Journal of the Meteorological Society of Japan

EARLY ONLINE RELEASE

The DOI for this manuscript is
DOI:10.2151/jmsj.2025-029
J-STAGE Advance published date: June 3, 2025
The final manuscript after publication will replace the preliminary version at the above DOI once it is available.

Data quality control and calibration for mini-radiosonde system “Storm Tracker” in Taiwan

Hung-Chi KUO

Department of Atmospheric Sciences, National Taiwan University, Taipei, Taiwan

Ting-Shuo YO

Department of Atmospheric Sciences, National Taiwan University, Taipei, Taiwan

Hungjui YU

Department of Atmospheric Science, Colorado State University, Fort Collins, Colorado, USA

Shih-Hao SU

Department of Atmospheric Sciences, Chinese Culture University, Taipei, Taiwan

Ching-Hwang LIU

Department of Atmospheric Sciences, Chinese Culture University, Taipei, Taiwan

and

Po-Hsiung LIN

Department of Atmospheric Sciences, National Taiwan University, Taipei, Taiwan

March 5, 2025

----

1) Corresponding author: Ting-Shuo Yo, Department of Atmospheric Sciences, National Taiwan University, No. 1, Section 4, Roosevelt Rd, Da’an District, Taipei City, 10617, TAIWAN
Email: tsyo@ntu.edu.tw
Tel: +886-975-157-101

Abstract

This study reports the correction methods of a newly introduced upper-air radiosonde instrument, "Storm Tracker" (ST), with more than one thousand co-launches of ST and Vaisala RS41-SGP (VS) data in field observations in the Taiwan area during 2016–2022. The co-launches provided more than a million comparable independent observations of wind, pressure, temperature, and humidity (PTU) data. Using the VS measurements as the reference, we use the statistical models, including the cumulative distribution function (CDF) matching method and generalized linear model (GLM), to correct the temperature and moisture fields of the ST sounding. Both approaches yield similar results. With a sounding-by-sounding comparison, the CDF-corrected ST soundings have a 1-K temperature and 7% relative humidity root mean square difference from the VS soundings. These error differences can be reduced to 0.66-K and 4.61% below the 700-hPa height. The GPS estimated a 0.05 ms^{-1} ST wind difference from the VS sounding. The biases of the corrected ST observations are slightly larger than the random errors, which were 0.24 K and 2.21% in the laboratory and 0.52 K and 2.23% in the field. The lower atmosphere in a region of complex terrain may have large wind, temperature, and moisture variations. With the relatively low cost, a high proportion of successful launches, and accuracy of wind, temperature, and moisture, ST can complement regular upper-air radiosonde observations for high-resolution observations in the lower troposphere. The high-

resolution lower troposphere observation is important for severe weather research in East Asia.

Keywords boundary layer; upper-air radiosonde observation; field campaign; data quality control and correction

1. Introduction

Upper-air radiosondes are one of the most important meteorological instruments for observing vertical profiles of atmospheric data at various altitudes. The measured pressure, temperature, and relative humidity (so-called “PTU”) data aids in weather forecasting, climate research, and the study of atmospheric dynamics. However, upper-air radiosondes are subject to certain biases due to instrument calibration, ascent rates, and environmental conditions. Collins (2001) distinguished the radiosonde observational errors into three types: random, rough, and systematic. According to Collins (2001), random error is caused by small-scale turbulence or unsystematic observational errors, and it is impossible to correct. The rough error can be introduced from observational protocol, computational error for data processing, or communication-related error. A properly defined operational procedure and automatic quality control process can minimize such errors. The third type of error, systematic error, is caused by insufficiencies in measurement devices or data processing procedures and persists in all observational data. This type of error can be detected and calibrated with statistical methods.

Nowadays, commercial radiosondes are often tested and corrected regarding these biases. However, they are typically characterized by their higher weight and cost, which limit the deployment of scientific field campaigns. The independently developed mini-radiosonde system – the “Storm Tracker (referred to as the ST, Figure 1b)” was created and first tested in 2016 (Hwang et al., 2020). The ST consists of a microcontroller

(ATMEGA328p), a GPS sensor (U-blox MAX7-Q), a pressure sensor (Bosch BMP280), a temperature–humidity sensor (TE-Connectivity HTU21D), and a transmitter (LoRa™). The sensors have an overall operation range from 1100 to 300 hPa in pressure and from -40℃ to 85℃ in temperature. The ST used a regular AAA battery for 2-4 hours of power; the total weight was 20g. More detailed hardware specifications can be found in Hwang et al., 2020. The design of ST aimed to leverage the low cost of sensors used in commercial electronics to enable high-frequency observations in the boundary layer. In addition, the receiver was designed to receive up to ten STs simultaneously. With such agility, using ST to gather supplemental data between regular sounding was ideal.

The ST was then put into intensive field observation operations for the first time during the Taipei Summer Storm Experiment (TASSE) in 2018–2020. The main goal of the field campaign is to investigate the thermal characteristics of the boundary layer in the Taipei Basin and local wind field variations to improve the forecasting ability of afternoon convection in the metropolitan area. Three advantages of using the ST for atmospheric field research were learned. First, the weight of ST with a battery is 20g, which helps to reduce the helium/hydrogen usage. Second, the commercial sensors, chips, and signal transmission components in the ST significantly reduce the cost and provide flexibility for multiple deployments and high spatial and temporal resolution observations. Lastly, the ST is easy to set up and can be quickly deployed or even mobile, which provides adaptability for different research needs and broadens the possibility for field campaign design.

The early work of ST by Hwang and colleagues (2020) showed an overall warm and dry bias in the troposphere compared to the VS, as shown in their figure 13. During TASSE, we discovered similar bias patterns and a typical example is shown in Figure 2. These common bias patterns motivated us to design a systematic approach to improve the data quality of ST. Our correction methods seek to align ST measurements as closely as possible with VS, enabling researchers to perform high-frequency, high-spatial resolution observations using ST with greater confidence and accuracy.

Many prior studies have recognized these biases and suggested that solar radiation can induce warm and dry bias for radiosonde measurements (Vömel et al. 2007). Similar daytime warm and dry biases have been reported in previous field experiments around the world that used relatively mature radiosonde systems (e.g., Wang et al., 2002; Ciesielski et al., 2009; Yu et al., 2015). Earlier studies indicated that radiosonde temperature biases are primarily contributed by radiative effects, with a minor proportion caused by the sensor response lag of the changing of temperatures as the radiosonde rises (e.g., McMillin et al., 1992; Sun et al., 2013).

The daytime temperature bias induced by solar heating was identified with various radiosonde systems (e.g., Luers, 1989, 1997; Luers and Eskridge, 1998; Sun et al., 2013; Lee et al., 2022; von Rohden et al., 2022). Their findings resulted in special surface coating over temperature sensors in most commercial radiosondes. Even though environmental parameters can still affect the observed temperature, all factors influencing radiative or sensible heat flux around the sensor, such as the sensor surface temperature, solar angle, cloud fraction, and ventilation velocity, can cause the sensor temperature bias (e.g., McMillin et al., 1992; Luers and Eskridge, 1995; Mattioli et al., 2007; Lee et al., 2022). Luers and Eskridge (1998) evaluated the impact of the environmental parameters on the radiosonde in detail. Their results suggested that the temperature bias is most sensitive to solar angle, while the cloud cover has a slight impact. Also, the ventilation effect may cause bias when the sensor is in the balloon wake zone. To study the source of bias under a controlled environment, von Rohden et al. (2022) presented the Simulator for the Investigation of Solar Temperature Error of Radiosondes (SISTER), and Lee et al. (2022) proposed the Upper Air Simulator (UAS), which allows precise control over temperature, pressure, ventilation, and irradiation. Such advanced setups can help researchers measure the measurement errors more accurately and verify the cause of the errors.

In addition to temperature bias, the humidity bias has been discussed in many studies (e.g., Vömel et al., 2007; Yoneyama et al., 2008; Nuret et al., 2008; Kizu et al., 2018; Lee et al., 2022; Sommer et al., 2023). Vömel et al. (2007) found that the solar-heating-induced dry bias increased with altitude in the troposphere, which means the humidity bias also depended on the temperature. This resulted in the relative humidity (RH) measured in the low-temperature environment being less accurate (Miloshevich et al., 2001). Miloshevich et al. (2004) also pointed out that the response delay in humidity sensors could cause measurement errors at low temperatures. The influence of these biases could be huge. For example, in the Tropical Ocean Global Atmosphere Coupled Ocean-Atmosphere Response Experiment (TOGA COARE,1992-1993), scientists have reported the observational error induced an unrealistically dry boundary layer and caused an underestimate of convective available potential energy (CAPE) (Miller et al., 1999; Lucas and Zipser, 2000). Although the primary observation targets of ST are the lower troposphere environmental conditions, we still noticed significant warm and dry deviations in the near-surface boundary layer in TASSE (Figure 2).

Many studies have attempted to remedy the systematic error in radiosonde data with statistical methods. Lesht and Richardson (2002) mentioned that Vaisala accounts for the sensitivity of the RH sensor to temperature by using a high-order polynomial function with empirical coefficients. Yoneyama et al. (2008) applied a polynomial fitting function of pressure for the relative difference of RH and used the solar zenith angle as a factor for bias corrections. Other studies leveraged the thermodynamic equation and provided the temperature correction table with empirical correction factors (Wang et al., 2013; Dzambo et al., 2016).

In past field campaigns, scientists have also developed statistical models of humidity correction based on probability matching. For example, Ciesielski et al. (2009) used the cumulative distribution function (CDF) matching method to correct the humidity bias for nearby soundings. The advantage of the CDF-based calibration method is that the calibration procedure is fast and straightforward. Building the correction table requires sufficient data to represent the statistical characteristics and questionable data can be adjusted to match the same distribution. The basic concept of the CDF matching calibration method is assuming the ambient atmospheric conditions are similar for all observation sites. In most field campaigns, the spatial distribution of upper-air radiosonde sites mostly satisfied such requirements, and hence, this method can efficiently adjust the data bias for most atmospheric conditions. However, such assumptions limit the generalizability of the CDF calibration models. Thus, the CDF models may not be directly applied to the data collected from different weather conditions, seasons, or climate regions with smaller sample sizes.

In this study, we focused on the calibration process of systematic error for ST temperature and moisture observations using the co-launch VS data. We use the co-launch data collected across several field campaigns in Taiwan to develop calibration methods for ST. Here, we proposed and evaluated two different calibration approaches. First, we followed the widely used CDF-matching approach and proposed a two-step CDF-based calibration scheme. Secondly, we incorporated the CDF-matching approach with modeling multivariate distributions, the central concept of machine learning, to introduce a novel correction method based on the generalized linear model (GLM). While the CDF approach discretized continuous variables, e.g., pressure and temperature, into bins to establish look-up tables, the machine-learning approach modeled a high-dimensional joint probability distribution with the same variables in their original forms. The latter approach allowed us to compress complicated look-up tables into a unified mathematical representation. Hence, we can adjust the models more easily for better performance, robustness, and generalizability.

Section 2 describes the co-launched radiosonde data and the pre-processing. Section 3 focuses on the data correction algorithms, and data calibration processing flow. Section 4 summarizes the ST calibration results and compares them to the benchmark. Finally, Section 5 discusses the feature importance analysis and other calibration issues, and Section 6 presents the conclusions.

2. Data and Preprocessing

2.1 Data Collection

In the previous years since 2018, we have co-launched the ST with the Central Weather Administration (CWA) operational Vaisala RS41-SGP radiosonde (Figure 1c). The co-launch was conducted during field campaigns in the Taiwan area, including the Taipei Summer Storm Experiment (TASSE), the Yilan Experiment of Severe Rainfall (YESR2020), the Taiwan-Area Heavy Rain Observation and Prediction Experiment (TAHOPE), the Northern Coast Observation, Verification, and Investigation of Dynamics (NoCOVID21), and the Mountain Cloud Climatology (MCC) project, We collected 1,029 co-launches of ST and VS from these field campaigns during 2018–2022. These co-launches provided more than 1,000,000 comparable independent observations of wind, pressure, temperature, and humidity (PTU) data. The co-launches of each campaign are summarized in Table 1, and the geographic locations of the co-launch sites are shown in Figure 3.

In 2018 and 2019, based on the scientific goals of TASSE, we established a standardized procedure for the co-launches, and the observations were primarily conducted in the daytime. Once the observational procedure matured, we performed the day and night co-launches evenly in 2020, 2021, and 2022. (Table 2). Eventually, we collected 625 daytime cases and 404 nighttime cases. Also, the pilot experiments were conducted in the summer, and in the latter field experiments, we performed the co-launches in other months. Though there were more cases in July and August, we still conducted at least 21 co-launches in May. As for the location, most co-launches were conducted at the Taipei weather station, while about 150 cases were in other cities in Taiwan. In these 1,029 co-launches, all STs successfully launched, and only 7 stopped sending signals after 300 seconds. The ST, designed with commercial hardware components, is reliable in field observations.

Note that the binding of ST and VS shown in Figure 1c differs from the instruments used in the Report of WMO's 2022 Upper-Air Instrument Intercomparison Campaign (IOM-143). The IOM-143 can be categorized into in-laboratory and in-field campaigns. The laboratory calibration techniques focus on understanding each instrument's characteristics regarding random errors, low-temperature performance, and solar radiation sensitivity. The field campaign calibration techniques emphasize ground checks. A major goal is to evaluate the observation difference between the radiosonde systems, including the VS.

The IOM-143 used a rig to hold multiple instruments together while avoiding interferences from ventilation and signals. Accordingly, the simple binding in our study may increase the random difference between ST and VS. However, this study aims to develop correction methods for ST to behave as close to VS as possible. Our simple binding co-launches in a consistent manner for several years are the only data we have. As presented in the following session, the biases of the corrected ST observations are slightly larger than the random errors. Hence, we used a relatively simple binding design in the co-launches before 2023. Future binding co-launches will be conducted according to the WMO standard.

2.2 Pre-processing of the co-launch data

The ST is with the wind estimated from GPS. We analyzed the difference in wind variables with the paired data of VS and ST. The mean deviation in zonal and meridional wind components, u and v, are 0.04 and 0.03 ms-1, respectively. The difference may come from the time lag of GPS signals between two sensors, which is small enough to ignore. In this paper, we emphasize the correction of temperature and humidity calibration.

The co-launch's primary purpose is to understand ST's performance further and develop a data correction scheme to approximate the VS's observations. The raw data collected often contains inconsistencies, inaccuracies, and outliers that can significantly distort analytical results and impede the accuracy of predictive modeling. Therefore, we need a proper procedure to process the raw data.

In the work of Ciesielski et al. (2012), the authors suggested four stages for developing research-quality radiosonde data (their figure 1). The first level requires a single unified data format. The second stage uses automated tools to remove unreliable data based on prior knowledge of quality control (QC) checks. Then, data biases are detected and corrected in the third level based on analysis or statistical methods. Finally, the fourth level dataset aims to be user-friendly, usually in uniform vertical resolution with QC flags.

Following the framework proposed by Ciesielski et al. (2012), our data correction method is applied in the third stage. Hence, we need a pre-processing scheme to derive a level 2 dataset from the raw co-launch data.

Figure 4 illustrates the preprocessing used in this study. In the first stage, we paired each ST and VS observation by nominal observation time and stored them in the same plain-text format, L1_ST and L1_VS. Then, in the second stage, we corrected known errors for both sensors, including missing values and outliers. After this stage, we derived the level 2 dataset, L2_ST and L2_VS. Finally, given the fact that both ST and VS radiosondes are attached during co-launch (as Figure 1c), we used “time after launch” (every second) in both profiles to pair the values of two sensors, and resulted in L2_ST-VS.

Based on the prior studies of ST (Hwang et al., 2020), we performed a “ground check” procedure to correct the pressure values of ST. This procedure adjusts the P_ST by a constant bias dP_0, which is the difference between the surface pressure of the standard instrument and the sensor of ST. Furthermore, we filtered out profiles with inconsistent timestamps and paired records less than 250 (366 out of 1,029). Finally, we derived a dataset of 663 merged profiles and 1,219,710 paired entries (up to every second) for further analysis.

3. Data Correction Methods

To develop a data correction scheme for ST, we first investigated the conventional CDF-based probability matching method (Ciesielski et al., 2009). Then, we extended this approach with direct modeling of multivariate distributions, which is the central concept of modern machine learning. We implemented the scheme with the basic generalized linear model (GLM) and compared the differences between the two approaches. Both CDF and GLM are simple statistical models. The CDF is based on a non-parametric approach, and the GLM is a parametric distribution (i.e., Gaussian distribution).

Before diving into the specific correction methods, we define the notations and symbols used in this study. While ST and VS represent the storm tracker and the Vaisala RS41-SGP radiosonde device, respectively, they are used as subscripts to denote the sensor of measurements. For example, P_{ST} means the pressure measured by ST, and T_{VS} is the temperature recorded by VS. The Δ(delta) symbol is used to denote the difference of the same variable between two sensors. Finally, the ’ (prime) represents the corrected measure.

3.1 CDF-based Probability Matching

CDF-based Probability matching, also known as histogram matching or quantile mapping, is a statistical technique used to adjust the distribution of a dataset (e.g., a forecast distribution) to match that of another dataset (e.g., an observed distribution). The primary objective of this method is not to directly correct individual data points but to ensure that the overall statistical properties, such as the frequency of occurrence of specific values, match between the two datasets. In radiosonde observation, CDF-based probability matching is commonly used as a quality control tool to ensure data quality consistency for field campaigns (Nuret et al., 2008; Ciesielski et al., 2009).

Based on the paired entries collected in co-launches, the two-step correction scheme starts with correcting temperature (ΔT) based on the ground-checked pressure (P'_{ST}) and the measured temperature (T_{ST}). Then, the adjusted temperature (T'_{ST}) is used together with the measured relative humidity (RH_{ST}) to estimate the correction (ΔRH).

We first discretize the pressure and temperature variables in temperature correction into bins. Pressure is divided into 50 hPa intervals from 975–1025 hPa to 175–225 hPa, denoted by their centers, 1000 hPa to 200 hPa. The cumulative distribution function of temperature measured by ST and VS for each pressure bin is calculated as follows. The observed temperature records are sorted in ascending order, and then the proportion of observations is derived for every 1-degree interval from -80 to 40 degrees Celsius as the probability density. Based on the assumption that two sensors have the same CDF within this specific range, we derived the correction values, ΔT, as a function of measured temperature, T_{ST}. Figure 5 demonstrates the CDF-based temperature correction of the pressure bin 475–525 hPa as an example. The upper panel shows the CDF of T_{VS} and T_{ST}, and the lower panel illustrates the correction (ΔT) as a function of the observed temperature (T_{ST}). We grouped the co-launches into daytime and night-time and performed the above procedure for each pressure bin. The results are shown in Figure 6, the complete temperature correction table used in this study.

As shown in Figure 6, the temperature sensor of ST consistently shows warm bias in all pressure bins, and the bias is stronger at high altitudes. The night-time warm bias exhibits similar patterns to the daytime but with a lower quantity.

The correction of relative humidity (RH) is derived in the same way as the temperature, except for the independent variables, which are the corrected temperature (T'_{ST}) and the measured relative humidity (RH_{ST}). The corrected temperature is discretized into 10-degree intervals from -65 to 35 degrees Celsius. The relative humidity values are then rounded to integers and form 1% intervals from 0 to 100. Like the temperature correction procedure, the correction value is derived based on the CDF probability matching as a function of RH within each temperature bin. Figure 7 illustrates the complete RH correction table used in this study. Figure 7 indicates that the ST shows dry-bias (wet-bias) in lower (higher) altitudes. ST is generally dryer during the daytime.

Using the correction tables shown in Figures 5 and 6, the temperature and relative humidity measured by ST are corrected and evaluated. Mathematically, this procedure can be expressed as:

$$\Delta T = T_{VS} - T_{ST} = f(P'_{ST}, T_{ST}, Day) \quad (1)$$

$$\Delta RH = RH_{VS} - RH_{ST} = f(T'_{ST}, RH_{ST}, Day) \quad (2)$$

, where Day is a binary variable representing the daytime or night-time, and f is the CDF-based probability matching. Because we first correct the temperature and then use the corrected temperature to correct the humidity, we call this approach a two-step CDF-base calibration.

3.2 Generalized Linear Model

Despite the robustness and ease of implementation of CDF-based probability matching, the discretization steps and the form of the look-up table limit its application. For example, the discretization of pressure and temperature is empirical. Though the resulting CDFs and correction tables look reasonable, it is hard to justify that this is the only way to split a continuous variable into bins. In other words, by focusing on matching the overall distribution, probability matching may overlook or alter some of the finer-scale details in the dataset. Furthermore, the look-up table makes adding extra independent variables more complicated. For example, we used daytime and night-time tables to simplify the influence of solar radiation so that we could use two tables for each correction. Another example is when we consider adding the effect of pressure in the correction of RH. In that case, we need to establish three-dimensional bins and justify whether the cut-off points are adequately selected. Therefore, we want to introduce the modeling of the multivariate probability distribution to our correction scheme.

In essence, modeling the joint probability distributions of multiple variables is fundamental in machine learning for capturing relationships and dependencies among numerous predictors. It forms the backbone for various algorithms and techniques to predict, generate, and understand multi-dimensional data. In equations (1) and (2), the mapping function, f, can be seen as a model of the joint probability distribution of the independent variables. While the CDF-based probability matching algorithm models this distribution by discretizing the independent variables, it can be replaced by different algorithms that keep the predictors in their continuous form.

The Generalized Linear Model (GLM, Nelder and Wedderburn, 1972) is a versatile statistical framework used for modeling the relationship between a dependent variable (response) and one or more independent variables (predictors) in a wide range of applications. GLMs extend the concept of linear regression to handle a broader array of data types and distributions. They are valuable for offering interpretable coefficients to understand the impact of predictors on the response. GLMs have become a fundamental tool in statistics and data analysis due to their flexibility and applicability across various fields. In this study, we used GLMs in three different settings: first, the same scheme as CDF-based probability matching (GLM1, as specified in equations (1) and (2)); second, using the same set of predictors for T and RH corrections (GLM2); and finally, replacing daytime with Julian-day and hour-of-day (GLM3).

To develop the GLM-based corrections, we used the paired entry dataset and the least squared algorithm to fit linear regression models for the response variables (ΔT and ΔRH) and the predictors (P'_{ST}, T_{ST}, RH_{ST}, and Day). This study used the Python algorithm implementation from scikit-learn (Pedregosa et al., 2011). The resulting regression equations are used to correct the storm tracker data.

In the second GLM configuration, we use the variables of P'_{ST}, T_{ST}, RH_{ST}, and Day to predict the corrections of temperature (ΔT) and relative humidity (ΔRH). The resulting models can be mathematically denoted as:

$$\Delta T = f(P'_{ST}, T_{ST}, RH_{ST}, Day) \tag{3}$$

$$\Delta RH = f(P'_{ST}, T_{ST}, RH_{ST}, Day) \tag{4}$$

Previous studies have suggested that solar radiation could be the leading cause of the warm bias in the radiosonde data. This is why we established correction tables for daytime and night-time separately. To simplify the correction process and limit the number of tables created, the solar radiation is represented by the binary variable of Day. However, with GLMs, we can easily use continuous variables in their original form. Hence, we used the “Julian day from the summer solstice” (Jday) and the “hour-of-day from noon” (Hour) to replace the Day variable. The resulting models are:

$$\Delta T = f(P'_{ST}, T_{ST}, RH_{ST}, Jday, Hour) \tag{5}$$

$$\Delta RH = f(P'_{ST}, T_{ST}, RH_{ST}, Jday, Hour) \tag{6}$$

These three settings are noted as GLM1, GLM2, and GLM3 in the later text.

Because all of our co-launches were conducted over the Taiwan area, the Julian day and the hour of the day can properly approximate the value of the clear day radiation. Though the resulting correction formula can be applied to other regions, the differences in the pressure-altitude relationship might slightly interfere with other predictors. Therefore, we recommend adding the location information (i.e., longitude and latitude) or directly using the derived values of clear-sky radiation to develop the correction formula in other regions.

4. Results

Figure 8 illustrates the patterns and deviations between ST and VS at various pressure levels. The panels (a), (b), and (c) demonstrate the temperature of VS and ST, and the differences between the two sensors. The relative humidity is shown in panels (d), (e), and (f). As shown in Figure 8, the ST exhibits warm and dry biases in general, and the biases increase as the altitude rises.

We applied the four correction methods described in the previous section, i.e., CDF, GLM1, GLM2, and GLM3, to the 663 sounding profiles. Using the VS as the reference observations, we calculated the root-mean-squared errors (RMSEs) as the evaluation metrics. We did not use the correlation coefficients for evaluation because two sensors have correlation coefficients higher than 0.99, even without corrections. The reason for this lies in the co-launching strategy, which ensures that both instruments endure the same environmental conditions. The means and standard deviations of RMSEs for all correction methods are shown in Table 3 and Figure 9. As shown in Figure 9, we can see a significant bias reduction for all correction methods. We performed t-tests on the raw and corrected values, and the improvement of all four methods is statistically significant (for p-values little than 10e-29). We also compared the CDF and GLM, and the results show that CDF correction is slightly better than GLMs for both temperature and relative humidity. The difference between CDF and GLMs is significant in the t-test, though the significant level is much lower than their bias reduction.

We also conducted t-tests on different GLM settings. The GLM1 and GLM2 did not show significant differences in temperature and relative humidity correction results. However, the GLM3 showed great improvement compared to GLM1 and GLM2. This suggested that solar radiation parameters can influence the correction more than a simple day/night indicator.

Table 3 and Figure 9 also show the evaluations for all records below 500- and 700-hPa heights. As shown in the results, ST can proximate the VS measurements with a temperature error of less than 1 degree Kelvin and a relative humidity error of less than 10%. Suppose we focus on the observations below 700 hPa. In that case, the averaged RMSE can be as low as 0.66-degree Kelvin for temperature and 4.61% for relative humidity, comparable to the uncertainties of VS temperature and relative humidity measurements (Vaisala, 2017). Such results suggested that the ST is sufficiently accurate, especially when focusing on the boundary layer and lower atmosphere.

In addition to the overall performance of ST, we illustrated the RMSEs distribution of the 663 soundings in Figure 10. The upper panel, (a), illustrates the distribution of RMSEs before correction, and the lower panel, (b), shows the results after the CDF-based correction. As shown in Figure 10, the proposed correction methods reduced both the biases and spreads. The reduction in the standard deviation of RMSE in Table 3 also represents this fact. Based on Figure 10, we selected three cases with low, middle, and high biases in RH before correction to discuss in the following section. The one-by-one comparison of the 633 profiles can be found in the supporting materials.

5. Discussion

5.1 The Random Errors of ST

The specifications of the temperature and humidity sensor used in the ST reported the accuracy range as ±0.3°C and ±2% (Huang et al., 2020). We examined the random errors with cloud chamber laboratory examination and field observation datasets with dual ST launching.

Six STs of the same batch used in the co-launches were measured in controlled chambers. Each sensor was repeatedly measured at 10°C, 20°C, 30°C, and 40°C, and at relative humidity of 30%, 50%, 70%, and 90%. The results are shown in Figure 11. The standard deviations of the measured differences are 0.24°C (temperature) and 2.21% (relative humidity), respectively. The results reasonably agreed with the random errors reported by the manufacturer.

To assess the random error in the field, we conducted 42 observations with dual-ST. We aligned the records of two instruments with timestamps and evaluated the differences in temperature and humidity. In the 42 launches, there were a total of 96,284 aligned entries. We used statistical fences to exclude extreme situations such as frozen or malfunctioning sensors (Everitt and Skrondal, 2010; Tukey, 1977). After applying this simple outlier removal technique, we have 85,641 temperature measurements and 81,616 pairs of relative humidity. The paired measurements are shown in Figure 12. The derived standard deviation for temperature is 0.52°C, and for relative humidity is 2.25%. The random errors measured in the field are slightly higher than those measured in the laboratory and reported by the manufacturer. There were 42 dual-ST attached to VS co-launches in the field, all of them were conducted during the day. We realized that the sensor performance could have diurnal variations and we had performed the correction according to the day-night difference. By following the types of errors defined in Collins (2001), we attributed the day-night variability as systematic error, which our correction methods can remedy. Hence, we didn't further distinguish the random errors for day and night.

The results of the sounding-by-sounding evaluation presented in the earlier section, 0.66-degree for temperature and 4.61% for humidity, are slightly larger than the random errors measured in the field. This suggests that there is room to develop more sophisticated correction methods.

According to previous studies, the ST sensor has a about 5-second response time

(Huang et al., 2020). Several time-lag analyses were conducted to verify this and the impact to the measurement correction, the results suggest insignificant changes to the bias correction. However, given that Miloshevich et al. (2001, 2004) discussed the errors introduced by the sensor's time lag and proposed a correction algorithm, we plan to incorporat further sophisticated time-lag correction approaches in the future.

5.2 General performance of ST

Figure 13 illustrates the paired entries of VS and ST before and after corrections. As described in the previous section, the ST exhibits correlation coefficients higher than 0.99 for temperature and RH even before any correction. Hence, the effect of corrections is represented by the narrower diagonals in the right panels in Figure 13.

Even though the statistical tests showed the significance of the correction results, they are not easily perceived. Hence, we selected a few sounding profiles to demonstrate the effectiveness of our correction methods. Figure 14 shows the T and RH profile of the sounding launched at 2021-08-03 12Z. This sounding was selected because of the overall low RH bias before and after correction. In Figure 14, the corrected temperature is adequately aligned to the reference (T_{VS}), and the corrected relative humidity (RH) is entirely satisfactory, particularly below 350hPa, covering most tropospheric levels with water vapor and clouds. Consistent findings are prevalent within our dataset, indicating that the adjusted ST measurements are reliable across various observational scenarios.

However, the corrected results may perform less when encountering extreme wet cases. Figure 15 is the sounding profile on 2018-08-27 06Z when the reference RH of VS is about 90% from ~850 to ~350-hPa heights. As shown in Figure 15, the temperature correction still works properly, except that the VS's temperature sensor showed much larger amplitude compared to VS. However, the RH measured by ST shows a dry bias of magnitude of 20% from ~850 to ~350-hPa heights while the patterns stay similar. The RH correction mechanisms adjust the RH toward the reference, but the deviations are still significant. Note that this observation occurred during a severe rainfall event caused by the convergence of the tropical depression and the southwest monsoon from August 23 to August 30, 2018. All fifteen co-launches conducted in this event exhibited high bias in RH, ranging from 10% to 24%, and five showed bias greater than 10% even after correction. This particularly biased case has RMSE ranked 99.93% in our dataset. Since such a large deviation rarely showed in the colaunches, we believe it could be caused by malfunction of this specific sensor.

In the left panel of Figure 15, we can also see a sudden change in GLM-corrected temperature around 310hPa. This should be caused by the missing values of ST in RH (see the missing yellow line section in the right panel). Because the GLM correction includes RH as an independent variable, when RH values are missing (treated as 0), the amount of correction can change accordingly.

Figure 16 illustrates the sounding profile on 2020-03-13 12Z. This is an average case with middle bias in RH before correction. Most of the 633 co-launches behave similarly to this case.

From the cases shown above, we also notice the characteristics of different correction methods. The GLM adjustments look like horizontal shifts of the original values due to the linearity of the model.

Despite the simplicity of our correction methods, the temperature bias between ST and VS can be reduced from 3.0 K to 0.9 K, and the RH bias from 8.5% to 6.9%. Note that our correction methods also reduce the standard deviations from 1.8 K to 0.6 K and 3.8% to 2.8%, respectively. Hence, we can expect 80% of ST observations to exhibit less than 1K bias in temperature and 8.8% bias in RH.

The corrected ST measurements aligned well with the VS data, especially when the sounding successfully reached an altitude higher than 300hPa. For those co-launches that ended early, though their bias is still low in statistics, their profiles usually looked problematic when visualized. We recommend further looking into the reasons that cause the sounding to end early.

5.3 A ST observation in afternoon thunderstorm study

The low cost of the ST can facilitate high spatial-temporal frequency of upper-air observations. While the ST provides reasonable measures after correction, its reliability in higher altitudes is still incompatible with the VS used in standard operation. Therefore, here, we demonstrate a use case to illustrate the strength of the ST. Figure 17 shows a set of continuous ST profiles on 2018-08-17 with one-hour intervals. This experiment used only

ST and was not included in the colaunch dataset. Figure 17 shows the evolution of a local convective system, which is not feasible in regular 12-hour interval radiosonde operation - the increase of atmospheric moisture at 1300 local time before the heavy rain occurrence is observed. Using the flexibility in deploying the ST during field campaigns allows us to capture vertical profiles in the lower troposphere at an hourly, or even a shorter time interval. This is notably advantageous for understanding the development of deep convection, which typically has a lifetime of 1 to 3 hours, and the surrounding environment, especially the lower boundary layer. A similar ST profile has been used in the study of the afternoon thunderstorm in Taipei compared to the results from CRESS cloud-resolving modeling (Tsujino et al. 2022). Note that the ST data here was corrected with the CDF-based method; better performance can be achieved with GLM-based methods.

6. Concluding Remarks

In this study, we assess the data quality control and calibration of the Storm Tracker (ST) with the co-launched Vaisala RS41-SGP (VS) in temperature, relative humidity, and winds for lower atmospheric observations. Although wind speed and direction are crucial information in radiosonde observation, we found from the co-launched data that the GPS-estimated ST wind differs from that of VS in insignificant magnitude. The GPS estimated ST wind error difference is about 0.05 ms^{-1}. To ensure the reliability of ST measurements in temperature and moisture, we conducted over a thousand co-launches of the ST and the VS, evaluating and refining the performance of the ST through developed correction methods for temperature and humidity measurements. Based on the sounding-by-sounding comparison, the corrected ST soundings have a 1-K temperature and 7% relative humidity root mean square difference from the VS soundings. These error differences can be reduced to 0.66-K and 4.61% below the 700-hPa height. The biases of the corrected ST observations are slightly larger than the random errors, which were 0.24 K and 2.21% in the laboratory and 0.52 K and 2.23% in the field.

Derived from the co-launch dataset, two correction methods based on CDF and GLM algorithms were implemented to enhance the quality of temperature and humidity observations in the ST. Both methods work comparably well to reduce the biases of the ST. While the CDF-based correction is robust and reliable, the GLMs easily model and change the predictors. The ST observations closely aligned with the VS after corrections, particularly in the lower atmospheric layers below 700hPa. For synoptic weather, geostrophic adjustment dynamics suggest that spatial temperature variations in the free atmosphere may not be significant, reducing the need for high-frequency upper-air radiosonde observations. Consequently, most operational radiosonde observations worldwide are conducted daily at 00Z and 12Z, with 12–24 hours intervals. However, atmospheric phenomena originating from the boundary layer are often smaller in scale and closely related to local terrain. For example, a single convective cell typically lasts minutes, while thunderstorms persist for a few hours. To better understand these types of weather, a low-cost and lightweight device capable of deploying multiple sensors simultaneously or at intervals of less than an hour can enhance field experiments. This approach provides valuable insights into the lower atmosphere's significant variations in temperature and moisture, especially for convective systems that may lead to disastrous rainfall or flash flooding. This positions the ST as a promising candidate for supplementing regular upper-air observations for high spatial and temporal resolution in the lower atmosphere. Our work also demonstrated that low-cost commercial sensor components can help high-frequency observations in specific targets with carefully developed correction methods.

Although we used the linear regression version of GLMs in this study, the concept of modeling the joint probability distribution can be extended to various statistical models such as decision trees, support vector machines (SVM), and artificial neural networks (ANN). The simple GLMs in this study assume the response is a Gaussian distribution of the linear combination of predictors. Other machine learning models can establish nonlinear mappings between the predictors and response without assuming any distributions. However, investigating more machine learning models is beyond the scope of this study.

In summary, while the VS remains the standard for upper-air observation, ST is suitable for Planetary Boundary Layer (PBL) or lower atmosphere studies in areas with complex terrain. The ST can complement the VS observation with high spatial and temporal resolution observation of the lower atmosphere. This may be useful for mesoscale storm observations in East Asia, where PBL conditions can vary significantly within short distances.

However, it is important to note that the correction results presented here are specific to Taiwan’s observation. This is especially true for the CDF method, as the variability of the CDF method data is height-dependent, so the direct use of our CDF calibration should be cautious. On the other hand, the GLM method may provide a reasonable calibration to the ST sounding when longitude and latitude are used as predictors or local clear-sky radiation is directly used. To ensure broader applicability, we suggest conducting co-launches during field campaigns. This approach would allow users to derive in-situ correction formulas using the proposed methods. Our experiments indicate that ST between VS launches may enhance meteorological data collection and analysis in the lower atmosphere.

Data Availability Statement

The data for this project is confidential but may be obtained with Data Use Agreements with the National Taiwan University. Researchers interested in access to the data may contact authors. It can take some months to negotiate data use agreements and gain access to the data. The author will assist with reasonable replication attempts for two years following publication.

Code for data cleaning and analysis is provided in a replication package. It is available at https://www.dropbox.com/scl/fo/ah7i6z4f7u2yzijfh7ua3/h?rlkey=ar4g2hq7hwkop2eyzw83el8ih&dl=0 for review. It will be uploaded to GitHub once the paper has been conditionally accepted.

Acknowledgments

This study was supported by the National Science and Technology Council (NSTC) of Taiwan for many years of support under Grants MOST 107-2628-M-002-016, MOST 108-2119-M-002-022, MOST 109-2111-M-002-008, MOST 110-2123-M-002-007, NSTC 111-2123-M-002-014, NSTC 112-2123-M-002-006, NSTC 113-2123-M-002-018, and NSTC 113-2124-M-002-015..

We want to express our sincere gratitude to Wei-Chun Huang, who contributed to developing the storm tracker in the first place. Our heartful appreciation to the Central Weather Administration (CWA) for most of the co-launch in Vaisala RS41-SGP (VS) and Storm Tracker (ST). We thank the research team members of the Taipei Summer Storm Experiment (TASSE), the Yilan Experiment of Severe Rainfall (YESR2020), the Taiwan-Area Heavy Rain Observation and Prediction Experiment (TAHOPE), the Northern Coast Observation, Verification, and Investigation of Dynamics (NoCOVID21), and the Mountain Cloud Climatology (MCC) project. Their dedication and commitment were instrumental in the realization of our research objectives.

We sincerely thank the reviewers for their thorough and insightful feedback, which immensely helped to improve the manuscript.

References

Ciesielski, P. E., Haertel, P. T., Johnson, R. H., Wang, J., and Loehrer, S. M.: Developing High-Quality Field Program Sounding Datasets. *Bulletin of the American Meteorological Society*, **93**(3), 325–336. https://doi.org/10.1175/BAMS-D-11-00091.1, 2012.

Ciesielski, P. E., Johnson, R. H., and Wang, J.: Correction of Humidity Biases in Vaisala RS80-H Sondes during NAME. *Journal of Atmospheric and Oceanic Technology*, **26**(9), 1763–1780. https://doi.org/10.1175/2009JTECHA1222.1, 2009.

Collins, W. G. (2001). The Operational Complex Quality Control of Radiosonde Heights and Temperatures at the National Centers for Environmental Prediction. Part II: Examples of Error Diagnosis and Correction from Operational Use. *Journal of Applied Meteorology and Climatology*, 40(2), 152–168.

Dirksen, R. J., Sommer, M., Immler, F. J., Hurst, D. F., Kivi, R., & Vömel, H. (2014). Reference quality upper-air measurements: GRUAN data processing for the Vaisala RS92 radiosonde. *Atmospheric Measurement Techniques,* **7**(12), 4463–4490. https://doi.org/10.5194/amt-7-4463-2014

Dzambo, A. M., Turner, D. D., and Mlawer, E. J.: Evaluation of two Vaisala RS92 radiosonde solar radiative dry bias correction algorithms. *Atmospheric Measurement Techniques*, **9**(4), 1613–1626. https://doi.org/10.5194/amt-9-1613-2016, 2016.

Everitt, B. S.; Skrondal, A. (2010), The Cambridge Dictionary of Statistics, Cambridge University Press.

Hoshino, S., Sugidachi, T., Shimizu, K., Kobayashi, E., Fujiwara, M., & Iwabuchi, M.

(2022). Comparison of GRUAN data products for Meisei iMS-100 and Vaisala RS92 radiosondes at Tateno, Japan. *Atmospheric Measurement Techniques*, **15**(20), 5917–5948. https://doi.org/10.5194/amt-15-5917-2022

Hwang, W.-C., Lin, P.-H., and Yu, H.: The development of the “Storm Tracker” and its applications for atmospheric high-resolution upper-air observations. *Atmospheric Measurement Techniques*, **13**(10), 5395–5406. https://doi.org/10.5194/amt-13-5395-2020, 2020.

Kizu, N., Sugidachi, T., Kobayashi, E., Hoshino, S., Shimizu, K., Maeda, R., & Fujiwara, M. (2018). Technical characteristics and GRUAN data processing for the Meisei RS-11G and iMS-100 radiosondes. GRUANLead Centre. https://www.gruan.org/documentation/gruan/td/gruan-td-5

Lee, S.-W., Kim, S., Lee, Y.-S., Choi, B. I., Kang, W., Oh, Y. K., Park, S., Yoo, J.-K., Lee, J., Lee, S., Kwon, S., & Kim, Y.-G. (2022). Radiation correction and uncertainty evaluation of RS41 temperature sensors by using an upper-air simulator. *Atmospheric Measurement Techniques*, **15**(5), 1107–1121. https://doi.org/10.5194/amt-15-1107-2022

Lesht, B., and Richardson, S.: The Vaisala RS80H Radiosonde Dry-Bias Correction Redux. *Proceedings of the ninth atmospheric radiation measurement (ARM) science team meeting 2002*. ARM-CONF-2002, St. Petersburg, Florida, 2002.

Lucas, C., and Zipser, E. J.: Environmental Variability during TOGA COARE. *Journal of the Atmospheric Sciences*, **57**(15), 2333–2350. https://doi.org/10.1175/1520-

0469(2000)057<2333:EVDTC>2.0.CO;2, 2000.

Luers, J. K.: The Influence of Environmental Factors on the Temperature of the Radiosonde Thermistor. Doctoral Dissertations. https://trace.tennessee.edu/utk_graddiss/1913, 1989.

Luers, J. K.: Temperature Error of the Vaisala RS90 Radiosonde. *Journal of Atmospheric and Oceanic Technology*, **14**(6), 1520–1532. https://doi.org/10.1175/1520-0426(1997)014<1520:TEOTVR>2.0.CO;2, 1997.

Luers, J. K., and Eskridge, R. E.: Use of Radiosonde Temperature Data in Climate Studies. *Journal of Climate*, **11**(5), 1002–1019. https://doi.org/10.1175/1520-0442(1998)011<1002:UORTDI>2.0.CO;2, 1998.

Mattioli, V., Westwater, E. R., Cimini, D., Liljegren, J. C., Lesht, B. M., Gutman, S. I., and Schmidlin, F. J.: Analysis of Radiosonde and Ground-Based Remotely Sensed PWV Data from the 2004 North Slope of Alaska Arctic Winter Radiometric Experiment. *Journal of Atmospheric and Oceanic Technology*, **24**(3), 415–431. https://doi.org/10.1175/JTECH1982.1, 2007.

McMillin, L., Uddstrom, M., and Coletti, A.: A Procedure for Correcting Radiosonde Reports for Radiation Errors. *Journal of Atmospheric and Oceanic Technology*, **9**(6), 801–811. https://doi.org/10.1175/1520-0426(1992)009<0801:APFCRR>2.0.CO;2, 1992.

Miller, E. R., Wang, J., and Cole, H. L.: Correction for dry bias in Vaisala radiosonde RH data. *Proceedings of the ninth atmospheric radiation measurement (ARM) science team* *meeting 1999*. ARM-CONF-1999, San Antonio, Texas, 1999.

Miloshevich, L. M., Paukkunen, A., Vömel, H., and Oltmans, S. J.: Development and Validation of a Time-Lag Correction for Vaisala Radiosonde Humidity Measurements. *Journal of Atmospheric and Oceanic Technology*, **21**(9), 1305–1327. https://doi.org/10.1175/1520-0426(2004)021<1305:DAVOAT>2.0.CO;2, 2004.

Miloshevich, L. M., Vömel, H., Paukkunen, A., Heymsfield, A. J., and Oltmans, S. J.: Characterization and Correction of Relative Humidity Measurements from Vaisala RS80-A Radiosondes at Cold Temperatures. *Journal of Atmospheric and Oceanic Technology*, **18**(2), 135–156. https://doi.org/10.1175/1520-0426(2001)018<0135:CACORH>2.0.CO;2, 2001.

Nelder, J. A., & Wedderburn, R. W. M.: Generalized Linear Models. *Journal of the Royal Statistical Society. Series A (General)*, **135**(3), 370–384. https://doi.org/10.2307/2344614, 1972.

Nuret, M., Lafore, J.-P., Guichard, F., Redelsperger, J.-L., Bock, O., Agusti-Panareda, A., and N'Gamini, J.-B.: Correction of Humidity Bias for Vaisala RS80-A Sondes during the AMMA 2006 Observing Period. *Journal of Atmospheric and Oceanic Technology*, **25**(11), 2152–2158. https://doi.org/10.1175/2008JTECHA1103.1, 2008.

Sommer, M., von Rohden, C., Simeonov, T., Oelsner, P., Naebert, T., Romanens, G., Jauhiainen, H., Survo, P., & Dirksen, R. (2023). GRUAN characterisation and data processing of the Vaisala RS41 radiosonde. GRUAN Lead Centre.

https://www.gruan.org/documentation/gruan/td/gruan-td-8

Sun, B., Reale, A., Schroeder, S., Seidel, D. J., and Ballish, B.: Toward improved corrections for radiation-induced biases in radiosonde temperature observations. *Journal of Geophysical Research: Atmospheres*, **118**(10), 4231–4243. https://doi.org/10.1002/jgrd.50369, 2013.

Tsujino, S., Kuo, H.-C., Yu, H., Chen, B.-F., and Tsuboki, K.: Effects of mid-level moisture and environmental flow on the development of afternoon thunderstorms in Taipei. *Terr. Atmos. Ocean. Sci.*, **32**, 497-518, https://doi: 10.3319/TAO.2021.11.17.01, 2021.

Tukey, J. (1977). Exploratory Data Analysis, Addison-Wesley

Vaisala.: Vaisala Radiosonde RS41 Measurement Performance. Vaisala. https://www.vaisala.com/en/whitepaper/vaisala-radiosonde-rs41-measurement-performance, 2017.

Vömel, H., Selkirk, H., Miloshevich, L., Valverde-Canossa, J., Valdés, J., Kyrö, E., Kivi, R., Stolz, W., Peng, G., and Diaz, J. A.: Radiation Dry Bias of the Vaisala RS92 Humidity Sensor. *Journal of Atmospheric and Oceanic Technology*, **24**(6), 953–963. https://doi.org/10.1175/JTECH2019.1, 2007.

von Rohden, C. von, Sommer, M., Naebert, T., Motuz, V., & Dirksen, R. J. (2022). Laboratory characterization of the radiation temperature error of radiosondes and its application to the GRUAN data processing for the Vaisala RS41. *Atmospheric Measurement Techniques*, **15**(2), 383–405. https://doi.org/10.5194/amt-15-383-2022

Wang, J., Zhang, L., Dai, A., Immler, F., Sommer, M., and Vömel, H.: Radiation Dry Bias Correction of Vaisala RS92 Humidity Data and Its Impacts on Historical Radiosonde Data. *Journal of Atmospheric and Oceanic Technology*, **30**(2), 197–214. https://doi.org/10.1175/JTECH-D-12-00113.1, 2013.

Yoneyama, K., Fujita, M., Sato, N., Fujiwara, M., Inai, Y., and Hasebe, F.: Correction for Radiation Dry Bias Found in RS92 Radiosonde Data during the MISMO Field Experiment. *Sola*, **4**, 13–16. https://doi.org/10.2151/sola.2008-004, 2008.

Yu, H., Ciesielski, P.E., Wang, J., Kuo, H.-C., Vömel, H., and Dirksen, R.: Evaluation of Humidity Correction Methods for Vaisala RS92 Tropical Sounding Data. *Journal of Atmospheric and Oceanic Technology*, **32**, 397–411, https://doi.org/10.1175/JTECH-D-14-00166.1, 2015.

Yu, H., Kuo, H.-C., Lin, P.-H., Huang, W.-C., Liu, C.-H., Su, S. H., and Yang, J.-H.: Quality-Controlled High-Resolution Upper-Air Sounding Dataset for TASSE: Development and Corrections of the “Storm Tracker” Observations. 2020 Conference on Weather Analysis and Forecasting. *2020 Conference on Weather Analysis and Forecasting*, Taipei, 2020.

List of Figures

Fig. 8 The boxplot of temperatures (upper) and RH (lower) of ST (left), VS (center), and their difference (right) without corrections.

Fig. 9 The mean RMSE of ST and VS with different correction methods for temperature (left) and RH (right). For each correction method, the mean RMSE is derived with all available records (blue), records below 500 hPa (orange), and records below 700hPa (green). The upper panel (a) showed the overall RMSE, and the middle (b) and lower panel (c) demonstrated the RMSE of daytime and nighttime, respectively

Fig. 10 The scatter plots of temperature (upper) and RH (lower) before and after correction.

Fig. 11 The temperature (left) and RH (right) measurements of STs in the controlled laboratory environment. Six STs were measured separately. The temperature (left panel) was measured repeatedly at 10°C, 20°C, 30°C, and 40°C. The relative humidity (right panel) was measured at 30%, 50%, 70%, and 90%. The derived random error for temperature is 0.24°C, and for relative humidity is 2.21%.

Fig. 12 The biases of 42 dual-ST launches. The figure shows 85,641 temperature measurements (left panel) and 81,616 pairs of relative humidity (right panel). The derived random error for temperature is 0.52°C, and for relative humidity is 2.25%.

Fig. 13 The histograms of the RMSEs of temperature (upper) and RH (lower) between ST and VS. The upper panel, (a), illustrates the distribution of RMSEs before correction, and the lower panel, (b), shows the results after the CDF-based correction.

Fig. 14 The temperature (left) and RH (right) of the co-launch sounding on 2021-0803 12Z. The reference (VS) is illustrated in blue, the ST in range, CDF-corrected in green, and GLM-corrected in red.

Fig. 15 The temperature (left) and RH (right) of the co-launch sounding on 2018-08-27 06Z. The reference (VS) is illustrated in blue, the ST in range, CDF-corrected in green, and GLM-corrected in red.

Fig. 16 The temperature (left) and RH (right) of the co-launch sounding on 2018-03-13 12Z. The reference (VS) is illustrated in blue, the ST in range, CDF-corrected in green, and GLM-corrected in red.

Fig. 17 The continuous ST observations of one-hour intervals on 2018-08-17 at Shezi. The soundings were corrected with CDF, and the derived specific humidity, q, is shown in panel (a) with the wind field. The derived equivalent potential temperature, Θe, is shown in panel (b).

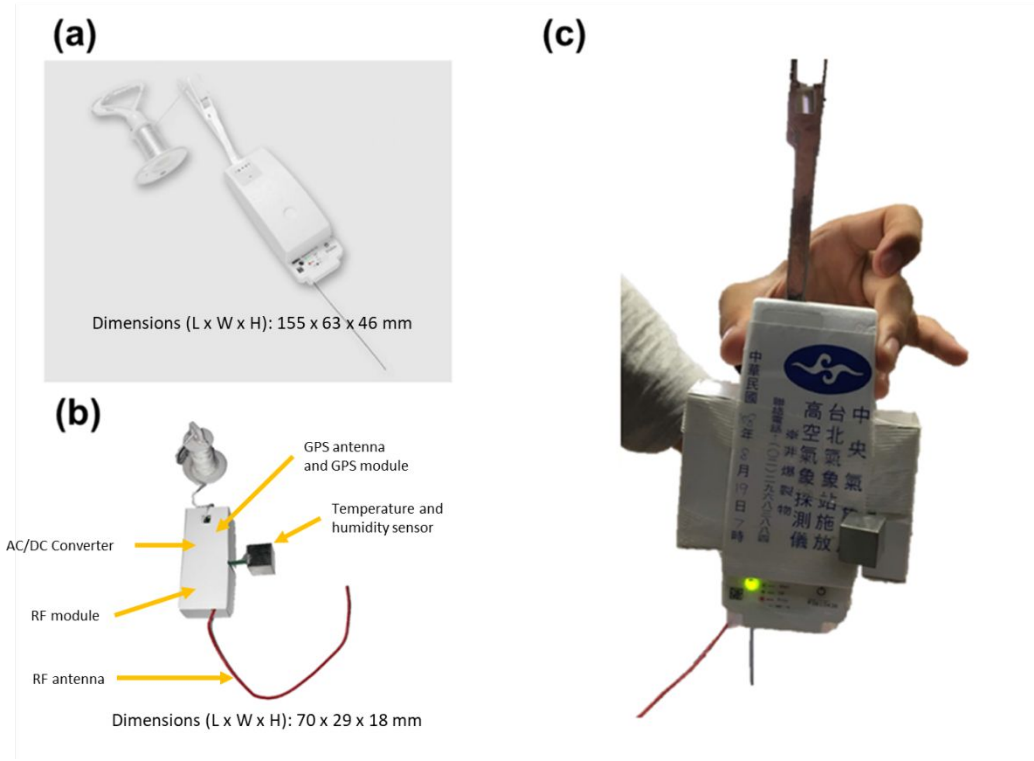


Fig. 1 (a) The Vaisala RS41-SGP radiosonde (weighted 84g, body dimension: 155 x 63 x 46 mm), (b) the storm tracker mini-radiosonde (weighted 20 g with battery, body dimension: 70 x 29 x 18 mm), and (c) an example of the co-launched soundings via the TASSE experiment. More ST hardware details are described in Hwang et al. (2020).

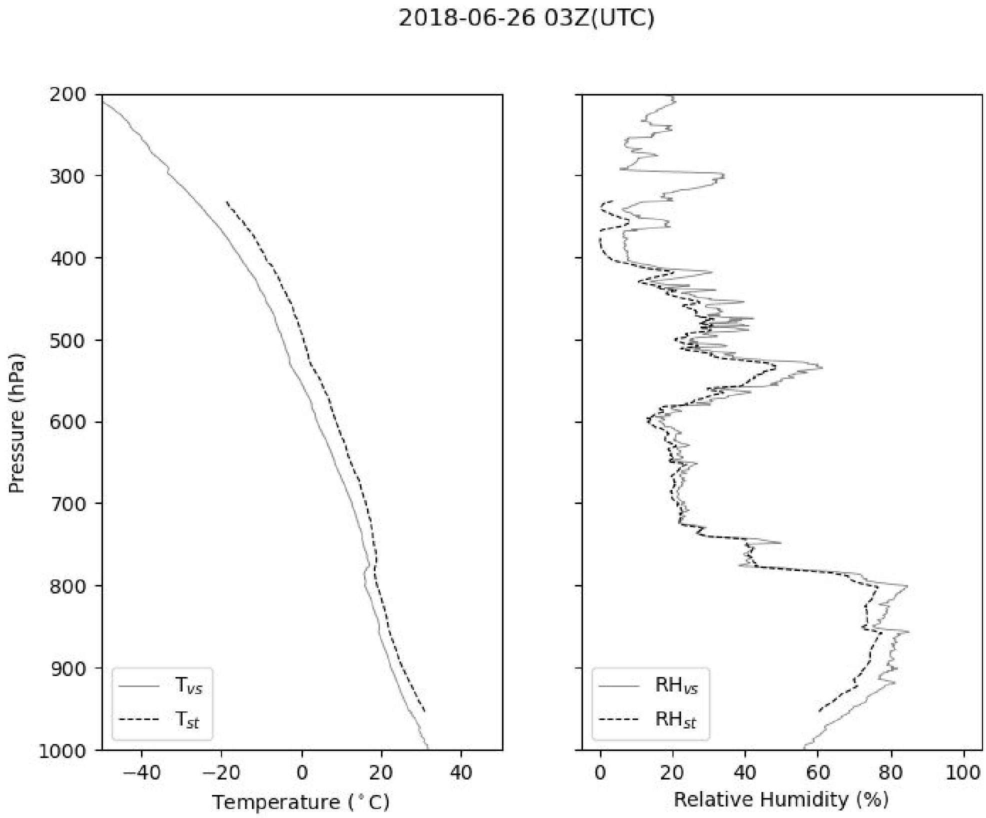


Fig. 2 The sounding of 2018-06-26 03:00 UTC (11:00 LST) by VS (solid lines) and ST (dashed lines). The ST profile showed warm and dry bias near the surface.

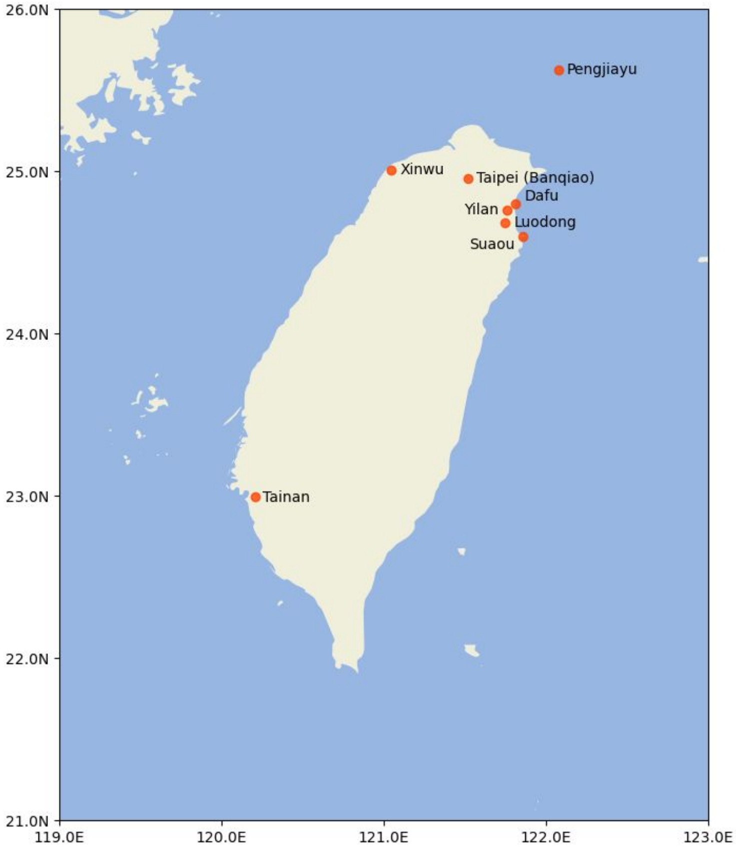


Fig. 3 The sites of the co-launch experiments. Most co-launches (909 out of 1,029) were conducted in the Taipei (Banqiao) station. The number of co-launches collected in each site can be found in Table 1.

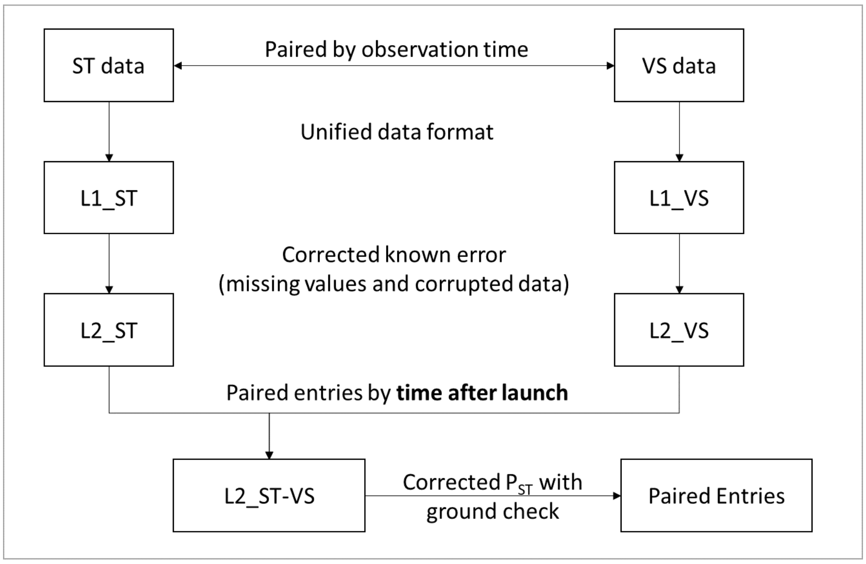


Fig. 4 The preprocessing for ST and VS data from raw to level 2.

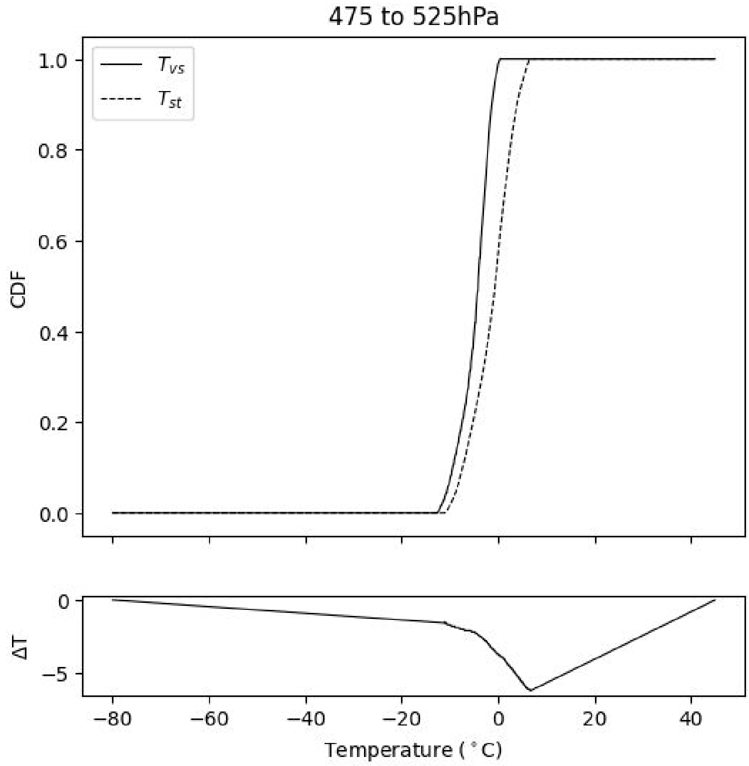


Fig. 5 The CDF-based temperature correction of the pressure bin 475 ~ 525 hPa. The upper panel shows the CDF of the temperature of two sensors, and the lower panel shows their difference as a function of temperature. The probability density is defined by the proportion of observations within every 1-degree interval from -80 to 40 degrees Celsius.

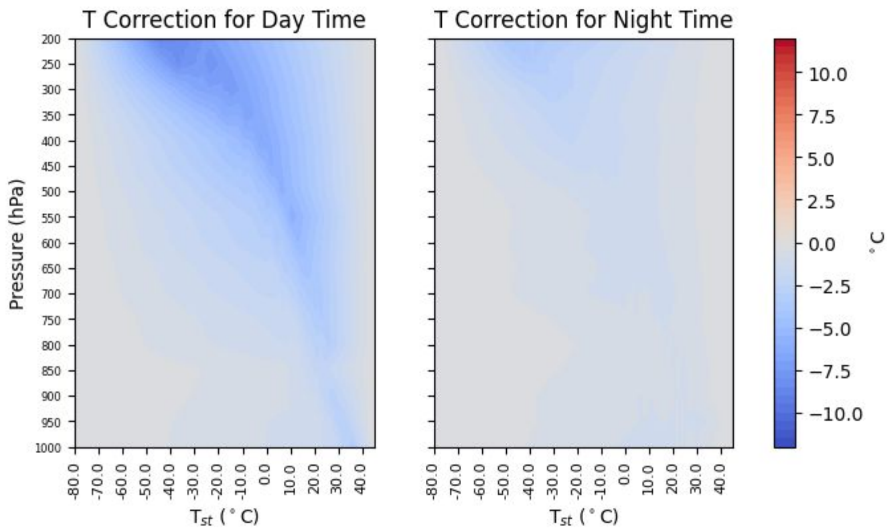


Fig. 6 The CDF-based temperature correction tables for daytime (00z-12z, left panel) and night-time (12z-00z, right panel).

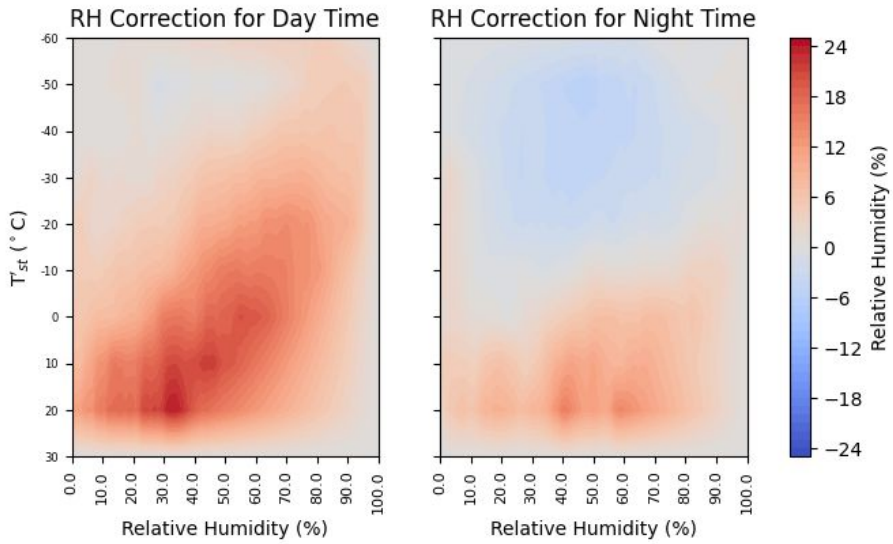


Fig. 7 The CDF-based RH correction tables for daytime (00z-12z, left panel) and night-time (12z-00z, right panel).

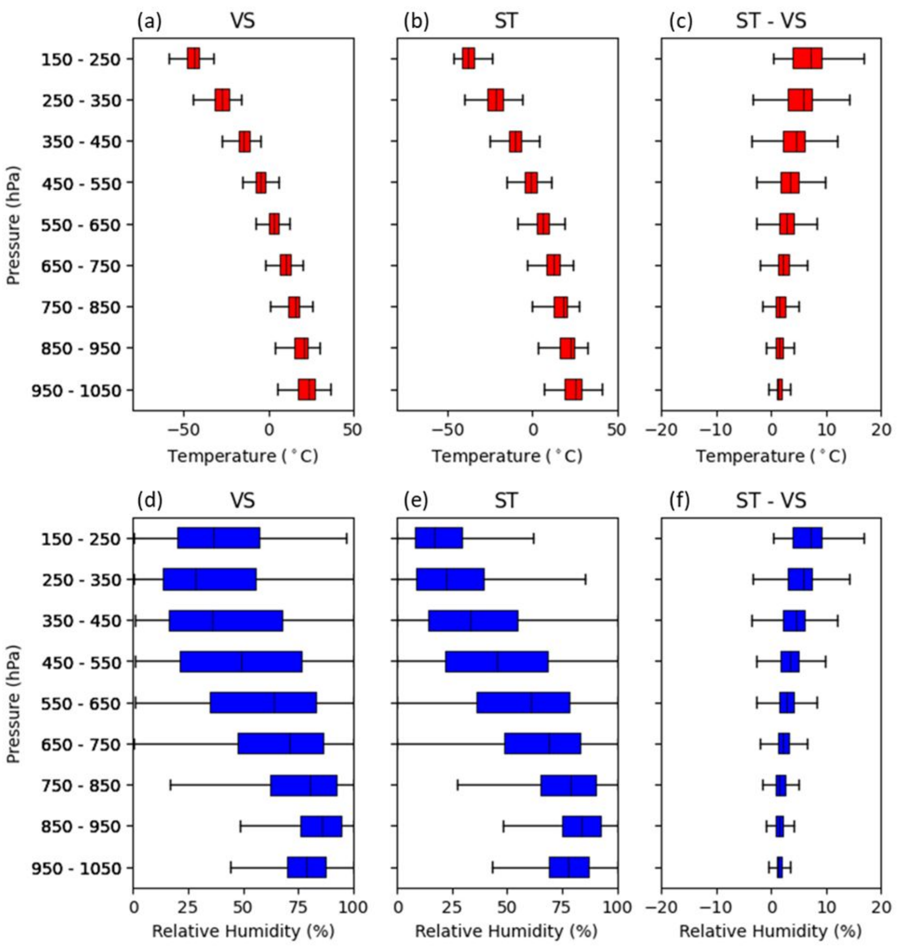


Fig. 8 The boxplot of temperatures (upper) and RH (lower) of ST (left), VS (center), and their difference (right) without corrections.

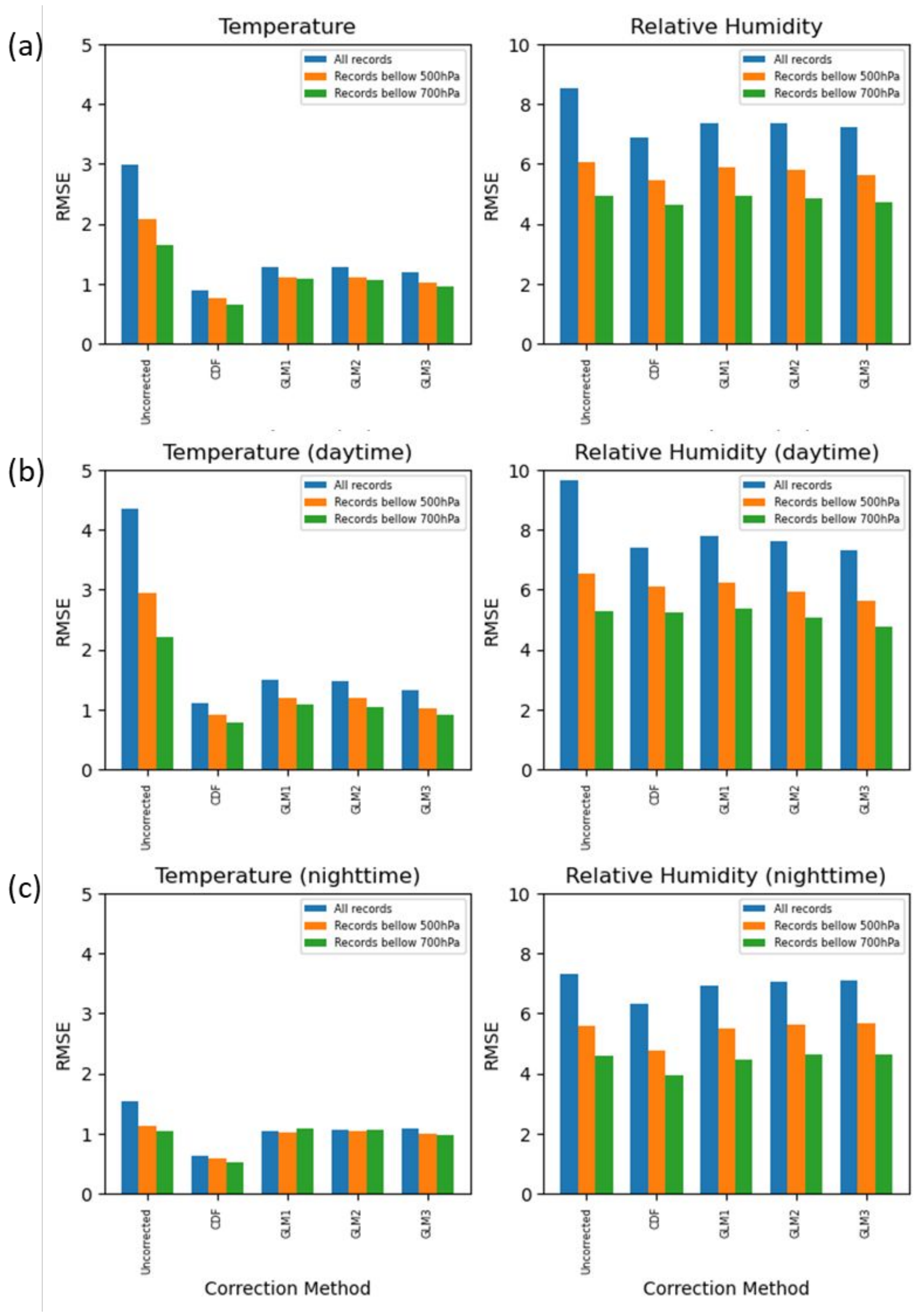


Fig. 9 The mean RMSE of ST and VS with different correction methods for temperature (left) and RH (right). For each correction method, the mean RMSE is derived with all available records (blue), records below 500 hPa (orange), and records below 700hPa (green). The upper panel (a) showed the overall RMSE, and the middle (b) and lower panel (c) demonstrated the RMSE of daytime and nighttime, respectively

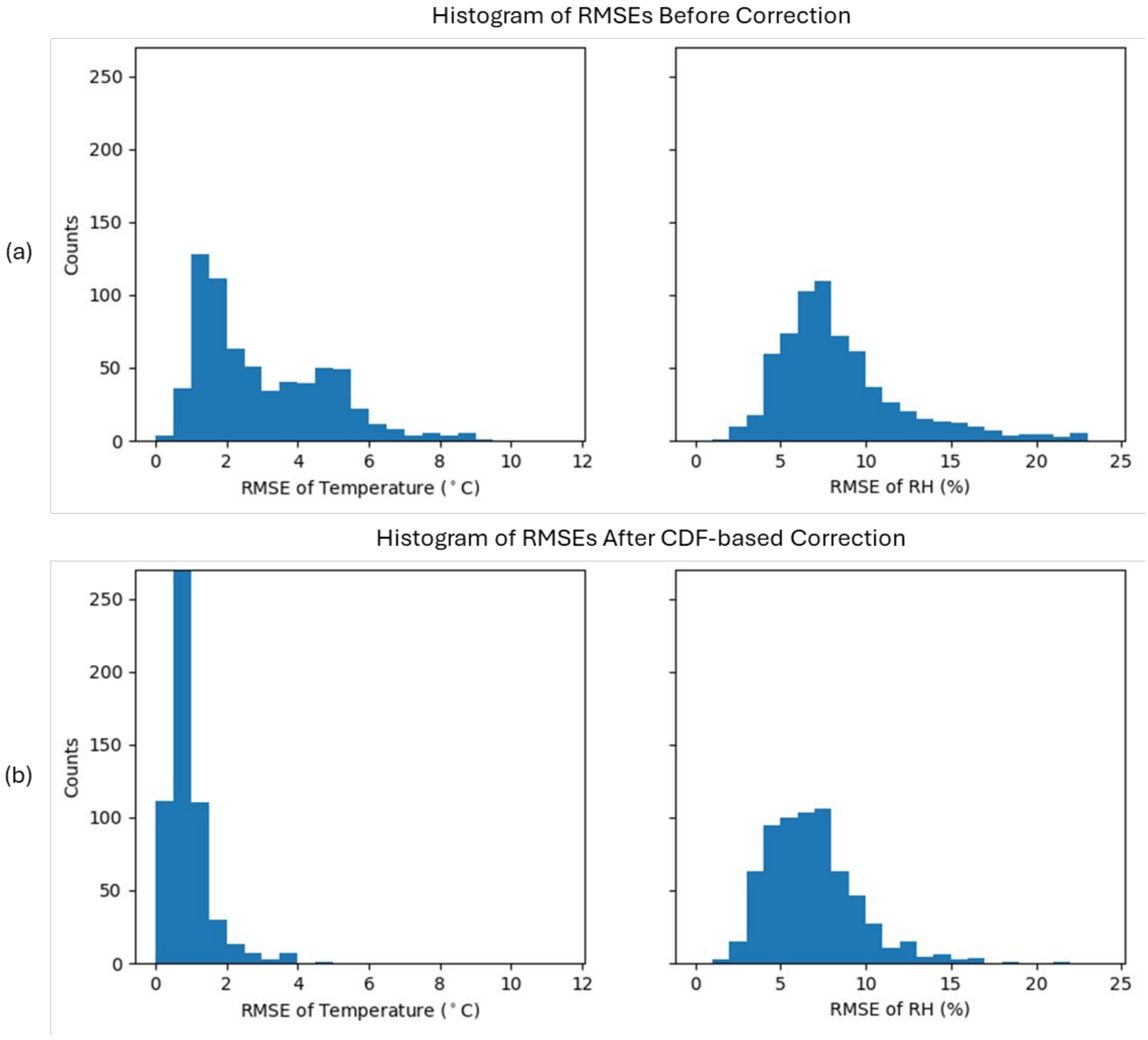


Fig. 10 The histograms of the RMSEs of temperature (upper) and RH (lower) between ST and VS. The upper panel, (a), illustrates the distribution of RMSEs before correction, and the lower panel, (b), shows the results after the CDF-based correction.

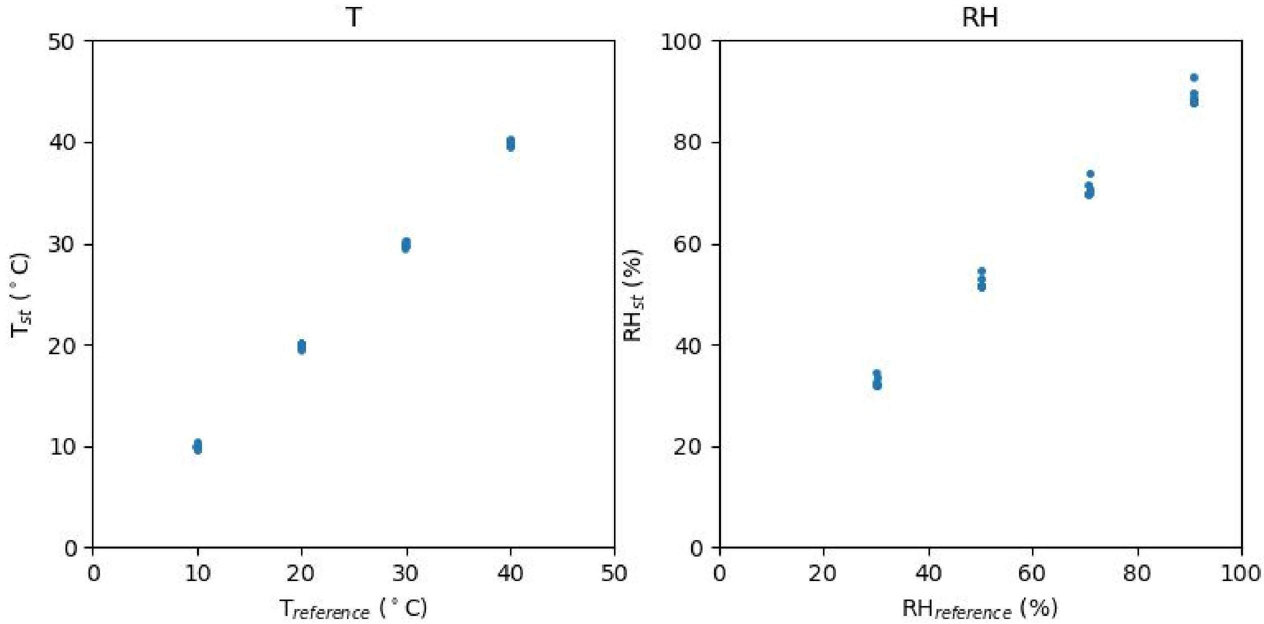


Fig. 11 The temperature (left) and RH (right) measurements of STs in the controlled laboratory environment. Six STs were measured separately. The temperature (left panel) was measured repeatedly at 10°C, 20°C, 30°C, and 40°C. The relative humidity (right panel) was measured at 30%, 50%, 70%, and 90%. The derived random error for temperature is 0.24°C, and for relative humidity is 2.21%

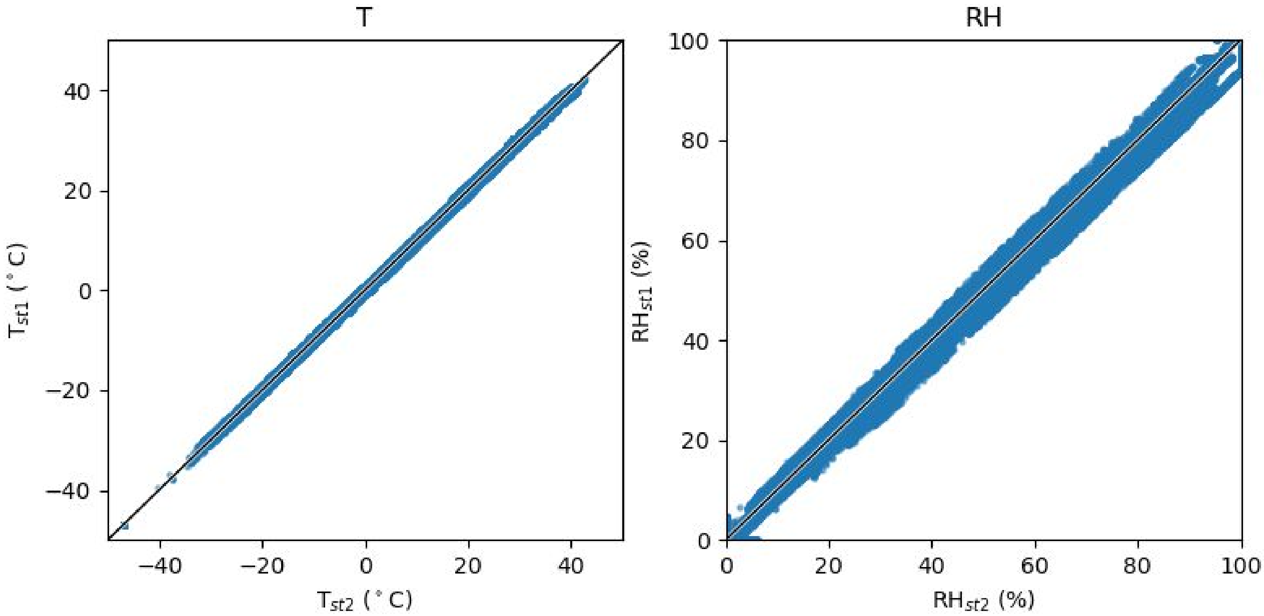


Fig. 12 The biases of 42 dual-ST launches. The figure shows 85,641 temperature measurements (left panel) and 81,616 pairs of relative humidity (right panel). The derived random error for temperature is 0.52°C, and for relative humidity is 2.25%.

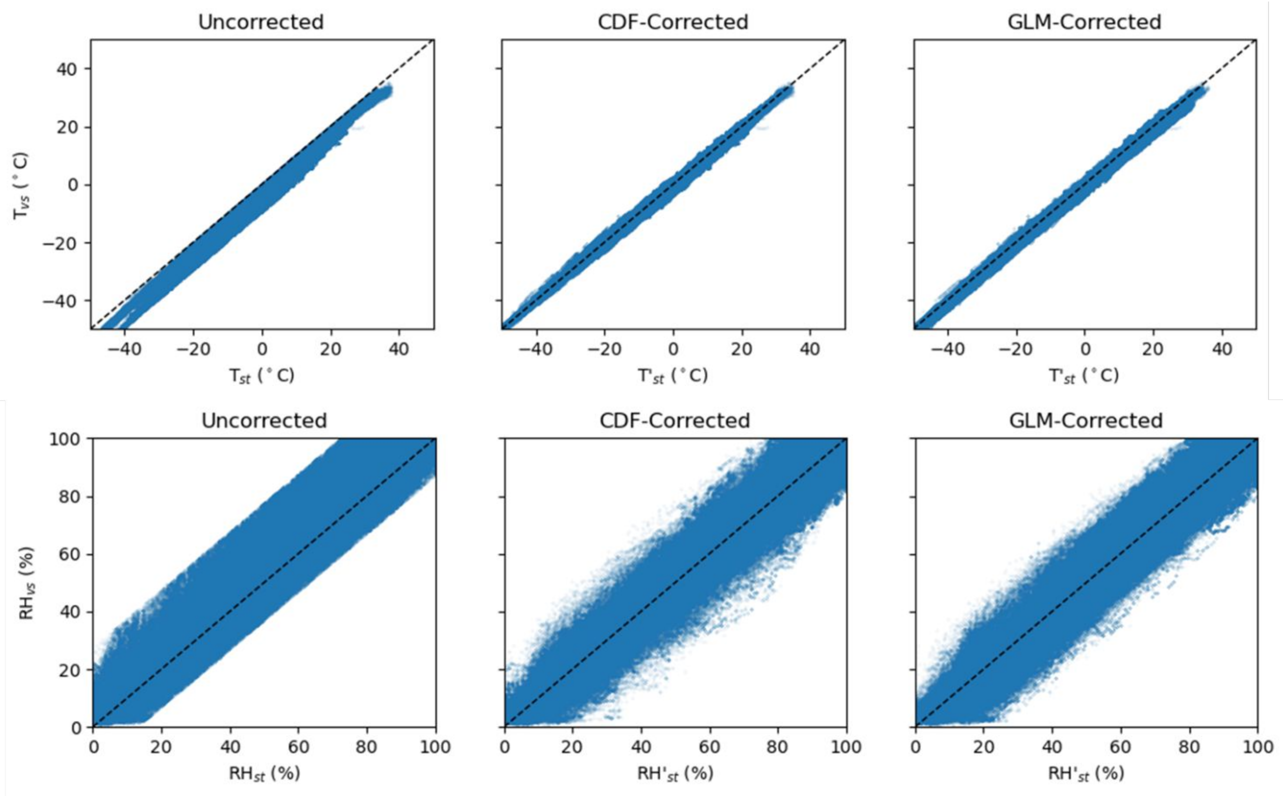


Fig. 13 The scatter plots of temperature (upper) and RH (lower) before and after correction. The dashed lines indicate 1-to-1 reference lines.

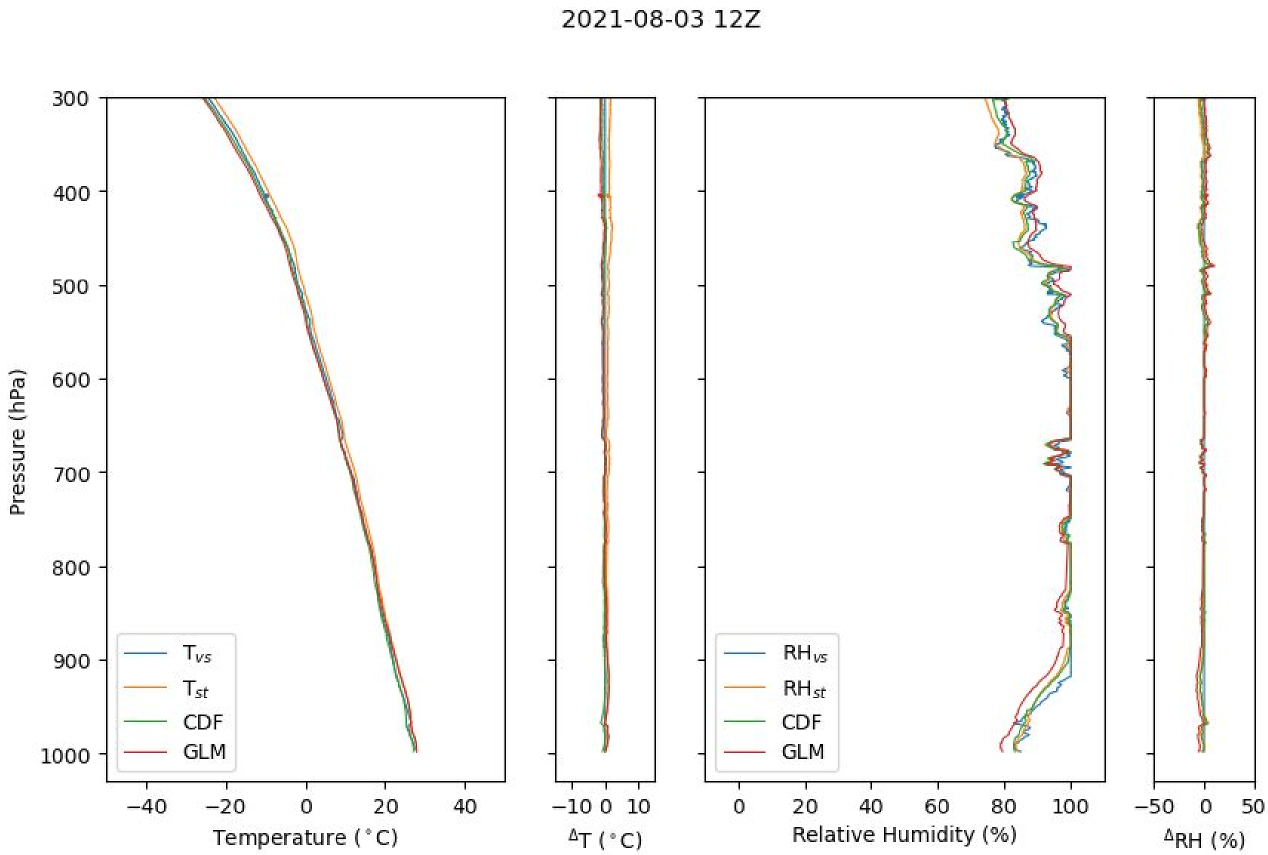


Fig. 14 The temperature (left) and RH (right) of the 2021-0803 12Z co-launch sounding. The reference (VS) is illustrated in blue, the ST in range, CDF-corrected in green, and GLM-corrected in red.

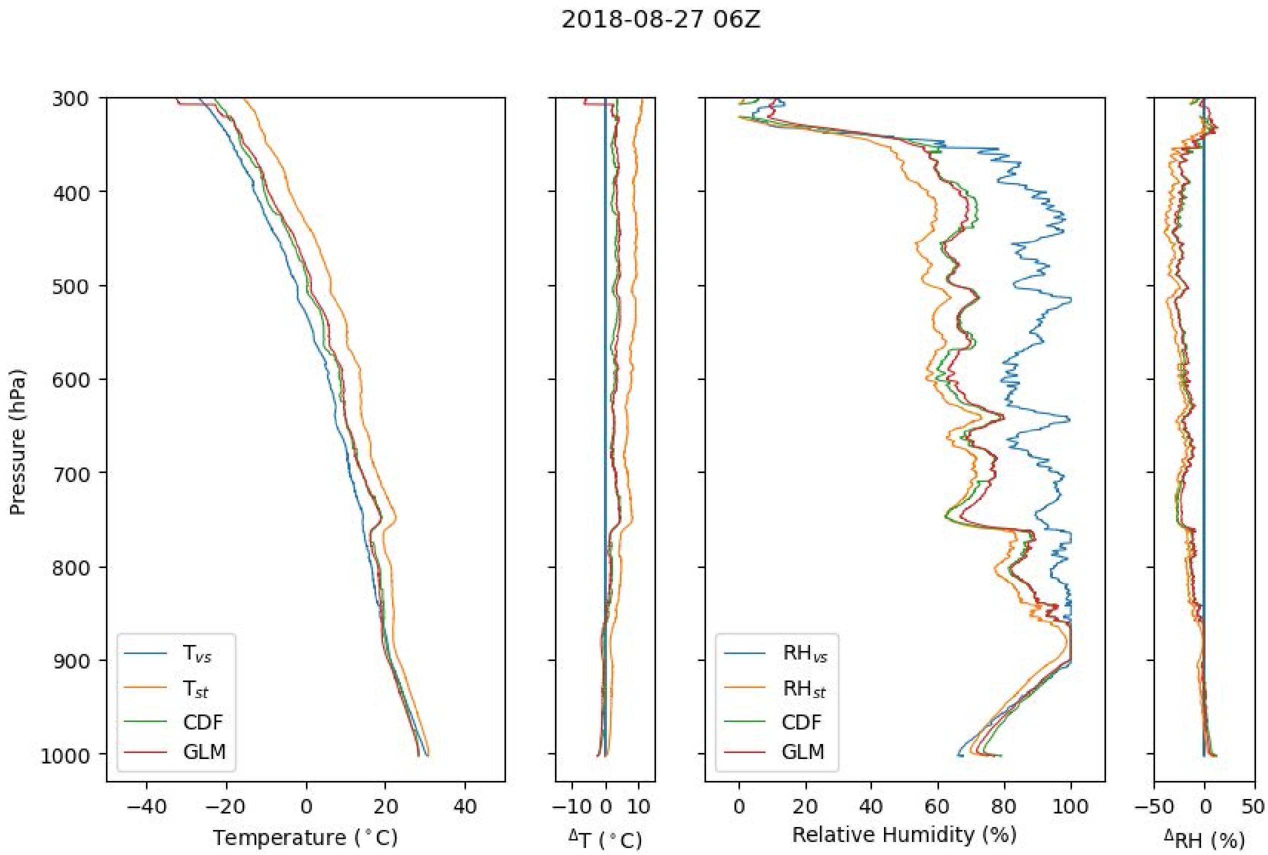


Fig. 15 The temperature (left) and RH (right) of the 2018-08-27 06Z co-launch sounding. The reference (VS) is illustrated in blue, the ST in range, CDF-corrected in green, and GLM-corrected in red.

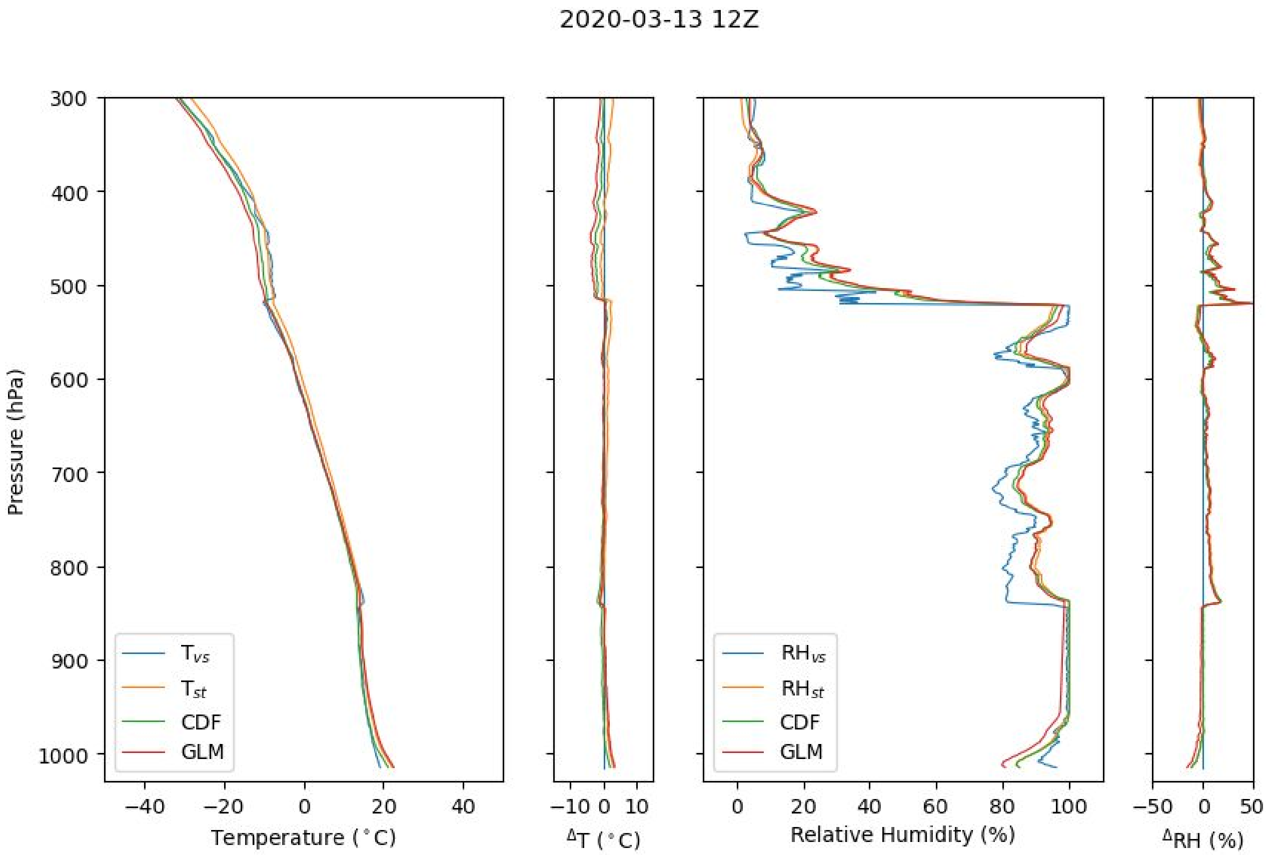
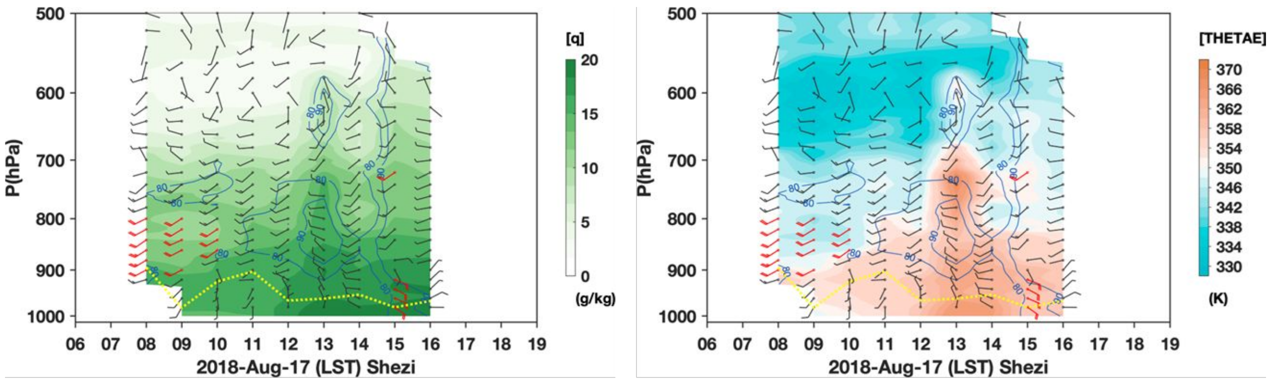


Fig. 16 The temperature (left) and RH (right) of the 2020-03-13 12Z co-launch sounding. The reference (VS) is illustrated in blue, the ST in range, CDF-corrected in green, and GLM-corrected in red.



(a) (b)

Fig. 17 The continuous ST observations of one-hour intervals on 2018-08-17 at Shezi. The soundings were corrected with CDF, and the derived specific humidity, q, is shown in panel (a) with the wind field. The derived equivalent potential temperature, Ɵe, is shown in panel (b). Note that this field experiment used only ST for observation and the data was not included in the colaunch dataset.

List of Tables

Table 1 The summary of the field experiments conducting ST-VS co-launches.

Experiment	Time	Location	Total Numbers of ST-VS Co-launch
Taipei Summer Storm Experiment (**TASSE**)	2018-2020	Taipei (Banqiao)	478
Yilan Experiment of Severe Rainfall (**YESR2020**)	2020. Nov	Yilan, Suaou, Luodong, Dafu	46
Taiwan-Area Heavy rain Observation and Prediction Experiment (**TAHOPE**)	2019-2022	Taipei (Banqiao), Pengjiayu	382
Northern Coast Observation, Verification, and Investigation of Dynamics (**NoCOVID21**)	2021. May-Jun	Taipei (Banqiao)	49
Mountain Cloud Climatology (**MCC**)	2022. Oct-Nov	Suaou	23
Other		Tainan, Xinwu	51

Table 2 The summary of the 1,029 co-launches.

Month	2018		2019		2020		2021		2022		Total	
	Day	Night	Day	Night	Day	Night	Day	Night	Day	Night	Day	Night
1					21	24			25	26	46	50
2					29	29					29	29
3					27	31					27	31
4					30	30	15	13			45	43
5					6	5	6	4			12	9
6	14		20				30	32			64	32
7	14		60	12			22	23			96	35
8	41		85				23	22			149	22
9							25	26			25	26
10							29	28	7	3	36	31
11					20	17	40	41	6	7	66	65
12							30	31			30	31
Total	69	0	165	12	133	136	220	220	38	36	625	404

Table 3 The RMSE of ST and VS with different correction methods for temperature and RH.

Variable	Correction Method	mean RMSE			stdev of RMSE		
		full	500hPa	700hPa	full	500hPa	700hPa
Temperature	Uncorrected	2.9969	2.0753	1.6446	1.8399	1.2291	0.8894
	CDF	0.8778	0.7568	0.6560	0.5579	0.4166	0.3367
	GLM1	1.2714	1.1126	1.0732	0.6612	0.4549	0.3682
	GLM2	1.2745	1.1128	1.0533	0.6625	0.4633	0.3693
	GLM3	1.1991	1.0105	0.9483	0.6284	0.4566	0.3579
Relative Humidity	Uncorrected	8.5265	6.0721	4.9336	3.8236	2.9284	2.3624
	CDF	6.8946	5.4707	4.6098	2.8107	2.7488	2.4442
	GLM1	7.4604	5.8673	4.9267	2.9158	2.6489	2.3084
	GLM2	7.4152	5.7997	4.8478	2.7785	2.4307	2.0590
	GLM3	7.2683	5.6355	4.7043	2.6668	2.3372	1.9878