Author's response in red.

K. Kreher

GENERAL COMMENTS
The goals of the GCOS (Global Climate Observing System) Reference Upper Air Network (GRUAN) are to

(i) provide long-term high quality climate records;
(ii) constrain and calibrate data from more spatially-comprehensive global observing systems (including satellites and current radiosonde networks); and
(iii) fully characterize the properties of the atmospheric column.

GRUAN achieves these goals by generating reference quality measurements of upper air essential climate variables. The uniformity and coherence of standard operating procedures at GRUAN sites and the resultant homogeneity of GRUAN climate data records not only provides a global reference standard for operational upper-air network sites, but improves the detection of changes in the climate of the troposphere and stratosphere. Reference measurements are characterised by being traceable to internationally accepted standards, having had all sources of systematic biases identified and removed, and having well characterized measurement uncertainties. It is essential that the procedures used to generate GRUAN data products are documented in the international peer reviewed literature and are subjected to the scrutiny of the international measurement community. The RS92 radiosonde data products of temperature, pressure, humidity and wind are the first official GRUAN data products and this paper is the seminal document fully describing the processing of the measurements into GRUAN data products. As is appropriate, a significant portion of the paper is dedicated to describing the calculation of the corrections applied to account for systematic biases and the calculation of the measurement uncertainties.

Below we provide some suggested minor corrections and points of clarification which we hope will improve the quality of the paper.

SPECIFIC COMMENTS
1) It would be useful to readers if you could include in the abstract and/or summary the number of GRUAN sites that use Vaisala RS92 radiosondes, and to also mention other radiosondes that are used at GRUAN sites.

We will include this in the abstract.
Currently, 13 GRUAN sites launch the RS92 on a regular basis. Other radiosondes include Meteolabor, Modem M10, and Meisei RS-11G.

2) Page 3732, line 23: The heating of the humidity sensor is switched off below -60◦C or above 100hPa. The reason for this could be explained more clearly.

Agreed. We will include a sentence explaining that the switch-off is to prevent overheating of the humidity sensors.

3) A list of collected metadata, or a pointer to a paper/web page where a complete list can be found, would be useful.

Such a document does unfortunately not yet exist. The list of collected metadata is rather extensive, covering several pages, and therefore would be too long to include in the paper.

4) Page 3737, line 19: If there is a reference to the correlation model it would be good if it was mentioned here.

The correction model is discussed in the next Section (5.2.1), and explicitly given by Equation 1.

5) Page 3738, line 24: Refer the modelling of the actinic flux to next section, or mention that this will be explained below.

We will add a statement that the modeling of the actinic flux will be explained below.

6) Page 3740, line 7 and line 13: Table 1 shows the elevation angles 90, 72 and 30 degrees for simulations with the Streamer model. The dates are 21st of June for 90 and 72 degrees and 13th of November for 30 degrees. Figure 4 shows vertical profiles of the simulated total actinic flux on the sensor for 21 June for 30, 72 and 90 deg. This might be a little bit confusing and should be mentioned somewhere in the text.

Thank you for pointing this out. The phrase '21 June' will be removed from the caption of Figure 4.

7) Page 3740, line 17: Is the surface albedo of 0.27 used everywhere in the simulations?

Yes.

8) Page 3749, line 18: This dry bias of 50% at 15 km is very surprising. Does this mean that the humidity sensor never measures humidity above 50%?

The quoted numbers (9% and 50%) refer to relative errors, as was mentioned in the beginning of the sentence. Relative humidity is referred to as %RH throughout the paper.

9) Page 3750, line 18: It is not clear how the humidity sensor affects the vertical resolution of the radiosonde. It may affect the vertical resolution of the humidity profile but not the vertical resolution of the radiosonde.

Thank you for pointing this out. The sentence will be changed to make clear that the time-lag of the humidity sensor affects the resolution of the humidity profile.

10) Page 3752, line 12: The performance of this time-lag correction for the humidity sensor is impressively good.

Thank you.

11) Page 3753, line 3. This division of ΔU1 and ΔU2 by 3 is sufficiently important to warrant just one sentence rather than referring the reader directly to the GUM.

A sentence explaining the origin of this factor 3 will be added.

12) Page 3756, line 12: It is stated in line 12 that the night time differences between GRUAN and Vaisala humidity increase from 0% at the surface to approximately 5% at 10km. However, in line 18 it is said that the night time difference is 7%, which is inconsistent with what is said in line 12.

Thank you for pointing this out. We will change line 9 to read "[...] surface to approximately 7% near the tropopause at 10km."

13) Page 3764, line 9: It would be helpful to the reader to understand how the statistical uncertainties δu and δv are obtained and therefore we would recommend referring to the Appendix or to previous sections where the smoothing procedure is explained.

We will add "(Sect. 6.5)" at the end of the sentence in line 7, and we will make a reference to Eq A5.

14) Section 9: It would be useful if the authors could say something about how the wind speed and directions are derived.

We will add a sentence that explains how the wind speed and direction are derived from the GPS data.

15) Page 3765 line 4-8: We suggest that the order of the sentences starting with Earlier. . . and A novel aspect. . ., respectively, are switched.

Agreed.

16) Page 3765, line 9: It would be good for understanding when reading the summary to know who performed the laboratory experiments.

Agreed. Sentence will be modified to clarify that the measurements were performed at Lindenberg.

17) Page 3766, line 7: It would be helpful if the authors could, again, mention that CFH refers to a frost point hydrometer.

Agreed. Sentence will be changed accordingly.

18) Following the results presented in this paper, the authors suggest several improvements that can be made to reference radio soundings which should be followed to reduce the uncertainties on the measurements, to ensure the high quality of the data and reduce systematic effects of instrumental origin. It would have been helpful if the authors would have mentioned that GC25 refers to a calibration unit.

We will add "calibration unit" after GC25 on page 3767

19) The paper finishes with planned future improvements to the GRUAN processing, namely to use the readings during additional ground check in the SHC (standard humidity chamber) to correct the humidity profile and to update the radiation temperature error correction. From the summary it is not clear what the first statement refers to and what it actually means ‘to use the readings’.

We will modify the text to clarify this.

GRAMMAR AND TYPOGRAPHICAL CORRECTIONS

All agreed.

1) Page 3733, line 4: Change the abbreviation cm to centimetre in the text.
2) Page 3733, line 10: - missing word: Currently, the SHC is in use at several GRUAN stations, and analysis of the SHC measurements showed that at one station the reconditioning of the RS92 was systematically skipped, which lead to a 1–5% RH dry bias.
3) Page 3749, line 7: Replace ’This dry bias and predominantly’ with ’This dry bias

M. Fujiwara

This paper describes and discusses a special data product of Vaisala RS92 radiosonde profiles using a special data processing developed by the GCOS Reference Upper Air Network (GRUAN) Lead Centre. This data product is named as the GRUAN RS92 data product. The details of the radiosonde data processing algorithms are usually not available for the user/science community, and therefore, the information on random errors (uncorrelated uncertainty) and systematic errors (correlated uncertainty) is very limited for measurements that a radiosonde covers. In this paper, the authors developed a full algorithm to process raw RS92 data (temperature, relative humidity, geopotential height, and horizontal winds), and also evaluated the vertical distribution of uncertainty for these variables. Where possible, intercomparisons with independent (manufacturer's) algorithm and measurements from other instruments were made to characterize the GRUAN RS92 data product and to confirm the uncertainty evaluation. This paper is also very valuable in that it discusses the details of the radiation error of the temperature and humidity sensors, which is a major source of uncertainty, and that it describes the detailed calculation procedures to obtain geopotential and geometric heights from pressure and GPS measurements.

One relatively major comment is that it would be helpful for the readers if the authors summarize the actual steps that the radiation correction is applied. The best place to do this is at the end of section 5.2. It was not very clear to me how the results from the field experiments (section 5.2.1) and from the radiative transfer model experiments (section 5.2.2) are used to obtain the GRUAN temperature data. (See below for some related comments.)

The following are all minor comments.

We will add a sentence to the end of Sect 5.2 that briefly describes the actual steps of the radiation temperature correction.

p. 3738, equation (1): Some discussion on the justification of this formulation is needed. Also, pressure is used as a variable; but, does this mean an air-density factor, i.e., the strength of convection (by comparison with radiation)? Actinic flux Ia also has a pressure dependence as the vertical dependence in this case. Please clarify what is meant by this pressure factor.

Infrared radiative heating and cooling are neglected here. The consequence of this treatment may be discussed. For example, the difference between the Vaisala product and GRUAN product shown in Fig. 9 might be in part due to this.

The pressure does indeed signify the strength of convection. We will insert the word "ambient" before pressure in line 24 to make this clear.

Furthermore, Figure 4 shows that above the tropopause the actinic flux I_a hardly varies with pressure. Long-wave radiative cooling is relatively small compared to the shortwave heating, as is illustrated by the 0.04K cooling at 5hPa that is applied by the Vaisala radiation correction for nighttime measurements. We will include a brief discussion of the long-wave effects, but we don't think it will account for the differences observed in Fig 9.

Section 5.2.2 about the Streamer radiative transfer model: The distance between the Sun and the Earth changes in the course of a year (longer for NH summer and shorter for SH summer). Is this considered in the model? Is the contribution of this factor quite small?

The Streamer model does indeed account for the seasonal variation in the Earth-Sun distance. The amplitude of this factor is approximately 3.4%. This is small compared to the uncertainties associated with the radiation correction itself, and the current version of the GRUAN dataprocessing does not explicitly take the variations in the Earth-Sun distance into account. The main purpose of using different dates is to vary the solar elevation angle in the calculations.

It is very interesting (and problematic) to see that the cloud albedo is a major factor for actinic flux uncertainty particularly in the stratosphere where the radiation error becomes large.

True. This inherently affects the uncertainty of the radiation correction because of the limited knowledge of the cloud configuration during the sounding.

p. 3740: The air mass factor of sin(solar elevation angle) may be a good approximation for 90-30 degree, but beyond that, consideration of the curvature of the atmospheric layers may be necessary. The formulation may be found, for example, in the following paper: Solomon et al., on the interpretation of zenith sky absorption measurements, JGR, 1987 http://onlinelibrary.wiley.com/doi/10.1029/JD092iD07p08311/abstract

Thank you for pointing this out. The air mass factor dependence will be incorporated in the next version of the data processing.

p. 3741, line 11: should introduce –> should not introduce?

Agreed

p. 3743, the second paragraph of section 5.3, and Figure 7 and Table 2: If the procedure of spikes removal actually removes only warm spikes (which may be reasonable), the error due to the spikes should be a systematic error and thus the uncertainty due to this should be "correlated." Also, in Table 2, uncertainty from spikes and from rotating radiosonde is separated and considered as independent. But, these two may be closely related.

Although of different origin, the effects of the spikes and the sonde rotation on the temperature profile are indeed hard to distinguish. In our interpretation, temperature spikes are caused by patches of warm air coming from the balloon or (neighboring) radiosonde housing, this is the reason why it is treated separately from the radiosonde rotation.
The fact that only positive spikes are removed does indeed give rise to a small bias. You are right that associated uncertainties of the current spike removal algorithm are correlated. In the next version of the GRUAN dataprocessing the negative spikes will also be removed.
We consider temperature spikes short, incidental events that are removed by the spike filter when they exceed a certain threshold. The rotation period of the radiosonde is variable. Temperature deviations due to radiosonde rotation that are shorter than 10 seconds will be filtered by the spike removal algorithm. The uncertainty associated with the sensor orientation is uncorrelated.

Section 5.6 and Fig. 8: The uncertainty from the radiation error is actually divided into three factors, i.e., radiation field, correction model (obtained from field experiments?), and sensor orientation (the factor g in equation (2)?). Does the "radiation field" here include the contribution from various (but limited) radiative model calculation settings including the cloud albedo setting? It would be good to have a summary about each component and its contribution around the end of this section.

The uncertainty due to the radiation field does indeed represent the uncertainty due to the scene's albedo (cloudy/cloud-free).
The various contributions to the temperature uncertainty are listed in Table 2 and discussed throughout Sect 5.6. If you mean summary of the three factors (radiation field, correction model and sensor orientation) we can do that.

p. 3747, line 7: due to ?

Agreed

p. 3762, line 8: Check the grammar.

We think the grammar is ok, this part of the sentence is equivalent to "where the random component is equal to the statistical noise of z_p."

p. 3766, line 2: It would be valuable to evaluate the uncertainty in IPW from RS92 measurements.

The major part of the precipitable water column is contained in the lower part of the troposphere (say < 3km) where effects like time-lag and radiation dry bias typically are small. The calibration uncertainty is the only relevant uncertainty component, which is composed of the absolute and relative calibration uncertainty. We will add 1-2 sentences mentioning the effect of the calibration uncertainty on the uncertainty in IPW.

p. 3767, line 3: an ascent p. 3767, line 11: "reversed" should be "removed"?

'reversed' and 'removed' are in principle equivalent in this case. We prefer 'reversed' as it expresses the fact that the recalibration that has been applied by Vaisala to the raw data (stored in table FRAWPTU) is being undone.

p.3768, equation (A4): j should be from -M to M?

Agreed

Captions for Figures 5, 6, 7, 8, 9: Please also specify the solar elevation angle as well. This angle depends on the station latitude, day of year, and local time (i.e., universal time and station longitude).

We will include the solar elevation angle in the captions.

P. Thorne

This paper describes important work. Formal reviews have highlighted any technical issues. Here I limit myself to questions of user comprehension. These are enumerated in the order they arise. I hope they are of some use to the authors in revising the text.

1. In the abstract the discussion in the final two sentences around the humidity product performance is a little confusing. The number 15% is given in two distinct contexts but could very easily be conflated by the unweary non-expert. Bearing in mind that many readers get little further than the abstract the mis-impression that this gives should be rectified by nuancing this segment of the abstract for clarity.

We will rephrase the final part of the abstract to make sure that it is clear to the reader that both occurrences of 15% refer to different topics/comparisons.

2. On p. 3729 ln. 13 this very limited set of references may cause some to look somewhat askance. There are many efforts that have been undertaken upon the homogenization of radiosonde temperature and humidity records many of which are better known than these. Consideration should be given to either a more holistic set of references which I could provide the authors upon request or referencing in addition some review paper or assessment product such as IPCC where these are discussed.

We will include additional references, thank you for the offer to supply these.

3. On p. 3729 ln. 16 consider referencing the GCOS ECV paper in press in BAMS by Bojinski and colleagues when mentioning ECVs to provide an easy reference for the interested reader.

Agreed. We will include this reference.

4. It seems odd that in the paragraph starting ln. 16 of p.3729 no explicit reference is made to the GUM and this only arises for the first time instead in Sect. 2. To the metrologically conversant reader it would seem worth making the point in the introduction section that what is outlined here is consistent with this best practices guidance documentation.

OK, we will refer to the GUM in the introduction.

5. p. 3731 ln 28 onwards this paragraph could more directly address that measures of the same measurand by a second method allows a degree of verification of both the processing and the uncertainty quantification. That information is in there but it's not as clear to the reader as it could be.

We will rephrase so that this becomes clearer.

6. p.3732 ln 11 is it worth making the point that the ground system can cope with more than one instrument at a time? This may help later on.

The DigiCora ground system can only deal with one instrument. For the dual RS92 soundings that are discussed in the paper, two separate ground systems were used.

7. p. 3734 lines 18-25 you could make the point that collection of such data and metadata is essential to enable future reprocessings of the data stream.

Reprocessing was already mentioned in Sect 1, but it is a good idea to mention it again in this section.

8. p. 3741 ln 11. Either you are missing the word 'not' somewhere here or I am failing to follow the logic that underlies this conclusion. Either way clarifying this text would seem advisable.

The word 'not' is missing indeed.

9. In several places mention is made of a next product version. A reader may logically infer that you should wait for that to write your paper so you may wish to reduce their propensity throughout.

We appreciate the feeling this may induce at the reader, however as version 2 of the GRUAN processing is an officially released product it is important to properly document it. This documenting includes discussing/mentioning the points where the processing can be improvemed.

10. p. 3747 ln. 20 - or sondes to other measurement systems capable of measuring temperature profile data (although few can measure with anywhere near the vertical fidelity).

Agreed, we will include a statement of this effect.

11. p. 3765 ln. 2 and traceability to calibrated measures I think?

Agreed, we will change the sentence accordingly.

12. Section 10 feels like it ends very abruptly. I thought I had lost a page. Perhaps thought could be given to ending in a slightly less abrupt manner with a more uplifting and definitive concluding paragraph that perhaps reminds readers of the essential GRUAN measurement characteristics and points them to where they can actually get hold of and play with the data.

Thanks for the suggestion. We will include such a paragraph.

I did not review the appendices due to closeness of deadline.
Figure 12 the left hand panel the key overlaps the y-axis making it hard to read. Consider changing the layout somehow to avoid this.

Figure 12 will be improved as suggested.

L. Miloshevich (Referee #1)

GENERAL COMMENT

Typical radiosonde data are not suitable for long-term trend monitoring or other climate-related or satellite validation purposes. This paper describes the GRUAN program's specialized method of processing raw data from Vaisala RS92 radiosondes to make it sufficiently accurate and well-characterized to serve these purposes, mainly by applying corrections that minimize known systematic measurement errors and by rigorously estimating the uncertainty in the corrected data on a point-by-point basis. Good attention was paid to thoroughly considering uncertainties and traceability. The approach that was used allows for reprocessing of the raw data if/when superior corrections or uncertainty estimates are derived. Suggestions for improving the paper are relatively minor.

SPECIFIC COMMENTS

- sec 3: Consider including a picture of an RS92, especially the sensor boom (or reference a picture in the 2011 WMO Intercomparison report).

We will include a picture of the sensor boom.

- p3733, line 16: Miloshevich et al. (2009) is a better calibration error reference than the 2006 paper.

Thank you. We will change the reference to Miloshevich et al. 2009.

- p3734, line 10-11: This sentence needs some editing for clarity. Also, you are suggesting that the reconditioning portion of the ground check procedure is important to eliminate a 1-5% RH dry bias, which I agree is important but also implies that ALL ground check elements in section 3.1 are being recommended...is that true (including the RH recalibration adjustment based on the desiccant)? Also, it might help readers to better understand the purpose of the reconditioning by mentioning in Sect 3.1 that the contaminants produce a dry bias in the RH measurements.
Same paragraph: It is a little confusing that both the GC and the additional GC are discussed in this same short paragraph, which makes it unclear as to what is being recommended.

We will mention the purpose of reconditioning in Sect 3.1
It is important that operators follow the Vaisala-prescribed operational procedure, to prevent mishaps as the earlier mentioned skipping of the reconditioning. This includes performing the ground check. During ground check in the GC25 not only the RH sensors are recalibrated, but the p and T sensors as well, which are important steps. To our knowledge the recalibration of the RH sensors can not be separated from the recalibration of the p and T sensors. Therefore we recommend performing the ground check in the GC25.

The purpose of Sect 3.2 is to emphasize the importance of an additional manufacturer-independent ground check. The additional ground check in the SHC provides traceability at 100%RH. Furthermore it serves to track instrument changes, and to identify possible sensor issues (i.e. defects), as both are more likely to show up at high RH values. The tracking of instrument changes is nicely illustrated by Fig 11. The identification of the dry bias due to skipping of the reconditioning underlines the value of additional testing in the SHC.

line 12: "budget" –> "estimate"?

Agreed

Same paragraph: Can a correction based on SHC measurements at 100% RH really ever be made? Wouldn't it require knowing a lot of proprietary information about the Vaisala calibration function, which is NOT simply a linear function of RH that could be scaled based on a measurement at one point?

It is true that scaling with the measurements at 100%RH will not completely correct the errors at other humidities. However, tests at various well-controlled humidities (standard saline solutions at 11, 33, 75%RH) show that scaling with the reading at 100%RH does reduce the error at other humidities, as is illustrated by the attached plots (figure1 & figure2)

line 21: remove "and". "include" –> "including"

"The metadata" remains subject of this sentence. We will place "and in addition to all instrument parameters" between commas to improve readability.

p3735, line 9: It might be more clear to readers if it was mentioned in Sect. 3.1 that the RH GC correction is not applied for GRUAN soundings due to its very large uncertainty. Is it always necessary to reverse the RH GC correction, or is it possible to use Digicora settings to simply not apply it in the first place?
It's very important that readers be clear that it IS recommended to apply the sensor reconditioning (heating) but NOT the RH GC correction; otherwise some users might skip the reconditioning.
Would it be advisable to inform readers specifically how the Digicora settings must be adjusted from their default settings in order to not apply the Vaisala corrections in a manner consistent with GRUAN soundings?

The RH data are always recalibrated when the ground check has been performed. To our knowledge this can't be circumvented by changing a parameter setting in the DigiCora program. However, the ground check should be applied as this includes the important recalibration of the temperature and pressure sensors.

p3736, line 19: Please indicate that this refers to "Vaisala-corrected" T measurements.

Agreed.

line 21: The wording of this sentence suggests that the satellite measurements are "correct" and the RS92 measurements have a 1 K bias. 1 K is a lot, and is inconsistent with the Luers, Vaisala, and WMO findings, and I would guess that the discrepancy could mostly be attributed to the satellite approach. Perhaps the sentence should be reworded, and if possible add some info about uncertainty in the satellite approach.

Sun et al. 2013 report a warm bias of 0.5-1K (depending on solar elevation angle) at 17 hPa for RS92 as compared to satellite observations. The accuracy of the satellite retrieval is approximately 0.2-0.3K in the middle stratosphere. We will change the sentence accordingly, and we will provide references on the accuracy of the satellite measurements.

Fig. 5: suggest minor tick marks at 0.1 K intervals (rather than 0.125 K).

Agreed

Sec 5.2.1: It would be helpful to see a photograph of this chamber.

For a picture of the chamber, see slide 5 of:
http://www.dwd.de/bvbw/generator/DWDWWW/Content/Projekte/Gruan/Downloads/ICM-6__2014/pres__304__Dirksen,templateId=raw,property=publicationFile.pdf/pres_304_Dirksen.pdf

p3738, line 18: "extend" –> "extent" (also p3744, line 6)

Agreed

p3738, line 21: It appears from Fig. 2 that 20 s is most but not all of the way to the ultimate (relatively stable) T difference. Does this choice of 20 s represent a component of uncertainty? For example, why not use the average over a time period once the T has stabilized? Is the idea here that a time is chosen that fully includes the rapid rise but does not include much of the slow rise due to uncertainty over its relevance to actual operational conditions? More importantly, please make clear whether this selection of 20 s varies with P (presumably it is longer at lower P).
Eq (1) and related text: Just a curiosity question...is there any advantage in accuracy to be gained by instead characterizing dT in terms of several (I/P) relationships for different v values, and interpolating between the v curves? I'm wondering if each curve would have a tighter grouping of points in Fig. 3.

Thank you for the suggestion. We will investigate this for the future version of the data processing.

p3739: It is not entirely clear whether consideration is made for the difference between the solar spectrum at ground level in Lindenberg where the experiments were done and the modeled spectrum as a function of height and geographic location (assuming that the T sensor is sensitive to the spectrum). Also, what is the spectral response of the filter (does grey mean that it affects all wavelengths equally)?

The GRUAN radiation correction only takes into account the total actinic flux, spectral changes are ignored. Spectral changes can occur, depending on latitude and altitude, with the most notable effect the increase of UV due to the reduction of absorption by ozone. However, the wavelengths below 310 nm represent approximately 1.5% of the total energy of the solar spectrum. Therefore, we think that this error is negligible compared to the uncertainties associated with other terms of the radiation correction.
The filters indeed have uniform wavelength attenuation between 350-2500 nm, see: http://marketplace.idexop.com/store/IdexCustom/PartDetails?pId=385

p3742, sec 5.2.4: Can the magnitude of the nighttime cooling correction be characterized, and can anything be said about how the Vaisala correction was determined (pure modeling?)?

With our set-up it is not possible to determine the nighttime cooling correction, therefore we rely on the Vaisala correction for nighttime measurements.
The Vaisala correction is based on a combination of modeling (comparable to, e.g., Luers1990) and measurements with temperature sensors with coatings of different reflectivity.

p3747, sec 5.7 and fig 9: Is it correct that there is statistical uncertainty in the mean differences (red curve) due to the sample size, given by the mean divided by sqrt(N)? If so, consider adding a red shaded region to Fig. 9 indicating this uncertainty. Also, why is only one sounding used to represent the modeled uncertainty, rather than the average of the actual GRUAN uncertainty estimate for the 29 soundings?

We will add a shaded region to indicate the statistical uncertainty of the differences.

p3747, line 25: Consider adding "about" before -40C, because there is also time-lag error at warmer temperatures that may not be totally negligible (e.g., at -35C when the humidity gradient is very steep).

Agreed.

p3748, line 4: Regarding the convention to express humidity as %RH over liquid water, the Associate Editor specifically asked me to comment on its appropriateness. It is indeed appropriate, as you know, but I recommend adding a sentence or two for clarity, and possibly a reference to Miloshevich et al. (2006, Appendix A) where this issue is discussed at length. Some may be confused that RHw has no physical meaning below the homogeneous ice nucleation temperature of -35 to -40C where even tiny droplets freeze spontaneously. Nonetheless, there are many expressions for RHw,
and the key point to make in the paper is that accurate conversion to other water vapor units such as vapor pressure, mixing ratio, or RH with respect to ice requires using the same saturation vapor pressure formulations as assumed by Vaisala in determining their RH calibration function, which are Wexler (1976) and Hyland and Wexler (1983). Errors from using other common SVP expressions are shown in Fig. A1 of the above reference. The recommendation to use Wexler (1976) was also made in the 2010 WMO Intercomparison report (Nash et al. 2011).

Thank you for clarifying this. We will include a reference to Miloshevich2006 and a few explanatory sentences.

p3748, Sec 6.2: The choice to not apply the Vaisala GC correction in GRUAN processing is important and well-described here. The accuracy of RS92 soundings globally is fundamentally tied to how much care operators take in changing the desiccant, and this paper could do a great service to the community by strongly advising all RS92 users to either abandon applying the GC correction or at least use a correction of 1% RH as a criterion for changing the desiccant. I have seen tropical soundings where the correction was allowed to reach 7% RH. GC corrections of even 2-3% RH are obviously in error because they lead to negative RH values in the stratosphere.

Also, it might be advisable in order to be clear, to repeat in this section that the other part of the launch prep, the heating of the sensor in the GC25, is necessary in order to drive off contaminants.
Also, if fresh desiccant generally indicates a GC correction of about 0.0% RH, apparently the Vaisala calibration is fundamentally based on the assumption that fresh desiccant really does produce an environment of 0.0% RH. However, laboratory desiccant experiments, plus private communication with Vaisala, suggest that in reality the environment above fresh desiccant is 0.3-0.5% RH. This suggests a possible systematic calibration bias of this magnitude (at 0% RH and relative warm T).

We agree with you that the ground check recalibration of the RH sensors using the desiccant is a source of systematic error, and that this is a reason for concern. The ground check needs to be performed because of the important recalibration of the p and T sensors that is performed as part of this procedure. However, we will include the recommendation to regularly check the quality of the desiccant and to replace it when the ground check correction exceeds 1% RH.
Thank you for sharing the information on the limitations of the desiccant.

p3749, Sec 6.3: This section would be more robust if the profile data upon which Table 3 is based were presented, for example nighttime profiles of RS92/CFH differences (perhaps 1 km vertical averages), after correcting for the time-lag error.
This correction may require more justification because it is empirical (not physically-based like the other corrections), and as such it may include other sources of error such as a constant calibration bias at 0% RH that is not well-represented linearly as a scale factor, or CFH-related considerations such as its suspected bias above -20C or so when there is liquid condensate on the mirror, or mismatched time responses between CFH and RS92 at low T. Such complexities may be responsible for a strong RH-dependence in the "calibration bias" from CFH/RS92 comparisons found by Miloshevich et al. (2009, Fig. 9a), even though as stated the calibration error should theoretically be just a function of T. This is also a less reliable correction because it will be in error if Vaisala ever changes their calibration function, since such changes by Vaisala affect even the raw data used by GRUAN. I suggest that the paper reflect more caution about how well understood and reliable this correction is. It is appropriate that the stated uncertainty is large in that it is roughly equal to the correction itself.

Thank you for pointing this out. We will present the data on which the correction is based, and we will expand the discussion of the issues concerning this correction.

p3749, Sec 6.4, line 18 or elsewhere: It may strengthen the case to also reference the radiation error found by Miloshevich et al. (2009, Fig. 9b), which was derived from a different CFH/RS92 dataset and showed a similar height-dependent bias.

We will include this reference.

p3750, line 1: This could be clarified a little, for example: "dT is THE RADIATIVE HEATING OF THE TEMPERATURE SENSOR calculated from...where f is an empirical scale factor that accounts...".

Agreed, we will change the sentence accordingly.

p3750, line 12: The radiation error for post-2009 sondes has been reduced by half from pre-2006 sondes? Since that is substantial, can the data be shown from which the post-2009 values in Table 4 were derived?

The attached plots (figure3 & figure4) show the response of the temperature and the humidity sensors to solar irradiation during experiments performed at approximately 250hPa. The heating of the humidity sensors is derived from the change in the measured RH, basically using Eq 5. The other plot (figure5) presents the ratio of the heating of the humidity and the temperature sensors, which is distributed around 13 for the B-series and distributed around 6 for the E-series.

p3751, line 9: Since the reference is to a website (non-permanent), the authors are invited to also include the figure and caption in their paper if they wish (or alternatively, add a sentence of explanation along the lines of the figure caption that the fit is based on the same time-constant data referred to earlier).

Thank you, we accept the offer to include the figure and caption in the paper.

p3752, line 20: I'm a little confused as to why the GC25 readings (terms 2 and 3 in Eq. 11) are included in the uncertainty estimate, when these same readings are deemed unreliable for applying as a correction due to the varying desiccant contamination. Isn't this essentially adding the amount of desiccant contamination as RH measurement uncertainty?

You are right that these terms only reflect the quality of the desiccant and should therefore not be included in the uncertainty estimate. This will be changed in the next version of the data processing.

p3755, line 21: Why is the Vaisala radiation correction not applied to the FLEDT data when it IS applied to the GRUAN data? Doesn't that make Fig. 14 an apples-to-oranges comparison when it doesn't need to be?

Not necessarily. The DigiCora settings that are employed in Lindenberg are such that the radiative heating correction of DigiCora version 3.63 (and earlier) is used. This version of the radiation correction was implemented and published in 2005. The 2005 version of the radiative heating correction still is in use at various stations, including several GRUAN sites.

p3756, line 2: So do the differences represent only differences in the processing of the raw data (mainly GRUAN spike removal)?

That is correct.

p3756, line 10: Rather than "due to the time-lag correction" should this be "due to differences between the GRUAN and Vaisala time-lag corrections"?

No, in the DigiCora settings used in Lindenberg the Vaisala time-lag correction is disabled.

p3757, line 11: This would seem to be an offset at -40C that is attributable to the reference instrument. Can any possible explanation be given as to why the FPH would exhibit a behavior below -40C that the CFH doesn't exhibit? This is potentially troubling as it calls into question the reliability of the reference instrumentation. Upon reflection, the different behavior of the FPH may be the portion that is warmer than -40C rather than colder. Another possible explanation would be differences in the typical RH values and hence the corrections in the lower levels of the Boulder soundings relative to the other sites.

The controller of the FPH is different from that of the CFH. Further investigation in this matter is needed.

p3757, line 20: It is implied here that there may be an underestimate in the calibration correction in a portion of the temperature range (though this correction appropriately has a large uncertainty). Note that the calibration correction reported by Vomel et al. (2007b) is smaller than the RH and T-dependent calibration correction reported by Miloshevich et al. (2009). The inclusion of an RH dependence in the latter correction is one possible explanation, namely that it is not purely a T-dependent bias because it is empirical and includes other small sources of error, and has an RH-dependence that explains differences between sites that have different typical RH profiles. This is just speculation.

Thank you for pointing this out. We will further investigate this.

p3761, line 4: Could there be a problem with using the sonde temperature measurement right at launch? The temperature may jump at launch because there is no ventilation so T will be too high for a couple seconds.

It is true that this can lead to a too high temperature measurement at launch. The resulting error on the altitude determination is small enough to be ignored:
A 1K temperature error leads to an error of approximately 0.4% in the air density, and consequently to an error of 0.4% in the altitude correction. The correction equals the altitude difference between the station's barometer and the launch site. At Lindenberg, where the station-launchsite altitude difference is on the order of 10m, the resulting error is a few cm.

p3766, sec 10.1: I suggest several changes regarding the RS92 launch preparations discussed in the second and last items in this list, first by moving the last item to the third item since they are related. You are recommending two inconsistent things: following the manufacturer-recommended launch prep procedure, but not applying the desiccant-based GC correction. It should be clarified that it is important for users to apply the recommended sensor reconditioning (heating) to drive out contaminants that give a dry bias, but users should not apply the GC correction (especially if it is >1% RH).
While GRUAN reverses the GC correction when reprocessing from raw data, everyone else uses the Vaisala-processed data product, and since this section is about recommendations to others for making reference-quality soundings, the recommendation should be to not apply the GC correction in the first place.
Another related recommendation to mention in this section is for users to select the FLEDT output file rather than the default EDT output file, because EDT has only integer values for RH, and this low precision results in poor accuracy for low RH values as well as difficulties and inaccuracies if a time-lag correction is

applied. It might even be helpful to describe how users select FLEDT output files with the Digicora.

We will move the last item in the list to position #3.
As stated before, the ground check involves an essential recalibration of the T and p sensors, therefore we recommend to follow Vaisala prescribed operational procedure which includes the ground check.
We will include a remark regarding the selection of the FLEDT output, however we would like to point out that the Vaisala data product does not qualify as a reference product because of the application of proprietary, black box algorithms.

p3767, line 16: Is this really possible? Perhaps Vaisala should be asked about this possibility. The RH calibration function is not linear; I think it is a non-linear fit through several RH calibration values and therefore a linear scaling would be inappropriate (except possibly above the highest RH calibration point). At a minimum the statement should be hedged by beginning the sentence with something like "Investigate using..."

See the answer to your specific comment #4 (p3734 line 12)

p3767, last sentence: curiosity question: changing the dT correction value would also affect the RH radiation correction...would the factor f need to be rederived?

In additional radiation experiments we will investigate and evaluate the effect on the humidity sensors as well. This could indeed lead to new values of the factor f.

Aside to authors: Just FYI, not for the paper, the RS92 heating cycle for the dual sensors effectively reduces sensor icing but does not entirely eliminate it in two situations. One is well-known (below -60C for ice-supersaturated conditions), but the second recently became apparent when someone was exploring curious results just above low-level cloud tops in Greenland that turned out to be instrumental artifacts (this interpretation was confirmed by Vaisala). For a portion of a heating cycle, the measurement sensor may be somewhat iced after emerging from cloud top and lead to elevated RH measurements and inconsistency of the RH and T measurements until the heated sensor becomes the measurement sensor. It might be worth trying to detect this condition from a relatively large jump in RH between the sensors when they switch dominance, at least for liquid water clouds when cloud top can be approximately determined within a heating cycle, and add a flag or increase the uncertainty.

Thank you for the background information and the useful suggestion. This effect seems to be closely related to the evaporative cooling of the temperature sensor. Perhaps the conjunction of data from the T & RH sensors can help in identifying, flagging, and perhaps, the correction of this effect.

Anonymous Referee #2

1 Overall recommendation

The manuscript by Dirksen et al. describes data processing and uncertainty estimation for Vaisala RS92 radiosondes as performed in the GRUAN network. Scope and content of the manuscript are very well suited for AMT. While the manuscripts does not present a lot of fundamentally new information, it summarizes correction methods and their uncertainties quite well. It is also very important to document these in the scientific literature. After a few minor revisions, I think the manuscript will be well suited for publication in AMT.

2 General points

I think it is very important to have the relevant final numbers for temperature uncertainty and humidity uncertainty, geopotential height uncertainty, pressure uncertainty (for day and night when they differ) mentioned in the abstract. This important information should be available from the abstract, without having to read the paper first. So please add these numbers to the abstract.

We will include these numbers in the abstract.

Also, I highly recommend adding a table (or figures) that summarizes the obtained final uncertainties for temperature, humidity, geopotential height and wind as a function of height / pressure. This table / figures should also include the uncertainties given by Vaisala (and flag where Vaisala processing and/or uncertainty is believed to be wrong or biased). While this information is scattered in the manuscript, it would be very helpful to have it collected in one easy place.

Thank you for the suggestion, we will include these in the manuscript.

Section 3 "Description of the RS92 Radiosonde" is not balanced. There is a lot of discussion (on pg 3732, 3733) of the humidity measurement/ sensor, but little or no corresponding discussion of the pressure and temperature sensors. Either discussion of the p and T sensors needs to be expanded here, or the details on humidity should be moved to Section 6 on humidity.
The same might be true for Section 3.2: I am not sure if this is that helpful here. It might be better to move it to Section 6. I agree with Larry Miloshevich's review, that the discussion of the ground-check is a bit scattered throughout the manuscript, and that the final recommendations on the ground-check and undoing of the Vaisala humidity correction may not be clear enough. The authors should address this. Since the authors think that one-point calibration of the humidity sensor by the ground-check is not useful, they should make this very clear (e.g. also in the recommendations in sections 10, 10.1).

We will move details on the humidity to section 6, including section 3.2.
We will concentrate the discussion of the ground check, and make the final recommendations clearer.

What is the measurement principle of the humidity sensor? Does it measure water vapor partial pressure that is then processed to humidity?

The humidity sensor does indeed measure the water vapor partial pressure (wvpp). The sensor's dielectric is sensitive to both wvpp and temperature, so that it effectively measures the relative humidity. An interesting physical interpretation of the change of the dielectric's permittivity is given in Sect 2.3 of Wildmann et al. 2014 (www.atmos-meas-tech-discuss.net/7/4407/2014/).

What is a likely physical mechanism / explanation for the "out of the air" humidity correction in Table 3 / Section 6.3. Additional explanation is needed here (also on page 3757).

The temperature dependent calibration correction of the humidity sensor is an empirical correction derived from comparisons with coincident frostpoint hygrometer observations, where the frostpoint hygrometer is considered a reference instrument. We interpret this bias between the RS92 and the frostpoint hygrometer data as an inaccuracy in the calibration function of the RS92.

Is it possible to use the humidity displayed in the stratosphere as 0% RH for sensor calibration? Better than the problematic ground check?

Stratospheric measurements are not suitable as a 0% RH reference because of the risk of sensor icing as the heating is switched off at -60C/100hPa. Furthermore, high RH values can occur in the stratosphere. For example, we have observed values as high as 20% RH over Sodankyla (at an ambient temperature of -80C).

Is it possible to give uncertainties on the wind data?

We will present uncertainties on the wind data.

3 Detailed Comments

pg. 3278, line 26: I very much doubt that routine radiosondes have measured "up to about 40 km" for decades. 30 km seems more realistic. Water vapour is measured only up to 10 km (at best).

40km denotes the upper limit for radiosoundings, it is not meant to reflect the daily operational practice.

pg. 3729, lines 13 to 15: Suggest to drop the sentence after Wang et al. (2013). There are lots of approaches for homogenizing. This should not be simplified and judged here.

We will drop the sentence as suggested. Furthermore we will add several references as a consequence of point 2 raised by P. Thorne.

pg. 3730, 3731: I have to admit that I am not happy about calling a systematic bias "correlated uncertainty" and random errors/uncertainty "uncorrelated uncertainty". Whenever I hear correlated and uncorrelated I have to ask "correlated with what?". But I guess this is GUM-speak and needs to be used now.

We will check the wording in this section in order to make clear that the correction for a systematic bias is associated with a correlated uncertainty, and that the correction for a random error leaves an uncorrelated uncertainty.
As a the GRUAN data product is a reference product we aim to be compliant with the terminology used in the GUM.

pg. 3731, line 20, "does not reduce its uncertainty" -> "does not reduce its systematic error" or "does not reduce its systematic uncertainty"

We will insert “systematic” before uncertainty.

pg. 3733, lines 1 to 3: Are the p and T sensor calibrations SI traceable like the humidity calibrations? This should be stated here. Generally, throughout the manuscript, the authors have to be more careful about too much focus on humidity, and the lack of similar detail on the temperature measurement (and pressure). Temperature is also a very important climate variable! Also a statement should be made about how good these Vaisala stated accuracies are. Are they very conservative? Are they consistent with the precisions found in twin flights? With the findings of the authors?

The calibrations of the T and p sensors in the Vaisala CAL-4 facility are SI traceable. A statement of this effect will be included in the manuscript. It is certainly not our intention to disregard the intricacies of the temperature measurements. For example, the approach for the correction of the radiation temperature error is discussed extensively. However, for the humidity measurements more correction steps are applied.
We will include an assessment of the Vaisala-stated accuracies. The uncertainty profile of the temperature measurements that we derived from twin flights is consistent with findings by Vaisala. Figure 9 in our paper shows approx. 0.8K (k=2) at 30km, Fig. 3 in [Vaisala2007] shows approx. 0.4K (k=1) at 10hPa.

pg. 3736, line 17: Add version for Digicora software.

Will add (< version 3.64) to the sentence

pg. 3736 lines 20 to 22: This 1 K warm bias needs to be qualified. Do the authors think that RS92s have a 1 K bias? If so, they should correct for it in the GRUAN processing. Does that mean that temperature measurements from Vaisala RS80s (which are colder in the stratosphere, Steinbrecht et al. 2008) were correct? Or is it likely that the radio occultation temperatures are too cold? Please clarify.

See the answer to the comment by L. Miloshevich who raised a similar question.

pg. 3737, line 6: add "(see Section 5.5)" after "negligible"

Agreed

pg. 3738, line 18: extend -> extent (also in other places!). It would make sense to run a spell-checker over the manuscript!!

We did use a spell checker. However, we could do with a grammar-checker that is capable of handling latex source files.

pg. 3742, section 5.2.4: Explain why there is no correction for long-wave cooling at night in the GRUAN processing. How big is the correction in the Vaisala software?

Apparently the sentence on line 10-12 leads to confusion. We have not derived a GRUAN correction for long-wave cooling, therefore we use the Vaisala correction at night (as is stated on p3743 l1-2). The Vaisala correction for nighttime long-wave cooling is 0.04K at 5hPa (source: Vaisala continuity website).
We will add a statement in Sect 5.2.4 to make clear that the radiation correction in the GRUAN data processing uses the Vaisala correction for the nighttime measurements.

Section 5: One thing that never became clear to me is how solar zenith angle (SZA) is accounted for in the GRUAN radiation correction. Is there a lookup table for the conditions given in Table 1? Is that interpolated to the actual SZA? Depending on generalized cloud conditions? Section 5.2.2 only clarifies (to me) what is used for uncertainty estimation, but not what is used for the actual radiation correction. I think this needs clarification (and maybe a repetition later in the manuscript).

RTF simulations are performed for the solar zenith angles listed in Table 1. These simulated actinic fluxes are stored in a look up table and the I_a for the actual SZA during the sounding indeed follows from linear interpolation between the reference angles (listed in Table 1). We do indeed use generalized cloud conditions, namely (p 3741 l5) "I_a is assumed to be the average of the cloudy and the cloud-free case" For clarity we will mention this again in Sect 5.6.

Section 5.3. Is it also possible that there are cold spikes at night, due to a cold balloon? (radiative cooling at night and large heat capacity making the balloon colder than ambient in the stratosphere)

Cold spikes are indeed possible. The current version of the GRUAN data processing does not correct for these, but it will be included in the next version.

page 3758, lines 26, 27: What is meant by calibrating? Is that not simply addition of a constant value, so that the GPS altitudes start with the altitude of the launch position? Please be more specific.

Correction is probably a better choice of word here.

Section 9: Please explain the principle of the wind measurement. Is wind derived from the change of GPS position over time (derivative), or is it derived from Doppler shifts of the GPS carrier frequencies? Also, it would be good to give some numbers for the statistical uncertainties (page 3764, lines 9, δu, δv and then for $u(s)$ and $u(\Phi)$)

Wind is derived from changes in the GPS position. We use the zonal and meridional wind vectors (u, v) that are derived by the Vaisala processing software from the changes in the GPS position. The statistical uncertainties on δu, δv, and $u(s)$ are 0.4-1 m/s (after smoothing by a low pass filter which reduces the temporal resolution to 40s). The statistical uncertainty on $u(\Phi)$ is approximately 1 degree (after smoothing).

Table 2, caption: please add "temperature" before "uncertainty" to the caption, so the reader knows what this table is for.

Agreed.

Table 7, caption: Typo "temperuture"

Agreed.