Authors’ Response to Referee #1 Comments on

“Completeness of radiosonde humidity observations based on the IGRA” by António P. Ferreira, Raquel Nieto, Luis Gimeno

Referee comments (highlighted in blue) are copied before the authors’ responses.

The manuscript by Ferreira et al. presents an analytical description and a rationalized statistical analysis of the radiosounding data archive available through the NOAA-IGRA initiative. The authors report an extensive and precious description of a huge number of information about the changes occurred in the number, geographical density and type of radiosoudings and payloads used around the globe since the beginning of last century to present for the measurement of atmospheric humidity. The authors focus they study on humidity measurements and they investigate the datasets in a critical way to show if, in term of continuity and coverage, the available data may support climate studies, though they do not assess this aspect on a scientific sound basis. The paper is well written and curated and the provided analysis and it update expected on a biannual basis may support many activities currently ongoing, not ultimately the Copernicus Climate Change Service (C3S). I am in favor of the manuscript publication. Nevertheless, I have to provide the authors with major revisions, because to my opinion a few concepts must be clarified and modified in the text of the manuscript, though this does not affect my positive view on the manuscript itself. Below I enclose my general comment and a bunch of minor issues to solve.

General comments

1. In the manuscript, starting form the abstract, it is mentioned that the IGRA V2 datasets is investigated according to various completeness criteria in order to show if the length of the available data records is sufficient for the purposed to estimate climate variability trends. The authors mainly refer to the temporal continuity of the observation and not the spatial gaps which are anyhow investigated in the statistics. Despite this purpose, the authors clarify in the conclusion that their work can only be supportive of climate studies or works related to the climate homogenization of the time series, and nor they do not discuss on a quantitative basis what is the length of data records required for climate studies neither they refer to past literature to asses this aspect. Assuming that the authors are not interested in the going ahead in the assessment of the effect of gaps in the data records on the estimation of climate signals (which is well beyond the scope of the paper), I’d ask them to change the tone of the manuscript, where needed, to better clarify the scope of the manuscript.

The abstract states the paper´s purpose: “The sounding data compiled in the Integrated Global Radiosonde (…) are examined here until the end of 2016, aiming to describe the completeness of humidity observations from radiosondes in different times and locations” [P10, L10-12], further expressing the expectation that “the derived metadata will help climate and environmental scientists to find the most appropriate radiosonde data for humidity studies by selecting upper-air stations, observing years or individual soundings according to various completeness criteria – even if differences in instrumentation and observing practices require extra attention” [P1, L25-27]. Similarly, the introductory section states: “The purpose of this paper is to study the completeness of humidity observations collected in IGRA according to various needs – number and latitudinal distribution of observing stations, fraction of observing days in a year, resolution and range of vertical levels, length and continuity of the time-series, minimal sampling between the surface and the 500-hPa level – aiming to facilitate the use of radiosonde humidity data by atmospheric and environmental scientists.” [P3, L14]. In short, our main purpose is not exactly to show if the available data may support climate studies. Still, the interest of our work to climate studies is stressed in several parts of the manuscript. The next comment summarizes how climate studies are addressed in the body of the text.

The use of radiosonde data for climate studies is mentioned in the Introduction (P2, L25-29) because radiosonde archives provide the longest atmospheric humidity records available. While the notion of climatic variability may refer to different time scales, it usually involves a few decades (to define a reference mean state). Concerning climatic trends, the longer and stable the time series, the better. Therefore, Section 2.3.4 draws attention to the importance of accessing large interruptions in long term humidity time series, by identifying the time series without gap years and measuring the continuity of data within each year. The availability of multidecadal humidity time series of upper-air humidity is illustrated in Figures 9 and 10 of Section 3.4, using different completeness requirements for time series without gap years (concerning the fraction of observing days in a year, the maximum size of gap days and the vertical sampling in the lower troposphere). As explained in that section, "There is no simple answer to the question of since when we have enough data to (in theory) perform climate studies from radiosonde humidity data: it depends on the strictness of completeness criteria." [P20, L4-6]. Figs. 10-11 simply indicates how much the availability of long term humidity time series can depend on specific data completeness requirements, besides the length of the period of record of course. We think that the summary results in the concluding Sect. 6 (last bullet paragraph on P23, L1-8) are sufficiently objective in this respect. The paper ends with a suggestion on the selection of time series for studying the homogenization of humidity data.

Regarding the availability of data for climate studies, the abstract reads: "For illustration, the study presents a global picture of the completeness of radiosonde humidity observations over the years, including their latitudinal coverage. This overview shows that the number of radiosonde stations having a long enough record length for studies on the climatic variability and trends of humidity-related quantities depends critically on the temporal continuity, regularity and vertical sampling of the humidity time-series." [P1, L22-25]. We understand that the latter sentence can give the wrong impression that our paper quantifies the completeness *requirements* for performing climate studies on humidity-related quantities. This misperception can be briefly solved. We further understand that the Short Summary appearing on the ESSDD publication can be slightly misleading about the scope of the paper. This can be changed too.

Changes in the Abstract:

Paragraph on P1, L23-25:
This overview shows that the number of radiosonde stations having a long enough record length for studies on the climatic variability and trends of humidity-related quantities depends critically on the temporal continuity, regularity and vertical sampling of the humidity time-series.

It will be changed as follows:
This overview indicates that the number of radiosonde stations having a record length potentially long enough (multidecadal) for climate studies involving humidity-related quantities depends not only on the temporal range, but also on the continuity, regularity and vertical sampling of the humidity time-series**.**

Amendment to the Short Summary:

Where it reads:
The work shows that the potential use of radiosonde humidity data for climate studies depends on the continuity, regularity and vertical sampling of long time series.

It should read:

The work illustrates how humidity data potentially available for climate studies depend on the length, continuity, regularity and vertical sampling of time series.

2. The authors make use of IGRA data V2, which is the most updated version of IGRA which embed several improvements compared to the previous data version (V1). Nevertheless, in my personal researches I had a chance to find many bugs in the IGRA V2 where many data present in the V1 are missing and this is not dependent on the extended quality check applied within the IGRA V2. I had also a chance to report this bug to Imke Durré (PI of IGRA). I got similar feedbacks from other EU and US colleagues during discussion at various workshops. The station of Lindenberg in Germany (WMO index=10393), so accurately described in its past history in the manuscript, is one those affected by this issue (at least until a few weeks ago, my last access to IGRA). Therefore, I am wondering whether a comparison between IGRA V1 and V2 has been carried out and if gaps have been found and fixed somehow in the datasets investigated in the manuscript.

Such apparent data loss during the processing of data sources in IGRA 2 is concerning. We have used IGRA 2 as is, because of the extended data coverage and improvements of quality assurance on humidity data. On the positive side, any future corrections to sounding data in IGRA will translate into the updated versions of the meta-dataset introduced in the present paper.

3. The description of statistical analysis of IGRA V2 data is very extensive way, providing several details and a long description of each figure and table. When reading, I have been very interested by the content, though sometimes the reader may get tired by the way the manuscript is written. In a way similar to the conclusions, I'd suggest to the authors to change the style of their writing and privilege a description in "bullets" to describe the results whenever possible. this will help also to clarify the text itself. For example, for the relative humidity observations, At pag.5, it is reported that the average number of non-standard levels in weather-balloon sounding reports increased from about zero by 1945 to about 30 by 2000, but later in the text it is said that RH observation were already available since the 1930, and again later on (pag. 18) it is stated that RH are becomes more abundant since 1949. Though all of this information are exact the reader may get confused and their comparability and usefulness could be limited if the statical analysis is not described in a more schematic way.

We believe that the largest sections (Sect. 2 and 3) will become much easy to read once the tables and figures are placed besides the text. Please note that paragraphs were used extensively to organize ideas around themes. Also, number bullets were used four times before the concluding section. Data analysis is schematized in the following way: Sect. 2.1, 2.2 and 2.3 describe, respectively, the IGRA sounding data set used in our work, the identification of mostly-radiosonde data series, and the main data analysis needed to compute the parameters used to describe 'completeness of humidity observations'. Figures 1 and 2 are auxiliary to Sect. 2.1 and 2.3. Each subsection of Sect 2.3 ('Analysis of humidity data') introduces the corresponding completeness parameters and ends with a description of the secondary statistical analysis used later in Sect. 3 ('Overview on the completeness of radiosonde humidity observations' – in which the several completeness measures are graphically illustrated by Figs. 3-10. The introductory part of section 3 and several parts of subsections 3.1 to 3.4 recall the data and methods used to draw the related figures, each time they are first mentioned in the text.

On P8 (not P18), L2-4, what we have said is that *until 1969 humidity was reported only in the form of RH* (not that RH data became more abundant since 1949), as documented in IGRA's Readme file. This is unrelated with the fact that the earliest upper-air humidity observations are from 1930 [see P5, L28; P9, L16]; or that in 1945 the average number of significant levels in weather-balloon reports was nearly zero [see P5, L15] – meaning that standard pressure levels were virtually the only ones reported by then. None of this information has to do with the data analysis explained later in Sect. 2.3.

4. The use of the metric present in the Eq.1, "zonal coverage index" at pag 11 is not clear to me. Have been it used in the past and its adde value with respect to other metrics was shown? What's the added value with respect to a station density per 1000 km, for example? Has the ocean surface been excluded from the global surface, given that this can be calculate more clearly for fixed stations? There are many concerns to me on the use of this index, which requires clarification from the authors. Personally, I think that the user can make use of a much simpler index or statistics (a few of these are used later on in the manuscript by the authors themselves) to show if a zonal belt are under and over- represented. This is also depending on the different atmospheric circulation occurring in the different zonal belts, so I am wondering what's the usefulness of adopting this metric. I ask the authors to clarify in the manuscript of the added value due to the use of this metric.

The "zonal coverage index" index defined by Eq. 1 was intended to compare the fraction of stations among in different climatic zones, normalized by the surface area fraction. To our knowledge it was never used before. [Note: by mistake, the modulus term in Eq. (1) is not raised to -1, as it should be]. Although that index was used to construct Fig. 4b, note that it does not form part of the parameters represented in the dataset introduced in the paper.

We agree that the average station density in different latitude belts – or alternatively the average spacing of stations – would give a more direct information about the data coverage. Moreover, the plot of Fig. 4a representing the fraction of humidity-reporting stations at different latitude bands is not essential, since Fig. 3 (which uses yearly absolute numbers) gives a detailed picture of the same distribution. Therefore, the index in question can be replaced by a more familiar metrics, as suggested in the Referee comment, without affecting the dataset introduced in the paper.

Planned changes in the manuscript:

Section 2.3.1 will be fully revised, with the purpose of replacing Eq (1) by a parameter describing the average separation between adjacent stations in a given zonal band. Separate calculations will be presented for land and ocean areas. The **plots in Figure 4a and 4b** will be replaced to represent both regions. The **comments to Fig. 4 in Sect. 3.1** and **the first bullet paragraph of Sect. 6** will be modified accordingly (but retaining the essential about the relative amount of stations in each climatic zone as shown in Fig. 3).

5. The average vertical resolution could be useful information, but to my opinion, thinking about different applications it could be useful to have a statistic of the available level for the different regions of the atmosphere: Planetary Boundary Layer, Free Troposhere, Upper Troposphere/Lower Stratosphere (UT/LS), Upper Stratosphere. This classification may have a

stronger impact to orienting users' application in the selection of the available data, e.g for trends calculation.

Figure 6 only shows the distribution of vertical resolution among humidity observations over time. The dataset introduced in the discussion paper, however, contains four related parameters for each IGRA-RS station:

i. Annual mean vertical resolution of yearly humidity obs. from each station
ii. Annual mean vertical resolution of yearly Sfc-to-500hPa humidity obs. from each station
iii. Vertical resolution of individual humidity obs. from each station
iv. Vertical resolution of individual Sfc-to-500hPa humidity obs. from each station

where 'resolution' stands for geometric mean resolution. Parameters (i) and (iii) refer to reported humidity data up to the highest measuring level. Parameters (ii) and (iv) refer to humidity data between the surface and the 500-hPa level, requiring humidity data at the surface and at a minimum of upper-air levels to estimate precipitable water vapor. Therefore, the dataset accompanying the paper provides specific information on the vertical resolution at the lower troposphere.

Moreover, the dataset also provides information on the highest measuring level:

v. Annual geometric mean pressure of the highest level with humidity data at each station
vi. Pressure of the highest level with humidity data in individual obs. at each station
vii. Altitude of the highest level with humidity data in individual obs. at each station

where (vii) is obviously only calculated if humidity is not missing at the surface level. Such information allows to select the soundings with humidity observations reaching the UT and the LS, as well as the stations and years that normally have observations in the UT/LS, even if the vertical resolution at those regions is not detailed.

Considering the parameters describing temporal completes of humidity observations [periods of record for humidity; fraction of observing days in a year; largest number of consecutive missing days], analysing vertical resolution for each separate atmospheric layer (PBL, lower/middle troposphere, UT, LS) – each one possessing its own temporal completeness – would imply extending the work beyond what might be expected from a first approach. Undoubtedly, a recommendation for future research.

6. Please check the use of the term "error" throughout the manuscript and replace it with "uncertainties" where more appropriated.

The term 'error' is used only twice in the manuscript: "Concerning humidity, radiosonde-based climatic studies are for now confined to the lower and middle troposphere, because of the large errors and insufficient data in the upper troposphere and lower stratosphere." [P2, L25-26] "For the purpose, it suffices to neglect moisture in the hypsometric equation; given that the virtual temperature is typically within 4 K above the actual temperature, the error amounts to less than 1 %." [P12, L28-29]

In the first case, we refer to measuring errors related to the poor accuracy of humidity sensors in the upper troposphere (UT) and lower stratosphere (LS). Given that different sensors have different accuracies under the same conditions, equal instruments behave differently in different conditions, and climate studies use data averaging from many observations, we must accept that the expression ‘uncertainty’ is more appropriate. Moreover, missing data arising from varying observation practices (low RH reporting, low-temperature cutoff) introduce uncertainties in the time series and their averages. This affect particularly the measurements above 500 hPa. (On passing, we should refer the issue of observation practices right away on P2, L22.) However, systematic biases in the UT are well known for certain radiosonde models. Thus, to include both situations, we would say ‘uncertainty of measurements and biases’. In the LS, most measurements show a wet RH bias which is very large in relative terms. We think it’s also adequate to add a few citations regarding accuracy and uncertainty of humidity measurements in the UT and LS.

In the second case, “error” is the error in calculating geopotential height; this will be clarified

Changes in the manuscript:

P2, L22
Where it reads
instrument changes and sampling differences
It will read
instrument changes, reporting practices and sampling differences

P2, L25-26
Where it reads:
the large errors and insufficient data in the upper troposphere and lower stratosphere
It will read:
the large uncertainty of measurements and biases in the upper troposphere stratosphere (Elliot and Gaffen, 1991; Soden and Lazante, 1996; Wang et al., 2003) and the extremely large relative biases and insufficient data in the lower stratosphere (Miloshevich et al., 2006; Nash et al. 2011)

P12, L29
Where it reads:
the error amounts to
It will read:
the error in calculating geopotential height amounts to

Additions to the reference list:

Wang, J., D. J. Carlson, D. B. Parsons, T. F. Hock, D. Lauritsen, H. L. Cole, K. Beierle, and E. Chamberlain (2003), Performance of operational radiosonde humidity sensors in direct comparison with a chilled mirror dew-point hygrometer and its climate implication, Geophys. Res. Lett., 30, 1860, doi:10.1029/2003GL016985, 16.

Soden, Brian & Lanzante, John. (1996). An Assessment of Satellite and Radiosonde Climatologies of Upper-Tropospheric Water Vapor. J. Climate. 9. 1235-1250. 10.1175/1520-0442(1996)009<1235:AAOSAR>2.0.CO;2.

Nash, J., T. Oakley, H. Vömel, and L. I. Wei, 2011: WMO intercomparison of high quality radiosonde systems (Yangjiang, China 12 June–3 August 2010). WMO Instruments and Observing Methods Rep. 107, 238 pp. [Available at www.wmo.int]

Specific comments

Pag.2, lines 10-12: the authors could mention that recent reanalysis products, for example ERA5 from ECWMF, will improve the 4 times daily frequency of the products up to hourly.

Perhaps, we should rather refer some ECMWF reanalysis as it is now, with a different period of record. This can be amended.

Pag.2, line 13: replace "not to mention : : :.." with "with not negligible : : :.."
Agree.

Pag.2, lines 17-19: uncertainties due to balloon drifting and observation time are considered negligible citing the publication by Kitchen (1989) as well as the radiosonde profile accuracy. To my opinion these cannot be considered minor, and anyhow if minor or not this is depending on the considered application. First of all there more recent papers by Seidel et al. (2011) dealing with radiosonde balloon drifting. Estimates of elapsed time from balloon launch to various pressure levels, due to vertical balloon rise, have median values increasing from about 5 min at 850 hPa to about 1.7 h at 10 hPa, with ranges of about 20% of median values. Observed elapsed times always exceed those estimated using assumed 5 or 6 m/s rise rates. Regarding the data data accuracy, if we are referring to the ensemble of effect which may alter the sensors' optimal measurement conditions, like the solar radiation effect to the effect of a sensor time-lag, these have been better quantified for the more recent radiosonde types and are quite relevant for any kind of climate application (see for example the GRUAN quantification by Dirksen et al. 2014). This is also the reason why many scientific groups have developed homogenized dataset of radiosonde data for climate application. For there reason above, I'd reformulate these line in the manuscript and I'd provide more updated references.

We understand that uncertainties related to the observation time and balloon drift are currently recognized to be relevant for data assimilation in models (both in numerical weather prediction and reanalysis), besides impacting forecast verification, satellite validation and climatic statistics, specially at high levels. Kitchen (1989) pointed out its importance for weather forecasting. We appreciate that the referee comment calls for more recent literature to substantiate the subject.

We remove "minor" and improve the whole sentence, adding several citations and new references to make a balance between the different problems mentioned. Dirksen et al. (2014) will be cited on P 6, L30.

Changes in the manuscript:

P2, L17-19
Where it reads:
despite sampling differences among stations and over time, minor uncertainties related to observation time and balloon drift (Kitchen, 1989), and differences in data accuracy – which depends on humidity sensors and varies with measured conditions (e.g., Nash, 2002).
It will read:
despite geographical-temporal sampling differences (Wallis, 1998), uncertainties related to observation time and balloon drift (Kitchen, 1989; McGrath et al., 2006; Seidel et al., 2011; Laroche and Sarrazin, 2013), differences in vertical coverage and data gaps related to reporting practices of humidity (Dai et al. 2011 and references thein) and differences in humidity data accuracy – which depends on humidity

sensors and varies with measured conditions (e.g., Nash, 2002; Sappuci et al., 2005; Moradi et al., 2013; Dirksen et al., 2014).

P6, L30
Insert reference: (Dirksen et al., (2014) and references therein)

<u>Additions to the list of references</u>:

Seidel, D. J., B. Sun, M. Pettey, and A. Reale, 2011: Global radiosonde balloon drift statistics. J. Geophys. Res., 116, D07102, Doi: 10.1029/2010JD014891.

Laroche, S. and R. Sarrazin, 2013: Impact of Radiosonde Balloon Drift on Numerical Weather Prediction and Verification. Wea. Forecasting, 28, 772–782, Doi: 10.1175/WAF-D-12-00114.1

McGrath, R., T. Semmler, C. Sweeney, and S. Wang, 2006: Impact of Balloon Drift Errors in Radiosonde Data on Climate Statistics. J. Climate, 19, 3430–3442, https://doi.org/10.1175/JCLI3804.1

Sapucci, L.F., L.A. Machado, R.B. da Silveira, G. Fisch, and J.F. Monico, 2005: Analysis of Relative Humidity Sensors at the WMO Radiosonde Intercomparison Experiment in Brazil, J. Atmos. Oceanic Technol., 22, 664–678, doi:10.1175/JTECH1754.1

Wallis, T.W., 1998: A Subset of Core Stations from the Comprehensive Aerological Reference Dataset (CARDS). J. Climate, 11, 272–282, doi: 10.1175/1520-0442(1998)011<0272:ASOCSF>2.0.CO;2

Dirksen, R. J., Sommer, M., Immler, F. J., Hurst, D. F., Kivi, R., and Vömel, H.: 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, 2014.

Pag.2, line 26: please replace "errors" with uncertainties and then add also that the radiosonde sensors may have a limited sensitive to ppm water vapor concentrations in the UT/LS as one of the main reason because humidity data above 300hPa are unreliable.

Regarding the word replacement, please <u>refer to our response to point 6 of the general comments</u>. The details about limitations of humidity sensors were given in Sect. 1.3. However, <u>we understand that the limited response to low water vapor pressures should be directly mentioned in Sect 1.3</u>.

<u>Change in the manuscript:</u>

P5, L29
Where it reads:
humidity has been always difficult to measure in very cold or dry air.
It will read:
humidity has been always difficult to measure in very cold or dry air due to the poor response of many instruments at very small vapor concentrations (by limiting RH, cold temperatures are associated with low water vapor pressures).

Pag.2, line 32: please change "lag" time "time-lag".
<u>Correct</u>.

Pag.3, lines 19: "the remainder of this section".
<u>Correct</u>.

Pag.3, lines 21: put "available" in between of "levels" and "in radiosonde"
Fine.

Pag3, line 26: ": : :.relevant for the study". Study of what?

It means it is relevant to our study, referring to the subset of IGRA without pilot-balloon stations (as explained in lines 23-25 right above). See rephrasing.

Change in the manuscript:

P23, L26
Where it reads:
from each IGRA station relevant to the study
It will read:
from each relevant IGRA station

Pag3, line 26: " please change "indicated" with "reported"
Okay (referring to P3, L28).

Pag4, lines 7-10: Please put a reference related to the importance of data continuity for climate studies (trends, annual cycles).

Previous literature regarding temporal inhomogeneities in radiosonde time series have mainly focused in instrument changes (responsible for the largest time-varying biases), paying less attention to temporal sampling. Nevertheless, this issue has been assumed relevant to climate studies, as detailed next.

Concerning standards for long-term monitoring of the climate system: Karl et al. (1995) puts the *continuity* and *frequency* of in-situ and other observations – as well as the maintenance of local observations with a long and uninterrupted record – among the "critical issues for long-term climate monitoring".

Concerning temporal homogenization of radiosonde time series for determining long-term temperature and humidity trends: Lanzante et al. (2003), mentioned the interest of "counts of numbers of observations per month as a function of time and by level; these aid in finding sampling biases or less reliable time periods". Moreover, they have computed separate monthly means at 0000 and 1200 UT from 1947-98 CARDS data with the requirement of at least 16 valid values per month. [Since this paper focus on the detection of changing points in temperature time series related to instrument changes and reporting practices, it will also be cited in Sect.1; see P2, L20-23]. McCarthy et al. (2009) [already cited on P7, L8] addresses the problem of "important sampling biases in the raw humidity data, from missing dry observations and missing cold observations"; they also require at least 15 days of observations within a month to calculate monthly mean temperatures.

Concerning the temporal completeness requirements in climate studies on humidity or integrated water vapor: Gaffen et al. (1991), on studying spatial-temporal variability of global tropospheric moisture, discuss temporal sampling, arguing that a minimum of three observations per month is required to obtain an estimate of the monthly mean specific humidity that falls within the 0.1 confidence bands. Using a 1978-85 dataset from 119 stations, they noted that almost 5% of the station months did not meet that criterion. Gaffen and Elliot (1992) [already cited in the manuscript] selected radiosonde stations with at least one observation per day to accurately estimate the seasonal cycle of RH at different locations of the globe. Zhai and Eskridge (1997) used Gaffen´s criterion to study changes in PW over China, further rejecting stations with more than 3 years of data missing at the same level. Ross and Elliot (1996) [already cited in the manuscript], on studying water vapor trends over North America have required at least two months to estimate seasonal anomalies and 10 months to annual anomalies, a month being considering as missing if less than 10 observations during the month were missing.

The manuscript changes to the lines in question consists of rephrasing and inserting the proper references.

Changes in the manuscript:

P4, L9-10

Where it reads:

in addition, the regularity and continuity of humidity profiles are important to trend analysis and to detail annual cycles of humidity or integrated water vapor

It will read:

furthermore, the period of record and the regularity and continuity of radiosonde data is a relevant issue for long-term monitoring of the climate system (Karl et al., 1995), as exemplified by temporal sampling requirements used in trend and seasonal analysis of temperature, humidity and integrated water vapor (Gaffen et al., 1991; Gaffen and Elliot, 1992; Karl et al., 1995; Ross and Elliot, 1996; Zhai and Eskridge 1997, Lazante et al. 2003; McCarthy et al., 2009)

P2, L23 | Add citation:

Lazante et al. (2003)

Additions to the list of references:

Gaffen, D.J., T.P. Barnett, and W.P. Elliott, 1991: Space and Time Scales of Global Tropospheric Moisture. J. Climate, 4, 989–1008, https://doi.org/10.1175/1520-0442(1991)004<0989:SATSOG>2.0.CO;2

Karl, T.F., Derr, V.E., Easterling, D.R., Folland, C.K., Hofmann, D.J., Levitus, S., Nicholls, N., Parker, D.E., & Withee, G.W. (1995). Critical issues for long-term climate monitoring. Climatic Change, 31(2/4), 185-221.. Doi:10.1007/BF01095146

Zhai, P. and R.E. Eskridge, 1997: Atmospheric Water Vapor over China. J. Climate, 10, 2643–2652, https://doi.org/10.1175/1520-0442(1997)010<2643:AWVOC>2.0.CO;2

Lanzante, J.R., S.A. Klein, and D.J. Seidel, 2003: Temporal Homogenization of Monthly Radiosonde Temperature Data. Part I: Methodology. J. Climate, 16, 224–240, doi:10.1175/1520-0442(2003)016<0224:THOMRT>2.0.CO;2

Pag4, Line 10-12: I tend to disagree with the introductory sentence while I like the authors approach in the manuscript; therefore, I think in this sentence you must report which is needed to investigate the simultaneous spatial and time sub-sampling on the data whatever challenging this might be in the practice.

We think that advancing what is needed to investigate the simultaneous spatial and time sub-sampling on the data is beyond the scope of our work. Furthermore, it depends on the application of radiosonde data. Anyway, knowing the content of a comprehensive radiosonde data archive such as IGRA 2 represents a first step. The present paper might be useful in two ways: first, it gives a general picture of the spatial-temporal distribution of in-situ humidity observations worldwide. Second, it provides a tool (dataset) that helps to select either humidity time-series or individual observations based on their temporal and vertical completeness of humidity (*i.e. of simultaneous observations of pressure, temperature and humidity*). It is left to the user to find the data coverage and spatial continuity from the geographical coordinates of stations [please note the final recommendation on P23, L20-23].

However, we agree that the sentence in question can be improved in the following way. First, a station's period of record cannot be separated from its temporal completeness. Secondly, it is appropriate to cite again the paper of Wallis (1998)* since it represents a major effort towards a selection of radiosonde stations for trend studies on a regional or global scale.

(*) Wallis, T.W., 1998: A Subset of Core Stations from the Comprehensive Aerological Reference Dataset (CARDS). J. Climate, 11, 272–282, doi: 10.1175/1520-0442(1998)011<0272:ASOCSF>2.0.CO;2

Changes in the manuscript:

P4, L10-12
Where it reads:
Although the vertical and temporal completeness of station-based humidity time-series can be treated separately from the issues of spatial coverage and record length of stations, it seems to us that studying the completeness of observations in a global, historical data set of radiosonde observations should address these issues too.
It will read:
Although the vertical and temporal completeness of station-based humidity time-series can be treated separately from the issue of spatial (horizontal) coverage, studying the completeness of observations in a global, historical data set of radiosonde observations should address that issue too. This is particularly true regarding the subsampling of radiosonde stations for studies of atmospheric temperature or water vapor trends on a regional or global scale (Wallis, 1997).

Pag. line 24: pressure in not measured anymore in the most recent radiosonde types (e.g. RS-41 operation since a couple of years); please for completeness you may mention this.

Right, not only in the most recent GPS radiosondes but also in some Russian radiosonde systems which used ground radar and a radiosonde without pressure sensor. Being an exception to common radiosondes, this detail deserves a footnote. As to a reference to modern GPS radiosondes (with or without a pressure sensor), Nash et al. (2011) will be also cited in the former footnote 1 on page 30 [as well on page 2 – please see response to general comment 6]

Changes in the manuscript:

P4, L22-24
Where it reads:

In radiosonde soundings, temperature, relative humidity (eventually dewpoint depression too) and wind speed and direction are measured together with atmospheric pressure; geopotential height is indirectly measured from hypsometric calculations but may be missing in radiosonde reports.

It will read:

In radiosonde soundings, temperature, relative humidity (and/or dewpoint depression) and wind speed and direction are measured together with atmospheric pressure, while geopotential height is indirectly measured from hypsometric calculations[1] (but may be missing in radiosonde reports).

______

([1]) Except in some Soviet/Russia radiosonde-radar systems and the last generation of GPS radiosondes – in which pressure is deduced from the (radar or GPS, respectively) profile of geometric height and the radiosonde profiles of temperature and humidity (Zaitseva, 1993; Nash et al., 2011)

P7, L30

Where it reads:

(Dabberdt et al. 2002)

It will read:

(Dabberdt et al. 2002; Nash et al., 2011)

Additions to the list of references:

Zaitseva, N. A., 1993: Historical developments in radiosonde systems in the former Soviet Union. Bull. Amer. Meteor. Soc., 74, 1893–1900, doi:10.1175/1520-0477(1993)074<1893:HDIRSI>2.0.CO;2.

Nash, J., T. Oakley, H. Vömel, and L. I. Wei, 2011: WMO intercomparison of high quality radiosonde systems (Yangjiang, China 12 June–3 August 2010). WMO Instruments and Observing Methods Rep. 107, 238 pp. [Available at www.wmo.int]

Pag.5, line 17-19: this sentence could be a good opportunity to claim for the importance of having high resolution measurements and, therefore, more levels available in the radiosounding report. This is in line with the high resolution BUFR files already flowing to the Met services from more than 100 station worldwide.

The current migration from TEMP to high-resolution BUFR reports will be referred in Sect. 2.1; the fact that BUFR data are not currently available in open archives (e.g. IGRA) should be noted.

Planned change to the manuscript:

P5, L19: Insert sentence about the migration to BUFR messages sent to the GTS, with a reference to Ingleby et al. (2016).

Addition to list of references:

Ingleby, B., P. Pauley, A. Kats, J. Ator, D. Keyser, A. Doerenbecher, E. Fucile, J. Hasegawa, E. Toyoda, T. Kleinert, W. Qu, J. St. James, W. Tennant, and R. Weedon, 2016: Progress toward High-Resolution, Real-Time Radiosonde Reports. Bull. Amer. Meteor. Soc., 97, 2149–2161, doi:10.1175/BAMS-D-15-00169.1

Pag. 7, line 13: "anything but uniform" can be modified "quite heterogeneous".

Correct. After all, there is substantial uniformity as to the launching time and STD levels.

Pag. 7, lines 22-24: these sentence is the "official" IGRA description, but going through the data the authors may realize that among the 2761 stations, many of them are "nearsurface"

stations and not radiosounding station. All the reports are empty (-9999) for many station. These must be clarified and the reported number estiamted in a more precise way.

The NOAA's IGRA web page says, "over 2700", without giving a precise number, since it may increase as new sources are added. Maybe what is lacking here is the time when the data were downloaded (or the date of the IGRA stations list, since the data files were downloaded using a script based on that list). Regarding "near-surface stations", we have checked on this and, based on the dataset complementary to the paper, it seems there are about 20 stations (among the IGRA-RS subset given in Table S1) in which *the first one or two years of record* contains only surface data. This happens mostly during 1946-48. This can be mentioned in Sect. 2.2.

Planned changes in the manuscript:

P7, L22: Indication of the time when the data (or the stations list) was downloaded.
P10, L18: Mention to only-surface data at the beginning of record at some stations.

Pag. 7, line 16: Is the reported typical average accuracy in the troposphere only related to the most recent radiosonde types? Please clarify

(Referring to P8, L16) Recent radiosondes achieve an accuracy of 2% RH, while very old radiosondes have worse accuracy (10%). 5% is roughly the accuracy accepted for the troposphere, the layer that presents the most favourable conditions for measurements. The remainder of the sentence specifies that accuracy can in fact deviate significantly from this round number. At extremely low temperatures is not uncommon to see errors of 15%. Accuracy is compromised in high resolution profiles if the response time turns too long. And then there is precision for RH or DPD. In 1991, Elliot and Gaffen reported that by that time radiosondes presented a precision of about 3.5% RH. But this depends on environment conditions too. Moreover, TEMP reports may have different restrictions for precision over time. For all these reasons, maybe it is wise to withdraw the "typical accuracy", and shorten the whole sentence keeping the main idea (Note: the related references were already cited in other parts, mainly in Sect. 1.3)

Changes in the manuscript:

P8, L15-18
Where it reads:
Note that RH measurements from radiosondes have a typical absolute accuracy of ∼ 5 % on average tropospheric conditions. However, accuracy varies substantially as a function of RH and temperature, degrading in dry or cold conditions to a greater or lesser extent depending on the radiosonde type (Nash, 2002; Miloshevich et al., 2006; Nash, 2015).
It will read:
Note that the precision and accuracy of RH and DPD data varies substantially as a function of RH and temperature, degrading in dry or cold conditions to a greater or lesser extent depending on the radiosonde type.

Pag.11, line 1: please put a descriptive reference for GRUAN, I suggest Bodeker et al., 2016 BAMS.
Pag.11 line 3: It is not true that all of the GRUAN sites are transmitting data to the GTS. A few sites are still working to establish this data flow. This is also connected to the sentence at line 9, reporting the fact that not all GRUAN station are present within IGRA.
Pag.11, line 6: The added value of GRUAN is not only to provide data the quality of which should be "above the average", but to provide traceable uncertainties and a fully disclosed data processing described in peer-reviewed literature. Please add more details to show the real added value coming from GRUAN

We will add the suggested reference, clarify that by "All GRUAN sites" we mean the ones present in IGRA, and add something about the real value of GRUAN.

Changes in the manuscript:

P11, L1-3
Where it reads:
Moreover, the IGRA-RS contains 16 stations that form part of the GCOS Reference Upper-Air Network (GRUAN), (half certified and half to be certified according to current GRUAN status), of which eight (half certified too) are also GUAN stations. All GRUAN sites report
It will read:
Moreover, the IGRA-RS contains 16 stations that form part of the GCOS Reference Upper-Air Network (GRUAN; Bodeker et al., 2016): half certified and half to be certified according to current GRUAN status, of which eight (half certified too) are also GUAN stations. Those specific GRUAN sites report

P11, L4 – Insert text:
GRUAN aims to serve as reference to other radiosonde networks, by providing long-term high-quality records of vertical profiles of selected essential climate variables, accompanied by traceable estimates of measurement uncertainties.

Addition to list of references:

G. E. Bodeker, S. Bojinski, D. Cimini, R. J. Dirksen, M. Haeffelin, J. W. Hannigan, D. F. Hurst, T. Leblanc, F. Madonna, M. Maturilli, A. C. Mikalsen, R. Philipona, T. Reale, D. J. Seidel, D. G. H. Tan, P. W. Thorne, H. Vömel, and J. Wang: Reference Upper-Air Observations for Climate: From Concept to Reality. Bull. Amer. Meteor. Soc., 97, 123–135, doi:10.1175/BAMS-D-14-00072.1, 2016.

Pag.11, eq.1: see my general comments above.

See our response to general comment 4.

Pag.12, line 18: what's the meaning of the "relevant" in this sentence?

It means of interest, i.e. a vertical level among the levels with non-missing humidity data. It is understood in the context, considering the preceding phrase in the whole sentence "M is the number of levels with humidity data above the lowest level", as well as the previous sentence which presents Eq. (1): "the mean vertical resolution of a single humidity sounding was defined by the geometric mean of $\{dzk\}$ for all levels with humidity data in the sounding profile"

Pag.12, line 19: How did you choose the value of the constant temperature T0, please clarify in the text.

The way Eq. (1) was written, the terms inside the product operator are non-dimensional. To prevent *underflow* in a calculation for high-resolution profiles and limited magnitude range for real numbers (because the logarithm terms are inferior to 1), T_0 should rather be chosen so that $RT_0/g \approx$ *average distance between vertical levels* (but this is precisely what we don't know).

Since we used double precision real numbers, and the number of levels with humidity data is not excessively large in IGRA, the result was insensible to the value used (250 K).

However, the term T_0 is not needed providing that Eq. (1) is reformulated in a way that (i) is mathematically equivalent and (ii) it can be safely implemented as is in a computer programme, without leading to underflow when the number of levels is too large. (Although we could alternatively define geometric mean by the anti-log of the arithmetic mean of log-transformed values, this is not needed.) In the revised and straightforward form, each individual term between curved brackets is raised to $1/M$ inside the product operator (instead of doing the product first and then the M^{th} root, as usually seen in the formal definition of geometric mean).

Changes in the manuscript:

P12, L15 | Reformulation of Eq(1):

$$mean\ vertical\ resolution = \frac{R_d}{g}\prod_{K=1}^{M}\left(\bar{T}_k ln\frac{p_{k-1}}{p_k}\right)^{1/M}$$

P12, L19-20 | Delete text:

~~T_0 is a constant temperature in the range of the values found in the troposphere (used in the calculation to avoid overflow in case M is too large),~~

Section 2.3.3: this section refers to the quantification of the number of soundings which can be qualified to estimate precipitable water vapour. I am not sure to what extent this section may really confuse the reader. From one side, I think this is redundant with assessment of other indicators in the manuscript and would add value if then the selected radiosoundings according the criteria reported in this action may ready represent soundings for which an accurate estimation of the water vapor is feasible. i think the authors should clarify that thorn the radiosoundings selected according to the presented criteria allow to calculate an estimation of water vapor content which is the closest possible to the true one given the small number of vertical level available. The accuracy of the calculation of precipitable water vapour for these radiosoundings is anyhow affected by many other aspects: presence of clouds affecting the measurement sensors, homogeneity the water vapor field close to the surface, non-linearity of the water vapor variability along the vertical profile and so on.

We know that the calculation of precipitable water vapour (PWV) is affected by radiosonde data imperfections like defective RH measurements inside (or on the way in/out to) clouds, poor resolution near the surface (where specific humidity is highest and it rapidly varying with height), and possibly insufficient resolution to describe the vertical variations of humidity in the troposphere. Leaving aside wet data biases, our approach deals precisely with vertical resolution.

To our knowledge, the effect of the vertical resolution of in-situ observations on the accuracy of PWV estimation is poorly, if ever, quantified. However, all practical studies try to guarantee that at least the surface and standard levels, normally up to 500 hPa (infrequently, up to 400 hPa) are represented in humidity soundings. This is also the requirement used by NOAA/NCEI in their IGRA-derived data set.

In the Abstract and in other parts of the manuscript, including the concluding section, we clarify that our Sfc-to-500hPa soundings are those with adequate/suitable vertical sampling (coverage and resolution) to *estimate* PWV. Our approach is this: 1) To allow an additional near-surface level to fill the lack of the 925 hPa level before 1992, on the condition that such additional level provide a similar resolution near the surface (firs 2 km); 2) To include soundings in which some standard level is missing, providing that an additional level is given in its close vicinity.

In short, the proposed definition of Sfc-to-500hPa eligible to estimate PWV [P6, L5-13] is more inclusive than the usual requirement; furthermore, it assures that the near-surface resolution is similar to that of current standard levels, regardless of the absence of 925 hPa in a large portion of past radiosonde data. The definition is explained in detail in Sect. 2.3.3, with the aid of Fig.2, and is recalled in a straightforward manner in the beginning paragraph of Sect. 3.3.

We must agree that the wording “eligible to estimate PWV”, as used in the beginning line of Sect 2-3-3, is better than “qualified to estimate PWV”. This applies to the Abstract and to the title of Sect. 2.3.3.

Changes in the manuscript:

P1, L19 and P13, L1: Replacement of “qualified” by “eligible”

Pag.16, line 26: Please provide more details to explain the differences in the maximum range covered by the measurement of these parameters. For example, the way wind and humidity have been measured in the past compared to temperature?

The limited vertical range of humidity measurements was mentioned throughout Sect. 1, and its relationship with the working range of humidity sensors was explained in Sect. 1.3. Regarding wind, we had in mind the limitations of PIBAL (optical theodolite and radar) wind observations, but we cannot ascertain the same about wind profiles using radio-theodolites, as in most of the radiosonde systems of the past.

In the section where page 16 is we focus on humidity data, which must be accompanied by collocated temperature data. So, we think it is wise to withdraw wind and to clarify the difference between the top of humidity and temperature data in RAOB reports. (Of course, the burst altitude evolved with rubber balloons technology; we avoid such detail, since the increase of the vertical range of mandatory levels in the early years must have adapted to the vertical reach of balloons.) For the sake of accuracy, several minor corrections will be made in other parts of the manuscript.

Changes in the manuscript:

P7, footnote 2 regarding PILOT
Insert text:
The common single theodolite technique requires the approximate ascent rate to obtain position, while the double-theodolite method allows a pure trigonometric calculation.

P16, L20-21
Where it reads:
of temperature measurements in RAOB soundings
It will read:
of temperature measurements (i.e. of RAOB)

P16, L26-27
Where it reads:
Contrary to wind and humidity, temperature is usually measured up to the height achieved by the sounding balloon (or close to it), i.e., the burst altitude.
It will read:
Contrary to temperature, which can be measured up to the maximum height achieved by the sounding balloon (burst altitude, although the highest reported level is usually limited by the standard levels in use), the vertical range of humidity measurements depend on the working range of humidity sensors (and, to some degree, on reporting practices).

P16, L28
Where it reads:
at present situated almost 5 km below the burst altitude
It will read:
at present situated almost 5 km below the maximum altitude of RAOB (close to the burst altitude)

P17, L3-4
Where it reads:
faster than the burst altitude
It will read:
faster than the RAOB-top

P22, L18
Where it reads:
to the burst altitude (denoted by the maximum height of the temperature measurements)
It will read:
to the maximum height of the temperature measurements (close to the burst altitude)

P34, L4-5 | Caption of Figure 6
Where it reads:
in the annual soundings of temperature (TMP) and humidity (HUM) across the globe: mean and quartiles. (b) As in (a), but for the vertical extent of soundings.
It will read:
in the temperature (TMP) and humidity (HUM) observations across the globe, calculated up to the highest measuring level for humidity: Mean and quartiles. (b) As in (a), but for the vertical extent of TMP and HUM observations.

Pag. 21, line 1: Did the metadata adhere to any international standards like ISO19115 or WIGOS? Please clarify this aspect.

The answer is no. In fact, such standards are of little use to our (meta) dataset for the following reasons. The present metadata do not substitute IGRA's own metadata (data sources, quality assurance, data format, stations list, periods of record, stations history). They are very specific in content, consisting on a set of values describing completeness of humidity data in a subset of the IGRA v2 main dataset – the sounding data from the stations reporting a minimum of RAOB data. Incidentally, as far as we understand, the metadata pertaining to IGRA do not adhere strictly to 'WIGOS metadata categories' or any other standard for geographic information. To avoid confusion with conventional metadata, we have named the dataset supplementary to the paper without using the word metadata: "A dataset of completeness of radiosonde humidity observations based on the IGRA".

As to metadata of the supplementary dataset, Sect. 4 and 5 of the manuscript provide primary information on the dataset, including a brief description of the represented parameters (see Tables 2 and 3), the spatial-temporal organization of data, data usability and access. The reader is referred to the Readme file for further details. Besides the short summary posted in the corresponding Zenodo's landing page, the Readme file provides the essential, mandatory elements*:* identification, purpose, data resources, sampling, spatial and temporal schema, data file description and format, authorship, contact. Ownership and data police are of course defined by Zenodo.

Pag.22, line 32: "exhibits" instead of the plural.
Correct.

Finally, a general comment about the plots in the different Figure. They are good and clear but the quality of the figures must be improved for the printing. Supplementary material is quite useful for the reader.

The original size of some figures is smaller than others because they were designed as in-text-figures, although observing the minimum recommended size (8 cm). If necessary, this issue can be corrected in due time.