

The new Radiosounding HARMonization (RHARM) dataset of homogenized radiosounding temperature, humidity and wind profiles with uncertainties. Part II: comparisons with reanalysis, satellite data and validation of uncertainties

Fabio Madonna¹, Souleymane SY², Emanuele Tramutola¹, Federico Serva³, Monica Proto¹, Marco Rosoldi¹, Francesco Amato¹, Fabrizio Marra¹, Simone Gagliardi¹, Alessandro Fassò⁴, Tom Gardiner⁵, and Peter William Thorne⁶

¹Consiglio Nazionale delle Ricerche - Istituto di Metodologie per l'Analisi Ambientale (CNR-IMAA)

²Italian National Research Council (CNR)

³CNR-ISMAR

⁴University of Bergamo

⁵National Physical Laboratory

⁶Maynooth University

November 23, 2022

Abstract

The RHARM (Radiosounding HARMonization) algorithm is the first to provide homogenized radiosonde-based records of temperature, relative humidity and wind profiles since 1978, alongside an estimation of the observational uncertainty for each observation and pressure level. The algorithm and the dataset are presented in the companion paper. In this paper we assess the performance of the dataset through comparison with some of the most widely used climate data records. The RHARM adjustments reduce the difference with the reanalysis especially in the northern hemisphere for temperature and relative humidity. The study of temperature trends at different pressure levels reveals a good agreement between RHARM, reanalysis and existing radiosounding homogenized datasets ($<0.1\text{K}$ per decade above 300 hPa, 0.25K per decade below). For relative humidity, the discrepancies among the datasets are more significant, although RHARM trends are most similar to the reanalysis. For wind speed, the comparison indicates a good agreement above 300 hPa. Compared to IGRA, RHARM also improves by 50% the agreement with the estimated trends in the lower stratosphere from MSU (Microwave Sounding Unit) deep layer averages. For water vapour, the good performance of post-2004 RHARM data is quantified from the comparison of the 300 hPa monthly means in tropics between RHARM and AURA/MLS, as the absolute mean difference is 0.01 g/kg for RHARM and 0.03 g/kg for IGRA, and correlation increases from 0.95 to 0.99. A validation of the observational uncertainties of RHARM is also presented showing that they provide a good estimate or overestimate the theoretical distribution.

The new Radiosounding HARMonization (RHARM) dataset of homogenized radiosounding temperature, humidity and wind profiles with uncertainties. Part II: comparisons with reanalysis, satellite data and validation of uncertainties

Fabio Madonna¹, Souleymane SY¹, Emanuele Tramutola¹, Federico Serva², Monica Proto¹, Marco Rosoldi¹, Simone Gagliardi¹, Francesco Amato¹, Fabrizio Marra¹, Alessandro Fassò³, Tom Gardiner⁴, and Peter William Thorne⁵

¹Consiglio Nazionale delle Ricerche - Istituto di Metodologie per l'Analisi Ambientale (CNR-IMAA), Tito Scalo (Potenza), Italy

²Consiglio Nazionale delle Ricerche – Istituto di Scienze Marine (CNR-ISMAR), Rome, Italy.

³University of Bergamo, Bergamo, Italy

⁴National Physical Laboratory, Teddington, UK

⁵Irish Climate Analysis and Research Units, Department of Geography, Maynooth University, Maynooth, Ireland

Abstract

The RHARM (Radiosounding HARMonization) algorithm is the first to provide homogenized radiosonde-based records of temperature, relative humidity and wind profiles since 1978, alongside an estimation of the observational uncertainty for each observation and pressure level. The algorithm and the dataset are presented in the companion paper. In this paper we assess the performance of the dataset through comparison with some of the most widely used climate data records. The RHARM adjustments reduce the difference with the reanalysis especially in the northern hemisphere for temperature and relative humidity. The study of temperature trends at different pressure levels reveals a good agreement between RHARM, reanalysis and existing radiosounding homogenized datasets (<0.1K per decade above 300 hPa, 0.25 K per decade below). For relative humidity, the discrepancies among the datasets are more significant, although RHARM trends are most similar to the reanalysis. For wind speed, the comparison indicates a good agreement above 300 hPa. Compared to IGRA, RHARM also improves by 50% the agreement with the estimated trends in the lower stratosphere from MSU (Microwave Sounding Unit) deep layer averages. For water vapour, the good performance of post-2004 RHARM data is quantified from the comparison of the 300 hPa monthly means in tropics between RHARM and AURA/MLS, as the absolute mean difference is 0.01 g/kg for RHARM and 0.03 g/kg for IGRA, and correlation increases from 0.95 to 0.99. A validation of the observational uncertainties of RHARM is also presented showing that they provide a good estimate or overestimate the theoretical distribution.

1. Introduction

Observational data records are influenced by instrumental effects which may erode their reliability for climate applications (Kremser et al., 2018). Among in-situ measurement techniques, radiosounding balloons are widely used for the study of climate changes in the atmosphere. Several climate studies use radiosounding time series homogenized at mandatory pressure levels, using a range of algorithms based on the use of reference time series (mainly atmospheric reanalysis data), information from the nearest neighboring stations or kriging techniques (e.g. Free et al. 2004; Thorne et al., 2005; McCarthy et al., 2008; Sherwood et al. 2008; Dai et al., 2011; Haimberger et al., 2012). Some of these methods include ensembles to quantify parametric uncertainties (Haimberger et al., 2012; Thorne et al., 2011). Nevertheless, none of these algorithms provides a full quantification of measurement uncertainties.

Homogenization algorithms aim at detecting and adjusting for quantifiable systematic inhomogeneities in the observational time series, which for radiosoundings strongly depend on the sensor type and altitude level and vary substantially through time (e.g changes in station location, instrumentation, calibration or drift issues, different instrument sensitivity with respect to different networks, changes in the measurement procedures, etc.). Metadata can support the design or the validation of the algorithms, although for historical in-situ observations the metadata availability is poor and known to be incomplete (Durre et al., 2006, 2012; Ferreira et al., 2019).

In a companion paper (Madonna et al. 2021), a novel algorithm named RHARM (Radiosounding HARMonization) for the homogenization of historical radiosounding data records available since 1978 was introduced. The RHARM algorithm, designed and implemented under the Copernicus Climate Change Service (C3S), is based on a novel hybrid approach, combining physical (metadata-based) and statistical adjustments of radiosonde instrumental effects. RHARM is the first algorithm to provide in a single package homogenized time series of temperature, relative humidity and wind profiles alongside an estimation of the observational uncertainty for each observation and pressure level. Physical adjustments for the most recent radiosonde measurements (generally post-2004) are calculated following the data processing of radiosounding data adopted by GRUAN (Global Reference Upper-Air Network, Bodeker et al., 2016) and inferring information from the WMO 2010 radiosonde intercomparison (Nash et al., 2011). Historical observations before 2004 are homogenized using a statistical methodology allowing quantification of the measurement uncertainties. A detailed description of the RHARM approach is provided in the companion paper by Madonna et al. (2021).

In this paper we investigate the quality of the RHARM dataset for climate applications for temperature, relative humidity and winds, via a number of comparisons with several widely used climate data records: the ERA5 reanalysis (Hersbach et al., 2020); the homogenized dataset RADiosonde OBServation CORrection using REAnalyses (RAOBCORE, Haimberger et al., 2008), the Radiosonde Innovation Composite Homogenization (RICH) by the University of Wien (Haimberger et al., 2012), the Iterative Universal Kriging v2 (IUKv2) by University of New South Wales (Sherwood and Nishant et al., 2015); the Homogenized RS92 radiosounding humidity measurements (HomoRS92) by the University of Albany (Dai et al., 2011); the Microwave Sounding Unit (MSU) elaborated, by the Global Hydrology and Climate Center at the University of Alabama in Huntsville (UAH – Spencer and Christy, 1990; Spencer, 2016) and by Remote Sensing Systems (RSS - Mears and Wentz, 2009), respectively, and the Microwave Limb sounder (MLS) water vapour data records (Yan et al., 2016).

In addition, a study has been performed using the approach described in Merchant et al. (2017) and exploiting the ERA5 background data as a diagnostic tool to validate the credibility of the estimated uncertainties.

The remainder of this paper is organized as follows. In section 2, the data sources used in the paper are outlined. In section 3, the comparison between RHARM and ERA5 is discussed. In section 4, intercomparisons of RHARM with the other existing homogenized radiosonde datasets mentioned above is undertaken. In section 5, RHARM is compared with MSU and MLS data. In section 6, the RHARM uncertainties are validated using the ERA5 forecast data. A discussion and conclusions are provided in Section 7.

2. Data sources and methodology

An important step in the performance assessment of the RHARM data is the comparison with independent datasets. The data records we consider do not include any truly “Reference” dataset (Thorne et al., 2017) and therefore none of the compared dataset can be assumed as the truth. Hence these comparisons can elucidate only relative rather than absolute characteristics of the RHARM product.

The RHARM dataset is described at length in the companion paper, and here we summarize its main features. The RHARM algorithm for the homogenization of global radiosounding temperature, humidity and wind profiles is applied to per-ascent (generally 00:00 and/or 12:00 UTC) radiosonde data on 16 mandatory pressure levels (10, 20, 30, 50, 70, 100, 150, 200, 250, 300, 400, 500, 700, 850, 925, 1000 hPa). Relative humidity (RH) adjustments are limited to 250 hPa owing to pervasive sensor performance issues at greater altitudes. Profiles are adjusted at these mandatory pressure levels. The applied adjustments are then interpolated to significant levels. Uncertainties are estimated for each processing step and then propagated to estimate the total uncertainty.

2.1 Homogenized radiosonde datasets

The Radiosonde Observation Correction Using Reanalyses (RAOBCORE) dataset (Haimberger, 2007) provides homogenized radiosounding time series at 0000 – 1200 UTC with bias adjustments based on the radiosonde O-B innovation statistics, i.e. the differences between observations and the state vector of the background forecast by the ECMWF ERA-interim reanalysis. RAOBCORE adjustments have been used for radiosonde bias correction in the interim ECMWF Re-Analysis (Dee et al. 2011), as well as in the Modern-Era Retrospective Analysis for Research and Applications (MERRA; Rienecker et al. 2011), and the Japanese 55-year reanalysis (JRA-55, Kobayashi et al. 2015). To overcome the dependence problem, Haimberger et al. (2008) further developed a method called Radiosonde Innovation Composite Homogenization (RICH), which uses the breakpoint date information from RAOBCORE, but calculates the break size estimates by comparison with neighboring radiosonde temperature records. Current versions of these datasets are produced using the ECMWF ERA5 background forecasts. These versions, labelled as ERA5_v4, have been used in this paper.

The IUKv2 is an updated version of the radiosonde temperature dataset homogenized by Iterative Universal Kriging (IUKv2), following the method used in the first version (Sherwood et al., 2008). The methodology statistically corrects for incomplete sampling and step changes in observational

biases. This method is based on a multiple linear regression of the data onto a structural model that includes natural variability, trends, and time-changing instrument biases. Homogenized winds are also provided.

The homoRS92 dataset contains daily radiosonde humidity data at standard pressure levels from January 1945 to the end of 2010. The approach homogenizes historical records of tropospheric (up to 100 hPa) dewpoint depression (DPD). Changepoints are detected using a variant of the Kolmogorov–Smirnov (K–S) test and the penalized maximal F test (PMFred) for mean shifts in the occurrence frequency for different bins of the daily DPD. Then the time series are adjusted using a quantile-matching algorithm (Dai et al., 2011). For 63 stations during the period when Vaisala RS92 was used, 2006-2010 (Wang et al. 2013), profiles are adjusted for the well-known (e.g., Dirksen et al., 2014) solar radiation dry bias using a method similar to the current GRUAN Data Processing (GDP). Then, the data before 2006 are homogenized using the quantile-matching algorithm.

All the dataset described above have been sub-sampled to match the location of RHARM stations using the nearest grid point to each station. For RICH and RAOBCORE, being not gridded, a station-by-station matching has been carried out, covering almost all the RHARM stations.

2.2 ERA5

The ECMWF ERA5 reanalysis (Hersbach et al., 2020) incorporates millions of observations into a data assimilation system, every 6-12 hours over the period being analyzed, providing a systematic approach to produce data sets for climate monitoring and research. ERA5 is the latest climate reanalysis produced by ECMWF providing hourly data on regular latitude-longitude grids at 0.25° x 0.25° resolution and on 37 pressure levels. ERA5 is publicly available through the Copernicus Climate Data Store (CDS, <https://cds.climate.copernicus.eu>). For the purposes of the validation of uncertainties, we also use the ERA5 background (6-hours forecast) as a reference value. The various reanalysis products have proven to be valuable when used appropriately (Dee et al., 2011). Nevertheless, reanalysis reliability can considerably vary depending on the location, time period, and variable considered (Dee and Fasullo, 2016). The changing mix of observations, and biases in observations and models, can introduce spurious variability and trends into reanalysis output (Dee and Fasullo, 2016). However, the ERA5 data is certainly one of the most widely used dataset for climate applications, with a growing popularity.

For the comparison with RHARM data, which are not gridded, ERA5 data obtained from the Copernicus Climate Data Store (CDS) have been sub-sampled to match the location of RHARM stations using the nearest grid point to each station. The same approach has been used for the ERA5 background data considered for the uncertainty validation.

2.3 Satellite data

Data from the combination of the Microwave Sounding Units (MSU) and its follow-on Advanced Microwave Sounding Unit (AMSU), have been used for comparison with the temperature anomalies in the lower stratosphere since December 1978. The MSUs measure radiation emitted by the Earth's atmosphere from NOAA polar orbiting satellites. with a spatial resolution on the ground of 2.5° in longitude and latitude (about 250 km before 1998 and 50 km afterwards). The different channels of the MSU measure different frequencies of radiation proportional to the temperature of broad vertical layers of the atmosphere. MSU temperatures are adjusted for time-dependent biases. The

two most used homogenized datasets are provided by the Global Hydrology and Climate Center at the University of Alabama in Huntsville (UAH), and by Remote Sensing Systems (RSS). For our analysis, lower stratospheric temperature (LS) data from MSU bias-adjusted data have been compared with IGRA and RHARM data at 50 hPa temperature time series, assuming this pressure level as the most representative for the comparison with the LS.

For water vapour, data from the Microwave Limb Sounder (MLS) on-board the Aura satellite have been used. For the comparison with RHARM, the MLS data version 4 have been used (Yan et al., 2016). For the comparison shown in this paper with RHARM data we considered only the MLS data at 316 hPa RH time series and these have been subsampled at the IGRA stations using the nearest grid points and compared to 300hPa standard pressure level reports.

3. Comparisons with ERA5

In this section, we compare the tropospheric interannual variations for the period 1979-2018 as represented in radiosonde datasets with that from the ERA5 reanalysis. Monthly anomalies of temperature, RH and wind speed from IGRA and RHARM are compared with the results obtained for ERA5, having selected the nearest grid-point to each station (Figures 4-6). Simultaneous vertical profiles on 12 UTC and 00 UTC at 300 hPa and 500 hPa pressure levels are selected. Considering the high spatial resolution of ERA5 and its representativeness, the uncertainty due to the use of the nearest grid-point interpolation should be comparable with other methods (e.g. kriging, bilinear interpolation, etc.).

For temperature (Figures 1,2) in the NH, IGRA, RHARM and ERA5 show a similar positive decadal trend of 0.38, 0.39 and 0.43 K/decade, respectively, while in the tropics at 300 hPa the trend is of 0.17, 0.25, 0.20 K/decade, with a more pronounced trend increase starting around 1997. Similar results have been obtained considering European stations only (Madonna, 2020c). In the NH, the comparison of the anomalies at 300 hPa shows the evident adjustment applied by RHARM on the IGRA data over 1996-2005 (corresponding to period of the RS90 radiosondes usage) which reduces the difference from ERA5 results. Differences between RHARM and ERA5 are generally smaller than 0.5 K in absolute value, in both the regions, with larger difference for the most extreme values especially at the tropics. At 500 hPa the situation is very similar although in the NH the differences are smaller, while at the tropics results are in line with those for the 300 hPa pressure level.

For relative humidity (Figures 3,4), in the NH the substantive adjustment applied to IGRA by RHARM at 300 hPa before 1986 (up to 10 %RH) largely improves the agreement with ERA5. In 1986, a few major changes occurred in the global radiosounding data, the most relevant of which are: changes in several radiosonde models, such for MARS/MRZ and VIZ radiosondes; the adoption of new manufacturers at some stations, mainly changes from another manufacturer to Vaisala, and changes in the dewpoint depression algorithm, for example at UK stations; and, maybe the most important, the introduction of "pre-baselined" radiosondes, i.e. removal of the practice of applying a manual baseline lock for all temperature and RH profiles which was discovered to be prone to producing a wet bias in all the RH values smaller than 60% (more details at https://library.wmo.int/doc_num.php?explnum_id=9592). Since 2004 the adjustment applied by RHARM is smaller and further improves the agreement with ERA5 with a negative trend of -0.8 % RH/da. In the last decade, the trends show a change with a slight increase which has been already quantified in the European domain (Madonna, 2020c). At 500 hPa the situation is very similar although the adjustments are much smaller. In the tropics, the adjustments applied by RHARM at 300 and 500 hPa are smaller than in the NH. The comparison with ERA5 shows that the largest differences are at 500 hPa (up to 4-5 %RH). The comparison highlights major differences in three

periods: before 1990, where ERA5 negative anomalies are smaller; after 2005, with RHARM anomalies larger than those of ERA5; and after 2015, when differences are larger especially at 500 hPa. The comparison in these three periods generally reflects the higher variability of the observational time series compared to ERA5 data.

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

For wind speed, the comparison between observational data and ERA5 (Figures 5,6) shows a good agreement in the NH at 300 hPa, with a positive decadal trends of 0.34, 0.33 and 0.07 m/(s-1 per decade) for IGRA, RHARM and ERA5, respectively. In the tropics, trends at 300 hPa are negative but in much better agreement, with values of -0.06, -0.10, 0.06 m/(s-1 per decade). Similar considerations can be made for the anomalies at 500 hPa. In the NH, there is a good agreement also in the long-term variability and peak values. In the tropics, IGRA and RHARM exhibit larger variability than ERA5 although with significant differences for the extreme values.

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

Despite a degree of non-independence, the comparison with ERA5 reveals discrepancies in the monthly anomalies and trends with both IGRA and RHARM, although the adjustments applied in the RHARM data allow to reduce the difference between ERA5 and the observations, especially for temperature and RH in the NH. ERA5 performances in reproducing the observed atmospheric variability appear to be higher in the NH than in the tropics, likely due to the stronger observational constraints. Considering the difference in the RHARM homogenization approach for the data before and after 2004 (based on a statistical method and on a GRUAN-like post-processing, respectively), the high-frequencies of time series anomalies are very similar for IGRA and RHARM: this is expected by construction for the data before 2004, where only the bias in between each pair of breakpoints is adjusted, while for data after 2004 the adjustments, although affecting the overall underlying data distribution, are typically of small magnitude due to the enhanced quality of recent radiosonde data compared to historical observations (Thorne et al., 2011).

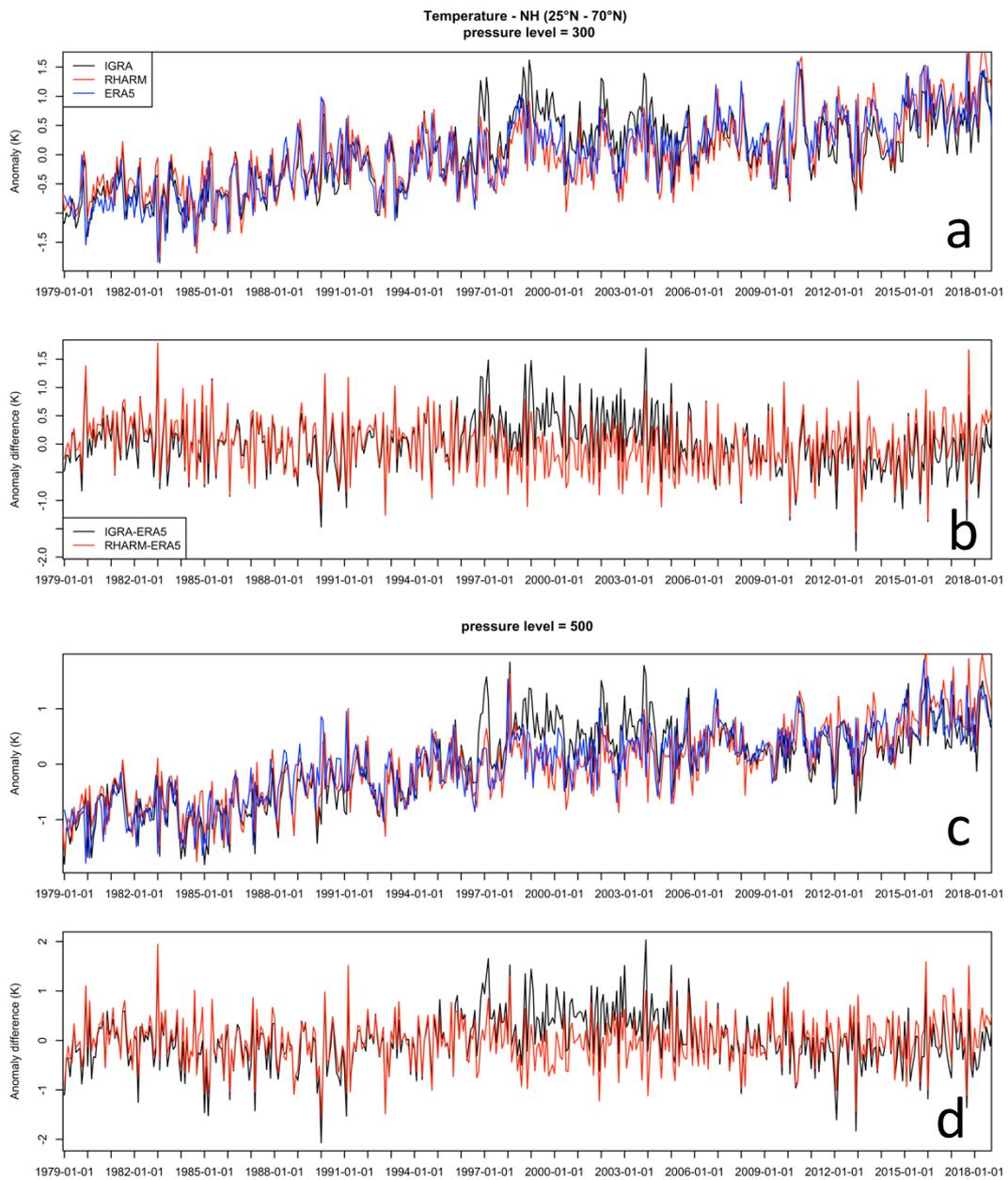


Figure 1: monthly temperature anomalies (i.e deviation from the mean created by subtracting climatological monthly means) and anomaly differences calculated between 01/01/2006 and 01/08/2019 for IGRA (black), RHARM (red) and ERA5 reanalysis (blue) at 300 and 500 hPa for Northern Hemisphere. Anomalies are shown in the panels a and c, while anomaly difference in the panels b and d.

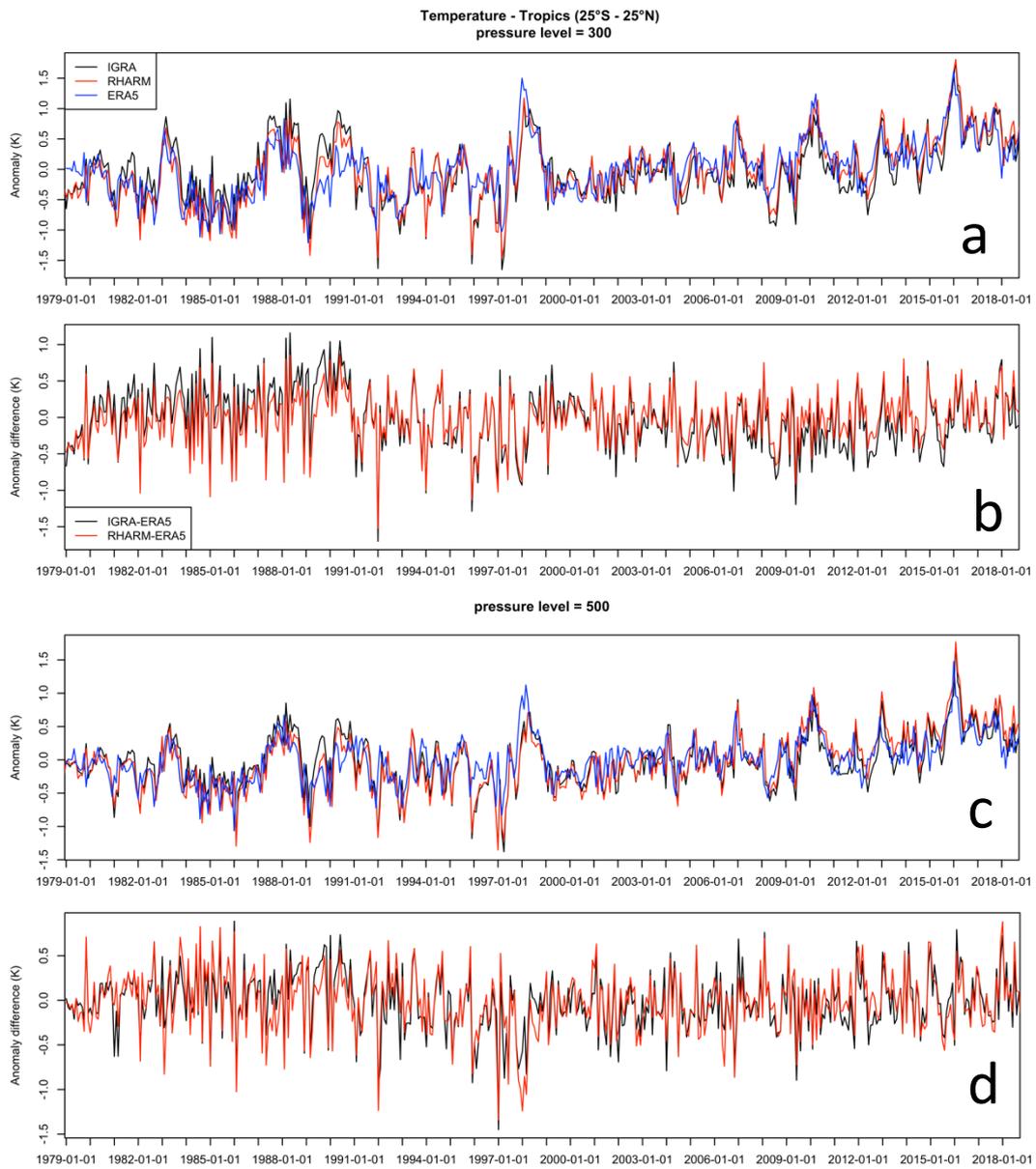


Figure 2: same as Figure 1 but for the tropics.

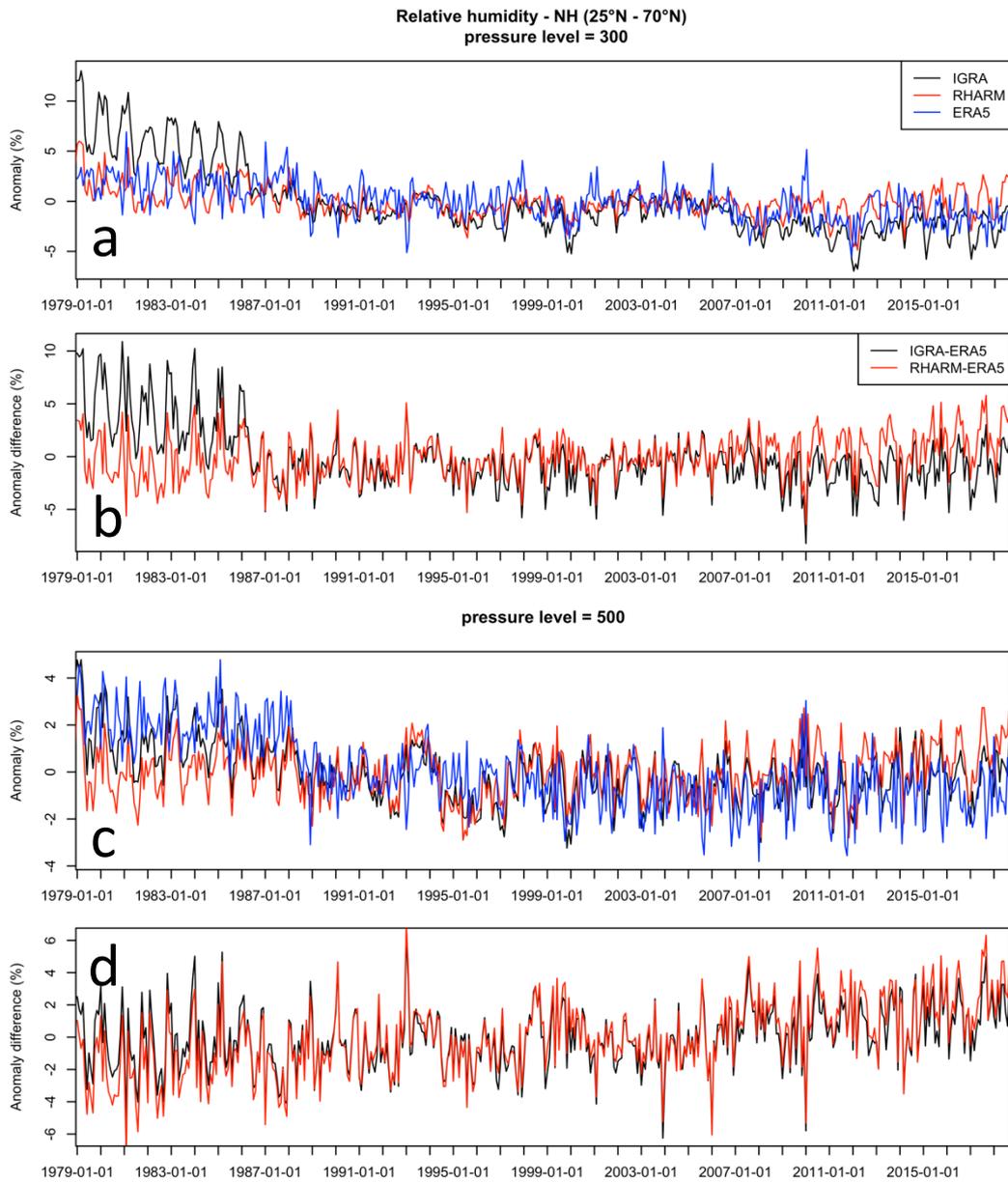


Figure 3: same as Figure 1 but for relative humidity.

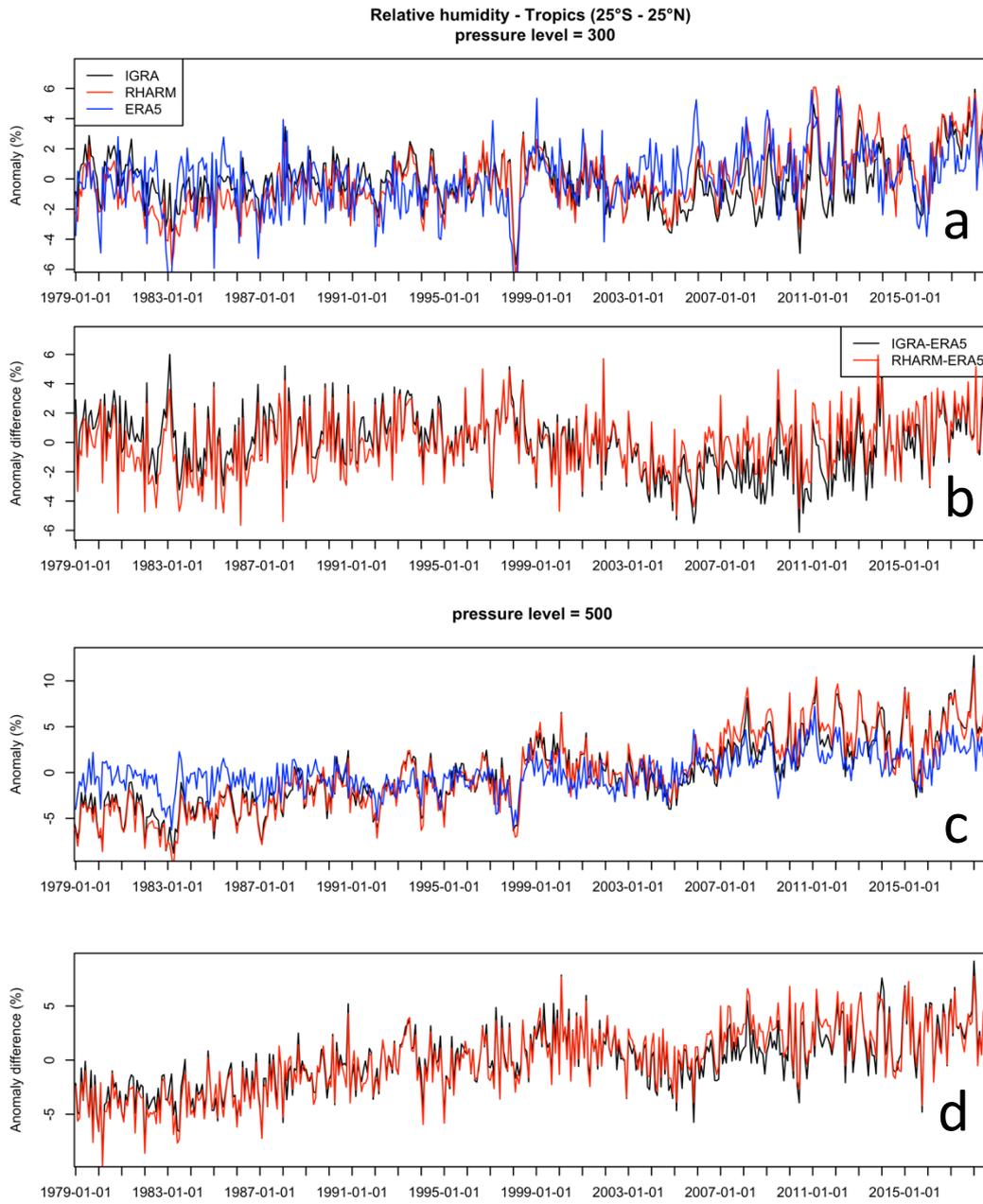


Figure 4: same as Figure 2 but for relative humidity.

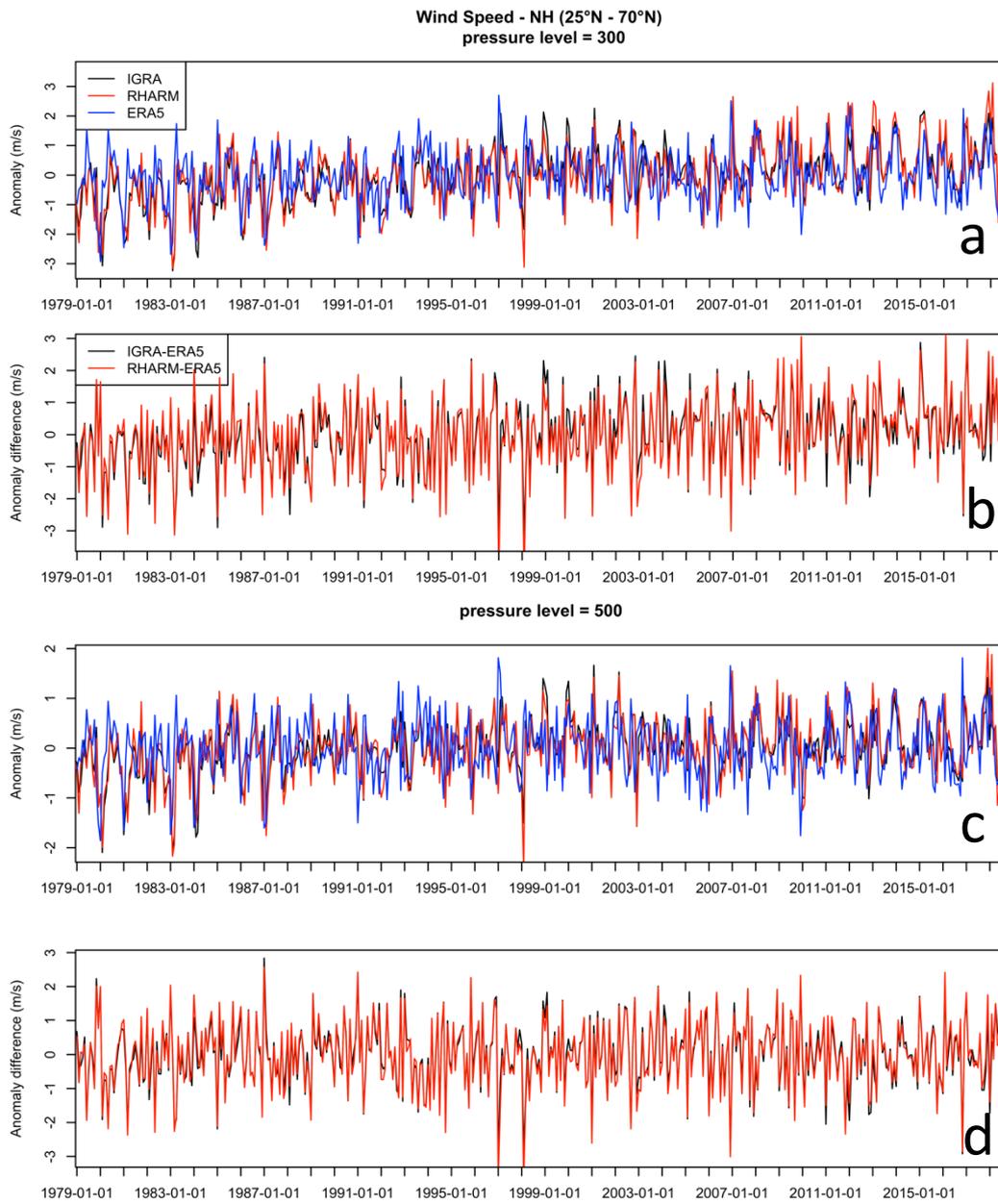


Figure 5: same as Figure 1 but for wind speed.

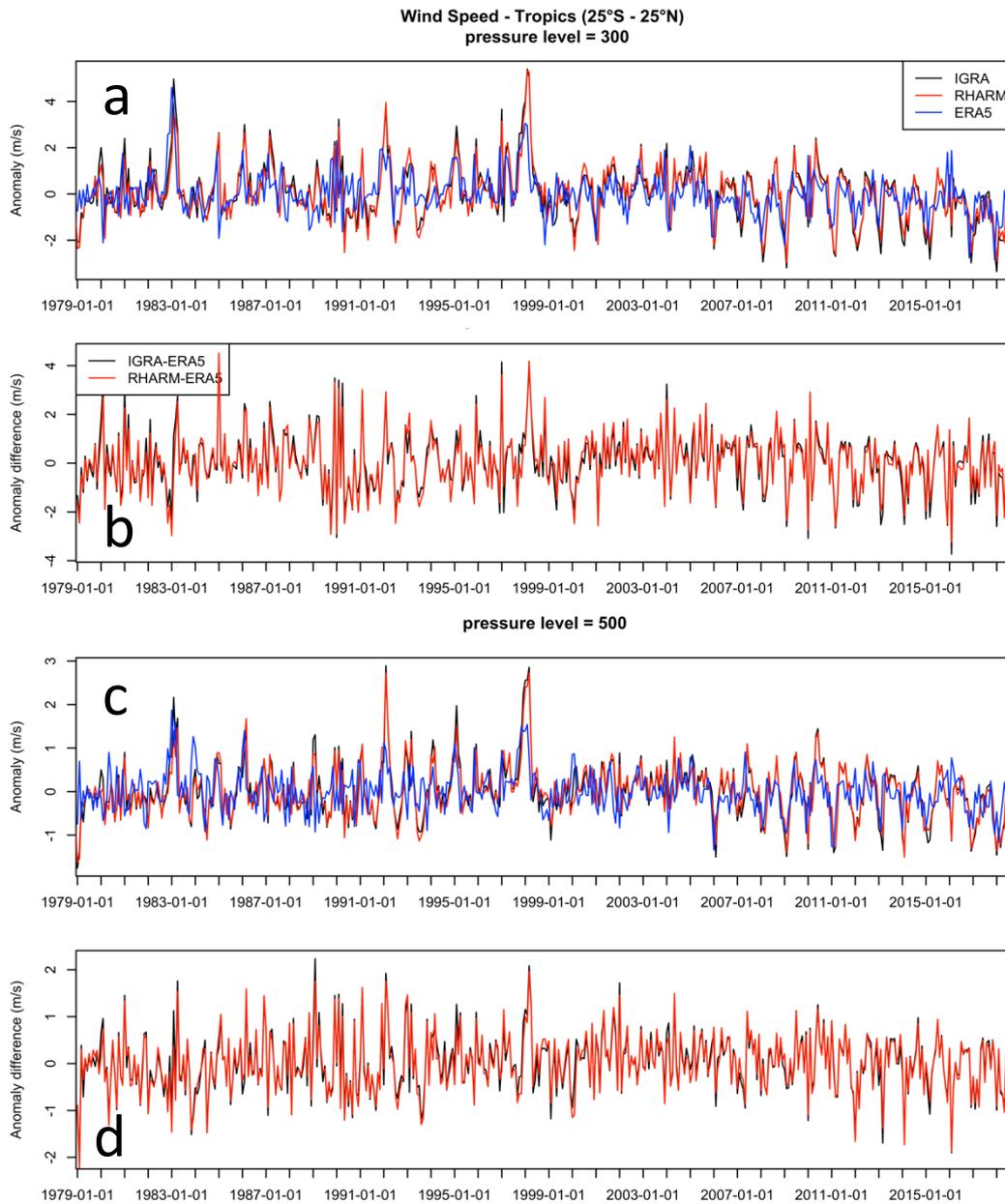


Figure 6: same as Figure 2 but for wind speed.

4. Comparisons with existing homogenized datasets

4.1 Temperature

In the NH, the comparison with existing homogenized datasets for temperature shows a very good agreement between all the datasets at pressure levels up to 200 hPa (Figure 7, left), with relative differences within 0.1 K/decade, except for IUKv2 which oscillates between -0.5 and 1.5 K/decade. Above 200 hPa, all the datasets are in agreement within 0.1 K/decade. Between 200 hPa and 70 hPa, RAOBCORE and RICH are the closest to ERA5, while above 70 hPa the ERA5 values overlap with IGRA. In the same altitude range RAOBCORE, RICH and RHARM are highly similar. IUKv2 data are

limited to 30 hPa and this is the only observational gridded dataset, with a consequent additional uncertainty due to the interpolation to the nearest grid point.

In the tropics (25°S-25°N), the shape of the trend vertical profiles is similar for all datasets, with IGRA the coldest and IUKv2 the warmest at all pressure levels. Up to 300 hPa, trends are positive (tropospheric warming) and their difference does not exceed 0.25 K/decade. In the range 300 hPa – 70 hPa RHARM, ERA5, RAOBCORE and RICH are very close, within 0.1 K/decade, while above this level ERA5 is closer to RAOBCORE and RHARM is closer to RICH, respectively. The IUKv2 is warmer than any other dataset above 300 hPa.

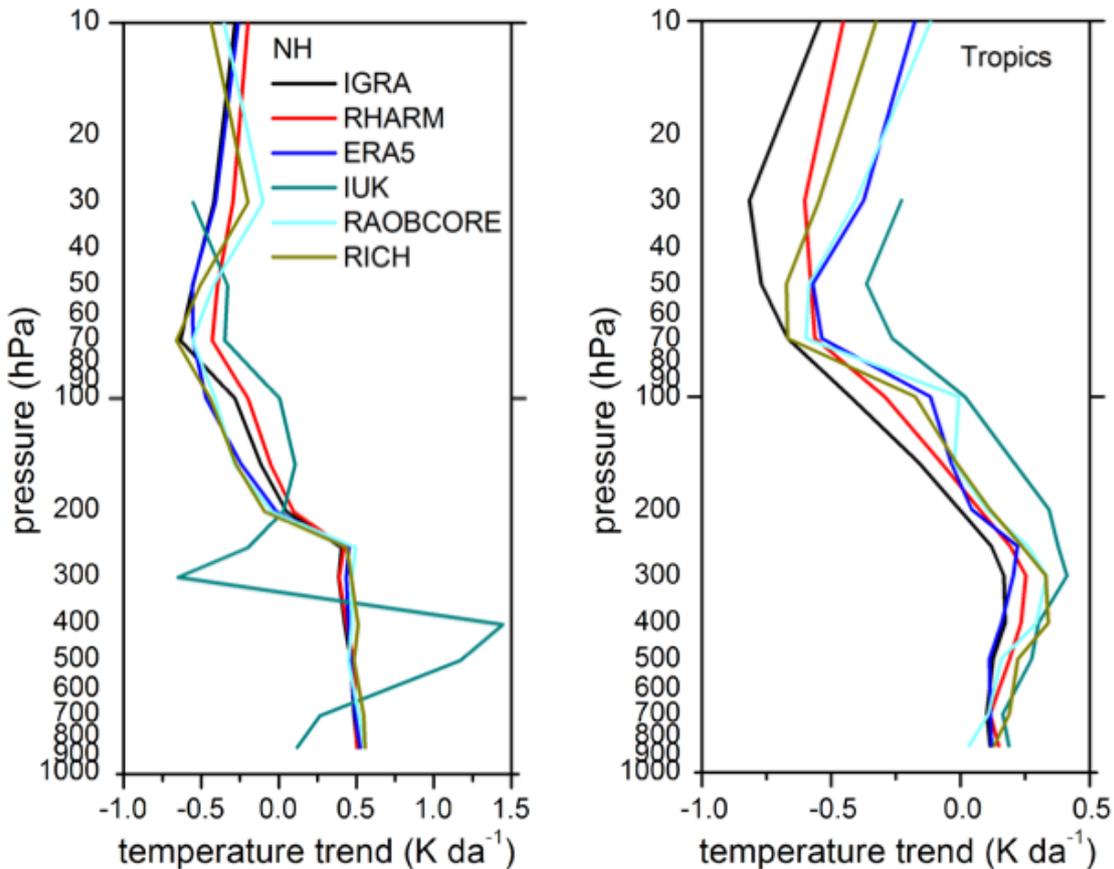


Figure 7: profiles of temperature decadal trends at mandatory pressure level between 850 and 10 hPa for the period 1979-2018, in the northern hemisphere (left panel) and in the tropics (right panel) for the unadjusted IGRA (black line), RHARM (red), ERA5 (blue), IUKv2 (green), RAOBCORE (cyan), RICH (dark yellow) datasets. The ordinate is logarithmic, and the abscissa differs between the two panels.

4.2 Relative humidity

In literature, there are two examples available in literature of homogenized datasets of relative humidity radiosounding profiles: these are HadTH (McCarthy et al., 2009) and the HOMORS92 dataset (Dai et al. 2011). HadTH provides a time series of anomalies relative to the monthly 1981-2000 climatology at standard pressure levels up to 300 hPa for stations in the NH only. As stated in McCarthy et al. (2009), only in the NH there is sufficient data for the application of the HadTH homogenisation method. Considering that HOMORS92 dataset has a longer time coverage and it is not limited to NH only, this is preferred to perform a comparison with RHARM.

Relative humidity (RH) is adjusted directly in RHARM whereas dew point temperature (DPT) is adjusted in HOMORS92. RH is derived from DPT, which is the variable measured by radiosondes. For this comparison, the values of relative humidity have been used: for HOMORS92 these are available in the dataset (albeit the formula to calculate DPT is not declared), while RHARM converts the IGRA DPT using the Magnus-Teten formula. Moreover, the HOMORS92 data are available at about 1500 IGRA stations, therefore for the comparison RHARM stations only have been selected.

In Figure 8, the comparison of the vertical profiles of RH trends, between 850 hPa and 300 hPa in the NH (left panel) shows that RHARM, ERA5 and HOMORS92 have a similar shape with relative differences around 1% RH/da throughout the entire vertical range, although RHARM is the only with positive values (near zero) above 500 hPa. Both datasets significantly differ from the unadjusted IGRA data for pressures below 500 hPa. In the tropics, HOMORS92 displays negative trends (slightly below zero) at all pressure levels, while RHARM and ERA5 shows a similar shape despite a difference up to 1.5% RH/decade, increasing with height. Differently from temperature, ERA5 RH assimilated data are not bias adjusted. HOMORS92 uses a quantile matching algorithm to adjust the DPT histograms in between different changepoints of the time series, not fully preserving seasonal variations (Wang et al., 2013). Instead, RHARM is designed to adjust systematic effects preserving trends and natural (i.e., internal) variability. The comparison cannot ascertain which of the datasets provides the best option to assess RH trends, although some features are apparent. The vertical profiles of ERA5 and RHARM are the most similar in terms of vertical structure while HOMORS92 and ERA5 have the smallest absolute differences, especially in the tropics, while in the NH they have opposite signs. Fundamentally, the paucity of available estimates makes it difficult to assess structural uncertainties.

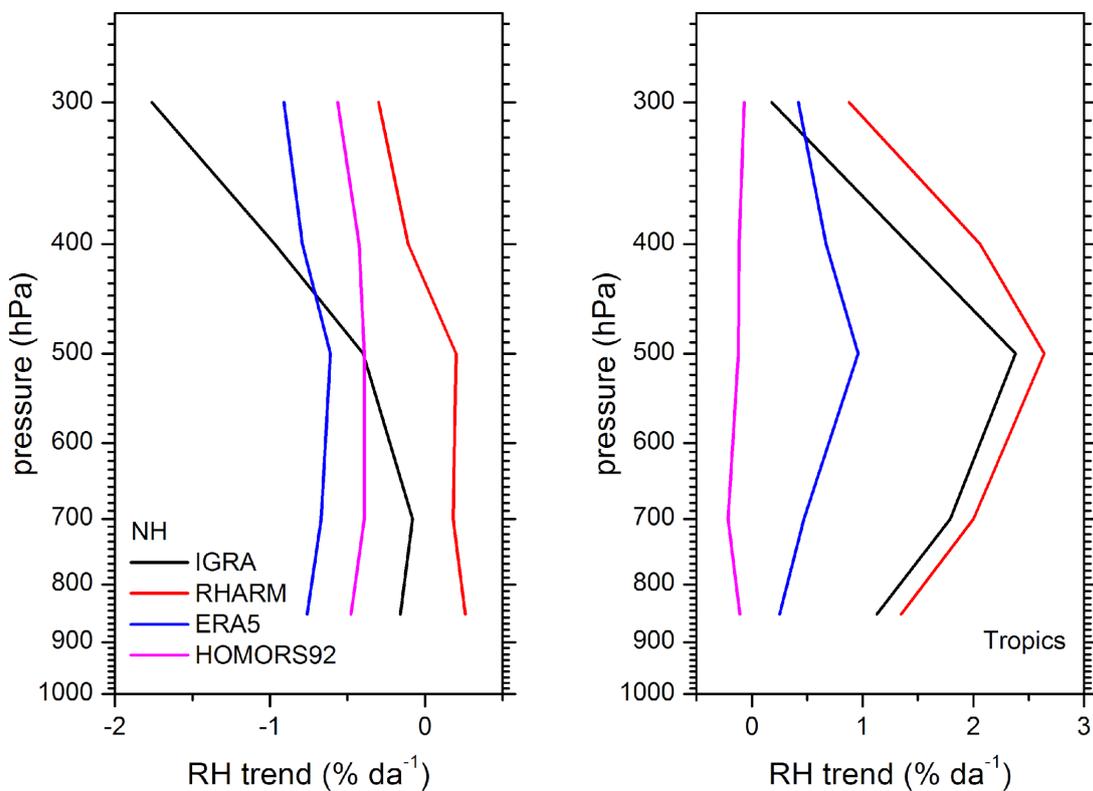


Figure 8: profiles of relative humidity decadal trends at mandatory pressure level between 850 and 300 hPa, in the period from 1979 to 2018, in the northern hemisphere (left panel) and in the tropics (right panel) for unadjusted IGRA (black line), RHARM (red), ERA5 (blue), and HOMORS92 (green). The ordinate is logarithmic, and the abscissa differs between the two panels. The comparison is limited to 300 hPa at lower pressure the comparison, and more in general the water vapour measurements, might be compromised by pervasive issues on radiosonde sensors' performance in the upper troposphere and lower stratosphere, even larger for the oldest measurements.

4.3 Wind speed

In the NH, below 70 hPa and above 300 hPa the comparison shows differences among the datasets (Figure 9) within $0.5 \text{ m}/(\text{s}^{-1} \text{ da}^{-1})$, with a better agreement between the observational datasets below 70 hPa. Between these pressure levels, the agreement between ERA5 and IUKv2 is generally better, especially between 70 hPa and 300 hPa where IGRA and RHARM show trends larger than $1.0 \text{ m}/(\text{s}^{-1} \text{ da}^{-1})$, with a small adjustment applied by RHARM.

In the tropics, the shape of the trend vertical profiles is almost the same for all datasets up to 300 hPa, while below this pressure the observational datasets shows similar positive trends, around $0.1 \text{ m}/(\text{s}^{-1} \text{ da}^{-1})$, differently from ERA5 which is more negative with values up to $-0.6 \text{ m}/(\text{s}^{-1} \text{ da}^{-1})$. Note that observational constraints in the reanalysis are weaker in this region (e.g., Kawatani et al., 2016), and artefacts due to changes in assimilated observations cannot be excluded. Below 50 hPa there is a very good agreement between RHARM and IUKv2.

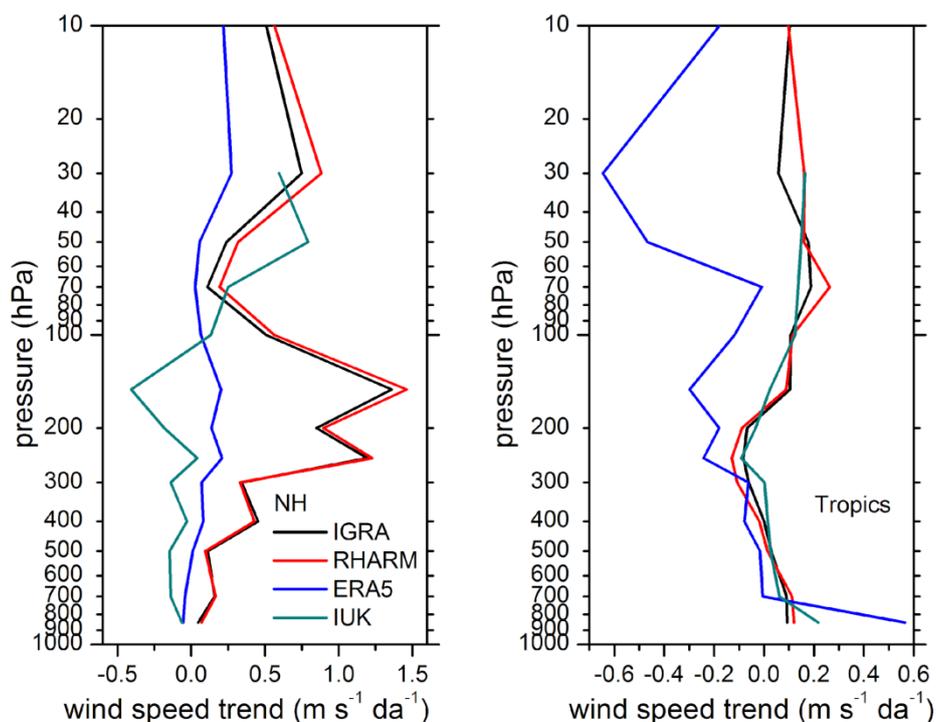


Figure 9: profiles of wind speed decadal trends at mandatory pressure level between 850 hPa and 10 hPa, in the period from 1979 to 2018, in the northern hemisphere (left panel) and in the tropics (right panel) for unadjusted IGRA (black line), RHARM (red), ERA5 (blue), and IUKv2 (green). The ordinate is logarithmic, and the abscissa differs between the two panels.

5. Comparisons with satellite data

5.1 Comparison with MSU

The comparison between IGRA, RHARM, UAH and RSS datasets has the aim to evaluate the consistency among the longest temperature data records in the lower stratosphere available from ground and satellite observations. Comparisons in the middle and low troposphere are not shown because, at the corresponding pressure levels, the RHARM adjustments are smaller and the effect

on the comparison with UAH and RSS less evident. The comparison shown in Figure 10 for the period 1979-2020 reveals a generally good agreement in the temperature variability between these datasets. All the time series show a cooling trend in stratospheric temperatures until 2000, followed by little further change. RHARM generally reduces the gap between satellite retrievals and radiosounding observations. To quantify the gap reduction between ground-based and satellite observations with RHARM, the linear decadal trends estimated using the LADFIT reported in Table 3 along with the related Mean Absolute Deviations (MADs). Differences in the trends estimated by radiosounding and satellite observations, both in the NH and at the tropics, are reduced by 50% for RHARM compared to IGRA. Decadal trends calculated over the entire time series are all negative with values ranging from -0.24 to -0.66 K/decade in the NH, and from -0.20 to -0.83 K/decade in the tropics. The MAD are larger for IGRA and RHARM than for satellite datasets indicating a larger variability in the anomalies measured with the radiosounding than with the satellite observations, likely related to effect of the coarser resolution of the satellite products which would smooth the most extreme values.

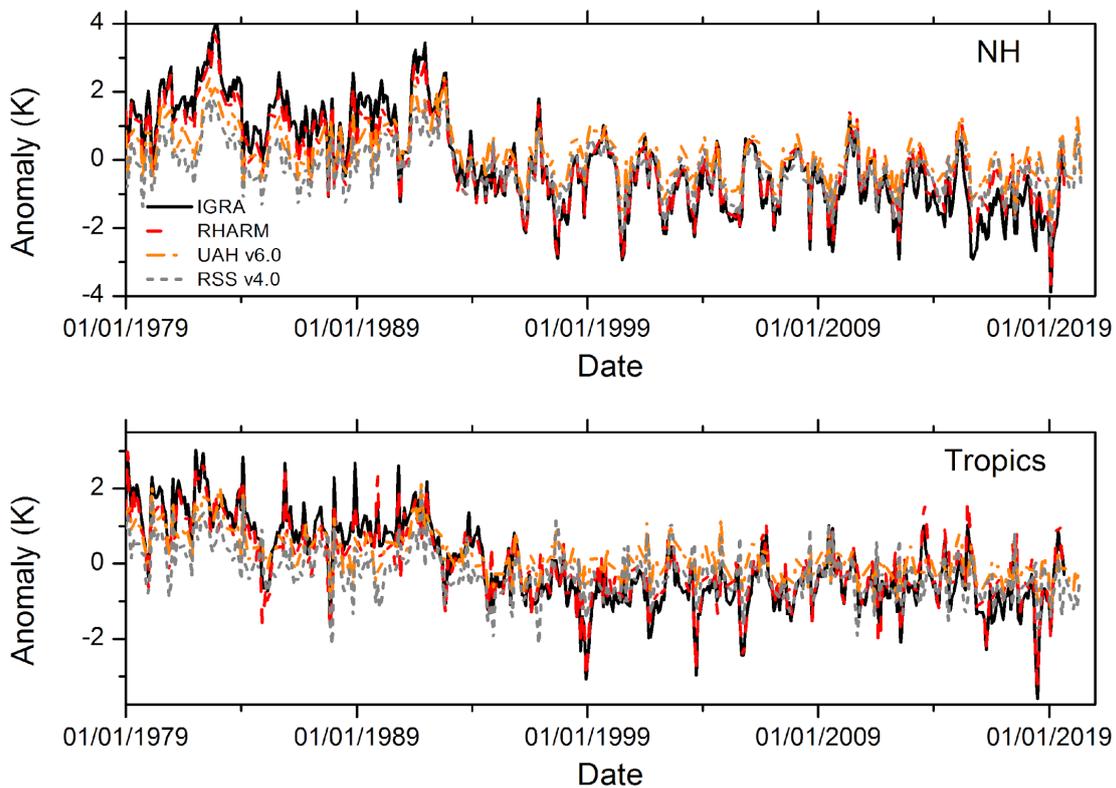


Figure 10: time series of the temperature monthly anomalies since 1979 to 2020 in the NH (top panels) and in the tropics (bottom panel) for the 50 hPa pressure level for IGRA (black line), RHARM (red dashed line), UAH v6.0 (orange dash-dotted line and RSS v4.0 (gray small-dashed line).

Table 1: Decadal linear trends estimated with the LADFIT regression method for the datasets shown in both panels of Figure 10. Median absolute deviations of the residuals are also reported for each trend.

NH	Trend	MAD
IGRA	-0.66	0.62
RHARM	-0.45	0.6
UAH	-0.28	0.37
RSS	-0.24	0.44
Tropics	Trend	MAD
IGRA	-0.83	0.78
RHARM	-0.59	0.77
UAH	-0.27	0.52
RSS	-0.2	0.53

5.2 Comparison with MLS

To quantify the adjustments of RHARM on RH, we compare its water vapour mixing ratio with the data from the Aura MLS instrument. The Aura MLS time series at the nominal 316 hPa pressure level is chosen being the closest to the mandatory level at 300 hPa for the IGRA/RHARM radiosondes, and because it is in good agreement with the Cryogenic FrostPoint Hygrometer (CFH) records from Boulder (Vömel et al., 2007). CFH is currently the only reference traceable instrument available for water vapour measurements in the Upper Troposphere/Lower Stratosphere (UT/LS). The possible difference in the water vapour content between the two levels at 316 and 300 hPa must be considered as an uncertainty contribution to the comparison. For the period 2006-2019, Figure 11 reveals a good agreement between RHARM and MLS, and it highlights the efficacy of the RHARM dry bias correction, especially tangible at the lowest water vapour concentrations. RHARM performs better in terms of mean offset, RMS differences and correlation than IGRA, as summarized in Table 2.

Tropics 0-25°N, 2006-2019

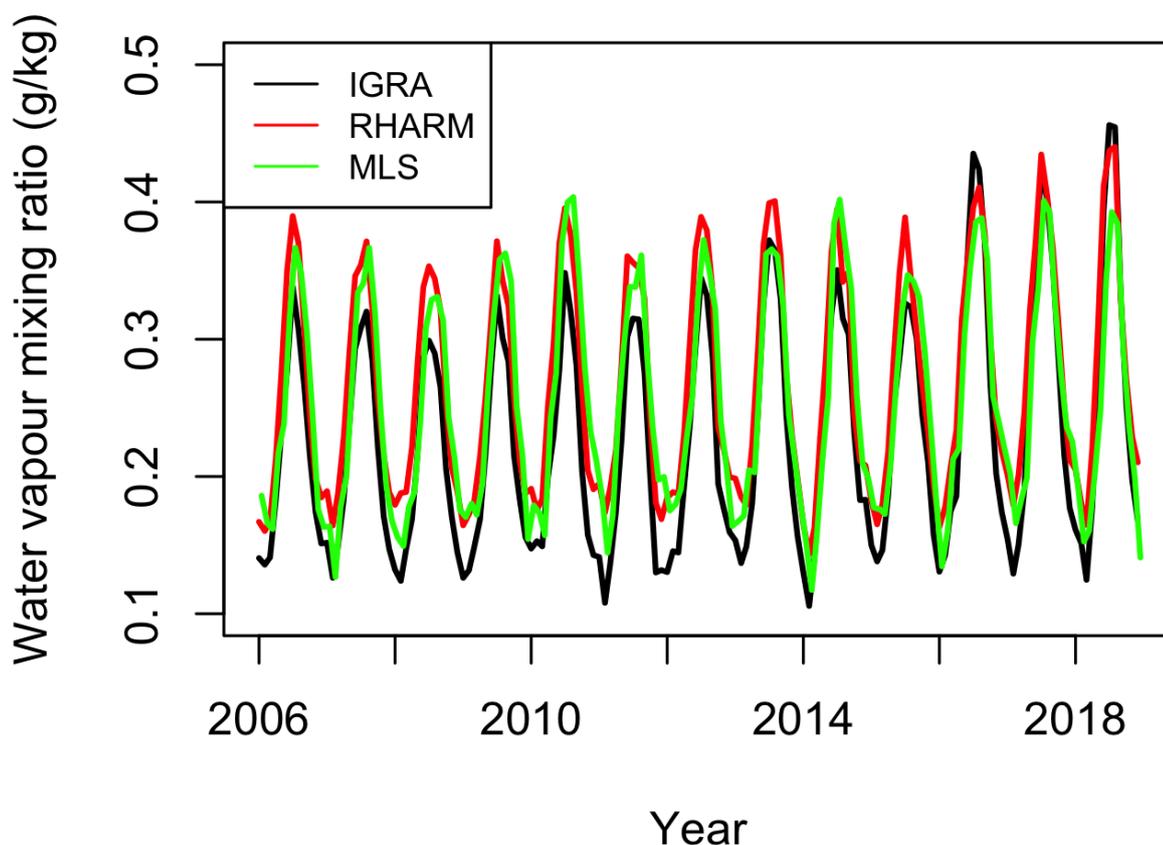


Figure 11: Comparison of monthly time series of the zonal mean water vapour mixing ratio in the northern tropics (0°N-25°N) from 2006 to 2019. The black lines show the IGRA time series, while the red and the green lines are the corresponding RHARM and MLS time series. AURA MLS data have been subsampled at the IGRA stations within the considered domain. The MLS data are version 4 have been used (Yan et al., 2016).

Table 2: statistics on the comparison between IGRA-MLS and RHARM-MLS for the data shown in Figure 11.

	mean difference (g/kg)	rms difference (g/kg)	correlation
IGRA-MLS	-0.03	0.03	0.95
RHARM-MLS	0.01	0.02	0.99

6. Validation of uncertainties

A major innovation of the RHARM dataset is the estimation of uncertainties for temperature, relative humidity and wind measurements for each pressure level of the adjusted IGRA profiles. Uncertainty estimation is fundamental for any type of observation, enabling a metrologically consistent comparison with other datasets. To this purpose, a proper validation of the uncertainties is also required. Validation of uncertainties means that these must be “evaluated by independent means to establish quantitative realism and the credibility of the uncertainty estimates” (Merchant et al., 2017). In order to provide a validation of the RHARM uncertainties, the methodology of

Merchant et al. (2017) has been applied. This is based on the study of the probability density function of the ratio:

$$\frac{x_{RHARM} - x_{ref}}{\sqrt{u_{RHARM}^2 + u_{ref}^2 + u_{mis}^2}} \text{ [Eq. 1]},$$

where x_{RHARM} is the RHARM estimate, x_{ref} indicates an independent estimate of the measurand, u denotes the uncertainty and u_{mis} is the geophysical variability arising from temporal, spatial, and definitional mismatch between RHARM and reference data. A correct quantification of random uncertainties and variability should be reflected in a normal distribution of the ratio defined in Eq. 1, with a standard deviation equal to unity. This approach does not allow to disentangle residual systematic (bias) uncertainties in the RHARM adjusted data. Bias between the reference and the RHARM dataset will affect the distribution derived from Eq. 1 mainly producing a non-zero mean value.

Acknowledging that the ideal solution for the validation should be based on independent reference measurements (Thorne et al., 2017) of the same measurand, GRUAN data would be the ideal candidate. However, RHARM has used information from and mimics part of the GDP, meaning that circularity considerations preclude its use for such a purpose. An alternative solution is adopted, which is to use the ERA5 background (6-hours forecast) as a reference value. Whilst this background is likely a reliable estimation of the atmospheric state, it is not a true reference measurement in that it is not itself an SI traceable measurement, nor does it have comprehensive uncertainty estimates. Observation minus Background (O-B) departures have previously been used as a diagnostic tool for different latitude belts (Ingleby et al., 2017). They also form the basis for the RAOBCORE / RICH family of dataset approaches (Haimberger et al., 2012). Therefore, the use of the ERA5 background as a reference for the test described in Eq. 1 appears viable as a community-accepted approach to infer quantitative information for validating the uncertainties. Other candidates exist, such as radio occultation (RO) satellite measurements (Bauer et al., 2013), which are a valuable solution for dry temperatures in the UTLS, however, for the mid and lower troposphere the deconvolution of temperature and RH in the retrieval is dependent on a first guess model. Furthermore, RO profiles can rarely provide information all the way down to the surface.

Using the background as the reference dataset in Eq. 1, u_{ref} has been estimated applying the Leave-One-Out Cross validation method, LOOCV (Stone, 1974), to the background. LOOCV involves one-fold per value of the background x_{ref} , i.e. each value of the background by itself plays the role of the validation set. The other (N-1) values are used as the training set. The predicted values x_{LOOCV} is estimated using a least-squares linear model and, therefore, u_{ref}^2 is obtained from the corresponding squared error. Instead, u_{mis} is evaluated as the standard deviation of the O-B climatology at each station. The uncertainty validation is carried out separately for two periods: 2004-2019, for which almost all IGRA data have been post-processed using a GRUAN-like adjustment, and 1980-2003, for which the uncertainty estimation of RHARM is obtained by constraining the residuals on a monthly basis (see Madonna et al., 2021, for more details). The validation is focused on temperature and RH uncertainties, because there is still limited information on the homogeneity of the ERA5 background wind data.

For the period 2004-2019 (Figure 12), in the NH, the ratio for temperature has a mean value of 0.18, while the standard deviation is 0.72 indicating, for linearity, that the uncertainty at 300 hPa for temperature is overestimated by about 28%. The overestimated values of the uncertainty decrease

the distribution near the central peak compared to the fitted curve, while it is slightly larger towards the tails. In the tropics, a mean value of the ratio of 0.44 and a standard deviation of 0.89 indicate that the uncertainty is overestimated by about 11%. For RH, in the NH the uncertainty is underestimated with the mean value of the ratio is -1.3 and the standard deviation of 1.2, while the uncertainty is overestimated in the tropics the mean value is -0.69 with a standard deviation of 0.78 and a larger number of overestimated values than in the NH. For the RH uncertainties, the distribution in the NH is bi-modal while at the tropics is negatively skewed. This might be related to systematic effects on the O-B comparison, possibly due to inhomogeneities in the O-B departures within an entire latitude belt, which could broaden the O-B distribution and influence the value of the validation using the model forecast as a reference.

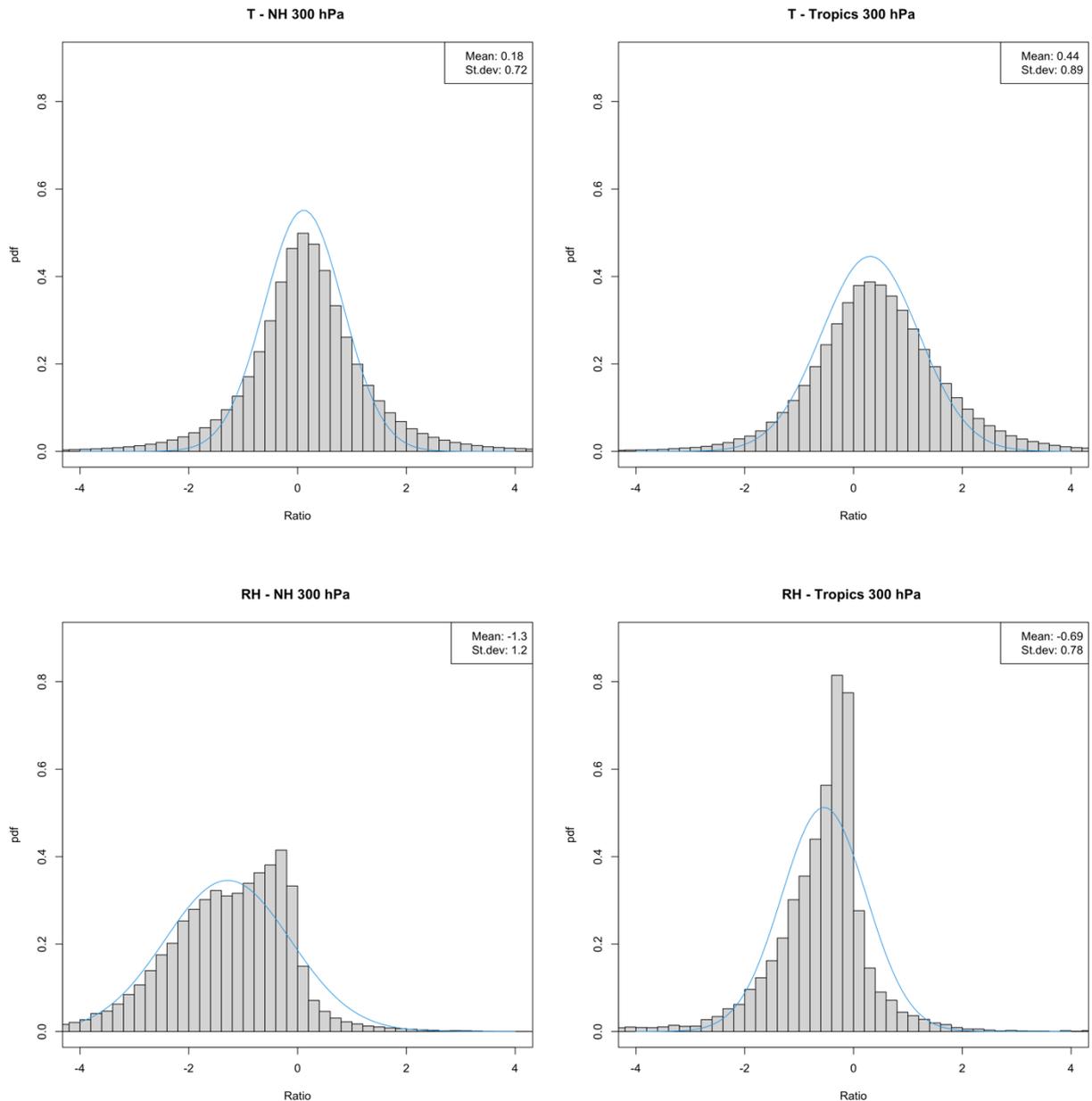


Figure 12: probability density functions (pdfs) of the ratio reported in Eq. 1 calculated using O-B data (RHARM-minus-Background) in the NH (left panels) and in the tropics (bottom panels) at 300 hPa for temperature (top panels) and for RH (bottom panels) to validate the uncertainties estimated using the RHARM approach. The pdfs refer to the RHARM uncertainty values estimated in the period 2004-2019. Background data are from the ERA5 6-hours forecast model. For comparison with ideal uncertainty estimates, the best fitted normal distribution to each dataset (blue line) is also shown.

In an ideal case where uncertainty would be properly estimated with the RHARM algorithm, the distribution should have a standard deviation equal to unity. Deviations from zero are due to the O-B discrepancy.

For the period 1980-2003 (Figure 13), in the NH the ratio for temperature is centred near zero, while the standard deviation is 0.97 indicating that the uncertainty at 300 hPa for temperature is well estimated. The same is true for the tropics where the mean value of the ratio is 0.06 and the standard deviation is 1.0. In both cases the ratios are normally distributed. For RH, both in the NH and at the tropics, the random uncertainty is overestimated by 15% and 12%, respectively, while the mean value of the ratio is 1.7 and 1.5, respectively at NH and tropics, revealing a systematic effect affecting the O-B comparison similarly to the period 2004-2019. The distribution in the NH is negatively skewed, while in the tropics it is positively skewed.

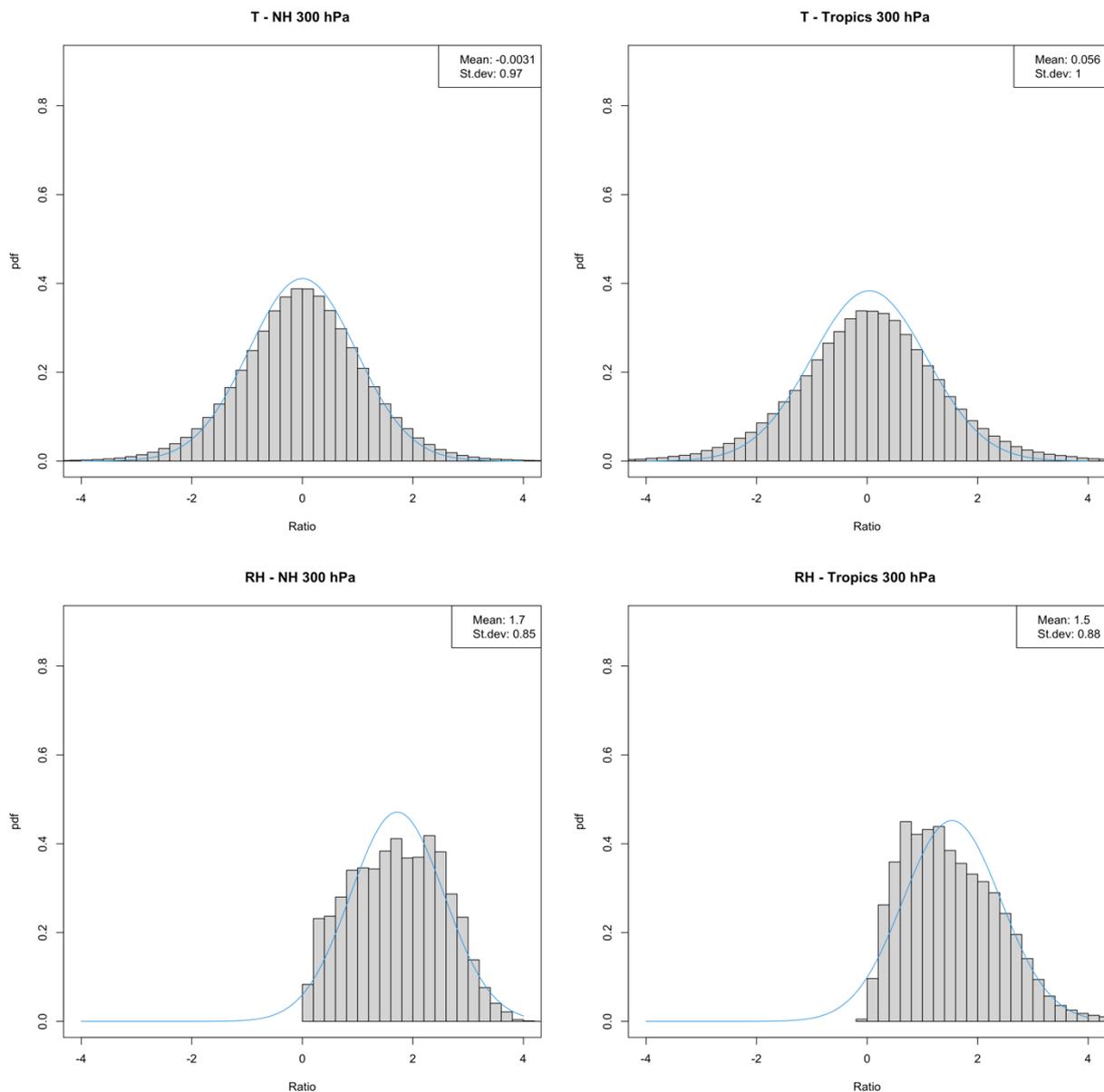


Figure 13: same as Figure 12, but for the period 1980-2003.

In general, the RHARM random uncertainties appear to be a reasonable estimate or a slight overestimate of the theoretical distribution, except for the RH in the NH after 2004 where they are an underestimate. It is arguably preferable to have an over-dispersive uncertainty estimate than an

under-dispersive estimate for most applications. Nevertheless, future versions of the RHARM dataset will be designed to improve the uncertainty estimation, also through the implementation of more sophisticated models, using techniques like kriging or modelling Gaussian processes. Temperature uncertainties are, in general, better behaved than RH uncertainties but it is unclear whether this relates to the uncertainty quantification or O-B field estimate issues.

To conclude the discussion on the RHARM uncertainties, we discuss a case-study where the importance of uncertainties in the comparison with other dataset is shown: while a westerly wind regime characterizes the boreal circulation in the winter stratosphere (e.g. Waugh et al., 2017), occasional flow reversal may occur during sudden stratospheric warmings (SSWs). Here we consider the winters of 2017 and 2018, as in February of both years' SSWs occurred (Knight et al., 2021). Figure 14 shows daily mean temperatures at 100 hPa over the European Polar domain (70°N-90°N; 10°E-50°E), showing sharp temperature increases. SSW are more evident at higher altitudes. The comparison has been carried out for the European Polar domain in the lower stratosphere at 100 hPa to benefit from the larger number of radiosounding data available than at higher pressure levels. For both the events, ERA5, IGRA and RHARM show a similar temperature variability although the reciprocal differences are much larger than the RHARM combined uncertainty (vertical dark gray lines) shown using $k = 3$, where k represents an uncertainty of 3 standard deviations equating approximately a 99% confidence level. Moreover, both the events are preceded by a strong cooling of the lower stratosphere which in 2017 is not well captured by IGRA, due to the warm bias affecting the unadjusted radiosoundings temperature profiles. In 2018, the discrepancy between RHARM and IGRA is similar to 2017 and due to the RHARM warm bias correction, while ERA5 is much warmer than both IGRA and RHARM. In this case, the estimation of the uncertainty is the only way to reveal how significant the difference is among the datasets. Other relevant differences among the three datasets can be pointed out in other months of the time series. In the right panel of Figure 14, a zoom for the orange window in the left plot is shown to provide a more detailed quantification of the discrepancies among the datasets compared to the uncertainty provided by RHARM.

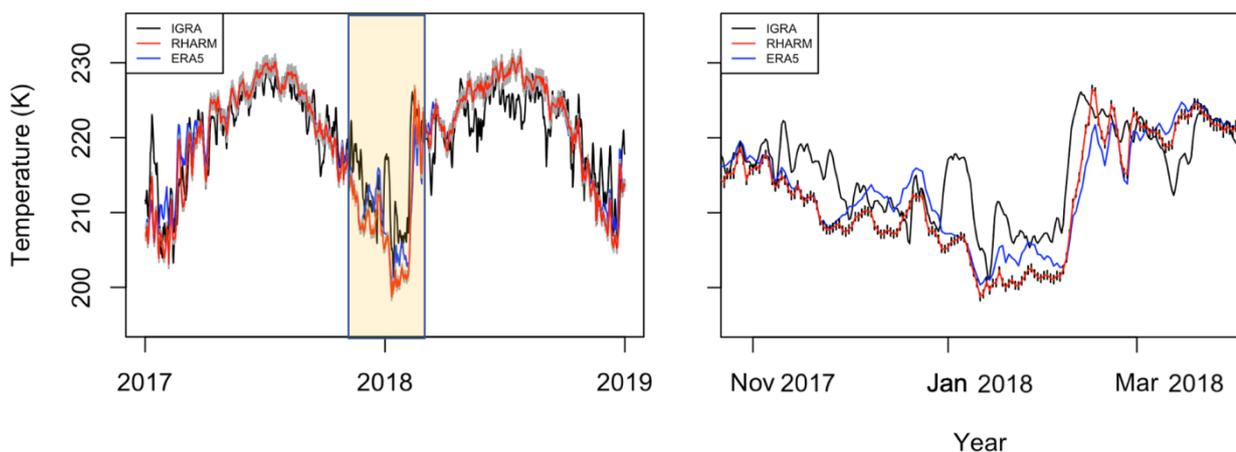


Figure 14: left panel, comparison of time series of daily mean temperature at 100 hPa in the Polar European domain from 1-1-2017 to 31-12-2018; right panel, same as left panel but only for the orange window reported in the left panel. The black lines are the ERA5 daily averages obtained by subsampling the reanalysis data at the IGRA stations within the considered domain; the blue lines show the IGRA daily averages while the red lines show the corresponding RHARM averages. The combined measurement uncertainties for RHARM is also shown (vertical grey error bars).

7. Conclusions and outlook

This work assesses the quality of the homogenized radiosounding measurements of temperature, relative humidity and wind provided in the RHARM dataset, described in detail in a companion paper by Madonna et al. (2021). RHARM is based on a hybrid approach combining physical (metadata-based) and statistical adjustments of radiosonde instrumental effects, along with an estimation of the observational uncertainties.

Existing homogenized datasets of temperature have been shown to largely improve the quality of the historical radiosonde data records, although they do not provide uncertainties which, if well quantified, enable the coherent use of a dataset for climate applications and to increase confidence in their use. RHARM is the only homogenized dataset of temperature providing uncertainties on a per profile basis. For relative humidity and wind, few homogenized datasets exist, and there is a clear benefit of having more datasets to compare and to quantify structural uncertainties. The validation analysis of the RHARM uncertainties, discussed in this paper, qualifies them as a good estimate or an overestimation of the theoretical distribution.

For temperature the comparison reveals that RHARM is consistent with the trend values from ERA5, RICH and RAOBCORE both in the NH and in the tropics with differences within 0.1 K/decade below 300 hPa and within 0.25 K/decade above this. Evident improvement compared to the unadjusted IGRA data can be identified. For relative humidity, the discrepancies among ERA5, HOMORS92 and RHARM are significant although ERA5 and RHARM provide a similar vertical variability of the trends, likely due to a better representation of the seasonality in the corresponding time series. Finally, for wind speed the comparison among ERA5, IUK and RHARM reveal a good agreement above 300 hPa with differences within 0.5 m/(s-1 per decade), while above in the tropics ERA5 has discrepancies from the other dataset larger than 1.0 m/(s-1 per decade).

The in-depth comparison of ERA5, IGRA and RHARM anomalies shows that the adjustments applied in the RHARM data processing reduce the differences between ERA5 and observational data, especially for temperature in the NH, where the most significant adjustment reduces the differences from 0.5 K to 0.1 K in the decade 1996-2005, on average, and for relative humidity, where the differences are adjusted from values of more than 10% RH to less than 5% RH in the decade 1978-1987. Relative humidity trends are negative in the NH and positive in the tropics: in particular, since the beginning of 2015, positive trends are observed in the tropics and in SH (not shown), which are correlated with the reported significant warming in the tropical Pacific Ocean. Adjustment to wind speed anomalies, although smaller, also show the improvement in the times series in comparison with ERA5.

To further demonstrate the value of the applied adjustments to temperature time series, RHARM has been compared also with the MSU data which is the longest satellite data record of temperature in the lower stratosphere. Differences in the trends estimated by radiosounding and satellite observations are about 50% lower for RHARM compared to IGRA. The RHARM temperature bias adjustments are also compared with ERA5 for the stratospheric warming events of 2017 and 2018, highlighting the critical role of uncertainties to assess the importance of differences among datasets.

For water vapour, the efficacy of RHARM processing for the period after 2004 is revealed from the comparison of the 300 hPa monthly means in tropics between RHARM and Aura MLS. The MAD is 0.01 g/kg for RHARM compared to 0.03 g/kg for IGRA, with a related reduction of the RMSE and increased correlation (from 0.95 to 0.99).

In conclusion, this paper demonstrates the good performance of the RHARM dataset in comparison with popular climate data records, and the added value that RHARM time series are provided with an estimation of the uncertainty for each observed value. The RHARM approach can therefore contribute to a more comprehensive exploration of structural uncertainties in historical records.

Future work should be oriented also in homogenizing other moments of the data probability density functions, like the standard deviation, which often appears strongly affected by the limited sensitivity of the radiosonde sensors used historically. Examples in this direction are offered for temperature by Della Marta and Wenner (2006), and more recently by Zhou et al. (2021). A relevant effort should be spent also to improve the uncertainty estimation, through the implementation of more sophisticated models, using techniques like kriging or modelling Gaussian processes. Similar statistical tools could be considered to extend the RHARM approach to those stations where metadata are not available. The extension of RHARM data backward in time will be also considered to improve the representation of the upper troposphere and lower stratosphere in future pre-satellite atmospheric reanalysis datasets, which are an important source of information for studies of stratosphere-troposphere coupling (Hitchcock et al., 2019).

The future availability of additional Reference data from GRUAN and from upcoming WMO/CIMO radiosonde intercomparisons will be fully exploited to improve the accuracy of the RHARM time series and the related uncertainties. The dataset maintenance and Improvement will continue also benefiting the support of climatologists and metrologists, and further comparisons with additional datasets, such as radio occultation measurements, will be carried out in critical regions, such as the tropical upper troposphere.

8. 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).

9. Acknowledgements

This work was done on behalf of the European Union's Copernicus Climate Change Service implemented by ECMWF. Use of 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.

10. References

- Bauer, P. et al. (2014) 'GNSS Radio Occultation Constellation Observing System Experiments', *Monthly Weather Review*, 142(2), pp. 555–572. doi: 10.1175/MWR-D-13-00130.1.
- Bodeker, G. E. et al. (2016) 'Reference Upper-Air Observations for Climate: From Concept to Reality', *Bulletin of the American Meteorological Society*, 97(1), pp. 123–135. doi: 10.1175/BAMS-D-14-00072.1.
- Dai, A. et al. (2011a) 'A New Approach to Homogenize Daily Radiosonde Humidity Data', *Journal of Climate*, 24(4), pp. 965–991. doi: 10.1175/2010JCLI3816.1.
- Dai, A. et al. (2011b) 'A New Approach to Homogenize Daily Radiosonde Humidity Data', *Journal of Climate*, 24(4), pp. 965–991. doi: 10.1175/2010JCLI3816.1.
- Dee, D. P. et al. (2011) 'The ERA-Interim reanalysis: configuration and performance of the data assimilation system', *Quarterly Journal of the Royal Meteorological Society*, 137(656), pp. 553–597. doi: 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>.
- Della-Marta, P. M. and Wanner, H. (2006) 'A Method of Homogenizing the Extremes and Mean of Daily Temperature Measurements', *Journal of Climate*, 19(17), pp. 4179–4197. doi: 10.1175/JCLI3855.1.
- Dessler, A. E. and Davis, S. M. (2010) 'Trends in tropospheric humidity from reanalysis systems', *Journal of Geophysical Research*, 115(D19), p. D19127. doi: 10.1029/2010JD014192.
- Dirksen, R. J. et al. (2014) 'Reference quality upper-air measurements: GRUAN data processing for the Vaisala RS92 radiosonde', *Atmospheric Measurement Techniques*, 7(12), pp. 4463–4490. doi: 10.5194/amt-7-4463-2014.
- Durre, I. et al. (2018) 'Enhancing the Data Coverage in the Integrated Global Radiosonde Archive', *Journal of Atmospheric and Oceanic Technology*, 35(9), pp. 1753–1770. doi: 10.1175/JTECH-D-17-0223.1.
- Durre, I., Vose, R. S. and Wuertz, D. B. (2006) 'Overview of the Integrated Global Radiosonde Archive', *Journal of Climate*, 19(1), pp. 53–68. doi: 10.1175/JCLI3594.1.
- Ferreira, A. P., Nieto, R. and Gimeno, L. (2019) 'Completeness of radiosonde humidity observations based on the Integrated Global Radiosonde Archive', *Earth System Science Data*, 11(2), pp. 603–627. doi: 10.5194/essd-11-603-2019.
- Free, M. et al. (2004) 'Using First Differences to Reduce Inhomogeneity in Radiosonde Temperature Datasets', *Journal of Climate*, 17(21), pp. 4171–4179. doi: 10.1175/JCLI3198.1.
- Haimberger, L. (2007) 'Homogenization of Radiosonde Temperature Time Series Using Innovation Statistics', *Journal of Climate*, 20(7), pp. 1377–1403. doi: 10.1175/JCLI4050.1.

- Haimberger, L., Tavalato, C. and 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(18), pp. 4587–4606. doi: 10.1175/2008JCLI1929.1.
- Haimberger, L., Tavalato, C. and Sperka, S. (2012) 'Homogenization of the Global Radiosonde Temperature Dataset through Combined Comparison with Reanalysis Background Series and Neighboring Stations', *Journal of Climate*, 25(23), pp. 8108–8131. doi: 10.1175/JCLI-D-11-00668.1.
- Hersbach, H. et al. (2020) 'The ERA5 global reanalysis', *Quarterly Journal of the Royal Meteorological Society*, 146(730), pp. 1999–2049. doi: 10.1002/qj.3803.
- Hitchcock, P. (2019) 'On the value of reanalyses prior to 1979 for dynamical studies of stratosphere–troposphere coupling', *Atmospheric Chemistry and Physics*, 19(5), pp. 2749–2764. doi: 10.5194/acp-19-2749-2019.
- Hu, S. and Fedorov, A. V. (2017) 'The extreme El Niño of 2015–2016 and the end of global warming hiatus: GLOBAL WARMING HIATUS AND 2015 EL NIÑO', *Geophysical Research Letters*, 44(8), pp. 3816–3824. doi: 10.1002/2017GL072908.
- Ingleby, B. (2017) 'An assessment of different radiosonde types 2015/2016'. doi: 10.21957/0NJE0WPSA.
- Kawatani, Y. et al. (2016) 'Representation of the tropical stratospheric zonal wind in global atmospheric reanalyses', *Atmospheric Chemistry and Physics*, 16(11), pp. 6681–6699. doi: 10.5194/acp-16-6681-2016.
- Knight, J. et al. (2021) 'Predictability of European Winters 2017/2018 and 2018/2019: Contrasting influences from the Tropics and stratosphere', *Atmospheric Science Letters*, 22(1). doi: 10.1002/asl.1009.
- Kobayashi, S. et al. (2015) 'The JRA-55 Reanalysis: General Specifications and Basic Characteristics', *Journal of the Meteorological Society of Japan*. Ser. II, 93(1), pp. 5–48. doi: 10.2151/jmsj.2015-001.
- Kremser, S. et al. (2018) 'Is it feasible to estimate radiosonde biases from interlaced measurements?', *Atmospheric Measurement Techniques*, 11(5), pp. 3021–3029. doi: 10.5194/amt-11-3021-2018.
- Madonna, F., Tramutola E. (2020a): RHARM (Radiosounding HARMonization) dataset - subset [Data set], Zenodo, <http://doi.org/10.5281/zenodo.3973353>
- Madonna, F (2020b). Can reference radiosounding measurements be used to improve historical time series? *Il Nuovo Cimento C*, 43, 1–10, doi:10.1393/ncc/i2020-20121-5c.
- Madonna, F. et al. (2020c) Radiosounding HARMonization (RHARM): a new homogenized dataset of radiosounding temperature, humidity and wind profiles with uncertainty. preprint. *Data, Algorithms, and Models*. doi: 10.5194/essd-2020-183.
- Madonna, F., Tramutola, E., Sy, S., Serva, F., Proto, M., Rosoldi, M., Gagliardi, S., Amato, F., Marra, F., Fassò, A., Gardiner, T., and Thorne, P. W.: The new Radiosounding HARMonization (RHARM)

dataset of homogenized radiosounding temperature, humidity and wind profiles with uncertainties: statistical analysis of the dataset, *J. of Geophys. Res.*, submitted, 2021.

McCarthy, M. P. et al. (2008) 'Assessing Bias and Uncertainty in the HadAT-Adjusted Radiosonde Climate Record', *Journal of Climate*, 21(4), pp. 817–832. doi: 10.1175/2007JCLI1733.1.

McCarthy, M. P., Thorne, P. W. and Titchner, H. A. (2009) 'An Analysis of Tropospheric Humidity Trends from Radiosondes', *Journal of Climate*, 22(22), pp. 5820–5838. doi: 10.1175/2009JCLI2879.1.

Mears, C. A. and Wentz, F. J. (2009) 'Construction of the Remote Sensing Systems V3.2 Atmospheric Temperature Records from the MSU and AMSU Microwave Sounders', *Journal of Atmospheric and Oceanic Technology*, 26(6), pp. 1040–1056. doi: 10.1175/2008JTECHA1176.1.

Merchant, C. J. et al. (2017) 'Uncertainty information in climate data records from Earth observation', *Earth System Science Data*, 9(2), pp. 511–527. doi: 10.5194/essd-9-511-2017.

Nash, J., T. Oakley, H. Vömel, LI Wei (2010). WMO intercomparison of high quality radiosonde systems, WMO/TD-No. 1580, Yangjiang, China, https://library.wmo.int/index.php?lvl=notice_display&id=15531#.YJp50C0QNQI, (last access 11/05/2021).

Rienecker, M. M. et al. (2011) 'MERRA: NASA's Modern-Era Retrospective Analysis for Research and Applications', *Journal of Climate*, 24(14), pp. 3624–3648. doi: 10.1175/JCLI-D-11-00015.1.

Sherwood, S. C., C. L. Meyer, R. J. Allen, and H. A. Titchner (2008). Robust tropospheric warming revealed by iterative homogenized radiosonde data, *J. Clim.*, 21, 5336–5352, doi:10.1175/2008JCL2320.1.

Sherwood, S. C. and Nishant, N. (2015) 'Atmospheric changes through 2012 as shown by iteratively homogenized radiosonde temperature and wind data (IUkv2)', *Environmental Research Letters*, 10(5), p. 054007. doi: 10.1088/1748-9326/10/5/054007.

Spencer, R. W. and Christy, J. R. (1990) 'Precise Monitoring of Global Temperature Trends from Satellites', *Science*, 247(4950), pp. 1558–1562. doi: 10.1126/science.247.4950.1558.

Spencer, R. W., 2016: UAH V6 global temperature update for January (2016). +0.54 deg C. [Available online at <http://www.drroyspencer.com/2016/02/uah-v6-global-temperature-update-for-january-2016-0-54-deg-c/>.]

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>

Thorne, P. W. (2005) 'Revisiting radiosonde upper air temperatures from 1958 to 2002', *Journal of Geophysical Research*, 110(D18), p. D18105. doi: 10.1029/2004JD005753.

Thorne, P. W. et al. (2011) 'A quantification of uncertainties in historical tropical tropospheric temperature trends from radiosondes', *Journal of Geophysical Research*, 116(D12), p. D12116. doi: 10.1029/2010JD015487.

- Thorne, P. W. et al. (2017) 'Making better sense of the mosaic of environmental measurement networks: a system-of-systems approach and quantitative assessment', *Geoscientific Instrumentation, Methods and Data Systems*, 6(2), pp. 453–472. doi: 10.5194/gi-6-453-2017.
- Vömel, H. et al. (2007) 'Validation of Aura Microwave Limb Sounder water vapor by balloon-borne Cryogenic Frost point Hygrometer measurements', *Journal of Geophysical Research*, 112(D24), p. D24S37. doi: 10.1029/2007JD008698.
- Wang, J. et al. (2013) 'Radiation Dry Bias Correction of Vaisala RS92 Humidity Data and Its Impacts on Historical Radiosonde Data', *Journal of Atmospheric and Oceanic Technology*, 30(2), pp. 197–214. doi: 10.1175/JTECH-D-12-00113.1.
- Waugh, D. W., Sobel, A. H. and Polvani, L. M. (2017) 'What Is the Polar Vortex and How Does It Influence Weather?', *Bulletin of the American Meteorological Society*, 98(1), pp. 37–44. doi: 10.1175/BAMS-D-15-00212.1.
- Yan, X. et al. (2016) 'Validation of Aura MLS retrievals of temperature, water vapour and ozone in the upper troposphere and lower–middle stratosphere over the Tibetan Plateau during boreal summer', *Atmospheric Measurement Techniques*, 9(8), pp. 3547–3566. doi: 10.5194/amt-9-3547-2016.
- Zhou, C. et al. (2021) 'A New Approach to Homogenize Global Subdaily Radiosonde Temperature Data from 1958 to 2018', *Journal of Climate*, 34(3), pp. 1163–1183. doi: 10.1175/JCLI-D-20-0352.1.