JGR Atmospheres

RESEARCH ARTICLE
10.1029/2021JD035220

The New Radiosounding HARMonization (RHARM) Data Set of Homogenized Radiosounding Temperature, Humidity, and Wind Profiles With Uncertainties

Fabio Madonna[1], Emanuele Tramutola[1], Souleymane SY[1], Federico Serva[2], Monica Proto[1], Marco Rosoldi[1], Simone Gagliardi[1], Francesco Amato[1], Fabrizio Marra[1], Alessandro Fassò[3], Tom Gardiner[4], and Peter William Thorne[5]

[1]Consiglio Nazionale delle Ricerche—Istituto di Metodologie per l'Analisi Ambientale (CNR-IMAA), Tito Scalo (Potenza), Italy, [2]Consiglio Nazionale delle Ricerche—Istituto di Scienze Marine (CNR-ISMAR), Rome, Italy, [3]University of Bergamo, Bergamo, Italy, [4]National Physical Laboratory, Teddington, UK, [5]Department of Geography, Irish Climate Analysis and Research Units, Maynooth University, Maynooth, Ireland

Key Points:

- The Radiosounding HARMonization (RHARM) data set provides homogenized time series of temperature, relative humidity and wind profiles for hundreds of radiosonde stations
- Metrologically accurate estimates of the observational uncertainty are provided together with each observation and variable
- For all variables, RHARM increases the geographical coherency of estimated trends and the agreement with a modern atmospheric reanalysis

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

F. Serva and F. Madonna,
federico.serva@artov.ismar.cnr.it;
fabio.madonna@imaa.cnr.it

Citation:

Madonna, F., Tramutola, E., SY, S., Serva, F., Proto, M., Rosoldi, M., et al. (2022). The new Radiosounding HARMonization (RHARM) data set of homogenized radiosounding temperature, humidity, and wind profiles with uncertainties. *Journal of Geophysical Research: Atmospheres*, *127*, e2021JD035220. https://doi.org/10.1029/2021JD035220

Received 11 MAY 2021
Accepted 20 DEC 2021

Author Contributions:

Conceptualization: Fabio Madonna, Alessandro Fassò
Data curation: Emanuele Tramutola, Souleymane SY, Federico Serva, Monica Proto, Simone Gagliardi, Francesco Amato, Fabrizio Marra
Formal analysis: Fabio Madonna, Emanuele Tramutola, Federico Serva, Fabrizio Marra

Abstract Observational records are more often than not influenced by residual non-climatic factors which must be detected and adjusted for prior to their usage. In this work, we present a novel approach, named Radiosounding HARMonization (RHARM), providing a homogenized data set of temperature, humidity and wind profiles along with an estimation of the measurement uncertainties for 697 radiosounding stations globally. The RHARM method has been used to adjust twice daily (0000 and 1200 UTC) radiosonde data holdings at 16 pressure levels in the range 1,000–10 hPa, from 1978 to present, provided by the Integrated Global Radiosonde Archive. Relative humidity (RH) data are limited to 250 hPa. The applied adjustments are interpolated to all reported levels. RHARM is the first data set to provide homogenized time series with an estimation of the observational uncertainty at each sounding pressure level. By construction, RHARM adjusted fields are not affected by cross-contamination of biases across stations and are fully independent of reanalysis data. Analysis of trends for temperature, RH and winds highlights increased geographical coherency of trends over 1978–2000 globally, but especially in the Northern Hemisphere and South America. RHARM shows warming trends of 0.39 K/decade at 300 hPa in the Northern Hemisphere and of 0.25 K/decade in the tropics. The RHARM adjustments also reduce differences with the European Centre for Medium-Range Weather Forecast ERA5 reanalysis, with the strongest effect in the Northern Hemisphere for temperature and relative humidity. For wind speed, the comparison indicates a good agreement with ERA5 in the troposphere.

Plain Language Summary Systematic observations of atmospheric parameters by balloon soundings have been collected regularly across the globe since decades. However, changes in the instrumentation, calculation algorithms, station re-locations and other factors can influence the measurement time series, making the identification of climate signals difficult. The availability of measurement metadata, cross-comparison with (sparser) reference networks and the application of statistical methods allows to greatly improve the physical consistency of the records. We developed the Radiosounding HARMonization (RHARM) methodology to provide users with homogenized radiosonde profiles of temperature, humidity, and wind, along with measurement uncertainties, for 697 locations globally and twice daily based on the Integrated Global Radiosonde Archive. Key features of RHARM are the provision of multiple fundamental atmospheric parameters, adjusted for spurious biases and the inclusion of measurement uncertainties estimated following the standards of the metrological community when available metadata permits. The current version of the RHARM data set can be used to investigate tropospheric weather and climate for the last four decades.

1. Introduction

Long-term homogeneous climate data records (CDRs) are essential to diagnose changes in our climate, understand its variability, and assess and contextualize future climate projections (Cramer et al., 2018). Use of CDRs influenced by residual non-climatic factors may lead to incorrect conclusions about the changing state of the climate (Kivinen et al., 2017). Therefore, when CDRs are used it is highly desirable to:

Methodology: Fabio Madonna, Federico Serva, Marco Rosoldi, Peter William Thorne
Resources: Fabio Madonna
Software: Emanuele Tramutola, Francesco Amato, Fabrizio Marra
Validation: Souleymane SY
Writing – original draft: Fabio Madonna, Marco Rosoldi
Writing – review & editing: Monica Proto, Simone Gagliardi, Tom Gardiner, Peter William Thorne

1. Detect and adjust for all the known and quantifiable systematic inhomogeneities in the observational record, arising from a variety of causes (changes in station location, instrumentation, calibration or drift issues, different instrument sensitivity across different networks, changes in the measurement procedures, etc.);
2. Establish measurement traceability ideally to an absolute reference (Système international, SI), or community acknowledged "standard" through an unbroken chain of calibrations, each contributing to the measurement uncertainty;
3. Quantify measurement uncertainties in any data where traceability was not properly established; in such cases, uncertainties must be inferred from the available metadata, results of sensors' intercomparisons, or information about the measurement process.

In practice, for historical in-situ observations it is often not easy to fulfill the above list of requirements, especially for global baseline or comprehensive networks (Thorne et al., 2017). Where commonly the metadata and original pre-processed data (e.g., digital sensor counts, hereinafter raw data) are either missing or retained solely by individual station PIs (if at all) and not routinely shared or stored in data archives.

This is the case for radiosounding measurements of temperature (T), relative humidity (RH), and wind which still represent anchor information for many meteorological applications, despite the advent of Global Navigation Satellite System-Radio Occultation (GNSS-RO) measurements which have proven valuable for data assimilation purposes (Bauer et al., 2014). Since the mid-20th century, radiosounding measurements are the only data source continuously available to study climate variability and change in the troposphere and lowermost stratosphere. They also constitute a valuable source of information for satellite cal/val activities (Calbet et al., 2017; Finazzi et al., 2019; Loew et al., 2017). In the ERA-Interim European Centre for Medium-Range Weather Forecasts (ECMWF) reanalysis (Dee et al., 2011), the conventional observing system which includes radiosoundings, despite proportionately low data volumes, still represents an indispensable constraint (Haimberger et al., 2012). A similar situation exists for the latest ECMWF ERA5 reanalysis (Hersbach et al., 2020) as well as for other recent global reanalyses (e.g., Gelaro et al., 2017; Kobayashi et al., 2015).

Quality and biases of radiosounding observations strongly vary with sensor type, altitude level, and through time. Many previous works described the adjustment of historical radiosounding temperature measurements to construct CDRs (e.g., Dai et al., 2011; Free et al., 2004; Haimberger, 2005; Haimberger et al., 2012; McCarthy et al., 2008; Sherwood et al., 2008; Thorne, Parker, et al., 2005; Zhou et al., 2021). These works have used a broad range of approaches enabling an exploration of structural uncertainty (Thorne, Christy, & Mears, 2005). Several products additionally include ensemble approaches to explore parametric uncertainty (Haimberger et al., 2012; Sherwood & Nishant, 2015; Thorne et al., 2011). Application of innovative statistical approaches has been recently proposed for the production of future datasets (Fassò et al., 2018). Datasets to date have not, however, taken direct benefit from either periodic intercomparisons (parallel measurement campaigns) or the work of the Global Climate Observing System (GCOS) Reference Upper Air Network (GRUAN).

Intercomparison datasets made available by various research organizations, institutions and manufacturers represent an invaluable source of information which improves the interpretation of effects, drifts and inhomogeneities in the recorded time series. Most notable are the periodic intercomparison campaigns that have been organized by the World Meteorological Organization/Commission for Instruments and Methods of Observation (WMO/CIMO), involving the vast majority of commercial manufacturers and providing a thorough periodical assessment of inter-sensor differences (e.g., Nash et al., 2006, 2011). These intercomparison exercises involve the flying of multiple sonde models on the same rig, enabling an evaluation of the relative performance of various sensors under the full range of conditions experienced at the location and time of the comparison. The most recent intercomparison, held at Yangjiang (China) in 2010, involved 11 manufacturers and for the first time three manufacturers from China. Different groups of radiosondes were intercompared by launching them on the same payload; permitting robust comparisons from at least 25 flights for each radiosonde type up to 20 km altitude.

To address the need of providing homogeneous and fully traceable upper-air measurements with quantified uncertainties, GRUAN was established in 2006 (Bodeker et al., 2016). GRUAN aims to provide reference-quality observations of Essential Climate Variables (ECVs, Bojinski et al., 2014) above the Earth's surface. GRUAN is providing long-term, high-quality radiosounding data at 30 sites (12 sites are certified to date) around the world with characterized uncertainties, ensuring the traceability to SI units or accepted standards, providing extensive metadata and comprehensive documentation of measurements and algorithms. GRUAN data processing starts

from the raw data and applies a number of SI-traceable adjustments (e.g., due to solar radiation, measurement sensors' time-lag, sonde pendulum motion, etc.), each with a quantified uncertainty contributing to the final uncertainty budget. As a reference network, GRUAN provides a potential basis for enhanced interpretation of broader radiosonde networks, for example, through the provision of instrumental corrections which can be extended to non-GRUAN stations to adjust quantifiable systematic effects compromising the quality of operationally processed data (JCGM100, 2008).

Taking advantage of GRUAN and intercomparison data we have designed and applied a novel algorithm for homogenizing historical radiosounding time series available since 1978. The new Radiosounding HARMonization (RHARM) approach discussed herein is a hybrid method based on two main steps:

1. Adjustment of radiosounding observations of temperature, humidity and wind from 2004 to present using the GRUAN data and algorithms, as well as the 2010 WMO/CIMO radiosonde intercomparison data set (hereinafter ID2010, Nash et al., 2011), with quantification of uncertainties;
2. Identification of change-points in the earlier portions of the time series (before 2004 and as early as 1978) and adjustment of non-climatic effects using statistical methods with related quantification of uncertainties.

The present paper provides an analytical description of the RHARM algorithm and an assessment of key characteristics of the data set comprising of 697 radiosounding stations available from the Integrated Global Radiosonde Archive (IGRA, Durre et al., 2006, 2018). Only data since 1978 are homogenized as before then radiosounding reports at mandatory pressure levels were not frequent and homogeneous.

The RHARM approach increases the limited number of existing homogenized datasets, which includes:

1. homogenized radiosounding temperature measurements: Radiosonde Atmospheric Temperature Products for Assessing Climate (RATPAC) by NOAA (Free et al., 2004), RAdiosonde OBservation COrrection using REanalyses (RAOBCORE), Radiosonde Innovation Composite Homogenization (RICH) by the University of Wien (Haimberger et al., 2012), Hadley Centre's radiosonde temperature product v2 (HadAT2) by Met Office (Thorne, Parker, et al., 2005), Iterative Universal Kriging v2 (IUKv2) by University of New South Wales (Sherwood & Nishant, 2015), the State University of New York Albany data set (Zhou et al., 2021);
2. homogenized radiosounding humidity measurements: the Homogenized RS92 radiosounding humidity measurements (HomoRS92) by State University of New York Albany (Dai et al., 2011) and the Hadley Centre's radiosonde temperature and humidity product (HadTH) (McCarthy et al., 2009); and
3. homogenized radiosounding wind datasets: IUKv2 and GRASPA (Ramella-Pralungo & Haimberger, 2014; Ramella Pralungo et al., 2014).

Distinct from previous efforts, RHARM is the first data set to provide homogenized time series of temperature, relative humidity and wind in the same package. Moreover, RHARM is based on the use of "reference measurements" to calculate and adjust for systematic effects, instead of using background information provided by meteorological reanalysis, autoregressive models or neighboring stations. RHARM adjusted fields are not affected by cross-contamination of biases across stations (Sherwood, 2007) and are fully independent of reanalysis data (Haimberger et al., 2012). In addition, each harmonized data series is provided with an estimation of the measurement uncertainty. RHARM is also valuable in providing adjustments for each individual radiosounding profile, not only at mandatory (https://glossary.ametsoc.org/wiki/Mandatory_level) but also at all the report significant levels (https://glossary.ametsoc.org/wiki/Significant_level).

The remainder of this paper is organized as follows. In Section 2, the data sources used are outlined. In Section 3, a detailed overview of the RHARM data processing for the observations post-2004 is provided followed by a description of the detection of breakpoints and the adjustments for the period before 2004. In Section 4, statistics of adjustments applied by RHARM in comparison with IGRA and ECMWF ERA5 reanalysis data and comparisons of the trend profiles in the troposphere and lower stratosphere are discussed for all the variables. In Section 5, statistics on the correlation of the identified breakpoints at different pressure levels is presented. Discussion and conclusions are provided in Section 6. Additional information on the consistency with GRUAN data and the validation of uncertainties as well as additional examples and comparisons between IGRA and RHARM are available in the appendices and Supporting Information S1.

2. Data Sources Used

RHARM is applied to IGRA Version 2 (Durre et al., 2018) which incorporates data from a considerably greater number of data sources with an increased data volume by 30% compared to Version 1. A subset of 697 radiosounding stations and radiosoundings from ships are retained based upon documented metadata (i.e., including the radiosonde code according to WMO table 3685, describing the radiosonde type) since 2000 and for a subset of these stations since 1978 (full metadata record provided by the station PIs). Depending on the radiosonde type, adjustments based on the application of GRUAN-like data processing or on the comparison between GRUAN data and ID2010 can be applied to the post-2004 period, for which several instrumental effects are already corrected (e.g., the well-known solar radiation dry bias, Dirksen et al., 2014).

The IGRA data v2 are the result of improved quality assurance procedures developed for the IGRA data v1 (Durre et al., 2006, 2008), which can be grouped into eight categories: fundamental "sanity" checks, checks on the plausibility and temporal consistency of surface elevation, internal consistency checks, checks for the repetition of values, checks for gross position errors in ship tracks, climatology-based checks, checks on the vertical and temporal consistency of temperature, and data completeness checks. The RHARM data set thus inherits the IGRA quality assurance procedures, and additional quality checks are then applied on: the metadata availability; physical plausibility; data completeness check; accuracy of the bias adjustment; removal of outliers; vertical correlation between structural breaks at the same station; coherency check for the adjustments applied at the significant levels.

As noted, the RHARM approach is applied on a subset of IGRA stations, depending on the availability of metadata (Durre et al., 2008; Ferreira et al., 2019). For these stations, a quality-enhanced data set with a sufficient number of radiosoundings available over 2004 to present are provided directly post-processing the profiles to account for several instrumental effects. The post-processed profiles are then used as reference information to adjust the systematic effects in the historical data before 2004. For those stations where the number of post-processed radiosoundings profiles is not sufficient for the purposes of the homogenization algorithm before 2004, the post-processed profiles since 2004 are provided only.

The GRUAN data v2 includes data from 26 sites providing radiosounding data since 2008, although with different lengths and completeness. Data availability can be found on the GRUAN website (www.gruan.org). The GRUAN data product is fully traceable to SI units and processed as described in Dirksen et al. (2014). As part of the product, GRUAN provides a full and extensive set of metadata which should enable to fully reprocess the raw data or to properly adjust unknown effects in future. For example, in the case of a radiosonde launch, a complete description of the set-up is required that includes the description of the balloon, the gas, filling weight, unwider type and length and so on. At present, there are only two GRUAN data products (GDPs), for the Vaisala RS92 and for Meisei RG11 sondes. RHARM applies adjustments to RS92 Vaisala sondes only, which represents a substantive portion of the global data. For the Meisei RG11 GDP, its recent introduction (Kobayashi et al., 2019) precluded its implementation within RHARM so far, but an update of the data processing will be implemented in the near future for any other radiosonde GDP which might become available.

The coverage of RHARM is reasonably homogeneous (Figure 1): it is one of the broadest datasets for the South America, while there is a sufficient coverage over Siberia. For the latter, limited information is available on the main radiosonde type used in the region since 2004 (AVZ), which cannot be adjusted using RHARM. The station density in Canada, Northeast Asia, and East Africa is lower than in Europe, U.S. and South America, but this is common to all datasets and reflects the inadequacies of the historical observing system. Table 1 confirms the lower number of measurements available in the Southern Hemisphere (SH) than at other latitudes, although the quantity of measurements alone cannot address the value of the data set for a specific study without considering representativeness (Weatherhead et al., 2017).

In the following sections, IGRA and RHARM datasets are also compared with the ECMWF ERA5 reanalysis (Hersbach et al., 2020). ERA5 is one of the most utilized datasets for climate studies and, although it cannot be considered a reference data set like GRUAN, reanalyses are often used to study datasets homogeneity (e.g., Haimberger et al., 2012; Zhou et al., 2021). ERA5 incorporates millions of observations into a data assimilation system, every 6–12 hr over the period being analyzed, providing a systematic approach to produce a data set for climate monitoring and research. ERA5 is the latest climate reanalysis produced by ECMWF providing hourly data on regular latitude-longitude grids at $0.25° \times 0.25°$ resolution and on 37 pressure levels. ERA5 is publicly

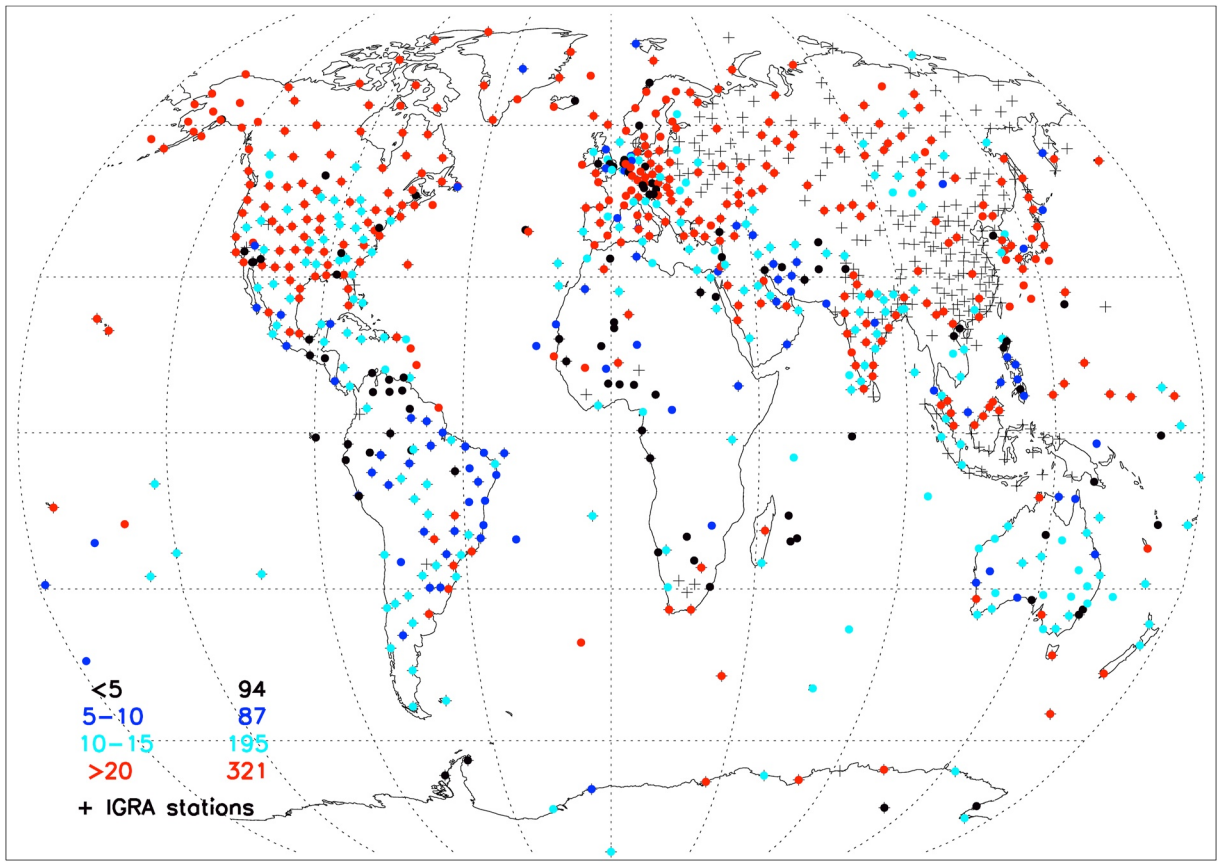

Figure 1. Global distribution and quantity of Radiosounding HARMonization homogenized profiles. The scale in the bottom left corner denotes the cumulative number of available radiosoundings at each station (in millions of ascents) from 1978 to present. The + symbol indicates Integrated Global Radiosonde Archive (IGRA) stations (1,156) reporting data since 1978 to present (last access to IGRA 31-12-2020).

available through the Copernicus Climate Data Store (CDS, https://cds.climate.copernicus.eu). For the purposes of the validation of uncertainties, described in Appendix A, we also use the ERA5 background (6-hr forecast) as a comparator value. The various reanalysis products have proven to be valuable when used appropriately (Dee et al., 2011). Nevertheless, reanalysis reliability can considerably vary depending on the location, time period, and variable considered (Dee et al., 2016). The changing mix of observations, and biases in observations and models, can introduce spurious variability and trends into reanalysis output (Dee et al., 2016).

For the comparison with RHARM data, which are not gridded, ERA5 fields have been sub-sampled to match the location of RHARM stations using the nearest grid point to each station. The same approach has been used for the ERA5 background data considered for the uncertainty validation. Considering the high spatial resolution of

Table 1
Number and Fraction of Launches in Different Latitude Bands for the Stations Shown in Figure 1

Region	Latitude range	Number of launches (thousands)	Fraction of total launches (%)
Arctic	70–90 N	316.1	2.5
Northern Hemisphere mid-latitudes	25–70 N	8203.7	65.4
Tropics	25 N–25 S	2979.3	23.8
Southern Hemisphere mid-latitudes	25–70 S	974.0	7.8
Antarctica	70–90 S	64.2	0.5
Total		12537.3	100

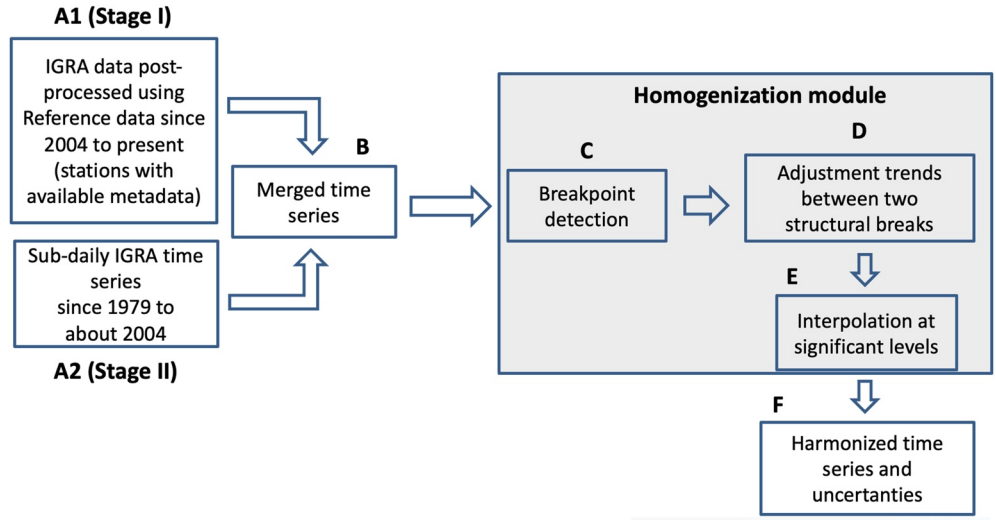

Figure 2. Schematic diagram describing the steps of the Radiosounding HARMonization approach.

ERA5 and its representativeness, the uncertainty due to the use of the nearest grid-point interpolation should be comparable with other methods (like kriging or bilinear interpolation).

3. Methodology

The RHARM homogenization of global radiosounding temperature, humidity and wind profiles is applied to each balloon launch (generally 00:00 and/or 12:00 UTC) data on 16 mandatory pressure levels (10, 20, 30, 50, 70, 100, 150, 200, 250, 300, 400, 500, 700, 850, 925, 1,000 hPa), because these levels are available from the stations for each ascent, whereas significant level reports vary by definition per profile. Relative humidity (RH) adjustments are limited to 250 hPa owing to pervasive sensor performance issues at greater altitudes in almost all commercial sondes (Miloshevich et al., 2004).

The RHARM algorithm works through the following steps:

1. For each time series (i.e., station), data since 2004 (with starting time station-dependent) are obtained by post-processing each single radiosounding profile using a GRUAN-like algorithm; these data are labeled "Stage I" time series;
2. The Stage I time series is merged with the preceding radiosounding time series, hereinafter named the "Stage II" time series;
3. The merged time series are then divided in two sub-series to separate the nominal 00 UTC and 12 UTC launches, which are the two most frequent launch times in IGRA;
4. Profiles are first adjusted at mandatory pressure levels and, therefore, a breakpoint detection method is applied to the night and daytime time series, at each mandatory level;
5. The trend of the Stage II time series is adjusted using the Stage I time series as a constraint;
6. The observational uncertainties are estimated for each data point;
7. Finally, adjustments and uncertainties estimated only at mandatory pressure levels are interpolated at significant levels; uncertainties are estimated for each processing step and propagated to estimate the total uncertainty.

The concatenation of Stage II and Stage I time series provides the entire time series for each station, and only those stations satisfying the requirements to produce a Stage I time series are considered for the Stage II time series calculations. An overall scheme of the RHARM approach is shown in Figure 2.

Local nighttime and daytime conditions for each radiosounding launch are identified by calculating the solar zenith angle using the LOWTRAN module (available at http://ethangutmann.com/pages/idl/Utilities/zensun.pro, last access on 31 December 2020), using as inputs each radiosonde launching time and the corresponding station

geographical coordinates. The small number of launches available at other synoptic hours have not been considered in the current RHARM data version. The step A1 in Figure 2 is critical mainly for temperature and humidity where radiation-heating effects can have substantive impacts on instrument performance (Dirksen et al., 2014; Miloshevich et al., 2004; Wang et al., 2013). The same separation is made for the wind profile to keep cross-variable processing consistency and because in many regions of the globe there exist marked diurnal and semi-diurnal components in the variability of winds (e.g., Harris et al., 1962). Nevertheless, either using GNSS or precursor radar tracking techniques, the effect of the separate daytime and nighttime post-processing for wind speed, wind direction and the related uncertainties is negligible (not shown).

In Section 3.1 and 3.2, the approach applied to obtain the Stage I time series (Step A1 in Figure 2) is outlined. The remaining subsections describe the adjustment of the Stage II time series (Step B-D) and subsequent adjustments to the significant levels of the radiosounding profiles and the estimation of measurement uncertainties (Steps E-F).

3.1. Adjustment of Vaisala Temperature, Humidity, and Wind Profiles Since 2004

During daytime, the sensor boom of any radiosonde type is heated by solar radiation which introduces biases in temperature and humidity (Wang et al., 2013). The net heating of the temperature sensor and the resulting dry bias affecting the relative humidity sensors depends on the amount of absorbed radiation and, therefore, the solar elevation angle (α), as well as on the cooling by thermal emission and ventilation by air flowing around the sensor (Dirksen et al., 2014).

To adjust this effect in the measured profiles of temperature and RH, the first step of the RHARM algorithm, involving only the Vaisala RS92 sondes, is to apply a solar radiation correction to the T vertical profiles (both for mandatory and significant levels) similarly to the metrologically traceable GRUAN processing. This is performed in two steps:

1. First, the radiation correction, $\Delta T_{\mathrm{VAISALA}}$, applied by the manufacturer to the temperature profiles is removed (i.e., $\Delta T_{\mathrm{VAISALA}}$ is added, because the correction is applied to decrease the measured value);
2. Second, a GRUAN-like radiation correction, $\Delta T_{\mathrm{GRUAN}}$ is applied using the values of the actinic flux modeled with the Streamer RTM (Key & Schweiger, 1998) following the approach documented in Dirksen et al. (2014). Where GRUAN-like corrections cannot be applied, the manufacturer correction is left unchanged.

$\Delta T_{\mathrm{VAISALA}}$ is derived from the tables provided by the manufacturer and accounts for changes in the RS92 data processing during the sonde model's production lifetime (see https://www.vaisala.com/en/sounding-data-continuity).

The GRUAN correction, $\Delta T_{\mathrm{GRUAN}}$, is defined as:

$$\Delta T_{\mathrm{GRUAN}}(I_a, p, v) = a x^b \tag{1}$$

$$x = \frac{I_a}{pv} \tag{2}$$

where I_a is the actinic flux at the solar zenith angle of the balloon release time, calculated using the LOWTRAN v7 solar position data; p is the pressure level; and v is the ascent speed in m s^{-1}. v cannot be directly ascertained from IGRA data as times of individual observations are, in general, not archived. For this reason, an average ascent speed of 5 m s^{-1} is assumed, based on the recommended ascent speed from WMO guidance, which corresponds well to typical measured ascent speeds (e.g., Madonna, Kivi, et al., 2020). The coefficients a and b in Equation 1 derived from laboratory experiments (Dirksen et al., 2014) are a = 0.18(±0.03) and b = 0.55(±0.06).

Once $\Delta T_{\mathrm{GRUAN}}$ is calculated, the final correction following Dirksen et al. (2014) is to derive a best estimate between the two approaches:

$$\Delta T = \frac{(\Delta T_{\mathrm{GRUAN}} + \Delta T_{\mathrm{VAISALA}})}{2} \tag{3}$$

Within RHARM, the final adjustment added to IGRA temperature profiles is correspondingly:

$$\Delta T_{\mathrm{RHARM},RS92} = \Delta T_{\mathrm{VAISALA}} - \Delta T + \Delta T_r \tag{4}$$

Table 2
List of the GRUAN Stations Used to Calculate the Additional Calibration Bias Applied in the RHARM Approach to Adjust the Vaisala RS92 Radiosoundings Available From IGRA

GRUAN code	Station name and country	Latitude	Longitude	Altitude	WMO index
CAB	Cabauw, Netherlands	51.97°	4.92°	1 m	06,260
LIN	Lindenberg, Germany	52.21°	14.12°	98 m	10,393
NYA	Ny-Ålesund, Norway	78.92°	11.92°	5 m	01,004
SGP	Lamont, OK, USA	36.60°	−97.49°	320 m	74,646
SOD	Sodankylä, Finland	67.37°	26.63°	179 m	02,836
TAT	Tateno, Japan	36.06°	140.13°	25 m	47,646

Note. GRUAN, Global Climate Observing System (GCOS) Reference Upper Air Network; IGRA, Integrated Global Radiosonde Archive; RHARM, Radiosounding HARMonization; WMO, World Meteorological Organization.

where ΔT_r is a residual calibration bias calculated from the mean difference of GRUAN and IGRA nighttime temperature profiles at mandatory pressure levels for the six GRUAN sites reported in Table 2. To calculate ΔT_r, outliers are filtered using a Z-score method removing values outside ±6 standard deviations. ΔT_r is added to both night and daytime profiles.

If the value of I_a in Equation 2 is equal to zero (i.e., $\Delta T = 0$), the manufacturer radiation correction applied to IGRA profiles is not modified and Equation 4 reduces to $\Delta T_{\mathrm{RHARM},RS92} = \Delta T_r$. Equation 4 removes the solar radiation correction applied by the manufacturer and adjusts the data using the GRUAN correction plus an additional term minimizing, on average, the difference with the GRUAN processing.

The standard uncertainty (coverage factor $k = 1$, confidence that 68% of values lie within one standard deviation) on $T_{\mathrm{RHARM},RS92}$, $\varepsilon(T_{\mathrm{RHARM},RS92})$, is calculated according to the following equation:

$$\varepsilon(T_{\mathrm{RHARM},RS92}) = \sqrt{\sum_i \varepsilon^i_{\mathrm{systematic}}(\Delta T)^2 + \varepsilon_R(\Delta T)^2} = \\ = \sqrt{\varepsilon_{c,I_a}(\Delta T)^2 + \varepsilon_{c,R_c}(\Delta T)^2 + \varepsilon_{\mathrm{vent}}(\Delta T)^2 + \varepsilon_r(\Delta T)^2 + \varepsilon_R(\Delta T)^2} \quad (5)$$

In Equation 5, $\varepsilon^i_{\mathrm{systematic}}(\Delta T)$ indicates a systematic uncertainty contribution, estimated using laboratory experiment, simulation and dual radiosoundings; $\varepsilon_{c,I_a}(\Delta T)$ is the uncertainty due to the estimation of the solar actinic flux (variable magnitude, typically <0.6 K); $\varepsilon_{c,R_c}(\Delta T)$ is the uncertainty due to parameters estimated in the radiation correction (typically <0.2 K) model reported in Equation 1. Formulas to calculate $\varepsilon_{c,I_a}(\Delta T)$ and $\varepsilon_{c,R_c}(\Delta T)$ are fully documented in Dirksen et al. (2014). $\varepsilon_{\mathrm{vent}}$ is the uncertainty due to the ventilation rate (including the effect of the pendulum motion of the radiosonde assumed as in GRUAN to be about 0.2 m s^{-1}); ε_r indicates the comparison uncertainties estimated from the standard deviation of ΔT_r. In RHARM, ε_R is the random uncertainty with a fixed value of 0.15 K chosen in agreement with the GRUAN approach (Dirksen et al., 2014). When the radiation correction of the manufacturer is left unchanged, $\varepsilon(T_{\mathrm{RHARM},RS92})$ is assumed to be the same as the closest temperature profile in time measured under the same meteorological conditions (i.e., clear sky or cloudy, when RH > 95% at least on one level).

Following the application of temperature adjustments, the measured value of the relative humidity, $RH_{\mathrm{RHARM},RS92}$, is adjusted for the solar radiation dry bias, estimated by the effect of the T warm bias on the saturation vapor pressure, using a correction factor:

$$RH_{\mathrm{RHARM},RS92} = f\,RH_{\mathrm{IGRA},RS92}\left(\frac{p_s(T_{\mathrm{RHARM},RS92} + g\Delta T_{\mathrm{RHARM},RS92})}{p_s(T_{\mathrm{RHARM},RS92})}\right) \quad (6)$$

where f is a scalar factor accounting for the temperature dependency of the sensor calibration estimated at night by a comparison with GRUAN measurements (Table 2); p_s is the saturation vapor pressure and g is a factor determined experimentally to weight the applied correction on different radiosonde batches (Dirksen et al., 2014). The factor f may embed a residual contribution from the sensors' time-lag which is typically small for the RH values up to 250 hPa. Known issues in radiosonde humidity data, such as humidity values under dry conditions

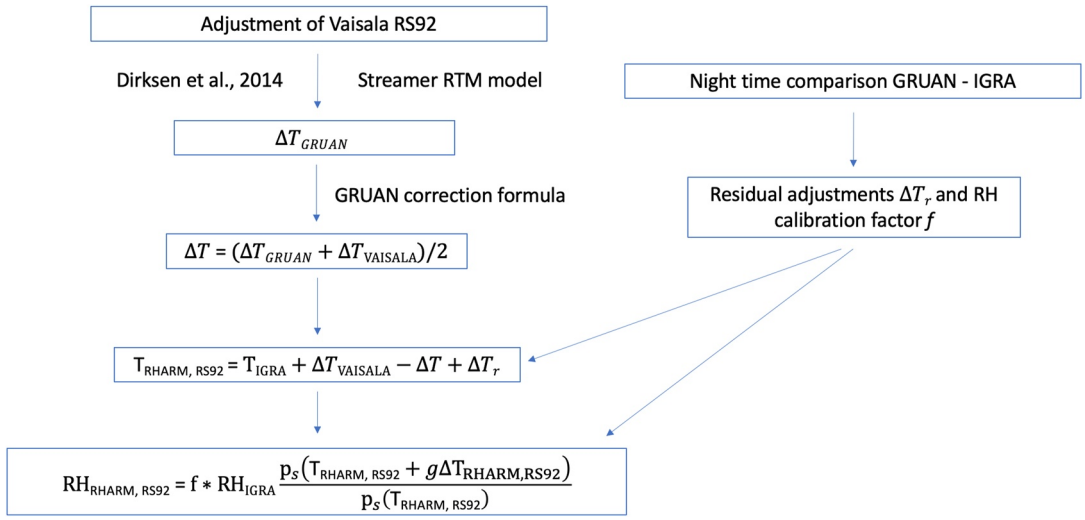

Figure 3. Flow diagram summarizing the post-processing steps of the Radiosounding HARMonization (RHARM) algorithm to adjust temperature and relative humidity profiles measured by the RS92 sondes from 2004. In the diagram, f is a calibration factor, p_s is the saturation vapor pressure, g is a factor determined experimentally to weight the applied correction on different radiosonde batches used over the years. ΔT indicates the adjustments applied to temperature, ΔRH to relative humidity. The subscripts refer to the Global Climate Observing System (GCOS) Reference Upper Air Network (GRUAN) adjustments, Integrated Global Radiosonde Archive (IGRA) adjustments (manufacturer based plus IGRA quality control), RHARM adjustments and to RS92 Vaisala sondes. The subscript "r" refers to a residual correction derived from the nighttime comparison between GRUAN and IGRA data at six GRUAN sites, reported in Table 2.

(RH < 20%) for U.S. stations which were set to a dewpoint depression of 30°C (or RH = 19%), have been managed according to McCarthy et al. (2009)

A flow diagram describing the application of the RHARM adjustments to both T and RH profiles from Vaisala RS92 instruments is shown in Figure 3.

Similarly to Equation 5 for temperature, the combined standard uncertainty for relative humidity is calculated as:

$$\varepsilon(RH_{\mathrm{RHARM},RS92}) = \sqrt{\sum_i \varepsilon^i_{\mathrm{systematic}}(\Delta RH)^2 + \varepsilon_R(\Delta RH)^2} = \\ = \sqrt{\varepsilon_{RC_T}(\Delta RH)^2 + \varepsilon_{RC_g}(\Delta RH)^2 + \varepsilon_f(\Delta RH)^2 + \varepsilon_R(\Delta RH)^2} \quad (7)$$

where $\varepsilon_{RC_T}(\Delta RH)$ is the uncertainty of dry bias correction; $\varepsilon_{RC_g}(\Delta RH)$ is the uncertainty of the radiation sensitivity factor g in Equation 5; ε_f is the uncertainty due to calibration factor f; ε_R is an additional random uncertainty of 2% RH. $\varepsilon(RH_{\mathrm{RHARM},RS92})$ is typically of the order 5–10%RH. In analogy with temperature, when the radiation correction of the manufacturer is left unchanged, $\varepsilon(RH_{\mathrm{RHARM},RS92})$ is assumed to be the same as the closest RH profile in time measured under the same meteorological conditions.

At the end of 2010, Vaisala processing software underwent a major change with the inclusion of humidity time-lag correction and an improved dry bias correction for RH (no specific details on the applied algorithm), but its uptake was heterogeneous across stations. For example, Germany and the UK started using it only in 2015, but this was not the case for other countries, due to choices by National Meteorological Services. In this version of RHARM it is very difficult to take into account such changes at each individual station, given the grossly insufficient metadata available. Nevertheless, this may be possible in future, for any such subsequent changes, using native BUFR reports which include the processing software version in their extra metadata. Storing of these files on a routine basis has been undertaken by ECMWF starting from 2016. Collaboration with Vaisala will also be undertaken to identify when individual stations switched, in order to improve future updates of the RHARM data set.

The GRUAN processing on wind profiles is more basic and does not apply as many corrections to the raw data. The manufacturer software retrieves the magnitudes of u and v from the Doppler shift in the GNSS carrier signal. In the GRUAN processing, these vectors are smoothed and converted into wind speed and direction. The noise in the raw zonal and meridional (u and v) data, due to the radiosonde's pendulum motion and the noise of the GNSS

data, is reduced by using a low-pass digital filter (Dirksen et al., 2014). This smoothing reduces the effective temporal resolution of the wind data to 40 s. Using statistical uncertainties calculated for u and v, the uncertainty of the wind direction φ is given by:

$$\varepsilon(\varphi) = \frac{180}{\pi} \frac{\sqrt{\delta_u^2 + \delta_v^2}}{\left(1 + \left(\frac{u}{v}\right)^2\right)|v|} \tag{8}$$

and the uncertainty of the wind speed w by

$$\varepsilon(w) = \sqrt{\frac{(u\delta_u)^2 + (v\delta_v)^2}{u^2 + v^2}} \tag{9}$$

δ_u and δ_v are the uncertainties of the u and v wind components. Typical values are between 0.4 and 1 ms^{-1} for $\varepsilon(w)$ and about 1° for $\varepsilon(\varphi)$. In the case of negligible wind, when u and v approach 0, the value of $\varepsilon(\varphi)$ becomes very large. For such cases, the absolute value of $\varepsilon(\varphi)$ is limited to 180° (Dirksen et al., 2014). The same limitation is applied to uncertainties estimated with RHARM. The RHARM algorithm converts wind direction and speed reported in IGRA data files into the vectorial components u and v. At time instant t and at a pressure level p, these variables are related as follows:

$$u(p,t) = w(p,t)\ \ sin\left(\frac{\pi}{180}\varphi(p,t)\right) \tag{10}$$

$$v(p,t) = w(p,t)\ \ cos\left(\frac{\pi}{180}\varphi(p,t)\right) \tag{11}$$

The conversion into u and v components avoids issues of interpretation over averages or differences associated with the use of the discontinuous wind direction scale. Nevertheless, to facilitate user applications preferring the use of wind speed and direction, vectors are converted back into wind speed and direction after uncertainty quantification. Equations 8 and 9 are then used also in RHARM to estimate the final uncertainty on w and φ.

To adjust the IGRA wind profiles, the daytime and nighttime differences for u and v between the GRUAN processed and the IGRA radiosounding wind profiles have been calculated using the stations in Table 2. The approach is the same as for temperature (Equation 4), although it is reduced to $\Delta u_{\mathrm{RHARM},RS92} = \Delta u_r$ and to $\Delta v_{\mathrm{RHARM},RS92} = \Delta v_r$, for each of the wind vectorial components. The standard deviation of the $\Delta u_{\mathrm{RHARM},RS92}$ and $\Delta v_{\mathrm{RHARM},RS92}$ are then used as the estimation of the combined standard uncertainties, which are expressed as $\varepsilon(\Delta u_{\mathrm{RHARM},RS92}) = \sqrt{(\varepsilon_r(\Delta u)^2 + \varepsilon_R(\Delta u)^2)}$ and $\varepsilon(\Delta v_{\mathrm{RHARM},RS92}) = \sqrt{(\varepsilon_r(\Delta v)^2 + \varepsilon_R(\Delta v)^2)}$. ε_R is a random uncertainty of 0.15 m s^{-1} for both u and v (https://www.vaisala.com/sites/default/files/documents/RS92SGP-Datasheet-B210358EN-F-LOW.pdf, last access 23 April 2021).

This adjustment can only partly reconcile the difference between GRUAN processing and manufacturer data processing, due to the differences in the low-pass filtering applied to reduce the effect of the radiosonde's pendulum motion. In Appendix B, discrepancies between the RHARM and GRUAN are quantified using the data from the six GRUAN stations reported in Table 2.

3.2. Adjustment of Other Radiosonde Types

For non-Vaisala radiosonde types, the adjustment estimation requires the adoption of a different approach given the lack of GRUAN reference products. To harmonize these time series, RHARM leverages the ID2010 intercomparison, from which estimations of the relative performance of operational radiosondes in 2010 were evaluated. ID2010 permits assessment of the systematic component of the inter-sensor differences, and does not contain strong outliers, but the post-processing applied may come at the cost of under-representing sonde-to-sonde random uncertainty effects (Nash et al., 2011). Furthermore, the use of complex multi-sonde rigs may alter the sonde characteristics compared to standard single-payload flights in important ways vis-a-vis aspects such as ventilation, thermal effects and the magnitude and periodicity of pendulum motion effects.

Among the radiosonde types involved in the intercomparison (Table 3), only those routinely employed at enough stations have been considered for calculating the adjustments for RHARM. The Vaisala RS92-SGP (WMO

radiosonde code = 80) was used as one of the common models during (almost) all flights, allowing us to tie each sonde to the RS92 (at least for the particular location, RS92 model version, the RS92 Vaisala data processing in operation at the time, and the season of the campaign). Therefore, the mean difference ΔT_{NORS92} between parallel profiles of RS92 and each radiosonde type was used for the adjustment (Figure 4). The Vaisala RS92 sondes available in ID2010 have been adjusted using the RHARM algorithm described in the prior sub section. The standard deviation $\sigma_{T_{NORS92}}$ is calculated from the spread of pairwise estimates of ΔT_{NORS92} estimated as $\sigma_{T_{NORS92}} = \sqrt{\sigma_{T_{NORS92}}{}^2 + \epsilon(T_{\mathrm{RHARM},RS92})^2}$ and used as the best estimate of the uncertainty for ΔT_{NORS92} assuming independence of the two components.

Due to the launch setup adopted during the WMO intercomparison, a few radiosonde types were compared less frequently than others on the same payload. Specifically, some models did not have a sufficient sample of coincident Vaisala RS92 sondes associated with them. In these cases, the Graw radiosondes, which flew on rigs both with RS92 sondes and the under-sampled sondes, have been used to make the bridge with the RS92 and to calculate statistics for a larger number of comparisons. $\sigma_{T_{NORS92}}$ have been recalculated accordingly to consider the additional contribution of the Graw radiosonde uncertainties and the two steps required to quantify the comparison.

Although the ID2010 have already been filtered for the presence of outliers, ΔT_{NORS92} and $\sigma_{T_{NORS92}}$ have been calculated using an outlier resistant algorithm where the mean trims away outliers using the median and the median absolute deviation (see https://idlastro.gsfc.nasa.gov/ftp/pro/robust/resistant_mean.pro, last access on 31-21-2020). This ensures that the most typical differences between any two radiosonde types are caught in the calculated differences, enabling their application as an average adjustment for a wide range of radiosondes. With the related considerations, the same approach used for temperature is applied to adjust also wind profiles.

For relative humidity, also in order to be consistent with the RHARM post-processing of RS92 sondes, instead of using the mean difference between pairwise profiles, relative humidity profiles have scaled using the factor f(RH) obtained as the mean ratio of $RH_{\mathrm{RHARM},RS92}$ and RH_{NORS92} (Figure 4). The related standard deviation, $\sigma_{f(RH)_{NORS92}}$, is calculated via error propagation. If the Graw radiosonde was considered as the link with the Vaisala RS92 f(RH) and $\sigma_{f(RH)}$ were rescaled accordingly.

To facilitate the application of the adjustments for all significant pressure levels available in the IGRA data set, the ID2010 profiles have been first smoothed to an effective resolution of 100 m (Iarlori et al., 2015), to reduce the uncertainties due to the limited sample size, and then interpolated at 0.1 hPa resolution. Interpolation has been performed to allow the processing chain to always get an exact match with any of the mandatory and significant levels available in the IGRA files. As for the RS92 case, the interpolation has been performed using a linear function for temperature, while a cubic spline interpolation has been applied to RH and wind component profiles. The interpolation uncertainty has been finally added to the final uncertainty budget (for T, $\sigma = 0.25$ K, for RH, $\sigma = 0.5\%$, for both u and v, $\sigma = 0.05$ ms^{-1}). All the profiles derived from the ID2010 with the corresponding standard deviations are shown in detail in the Supporting Information S1.

Table 4 gives the number and percentage of radiosonde launches adjusted by RHARM since 2004 with the approach generating the Stage I time series: it shows that more than 85% of RHARM adjusted radiosondes are manufactured by Vaisala. This increases the homogeneity of the data set globally, but on the other hand it makes the data set more prone to the impacts of unquantified random and systematic effects unique to Vaisala sondes. The radiosoundings reported in Table 4 include about 40,000 launches from 37 ships (mostly traveling in the Atlantic Ocean).

3.3. Detection and Adjustment of Early Period Breakpoints Using the CUSUM

This section discusses the homogenization module (Steps C, D, and E in Figure 2), applied separately to the daily time series of daytime and nighttime observations at mandatory pressure levels. Differently from previous efforts, data in RHARM are not monthly or annually aggregated.

The output of this module is the homogenized record before the year 2004, denoted above Stage II time series. Nonetheless, the algorithm below uses both Stage I and Stage II data with different aims.

For mandatory levels, the step C of the homogenization module is made by a sequence of four substeps (a–d), then followed by step D. These steps work iteratively along with each time series for a fixed mandatory pressure

level. After the quality check of substep (a), a LOESS transform is applied to the ECV time series in substep (b). Then, substep (c) uses the Cumulative Sum change detection algorithm (CUSUM) to estimate breakpoints. After that, data between two breakpoints are considered. In substep (d) the outliers are removed and in step D, adjustments are estimated and applied. In step E, after merging all the time series and reconstructing the atmospheric profiles, significant levels are adjusted by interpolation along with every single profile. Note that our CUSUM approach builds upon previous successful ECV literature (Peterson & Vose, 1997; Rhoades & Salinger, 1993), and it is used here to define a new integrated algorithm.

3.3.1. Preliminary Quality Check

First, the IGRA data set is filtered by a comprehensive set of quality control procedures to remove gross errors without removing jumps and other discontinuities caused by changes in instrumentation, observing practice (Durre et al., 2008, 2018). RHARM exploits the IGRA flagging system to eliminate questionable data. In addition to the IGRA quality checks, RHARM preliminarily verifies the physical plausibility of the values reported at each pressure level, that is, temperature values 170 K $< T <$ 350 K, relative humidity 0.01% $<$ RH $<$ 100%, wind speed 0 m/s $< w <$ 250 m/s, and wind direction $0° < \phi < 360°$. The missing data are not explicitly considered in the formulas below for the sake of simplicity, but note that data gaps larger than 10 days cause CUSUM to restart.

After filtering out the above unphysical values, we describe each time series for a fixed station and pressure level using the following additive model:

$$x(t) = B(t) + Tr(t) + S(t) + z(t) + \delta(t) \tag{12}$$

where $t = 1, \ldots, T$ is the time index in days. In the sequel, we use a backward time approach. Hence, we let $t = 1$ represent the most recent observation and $t = T$ the oldest one. In between, we have $t = T_1$ denoting the oldest observation in Stage I data, and $t = T_1 + 1$ denoting the most recent observation in Stage II data.

In the left-hand side of Equation 12, $x(t)$ is the observed time series of a specific ECV, with mean μ_x and standard deviation σ_x. On the right-hand side, B (t) is the instrumental bias component modeled by a step function characterizing the behavior of the different sensors used in different periods. The long-term climate trend Tr (t) is a slowly varying function. The seasonal component $S(t)$ is a quasi-periodic term with a yearly period. The stochastic component $z(t)$ represents the local meteorological variability, namely a zero-mean colored noise with a spectrum dominated by high frequencies. Eventually, $\varepsilon(t)$ is the zero mean measurement error with the standard deviation σ_ε representing the measurement uncertainty.

The subsequent harmonization steps are applied to Stage II data only if the corresponding Stage I time series is informative enough to learn the trend $Tr(t)$ and the bias $B(t)$. Namely, only if the Stage I time series meets the following two conditions:

1. Minimum length of 5 years;
2. At least 60 measurements per year.

If these two requirements are not met, the Stage II time series is not harmonized and provided as-is in IGRA after passing through the quality check mentioned above.

3.3.2. LOESS Smoothing

To filter the quasi-periodic and high-frequency component given by $S(t) + z(t) + \delta(t)$, the nonlinear trend component $Tr(t) + B(t)$ is preliminarily estimated by applying a locally weighted smoothing (LOESS) with a smoothing window equal to 30% of the length of the overall time series. Due to its nature, LOESS enables an efficient propagation into the smoothed times series of the bias present in the non-smoothed time series, removing the seasonality due to the large smoothing window used. Nonetheless, it is only a preliminary estimate of the instrument bias $B(t)$ as it tends to smooth the steps of $B(t)$ itself.

3.3.3. CUSUM Breakpoint Detection

Denoting the LOESS output by $x_L(t)$, breakpoints detection is based on the following CUSUM statistics:

$$S(t) = \max(0, S(t-1) + x_L(t) - \mu_{x_L} - k) \tag{13}$$

$$S'(t) = \max(0, S'(t-1) + \mu_{x_L} - x_L(t) - k). \tag{14}$$

where S and S' are equal to zero at the time $t = 0$, S is used to signal increasing changes, and S' is used to signal decreasing ones. When either S(t) or S'(t) exceeds the threshold value h, a break is detected at time t.

In general, by appropriate selection of the "allowance" k and the threshold h, it is possible to design CUSUM to be highly effective at detecting shifts of all sizes, even for highly skewed and extremely heavy-tailed process distributions (Stoumbos & Reynolds, 2004). However, the use of smoothed time series may affect the exact identification of the break occurrence (before or after the "real" occurrence). RHARM has been tuned to find a balance among the appropriate allowance value of the CUSUM, the LOESS smoothing window, and the timing ambiguity in the identification of breaks in the time series.

Consolidated literature (Woodall & Adams, 1993) has been integrated by manual investigation of selected stations with comprehensive metadata from 1978 to present (e.g., Lindenberg WMO index = 10,393 and Sodankyla WMO index = 2,836) to tune CUSUM parameters in RHARM. Moreover, unreported synthetic time series with artificial systematic effects $B(t)$ have been considered.

As a result, the allowance $k = 0.1\sigma_{x_L}$ and the threshold value $h = 0.4\sigma_{xL}$ are used. Note that both the mean μ_{x_L} and standard deviation σ_{xL} are estimated using the entire time series, before and after 2004, to avoid variance inflation and level bias due to instrument changes.

As mentioned above, CUSUM has applied backwards in time and the breakpoints denoted by $t_1^* < \cdots < t_k^*$ are detected iteratively from the most recent one of Stage II data.

Among the breakpoints $t_{i-1}^* < t_i^*$, we have and i-th adjustment period. If an adjustment period is shorter than one year, the corresponding breakpoint is skipped, and the period is merged with the previous one. As a result, we have that the i-th adjustment period defined by $t_i^* - t_{i-1}^* > 365$ and t_0^* located at the beginning of Stage II time series: $t_0^* = T_1 + 1$.

3.3.4. Outlier Removal

Once an adjustment period is identified, outlier detection and removal are performed between the two corresponding breaks. For temperature, data with a mean deviation larger than six times the standard deviation are rejected. For RH, values with a median deviation exceeding three times the interquartile range are rejected; for wind data, no outliers' removal is applied because, due to their highly non-Gaussian distribution, all tested criteria result in filtering too many plausible values. All the preliminary and post-processing data checks lead to the removal of an additional 0.2% of data points from the original IGRA collection.

3.3.5. Adjustment

To eliminate the instrument bias in the i-th adjustment period, two log-linear trends are considered. One, say $f_0(t) = \exp(a_0 + b_0 t)$, assumes no changes, while the other one, say $f_{B,i}(t) = \exp(a_i + b_i t)$, is sensitive to instrument bias of the i-th period.

In particular, for the first trend, a_0 and b_0 are estimated using the data before the i-th break, namely $x(1), \ldots, x(t_{i-1}^*)$. For the second trend, a_i and b_i are estimated using the data in the i-th period, namely $x(t_{i-1}^* + 1), \ldots, x(t_i^*)$. In both cases, a robust least absolute deviation method, known as LADFIT (Barrodale & Roberts, 1974), is applied to the log-transformed ECVs. For other climate variables, such as sea surface temperature, it is common practice to represent anomaly decay over time using exponential functions (Bulgin et al., 2020).

Recalling Equation 12 and using the first model, the average trend in the i-th period, say $\overline{Tr_i}$, free of instrumental bias, is obtained by averaging $f_0(t)$ in the i-th period, say $\overline{f_{0,i}}$. Using the second model, the average trend plus the bias, $\overline{Tr_i} + B(t)$ is obtained by A averaging $f_{B,i}(t)$ in the i-th period, say $\overline{f_{B,i}}$.

The i-th adjustment is then given by:

$$\Delta_i = \overline{f_{0,i}} - \overline{f_{B,i}} \tag{15}$$

and the homogenized data $x_H(t)$ in the i-th period is given by:

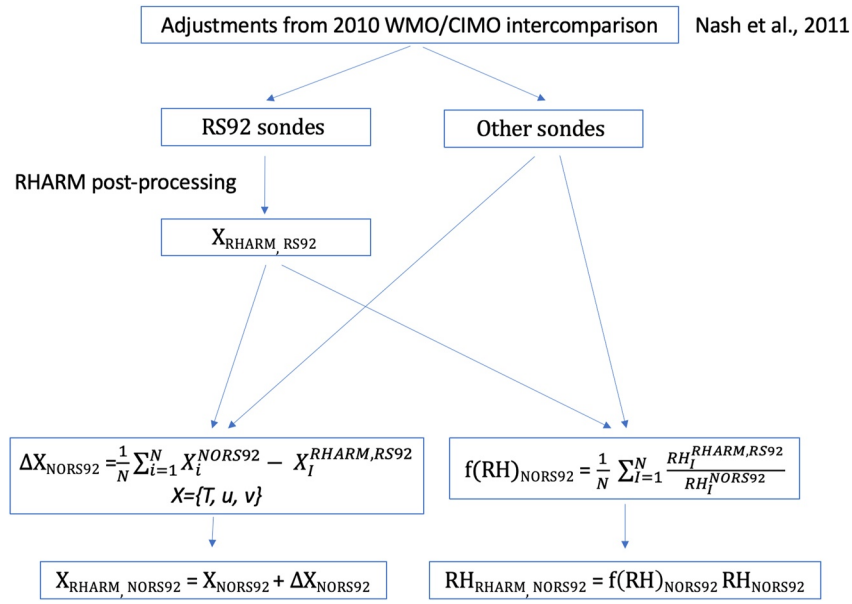

Figure 4. Flow diagram summarizing the post-processing steps of the Radiosounding HARMonization (RHARM) algorithm to adjust the temperature and relative humidity profiles measured for all radiosonde types other than RS92 reported in Table 3 in the period from 2004 onward. In the diagram, "X" stands for T, *u* or *v*. The subscript RHARM refers to the output adjusted variable and the subscripts RS92/NORS92 refer to the input radiosonde type: RS92 Vaisala or other.

$$x_H(t) = x(t) + \Delta_i \tag{16}$$

Note that, if the two intercepts a_0 and a_i have opposite signs, the former is replaced by the first, while the slope is not adjusted. This condition is mainly related to the presence of instrumental calibration drifts in a small number of time series.

Figure 5 shows an example of the breakpoints and adjustments for the nighttime relative humidity (over liquid water) measured at 300 hPa in Sodankyla from 1978 to the present: RHARM approach can identify the main documented breakpoints shown. Note that the most recent breakpoint is in the year 2005. As already mentioned, the separation date between Stage I and II data set in the text in 2004 is station dependent must be interpreted as an average.

3.3.6. Harmonization at Significant Pressure Levels

The significant pressure levels are known to show extreme variability over time and are provided for the reasonably accurate reproduction of the radiosonde profiles. By construction, the pressure of significant levels and their vertical randomness are different for different radiosonde launches. This hampers the use of time series techniques. Indeed, RHARM uses a profile-based approach to compute adjustments at each significant pressure level. After reconstructing the atmospheric profiles at mandatory levels, the RHARM adjustment Δ of Equation 16 at a significant level is calculated by interpolating Δs at the two closest mandatory levels, p_A and p_B (step E in Figure 2). The interpolation is performed using a linear function for temperature, while a cubic spline interpolation has been applied to RH and wind component profiles. The resulting interpolation uncertainty has been evaluated using the comparison of the effect of the interpolation at GRUAN stations where high-resolution profiles are available. As discussed in the next section, this interpolation uncertainty has been added to the final uncertainty budget (0.25 K for temperature, 0.5%RH for relative humidity, 0.05 ms^{-1} for both the wind components).

The above overall approach has exceptions at the approximately 30 stations where the Stage I data has more than one type of sonde post-processed by RHARM (see Table 4). These are handled with an ad-hoc application of the RHARM algorithm. Due to the above checks and constraints applied in the RHARM algorithm, the total amount of RHARM data represents 92% of those available in IGRA from surface to 100 hPa and 94%–95% at altitudes above.

Table 3
List of the Operational Radiosondes Involved in the 2010 WMO/CIMO Radiosonde Intercomparison Used to Calculate the RHARM Adjustments

Abbrev.	Name	WMO radiosonde code
RS92	VAISALA RS92 SGP	80
Graw	DMF-09 Graw	17
Modem	M10, Modem	57
LM	LMS6	11 (01/01/2008), 82 (07/11/2012)
Meisei	Meisei	30 (01/01/2010)
JinYang	JinYang	21
IntermSA	iMet-2 InterMet	97, 98, 99
Daqiao	Nanjing GTS1-2/GFE(L)	33 (03/11/2011)
Huayun	Taiyuan GTS1-1/GFE(L)	31 (03/11/2011)
Changf	Beijing Changfeng CF-06	45 (07/05/2014)
ML	Meteolabor	26

Note. Dates in brackets refer to the date of assignment for the WMO radiosonde code. Note that also RS92 is included in the list. Adjustments have been calculated using the RS92-SGP sondes as the comparator, in order to be physically consistent with the GRUAN product. For consistency, RS92-SGP sondes launched during the intercomparison have been reprocessed using the RHARM approach. CIMO, Commission for Instruments and Methods of Observation; GRUAN, Global Climate Observing System (GCOS) Reference Upper Air Network; RHARM, Radiosounding HARMonization; WMO, World Meteorological Organization.

Table 4
Number and Percentage of the Radiosonde Launches Available Since 2004 Adjusted Using the RHARM Approach

Radiosonde type	Launches	Percentage
LMS6	29,148	1.3
DMF-09 Graw	16,736	0.8
VIZ/JinYang	33,721	1.5
Taiyuan GTS1-1/GFE(L)	13,409	0.6
Nanjing GTS1-2/GFE(L)	17,406	0.8
Meteolabor	436	0.0
Meisei	16,179	0.7
Beijing Changfeng CF-06	36,393	1.7
M10, Modem	121,446	5.5
Vaisala RS92/RS41	1,893,805	85.9
Intermet	26,505	1.2
Total	2,205,183	100

Note. The total number of soundings available within IGRA from 2004 for the stations adjusted using RHARM is 4,785,543. These include 55,325 balloon launches with a Vaisala RS41 sonde, currently not adjusted within RHARM.

3.4. Estimation of Uncertainties

As for the Stage I time series, an uncertainty is attributed to each value of the Stage II time series using the following formula:

$$\varepsilon_H(p,t) = \sqrt{(\varepsilon(X_{\text{Stage I}}))^2 + (\varepsilon(X_{\text{Stage II}}))^2} \tag{17}$$

In Equation 17 (under the square root, dependencies for $(\varepsilon(X_{\text{Stage II}}))^2$ on p and t are omitted), $\varepsilon_H(p,t)$ is the total uncertainty for the homogenized IGRA time series calculated at the pressure level p and the time instant t, $\varepsilon(X_{\text{Stage I}})$ is the average uncertainty of the Stage I time series at the selected station, and $\varepsilon(X_{\text{Stage II}})$ is estimated using the residuals of each time series with respect to a predictor model (i.e., the smoothed time series obtained by applying a LOESS smoother):

$$\varepsilon(X_{\text{Stage II}}) = x_i - q_i \quad t = 1,\ 2, \ldots\ldots, T \tag{18}$$

where x_i is the measurement for the variable x at the instant i, q_i is the LOESS modeled value and T is time length of the time series.

In order to tune the statistical model and obtain a reliable estimation of the uncertainty, the LOESS smoothing parameter is optimized, for each individual station, to match the residuals to the average values of $\varepsilon(X_{\text{RHARM},RS92/NORS92})$, in the time period when this is available (approximately after 2004, depending on the station). The obtained smoothing parameter is then assumed to be optimal for the entire time series and the final value of the uncertainty is obtained by averaging the residuals on a monthly time scale. The uncertainty is not estimated for months with fewer than 15 radiosonde launches. The Stage I time series portions are built upon the most recent radiosounding instruments which should logically be better performing or, at least, better characterized through the outcome of the intercomparison experiments and it is therefore assumed to be a good constraint to the estimation of the uncertainties in the historical measurements which should be larger.

At each significant pressure level (p'), also the uncertainty is interpolated between the two closest mandatory levels, p_A and p_B. Similarly to what is reported in Section 3.2, an interpolation uncertainty term A is also added to the interpolated uncertainty values ($\varepsilon_I(p_A, p_B, t)$).

For these levels indicated with p′, Equation 17 becomes:

$$\varepsilon_H(p',t) = \sqrt{(\varepsilon_I(p_A,p_B,t))^2 + (\varepsilon_{int}(p',t))^2} \tag{19}$$

After interpolation of adjustments and uncertainties at the significant levels, nighttime and daytime time series are merged to provide the final homogenized time series (step F in Figure 2).

For wind, the formula to obtain the harmonized time series of wind speed (w) and direction (φ) (i.e., intensity and direction of the wind vector) once the u and v component have been homogenized. The following formulas are applied:

$$W = \sqrt{u^2 + v^2} \tag{20}$$

$$\varphi = a\tan2(-u;-v)\frac{180}{\pi} = 180 + a\tan2(v;u)\frac{180}{\pi} \tag{21}$$

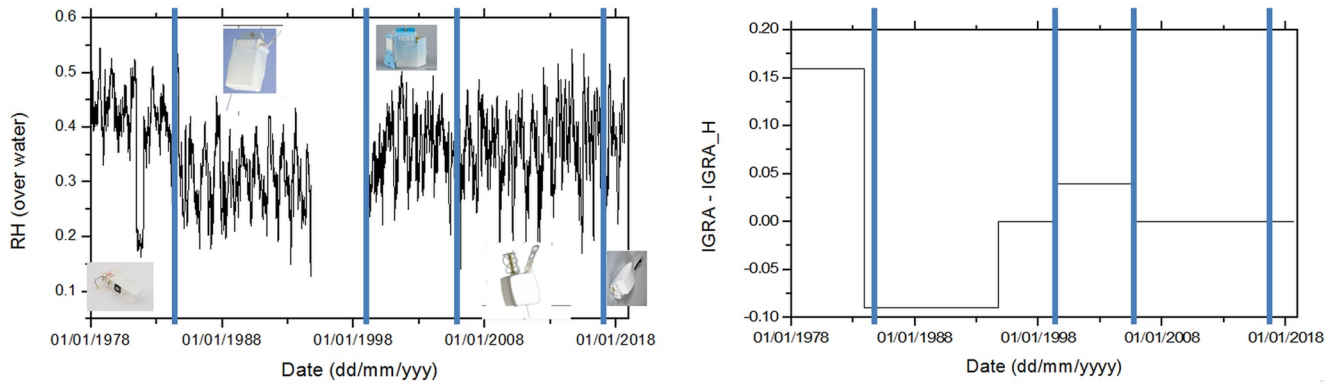

Figure 5. Left panel, relative humidity monthly time series for the station of Sodankyla (World Meteorological Organization [WMO] index = 2,836, 67.3667 N 26.6289 E, 179 m asl) as available from Integrated Global Radiosonde Archive (IGRA), reporting the radiosonde pictures used in different periods delimited by blue lines. Right panel, adjustments applied by the Radiosounding HARMonization (RHARM) algorithm (IGRA minus RHARM) with blue lines indicating sensor changes as in the left panel.

The second equation also enables the conversion of the wind vector to the meteorological convention of the direction the wind is coming from.

The estimated u and v uncertainties are then propagated to obtain the w and φ uncertainties using the following formulas (based on the trigonometric definition of the partial derivatives of the function atan2):

$$\sigma_W = 2\sqrt{\frac{u^2}{u^2+v^2}\sigma_u{}^2 + \frac{v^2}{u^2+v^2}\sigma_v{}^2 + 2\frac{uv}{(u^2+v^2)^2}\sigma_{uv}} \tag{22}$$

$$\sigma_\varphi = \frac{180}{\pi}\left(\sqrt{\left(-\frac{uv}{u^2+v^2}\right)^2\sigma_u{}^2 + \left(\frac{vu}{u^2+v^2}\right)^2\sigma_v{}^2 - 2\left(\frac{uv}{u^2+v^2}\right)^2\sigma_{uv}}\right) \tag{23}$$

In Appendix A, Figure A1 shows an example of a wind time series (for the both the u and v components) reporting also the uncertainties calculated according to the approach discussed in this section.

The RHARM data set is calculated assuming that adjustments of systematic effects do not affect the total uncertainty budget and, therefore, when false positives are detected, the uncertainty might be underestimated. The autocorrelation between the data, at night and day separately, of each time series has been estimated and found to be generally small at all pressure levels (<0.35). Therefore, autocorrelation has not been included in the estimation of trends.

The use of smoothed time series may affect the precise identification of the break timing. RHARM has been tuned to find a balance among the appropriate allowance value of the CUSUM, the LOESS smoothing window, and the timing ambiguity in the identification of breaks in the time series. Section 4.3 provides an assessment of the discrepancy between the breakpoints detected in the RHARM time series and the incomplete metadata available since 2000.

Finally, the RHARM algorithm cannot distinguish two consecutive systematic effects generating a monotonic increase of the CUSUM functions: these situations are adjusted as one single period affected by the mean of the real systematic errors. RHARM is currently run independently for each pressure level: correlations in breakpoint detection at different levels are discussed in Section 4.3.

4. Results

4.1. Overall Adjustments

The statistical properties of the adjusted records, that is, merged Stage I and Stage II time series (Figure 6), show that RHARM is warmer than the IGRA in the NH, by 0.6 K on average (difference of median values), while in the tropics RHARM is slightly cooler than IGRA by 0.1 K. For the most recent observations (since 2004), the

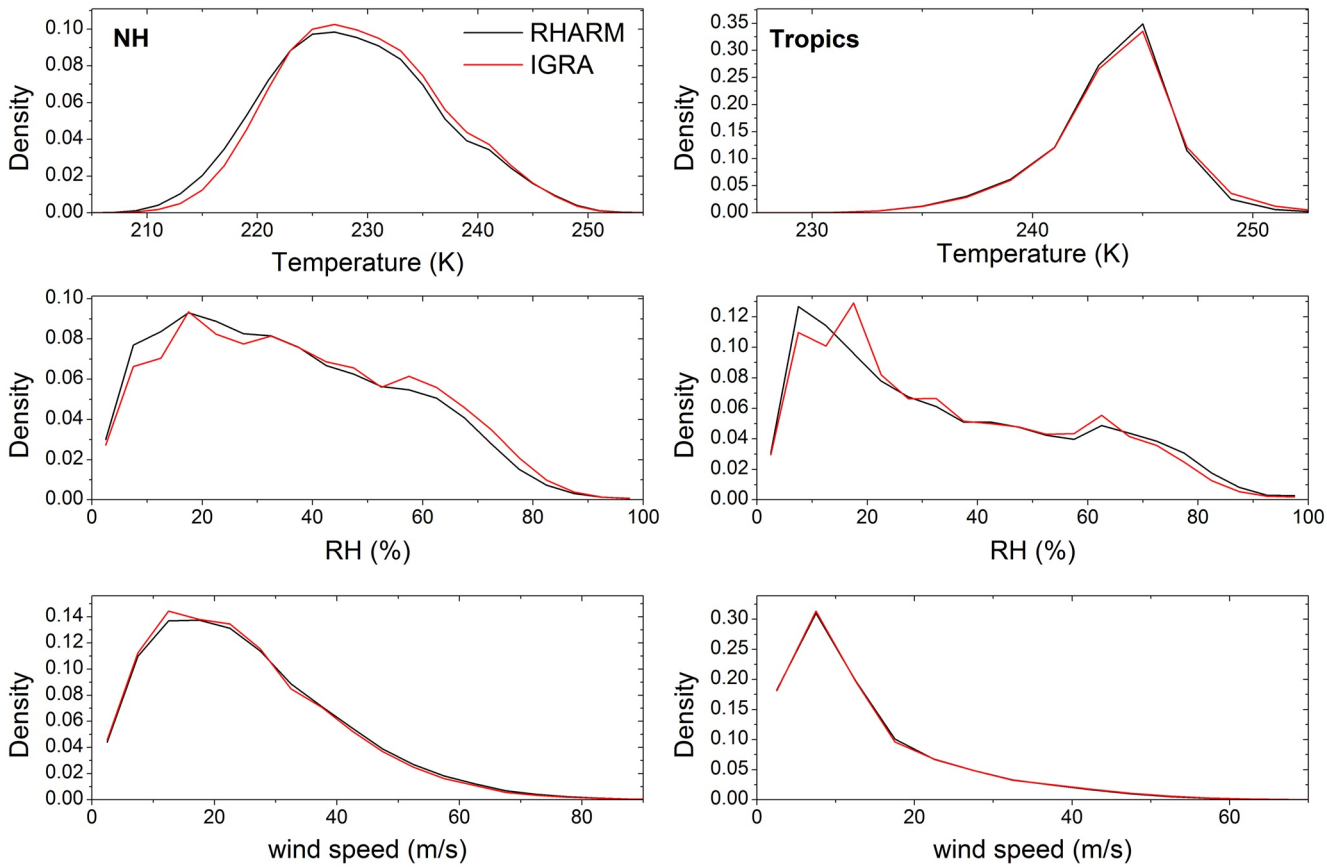

Figure 6. Probability density functions (pdfs) calculated in the northern hemisphere (NH) and in the tropics (±25° latitude) at 300 hPa for the Integrated Global Radiosonde Archive and Radiosounding HARMonization (RHARM) datasets of temperature (top panels), RH (middle panels), wind speed (bottom panel) for matching ascents at all the 697 stations shown in Figure 1. RHARM data refers to the merged time series (Stage I + Stage II time series). The median, the first and third quartiles of the pdfs are reported in Tables 5 and 6 for convenience.

magnitude of the RHARM adjustments for temperature is typically small, which is expected due to the enhanced quality of recent radiosonde data compared to historical observations (Thorne et al., 2011). This result is also consistent with existing comparisons (e.g., Dirksen et al., 2014).

For RH, RHARM is drier by 2.1% than IGRA in the NH, while in the tropics the profiles are moister by 0.3%. The adjustments are most noticeable for RH values below 20%–30% and above 52%, both in the NH and the tropics. For wind speed, as anticipated, the systematic effects have a smaller magnitude than for temperature and RH. Tables 5 and 6 further summarize the main characteristics of adjustments. The first and third quartiles for RHARM temperatures are 0.9 and 0.3 K higher than IGRA, respectively, revealing the predominance of cold biases in the IGRA data since 1978; for RH, the first and third quartiles of the RHARM probability density function (pdf) are 1.9% RH and 2.5% RH smaller than IGRA, respectively, corresponding to the predominance of a moist bias in IGRA.

Table 5
First, Second (Median), and Third Quartiles of the Northern Hemisphere Pdfs Shown in Figure 7

NH	1st quartile (Q1)	Median	3rd quartile (Q3)
T IGRA (K)	223.2	228.4	234.1
T RHARM (K)	224.1	229.0	234.4
RH IGRA (%)	19.6	35.1	53.5
RH RHARM (%)	18.0	33.0	51.0
w IGRA (m s^{-1})	13.4	22.0	33.3
w RHARM (m s^{-1})	13.6	22.6	34.1

Note. IGRA, Integrated Global Radiosonde Archive; RHARM, Radiosounding HARMonization.

In terms of the results at 850 hPa (Figure S5 in Supporting Information S1) corresponding roughly to the top of the planetary boundary layer except in regions of high topography, the comparison of temperature trends (per decade, abbreviated as "da") shows enhanced homogeneity for RHARM, especially in Europe and South America. The general tendency is for a warming in NH and tropics and for moderate cooling in SH. For relative humidity, the variability of the trends is larger than for temperature: the adjustments applied by RHARM reduces heterogeneity, in particular in Europe and the tropics. The

Table 6
First, Second (Median), and Third Quartiles of the Tropics Pdfs Shown in Figure 7

Tropics	1st quartile (Q1)	Median	3rd quartile (Q3)
T IGRA (K)	240.2	242.1	243.4
T RHARM (K)	240.2	242.0	243.2
RH IGRA (%)	15.5	28.9	50.7
RH RHARM (%)	13.9	29.2	52.4
w IGRA (m s^{-1})	6.0	10.2	17.6
w RHARM (m s^{-1})	6.0	10.2	17.6

Note. IGRA, Integrated Global Radiosonde Archive; RHARM, Radiosounding HARMonization.

overall tendencies show a moderate positive trend in the NH which becomes stronger in the SH. For wind speed, RHARM generally improves the estimation for several isolated and obviously spurious large station trends.

At 300 hPa (Figure 7), improvements in the homogeneity of temperature trends are mainly visible in parts of the NH and South America. For RH, improvements are observed mainly in Europe and in the tropics. For both temperature and RH, overall trends in the NH and SH observed at 300 hPa agree with trends at 850 hPa. In the most recent decades, a positive trend in RH was also identified by Madonna, Tramutola, et al. (2020) in Europe, in the SH and the tropics. Wind speed shows improvements especially in the NH.

Finally at 100 hPa (Figure S6 in Supporting Information S1), temperature trends are more homogeneous globally (in particular in the NH) and cooler than IGRA data. For wind speed, RHARM adjustments bring the major improvement in the NH with an overall negative trend in the NH and Australia and from neutral to moderately positive trend in the tropics. RH trends are not reported because these are not available from RHARM (limited at 250 hPa).

To support the interpretation of global trend maps, especially where improvements in spatial homogeneity are small, in Table 7 standard deviations of the distribution of trend values for the RHARM stations in NH, tropics, SH, and globally are reported. The aim is to provide a dispersion indicator of the trend values around the related zonal mean trend and, therefore, of their homogeneity. Due to the small number of stations, polar stations are included in the NH and SH, respectively. Cells in Table 7 are colored according to the outcome of an *F* test applied to assess the equality of variance estimates (homoscedasticity) of IGRA and RHARM for each latitude belt. A Levene's test has been also applied to confirmed results from the *F* test and because the assumption of independent datasets needed for the *F* test, is not respected for IGRA and RHARM. It is important to note that results from the *F* test are valid for the trends homogeneity within each latitude belt or globally, which does not exclude that homogeneity can be greater in specific regions, as discussed above.

For all the variables, the values of the global standard deviation of trends show improvements in the spatial homogeneity with RHARM data for all the variables, although they are larger for RH and wind speed. For RH, there is an overall enhanced homogeneity with a significant difference between the IGRA and RHARM variances at all latitudes and pressures. For wind, the variance difference is significant at all pressures in the NH and globally. For temperature, the difference in the variances is significant at 300 hPa in the tropics and at 100 hPa in the NH. In the SH, the homogeneity of RHARM is lower than IGRA: this is mainly due to the difference between the trends of the Australian stations and all the others in the SH. This difference will be further investigated in future releases of RHARM. In the Figures S7–S12 in Supporting Information S1, examples of anomalies at three pressure levels (850, 500, 300 hPa) for a few stations, selected among those with the longer data records in RHARM, are shown.

To illustrate the temporal evolution of the datasets considered, we further discuss tropospheric interannual variations and trends of temperature, relative humidity and wind for the period 1979–2018 obtained by aggregating data at different latitudes. For temperature (Figure 8) in the NH, IGRA, RHARM, and ERA5 show a similar positive trend of 0.38, 0.39 and 0.43 K da^{-1}, respectively, while in the tropics at 300 hPa the trend is of 0.17, 0.25, 0.20 K da^{-1}, with a more pronounced trend increase starting around 1997. Similar results have been obtained considering European stations only (Madonna, 2020). In the NH, the comparison of the anomalies at 300 hPa shows the evident adjustment applied by RHARM on the IGRA data over 1996–2005 (corresponding to the period of RS80 and RS90 Vaisala radiosondes) which reduces the difference from ERA5 results. Differences between RHARM and ERA5 are generally smaller than 0.5 K in absolute value. In late 1990, RHARM successfully adjusts a clear shift affecting the IGRA data in the tropics: this is evident from the comparison with ERA5 anomalies revealing a smaller difference with RHARM than with IGRA in the same period (Figure 8d). The shift is likely related to the mass adoption of RS80 sondes. An issue for a few tropical stations in late 90s was also discussed by Angell (2003). At 500 hPa (Figure S12 in Supporting Information S1) the situation is very similar although in the NH the differences are smaller, while in the tropics results are in line with those for the 300 hPa pressure level.

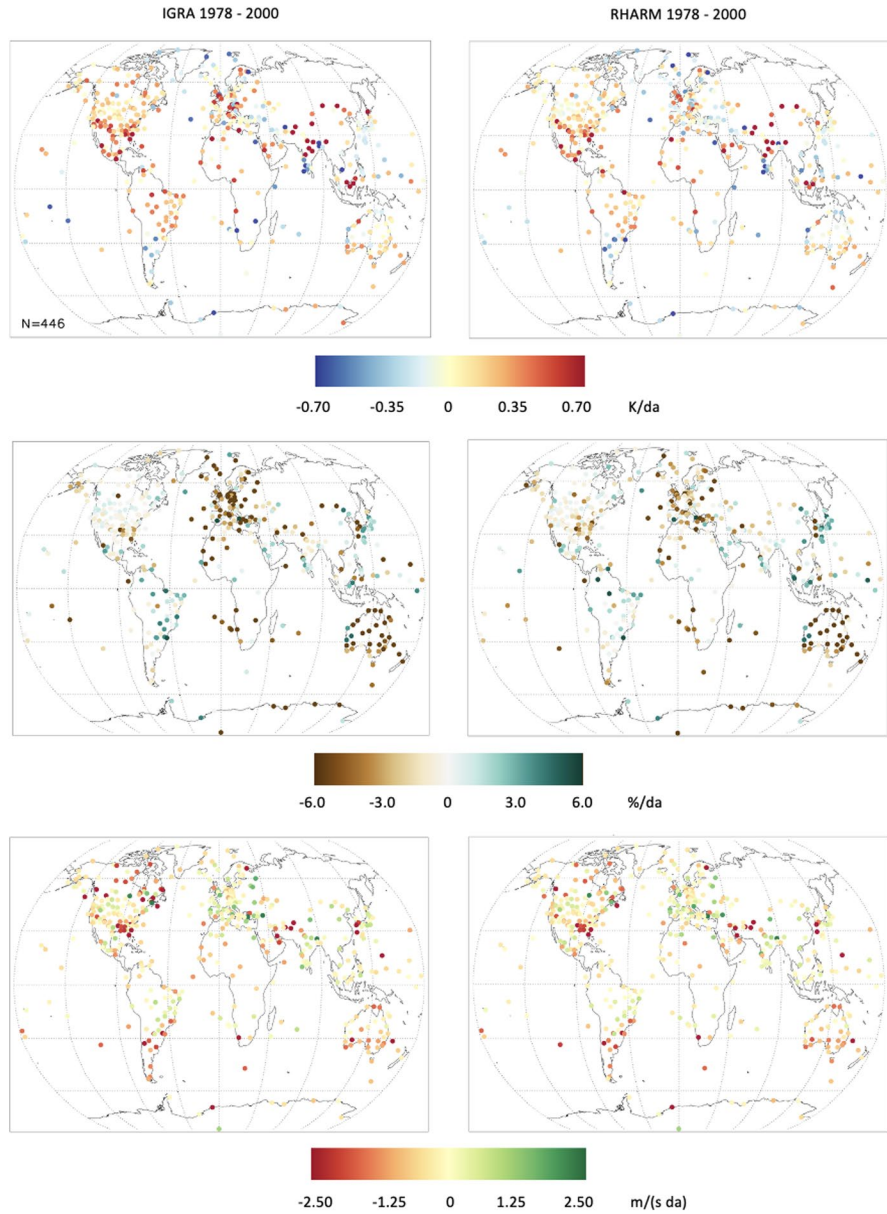

Figure 7. Global maps of the trends (per decade) at 300 hPa of temperature (top), relative humidity (middle) and wind speed (bottom) estimated from Integrated Global Radiosonde Archive (left panels) and Radiosounding HARMonization (right panels) stations. Trends have been estimated for each station in the period 1978–2000.

For relative humidity (Figure 9), in the NH the substantive adjustment applied to IGRA by RHARM at 300 hPa before 1986 (up to 10%RH) largely improves the agreement with ERA5. In 1986, a few major changes occurred in the global radiosounding data, the most relevant of which are: changes in several radiosonde models, such for MARS/MRZ and VIZ radiosondes; the adoption of new manufacturers at some stations, mainly changes from another manufacturer to Vaisala, and changes in the dewpoint depression algorithm, for example, at UK stations; and, maybe the most important, the introduction of "pre-baselined" radiosondes, that is, removal of the practice of applying a manual baseline lock for all temperature and RH profiles which was discovered to be prone to producing a wet bias in all the RH values smaller than 60% (more details at https://library.wmo.int/doc_num.php?explnum_id=9592). Since 2004 the adjustment applied by RHARM is smaller and further improves the agreement with ERA5 and shows a negative trend of −0.8% RH da^{-1}. In the last decade, the trends show a change with a slight increase which has been already quantified in the European domain (Madonna, 2020). At 500 hPa (Figure S13 in Supporting Information S1) the situation is very similar although the adjustments are much smaller.

Table 7
Standard Deviation of the Distributions of Trend Values for Different Latitude Belts and Globally

	IGRA T	RHARM T	IGRA RH	RHARM RH	IGRA W	RHARM W
100 hPa						
NH	**0.56**	**0.48**	N/A	N/A	**0.91**	**0.77**
TRO	0.77	0.73	N/A	N/A	0.89	0.86
SH	0.64	0.70	N/A	N/A	1.02	0.95
Global	0.51	0.48	N/A	N/A	**0.99**	**0.90**
300 hPa						
NH	0.39	0.37	**4.17**	**2.98**	**1.25**	**1.07**
TRO	**0.57**	**0.52**	**3.99**	**3.58**	0.78	0.73
SH	0.32	0.37	**4.49**	**4.15**	1.02	0.94
Global	0.40	0.37	**4.94**	**3.77**	**1.01**	**0.97**
850 hPa						
NH	0.42	0.41	**2.15**	**1.74**	**0.40**	**0.37**
TRO	0.23	0.21	**2.48**	**2.53**	0.36	0.32
SH	0.25	0.30	**3.00**	**2.68**	0.33	0.29
Global	0.32	0.29	**2.79**	**2.24**	**0.46**	**0.42**

Note. Pairs of cells with significant differences in their variances are in bold. At 100 hPa RH values from Radiosounding HARMonization are not available.

In the tropics, the adjustments applied by RHARM at 300 and 500 hPa are smaller than in the NH. The comparison with ERA5 shows that the largest differences are at 500 hPa (up to 4–5%RH). The comparison highlights major differences in three periods: before 1990, where ERA5 negative anomalies are smaller; after 2005, with RHARM anomalies larger than those of ERA5; and after 2015, when differences increase especially at 500 hPa.

The strong positive humidity anomalies observed in the tropics for the period 2015–2019 appear to be correlated with significant positive anomalies of the bi-monthly multivariate El Niño/Southern Oscillation (ENSO) index (Hu & Fedorov, 2017, available at https://www.esrl.noaa.gov/psd/enso/mei) which started in January 2015 and reaches within the same year values larger than 2.0. Boosted by the major El Niño event, 2015 was the first of five consecutive years among the six warmest years in the 140-year observational record (e.g., https://www.ncdc.noaa.gov/sotc/global), which may be related to the observed strong positive anomalies of relative humidity in the tropics and in the SH. A possible positive trend in upper-tropospheric absolute humidity has been noted in previous work (e.g., Dessler & Davis, 2010).

For wind speed, the comparison between observational data and ERA5 (Figures S12 and S13 in Supporting Information S1) shows a in the NH at 300 hPa positive trends of 0.34, 0.33, and 0.07 m s^{-1} da^{-1} for IGRA, RHARM, and ERA5, respectively. In the tropics, trends at 300 hPa are negative, but with a narrower difference than in the NH, with values of −0.06, −0.10, 0.06 m s^{-1} da^{-1}. Similar considerations can be made for the anomalies at 500 hPa. In the NH, there is a good agreement also in the long-term variability and peak values. In the tropics, IGRA and RHARM exhibit larger variability than ERA5 although with significant differences for the extreme values.

In the Southern Hemisphere (SH), where 66 stations are available in RHARM, the comparison (not shown) results are similar to the tropics, with the same strong positive humidity anomalies after 2015.

Despite a degree of non-independence, the comparison with ERA5 reveals discrepancies in the monthly anomalies and trends with both IGRA and RHARM, although the adjustments applied in the RHARM data serves to somewhat reduce the differences between ERA5 and the observations, especially for temperature and RH in the NH. ERA5 performances in reproducing the observed atmospheric variability appear to be higher in the NH than in the tropics, likely due to the stronger observational constraints.

4.2. Comparison of Trend Profiles

In this section, the comparison of the trend profiles for IGRA, RHARM, and ERA5 is discussed for the NH and the tropics, at all mandatory levels from 850 hPa to 10 hPa for the three ECVs (temperature, RH, and wind speed).

4.2.1. Temperature

In the NH, the comparison for temperature shows a very good agreement overall at pressure levels up to 200 hPa (Figure 10, left), with relative differences within 0.1 K da^{-1}. At pressure above 200 hPa, all the datasets agree within 0.1 K da^{-1}. In the tropics (25°S–25°N), the shape of the trend vertical profiles (Figure 10, right panel) is similar for all datasets, with IGRA the coldest. Up to 300 hPa, trends are positive (tropospheric warming) and their difference does not exceed 0.2 K da^{-1}. In the range 300–70 hPa cooling trends for RHARM and ERA5 are very close, within 0.1 K da^{-1}, while at lower pressures ERA5 is warmer than both IGRA and RHARM.

4.2.2. Relative Humidity

In Figure 11, the comparison of the vertical profiles of RH trends, between 850 hPa and 300 hPa in the NH (left panel) shows that RHARM and ERA5 have a similar shape with relative differences around 1% RH/da throughout the entire vertical range, although RHARM is the only with positive values (near zero) at pressures higher

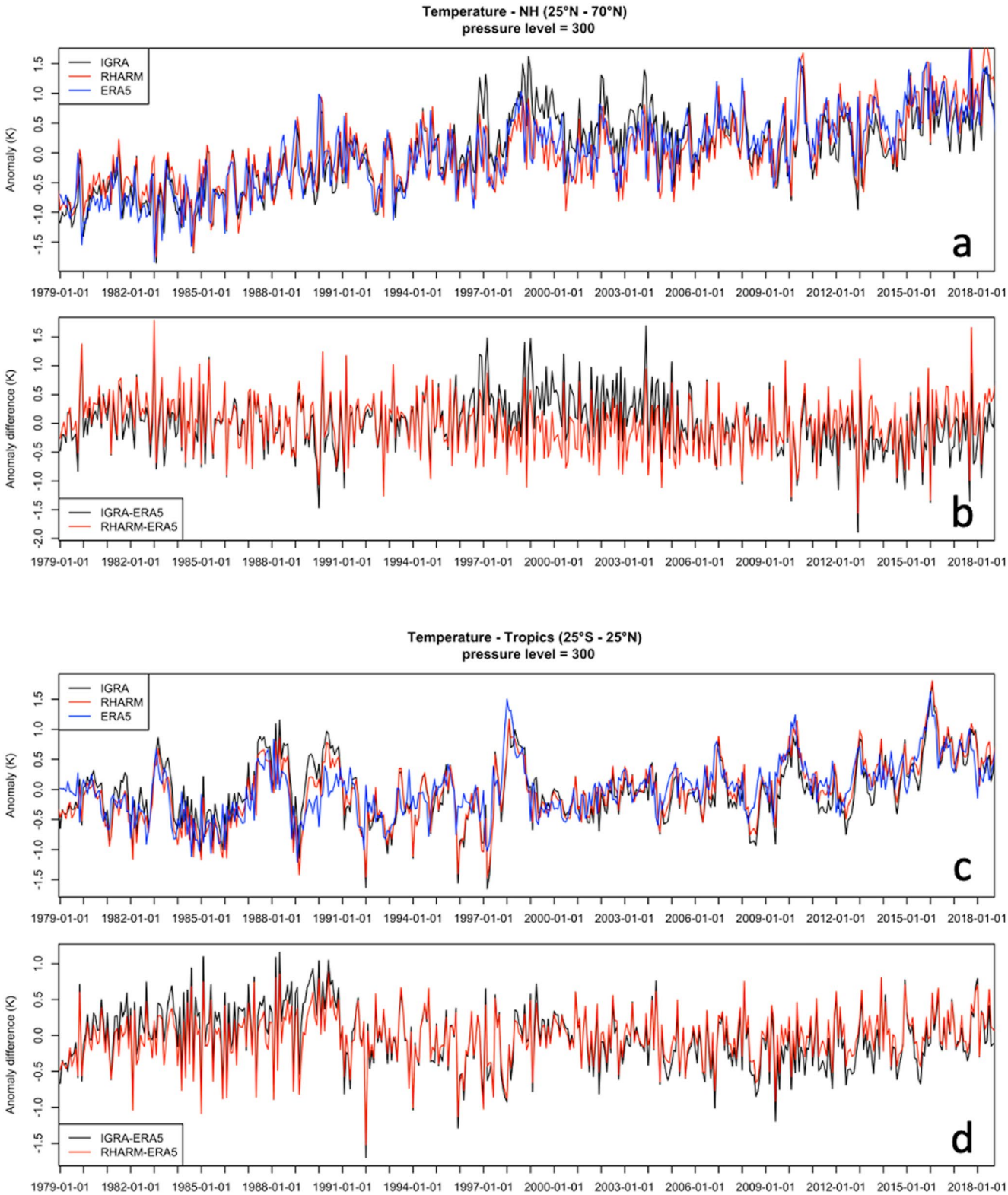

Figure 8. Upper tropospheric interannual variations of 300 hPa temperature for the period 1979–2018 for Integrated Global Radiosonde Archive (black), Radiosounding HARMonization (red), and ERA5 reanalysis (blue) in the Northern Hemisphere and in the tropics. Anomalies are shown in the panels a and c, while in panels b and d differences (to ERA5) are shown. For ERA5 the nearest grid-point to each station and simultaneous vertical profiles on 12 UTC and 00 UTC are selected.

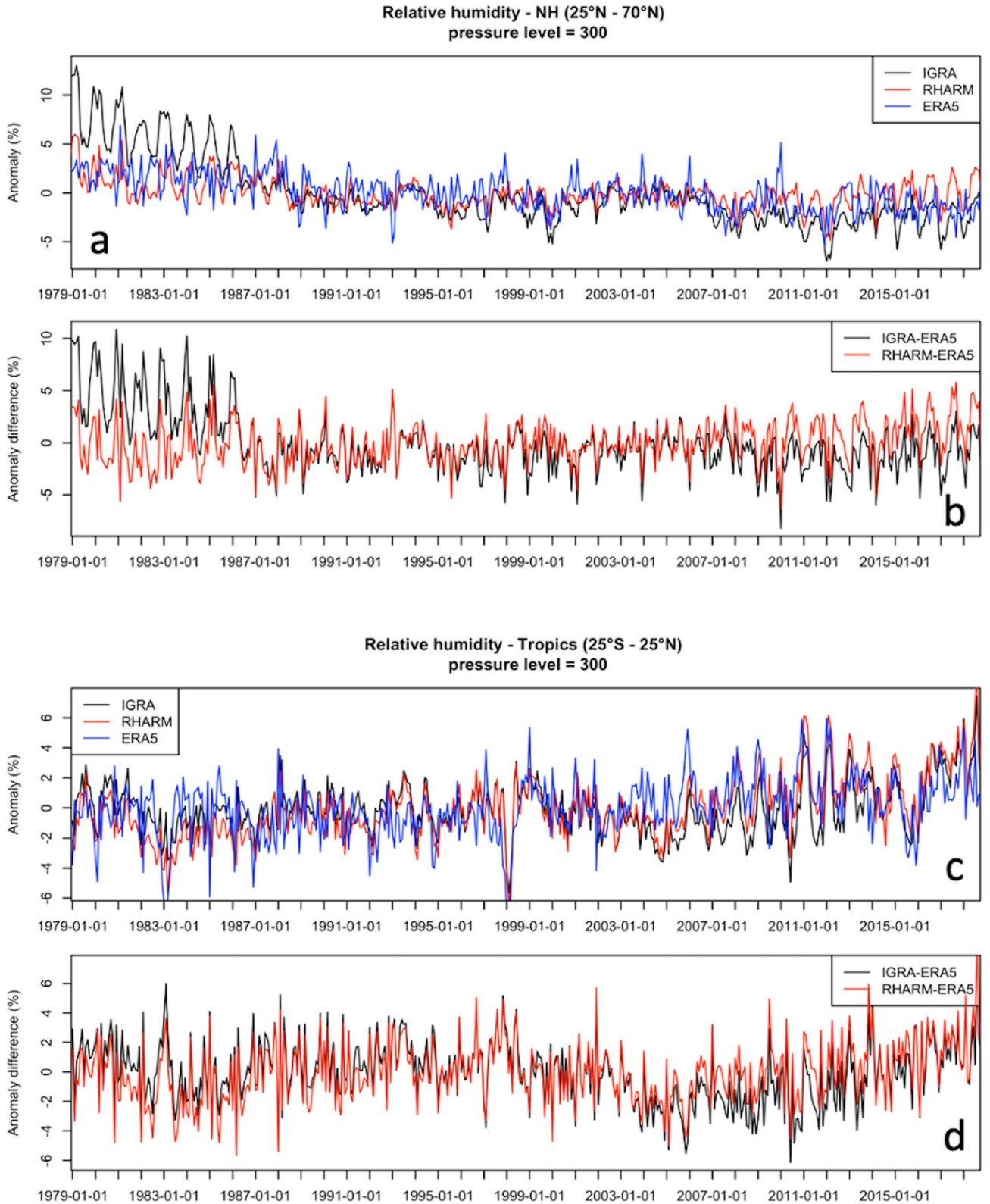

Figure 9. Same as Figure 8 but for relative humidity.

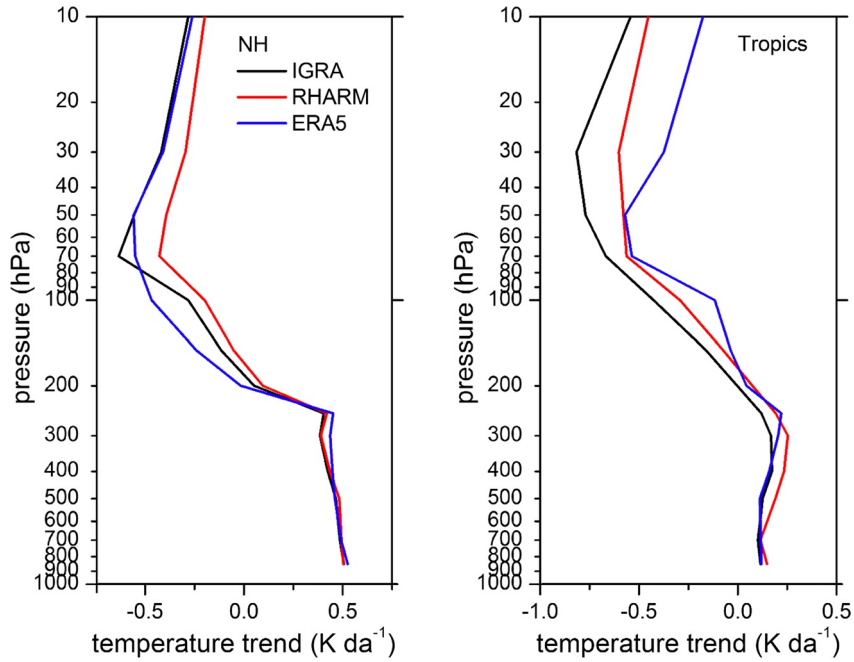

Figure 10. Profiles of temperature trends at mandatory pressure level between 850 and 10 hPa for the period 1979–2018, in the northern hemisphere (left panel) and in the tropics (right panel) for the unadjusted Integrated Global Radiosonde Archive (black line), Radiosounding HARMonization (red), ERA5 (blue) datasets. The ordinate is logarithmic, and the abscissa differs between the two panels.

than 500 hPa. Both datasets significantly differ from the unadjusted IGRA data for pressures below 500 hPa. In the tropics, RHARM and ERA5 shows a similar shape despite a difference up to 1.5% RH da^{-1}, increasing with height. Differently from temperature, ERA5 RH assimilated data are not bias adjusted. It must be remarked that

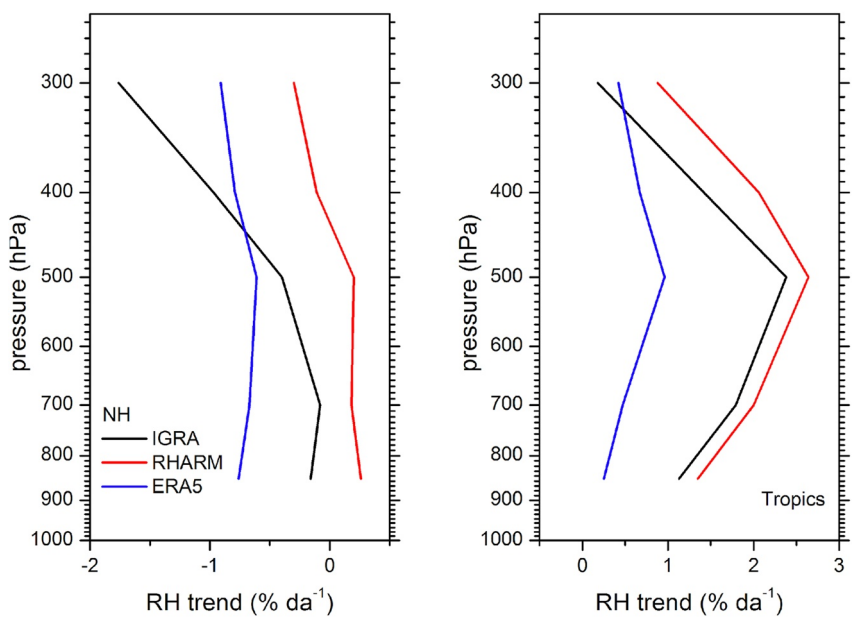

Figure 11. Profiles of relative humidity trends at mandatory pressure level between 850 and 300 hPa, in the period from 1979 to 2018, in the northern hemisphere (left panel) and in the tropics (right panel) for unadjusted Integrated Global Radiosonde Archive (black line), Radiosounding HARMonization (red), ERA5 (blue). The ordinate is logarithmic, and the abscissa differs between the two panels. The comparison is limited to 300 hPa, as water vapor measurements are not always reliable for lower pressures.

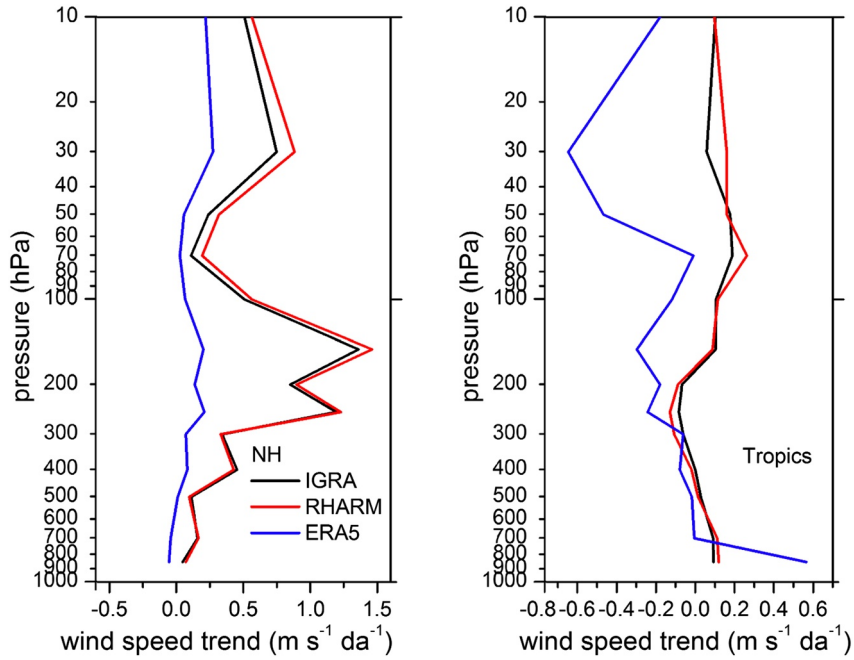

Figure 12. Profiles of wind speed trends at mandatory pressure level between 850 and 10 hPa, in the period from 1979 to 2018, in the northern hemisphere (left panel) and in the tropics (right panel) for unadjusted Integrated Global Radiosonde Archive (black line), Radiosounding HARMonization (red), ERA5 (blue). The ordinate is logarithmic, and the abscissa differs between the two panels.

the comparison cannot ascertain which of the datasets provides the best option to assess RH trends. Fundamentally, the paucity of available estimates makes it difficult to assess structural uncertainties.

4.2.3. Wind Speed

In the NH, below 70 hPa and above 300 hPa the comparison shows differences among the datasets (Figure 12) within 0.5 m s^{-1} $da^{-1}$with a better agreement between the observational datasets below 70 hPa. Between 70 and 300 hPa, IGRA and RHARM show trends larger than 1.0 m s^{-1} da^{-1}, with a small adjustment applied by RHARM. In the tropics, the shape of the trend vertical profiles is almost the same for all datasets from 850 to 300 hPa, while at lower pressures the observational datasets shows similar positive trends, around 0.1 m s^{-1} da^{-1}, differently from ERA5 which is negative with values up to −0.6 m s^{-1} da^{-1} Note that observational constraints in the reanalysis are weaker in this region (e.g., Kawatani et al., 2016), and artefacts due to changes in assimilated observations cannot be excluded.

4.3. Statistics of Early Period Breakpoints and Their Vertical Coherency in Stage II

The homogenization of the Stage II time series for each station, that is, data before 2004, is applied at each mandatory pressure level separately. Therefore, it is informative to study the distribution per year of the breakpoints detected in the Stage II time series for the measured ECVs as well as the correlation of the percentage of breakpoints per ECV at different pressure levels. For the latter purpose, the 100, 300, and 500 hPa levels have been selected as representative of different atmospheric regions (lower stratosphere, upper troposphere, free troposphere, respectively) where different types of biases and resulting adjustments, either height-dependent (solar radiation correction, time-lag correction) or correlated within the entire vertical profile (e.g., sensor calibration), are applied in the data processing.

The decrease in breakpoint detection after 2004 (Figure 13) is due to the progressive introduction of the most recent radiosonde types for which the Stage I RHARM approach can be applied. Furthermore, the percentage of breakpoints decreases going toward the past and this may be both related to a reduction in the number of stations and data available as well as to the adoption of the same type of measurement sensors for longer time periods. It can be noted that Sherwood et al. (2008), instead, showed a decrease of the number of detected breakpoints after

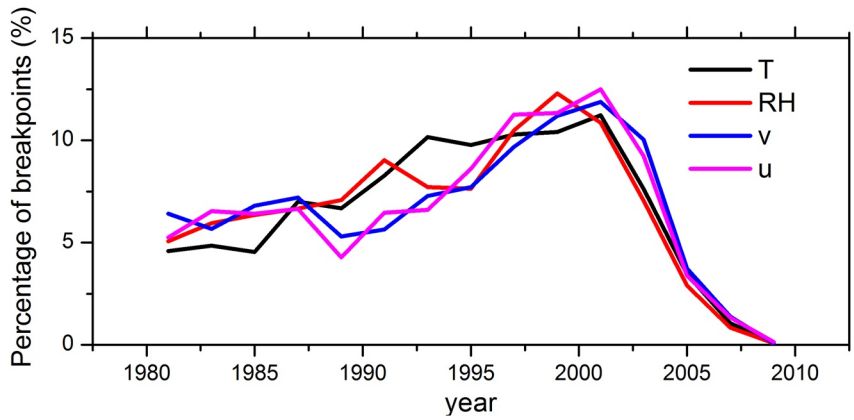

Figure 13. Percentage of breakpoints per year detected by the Radiosounding HARMonization approach for each of temperature, relative humidity, meridional and zonal wind speed, cumulated for all the homogenized time series and at all mandatory pressure levels.

1990–1995, using data from IGRA until 2000, while Haimberger (2005) showed a slight decrease in the number of detected breakpoints after 1990 at 500 hPa but an increase during the same period at 10 hPa. Figure 13 also reveals the good agreement in the percentage of breakpoints identified across the different ECVs.

Breakpoint percentages (Figure 14) show a similar distribution for RH, *u* and *v*, while for temperature the correlation is larger at 300 and 500 hPa. At 100 hPa, instead, in the period 1990–1997 there is a higher occurrence of breakpoints than at other levels indicating either a larger effect of the radiation bias for the sonde models operated

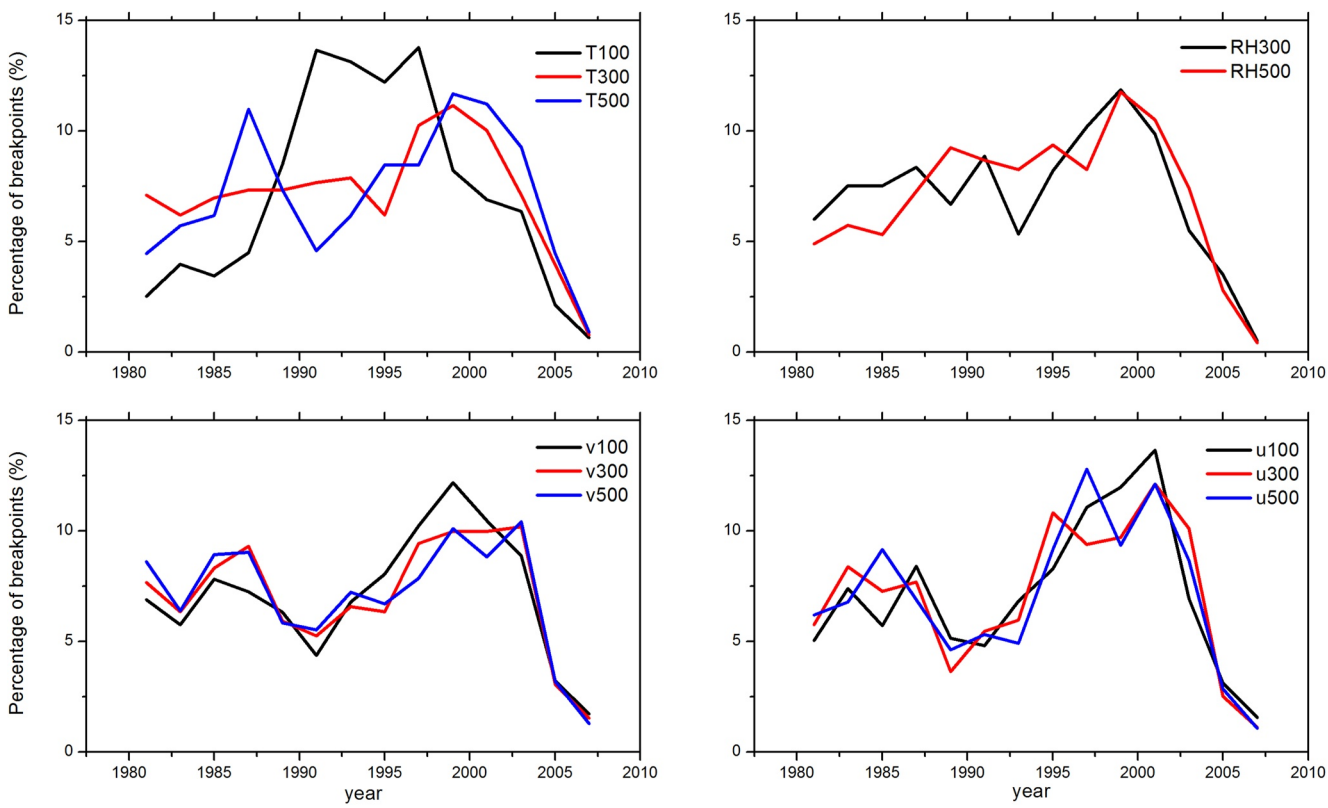

Figure 14. Percentage of breakpoints per year detected by the Radiosounding HARMonization approach for temperature (upper left panel), relative humidity (upper right panel), meridional (bottom left panel) and zonal wind speed components (bottom right panel). Each panel reports the frequency of occurrence per year at three pressure levels, 100 hPa (except for RH), 300 hPa, 500 hPa.

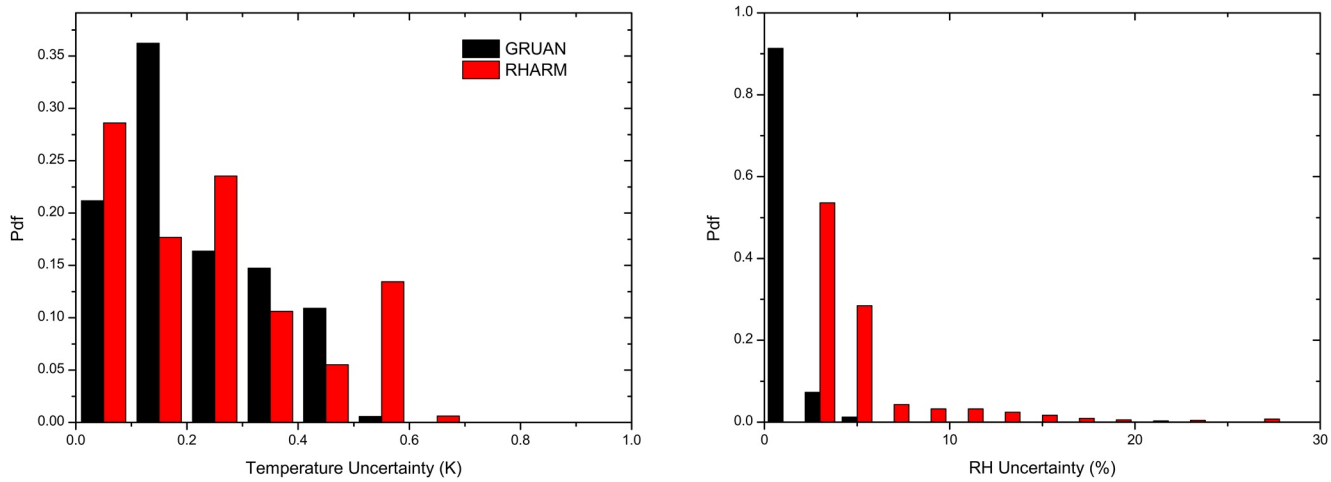

Figure 15. Comparison of pdfs of the uncertainty calculated using the Global Climate Observing System (GCOS) Reference Upper Air Network data processing and the Radiosounding HARMonization approach at the six stations shown in Table 1. Pdfs are relative to temperature (panel a) and relative humidity (panel b).

in this period, or a larger number of false positives than at other levels. This can be linked to the much smaller number of observations available at 100 hPa within IGRA, due to balloon burst which can also impart sampling effects (McCarthy et al., 2008; Sy et al., 2021).

To assess the coherence of breaks within individual stations between significant levels, an analysis was carried out on the correlation between occurrence dates at 300 hPa with respect to the dates at 100 and 500 hPa. Within a window of 2 months, correlation for temperature and RH breakpoints at 500–300 hPa is about 0.2, while it rises to 0.36 within 6 months and to 0.6 within a year. For wind vector components, within a time difference of 2 months, correlation for temperature breakpoints at 500–300 hPa is 0.26, while correlation is 0.52 within 6 months and 0.81 within 1 year. Very similar values are obtained for 300–100 hPa, except they were somewhat smaller for temperature. These results may indicate a temporal mismatch in the detection of the same breakpoint at different pressure levels. However, depending on the nature of the systematic effect, more or less significant biases may be present in different atmospheric ranges and, therefore, correlation in breakpoint detection among the selected levels would not be expected to be perfect. There are homogenization methods assigning a breakpoint to all pressure levels irrespective of whether a break is detected at a given level, assuming biases due to instrumental effects are vertically correlated (e.g., Sherwood et al., 2008). Although in previous studies based on monthly averages breakpoints at multiple levels were considered to discard false positives (if they only appeared at one or two levels), this choice was not considered for the current version of RHARM.

5. Quantification and Presentation of Uncertainties

A unique value of RHARM compared to other homogenized datasets is that, for the first time, an estimation of the uncertainty is provided for each single observation (i.e., at each pressure level). Other existing homogenized radiosounding datasets were, instead, mainly focused on estimating trend uncertainties using error models or providing observation uncertainties using, for example, Desroziers' method (Desroziers et al., 2005). In this section, statistics on the RHARM estimated uncertainties are provided.

Considering data at the six stations shown in Table 1 only in the GRUAN era, the uncertainty for RHARM is generally larger than the uncertainties obtainable using the GRUAN processing as expected given the methodological considerations outlined in Section 3. In particular, for temperature (Figure 15, left panel), the median value of the GRUAN uncertainty is 0.16 K compared to 0.22 K of RHARM (median values are considered for the analysis, given the shape of the pdf). The interquartile range (IQR) for GRUAN is 0.20 K, while for RHARM it is 0.26 K. These numbers confirm that on average the uncertainty estimation obtained for RHARM is somewhat greater than the GRUAN uncertainty. Nevertheless, due to the nature of the assumptions made within RHARM, in some cases its uncertainty may be underestimated compared to that of GRUAN, as seen for values below 0.1 K. These values are mainly related to nighttime measurements.

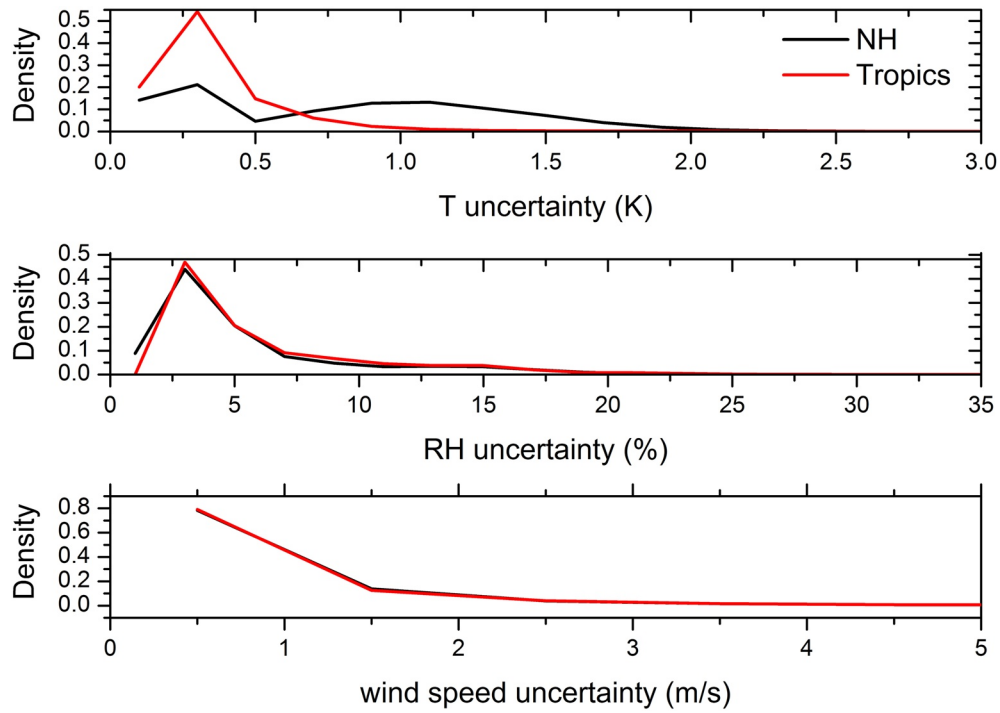

Figure 16. Comparison of the probability density function of the total uncertainties estimated for all the Radiosounding HARMonization temperature (T), relative humidity (RH) and wind speed (w) data since 1978 for the stations in the tropics and in the Northern Hemisphere (NH).

For RH (Figure 15, right panel), the median value of the GRUAN uncertainty pdf is about 1.1% versus 3.6% for RHARM, with an IQR for GRUAN of 0.1% and 3.0% for RHARM. Maximum values observed with GRUAN are less than 8% while RHARM has values larger than 10% and a very few values larger than 20%.

In Figure 16, the density function of the uncertainties estimated for the RHARM data are shown for the NH and the tropics. The comparison for the temperature uncertainties shows that the density function in the NH is bi-modal with modes centered around 0.5 and 1.0 K, with most values smaller than 2.0 K. In the tropics, values are smaller than 1.5 K. A large fraction of the values in both regions is around 0.25 K and these values are referring to the values of the Stage I time series. For the relative humidity, both the distributions are bimodal with values of the uncertainties larger in the tropics than in the NH. A large fraction of the RH uncertainty values is smaller than 10%, while the second distribution mode is 14%–15% RH uncertainty. Finally, for the wind speed uncertainty the distributions overlap in the selected latitude belts.

6. Discussion and Conclusions

RHARM provides a new approach for the homogenization of radiosounding temperature, RH and wind measurements. RHARM differs from previous efforts due to the use of reference measurements to calculate and adjust for systematic effects in the most recent portion of the time series, when metadata are available, and used the same adjustments as a constraint for homogenizing historical time series. RHARM adjusted fields are not affected by cross-contamination of biases across stations and are fully independent on atmospheric reanalysis data. This gives the data set independence properties across stations that may not have accrued to previous datasets. A significant benefit is that each harmonized time series is provided with an estimation of the uncertainty for each observation. The RHARM approach enables a more comprehensive exploration of uncertainties in historical time series.

Results from the analyses about the applied RHARM algorithm show that:

1. RHARM temperature data distribution is warmer than IGRA in the NH, due to the predominance of cooling biases affecting the IGRA time series since 1978, while RHARM is slightly cooler than IGRA in the tropics. For RH, RHARM adjusts the IGRA data dry bias, in particular below 20%–30% RH and above 52% RH, both

in the NH and at the tropics. For wind speed the systematic effects have a smaller magnitude than for temperature and RH, and IGRA and RHARM data distributions are fairly similar.

2. RHARM increases the spatial homogeneity of trends compared to IGRA (examples are provided at 850, 300, and 100 hPa) over 1978–2000. The reduction in the geographic spread of trend values is stronger for RH and wind than for temperature, although in the NH, especially at 100 hPa, and in South America also temperature trends are more homogeneous.
3. The comparison of RHARM with GRUAN shows that RHARM-GRUAN temperature difference is much reduced compared to the GRUAN-IGRA difference at all levels and for all ECVs, as expected given the RHARM methodology.
4. The comparison between RHARM, IGRA, and ERA5 shows that the adjustments applied in the RHARM data processing reduce the differences between ERA5 and observational data, especially for temperature in the NH, where the most significant adjustment reduces the differences from 0.5 to 0.1 K in the decade 1996–2005, on average, and for relative humidity, where the differences are adjusted from values of more than 10% RH to less than 5% RH in the decade 1978–1987. Adjustments to wind speed anomalies, although smaller, also show the improvement in the times series in comparison with ERA5.
5. The study of the vertical correlation of the breakpoints identified by RHARM at three mandatory pressure levels (100, 300, 500 hPa) shows that 60% of the changepoints are correlated within 1 year for T and RH, while this value increases to 81% for wind. RHARM uncertainties are generally larger than GRUAN.

Analyses of interannual variations and trends from RHARM data in the period 1979–2018 shows:

1. Warming trends of temperature are smaller than 0.5 K da^{-1} at pressures higher than 250 hPa, while trends are cooling up to 0.25 K da^{-1} below that level. In the tropics, trends are smaller 0.25 K da^{-1} at pressure higher than 250 hPa, while colling and within 0.5 K da^{-1} below. Results are in good agreement with ERA5, especially for pressures higher than 200 hPa in the NH and higher than 50 hPa in the tropics.
2. For RH, in the NH RHARM shows slightly positive or near-zero trends at pressures higher than 500 hPa, while trends are negative (up to 0.2% RH da^{-1}) below. In the tropics, trends are positive and higher than to 1.0% RH da^{-1} over the entire vertical range (larger than 2.0% RH da^{-1} at 500 hPa). Comparisons with ERA5 show differences probably due also the fact that, differently from temperature, ERA5 RH assimilated data are not bias adjusted. The increasing humidity anomalies in the period 2015–2019 and the positive RH trend observed in RHARM data at the tropics appear to be correlated with the warm ENSO event in 2015/16.
3. For wind speed, trends in the NH are smaller than 0.2 m s^{-1} da^{-1} at pressures higher than 300 hPa, whereas trends are greater below and, in particular, larger than 1.0 m s^{-1} da^{-1} in the 100–300 hPa interval. In the tropics, trends are smaller than 0.2 m s^{-1} da^{-1} along the entire vertical range. The comparison with ERA5 indicates a good agreement in the troposphere, mainly in the tropics.

From a technical point of view, it is useful to remark that wind radiosounding data are processed with proprietary software routines from the respective manufacturers which apply distinct smoothing to the data, the RHARM wind profile may have a different effective vertical resolution (Iarlori et al., 2015). The unavailability of the raw (manufacturer pre-processed) data inhibits reprocessing of the data to provide data at a common resolution or even at a known resolution, which could be controlled in the RHARM software in order to remove spurious effects on the wind measurement between the radiosondes.

In an ideal world, the collection and preservation of raw data by all radiosounding stations would allow to build the highest possible quality data set of radiosounding measurements by reprocessing all the data consistently to metrologically traceable standards. In the real world, save for GRUAN sites and intercomparison campaigns, we do not have such an option. There is an action currently under discussion in GCOS in its most recent Implementation Plan (personal communication by GCOS secretariat) to explore the possibility to collect and reprocess data from those sites who usually hold the original raw count data locally, although the timeline and the resources to start the action are still uncertain. The final goal of RHARM is to calculate average adjustments which should result in an improved estimation of the climatological variability for temperature, humidity and wind profiles The RHARM approach represents an innovative solution that is closer to a "traceable" estimate with uncertainties, in particular for the data after 2004. Considering that RHARM is an algorithm aiming at adjusting radiosounding time series without using model-based data or ancillary information from neighboring stations, the adjustments applied to stations in each climate region may be less homogeneous than for other homogenization approaches.

This is certainly one of the main reasons for the small contrast between IGRA and RHARM for temperature (Figures 6 and 7). The impact of LOESS filtering of each times series, the timeliness of CUSUM in the changepoint detection and the ability of RHARM to preserve the natural climate variability are other factors impacting the algorithmic efficiency.

Any future availability of new WMO/CIMO intercomparison data will enhance the capability of the RHARM approach to improve the quality of both near-real time and historical radiosoundings data. Moreover, the availability of the enhanced BUFR data reports (BTEM/BTEF files replacing TEMP and previous BUFR version), for radiosounding measurement submitted to the WMO Information System (WIS), to foster the reporting of high-resolution vertical profiles with improved metadata, will help reduce the gap between files reported by reference and baseline networks. These files are made available upon request by ECMWF (P.I. Bruce Ingleby) and their metadata are already incorporated in the latest version of RHARM. The availability of metadata from 2016 onwards, when enhanced BUFR files start to be available, will also improve near real-time data availability. New GRUAN data products, such as for the Meisei iMS-100 sonde (Kobayashi et al., 2019) and Vaisala RS41 (von Rohden et al., 2021) will be incorporated into subsequent versions of RHARM. This is in line with the design of the RHARM algorithm which allows continuous improvements exploiting new improved radiosonde sensors technology and processing algorithms as they become available in the future.

Appendix A: Validation of Uncertainties

A major innovation of the RHARM data set is the estimation of uncertainties for temperature, relative humidity and wind measurements for each pressure level of the adjusted IGRA profiles. Uncertainty estimation is fundamental for any type of observation, enabling a metrologically consistent comparison with other datasets. To this purpose, a proper validation of the uncertainties is also required. In Figure A1, it is shown an example of a wind time series (for the both the u and v components) reporting also the uncertainties calculated according to the RHARM approach.

Following the principles of metrology, RHARM estimated uncertainties must be also validated. Validation of uncertainties means that these must be "evaluated by independent means to establish quantitative realism and the credibility of the uncertainty estimates" (Merchant et al., 2017). To provide a validation of the RHARM uncertainties, the methodology of Merchant et al. (2017) has been applied. This is based on the study of the probability density function of the ratio:

$$\frac{x_{RHARM} - x_{ref}}{\sqrt{u^2_{RHARM} + u^2_{ref} + u^2_{mis}}} \tag{A1}$$

where x_{RHARM} is the RHARM estimate, x_{ref} indicates an independent estimate of the measurand, u denotes the uncertainty and u_{mis} is the geophysical variability arising from temporal, spatial, and definitional mismatch between RHARM and reference data. A correct quantification of uncertainties and variability should be reflected in a normal distribution of the ratio defined in Equation A1, with a standard deviation equal to unity. Deviations from zero are due to discrepancies between RHARM and the reference.

Acknowledging that the ideal solution for the validation should be based on independent reference measurements (Thorne et al., 2017) of the same measurand, GRUAN data would be the ideal candidate. However, RHARM has used information from and mimics part of the GRUAN processing, meaning that circularity considerations preclude its use for such a purpose. An alternative solution is adopted, which is to use the ERA5 background (6-hr forecast) as a reference value. Whilst this background is likely a reliable estimation of the atmospheric state, it is not a true reference measurement in that it is not itself an SI traceable measurement, nor does it have comprehensive uncertainty estimates. Observation minus Background (O-B) departures have previously been used as a diagnostic tool for different latitude belts (Ingleby, 2017). They also form the basis for the RAOBCORE/RICH family of data set approaches (Haimberger et al., 2012). Therefore, the use of the ERA5 background as a reference for the test described in Equation A1 appears viable to infer quantitative information for validating the uncertainties. Other candidates exist, such as radio occultation (RO) satellite measurements (Bauer et al., 2014), which are a valuable solution for dry temperatures in the UTLS, while for the mid and lower troposphere the deconvolution of temperature and RH in the retrieval is dependent on a first guess model. Furthermore, RO profiles can rarely provide information all the way down to the surface.

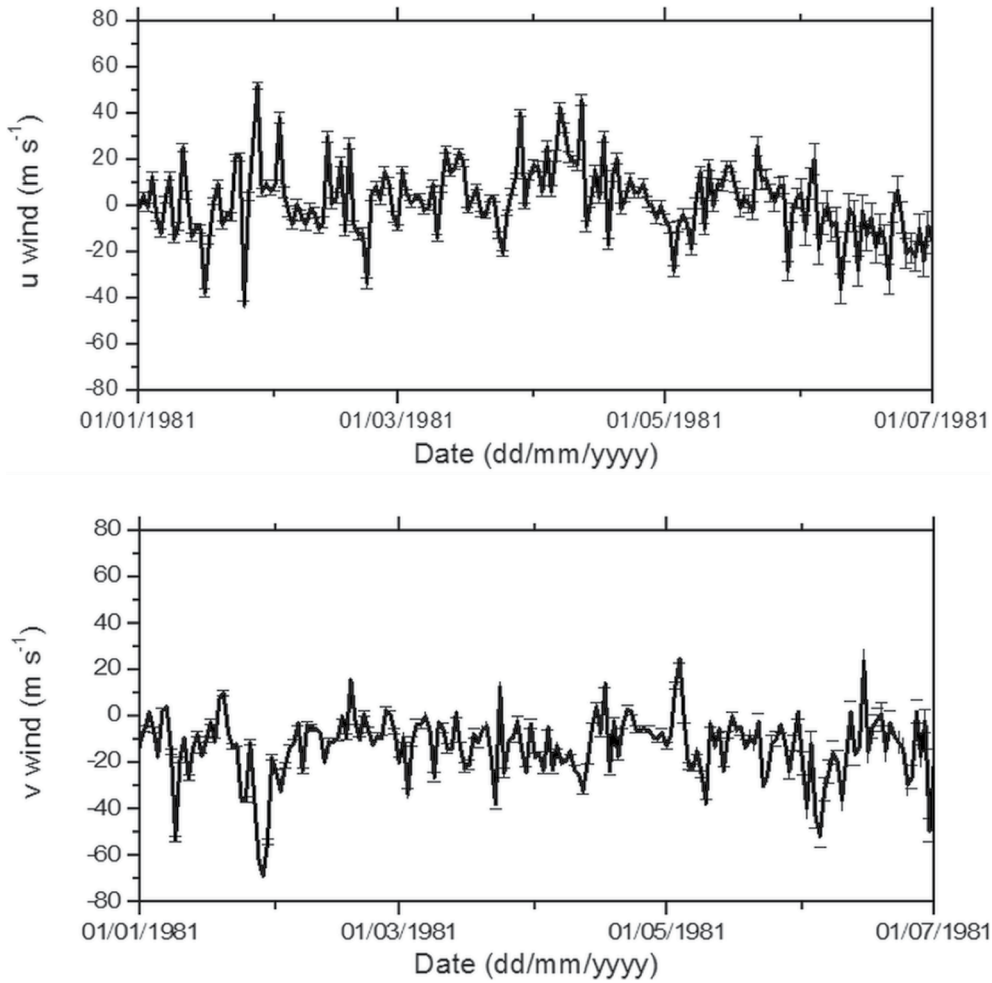

Figure A1. Top panel, zonal wind component (u) time series at 300 hPa (only nighttime) for the Sodankyla station with the uncertainties calculated using Radiosounding HARMonization for the period from 01/01/1981 to 01/07/1981. Bottom panel, same as top panel but for meridional component (v). The vertical bars show the random uncertainties quantified using the statistical method, and their plotting has been reduced to one value each two of the time series.

Using the background as the reference data set in Equation A1, u_{ref} has been estimated applying the Leave-One-Out Cross validation method, LOOCV (Stone, 1974), to the background while u_{mis} is evaluated as the standard deviation of the O-B climatology at each station. The uncertainty validation is carried out separately for two periods: 2004–2019, for which almost all IGRA data have been post-processed using a GRUAN-like adjustment, and 1980–2003, for which the uncertainty estimation of RHARM is obtained by constraining the residuals on a monthly basis. The validation is focused on temperature and RH uncertainties because there is still limited information on the homogeneity of the ERA5 background wind data.

For the period 2004–2019 (Figure A2), in the NH, the ratio for temperature has a mean value of 0.18, while the standard deviation is 0.72 indicating that the uncertainty at 300 hPa for temperature is overestimated by about 28%. The overestimated values of the uncertainty decrease the distribution near the central peak compared to the fitted curve, while it is slightly larger toward the tails. In the tropics, a mean value of the ratio of 0.44 and a standard deviation of 0.89 indicate that the uncertainty is overestimated by about 11%. For RH, in the NH the uncertainty is underestimated with the mean value of the ratio is −1.3 and the standard deviation of 1.2, while the uncertainty is overestimated in the tropics the mean value is −0.69 with a standard deviation of 0.78 and a larger number of overestimated values than in the NH. For the RH uncertainties, the distributions are negatively skewed and almost normally distributed. This might be related to systematic effects on the O-B comparison, possibly due to inhomogeneities in the O-B departures within an entire latitude belt, which could broaden the O-B distribution and influence the value of the validation using the model forecast as a reference.

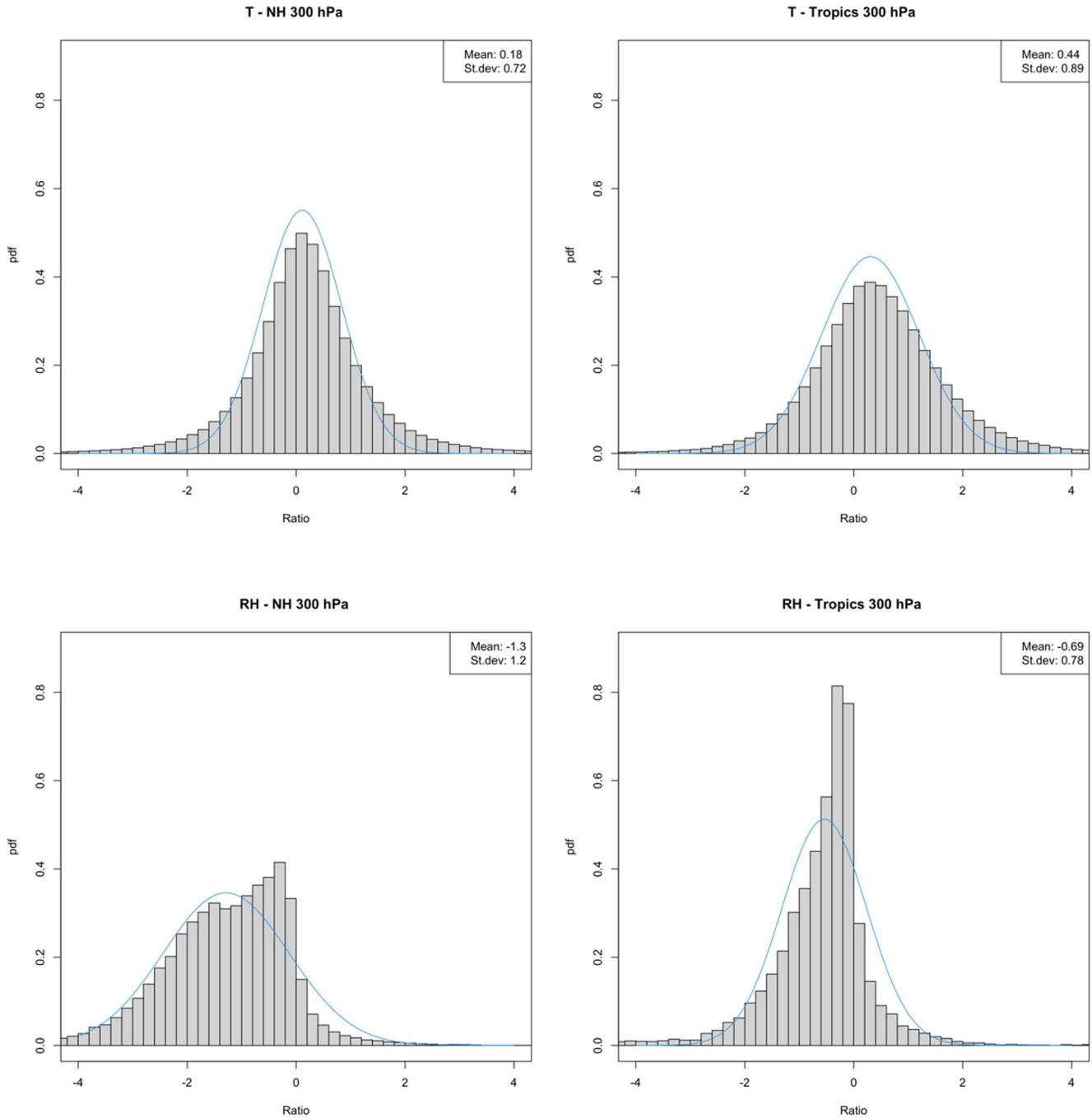

Figure A2. Probability density functions (pdfs) of the ratio reported in Equation A1 calculated using O-B data (Radiosounding HARMonization [RHARM]-minus-Background) in the NH (left panels) and in the tropics (bottom panels) at 300 hPa for temperature (top panels) and for RH (bottom panels) to validate the uncertainties estimated using the RHARM approach. The pdfs refer to the RHARM uncertainty values estimated in the period 2004–2019. Background data are from the ERA5 6-hr forecast model. For comparison with ideal uncertainty estimates, the best fitted normal distribution to each data set (blue line) is also shown. In an ideal case where uncertainty would be properly estimated with the RHARM algorithm, the distribution should have a standard deviation equal to unity. Deviations from zero are due to the O-B discrepancy.

For the period 1980–2003 (Figure A3), in the NH the ratio for temperature is near zero, while the standard deviation is 0.97 indicating that the uncertainty at 300 hPa for temperature is well estimated. The same is true for the tropics where the mean value of the ratio is 0.06 and the standard deviation is 1.0. In both cases the ratios are normally distributed. For RH, both in the NH and at the tropics, the uncertainty is overestimated by 15% and

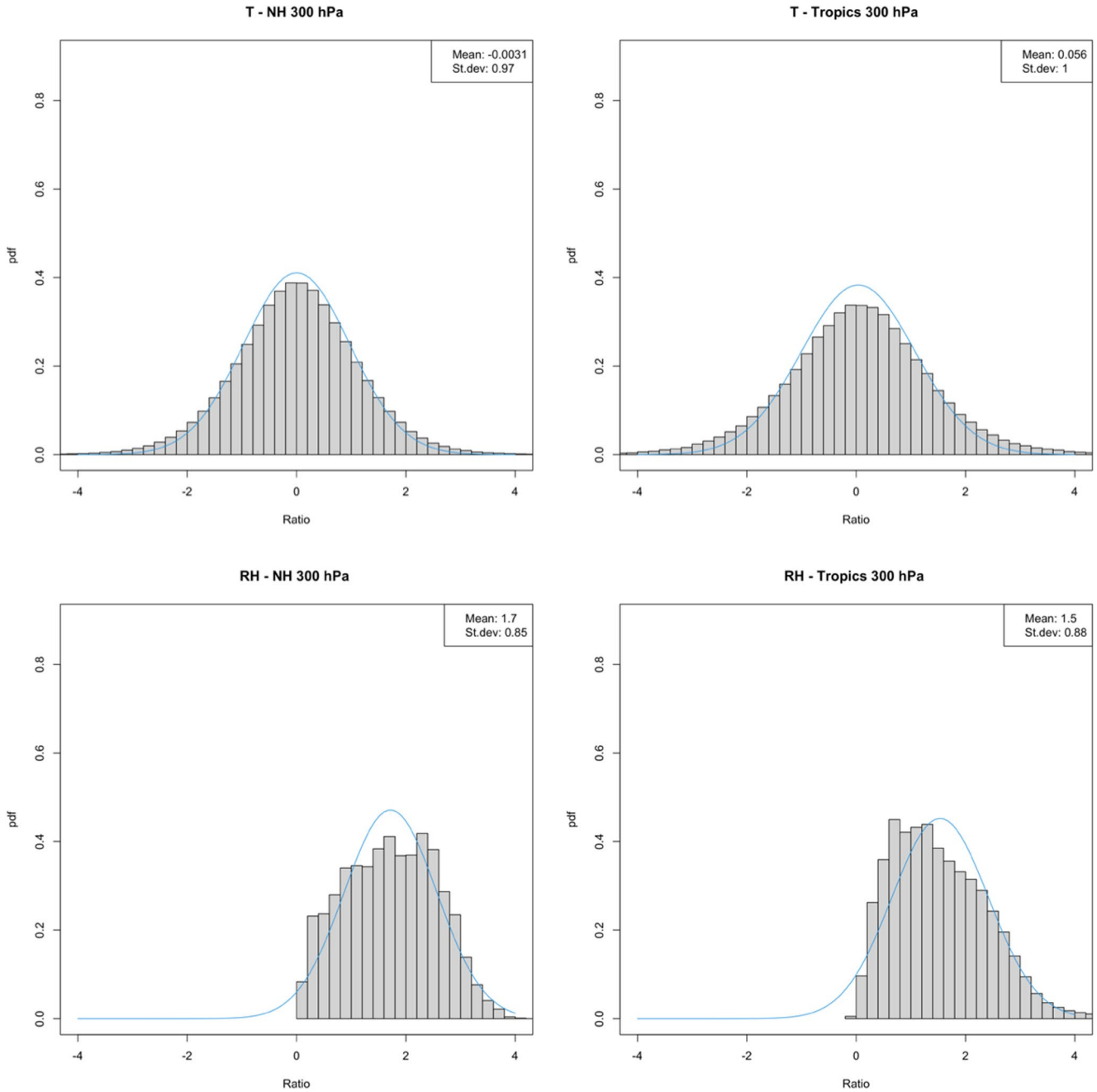

Figure A3. Same as Figure A2, but for the period 1980–2003.

12%, respectively, while the mean value of the ratio is 1.7 and 1.5, revealing a systematic effect affecting the O-B comparison like the period 2004–2019. The distribution in the NH is negatively skewed, while in the tropics it is positively skewed.

In general, the RHARM uncertainties appear to be a reasonable estimate or a slight overestimate of the theoretical distribution, except for the RH in the NH after 2004 where they are an underestimate. It is arguably preferable to have an over-dispersive uncertainty estimate than an under-dispersive estimate for most applications. Nevertheless, future versions of the RHARM data set will be designed to improve the uncertainty estimation, also through the implementation of more sophisticated models, using techniques like kriging or modeling Gaussian processes.

Temperature uncertainties are, in general, better behaved than RH uncertainties but it is unclear whether this relates to the uncertainty quantification or O-B field estimate issues.

Appendix B: RHARM Consistency With GRUAN

Although built to mimic the GRUAN processing, the RHARM algorithm, discussed in Section 3, is not applied to the raw unprocessed radiosonde data, because these are not available from any global repository and often not retained by the station managers or National Meteorological or Hydrometeorological Services (NMS). This difference with the GRUAN data processing may generate discrepancies between the RHARM and GRUAN data, which must be quantified. By construction, the RHARM approach is expected to be similar on average to the GRUAN products. For

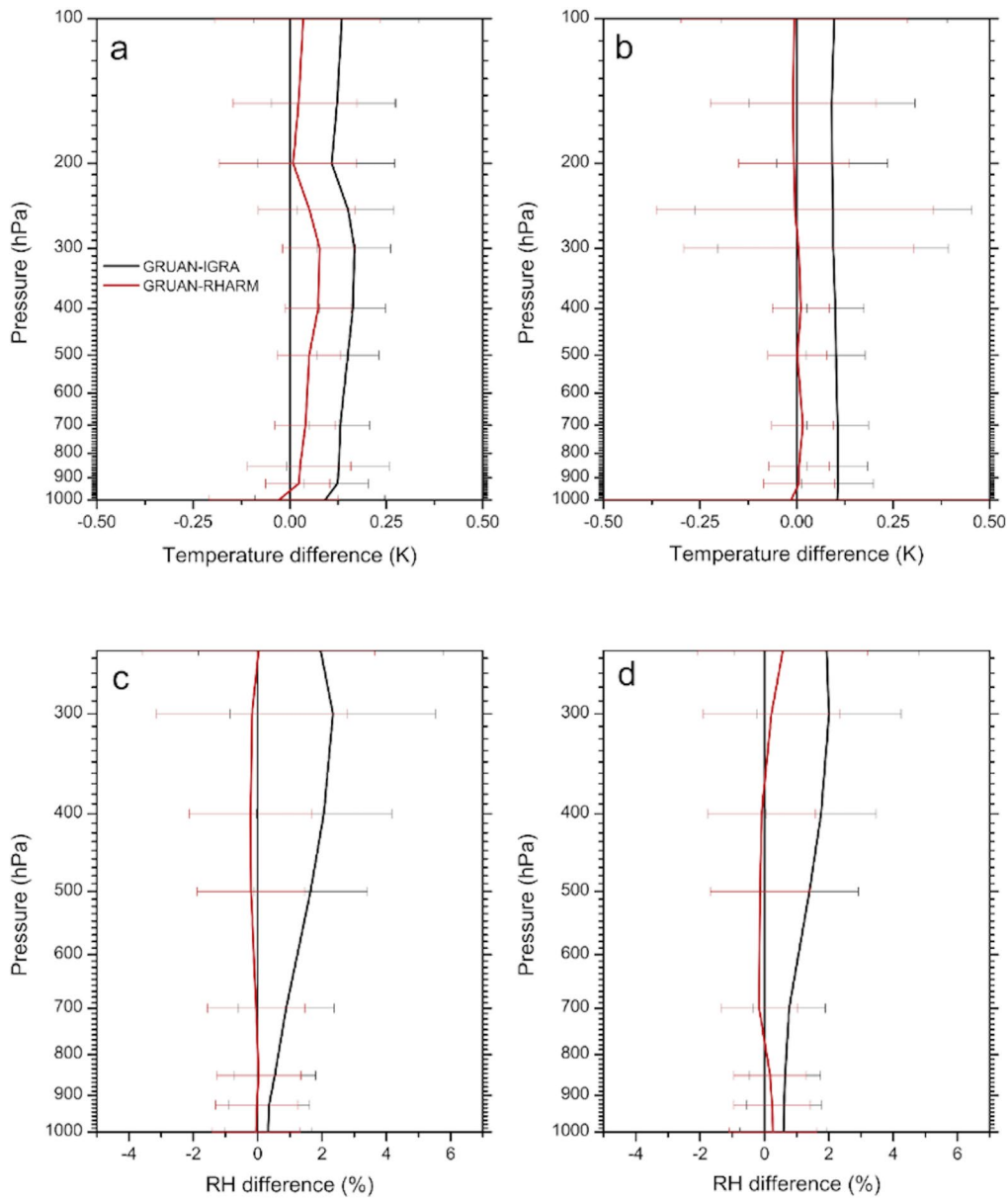

Figure B1. Mean difference profiles of temperature (top panels) and relative humidity (bottom panels) with the corresponding standard deviations (horizontal bar) calculated from the comparison of the nighttime (panels a and c) and daytime (panels b and d) difference "Global Climate Observing System (GCOS) Reference Upper Air Network (GRUAN) minus Integrated Global Radiosonde Archive" (black lines) and "GRUAN minus Radiosounding HARMonization" (red lines) for the profiles available at all GRUAN stations, in the period 2008–2018.

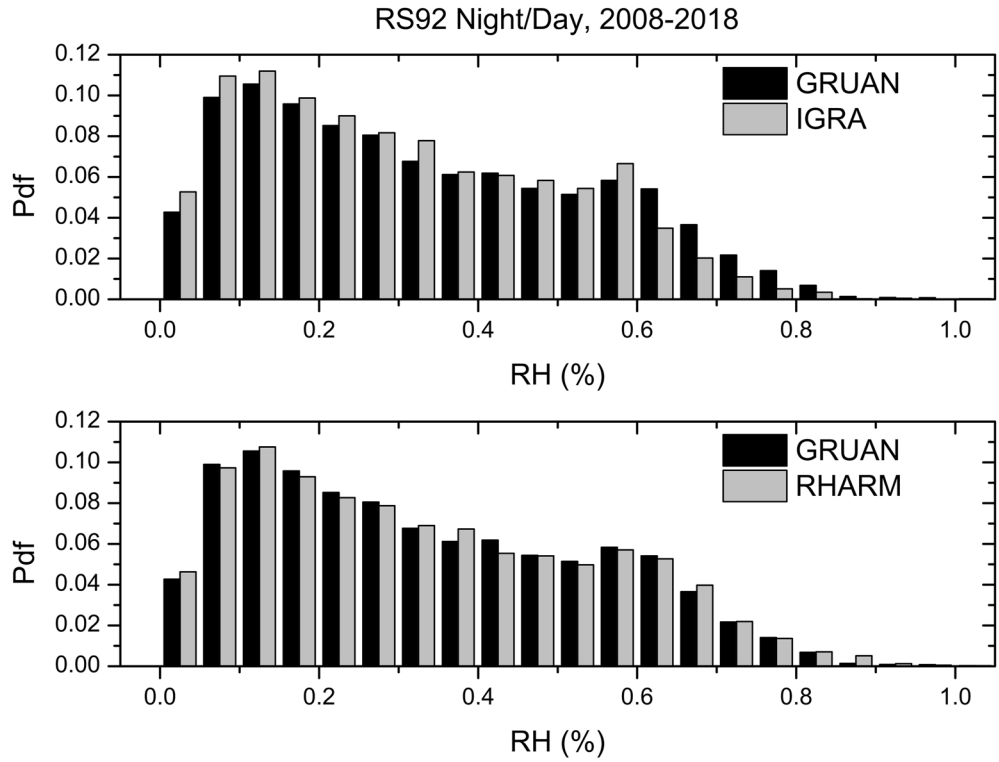

Figure B2. Top panel, comparison between Global Climate Observing System (GCOS) Reference Upper Air Network (GRUAN) (black) and Integrated Global Radiosonde Archive (gray) RH measurements at 300 hPa for the profiles available at all GRUAN stations (only RS92 sondes), in the period 2008–2018 The comparison comprises all the night and daytime observations on 00:00 and 12:00 UTC. Bottom panel, same as top panel but for GRUAN (black) and Radiosounding HARMonization (gray).

temperature at night, the difference GRUAN-IGRA is almost constant from the surface up to 300 hPa with a value of 0.12–0.13 K, while at lower pressure it is a slightly smaller with values of 0.1 K (Figure B1). In this comparison data from stations in Table 2 only in the GRUAN era (since 2008) have been considered. Conversely, the GRUAN-RHARM difference is close to zero at all levels, with values smaller than 0.05 K up to 250 hPa and close to zero at higher altitudes. During the day, the RHARM-GRUAN difference is near zero at all levels, while the GRUAN-IGRA difference is nearly constant at all the pressure levels at c.0.12 K. The standard deviations for both the differences are very similar and for both night and day show increasing values toward lower pressures from 0.2 to 0.3 K.

For RH at night, the GRUAN-IGRA difference increases with height from less than 0.5%–2.0% and, during the day, from 0.7% to 1.8% (Figure B1). The RHARM adjustments reduce these differences on average near zero, both during night and day. The standard deviation of the RH difference is similar for both the difference profile at night and day with values ranging between 1.5% and 5.0% RH, increasing with decreasing pressures.

In contrast to temperature and RH, the wind speed mean differences (not shown) for both the GRUAN-IGRA and GRUAN-RHARM difference profiles are very close to zero from 1,000 to 300 hPa. Above this altitude, the GRUAN-RHARM difference is smaller than IGRA-GRUAN difference and consistently below 0.05 m/s, while IGRA shows differences with GRUAN within about ±0.3 m/s. The larger variability of GRUAN-IGRA at altitudes above 250 hPa are due to the different data interpolation carried out in the GRUAN and IGRA profiles in a region where the wind speed variability is high (i.e., small differences in the pressure data interpolation are representative of large altitude differences and may generate large differences in the wind speed values to compare).

The RHARM RH values become considerably more similar to GRUAN, especially for values higher than 55% RH (Figure B2). These results imply that manufacturer data processing applied to the RH radiosounding profiles measured by Vaisala RS92 radiosondes is not adequate to compensate for instrumental effects, as it is inducing an apparent dry bias compared to the metrologically traceable GRUAN processing. The RHARM procedures are able to mimic, at the aggregated level, the GRUAN processing adjustments.

Data Availability Statement

The RHARM dataset is provided in textual format (comma-separated values) and is available through the Climate Data Store (CDS) at https://cds.climate.copernicus.eu/cdsapp#!/dataset/insitu-observations-igra-baseline-network?tab=overview.

Acknowledgments
This work was done on behalf of the European Union's Copernicus Climate Change Service implemented by ECMWF. Use of the RHARM data as stated in the Copernicus license agreement is acknowledged. Thanks to the GRUAN Lead Center for sharing the Look-up table of the Streamer RTM. The Yangjiang Intercomparison Data set (ID2010) has been released upon agreement with the WMO YID protocol, signed by CNR-IMAA and WMO on 27 July 2017. We would also like to thank the reviewers and editors for their suggestions, which improved the manuscript.

References

Angell, J. K. (2003). Effect of exclusion of anomalous tropical stations on temperature trends from a 63-station radiosonde network, and comparison with other analyses. *Journal of Climate*, *16*(13), 2288–2295. https://doi.org/10.1175/2763.1

Barrodale, I., & Roberts, F. D. K. (1974). Solution of an overdetermined system of equations in the *l* 1 norm [F4]. *Communications of the ACM*, *17*, 319–320. https://doi.org/10.1145/355616.361024

Bauer, P., Radnóti, G., Healy, S. B., & Cardinali, C. (2014). GNSS radio occultation constellation observing system experiments. *Monthly Weather Review*, *142*(2), 555–572. https://doi.org/10.1175/MWR-D-13-00130.1

Bodeker, G. E., Bojinski, S., Cimini, D., Dirksen, R. J., Haeffelin, M., Hannigan, J. W., et al. (2016). Reference upper-air observations for climate: From concept to reality. *Bulletin of the American Meteorological Society*, *97*, 123–135. https://doi.org/10.1175/BAMS-D-14-00072.1

Bojinski, S., Verstraete, M., Peterson, T. C., Richter, C., Simmons, A., & Zemp, M. (2014). The concept of essential climate variables in support of climate research, applications, and policy. *Bulletin of the American Meteorological Society*, *95*, 1431–1443. https://doi.org/10.1175/BAMS-D-13-00047.1

Bulgin, C. E., Merchant, C. J., & Ferreira, D. (2020). Tendencies, variability and persistence of sea surface temperature anomalies. *Scientific Reports*, *10*, 7986. https://doi.org/10.1038/s41598-020-64785-9

Calbet, X., Peinado-Galan, N., Ripodas, P., Trent, T., Dirksen, R., & Sommer, M. (2017). Consistency between GRUAN sondes, LBLRTM and IASI. *Atmospheric Measurement Techniques*, *10*, 2323–2335. https://doi.org/10.5194/amt-10-2323-2017

Cramer, W., Guiot, J., Fader, M., Garrabou, J., Gattuso, J.-P., Iglesias, A., et al. (2018). Climate change and interconnected risks to sustainable development in the Mediterranean. *Nature Climate Change*, *8*, 972–980. https://doi.org/10.1038/s41558-018-0299-2

Dai, A., Wang, J., Thorne, P. W., Parker, D. E., Haimberger, L., & Wang, X. L. (2011). A new approach to homogenize daily radiosonde humidity data. *Journal of Climate*, *24*, 965–991. https://doi.org/10.1175/2010JCLI3816.1

Dee, D., Fasullo, J., Shea, D., Walsh, J., & National Center for Atmospheric Research Staff. (Eds.). (2016). *The climate data guide: Atmospheric reanalysis: Overview & comparison tables*. Retrieved from https://climatedataguide.ucar.edu/climate-data/atmospheric-reanalysis-overview-comparison-tables

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., et al. (2011). The ERA-interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, *137*, 553–597. https://doi.org/10.1002/qj.828

Desroziers, G., Berre, L., Chapnik, B., & Poli, P. (2005). Diagnosis of observation, background and analysis-error statistics in observation space. *Quarterly Journal of the Royal Meteorological Society*, *131*, 3385–3396. https://doi.org/10.1256/qj.05.108

Dessler, A. E., & Davis, S. M. (2010). Trends in tropospheric humidity from reanalysis systems. *Journal of Geophysical Research*, *115*, D19127. https://doi.org/10.1029/2010JD014192

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*, 4463–4490. https://doi.org/10.5194/amt-7-4463-2014

Durre, I., Vose, R. S., & Wuertz, D. B. (2006). Overview of the integrated global radiosonde archive. *Journal of Climate*, *19*, 53–68. https://doi.org/10.1175/JCLI3594.1

Durre, I., Vose, R. S., & Wuertz, D. B. (2008). Robust automated quality assurance of radiosonde temperatures. *Journal of Applied Meteorology and Climatology*, *47*(8), 2081–2095. https://doi.org/10.1175/2008jamc1809.1

Durre, I., Yin, X., Vose, R. S., Applequist, S., & Arnfield, J. (2018). Enhancing the data coverage in the integrated global radiosonde archive. *Journal of Atmospheric and Oceanic Technology*, *35*, 1753–1770. https://doi.org/10.1175/JTECH-D-17-0223.1

Fassò, A., Finazzi, F., & Madonna, F. (2018). Statistical issues in radiosonde observation of atmospheric temperature and humidity profiles. *Statistics & Probability Letters*, *136*, 97–100. https://doi.org/10.1016/j.spl.2018.02.027

Ferreira, A. P., Nieto, R., & Gimeno, L. (2019). Completeness of radiosonde humidity observations based on the integrated global radiosonde archive. *Earth System Science Data*, *11*, 603–627. https://doi.org/10.5194/essd-11-603-2019

Finazzi, F., Fassò, A., Madonna, F., Negri, I., Sun, B., & Rosoldi, M. (2019). Statistical harmonization and uncertainty assessment in the comparison of satellite and radiosonde climate variables: Harmonization of satellite and radiosonde climate variables. *Environmetrics*, *30*, e2528. https://doi.org/10.1002/env.2528

Free, M., Angle, J. K., Durre, I., Lanzante, J., Peterson, T. C., & Seidel, D. J. (2004). Using first differences to reduce inhomogeneity in radiosonde temperature datasets. *Journal of Climate*, *17*(21), 4171–4179. https://doi.org/10.1175/jcli3198.1

Gelaro, R., McCarty, W., Suárez, M. J., Todling, R., Molod, A., Takacs, L., et al. (2017). The modern-era retrospective analysis for research and applications, version 2 (MERRA-2). *Journal of Climate*, *30*, 5419–5454. https://doi.org/10.1175/JCLI-D-16-0758.1

Haimberger, L. (2005). *Homogenization of radiosonde temperature time series using ERA-40 analysis feedback information. ERA-40 Project Report Series*, No. 23, ECMWF. Retrieved from https://www.ecmwf.int/node/9738

Haimberger, L., Tavolato, C., & Sperka, S. (2012). Homogenization of the global radiosonde temperature dataset through combined comparison with reanalysis background series and neighboring stations. *Journal of Climate*, *25*, 8108–8131. https://doi.org/10.1175/JCLI-D-11-00668.1

Harris, M. F., Finger, F. G., & Teweles, S. (1962). Diurnal variation of wind, pressure, and temperature in the troposphere and stratosphere over the azores. *Journal of the Atmospheric Sciences*, *19*(2), 136–149. https://doi.org/10.1175/1520-0469(1962)019<0136:dvowpa>2.0.co;2

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., et al. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, *146*, 1999–2049. https://doi.org/10.1002/qj.3803

Hu, S., & Fedorov, A. V. (2017). The extreme El Niño of 2015-2016 and the end of global warming hiatus: Global warming HIATUS and 2015 EL NIÑO. *Geophysical Research Letters*, *44*, 3816–3824. https://doi.org/10.1002/2017GL072908

Iarlori, M., Madonna, F., Rizi, V., Trickl, T., & Amodeo, A. (2015). Effective resolution concepts for lidar observations. *Atmospheric Measurement Techniques*, *8*, 5157–5176. https://doi.org/10.5194/amt-8-5157-2015

Ingleby, B. (2017). *An assessment of different radiosonde types 2015/2016, technical memorandum* (Vol. 807). ECMWF Research Department. https://doi.org/10.21957/0nje0wpsa

JCGM100. (2008). *Evaluation of measurement data—Guide to the expression of uncertainty in measurement*. Retrieved from https://www.iso.org/sites/JCGM/GUM/JCGM100/C045315e-html/C045315e.html?csnumber=50461

Kawatani, Y., Hamilton, K., Miyazaki, K., Fujiwara, M., & Anstey, J. A. (2016). Representation of the tropical stratospheric zonal wind in global atmospheric reanalyses. *Atmospheric Chemistry and Physics*, *16*, 6681–6699. https://doi.org/10.5194/acp-16-6681-2016

Key, J. R., & Schweiger, A. J. (1998). Tools for atmospheric radiative transfer: Streamer and FluxNet. *Computers & Geosciences*, *24*, 443–451. https://doi.org/10.1016/S0098-3004(97)00130-1

Kivinen, S., Rasmus, S., Jylhä, K., & Laapas, M. (2017). Long-term climate trends and extreme events in northern Fennoscandia (1914–2013). *Climate*, *5*, 16. https://doi.org/10.3390/cli5010016

Kobayashi, E., Hoshino, S., Iwabuchi, M., Sugidachi, T., Shimizu, K., & Fujiwara, M. (2019). Comparison of the GRUAN data products for Meisei RS-11G and Vaisala RS92-SGP radiosondes at Tateno (36.06°N, 140.13°E), Japan. *Atmospheric Measurement Techniques*, *12*, 3039–3065. https://doi.org/10.5194/amt-12-3039-2019

Kobayashi, S., Ota, Y., Harada, Y., Ebita, A., Moriya, M., Onoda, H., et al. (2015). The JRA-55 reanalysis: General specifications and basic characteristics. *Journal of the Meteorological Society of Japan*, *93*(1), 5–48. https://doi.org/10.2151/jmsj.2015-001

Loew, A., Bell, W., Brocca, L., Bulgin, C. E., Burdanowitz, J., Calbet, X., et al. (2017). Validation practices for satellite-based Earth observation data across communities: EO validation. *Review of Geophysics*, *55*, 779–817. https://doi.org/10.1002/2017RG000562

Madonna, F. (2020). Can reference radiosounding measurements be used to improve historical time series? *Il Nuovo Cimento C*, *43*, 1–10. https://doi.org/10.1393/ncc/i2020-20121-5c

Madonna, F., Kivi, R., Dupont, J.-C., Ingleby, B., Fujiwara, M., Romanens, G., & , et al. (2020). Use of automatic radiosonde launchers to measure temperature and humidity profiles from the GRUAN perspective. *Atmospheric Measurement Techniques*, *13*, 3621–3649. https://doi.org/10.5194/amt-13-3621-2020

Madonna, F., Tramutola, E., Sy, S., Serva, F., Proto, M., Rosoldi, M., et al. (2020). Radiosounding HARMonization (RHARM): A new homogenized dataset of radiosounding temperature, humidity and wind profiles with uncertainty. *Data, Algorithms, and Models*. https://doi.org/10.5194/essd-2020-183

McCarthy, M. P., Thorne, P. W., & Titchner, H. A. (2009). An analysis of tropospheric humidity trends from radiosondes. *Journal of Climate*, *22*, 5820–5838. https://doi.org/10.1175/2009JCLI2879.1

McCarthy, M. P., Titchner, H. A., Thorne, P. W., Tett, S. F. B., Haimberger, L., & Parker, D. E. (2008). Assessing bias and uncertainty in the HadAT-adjusted radiosonde climate record. *Journal of Climate*, *21*, 817–832. https://doi.org/10.1175/2007JCLI1733.1

Merchant, C. J., Paul, F., Popp, T., Ablain, M., Bontemps, S., Defourny, P., et al. (2017). Uncertainty information in climate data records from Earth observation. *Earth System Science Data*, *9*, 511–527. https://doi.org/10.5194/essd-9-511-2017

Miloshevich, L. M., Paukkunen, A., Vömel, H., & Oltmans, S. J. (2004). 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

Nash, J., Oakley, T., Vömel, H., & Wei, L. I. (2011). *WMO Intercomparison of High Quality Radiosonde Systems (12 July – 3 August 2010; Yangjiang, China), WMO/TD-No. 1580; IOM Report-No. 107*. https://library.wmo.int/index.php?lvl=notice_display&id=15531#.YdVW5y1aZQI

Nash, J., Smout, R., Oakley, T., Pathack, B., & Kurnosenko, S. (2006). *WMO intercomparison of radiosonde systems, Vacoas, Mauritius, 2–25 February 2005*. Tech. Rep. WMO, WMO/TD-No. 1303, Report No. 83. https://library.wmo.int/doc_num.php?explnum_id=9312

Peterson, T. C., & Vose, R. S. (1997). An overview of the Global Historical Climatology Network temperature data base. *Bulletin of the American Meteorological Society*, *78*(12), 2837–2850. https://doi.org/10.1175/1520-0477(1997)078<2837:aootgh>2.0.co;2

Ramella-Pralungo, L., & Haimberger, L. (2014). A global radiosonde and tracked balloon archive on 16 pressure levels (GRASP) back to 1905 – Part 2: Homogeneity adjustments for PILOT and radiosonde wind data. *Earth System Science Data Discussions*, *6*, 297–316. https://doi.org/10.5194/essdd-7-335-2014

Ramella Pralungo, L., Haimberger, L., Stickler, A., & Brönnimann, S. (2014). A global radiosonde and tracked balloon archive on 16 pressure levels (GRASP) back to 1905 – Part 1: Merging and interpolation to 00:00 and 12:00 GMT. *Earth System Science Data*, *6*, 185–200. https://doi.org/10.5194/essd-6-185-2014

Rhoades, D. A., & Salinger, M. J. (1993). Adjustment of temperature and rainfall records for site changes. *International Journal of Climatology*, *13*, 399–913. https://doi.org/10.1002/joc.3370130807

Sherwood, S. C. (2007). Simultaneous detection of climate change and observing biases in a network with incomplete sampling. *Journal of Climate*, *20*, 4047–4062. https://doi.org/10.1175/jcli4215.1

Sherwood, S. C., Meyer, C. L., Allen, R. J., & Titchner, H. A. (2008). Robust tropospheric warming revealed by iteratively homogenized radiosonde data. *Journal of Climate*, *21*(20), 5336–5352. https://doi.org/10.1175/2008jcli2320.1

Sherwood, S. C., & Nishant, N. (2015). Atmospheric changes through 2012 as shown by iteratively homogenized radiosonde temperature and wind data (IUKv2). *Environmental Research Letters*, *10*, 054007. https://doi.org/10.1088/1748-9326/10/5/054007

Stone, M. (1974). Cross-validatory choice and assessment of statistical predictions. *Journal of the Royal Statistical Society: Series B*, *36*, 111–133. https://doi.org/10.1111/j.2517-6161.1974.tb00994.x

Stoumbos, Z. G., & Reynolds, M. R. (2004). The robustness and performance of CUSUM control charts based on the double-exponential and normal distributions. In H.-J. Lenz, & P.-T. Wilrich (Eds.), *Frontiers in statistical quality control 7* (pp. 79–100). Physica-Verlag HD. https://doi.org/10.1007/978-3-7908-2674-6_6

Sy, S., Madonna, F., Rosoldi, M., Tramutola, E., Gagliardi, S., Proto, M., & Pappalardo, G. (2021). Sensitivity of trends to estimation methods and quantification of subsampling effects in global radiosounding temperature and humidity time series. *International Journal of Climatology*, *41*. https://doi.org/10.1002/joc.6827

Thorne, P. W., Brohan, P., Titchner, H. A., McCarthy, M. P., Sherwood, S. C., Peterson, T. C., et al. (2011). A quantification of uncertainties in historical tropical tropospheric temperature trends from radiosondes. *Journal of Geophysical Research: Atmospheres*, *116*, D12116. https://doi.org/10.1029/2010JD015487

Thorne, P. W., Christy, J. R., & Mears, C. A. (2005). Uncertainties in climate trends: Lessons from upper-air temperature records. *Bulletin of the American Meteorological Society*, *86*, 1437–1442. https://doi.org/10.1175/bams-86-10-1437

Thorne, P. W., Madonna, F., Schulz, J., Oakley, T., Ingleby, B., Rosoldi, M., et al. (2017). Making better sense of the mosaic of environmental measurement networks: A system-of-systems approach and quantitative assessment. *Geoscientific Instrumentation, Methods and Data Systems*, *6*, 453–472. https://doi.org/10.5194/gi-6-453-2017

Thorne, P. W., Parker, D. E., Tett, S. F. B., Jones, P. D., McCarthy, M., Coleman, H., & Brohan, P. (2005). Revisiting radiosonde upper-air temperatures from 1958 to 2002. *Journal of Geophysical Research*, *110*, D18105. https://doi.org/10.1029/2004JD005753

von Rohden, C., Sommer, M., Naebert, T., Motuz, V., & Dirksen, R. J. (2021). Laboratory characterisation of the radiation temperature error of radiosondes and its application to the GRUAN data processing for the Vaisala RS41. *Atmospheric Measurement Techniques Discussions*. https://doi.org/10.5194/amt-2021-187

Wang, J., Zhang, L., Dai, A., Immler, F., Sommer, M., & Vömel, H. (2013). 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

Weatherhead, E. C., Bodeker, G. E., Fassò, A., Chang, K.-L., Lazo, J. K., Clack, C. T. M., et al. (2017). Spatial coverage of monitoring networks: A climate observing system simulation experiment. *Journal of Applied Meteorology and Climatology*, *56*, 3211–3228. https://doi.org/10.1175/JAMC-D-17-0040.1

Woodall, W. H., & Adams, B. M. (1993). The statistical design of cusum charts. *Quality Engineering*, *5*(4), 559–570. https://doi.org/10.1080/08982119308918998

Zhou, C., Wang, J., Dai, A., & Thorne, P. W. (2021). A new approach to homogenize global subdaily radiosonde temperature data from 1958 to 2018. *Journal of Climate*, *34*, 1163–1183. https://doi.org/10.1175/JCLI-D-20-0352.1