater than 1. Overall, the level-drift temperature observations from the morning session mainly exhibit positive biases relative to the ERA5 temperature data. In terms of the afternoon observations, the correlation between the two datasets is stronger (a correlation coefficient of 0.89), and the fitted linear coefficient of slightly less than 1 indicates that the overall ADDRS observations are slightly lower than the ERA5 temperature data. Generally, the level-drift temperature observations from the afternoon session show weak negative biases relative to the ERA5 data. Notably, both the morning and afternoon observations exhibit two distinct dense bands of scatter points, with the secondary bands consistently characterized by negative biases.



165 **Figure 5: Temperature observations from the level-drift phase of ADDRS versus the ERA5 temperature data: (a) morning observations and (b) afternoon observations. The dashed lines represent the fitted lines.**

### 3.2 Variation of ADDRS temperature biases with solar elevation angle

Qian (2019) and Rong et al. (2023), in their analyses of the variations in temperature biases between the ADDRS observations and the ERA5 data as a function of local time, both noted that daytime temperature during the level-drift phase exhibits positive biases due to solar radiative heating and insufficient ventilation during detection processes. Furthermore, the influence of radiative heating becomes more pronounced with increasing observation elevation. The results in the previous section also indicate overall positive temperature biases during the morning observation period. Accordingly, this study introduces solar elevation angle (SEA) to further investigate the variation of temperature biases, thereby elucidating



the relationship between level-drift temperature biases and solar radiation.

175 The SEA varies with local time ( $t$ ) and solar declination ( $\delta$ ). The solar declination is equal to the latitude of the subsolar point. First, the solar angle ( $\Gamma$ ) is calculated based on the observation date, as shown in Eq. (1).

$$\Gamma = 2\pi(N - 79.6764 + 0.2422 \cdot (y - 1985) - (y - 1985)/4) / 365.2422, \quad (1)$$

where  $N$  indicates the day of year ( $N=1$  for January 1st), and  $y$  denotes the observation year. The calculation for solar declination angle ( $\delta$ ) is expressed by Eq. (2).

$$\delta = 0.3723 + 23.2567 \sin(\Gamma) + 0.1149 \sin(2\Gamma) - 0.1712 \sin(3\Gamma) - 0.758 \cos(\Gamma) + 0.3656 \cos(2\Gamma) + 0.0201 \cos(3\Gamma), \quad (2)$$

Formula for calculating parallax ( $E_t$ ):

$$E_t = 0.0028 - 1.9857 \sin(\Gamma) + 9.9059 \sin(2\Gamma) - 7.0924 \cos(\Gamma) - 0.6882 \cos(2\Gamma), \quad (3)$$

Formula for calculating local time ( $t$ ):

$$t = h + m/60 + s/3600 + (\text{lon} - 15 \times T_Z) / 15 + E_t / 60, \quad (4)$$

185 where  $h$ ,  $m$ , and  $s$  represent the hour, minute and second of the observation time, respectively, “lon” denotes longitude, and  $T_Z$  indicates the local time zone of the observation site.

The hour angle  $H$  is then calculated from the local time:

$$H = 15 \times (t - 12), \quad (5)$$

where  $t$  is the local time in hours and  $H$  is in degrees.  $H$  is converted to radians before being substituted into Eq. (6).

190 Finally, SEA ( $\theta$ ) is calculated by Eq. (6).

$$\sin \theta = \sin \varphi \sin \delta + \cos \varphi \cos \delta \cos H, \quad (6)$$

where  $\varphi$  denotes the geographic latitude of the observation site. Note that both solar declination ( $\delta$ ) and geographic latitude ( $\varphi$ ) are positive in the northern hemisphere and negative in the southern hemisphere.

Using the SEA calculated from the above formula, we analyze the variation of temperature biases ( $T_A - T_E$ ) between the

195 ADDR5 level-drift observations and the ERA5 data as a function of the SEA. Figure 6a shows the scatter distribution of temperature biases ( $T_A - T_E$ ) for the four stations versus the SEA over the 13-month period.

During nighttime, when the SEA is less than 0,  $T_A - T_E$  values hover around 0 K with a slight prevalence of negative biases, predominantly ranging from  $-5$  K to  $-3$  K. A secondary bias concentration band is also evident between  $-8$  K and  $-12$  K, exhibiting a relatively linear horizontal distribution (i.e., biases show little dependence on the SEA). During daytime, when

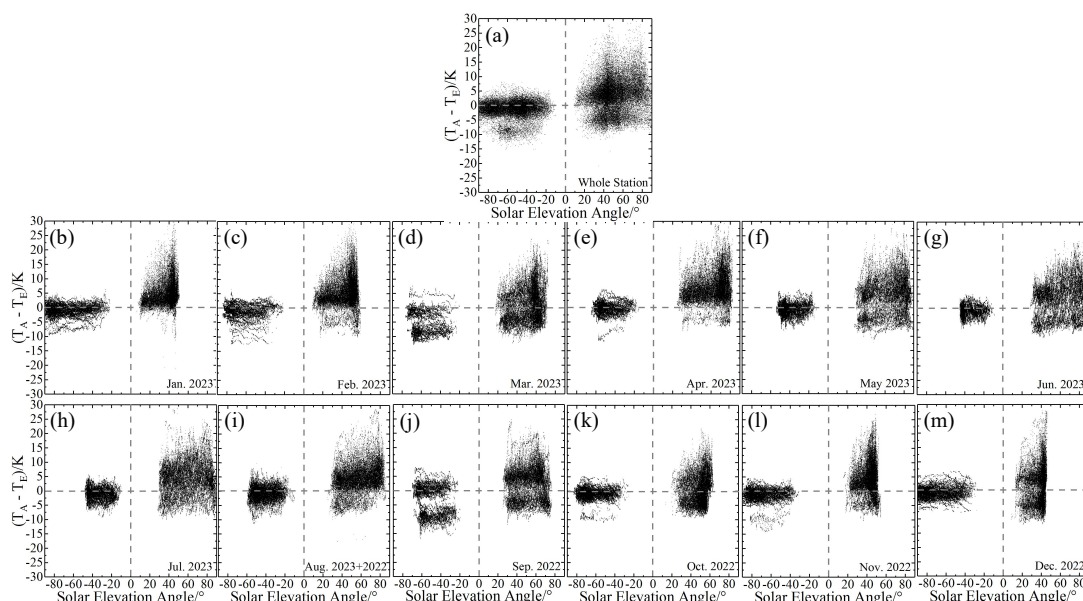
200 the SEA is greater than 0, the number of positive  $T_A - T_E$  values increases notably. As the SEA increases, the received solar radiation intensifies, and thus positive biases exhibit an increasing trend, generally remaining below 15 K and mainly concentrated in the range of 0–10 K. A secondary bias concentration band also exists between  $-7.5$  K and 0 K.

Figure 7 shows the probability distribution functions (PDFs) of  $T_A - T_E$  values before and after sunrise, mirroring the characteristics described above. Specifically, two peaks appear near  $-0.75$  K and  $-8$  K before sunrise, and two peaks are

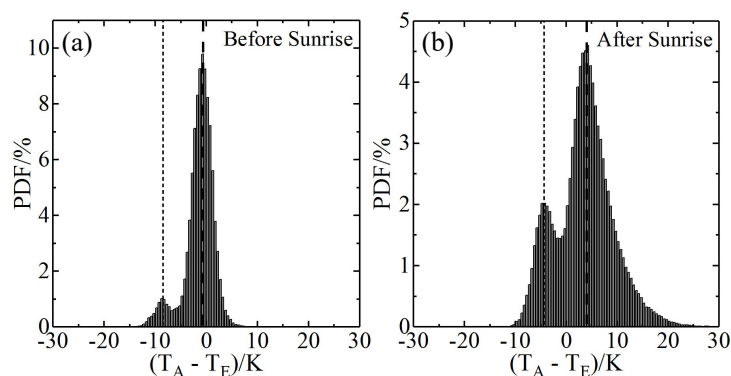
205 present near 4 K and  $-5$  K after sunrise. Notably, the two secondary peaks remain consistently in the negative range. The



variations of monthly  $T_A - T_E$  values with the SEA at all stations (Figs. 6b–m) are generally consistent with the variations of annual average biases (Fig. 6a). Before sunrise, the biases are near 0 K, with slightly more negative values. After sunrise, the biases are mainly positive, and with the increasing SEA, the positive biases tend to increase. At the maximum observed SEA, the positive biases mostly reach their maximum, and the scatter points become more concentrated. Some months show distinct secondary peak bands. From spring to winter, due to the seasonal shifting of sunrise and sunset times (sunrise times gradually advance and then delay, and sunset times gradually delay and then advance), the SEA corresponding to the beginning of the morning observation in the level-drift phase and that corresponding to the end of the afternoon observation in the level-drift phase show the variation pattern of first increasing and then gradually decreasing. Moreover, influenced by the seasonal cycle, the level-drift observations in the morning during winter and summer generally end when the SEA is around  $50^\circ$  and  $80^\circ$ , respectively. Therefore, two local maxima of positive biases appear near these two SEAs after sunrise.



**Figure 6:** Variations of the temperature biases between the ADDR5 level-drift observations and ERA5 data as a function of the solar elevation angle (SEA): (a) 13-month average at all stations, and (b–m) monthly variations at all stations from January to December. The gray dashed lines represent the positions where SEA =  $0^\circ$  and temperature bias = 0 K.





**Figure 7: Probability distribution functions (PDFs) of temperature biases in the level-drift phase (a) before and (b) after sunrise. The gray dashed lines represent the peak positions.**

### 3.3 Statistical characteristics of ADDRS temperature biases

Traditional statistical methods are employed to calculate the mean, standard deviation (SD), root mean square (RMS) of  $T_A - T_E$  values at the  $10^\circ$  interval of the SEA, as well as the SD of the observed temperature ( $O\_SD$ ) in the level-drift phase. Figure 8a shows the variations of these statistical metrics with the SEA, and Fig. 8b displays the sample sizes within each SEA bin. Notably, the sample size markedly drops to fewer than 1,000 within the SEA range of  $-10^\circ$  to  $0^\circ$ , and thus these small-sample data are excluded from the analysis. As shown in Fig. 8a, the mean temperature biases consistently exhibit weak negative values when the SEA is less than  $0^\circ$  (nighttime) and positive values when the SEA is greater than  $0^\circ$  (daytime). The positive biases gradually increase with the SEA, consistent with the results shown in Fig. 6. Two maxima of temperature biases appear near the SEAs of  $50^\circ$  and  $80^\circ$ . At an SEA of  $20^\circ$ , a third maximum is observed, where the influence of the secondary peak is minimal. The SD of temperature biases (solid lines) is smaller than that of the observed temperature ( $O\_SD$ ; dash-dotted lines). Previous studies have shown that in traditional statistical analysis, outliers and anomalies can substantially influence the mean and SD of a sample population, preventing the statistical results from accurately reflecting the true characteristics of the raw data. Therefore, a biweight algorithm is adopted to effectively reduce the influence of outliers (Zou et al., 2026). The biweight mean ( $\bar{X}$ ) and biweight standard deviation ( $D_{BST}$ ) are calculated using Eqs. (7)–(8).

$$\bar{X} = M + \frac{\sum_{i=1}^{n-1} (X_i - M) (1 - W_i^2)^2}{\sum_{i=1}^{n-1} (1 - W_i^2)^2}, \quad (7)$$

$$D_{BST}(X) = \left[ \frac{n \sum_{i=1}^{n-1} (X_i - M)^2 (1 - W_i^2)^4}{\sum_{i=1}^{n-1} (1 - W_i^2) (1 - 5W_i^2)} \right]^{1/2}, \quad (8)$$

In Eqs. (7)–(8),  $M$  denotes the sample median, and  $W_i$  represents the weighting factor, which is defined as

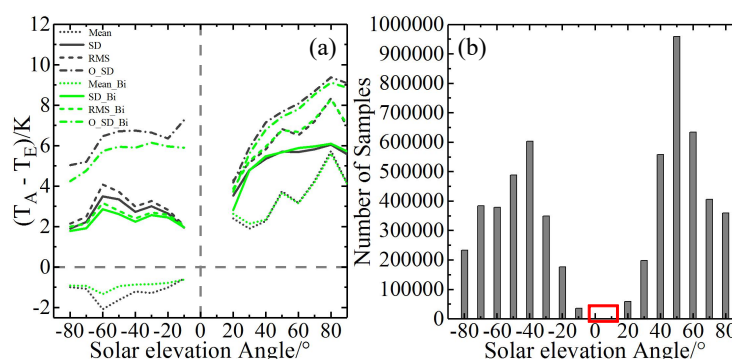
$$W_i = \frac{X_i - M}{7.5 M_{AD}}, \quad (9)$$

where  $M_{AD}$  indicates the median of  $|X_i - M|$ . When  $|W_i| > 1$ ,  $W_i$  is set to 1.

Statistical analysis based on the biweight algorithm (green lines in Fig. 8a) indicates that the biweight mean ( $Mean\_Bi$ ), biweight SD ( $SD\_Bi$ ), biweight RMS ( $RMS\_Bi$ ) of temperature biases, as well as the biweight SD of the observed temperature in the level-drift phase ( $O\_SD\_Bi$ ), all exhibit smoother variation trends, particularly when the SEA is less than 0. These biweight metrics effectively capture the true statistical characteristics of the original data. Specifically, the variation of  $Mean\_Bi$  with the SEA is basically consistent with that of mean temperature biases, but with a smoother overall curve. At night, due to the influence of secondary peaks, the  $Mean\_Bi$  (about  $-1.34$  K) is slightly pronounced at an SEA of  $-60^\circ$ , while it remains within  $-1$  K at the other SEAs. The overall nighttime  $Mean\_Bi$  is  $-0.91$  K. During daytime, the  $Mean\_Bi$  shows positive values and varies between  $2.14$  K and  $5.6$  K, with a mean value of  $3.48$  K. The  $RMS\_Bi$  ranges between  $2.0$  K and  $8.28$  K, and the average  $SD\_Bi$  is  $5.3$  K during daytime and  $2.3$  K at night, both of which are smaller than the  $O\_SD\_Bi$ .



According to the WMO guidelines for temperature uncertainty, the ideal target is 0.6–1 K, the breakthrough target is 1.8–2.8 K, and the “threshold” target is 6 K. Accordingly, the average SD\_Bi values during nighttime and daytime meet the WMO breakthrough target and the “threshold” target, respectively.



**Figure 8: (a) Variations of the statistics of temperature biases with the SEA, and (b) the sample sizes in the SEA bins. The black lines represent the statistics from the traditional statistical algorithm, and the green lines represent the statistics from the biweight algorithm (Bi). “Mean”, “SD” and “RMS” indicates mean, standard deviation, root mean square values of temperature biases, and “O\_SD” represents the standard deviation of level-drift temperature observations.**

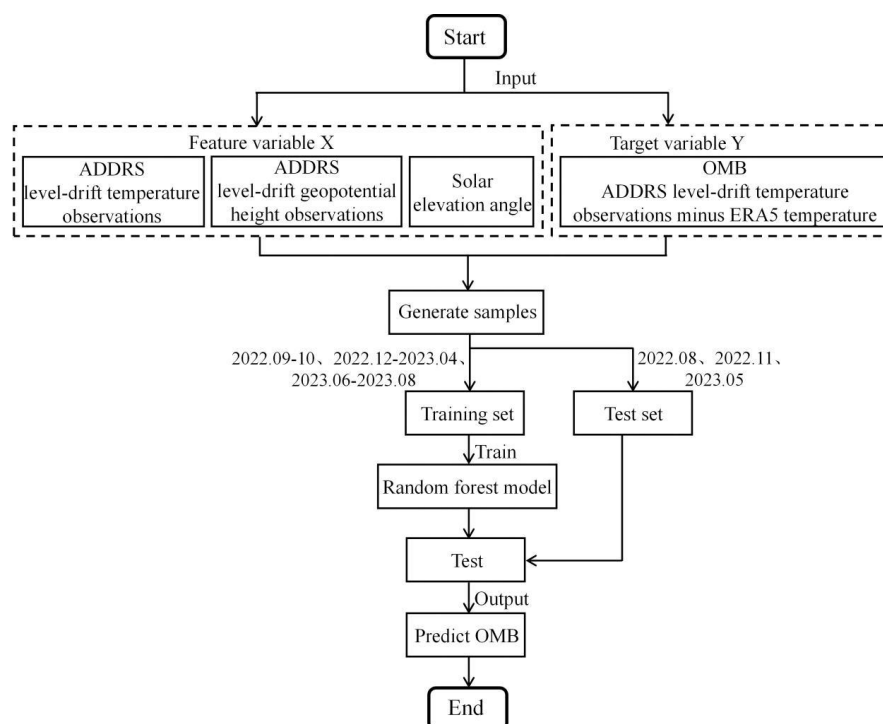
#### 4 Bias correction

Breiman (2001) proposed a random forest algorithm (statistical learning theory), whose basic building blocks are decision trees. Since decision trees are prone to overfitting, random forests combine multiple decision tree models to reduce overfitting and improve prediction accuracy. The process of establishing a random forest model involves several key steps. First, the Bootstrap resampling method is employed to extract multiple subsets from the original data. Second, a decision tree is built for each subset. Third, the Bagging method is applied to combine decision trees. Finally, the final prediction is determined through a voting mechanism (Fang et al., 2011; Xu et al., 2023; Zhao et al., 2022; Zhi et al., 2020). Random forest algorithms exhibit strong robustness to outliers and noise and do not suffer from overfitting with the increase in the number of decision trees, offering superior adaptability and accuracy compared with traditional methods such as support vector machines and single decision trees.

In this study, based on the random forest framework, a temperature bias correction model is constructed using the Scikit-learn module in Python (Fig. 9). The selection of feature variables and model parameters is crucial for model training. Through comparative analysis, the variables during the level-drift phase of ADDR5, i.e., temperature, geopotential height and the SEA, are selected as the input features of the model, and the temperature biases relative to the ERA5 data serve as the target variable. Following parameter tuning experiments, the optimal hyperparameters are determined as:  $n\_estimators=66$ ,  $max\_depth=11$ ,  $min\_samples\_split=30$  and  $min\_samples\_leaf=20$ . The remaining parameters are set to their defaults in the Scikit-learn module. The observations from August 2022, November 2022 and May 2023 are used as the



test set, while the observations from the remaining months are used as the training set. The random forest algorithm is employed to train the temperature bias correction model, and the prediction results from the model are verified using the test set. The calculated residual coefficient of determination, mean squared error, root mean square error and mean absolute error are 0.86, 3.55 K, 1.88 K and 1.48 K, respectively, indicating overall good model performance.

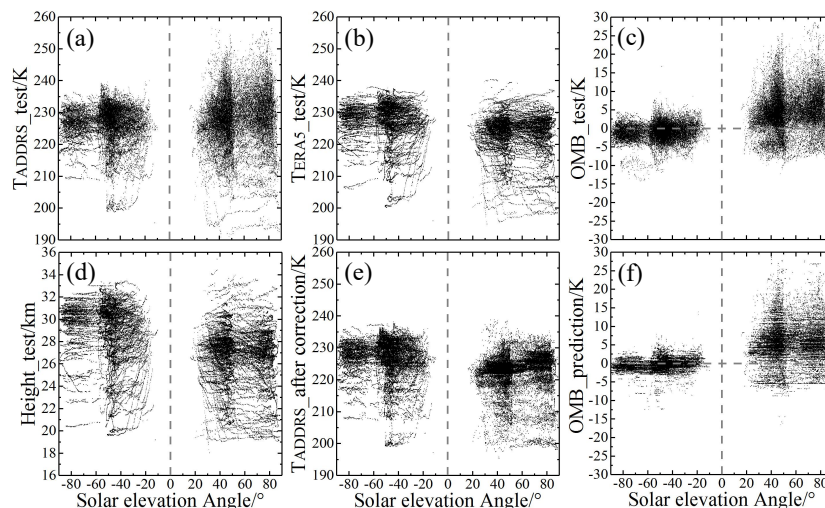


**Figure 9: Flowchart of temperature bias correction model. “OMB” indicates the ADDR5 level-drift temperature observation minus ERA5 temperature .**

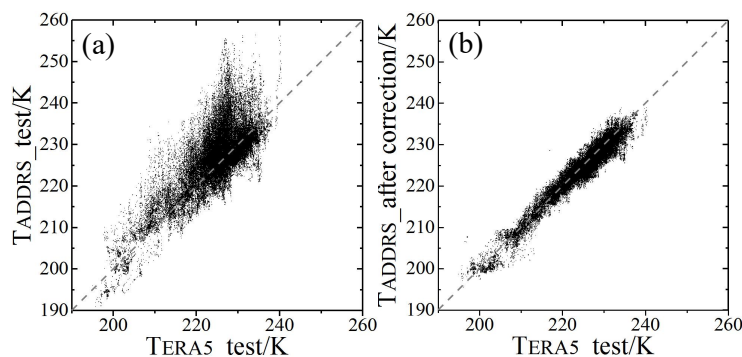
Comparing the temperature biases (OMB) from the model prediction (Fig. 10f) with those from the model test set (Fig. 10c), we find that their overall distributions are similar. Both show predominantly positive biases during daytime that increase with the SEA, alongside a secondary peak between  $-7.5$  K and 0 K. During nighttime, weak negative biases prevail. However, the model tends to overestimate daytime temperature biases and underestimate nighttime biases. Therefore, different weight coefficients are assigned to the model-predicted temperature biases to correct the level-drift temperature. The weight coefficients for the predicted negative and positive biases are set to 1.4 and 0.8, respectively. The corrected temperature in the level-drift phase (Fig. 10e) is calculated by subtracting the model-predicted temperature biases from the raw temperature in the test set (Fig. 10a). Compared with the observed level-drift temperature in the test set, the corrected temperature shows a distribution with respect to the SEA and a degree of data clustering that are much closer to those of the ERA5 reference temperature (Fig. 10b). The larger positive biases during daytime are effectively reduced. Figure 10d shows that the level-drift geopotential height during daytime is lower than that during nighttime. The comparisons of the test set



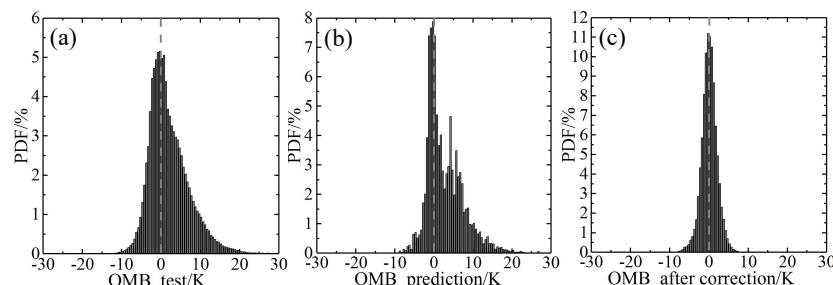
and corrected level-drift temperature with the ERA5 temperature (Fig. 11) indicate that the corrected level-drift temperature is closer to the ERA5 temperature than the test-set temperature. From the PDF of the temperature biases in the test set (Fig. 12a), it is evident that there are two peaks near  $-0.25$  K and  $4.75$  K. After correction, the PDF (Fig. 12c) more closely approximates a normal distribution, indicating the effectiveness of the bias correction model.



**Figure 10: Variations of various variables with the SEA: the level-drift temperature from the (a) test dataset and (b) ERA5 dataset, the (c) temperature biases (OMB) and (d) geopotential height from the test dataset, (e) the corrected level-drift temperature, and (f) the OMB from the prediction set. The gray dashed lines represent the positions where SEA =  $0^\circ$  and temperature bias =  $0$  K.**



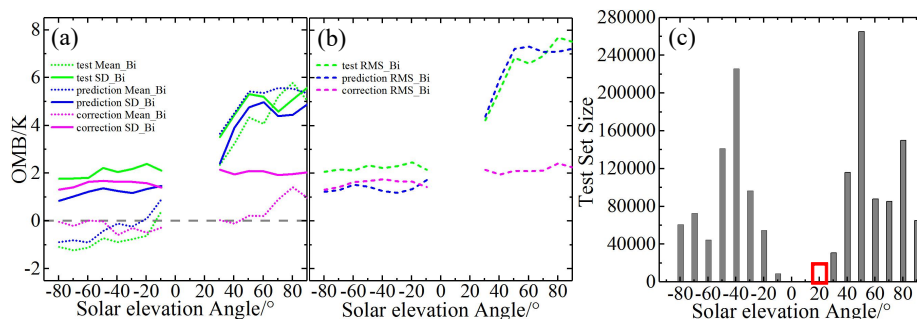
**Figure 11: Comparisons of the level-drift temperature from the (a) test set and (b) corrected results with the ERA5 temperature.**



**Figure 12: PDF distributions of the OMB values from the (a) test, (b) prediction set and (c) corrected results.**



Furthermore, statistical evaluations of the temperature biases from the test set, prediction set and corrected results are conducted at  $10^\circ$  SEA intervals using the biweight algorithm. Figure 13c displays the sample size within each  $10^\circ$  SEA bin for the test set, and the overall variation trend of the sample size is consistent with that of all samples shown in Fig. 8b. Therefore, small-sample data ( $<1000$ ) at the  $20^\circ$  SEA are also excluded. The biweight means of the OMB from the prediction and test sets show consistent variation trends with the SEA, characterized by predominantly positive values during daytime and negative values during nighttime, again indicating that the prediction set underestimates the negative biases and overestimates the positive biases. For the prediction set, the biweight mean values of the daytime and nighttime OMB range from 3.65 K to 5.56 K and from  $-0.91$  K to  $0.91$  K, respectively; the biweight SD values range from 2.4 K to 4.98 K and from 0.85 K to 1.46 K; and the biweight RMS values range from 4.37 K to 7.31 K and from 1.19 K to 1.72 K. After correction, the biweight mean, biweight SD and biweight RMS curves of the OMB become notably smoother. The averages of the biweight mean values are 0.52 K during daytime and  $-0.23$  K during nighttime, those of the biweight RMS values are 2.15 K during daytime and 1.57 K during nighttime, and those of the biweight SD values are 2.03 K during daytime and 1.54 K during nighttime. All these average values meet the WMO temperature breakthrough target, with the nighttime metrics far exceeding the breakthrough target. These quantitative results further support the validity of the bias correction.



**Figure 13: (a, b) Variations of the biweight statistics of the temperature biases with the SEA, and (c) the sample sizes for the test set in the SEA bins. The green, blue and pink lines represent the statistics from the test set, the prediction set and the corrected results, respectively.**

## 5 Attribution analysis

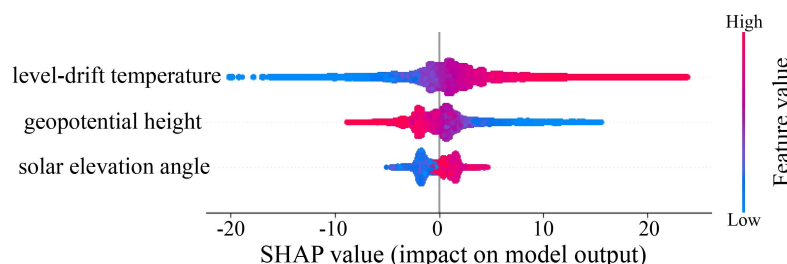
To further interpret the contribution of each feature variable to model predictions, the SHapley Additive exPlanations (SHAP) method is employed through the FastSHAP tool. Figure 14 presents a summary plot of the SHAP value distributions for all feature variables. Each point represents an individual sample, and the magnitude of SHAP values indicates the degree of influence of a given feature on the model output. A larger absolute SHAP value signifies a greater influence on model predictions and thus higher feature importance. Specifically, the ranking of feature importance for the bias correction model is as follows: level-drift temperature  $>$  geopotential height  $>$  the SEA. A positive SHAP value indicates that the feature contributes positively to the target prediction, while a negative SHAP value indicates the opposite.



The results indicate that the higher the level-drift temperature, the larger the predicted values. Conversely, lower level-drift temperatures correspond to lower predicted values. The SHAP values of the SEA exhibit a roughly symmetrical distribution around zero, and the distributions of high and low feature values are also largely symmetric. This indicates that the high and low SEAs exert opposite influences on the predictions, i.e., the SHAP values are positive during daytime and negative at night. This result is consistent with the previous analysis, where solar radiative heating leads to positive temperature biases during daytime and negative biases appear at night.

Moreover, lower geopotential height values correspond to positive SHAP values, while higher geopotential height values correspond to negative SHAP values. This implies that lower geopotential height tends to cause positive biases, and higher geopotential height generates negative biases. This result aligns with the previously observed scatter distribution of geopotential height versus the SEA. Specifically, during daytime, the geopotential height is lower in the level-drift phase, while at night, it is higher.

Consequently, the influence of each feature variable in the bias correction model on the predictions aligns with physical expectations, further substantiating the rationality of the model.



**Figure 14. Distribution of the SHapley Additive exPlanations values for each feature variable in the temperature bias correction model.**

## 6 Summary and conclusions

In this study, a detailed quality evaluation and analysis of ADDRS level-drift temperature data were conducted using ERA5 reanalysis fields as a reference. Furthermore, a temperature bias correction model for ADDRS level-drift observations was established based on the random forest method. The main conclusions are as follows.

Overall, comparison of temperature data between ADDRS level-drift observations and ERA5 shows that morning level-drift temperature exhibit mainly positive biases. In contrast, afternoon observations show weak negative biases and correlate more strongly with ERA5 temperature.

By further introducing the SEA, analysis of the temperature biases between ADDRS level-drift observations and ERA5 data ( $T_A - T_E$ ) reveals that the biases predominantly exhibit negative during nighttime ( $-5$  K to  $0$  K) and positive during daytime ( $0$ – $10$  K). As SEA increases, the enhanced solar radiation leads to an increasing trend in the temperature biases. Additionally, a secondary negative-bias band is observed both during daytime and nighttime.



To mitigate the influence of outliers on statistical results and better reflect the true characteristics of the original data, both the traditional statistical method and the biweight algorithm were used to analyze temperature biases between ADDRS level-drift observations and ERA5 data. The results indicate that for daytime and nighttime temperature biases, the biweight RMS values are 6.38 K and 2.48 K, the biweight mean values are 3.48 K and  $-0.91$  K, and the biweight SD values are 5.3 K and 2.3 K, respectively. According to the WMO guidelines on temperature uncertainty, the daytime biweight SD (5.3 K) meets the WMO “threshold” target (6 K), and the nighttime biweight SD (2.3 K) meets the breakthrough target (1.8–2.8 K). A temperature bias correction model for ADDRS level-drift observations was developed based on a random forest algorithm. Both qualitative and quantitative evaluations demonstrate that the model is rational, reliable, and effective. After bias correction, the biweight mean of temperature biases decreases from  $-0.75$  K to  $-0.23$  K at night and from 4.29 K to 0.52 K during the day. The biweight RMS decreases from 2.23 K to 1.57 K at night and from 6.47 K to 2.15 K during the day. The biweight SD decreases from 2.03 K to 1.54 K at night and from 4.82 K to 2.03 K during the day. The temperature uncertainties for both daytime and nighttime satisfy the WMO temperature breakthrough target, with nighttime metrics well exceeding this target.

Attribution analysis indicates that the feature importance ranking for the bias correction model is as follows: level-drift temperature > geopotential height > SEA. Both level-drift temperature and SEA exert a positive influence on the predictions at high values and a negative influence at low values, confirming that the positive temperature biases observed during daytime are primarily caused by solar radiative heating.

Moreover, temperature biases between ADDRS level-drift observations and the ERA5 data exhibit secondary peaks in the ranges of  $-8$  K to  $-12$  K at night and 0 K to  $-7.5$  K during the day. The exact causes of these secondary peak bands remain undetermined, and further research will focus on identifying their origin and mitigating their impacts. As the current bias correction model was trained on 10 months of data and tested on 3 months, statistical analysis indicates overfitting. Therefore, future work should incorporate larger datasets to optimize the temperature bias correction model, thereby facilitating more effective utilization of temperature observations from the level-drift phase of ADDRS.

*Code and data availability.* The datasets used and/or analyzed in the research are available from the corresponding author (Qiyun GUO (feng\_cloudy@163.com) and Xin SUN (sunx\_beht@163.com)) on reasonable request.

*Author contributions.* QG and JW conceived and designed the study. QG reveals the key physical influencing factors of temperature bias in ADDRS level-drift phase. XS supervised the project and developed the bias correction methodology. XY conducted the data and model analysis with contributions from YJ, LL, DW, FZ and KL. XY and JW wrote the manuscript with input from all authors.

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



390 *Disclaimer:* Publisher's note: Copernicus Publications remains neutral with regard to jurisdictional claims made in the text, published maps, institutional affiliations, or any other geographical representation in this paper. The authors bear the ultimate responsibility for providing appropriate place names. Views expressed in the text are those of the authors and do not necessarily reflect the views of the publisher.

395 *Acknowledgments.* The authors are thankful to the editor and all anonymous reviewers for their help in improving this manuscript. This research was supported by the National Natural Science Foundation of China (42375150), Inner Mongolia Autonomous Region Natural Science Foundation General Joint Fund (2025LHMS04007) and Open Project of the State Key Laboratory of Severe Weather Science and Technology (2025KFA20).

## References

- Breiman, L.: Random forest. *Machine Learning*, 45, 5-32, <https://doi.org/10.1023/A:1010933404324>, 2001.
- Chen, M., Fan, S. Y., Zheng, Z. F., and Zhong, J. Q.: The performance of the proximity sounding based on the BJ-RUC  
400 system and its preliminary implementation in the potential forecast, *Acta. Meteor. Sinica.*, 69, 181-194, <https://doi.org/10.11676/qxxb2011.016>, 2011. (in Chinese)
- Cao, X. Z., Guo, Q. Y., and Yang, R. K.: Research of rising and falling twice sounding based on long-time interval of flat-floating, *Chinese J. Sci. Instrum.*, 40, 198-204, <https://doi.org/10.19650/j.cnki.cjsi.J1803748>, 2019. (in Chinese)
- Cao, X. Z., Guo, Q. Y., Luo, H. W., Wang, J. C., Yang, R. K., Xiao, D., Liu, Y. F., Sun, Z. L., Liu, S. J., Chen, S. J., Huang,  
405 A. F., Guo, J. P., and Zhang, P.: Development and application of the Ascent-Drift-Descent Radiosonde System (ADDRS), *Atmos. Meas. Tech.*, 19, 3231-3251, <https://doi.org/10.5194/amt-19-3231-2026>, 2026.
- Fang, K. N., Wu, J. B., Zhu, J. P., and Xie, B. C.: A review of technologies on random forests, *Statistics & Information Forum*, 26, 32-38, <https://doi.org/10.3969/j.issn.1007-3116.2011.03.006>, 2011. (in Chinese)
- Faccani, C., Rabier, F., and Fourrié, N.: The impacts of AMMA radiosonde data on the French global assimilation and  
410 forecast system, *Wea. Forecasting*, 24, 1268-1286, <https://doi.org/10.1175/2009waf2222237.1>, 2009.
- Guo, Q. Y., Yang, R. K., Qian, Y., Cao, X. Z., and Lei, Y.: Full-range sounding comparison analysis of ballon borne radiosonde rising and parachute carrying radiosonde descending, *Meteorological Monthly*, 44, 1094-1103, <https://doi.org/10.7519/j.issn.1000-0526.2018.08.011>, 2018. (in Chinese)
- Hao, M., Tian, W. H., and Gong, J. D.: Study of L-band second-level radiosonde data applied in GRAPES assimilation  
415 system, *Meteorological Monthly*, 40, 158-165, <https://doi.org/10.7519/j.issn.1000-0526.2014.02.20140303>, 2014. (in Chinese)
- He, W. Y., Ho, S. P., Chen, H., Zhou X., Hunt, D., and Kuo, Y. H.: Assessment of radiosonde temperature measurements in the upper troposphere and lower stratosphere using COSMIC radio occultation data, *Geophysical Research Letters*, 36, 1397-1413, <https://doi.org/10.1029/2009GL038712>, 2009.



- 420 Liu, S. J., Yang, R. K., Cao, X. Z., Guo, Q. Y., Cheng, K. Q., Shen, Z. P., and Wang, J. C.: Analysis and numerical experiment of the horizontal drift round-trip sounding balloon's dynamic and thermal process in the adjacent space, Chinese Journal of Atmospheric Sciences, 46, 17, <https://doi.org/10.3878/j.issn.1006-9895.2110.20252>, 2022. (in Chinese)
- Qian, Y.: Quality control and evaluation of round-trip flat drift sounding data, M.S. thesis, School of atmospheric sciences, Nanjing University of information science & technology, 79pp, <https://doi.org/10.27248/d.cnki.gnjqc.2019.000327>, 2019. (in Chinese)
- 425 Rong, N., Yang, S. P., Wang, J. C., and Wang, D.: The evaluation of return radiosonde temperature using GNSS radio occultation retrieval temperature, Plateau Meteorol., 42, 221-232, <https://doi.org/10.7522/j.issn.1000-0534>, 2023. (in Chinese)
- 430 Wang, R.W., Wang, J. C., Wang, D., Tao, T. Y., and Tian, W. H.: Study on the influence of return sounding observation system based on CMA-MESO, Meteorol. Monthly, 49, 52-61, <https://doi.org/10.7519/j.issn.1000-0526.2022.032601>, 2023. (in Chinese)
- Wang, J. C., Wang, D., Yang, R. K., Cao, X. Z., and Guo, Q. Y.: A return radiosonde trajectory forecast method and its preliminary evaluation based on high resolution numerical weather prediction model, Chinese J. Atmos. Sci., 45, 651–663, <https://doi.org/10.3878/j.issn.1006-9895.2012.20186>, 2021. (in Chinese)
- 435 Wang, D., Wang, J. C., Tian, W. H., and Guo, Q. Y.: Quality control and uncertainty analysis of return radiosonde data, Chinese J. Atmos. Sci., 44, 865-884, <https://doi.org/10.3878/j.issn.1006-9895.1912.19203>, 2020. (in Chinese)
- Xu, J. F., Song, L. Y., Chen, M. X., Yang, L., and Han, L.: Comparative machine learning-based correction experiment for a 10m wind speed forecast at a 100m resolution in complex mountainous areas of the winter olympic games, Chinese Journal of Atmospheric Sciences, 47, 805-824, <https://doi.org/10.3878/j.issn.1006-9895.2209.22117>, 2023. (in Chinese)
- 440 Yang, C. Y., Cao, X. Z., Guo, Q. Y., and Yuan, Y.: Feature extraction and analysis of atmospheric turbulence based on new round-trip radiosonde, Chinese Journal of Atmospheric Sciences, 47, 1967-1982, <https://doi.org/10.3878/j.issn.1006-9895.2202.21103>, 2023. (in Chinese)
- Zhou, X.S., Guo, Q. Y., Xia, Y. C., and Tian, H.: Inspection of FY-3D satellite temperature data based on horizontal drift round-trip sounding data, J. Appl. Meteorol. Sci., 34, 52-64, <https://doi.org/10.11898/1001-7313.202301105>, 2023. (in Chinese)
- 445 Zhang, X. P.: Research on data assimilation and targeted observation of novel round-trip level-flight sounding data, M.S. thesis, School of atmospheric sciences, Nanjing University of information science & technology, <https://doi.org/10.27248/d.cnki.gnjqc.2022.000203>, 2022. (in Chinese)



- 450 Zhang, X.P., Guo, Q.Y., Yang R. K., Ma, X. L., and Cao, X. Z.: Assimilation experiment of rainstorm in the middle and  
lower reaches of the Yangtze river based on “Up-Drift-Down” sounding data, *Meteorol. Monthly*, 47, 1512-1524,  
<https://doi.org/10.7519/j.issn.1000-0526.2021.12.007>, 2021. (in Chinese)
- Zou, X. L., Zeng, Z.: A quality control procedure for GPS radio occultation data, *Journal of Geophysical Research*  
*Atmospheres*, 111, <https://doi.org/10.1029/2005JD005846>, 2026.
- 455 Zhao, S. H., Tan, B. Q., and Du, L.: Warning method of lightning nowcasting based on machine learning and single station  
ground meteorological elements, *Meteorological Science And Technology*, 50, 121-128,  
<https://doi.org/10.19517/j.1671-6345.20210210>, 2022. (in Chinese)
- Zhi, X. F., Wang, T., and Ji, Y.: Multimodel ensemble forecasts of surface air temperature over China based on deep  
learning approach, *Trans Atmos Sci*, 43, 435-446, <https://doi.org/10.13878/j.cnki.dqkxxb.20200219003>, 2020. (in  
460 Chinese)