The new Radiosounding HARMonization (RHARM) dataset of homogenized radiosounding temperature, humidity and wind profiles with uncertainties. Part I: dataset description and characterisation.

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]Irish Climate Analysis and Research Units, Department of Geography, Maynooth University, Maynooth, Ireland

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. Moreover, measurement uncertainties should be properly quantified and validated. In this work we present a novel approach, named RHARM (Radiosounding HARMonization), to provide a homogenized dataset of temperature, humidity and wind profiles along with an estimation of the measurement uncertainties for 700 radiosounding stations globally. The RHARM method has been used to adjust twice daily (0000 and 1200 UTC) radiosonde data holdings at 16 pressure levels from 1000 to 10 hPa from 1978 to the present from the Integrated Global Radiosonde Archive (IGRA). Relative humidity (RH) data are limited to 250 hPa. The applied adjustments are interpolated to all reported significant levels. RHARM is the first dataset to provide homogenized time series of temperature, relative humidity and wind profiles alongside an estimation of the observational uncertainty for each observation at each pressure level.

The comparison of RHARM and unadjusted profiles highlights a median temperature warmer by 0.6 K in the boreal hemisphere, while in the tropics RHARM is cooler by 0.1 K. For RH, the difference is -2.1%, while in the tropics it is reduced to 0.3%. For wind speed, adjustments largely improve the data homogeneity locally. Analysis of decadal trends for temperature, RH and winds highlights increased the geographical coherency of trends.

In a companion paper, the performances of the RHARM dataset are assessed through comparison with the reanalysis, satellite and other homogenized radiosonde datasets.

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:

- 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.);
- 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;
- 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 fulfil 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) are either missing or retained solely by individual station PIs (if at all) and not routinely shared or stored in their 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 GNSS-RO (Global Navigation Satellite System - Radio Occultation) measurements which have proven valuable for data assimilation purposes (Bauer et al. 2013). Radiosounding measurements are the only available data source continuously available to study climate variability and change in the troposphere and lowermost stratosphere since the mid-20th century. They also constitute a valuable source of information for satellite cal/val activities (Calbet et al., 2016, Loew et al., 2017, Finazzi et al., 2019). 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. Kobayashi et al., 2015; Gelaro et al., 2017).

Quality and biases of radiosounding observations strongly varies with sensor type, altitude level, and through time. Many previous works described the adjustment of historical radiosounding temperature measurements to construct CDRs (e.g. Free et al. 2004; Thorne et al., 2005a; McCarthy et al., 2008; Sherwood et al. 2008; Dai et al., 2011; Haimberger et al., 2012, Zhou et al., 2020). These works have used a broad range of approaches enabling an exploration of structural uncertainty (Thorne et al., 2005b). Several products additionally include ensemble approaches to explore parametric uncertainty (Haimberger et al., 2012; Thorne et al., 2011, Sherwood and Nishant, 2015). Application of innovative statistical approaches has been recently proposed for the production of future datasets (Fassò et al., 2018).

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 (e.g. Nash et al., 2006, Nash et al., 2011) and providing a thorough periodical assessment of inter-sensor differences. 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.

To address the need of providing homogeneous and fully traceable upper-air measurements with quantified uncertainties, the Global Climate Observing System (GCOS) Reference Upper-Air Network (GRUAN) was established in 2006 (Bodeker et al., 2018). 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) 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. Such reference-observing networks can provide metrologically traceable observations, with quantified uncertainty, at a small number of stations. Whereas baseline-observing networks provide long-term records that are capable of catching regional, hemispheric and global-scale features, though they lack absolute traceability (Thorne et al., 2017). 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 (JCGM100, 2008) compromising the quality of operationally processed radiosoundings.

Under the Copernicus Climate Change Service (C3S) activities, we have designed and applied a novel algorithm for homogenizing historical radiosounding data records available since 1978 (earlier records are not assessed due to various reasons and mainly the more heterogeneous data availability at mandatory levels and the use of more heterogeneous instrumentation before 1978). This new approach discussed herein is named RHARM (Radiosounding HARMonization), and it is a hybrid method based on two main steps:

a. Adjustment of systematic effects and quantification of uncertainties by adjusting the 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 dataset (hereinafter ID2010, Nash et al. 2011);

b. Identification of change-points in the earlier portions of the time series (before 2004) and adjustment of non-climatic (systematic) 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 dataset, via comparisons with the 'raw' data and GRUAN, for a subset of 700 radiosounding stations available from the Integrated Global Radiosonde Archive (IGRA - Durre et al., 2006; Durre et al., 2018). In a companion paper, the performance of the RHARM dataset is evaluated through comparisons with the ECMWF reanalysis products, pre-existing homogenized datasets and satellite observations.

RHARM provides another option within the limited number of existing datasets, such as:

- homogenized radiosounding temperature measurements, e.g. 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 et al., 2005), Iterative Universal Kriging v2 (IUKv2) by University of New South Wales (Sherwood and Nishant et al., 2015), the State University of New York Albany dataset (Zhou et al., 2020);
- homogenized radiosounding humidity measurements, e.g. 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
- homogenized radiosounding wind datasets, e.g. IUKv2 and GRASPA (Ramella-Pralungo et al., 2014a,b).

Distinct from previous efforts, RHARM is the first dataset 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. 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.

The remainder of this paper is organized as follows. In section 2, the data sources used in the paper 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 adjustment of the time series for the period before 2004. Section 3 is corroborated by the information provided in Appendix A. In section 4, statistics of adjustments applied by RHARM in comparison with IGRA and GRUAN data are discussed. 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.

2. Data sources used

The RHARM approach is applied to the IGRA database which is the most comprehensive global collection of original ‘raw’ historical and near-real-time radiosonde and pilot balloon observations. 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. RHARM is applied to a subset of 700 radiosounding stations and radiosoundings from ships. We selected only the records with documented metadata (i.e. including the radiosonde code according to WMO table 3685, describing the radiosonde type) since 2000 (for most of the stations) and for fewer stations since 1978. For these stations, 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).

The IGRA data v2 are the result of improved quality assurance procedures developed for the IGRA data v1 (Durre et al. 2006; Durre et al. 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 dataset thus inherits the IGRA quality assurance procedures, and additional quality checks are then applied. We perform tests 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 records, depending on the availability of the required metadata (Durre et al. 2008; Ferreira et al., 2019). For these stations, a quality-enhanced dataset with a sufficient number of radiosoundings available since 2004 to present can be provided directly post-processing the profiles to account for several instrumental effects (e.g. the well-known solar radiation dry-bias). 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. For the selected 700 IGRA stations, only measurements with the highest data quality according to the IGRA data quality system at each pressure level have been processed with the RHARM algorithm.

Figure 1 shows the locations of the stations processed herein and number of launches available. In addition, the 1156 IGRA stations reporting data since 1978 to present are also shown. The coverage of RHARM is reasonably homogeneous, except for Siberia where a smaller number of launches is available. This is due to the limited information available on the main radiosonde type used in the region since 2004 on (AVZ), which cannot be adjusted using RHARM to achieve the same quality as for other radiosonde types. The station density in Canada, North East 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 low number of measurements available in the southern hemisphere (SH), although the quantity of measurements alone cannot address the value of the dataset for a specific study without considering representativeness (Weatherhead et al., 2017).

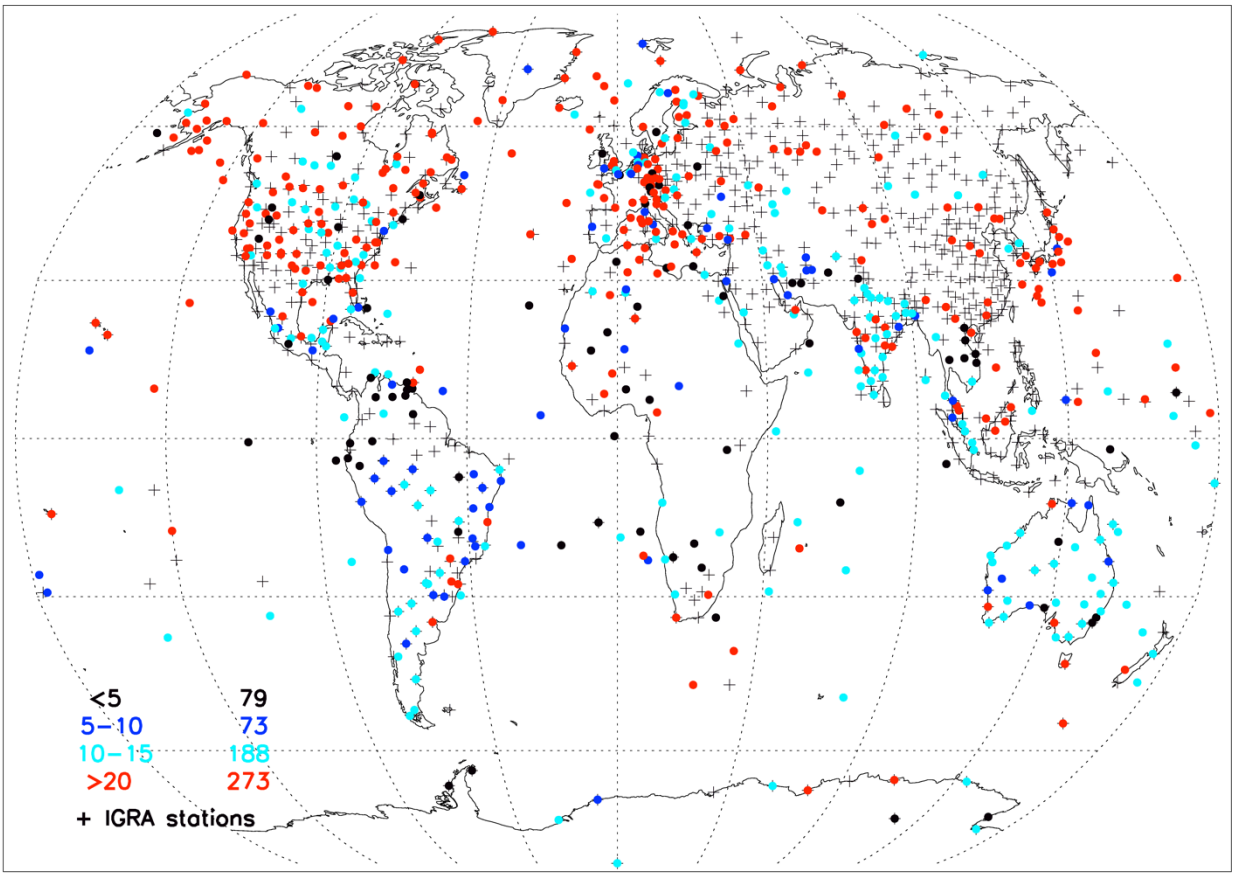

Figure 1: Global distribution and quantity of RHARM homogenized profiles. The scale in the bottom left corner denotes the available radiosoundings at each station (in millions of ascents) from 1978 to present. The + symbol indicates IGRA stations (1156) reporting data since 1978 to present (last access to IGRA 31-12-2020).

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

Table 1: Number and percentage of launches in different latitude bands for the stations shown in Figure 1.

3. Methodology

The RHARM homogenization of global radiosounding temperature, humidity and wind profiles is applied to per-ascent (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, 1000 hPa), because the frequency of reports from the stations is typically per 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 (Miloshevic et al., 2004). Profiles are adjusted at these mandatory pressure levels. The applied adjustments are then interpolated to the reported significant levels. Uncertainties are estimated for each processing step and propagated to estimate the total uncertainty.

For the sake of clarity, the RHARM-adjusted time series since 2004 (with starting time station-dependent) obtained by post-processing of each single radiosounding profile using a GRUAN-like algorithm is labelled PPTS (Post-Processed Time Series). The PPTS is then used as a constraint for adjusting the preceding radiosounding time series, hereinafter HST (Homogenized Time Series). The concatenation of HTS and PPTS records provides the entire time series for each station, and only those stations satisfying the requirements for the production of a PPTS are considered for the HTS calculations. An overall scheme of the RHARM approach is shown in Figure 2.

The PPTS produced as step A1 for each station is merged with the prior part of the record (step A2). The resulting time series (step B) are firstly 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. 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 31-12-2020), using as inputs each radiosonde launching time and the corresponding station geographical coordinates. The small number of radiosonde launches available at other synoptic hours have not been considered in the current RHARM data version. Such a step is critical mainly for temperature and humidity where radiation-heating effects can have substantive impacts on instrument performance (Miloshevic et al., 2004; Wang et al., 2013; Dirksen et al., 2014). It is less likely that there will be effects on the uncertainties for winds either using GNSS or precursor radar tracking techniques, but the same separation is made for 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).

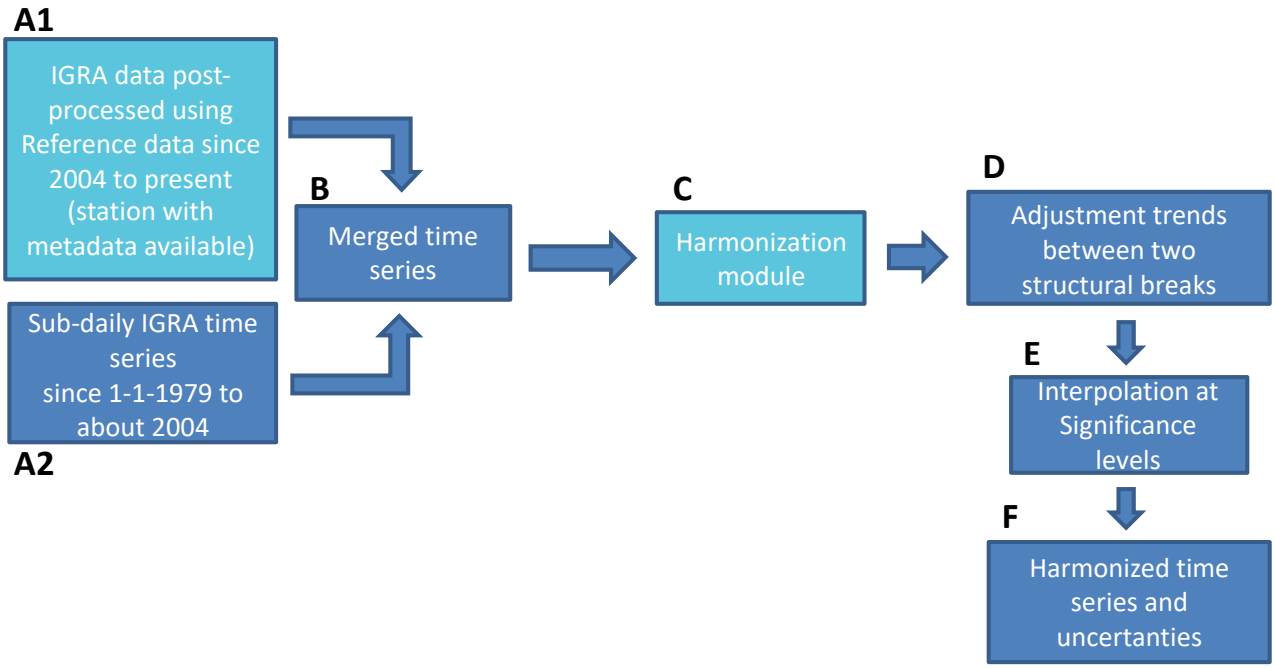

Figure 2: Schematic diagram describing the steps of the RHARM approach.

In section 3.1 and 3.2, the approach applied to obtain the PPTS (Step A1 in Figure 2) is outlined. The remaining sections describe the adjustment of the HTS (Step A2) and subsequent adjustments to the significant levels in the radiosounding profiles (Steps B-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 Data Processing (GDP). This is performed in two steps:

1. first, the radiation correction, $\Delta T_{VAISALA}$, applied by the manufacturer to the temperature profiles is removed;
2. second, a GRUAN-like radiation correction, ΔT_{GRUAN} is applied using the values of the actinic flux modelled with the Streamer RTM (Key and Schweiger, 1998) following the approach documented in Dirksen et al (2014). Where GRUAN-like corrections cannot be applied, the manufacturer correction is left unchanged.

$\Delta T_{VAISALA}$ 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, ΔT_{GRUAN}, is defined as:

$$\Delta T_{GRUAN}(I_a, p, v) = ax^b \quad \text{[Eq.1]}$$

$$x = \frac{I_a}{pu_a} \quad \text{[Eq.2]}$$

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 (taken from https://code.arm.gov/vap/mfrsrod1barnmich/blob/ed71a3666e8e1781ed8d753e859b284f3b7dcc2e/src/zensun.pro, last access on 31-12-2020); p is the pressure level; and u_a is the ascent speed in m s^{-1}. The ascent speed 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 et al., 2020b). The coefficients a and b in Eq.1 are fit parameters arising from laboratory experiments (Dirksen et al., 2014) yielding a = 0.18(±0.03) and b = 0.55(±0.06).

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

$$\Delta T = \frac{(\Delta T_{GRUAN} + \Delta T_{VAISALA})}{2} \quad \text{[Eq.3]}$$

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

$$\Delta T_{RHARM,RS92} = \Delta T_{VAISALA} - \Delta T + \Delta T_r \text{ [Eq.4]}$$

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 robust Z-score method. ΔT_r is added to both night and daytime profiles. If the value of Ia in equation 2 is equal to zero (i.e. $\Delta T{=}0$), the manufacturer radiation correction applied to IGRA profiles is not modified and Eq.4 reduces to $\Delta T_{RHARM,RS92} = \Delta T_r$. Eq. 4 removes the solar radiation correction applied by the manufacturer and adjusts the data using the GRUAN correction plus an additional term reducing, on average, the gap with the GDP.

GRUAN code	Station name and country	Latitude	Longitude	Altitude	WMO index
CAB	Cabauw, Netherlands	51.97°	4.92°	1 m	06260
LIN	Lindenberg, Germany	52.21°	14.12°	98 m	10393
NYA	Ny-Ålesund, Norway	78.92°	11.92°	5 m	01004
SGP	Lamont, OK, USA	36.60°	-97.49°	320 m	74646
SOD	Sodankylä, Finland	67.37°	26.63°	179 m	02836
TAT	Tateno, Japan	36.06°	140.13°	25 m	47646

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.

The standard uncertainty (k=1) on $T_{RHARM,RS92}$, $\varepsilon\left(T_{RHARM,RS92}\right)$, is calculated according to the following equation:

$$\varepsilon\left(T_{RHARM,RS92}\right) = \left(\sum_i \quad \varepsilon^i_{systematic}(\Delta T)^2 + \varepsilon_R(\Delta T)^{\,2}\right)^{\frac{1}{2}} = \left(\varepsilon_{c,I_a}(\Delta T)^2 + \varepsilon_{c,R_c}(\Delta T)^2 + \varepsilon_{vent}(\Delta T)^2 + \varepsilon_r(\Delta T)^2 + \varepsilon_R(\Delta T)^2\right)^{\frac{1}{2}} \text{ [Eq.5]}$$

In Eq. 5, $\varepsilon^i_{systematic}(\Delta T)$ indicates a systematic uncertainty contribution;$\varepsilon_{c,I_a}(\Delta T)$ is the uncertainty due to the estimation of the solar actinic flux; $\varepsilon_{c,R_c}(\Delta T)$ is the uncertainty due to parameters estimated in the radiation correction model reported in Eq. 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). ε_{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 GDP approach (Dirksen et al., 2014). When the radiation correction of the manufacturer is left unchanged, $\varepsilon\left(T_{RHARM,RS92}\right)$ 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, $RH_{RHARM,RS92}$, 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:

$$RH_{RHARM,RS92} = cf\ RH_{IGRA,RS92}\left(\frac{p_s(T_{RHARM,RS92}+g\Delta T_{RHARM,RS92})}{p_s(T_{RHARM,RS92})}\right) \text{ [Eq.6]}$$

where cf is a scalar factor accounting for the temperature dependency of the sensor calibration estimated at night by a comparison with GRUAN measurements (see 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., 2014). The factor *cf* 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 (RH < 20%) for U.S. stations which were set to a dewpoint depression of 30°C (or RH = 19%), have been properly managed (McCarthy et al., 2009). For the sake of clarity, 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.

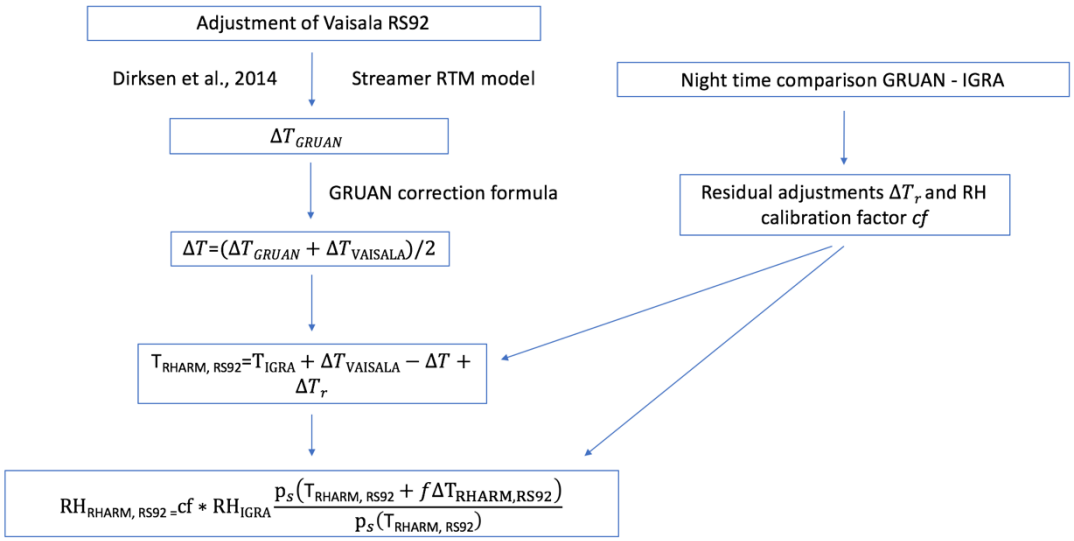

Figure 3: Flow diagram summarizing the post-processing steps of the RHARM algorithm to adjust temperature and relative humidity profiles measured by the RS92 sondes from 2004. In the diagram, cf is a calibration factor, p_s is the saturation vapor pressure, g is a factor determined experimentally to 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 GRUAN adjustments, 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.

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

$$\varepsilon\left(RH_{RHARM,RS92}\right) = \left(\varepsilon_{RC_T}(\Delta RH)^2 + \varepsilon_{RC_f}(\Delta RH)^2 + \varepsilon_{cf}(\Delta RH)^2 + \varepsilon_R(\Delta RH)^2\right)^{\frac{1}{2}} \text{ [Eq.7]}$$

where $\varepsilon_{RC_T}(\Delta RH)$ is the uncertainty of dry bias correction; $\varepsilon_{RC_f}(\Delta RH)$ is the uncertainty of the radiation sensitivity factor *g* in Eq. 5; ε_{cf} is the uncertainty due to calibration factor *cf*; ε_R is an additional random uncertainty of 2% RH. In analogy with temperature, when the radiation correction of the manufacturer is left unchanged, $\varepsilon\left(RH_{RHARM,RS92}\right)$ 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, 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 dataset.

The GDP 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:

$$\varepsilon(\phi) = \frac{180}{\pi} \frac{\sqrt{\delta_u^2 + \delta_v^2}}{\left(1 + \left(\frac{u}{v}\right)^2\right)|v|} \text{ [Eq. 8]}$$

and the uncertainty of the wind speed w by

$$\varepsilon(w) = \sqrt{\frac{(u\delta_u)^2 + (v\delta_v)^2}{u^2 + v^2}} \text{ [Eq. 9]}$$

Typical values are between 0.4 and 1 ms^{-1} for $\varepsilon(w)$ and about 1° for $\varepsilon(\phi)$. In the case of negligible wind, when u and v approach 0, the value of $\varepsilon(\phi)$ becomes very large. For such cases, the absolute value of $\varepsilon(\phi)$ 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:

$$u(p,t) = w(p,t)\, sin\left(\frac{\pi}{180}\phi(p,t)\right) \text{ [Eq. 10]}$$

$$v(p,t) = w(p,t)\, cos\left(\frac{\pi}{180}\phi(p,t)\right) \text{ [Eq. 11]}$$

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. Eqs. 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 1. The approach is the same as for temperature (Eq.4), although it is reduced to $\Delta u_{RHARM,RS92} = \Delta u_r$ and to $\Delta v_{RHARM,RS92} = \Delta v_r$, for each of the wind vectorial components. The standard deviation of the $\Delta u_{RHARM,RS92}$ and $\Delta v_{RHARM,RS92}$ are then used as the estimation of the combined standard uncertainties, which are expressed as $\varepsilon\left(\Delta u_{RHARM,RS92}\right) = (\varepsilon_r(\Delta u)^2 + \varepsilon_R(\Delta u)^2)^{\frac{1}{2}}$ and $\varepsilon\left(\Delta v_{RHARM,RS92}\right) = (\varepsilon_r(\Delta v)^2 + \varepsilon_R(\Delta v)^2)^{\frac{1}{2}}$. 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 23/04/2021).

This adjustment can only partly reconcile the difference between GDP 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 the final step of RHARM, the adjustments applied to temperature, humidity and wind profiles at the mandatory levels as well as the corresponding uncertainties are interpolated to the significant levels available in the IGRA files, which vary from profile-to-profile and are used to mark key geophysical points in the profile, such as temperature or humidity profile inflections. 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. This interpolation uncertainty has been added to the final uncertainty budget (for T, σ=0.25 K, for RH, σ=0.5 %, for both u and v, σ=0.05 ms^{-1}).

At present, there are only two GRUAN data products, 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.

3.2 Adjustment of other radiosonde types

For non-Vaisala radiosonde types, the adjustment estimation requires the adoption of a different approach due to the unavailability of GRUAN reference products. To harmonize these records, RHARM makes primary recourse to the ID2010, from which estimations of the relative performance of operational radiosondes in 2010 were evaluated through a joint effort between the scientific community and the radiosonde manufacturers. ID2010 allows us to assess 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 vis-a-vis aspects such as ventilation, thermal effects and the magnitude and periodicity of pendulum motion effects.

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

Table 3: List of the operational radiosondes involved in the 2010 WMO/CIMO radiosonde intercomparison used to calculate the RHARM adjustments. 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.

Among the radiosonde types involved in the intercomparison, only those routinely employed at a sufficient number of 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). The Vaisala RS92 sondes available in ID2010 have been adjusted using the RHARM algorithm described in the prior sub-section. The list of the selected radiosonde types is given in Table 3.

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. Standard deviations have been recalculated accordingly to consider the additional contribution of the Graw radiosonde uncertainties and the two steps required to quantify the comparison. The mean difference over *N* ascents between RS92 temperature profiles and the profiles measured by each of the sondes listed in Table 3 (hereinafter named as "NORS92") has been quantified as (Figure 4):

$$\Delta T_{NORS92} = \frac{1}{N}\sum_{i=1}^{N} \quad T_i^{NORS92} - T_i^{RHARM,\,RS92} \text{ [Eq. 12]},$$

and the standard deviation $\sigma_{T_{NORS92}}$ is calculated from the spread of pairwise estimates of ΔT_{NORS92} arising from the RHS term of equation 12. $\sigma_{T_{NORS92}} = \sqrt{{\sigma_{T_{NORS92}}}^2 + \varepsilon\left(T_{RHARM,RS92}\right)^2}$ is used as the best estimate of the uncertainty for ΔT_{NORS92} . If the Graw radiosonde is considered as the link with the Vaisala RS92 (with *M* ascents in common), Eq.12 becomes:

$$\Delta T_{NORS92} = \left(\frac{1}{N}\sum_{i=1}^{N} \quad T_i^{NORS92} - T_i^{GRAW}\right) - \left(\frac{1}{M}\sum_{j=1}^{M} \quad T_j^{GRAW} - T_j^{RHARM,\,RS92}\right) \text{ [Eq. 13]},$$

and the standard deviation $\sigma_{T_{NORS92}}$ is derived assuming independence of the two components so the individual estimates are combined in quadrature.

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 also https://idlastro.gsfc.nasa.gov/ftp/pro/robust/resistant_mean.pro, last access on 31-21-2020). This allows us to ensure that the most typical differences between any two radiosonde types are caught in the calculated differences, enabling their application as an average adjustment on a wide range of radiosondes. Eqs. 12 and 13, with the related considerations, are applied also to wind profiles.

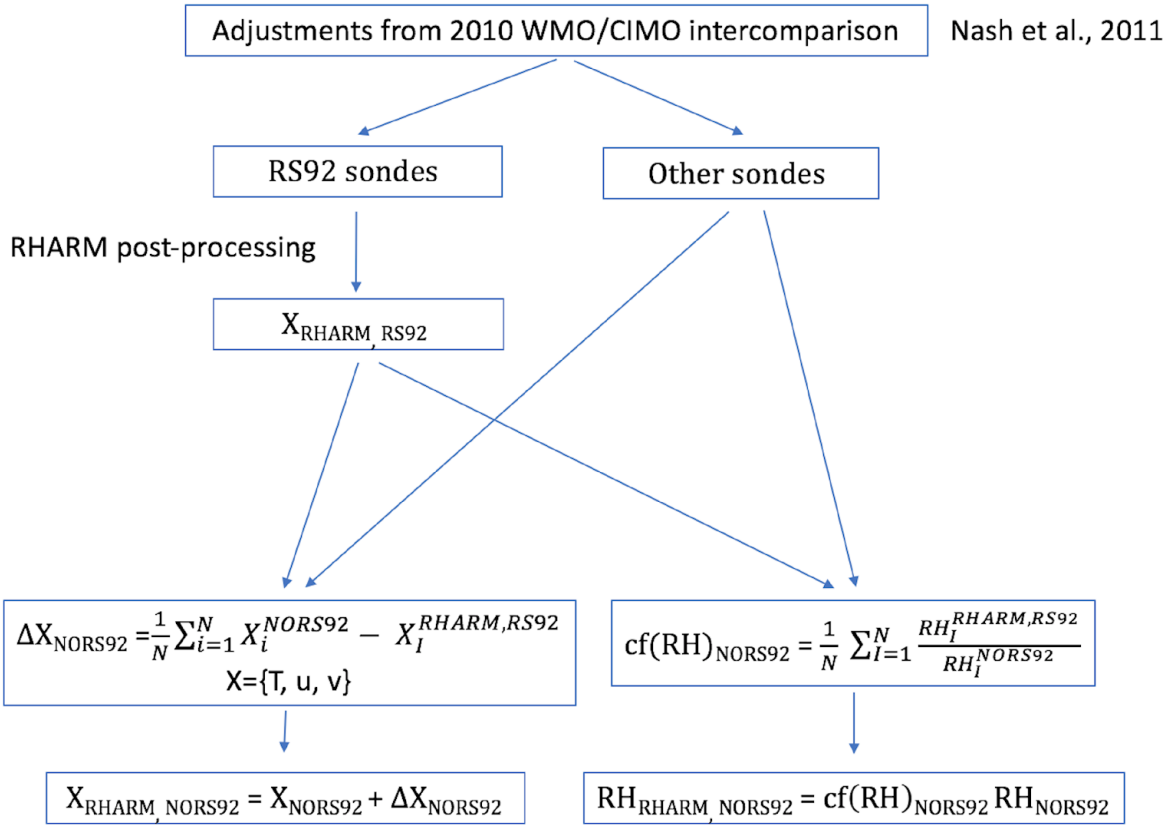

Figure 4: Flow diagram summarizing the post-processing steps of the RHARM algorithm to adjust the temperature and relative humidity profiles measured for all radiosonde types other than RS92 reported in Table 3 in the period from 2004 onward. In the diagram, "X" stands for T, u or v. The subscript RHARM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

For relative humidity, also in order to be consistent with the RHARM post-processing of RS92 sondes, instead of Eq. 12 the following is used:

$$cf(RH)_{NORS92} = \frac{1}{N}\sum_{i=1}^{N} \frac{RH_i^{RHARM,RS92}}{RH_i^{NORS92}} \quad \text{[Eq. 14]},$$

where $cf(RH)_{NORS92}$ is a scalar calibration factor to remove systematic effects on the NORS92 radiosondes; the related standard deviation, $\sigma_{cf(RH)_{NORS92}}$, is calculated via error propagation. If the Graw radiosonde is considered as the link with the Vaisala RS92, Eq.14 becomes:

$$cf(RH)_{NORS92} = \frac{1}{N} \sum_{i=1}^{N} \frac{cf(RH)_{GRAW}\, RH_i^{RHARM,GRAW}}{RH_i^{NORS92}}$$ [Eq. 15],

To facilitate the application of the adjustments for all significant pressure levels available in the IGRA dataset, the profiles obtained from the Eqs. 12, 13, 14 and 15, including all the available (mandatory and significant) levels, 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, σ=0.25 K, for RH, σ=0.5 %, for both u and v, σ=0.05 ms^{-1}). All the profiles derived from the ID2010 with the corresponding standard deviations are shown in detail in Appendix A.

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

Radiosonde type	Launches	Percentage
LMS6	29148	1.3
DMF-09 Graw	16736	0.8
VIZ/JinYang	33721	1.5
Taiyuan GTS1-1/GFE(L)	13409	0.6
Nanjing GTS1-2/GFE(L)	17406	0.8
Meteolabor	436	0.0
Meisei	16179	0.7
Beijing Changfeng CF-06	36393	1.7
M10, Modem	121446	5.5
Vaisala RS92/RS41	1893805	85.9
Intermet	26505	1.2
Total	2205183	100

Table 4: Number and percentage of the radiosonde launches available from 2004 and adjusted using the RHARM approach. 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.

Considering that the wind data collected with the radiosonde types reported in Table 3 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 la., 2015). The unavailability of the raw 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.

3.3 Detection of early period breakpoints using the CUSUM method

The detection of breakpoints in the IGRA radiosonde times series is carried out at each single station and is limited to the historical record HTS (i.e. does not affect the post-2004 PPTS at the same station). Time series are used at their full-time resolution (1-2 launches per day on average), i.e. data are not monthly or annually aggregated.

Before homogenizing the time series, a few checks are applied to the data. The IGRA dataset is provided after the application of a comprehensive set of quality control procedures to remove gross errors without removing jumps and other discontinuities caused by changes in instrumentation, observing practice, or station location (Durre et al., 2008 and Durré et al 2018). The IGRA flagging system is exploited by RHARM to filter out data which are considered not quality assured. IGRA quality checks (QC) have been also tested before the production of the RHARM dataset and revealed their robustness in detecting outliers and unphysical values, and this is one of the most important features which gives IGRA an added value compared to other radiosounding data sources for the different users' applications.

In addition to the IGRA QCs applied by IGRA, RHARM preliminarily verifies the physical plausibility of the values reported at each pressure level, i.e. 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°< ϕ <360°.

Once diurnal variability has been accounted for, data are processed through the homogenization module (Step C in Figure 2). After filtering out unphysical values, each time series can be described using an additive model, assuming that the profiles for each ECV arise at each pressure level from a normal distribution:

$$x(p,t) = Tr(p,t) + S(p,t) + B(p,t) + \varepsilon(p,t) \text{ [Eq. 16]}$$

where *x* is the time series of temperature, relative humidity or wind components, *Tr* is the unknown climate trend, *S* is the climate variability, B is an instrumental bias component and $\varepsilon \sim N(0,\sigma^2)$ represent the residuals which depends on the local meteorological variability and on the measurement uncertainties. Although classified as a systematic component, the term *B* is a function of time because it changes for different periods of a time series characterized by the use of different sensors.

Within the RHARM algorithm, the detection of breakpoints in the time series is based on the Cumulative Summing test (CUSUM test). The CUSUM test (Aue and Horváth, 2013) looks for a change in the mean of a stationary time series. It has been already used in the past to determine the homogeneity of a station for different ECVs (Rhoades and Salinger, 1993; Peterson et al., 1997). Within RHARM, the CUSUM test is applied to temperature, humidity and wind components (i.e. u and v) at each pressure level, separately for day and night time data. For the CUSUM test we define two variables:

$$S_i = \sum_{j=i}^{1} x_j^S - \mu - k \text{ [Eq. 17]}$$

$$S_i' = \sum_{j=i}^{1} \mu - k - x_j^S \text{ [Eq. 18]}$$

where x_j^S is representative of time series of the ECV (x) to homogenize at the instant i, while μ is the mean of the process when this is considered "under control" (i.e. with a change of the mean

within an acceptable tolerance). In our case, this corresponds to the mean and standard deviation of the entire time series (i.e. merging HTS and PTTS). *k* is the so-called "allowance" and represents the maximum allowed change for x_i^S. In our setup:

$$k = 0.1\sigma^S(p,t) \text{ [Eq. 19]}$$

k is typically related to the minimum detectable shift δ to the mean, expressed in terms of standard deviation units by the equation $k = 0.5\delta\sigma^S(p,t)$ and σ_x is the standard deviation of the process "under control". Therefore, in our setup $\delta = 0.2$. This choice has been optimised for detecting breakpoints in the time series by investigating manually selected stations where comprehensive metadata were available since 1978 to present (e.g. Lindenberg WMO index=10393, and Sodankyla WMO index=2836). Synthetic time series with artificial systematic effects (i.e step functions) have also been used to tune the CUSUM parameters.

The use of two variables S_i and S_i' allows us to detect both positive and negative changes in the cumulative sum (two-sided CUSUM). In order to identify breaks in the time series, the two following CUSUMs are calculated:

$$S_{hi}(i) = max(0, S_{hi}(i-1) + x_i^S - \mu - k) \text{ [Eq. 20]}$$

$$S_{lo}(i) = \left(0, S_{lo}(i-1) + \mu - k - x_i^S\right) \text{ [Eq. 21]}$$

where $S_{hi}(0)$ and $S_{lo}(0)$ are equal to 0 at the time instant t=0. When either $S_{hi}(i)$ and $S_{lo}(i)$ exceeds a certain threshold value *h*, a break is detected. According to several applications available in the literature, a good threshold value is *h*=4*k* (Woodall and Adams, 1993).

In general, it is possible to design the standard CUSUM test in order to be highly robust to non-normality and it is very effective at detecting shifts of all sizes, even for highly skewed and extremely heavy-tailed process distributions (Stoumbos and Reynolds, 2004). Nevertheless, the CUSUM is typically more efficient in cases of normally distributed variables and stationary phenomena. The non-stationarity of temperature time series, the non-normality of the RH time series, and the large variability of winds over time including a significant number of "extreme" events may affect the CUSUM break detection and increase the number of false positives. For this reason and considering that RHARM works to detect systematic effects generating "step-changes" in the time series, the CUSUM is not applied on the raw time series $x(p,t)$ of Eq.[16], but rather is applied to the non-linear trend component $Tr(p,t)$, estimated by applying a locally weighted smoothing (LOESS) with a smoothing window equal to the 30% of the overall length of the time series (for both night and day). Due to its nature, LOESS enables an efficient propagation into the smoothed times series of any systematic effect present in the original times series, removing the seasonality if applied over an appropriate smoothing window.

Figure 5 shows an example of the breakpoints detected for the night time relative humidity (over liquid water) measured at 300 hPa in Sodankyla from 1978 to present: RHARM approach is able to identify the main documented breakpoints in Sodankyla time series.

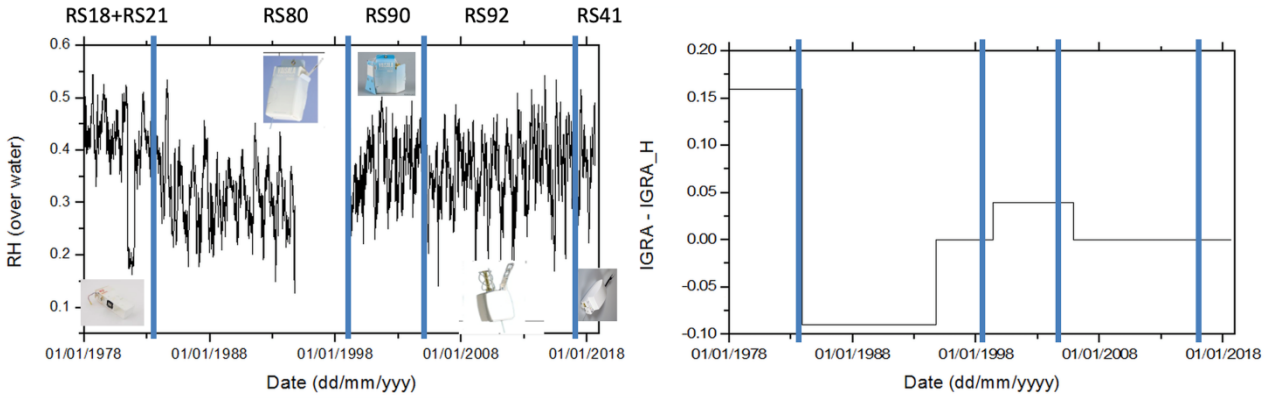

Figure 5: left panel, relative humidity monthly time series for the station of Sondakyla (WMO index=2836, 67.3667N 26.6289E, 179 m asl) as available from IGRA, reporting the types of radiosondes utilized at stations (name and picture of the sonde) in different time periods denoted in between of the blue lines. Right panel, adjustments applied by the RHARM algorithm (IGRA minus RHARM) with blue lines indicating sensor changes over the time.

Breakpoints in the HTS for each station are found only if the following two conditions are met:
- The minimum amount of available data (in years) to estimate adjustments and homogenize the times series at each station (PPTS must have a minimum length of 5 years);
- The minimum amount of data available per year to estimate adjustments (60 measurements per years);

Detection of breakpoints is limited to the time interval of a year, i.e. only if in between of two breakpoints there is a time interval longer than one year. If a breakpoint is found at a temporal distance shorter than 1 years this is skipped and treated with the previous breakpoint as a single break in the time series.

Once the structural breaks are identified in each time series, outliers are removed: for temperature the values exceeding 6 standard deviations of the data distribution in between of two breaks are rejected; for RH, the values exceeding 3 times the interquartile range of the data distribution in between of two breaks are rejected; for wind data, no outliers removal criteria are applied because the tested criteria resulted in filtering also plausible values.

After removal of outliers, at each mandatory pressure level, the corresponding adjustments are quantified and applied (step D in Figure 2). Given a time period Δt between two consecutive breaks, where $\Delta t = t_i - t_j$ ($i > j > 0$), the homogenized value of the generic variable *x* at the pressure levels p in the temporal window Δt, $x^H(p, \Delta t)$, is calculated as:

$$x^H(p, \Delta t) = x(p, \Delta t) - \left(\underline{expTr(p, \Delta t)} - \underline{expT_H(p, \Delta t)} \right) \text{ [Eq. 22]}$$

whereat a certain pressure level p:

- $expTr(p, \Delta t)$ is the exponential trend calculated in the time interval Δt, and
- $expT_H(p, \Delta t')$ is the predicted exponential trend extrapolated assuming the same exponential trend fitted to the homogenized portion of the time series at the step i and time instant t_i. The homogenization algorithm is recursive backward in time from present to the beginning of the time series.

If the two exponential trends, $expTr(p, \Delta t)$ and $expT_H(p, \Delta t')$ have opposite signs, the intercept of $expTr(p, \Delta t)$ is suitably replaced by the intercept of $expT_H(p, \Delta t')$, while the slope of the regression in the time interval Δt is not adjusted. This condition is mainly related to the presence of instrumental calibration drifts in the time series.

The exponential trend calculation is implemented by linearizing the exponential relationship and using a robust linear parametric fitting method, known as LADFIT (Barrodale and Roberts, 1974), which also uses a threshold on the accuracy of the fit residuals.

Exceptions due to stations (about 30) where the PPTS has more than one type of sonde post-processed by RHARM (see Table 4) are handled with an ad-hoc application of the RHARM algorithm.

For the significant pressure levels, their extreme heterogeneity over time (both the number of levels and what constitutes a significant level, as well as their vertical randomness) prevents the application of any sequential approach for the detection of the breaks in the time series. Within RHARM, adjustments at each significant pressure level (p') are calculated as the interpolated values between the two closest mandatory levels, above (p_A) and below (p_B) the considered pressure level (step E in Figure 2).

3.4 Estimation of uncertainties

As for the PPTS record, an uncertainty is attributed to each value of the HTS using the following formula:

$$\varepsilon_H(p,t) = \sqrt{(\varepsilon(X_{PPTS})\,)^2 + \left(\varepsilon(X_{HTS})\right)^2} \quad \text{[Eq. 23]}$$

In Eq. 23 (under the square root, dependencies 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*, $\varepsilon(X_{PPTS})$ is the average uncertainty of the PPTS at the selected station, and $\varepsilon(X_{HTS})$ is estimated using the residuals of each time series with respect to a predictor model, obtained by applying a LOESS smoother:

$$\varepsilon(X_{HTS}) = x_t - q_t \quad t = 1, 2, \ldots\ldots, T \quad \text{[Eq. 24]}$$

where x_t is the measurement for the variable x at the instant t, q_t is the LOESS modelled 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, at each individual station, to match the residuals to the average values of $\varepsilon\left(X_{RHARM,RS92/NORS92}\right)$, 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 PPTS 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.

At each significant pressure level (p'), similarly to the adjustment, the uncertainty is estimated as the interpolated value between the two closest mandatory levels, above (p_A) and below (p_B) the considered pressure level. An additional term, $\varepsilon_{int}(p', t)$, due to the interpolation (for T, σ=0.25 K, for RH, σ=0.5 %, for both u and v, σ=0.05 ms^{-1}, (https://www.vaisala.com/sites/default/files/documents/RS92SGP-Datasheet-B210358EN-F-LOW.pdf, last access 23/04/2021) is added to the interpolated uncertainty values ($\varepsilon_I(p_A, p_B, t)$). For temperature the interpolation is linear, while cubic splines are used for relative humidity and wind components.

For these levels indicated with p', Eq.8 becomes:

$$\varepsilon_H(p', t) = \sqrt{\left(\varepsilon_I(p_A, p_B, t)\right)^2 + (\varepsilon_{int}(p', t)^2)} \text{ [Eq. 25]}$$

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:

$$W = \sqrt{u^2 + v^2} \text{ [Eq. 9]}$$

$$\varphi = atan2(-u;\ -v)\frac{180}{\pi} = 180 + atan2(v;\ u)\frac{180}{\pi} \text{ [Eq. 26]}$$

The definition of atan2 can be found on Wikipedia (https://en.wikipedia.org/wiki/Atan2, last accessed on 31-12-2020). 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):

$$\sigma_W = 2\sqrt{\frac{u^2}{u^2+v^2}{\sigma_u}^2 + \frac{v^2}{u^2+v^2}{\sigma_v}^2 + 2\frac{uv}{(u^2+v^2)^2}\sigma_{uv}} \text{ [Eq. 27]}$$

$$\sigma_\varphi = \frac{180}{\pi}\left(\sqrt{\left(-\frac{uv}{u^2+v^2}\right)^2{\sigma_u}^2 + \left(\frac{vu}{u^2+v^2}\right)^2{\sigma_v}^2 - 2\left(\frac{uv}{u^2+v^2}\right)^2\sigma_{uv}}\right) \text{ [Eq. 28]}$$

In Figure 6, 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 approach discussed in this section. In RHARM any type of correlation between the uncertainties has been neglected. Were correlation to exist, the uncertainties would have need to be inflated commensurately.

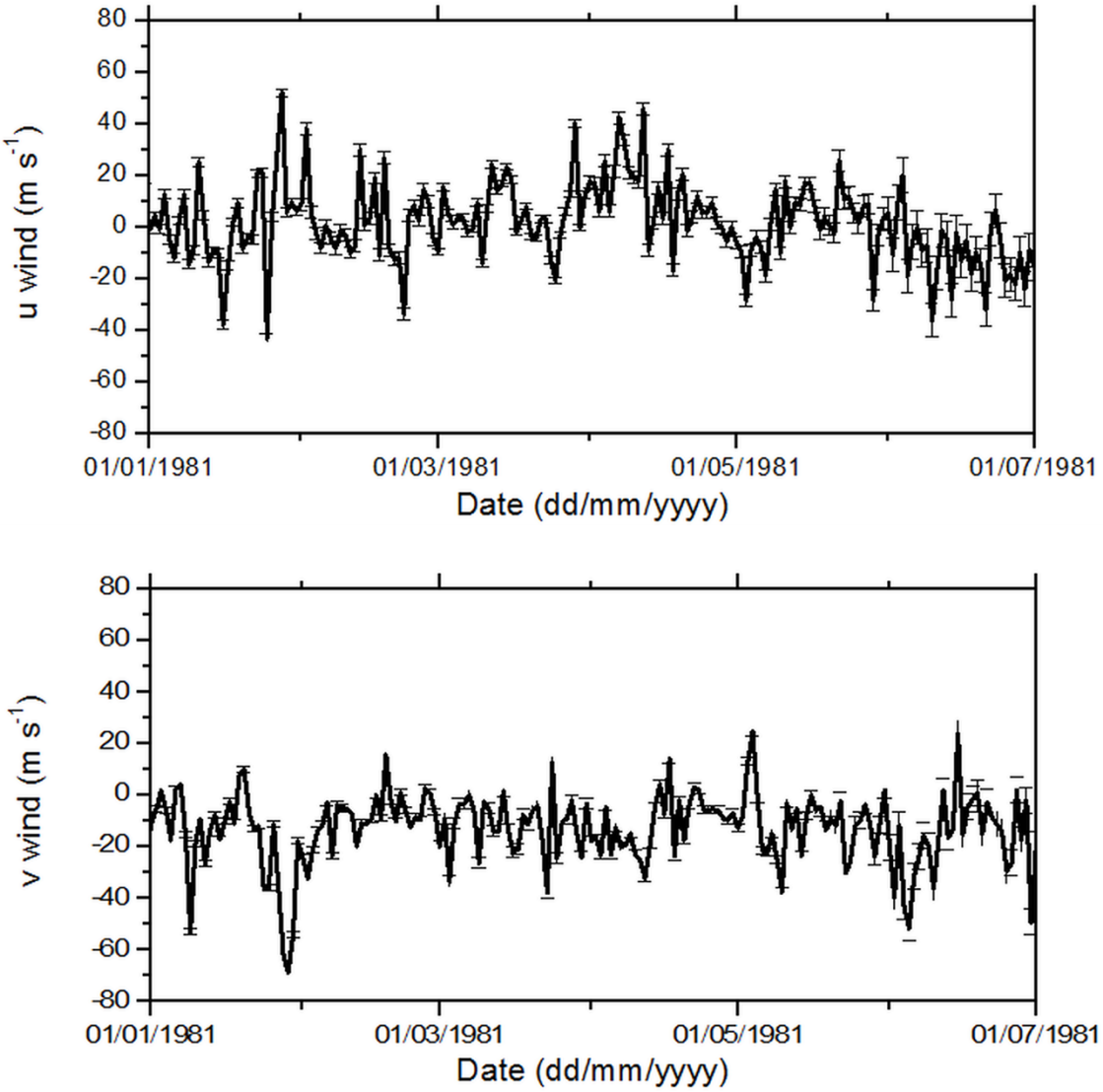

Figure 6: Top panel, zonal wind component (u) time series at 300 hPa (only night time) for the Sodankyla station with the uncertainties calculated using RHARM for the period from 01/01/1981 to 01/07/1981. Bottom panel, same as top panel but for meridional component (v). The vertical bars show the random uncertainties quantified using the statistical method, and their plotting has been reduced to one value each two of the time series.

The RHARM dataset 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 estimation of trends.

The CUSUM test is extremely efficient in the detection of changes in a time series. However, the use of smoothed time series may affect the exact identification of the break occurrence (in terms of days, before or after the real occurrence). RHARM has been designed 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

Comparisons for the adjustments applied to the entire time series, i.e. merged PPTS and HTS (Figure 7), 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 more recent observations (since 2004), the 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., 2012). 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 at the tropics. For wind speed, as anticipated, the systematic effects have a smaller magnitude than for temperature and RH. Tables 5 and 6 further summarise the main characteristics of adjustments. The 1st and 3rd quartiles for RHARM temperatures are 0.9 K and 0.3 K higher than IGRA, respectively, revealing the predominance of cold biases in the IGRA data since 1978; for RH, the 1st and 3rd 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.

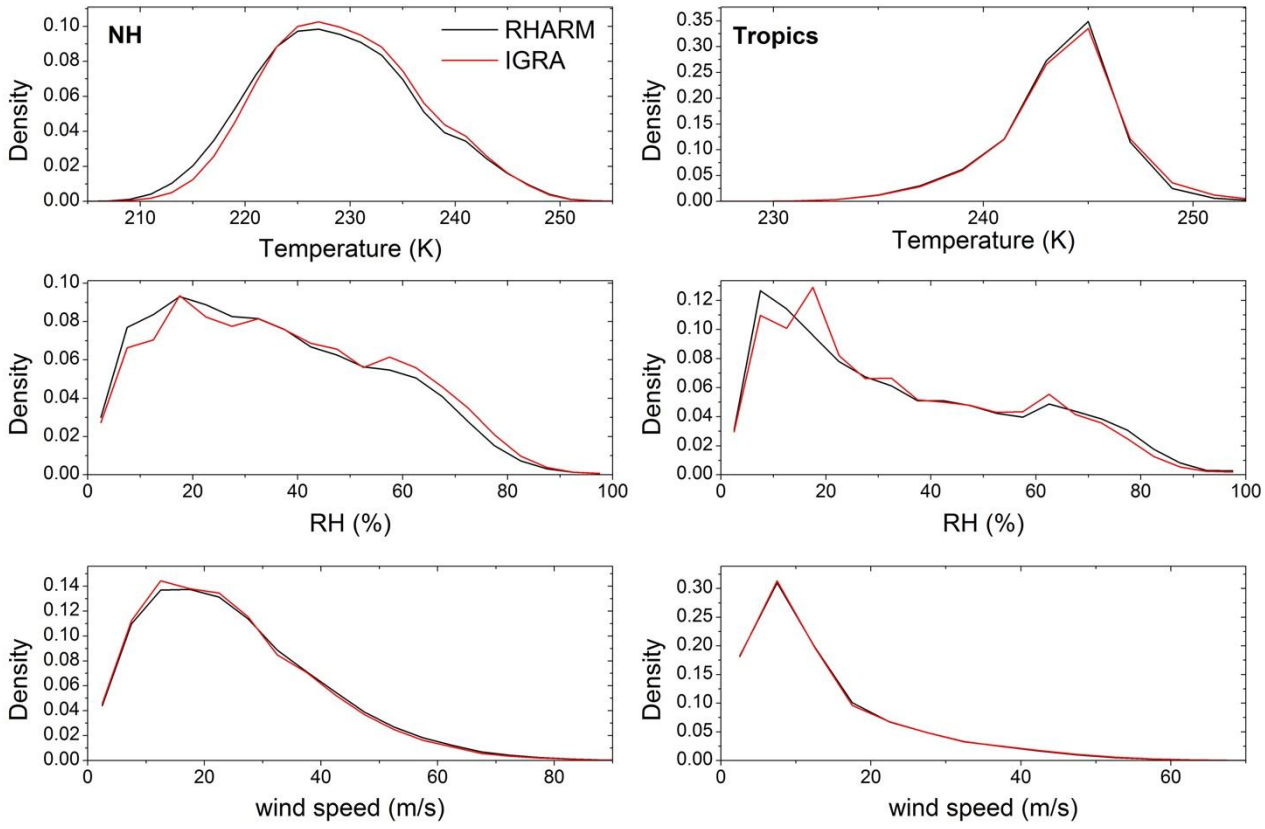

Figure 7: probability density functions (pdfs) calculated in the northern hemisphere (NH) and in the tropics (±25° latitude) at 300 hPa for the IGRA and RHARM datasets of temperature (top panels), RH (middle panels), wind speed (bottom panel) using the stations shown in Figure 1. RHARM data refers to the merged time series (PPTS + HTS). The median, the first and third quartiles of the pdfs are reported in Tables 5 and 6 for convenience.

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

Table 5: first, second (median) and third quartiles of the Northern Hemisphere 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

Table 6: first, second (median) and third quartiles of the tropics pdfs shown in Figure 7.

The pdfs shown in Figure 7 are representative of the overall dataset in the NH and the tropics and, although useful to show the effect of the RHARM adjustment on the original data distribution, they cannot show alone the effect of the applied adjustment in terms of reducing the instrumental effects on the anomalies and climate trends.

To this purpose, Figures 8-10 show global maps of trends of temperature, relative humidity and wind speed at three pressure levels (850 hPa, 300 hPa and 100 hPa, no relative humidity for the latter) estimated from IGRA and RHARM.

In term of the results at 850 hPa (Figure 8) covering roughly to top of the planetary boundary layer except in regions of high topography, the comparison of temperature trends shows enhanced homogeneity for RHARM, especially in northern Europe and Siberia, with additionally the removal of a few isolated and obviously spurious large station trends. The general tendency is for a cooling in North America, a moderate cooling in the SH, and a moderate warming in Europe, Middle East and Asia. For relative humidity, the variability of the trends is larger than for temperature: the adjustments applied by RHARM reduces heterogeneity in Siberia, South America and Northern Europe. The overall tendencies show a positive trend in the NH and a negative trend in the SH. For wind speed, RHARM improves the homogeneity of the trends globally, in particular in Siberia, Southern Asia and Central-South America, adjusting some of the most negative and positive values in both the hemispheres. An overall neutrality of the trends is obtained with most values ranging within -0.4 to 0.4 m/s.

At 300 hPa (Figure 9), improvements in the homogeneity of temperature trends are mainly visible in Siberia and Central-South America. For RH, improvements are observed mainly in Siberia, South

America, and South Africa. 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 et al. (2020c, 2020d) in Europe, in the SH and the tropics. Wind speed shows results similar to trends at 850 hPa.

Finally at 100 hPa (Figure 10), temperature trends appear more homogeneous globally with the adjustment in various regions of a number of spurious values, above all in Siberia and Northern America. For wind speed, the regions where RHARM adjustments bring the major improvement are again Siberia and Central-Southern America with an overall positive trend in the NH and a neutrality in the SH. The lack of values of trends at stations in the southeast regions and at the poles is related to the small number of ascents reaching 100 hPa at the stations selected by RHARM. RH trends are not reported because these are not available from RHARM (limited at 250 hPa).

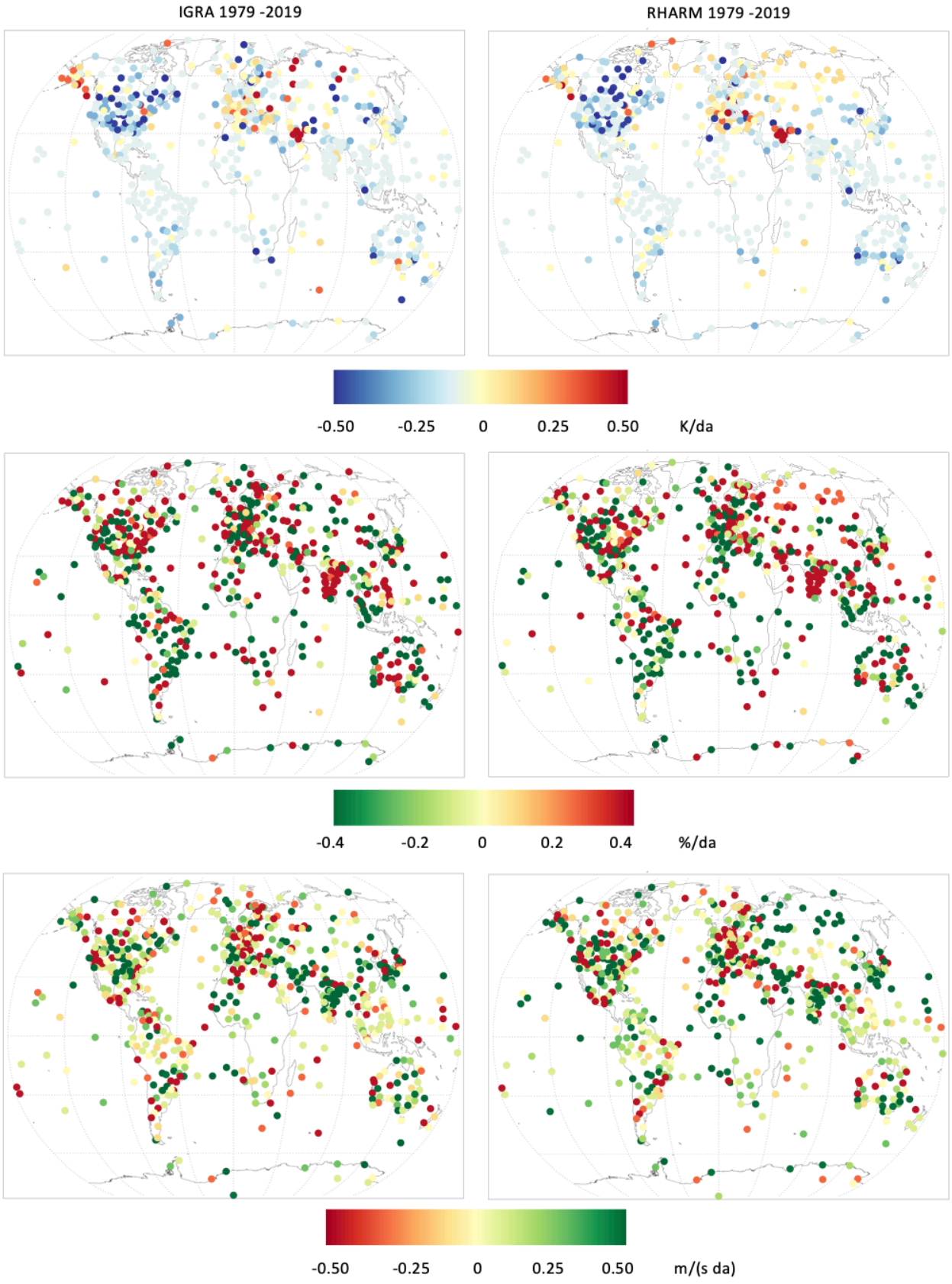

Figure 8: global maps of the decadal trends at 850 hPa of temperature (top), relative humidity (middle) and wind speed (bottom) estimated from IGRA (left panels) and RHARM (right panels) common stations. Decadal trends have been estimated over the entire time series for each station in the period 1979-2019.

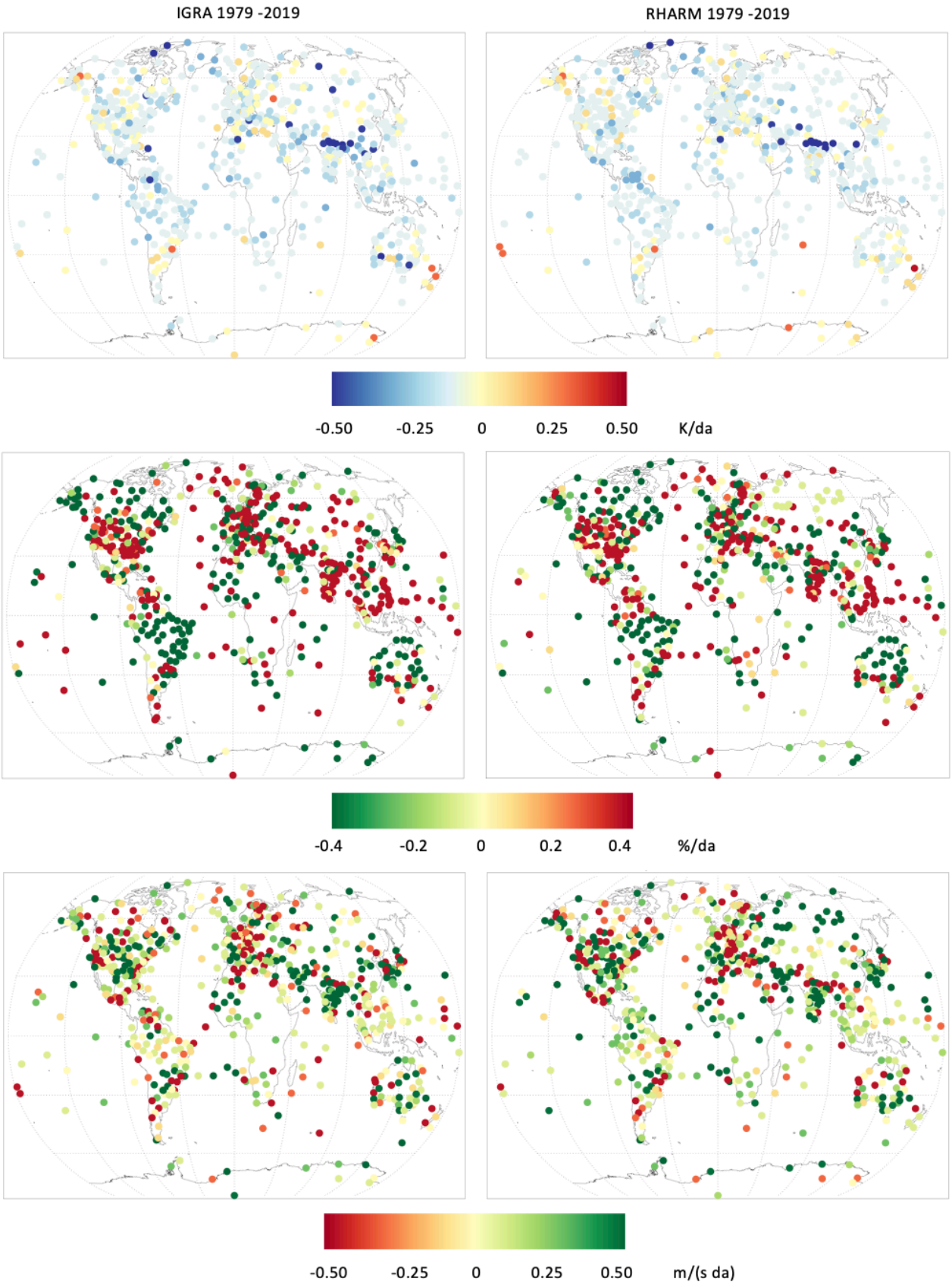

Figure 9: same as Figure 8 at 300 hPa.

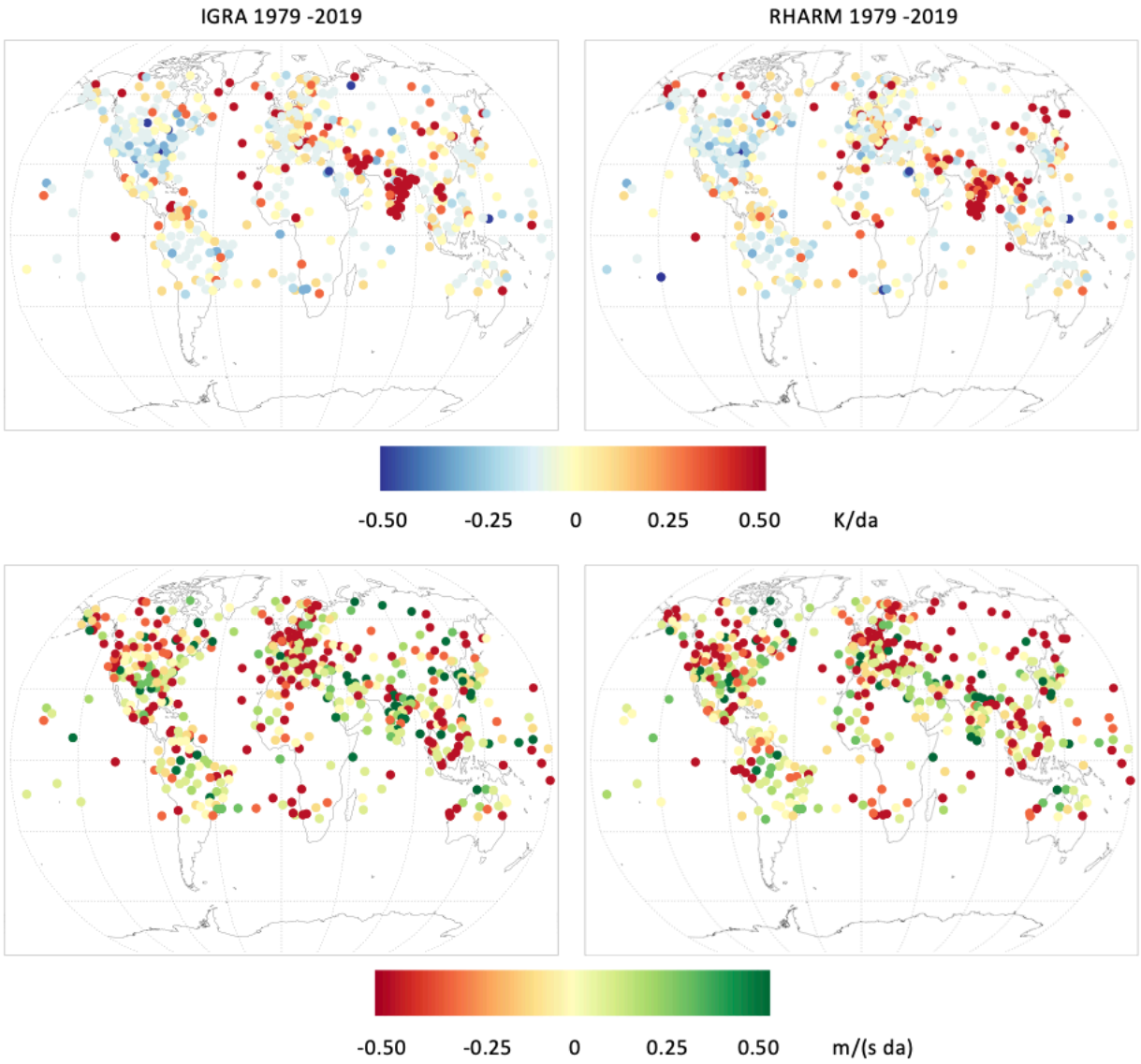

Figure 10: same as Figure 8 at 100 hPa for temperature and wind speed only.

Figures 11-13 provide the comparison of the time series of the monthly anomalies in the NH for the IGRA and RHARM temperature, relative humidity and wind measurements at 300, 400, 500, 850 hPa. In Figure 11, temperature monthly anomalies for RHARM are significantly reduced in the period from 1985 to 2005 on all levels.

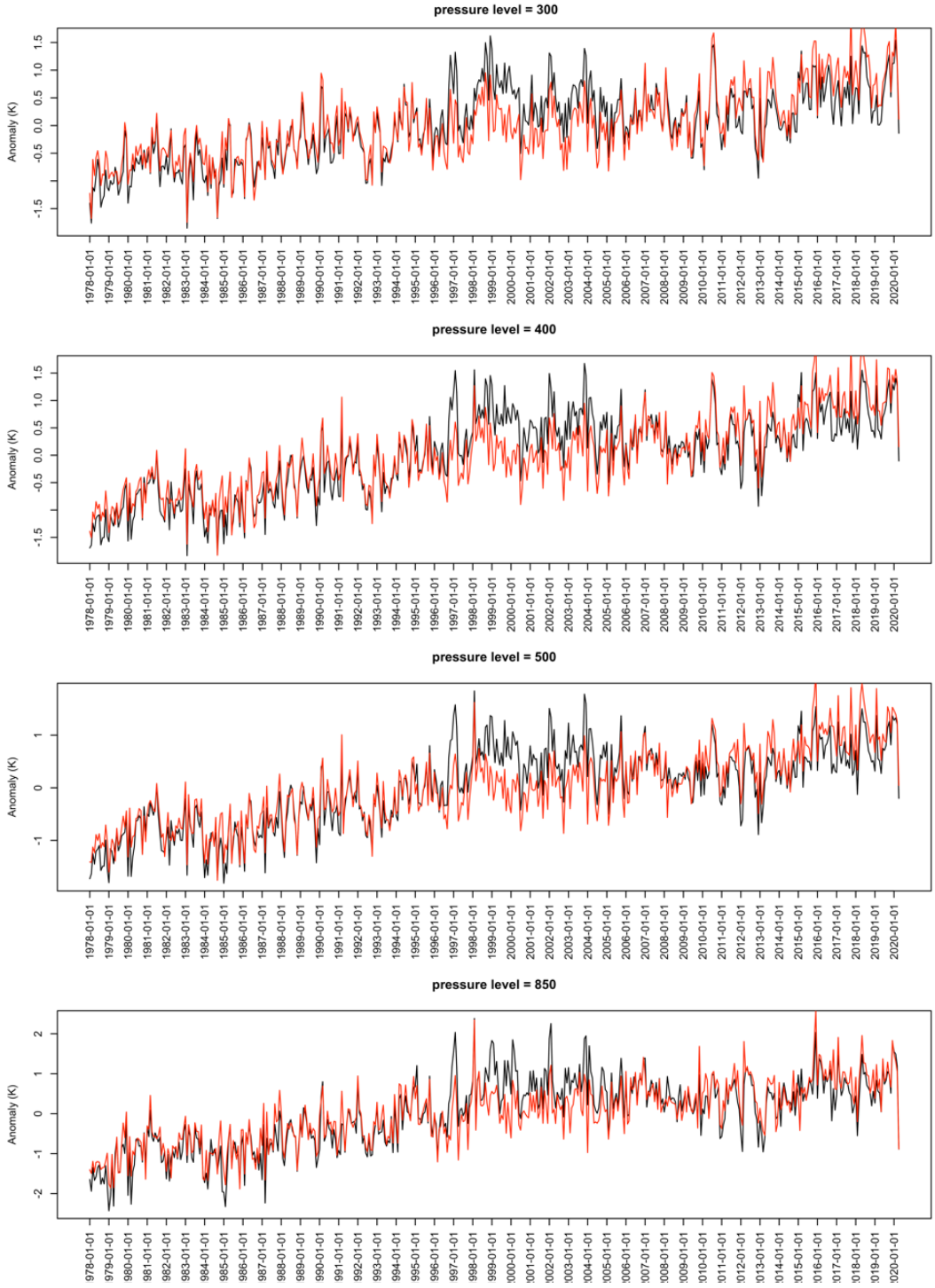

Figure 11: temperature monthly anomalies at 300, 400, 500, 850 hPa in the NH for IGRA (black line) and RHARM (red line) radiosounding measurements from 1978 to 2020.

In Figure 12, the RH monthly anomalies for RHARM are significantly adjusted before 1987 with a reduction from more than 10% RH to less than 6% RH at 300 hPa. Similar adjustments are applied at 400 hPa. In both the time series, the anomaly reduction is related to the wet bias, affecting the sonde used in the corresponding period: The RS18 and RS21 used at Sodankyla station (Figure 5) provide two examples of this wet bias. The RKS and MARZ Russian sondes were affected by a similar issue, and they were widely used at eastern Germany stations in the 1980s.

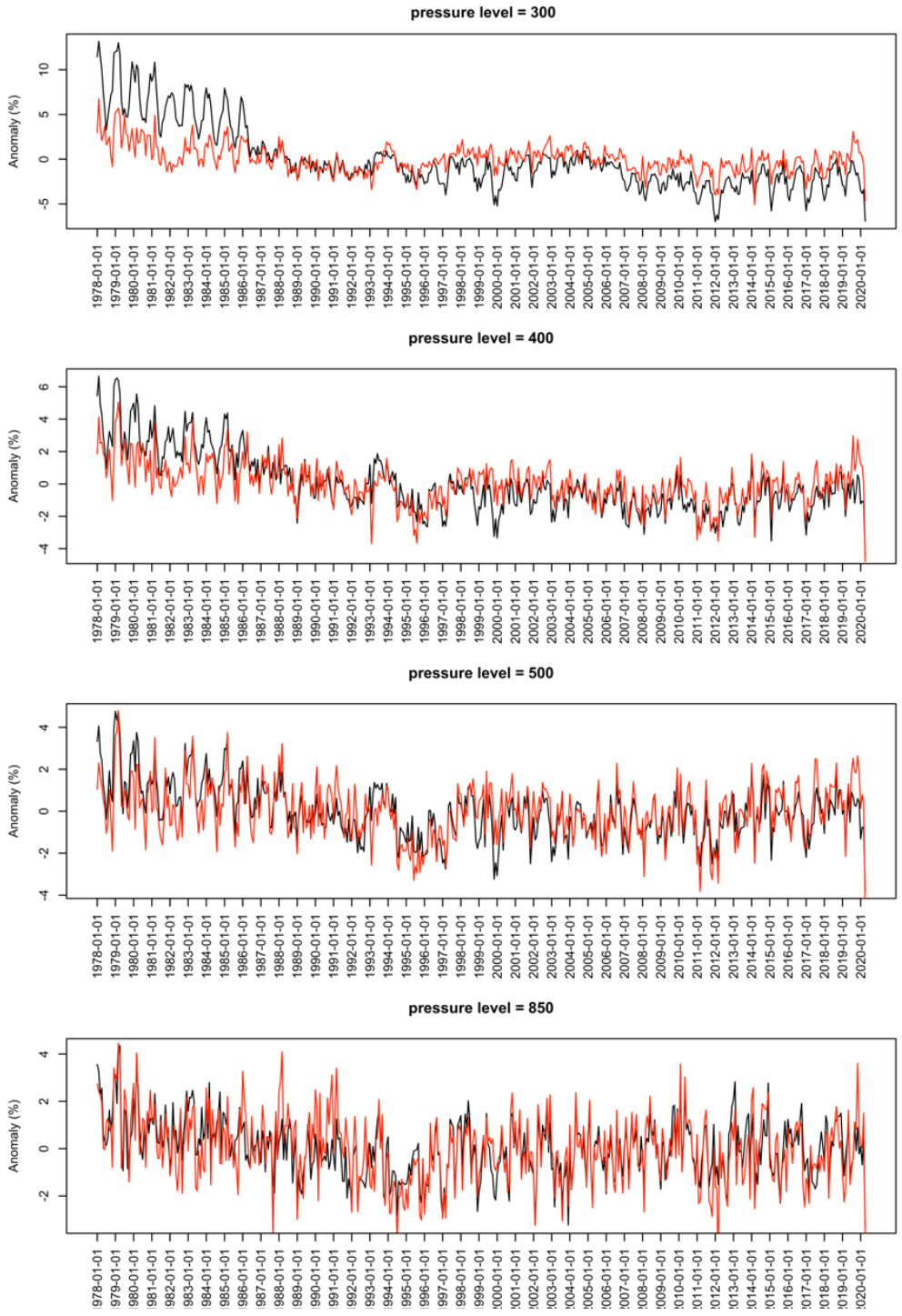

Figure 12: same as Figure 11 for relative humidity.

In Figure 13, the wind speed monthly anomalies show, as expected, smaller adjustments to the anomalies than temperature and relative humidity, which anyhow allows reduction of the number of stations affected by an anomalous bias, as shown in the map in Figure 10.

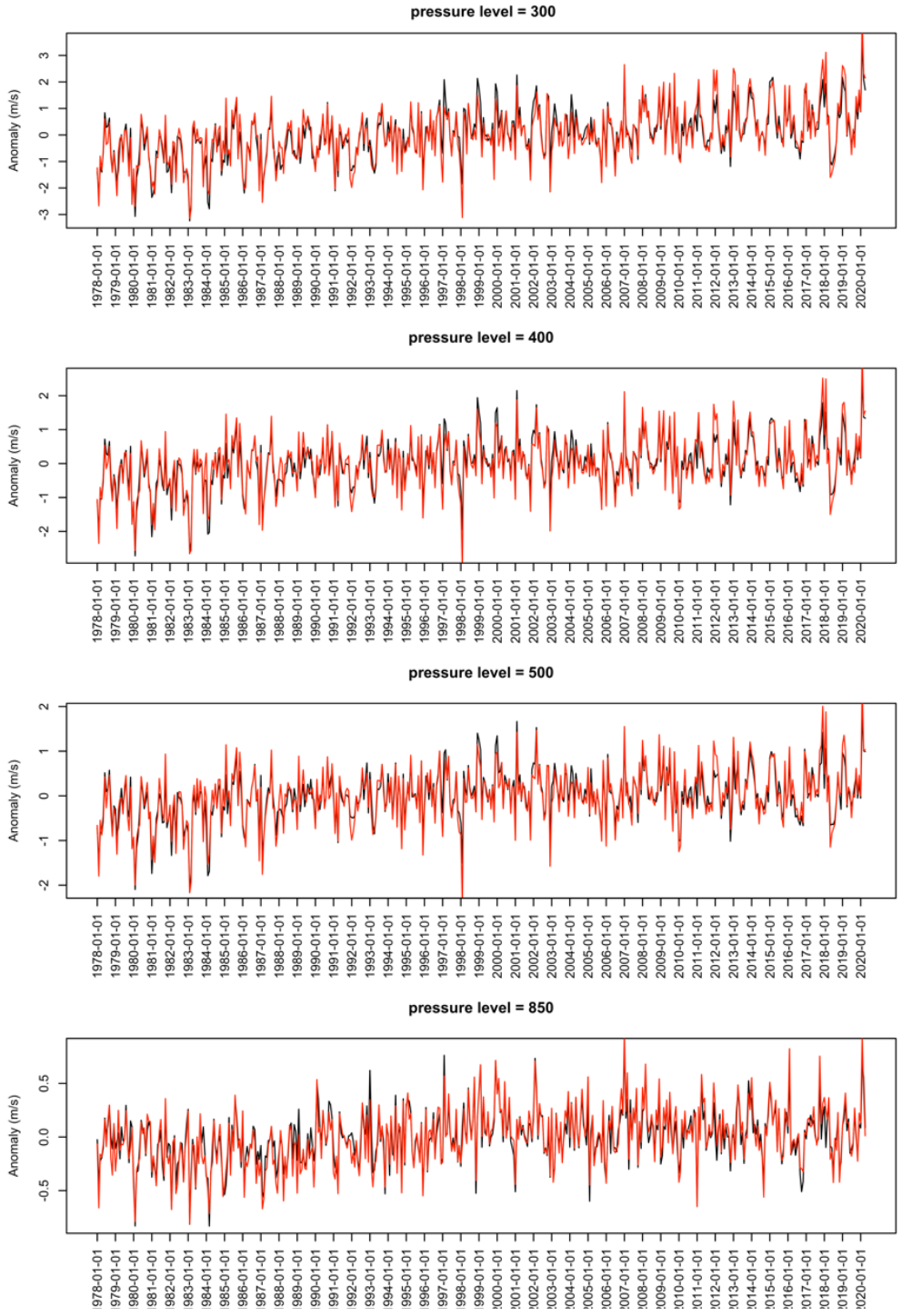

Figure 13: same as Figure 11 for wind speed.

4.2 RHARM consistency with GRUAN

Although built to mimic the GDP procedures, the RHARM approach is not applied to the raw radiosonde data. This may generate discrepancies between the RHARM and the GDP which must be quantified. By construction, the RHARM approach is expected to be similar on average to the GDP. For 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 14). 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 0.07 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 towards 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% to 2.0% and, during the day, from 0.7% to 1.8% RH (Figure 14). The RHARM adjustments are able to 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 have been composited for night and daytime observations, because there is no difference in the data processing applied in the products. Both the GRUAN-IGRA and GRUAN-RHARM difference profiles, shown in Figure 15, are very close to zero from 1000 to 300 hPa. Above this altitude, the RHARM-GRUAN difference is smaller than IGRA-GRUAN difference and below 0.05 m/s, while IGRA shows differences with GRUAN within about ±0.3 m/s.

Summarizing, the discussed consistency check confirms the overall reduction of the bias with GRUAN if RHARM data are used, as expected by construction, with small residual differences with GRUAN, likely due to rounding problems or differences in the smoothing window used in the manufacturers' and GRUAN data processing.

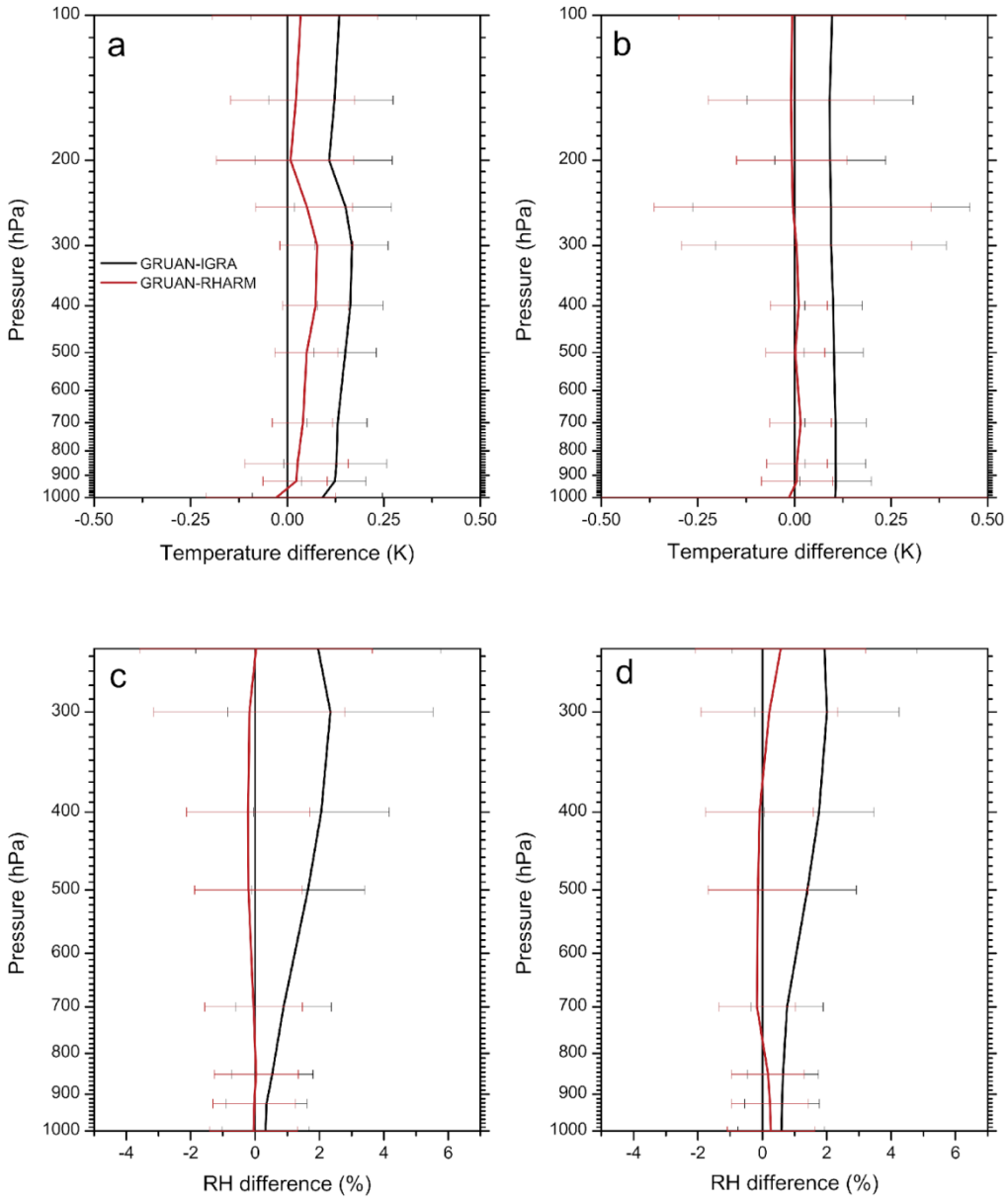

Figure 14: Mean difference profiles of temperature (top panels) and relative humidity (bottom panels) with the corresponding standard deviations (horizontal bar) calculated from the comparison of the nighttime (panels a and c) and daytime (panels b and d) difference “GRUAN minus IGRA” (black lines) and “GRUAN minus RHARM” (red lines) for the profiles available at all GRUAN stations, in the period 2008-2018.

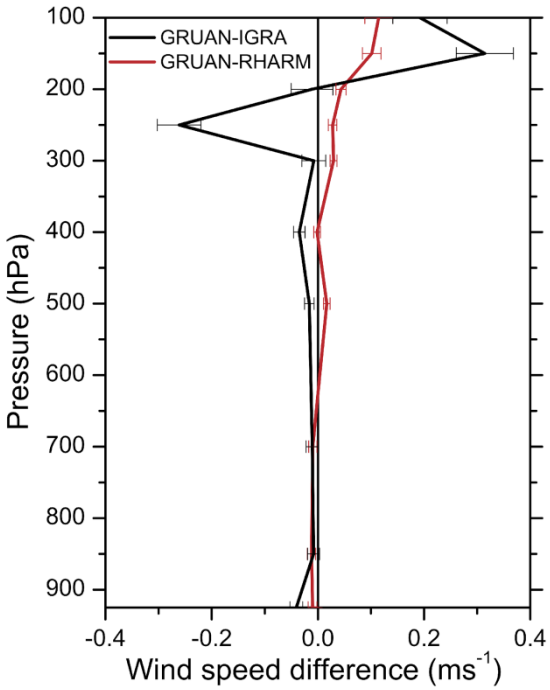

Figure 15: Same as panels in Figure 14 but for wind speed including both night and daytime observations.

To further show the effect of the adjustments using the RHARM approach, we show a comparison between GRUAN, IGRA and RHARM RH values for all the GRUAN stations over 2008-2018 (Figure 16). The RHARM RH values become considerably more similar to GRUAN, especially for values higher than 55% RH. 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.

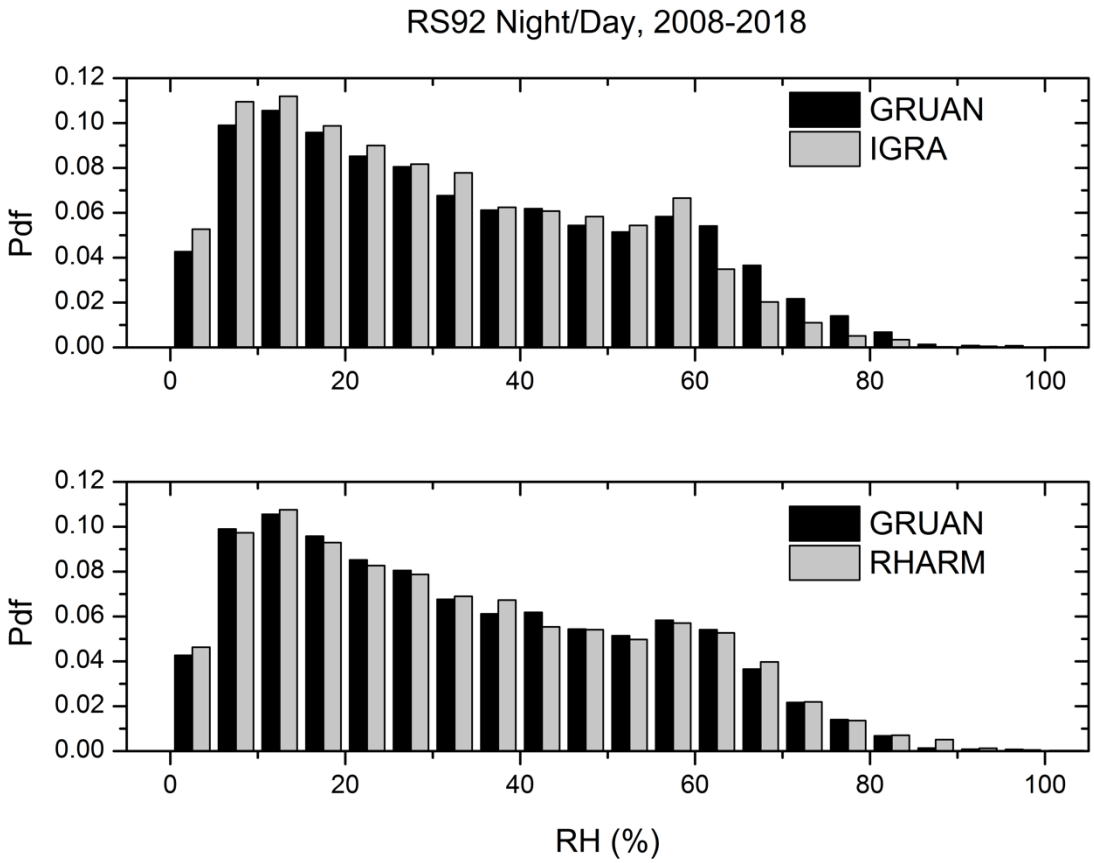

Figure 16: top panel, comparison between GRUAN (black) and IGRA (grey) RH measurements at 300 hPa for the profiles available at all GRUAN stations (only RS92 sondes), in the period 2008-2018 The comparison comprises all the night and daytime observations on 00:00 and 12:00 UTC. Bottom panel, same as top panel but for GRUAN (black) and RHARM (grey).

4.3 Statistics of breakpoints and vertical correlation

In the previous sections, we have clarified how the RHARM algorithm applies a post-processing of the data, and not a pure statistical homogenization of the IGRA data approximately since 2004, i.e. in the PTTS portion of each time series. The homogenization of HTS portion of each time series, i.e. data typically before 2004 on average, is applied at each mandatory pressure level separately. Therefore, it is informative to study the distribution per year of the breakpoints detected in the HTS for the measured ECVs as well as the correlation of the percentage of breakpoints per variable at different pressure levels. For the latter purpose, the 100 hPa, 300 hPa, 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 with the entire vertical profile (e.g., factory calibration), are applied in the data processing.

In Figure 17, the comparison between the percentage of breakpoints identified by RHARM from 1980 to 2010 for all homogenized series and at all mandatory pressure levels is shown; the detail for the three selected pressures is provided in Figure 18. In Figure 17, the percentage decrease after 2004 is due to the progressive introduction of the most recent radiosonde types for which the GRUAN-like component of RHARM can be applied. Furthermore, the percentage of breakpoints decreases going towards the past and this may be overall compatible with the use of the same type of radiosondes and data processing for longer periods at the selected stations. Figure 17 also reveals

the good agreement in the percentage of breakpoints identified across the different ECVs. For RH, we remark that the RHARM algorithm homogenizes only the levels at pressures higher than 250 hPa, thus determining a lower absolute number of breakpoints.

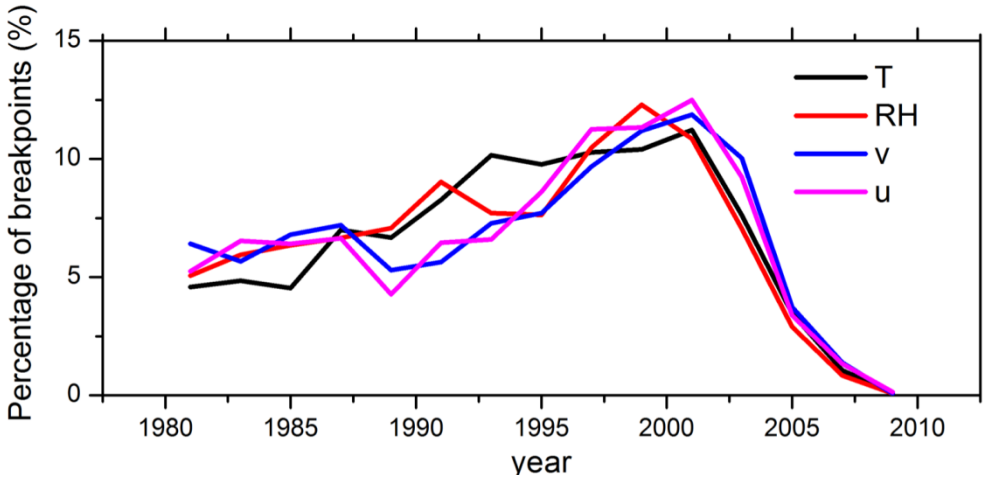

Figure 17: Percentage of breakpoints per year detected by the RHARM approach for each of temperature, relative humidity, meridional and zonal wind speed, cumulated for all the homogenized time series and at all mandatory pressure levels

Comparisons of breakpoint percentages (Figure 18) show a similar distribution for RH, u and v, while for temperature the correlation is larger at 300 hPa 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 in this period, or a larger number of false positives than 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. 2020).

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 hPa and 500 hPa. Within a window of 2 months, correlation for temperature and RH breakpoints at 500 hPa - 300 hPa is about 0.2, while it rises to 0.36 within 6 months and to 0.6 within a year. For wind vectorial components, within a time difference of 2 months, correlation for temperature breakpoints at 500 hPa - 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 hPa - 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. Besides, 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 guaranteed. 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. This choice was not preferred for RHARM because even though it has a high potential to capture all the true breakpoints in a time series, it may also increase the number of false positives and over-adjust the original data.

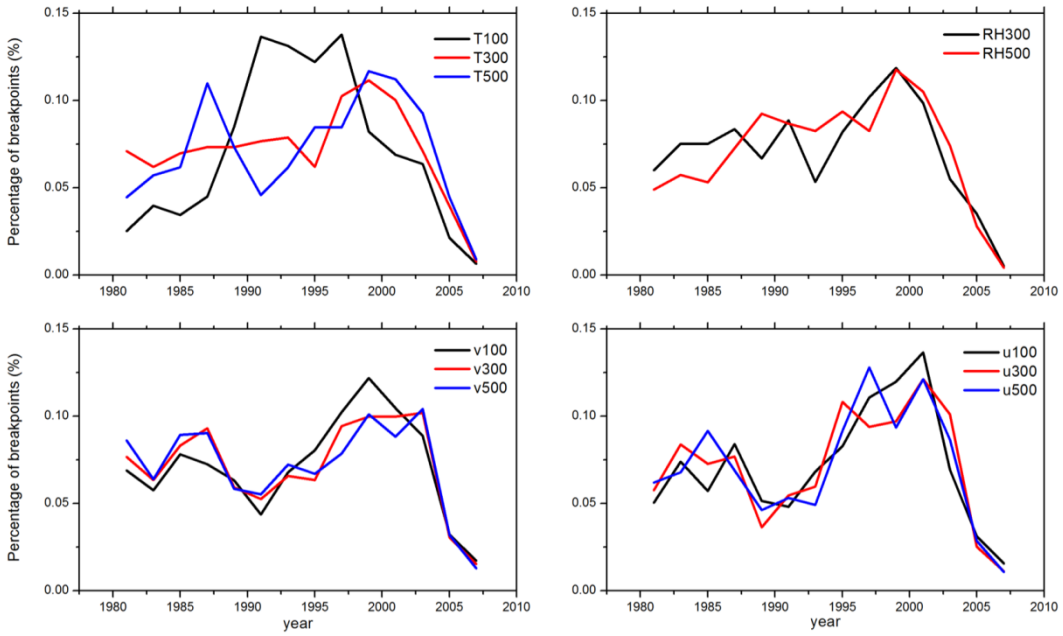

Figure 18: Percentage of breakpoints per year detected by the RHARM approach for temperature (upper left panel), relative humidity (upper right panel), meridional (bottom left panel) and zonal wind speed components (bottom right panel). Each panel reports the frequency of occurrence per year at three pressure levels, 100 hPa (except for RH), 300 hPa, 500 hPa.

5. Quantification and presentation of uncertainties

A unique value of RHARM compared to other harmonized datasets is that, for the first time, an estimation of the uncertainty is provided for each single observation (i.e. at each pressure level). In this section, statistics on the 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 GDP as expected given the methodological considerations outlined in section 3 (Figure 19). In particular, for temperature (Figure 19, 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 underestimate compared to that of GRUAN, as seen for values below 0.1 K. These values are mainly related to nighttime measurements.

For RH (Figure 19, right panel), the median value of the GRUAN 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 %.

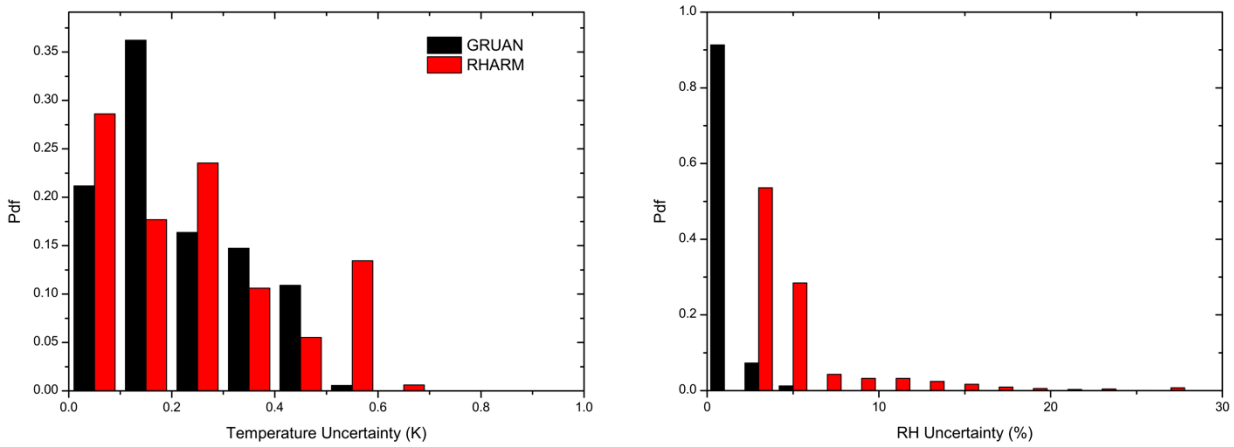

Figure 19: Comparison of pdfs of the uncertainty calculated using the GRUAN data processing (GDP) and the RHARM approach at the six stations shown in Table 1. Pdfs are relative to temperature (panel a) and relative humidity (panel b)

In Figure 20, 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 bimodal with modes centered around 0.5 K 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 PPTS. 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. A more comprehensive validation of the uncertainty estimates for T and RH is provided in the companion paper.

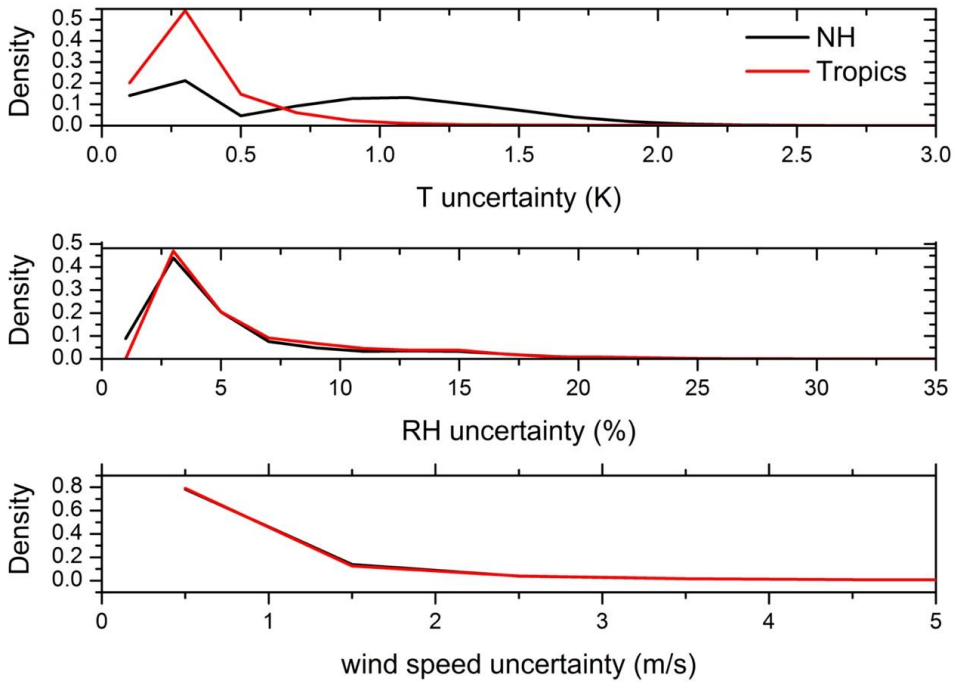

Figure 20: Comparison of the density function of the total uncertainties estimated for all the RHARM temperature (T), relative humidity (RH) and wind speed (w) data since 1978 for the stations in the tropics and in the Northern Hemisphere (NH).

6. Discussion and conclusions

The RHARM dataset provides a new homogenization option, complementary to existing datasets of homogenized radiosounding temperature measurements and to the handful of existing products for RH and wind. RHARM differs from these previous efforts due to the use of reference measurements to calculate and adjust for systematic effects in the most recent portion of records when possible. A significant benefit is that each harmonized data series is provided with an estimation of the uncertainty for each observation. The novel approach enables a more comprehensive exploration of uncertainties in historical records.

Results from the analyses in this paper show that:

- RHARM temperature data distribution is warmer than IGRA in the NH, due to the predominance in cold biases affecting the IGRA time series since 1978, while RHARM is slightly cooler than IGRA in the tropics. For RH, RHARM successfully adjusts IGRA data affected by 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.
- RHARM is able to increase the homogeneity of decadal trends compared to IGRA (examples are provided at 850 hPa, 300hPa and 100 hPa) with the largest effect for all ECVS on Siberian stations, historically affected by significant biases, and specifically for temperature in the NH and Central-Southern America. The largest effects for relative humidity are in Siberia, South America and Northern Europe. For wind speed, RHARM also improves the homogeneity of the trends in Siberia and Central-Southern America.
- The RHARM-GRUAN temperature difference is much reduced compared to the GRUAN-IGRA difference at all levels and for all the ECVS, as expected due the RHARM methodology.
- 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.

The assessment of the RHARM data quality and the related improvements compared to IGRA using several comparison datasets, such as the ERA5 reanalysis, existing homogenized datasets and satellite observations is presented in a companion paper (Madonna et al., 2021).

In an ideal world, the collection and preservation of raw data by all radiosounding stations would allow to build the highest possible quality dataset 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. This means that on an individual station basis, the benefit of applying the proposed adjustment could be limited or could even increase the difference with the “true” value for individual values or not properly estimate the uncertainty. This is different from the solar radiation correction discussed in Section 3 which, though not exactly the same as GDP, adjusts the data distribution, being applied as post-processing of the data and not only as an average correction.

The RHARM dataset is based on a hybrid approach which, by implementing a GRUAN-like data processing for the adjustment of most recent radiosounding ascents, represents an innovative solution that is closer to a ‘traceable’ estimate. For historical upper-air data before 2004, when only statistically based breakpoint detection and adjustment is possible, it is an alternative option to the

few existing homogenized datasets, but with the added value of providing an estimation of the measurement uncertainty.

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) 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

7. Data availability

A copy of the RHARM dataset is stored in the Copernicus Climate Data Store (CDS) although not publicly available yet. For review purposes only, a subset has been made available at http://doi.org/10.5281/zenodo.3973353 (Madonna et al., 2020a). The current version of the RHARM dataset is provided in textual format (comma-separated values).

8. Acknowledgements

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 Dataset (ID2010) has been released upon agreement with the WMO YID protocol, signed by CNR-IMAA and WMO on 27/07/2017.

9. Appendix A: WMO/CIMO 2010 radiosonde intercomparison data

In section 3.2, an outline of the adjustment estimation of all the radiosonde types involved in WMO/CIMO 2010 radiosonde intercomparison, reported in Table 1 was provided. The description of the temperature and humidity vertical profiles obtained from Eqs. 12, 13, 14, 15 reported in Section 3 are shown in this appendix.

In Figure 21, $\Delta T_{RS92,NORS92}$ is shown with the corresponding standard deviations $\sigma_{\Delta T_{RS92,NORS92}}$ for ten radiosonde types during night (upper panels) and day (lower panels) up to 50 hPa. $\Delta T_{RS92,NORS92}$ ranges between -0.2 K and 0.3 K up to 200 hPa, both at night and day. At higher altitudes, $\Delta T_{RS92,NORS92}$ increases with values between -0.3 K and 0.6 K. For a few radiosonde types, the ID2010 provides only a few profiles to calculate the adjustments up to 50 hPa and beyond. This may bias the value of $\Delta T_{RS92,NORS92}$ and inflate the standard deviation.

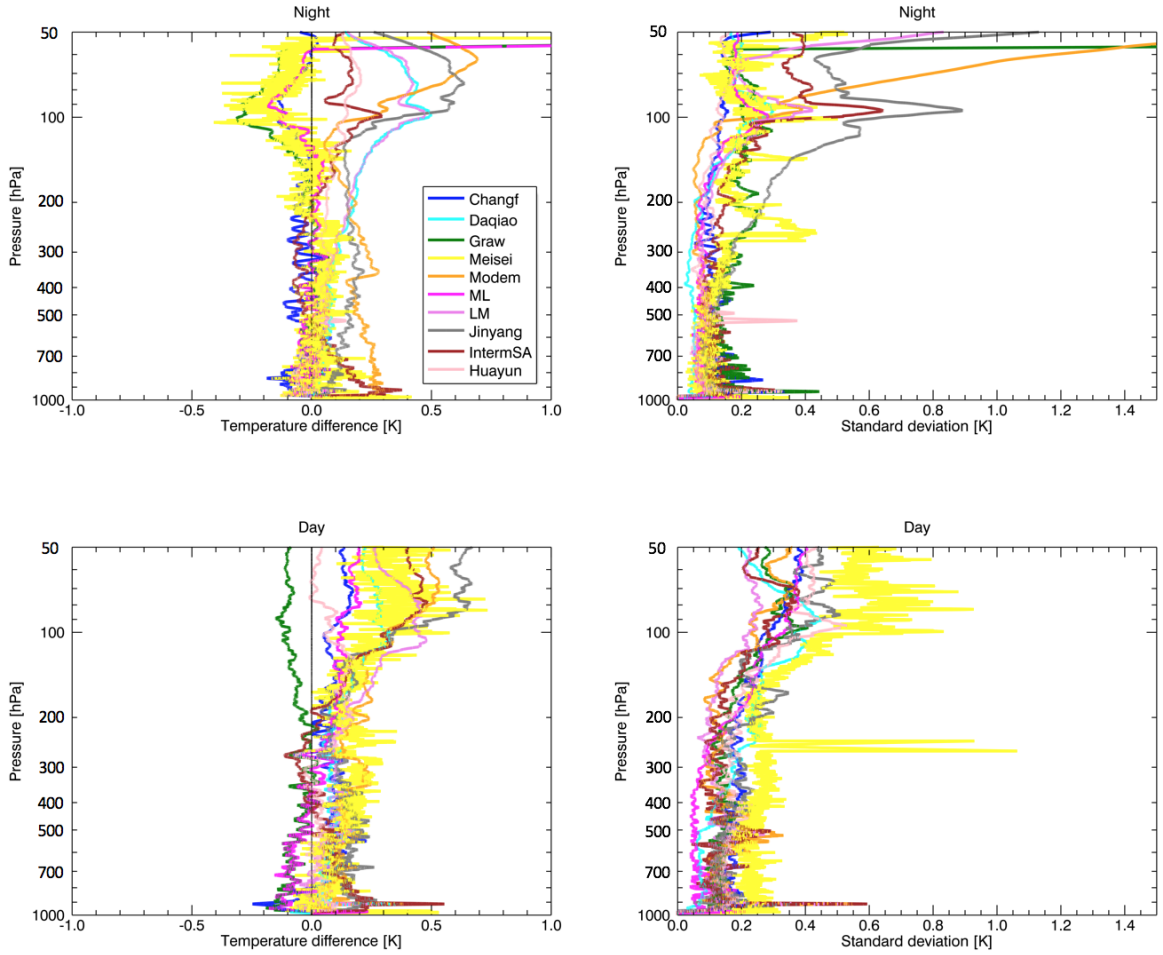

Figure 21: Left panels, nighttime and daytime profiles of the mean differences between RS92 temperature profiles and the profiles measured by all the other radiosonde types listed in Table 4; right panels, profiles of the standard deviation of the mean difference, reported in the corresponding left panels.

For this reason, the profiles in Figure 21 have been cut at tailored pressure levels p_t (ranging between 30 hPa and 100 hPa) and at pressures lower than p_t the adjustment applied in RHARM is equal to the value of $\Delta T_{RS92,NORS92}$ at p_t. $\sigma_{\Delta T_{RS92,NORS92}}$ is within 0.2 K at night up to 200 hPa and increases to 0.3-0.4 K at 100 hPa. A couple of radiosonde types show a larger standard deviation (e.g. JinYang). During daytime $\sigma_{\Delta T_{RS92,NORS92}}$ is larger than at night but is still less than 0.3K up to 200 hPa, while values above this level are very similar to nighttime. The Meisei comparison profiles appear to be generally noisier than the other types, particularly during the day. Some portion of the

apparent periodicity in the left panels of Figure 21 likely relates to manufacturer-to-manufacturer differences in accounting for the effect of the pendulum motion of the radiosondes.

In Figure 22, the mean difference $\Delta RH_{NORS92} = \frac{1}{N}\sum_{i=1}^{N} RH_i^{RHARM,\,RS92} - RH_i^{NORS92}$ is shown with the corresponding standard deviation. The values of ΔRH_{NORS92} are shown instead of $cf(RH)_{NORS92}$, which is the factor calculated in Eq.15, to give a clearer quantitative representation of the difference among the various radiosonde types for the ID2010. The plots in Figure 22 are shown up to 250 hPa which is the maximum altitude at which the RHARM approach performs the post-processing. ΔRH_{NORS92} ranges within about ±10% from the surface up to 500 hPa, both at night and day, although it is mostly positive for all radiosonde types during the day. This indicates that the adjustments applied to correct the effect of solar radiation by most of the manufacturers underestimates the RH profiles compared to the RHARM processed Vaisala RS92 profiles. At pressure levels above 500 hPa, ΔRH_{NORS92} generally increases with altitude and is positive during the day. The only exception is the Modem radiosondes which at night exhibit negative values of ΔRH_{NORS92}, smaller than -15%, and Daqiao and Meteolabor for some levels at pressures higher than 300 hPa. $\sigma_{\Delta RH_{NORS92}}$ is smaller than 10% at night and day, except for a few larger values at levels below 400 hPa reported for the Daqiao, Huayun and Meteolabor radiosondes.

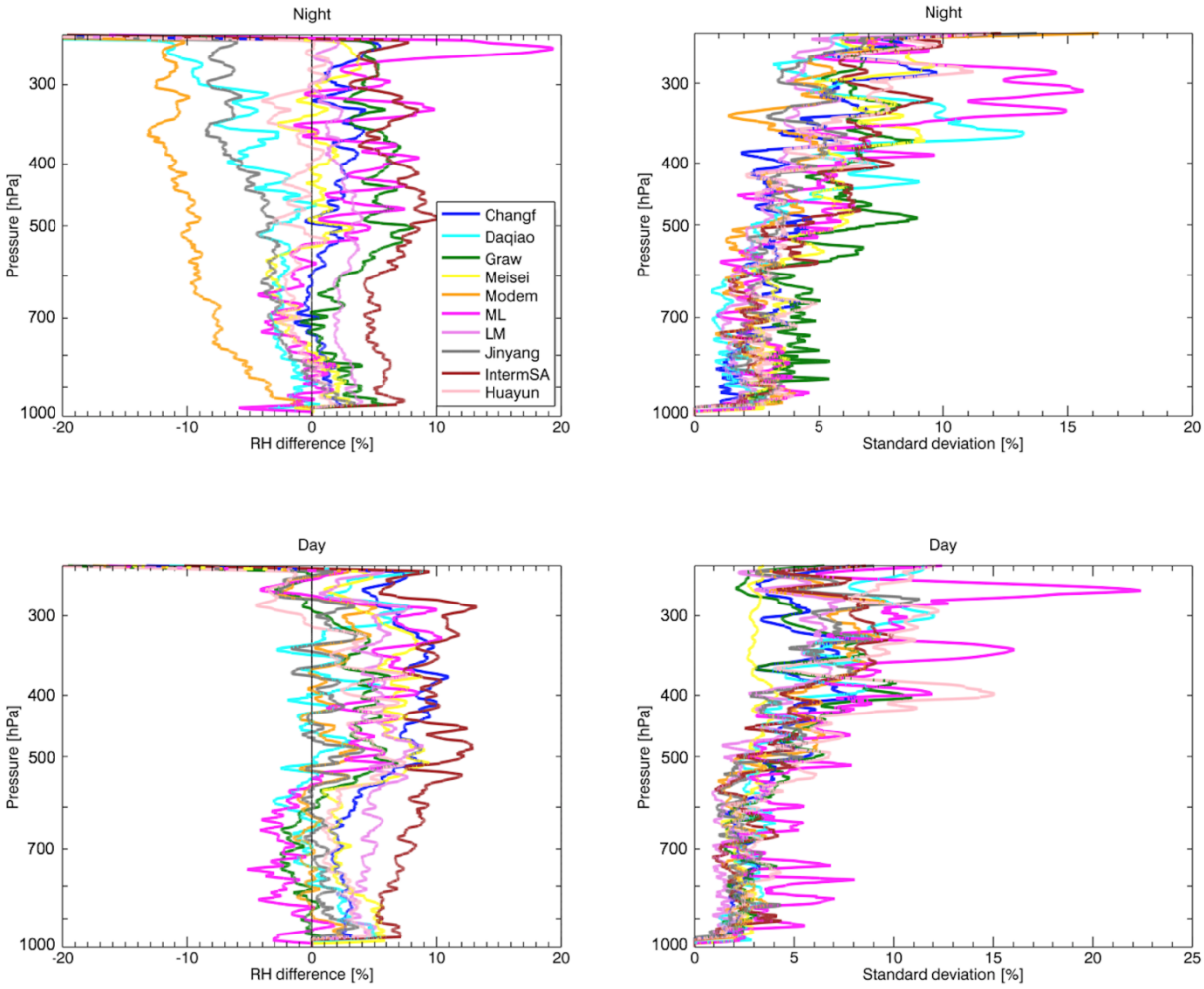

Figure 22: Same as Figure 21 but for RH.

In analogy with Figures 21 and 22, Figure 23 shows the profiles of $\Delta u_{RS92,NORS92}$ with the corresponding standard deviations $\sigma_{\Delta u_{RS92,NORS92}}$. The ID2010, apart from the Daqiao sondes, includes only winds measurements based on GNSS tracking of the radiosonde. Moreover, in the

ID2010 daytime and nighttime measurements were treated together as no significant difference could be found between the two categories. Nevertheless, considering that a different approach to the processing of the ID2010 is adopted by RHARM (i.e. decomposition into vectorial wind components u and v) and that here only one radiosonde model (e.g. RS92) is assumed as the reference for the calculation of adjustment profiles for all other sonde types of the ID2010, we treated daytime and nighttime data separately in order to check the robustness of the estimated adjustments. At night, $\Delta u_{RS92,NORS92}$ is predominantly negative throughout the profile for all manufacturers but is smaller than -0.5 ms^{-1} up to 400 hPa, then increases up to -2.0 ms^{-1} at 100 hPa reaching its maximum value. During the day, the same behavior is observed although the values from the surface to 400 hPa show greater spread. $\sigma_{\Delta u_{RS92,NORS92}}$ is lower than 2.0 ms^{-1} for both day and night, except for Graw and Modem radiosondes above 100 hPa and 50 hPa heights, respectively. Figure 24 shows the same as Figure 23 but for $\Delta v_{RS92,NORS92}$. Both at night and day, $\Delta v_{RS92,NORS92}$ is negative and smaller than -0.5 ms^{-1} up to 400 hPa while it is positive at lower pressure levels with values lower than 1.0 ms^{-1}. The small sample size for the comparison clearly affects the values of $\Delta v_{RS92,NORS92}$ at levels above 100 hPa. The same is true for $\sigma_{\Delta v_{RS92,NORS92}}$ for Graw and Modem sondes at night. $\sigma_{\Delta u_{RS92,NORS92}}$ is generally lower than 1.0 ms^{-1} both at night and day.

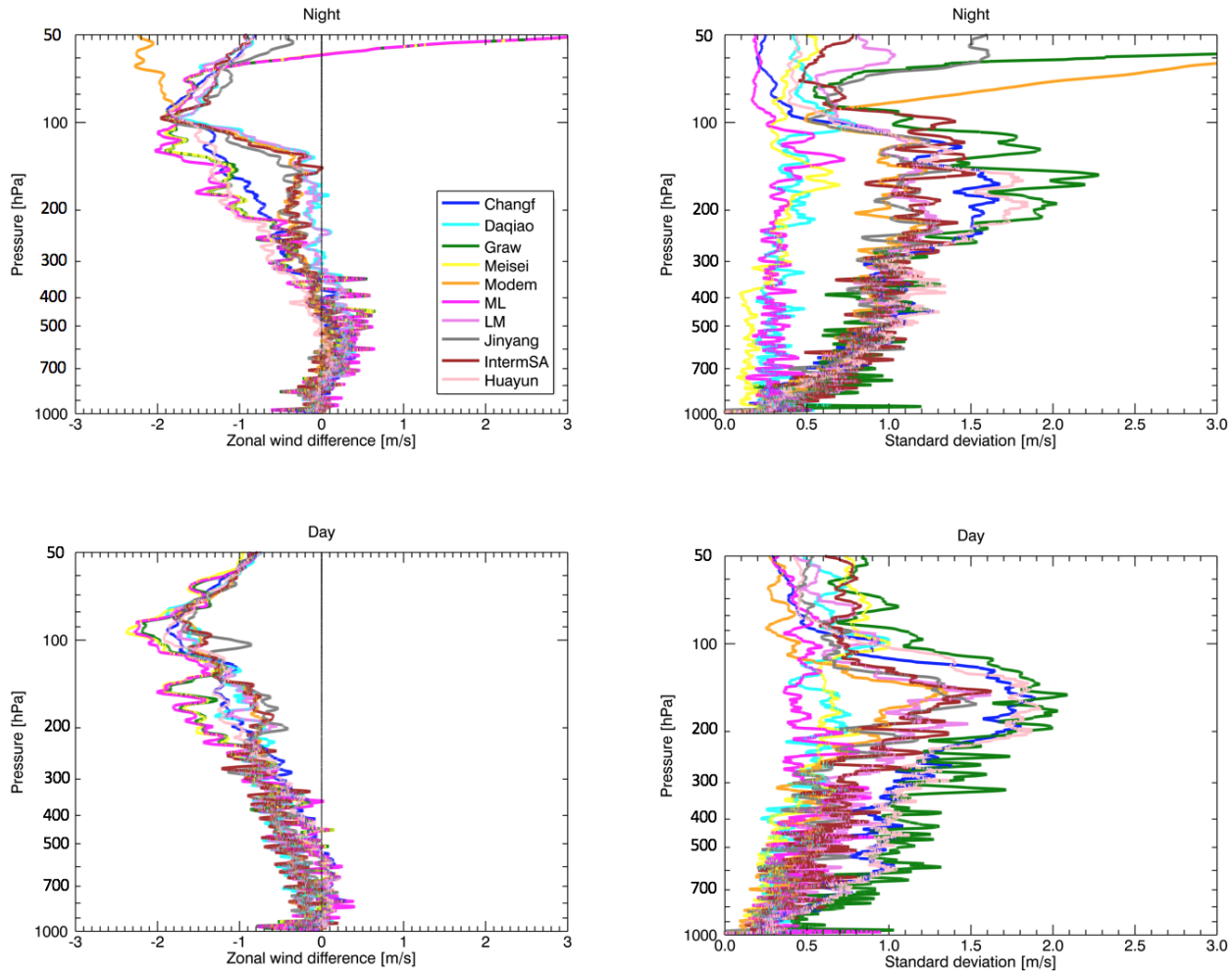

Figure 23: Same as Figure 21 but for the zonal wind component (u).

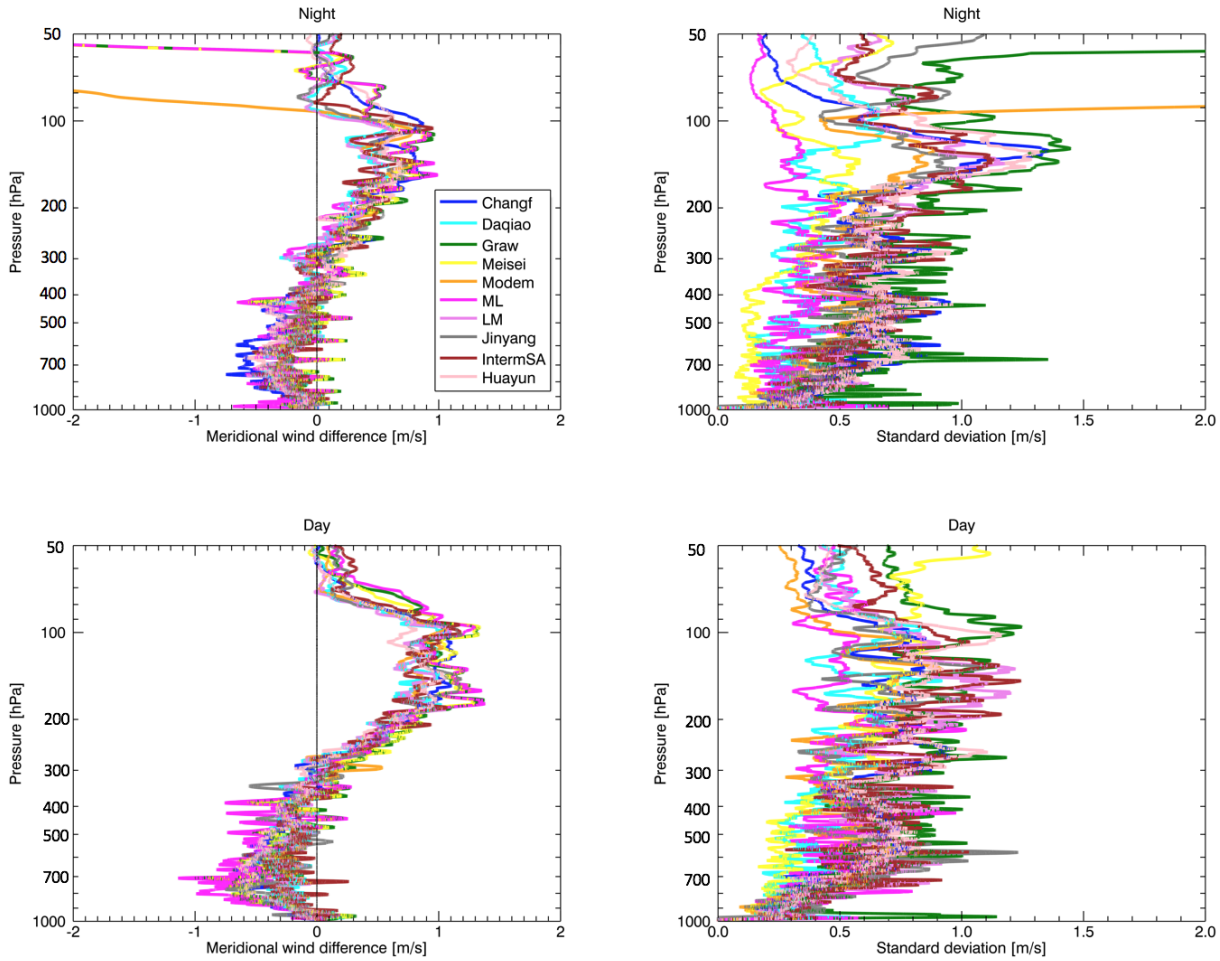

Figure 24: Same as Figure 21 but for meridional wind component (v).

10. References

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

Bauer, P., Radnóti G, Healy SB, Cardinali C.: GNSS radio occultation constellation observing system experiments. Mon. Weather Rev.. doi: 10.1175/MWR-D-13-00130.1.

Bodeker, G.E., Bojinski, S., Cimini, D., Dirksen, R.J., Haeffelin, M., Hannigan, J.W., Hurst, D.F., Leblanc, T., Madonna, F., Maturilli, M., Mikalsen, A.C., Philipona, R., Reale, T., Seidel, D.J., Tan, D.G.H., Thorne, P.W., Vömel, H., Wang, J., 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

Calbet, X., Peinado-Galan, N., Ripodas, P., Trent, T., Dirksen, R., Sommer, M., 2016. Consistency between GRUAN sondes, LBLRTM and IASI (preprint). Gases/Remote Sensing/Validation and Intercomparisons. https://doi.org/10.5194/amt-2016-344

Cramer, W., Guiot, J., Fader, M., Garrabou, J., Gattuso, J.-P., Iglesias, A., Lange, M.A., Lionello, P., Llasat, M.C., Paz, S., Peñuelas, J., Snoussi, M., Toreti, A., Tsimplis, M.N., Xoplaki, E., 2018. Climate change and interconnected risks to sustainable development in the Mediterranean. Nature Clim 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.P., Uppala, S.M., Simmons, A.J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M.A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A.C.M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A.J., Haimberger, L., Healy, S.B., Hersbach, H., Hólm, E.V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A.P., Monge-Sanz, B.M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N., Vitart, F., 2011. The ERA-Interim reanalysis: configuration and performance of the data assimilation system. Q.J.R. Meteorol. Soc. 137, 553–597. https://doi.org/10.1002/qj.828

Dee, D., J. Fasullo, D. Shea, J. Walsh, & National Center for Atmospheric Research Staff (Eds). Last modified 12 Dec 2016. "The Climate Data Guide: Atmospheric Reanalysis: Overview & Comparison Tables." Retrieved from https://climatedataguide.ucar.edu/climate-data/atmospheric-reanalysis-overview-comparison-tables.

Dessler, A.E., Davis, S.M., 2010. Trends in tropospheric humidity from reanalysis systems. J. Geophys. Res. 115, D19127. https://doi.org/10.1029/2010JD014192

Dirksen, R.J., Bodeker, G.E., Thorne, P.W., Merlone, A., Reale, T., Wang, J., Hurst, D.F., Demoz, B.B., Gardiner, T.D., Ingleby, B., Sommer, M., von Rohden, C., Leblanc, T., 2019. Progress in managing the transition from the RS92 to the Vaisala RS41 as the operational radiosonde within the GCOS Reference Upper-Air Network (preprint). Atmospheric instruments. https://doi.org/10.5194/gi-2019-36

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. Atmos. Meas. Tech. 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., 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

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., J. K. Angle, I. Durre, J. Lanzante, T. C. Peterson, and D. J. Seidel, 2004: Using first differences to reduce inhomogeneity in radiosonde temperature datasets, J. Clim., 17, 4171–4179.

GCOS. 2016. The Global Observing System for Climate: Implementation Needs, GCOS-200, Secretariat of the World Meteorological Organization: Geneva, Switzerland, pp. 315.

Gelaro, R., McCarty, W., Suárez, M.J., Todling, R., Molod, A., Takacs, L., Randles, C.A., Darmenov, A., Bosilovich, M.G., Reichle, R., Wargan, K., Coy, L., Cullather, R., Draper, C., Akella, S., Buchard, V., Conaty, A., da Silva, A.M., Gu, W., Kim, G.-K., Koster, R., Lucchesi, R., Merkova, D., Nielsen, J.E., Partyka, G., Pawson, S., Putman, W., Rienecker, M., Schubert, S.D., Sienkiewicz, M., Zhao, B., 2017. The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2). J. Climate 30, 5419–5454. https://doi.org/10.1175/JCLI-D-16-0758.1

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

Haimberger, L., Tavolato, C., Sperka, S., 2008. Toward Elimination of the Warm Bias in Historic Radiosonde Temperature Records—Some New Results from a Comprehensive Intercomparison of Upper-Air Data. Journal of Climate 21, 4587–4606. https://doi.org/10.1175/2008JCLI1929.1

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R.J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., Thépaut, J., 2020. The ERA5 global reanalysis. Q.J.R. Meteorol. Soc. 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. Geophys. Res. Lett. 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. Atmos. Meas. Tech. 8, 5157–5176. https://doi.org/10.5194/amt-8-5157-2015

Ingleby, B.: An assessment of different radiosonde types 2015/2016, Technical Memorandum, ECMWF Research Department, August 2017.

JCGM 100:2008 Evaluation of Measurement Data—Guide to the Expression of Uncertainty in Measurement. https://www.iso.org/sites/JCGM/GUM/JCGM100/C045315e-html/C045315e.html?csnumber=50461 (last access, 23/04/2021).

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. Atmos. Meas. Tech. 12, 3039–3065. https://doi.org/10.5194/amt-12-3039-2019

Kobayashi, S., Ota, Y., Harada, Y., Ebita, A., Moriya, M., Onoda, H., Onogi, K., Kamahori, H., Kobayashi, C., Endo, H., Miyaoka, K., Takahashi, K., 2015. The JRA-55 Reanalysis: General Specifications and Basic Characteristics. Journal of the Meteorological Society of Japan 93, 5–48. https://doi.org/10.2151/jmsj.2015-001

Loew, A., Bell, W., Brocca, L., Bulgin, C.E., Burdanowitz, J., Calbet, X., Donner, R.V., Ghent, D., Gruber, A., Kaminski, T., Kinzel, J., Klepp, C., Lambert, J.-C., Schaepman-Strub, G., Schröder, M., Verhoelst, T., 2017. Validation practices for satellite-based Earth observation data across communities: EO VALIDATION. Rev. Geophys. 55, 779–817. https://doi.org/10.1002/2017RG000562

Madonna, F., Kivi, R., Dupont, J.-C., Ingleby, B., Fujiwara, M., Romanens, G., Hernandez, M., Calbet, X., Rosoldi, M., Giunta, A., Karppinen, T., Iwabuchi, M., Hoshino, S., von Rohden, C., Thorne, P.W., 2020a. Use of automatic radiosonde launchers to measure temperature and humidity profiles from the GRUAN perspective. Atmos. Meas. Tech. 13, 3621–3649.

Madonna, F., Tramutola E.: RHARM (Radiosounding HARMonization) dataset - subset [Data set], Zenodo, http://doi.org/10.5281/zenodo.3973353, 2020a.

Madonna, F., Tramutola, E., Sy, S., Serva, F., Proto, M., Rosoldi, M., Gagliardi, S., Amato, F., Marra, F., Fassò, A., Gardiner, T., Thorne, P.W., 2020b. Radiosounding HARMonization (RHARM): a new homogenized dataset of radiosounding temperature, humidity and wind profiles with uncertainty (preprint). Data, Algorithms, and Models. https://doi.org/10.5194/essd-2020-183

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

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., Hollmann, R., Lavergne, T., Laeng, A., de Leeuw, G., Mittaz, J., Poulsen, C., Povey, A.C., Reuter, M., Sathyendranath, S., Sandven, S., Sofieva, V.F., Wagner, W., 2017. Uncertainty information in climate data records from Earth observation. Earth Syst. Sci. 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.

Nash, J., Smout, R., Oakley, T., Pathack, B., and Kurnosenko, S.: WMO Intercomparison of Radiosonde Systems, Vacoas, Mauritius, 2–25 February 2005, Tech. rep., WMO, WMO/TD-No. 1303, 2006.

Nash, J., T. Oakley, H. Vömel, LI Wei: WMO intercomparison of high quality radiosonde systems, WMO/TD-No. 1580, Yangjiang, China, 12 July – 3 August 2010. 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 Syst. Sci. Data 6, 185–200. https://doi.org/10.5194/essd-6-185-2014

Stone, M., 1974. Cross-Validatory Choice and Assessment of Statistical Predictions. Journal of the Royal Statistical Society: Series B (Methodological) 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: Lenz, H.-J., Wilrich, P.-T. (Eds.), Frontiers in Statistical Quality Control 7. Physica-Verlag HD, Heidelberg, pp. 79–100. 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. Int J Climatol 41. https://doi.org/10.1002/joc.6827

Thorne, P. W., D. E. Parker, et al., 2005. "Uncertainties in climate trends - Lessons from upper-air temperature records." Bulletin of the American Meteorological Society 86(10): 1437-+, http://journals.ametsoc.org/doi/abs/10.1175/BAMS-86-10-1437.

Thorne, P. W., P. Brohan, et al., 2011. "A quantification of uncertainties in historical tropical tropospheric temperature trends from radiosondes" Journal of Geophysical Research - Atmospheres, doi:10.1029/2010JD015487, http://onlinelibrary.wiley.com/doi/10.1029/2010JD015487.

Thorne, P.W., Madonna, F., Schulz, J., Oakley, T., Ingleby, B., Rosoldi, M., Tramutola, E., Arola, A., Buschmann, M., Mikalsen, A.C., Davy, R., Voces, C., Kreher, K., De Maziere, M., Pappalardo, G., 2017. Making better sense of the mosaic of environmental measurement networks: a system-of-systems approach and quantitative assessment. Geosci. Instrum. Method. Data Syst. 6, 453–472. https://doi.org/10.5194/gi-6-453-2017

Wang, J., L. Zhang, A. Dai, F. Immler, M. Sommer, and H. Vömel: Radiation dry bias correction of Vaisala RS92 humidity data and its impacts on historical radiosonde data. J. Atmos. Oceanic Technol., 30, 197–214, 2013, doi: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., Hurst, D.F., Hassler, B., English, J.M., Yorgun, S., 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

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