RC3 Review of manuscript by Ingleby et al., "On the quality of RS41 radiosonde descent data", submitted to Atmospheric Measurement Techniques

This paper presents interesting results describing the quality of RS41 radiosonde measurements made during descent after balloon burst, with and without a parachute. One finding that should be repeatedly stated in the manuscript is the reduction in positive temperature biases during descent when a parachute is used. An attempt is made to develop and apply a correction to these biases, based on descent rates, that includes descents with and without parachutes, but it seems like the very wide dynamic range of descent rates may require two correction equations - one for very fast descent rates (no parachute) and one for slower descent rates (parachute).

I had hoped for some deeper and more quantitative discussion of how bias-corrected RS41 descent data help improve NWP forecasts, but the paper is already quite long. I was also hoping for more results and discussion of ascent/descent RH comparisons, even though these are restricted to the troposphere and are therefore made more difficult by the high variability of RH below the tropopause.

The impact on NWP forecasts seen was quite small – in part because Europe is already well-observed. Descent data from ships and remote islands should have more impact per profile, but would need quite long experiments to show a clear signal.

RH: we wondered if time in the stratosphere would adversely affect the RH sensor, but were pleasantly surprised to see that it doesn't. Apart from the Finnish descent RH results being slightly worse we didn't see any systematic differences (unlike winds and temperature). Yes, the variability of humidity in the troposphere is a complicating factor.

In general, there are too many figures (27) in this manuscript, making it a lengthy and arduous read. Perhaps photographs of bursting (Fig 3) and burst (Fig 4) balloons can be moved to the supplement? I have also indicated in my comments where I think some figures can be eliminated.

We know that the manuscript is long. We have moved Figure 4 to the supplement but included two images from it in a revised Figure 3 – we were reluctant to move all the photographs to the supplement. We have moved two other figures out of the main manuscript as suggested.

My specific comments are:

Table 1: For me, numerical station codes are meaningless unless they are further identified by location in the caption. It would be interesting to know which stations launch the larger balloons, and why.

Added "The names/locations of the stations can be found in https://oscar.wmo.int/oscar/vola/vola_legacy_report.txt ." (I wasn't sure if the reviewer wanted names or lat/long or both.)

Balloon size is a compromise between cost and the desire to get profiles as high as possible – different NMSs make different decisions. Some stations use automated launchers, the earlier systems could not cope with larger balloons, the latest systems can use most balloon sizes.

Figure 7: This would be an excellent place to show the variability of descent rates in each country. A horizontal error bar (±1 std deviation) on each mean profile at 2-3 altitudes would be illustrative. Slightly offsetting the error bars in the vertical will allow them all to be viewed with clarity.

I tried showing error bars for all countries but the plot became too 'busy'. As a compromise I have included the error bars for German stations (and changed the text accordingly).

Line 183: When I calculate the reciprocal of 14.9 seconds I get 0.07 Hz (not 0.06 Hz).

Changed to 0.07 Hz.

Line 184: This sentence tends to imply that the increase (decrease) in period (±0.6 s) is linear with an increase (decrease) in tether length, which it is not.

Sentence added to make the nonlinearity clear.

Line 228: Here I think "ascent" should be changed to "descent"

Changed to 'descent'.

Line 239: The errors introduced into descent wind data are largely dependent on the vertical (temporal) averaging window. Errors can be made very small by using an averaging window that is similar to the period of the pendulum motion. If higher resolution wind data are required, the magnitudes of the errors become quite dependent on the averaging window.

We agree, but using an optimal averaging window is non-trivial. We have added (in italics here) a note to the end of the previous paragraph: 'The amplitude of oscillations seen under this scenario could introduce error in the processed winds and is a possible area for future study *as is the optimal filtering for descent winds*.'

Figure 11 and Line 283: To me, the "fit the background more closely" in line 283 implies that the mean O-B differences for descent in Figure 11 should be smaller than those for ascent, particularly at upper altitudes, which they are not. I think your statement in line 283 is instead referring to the variability of O-B differences, not the mean differences. "More closely" makes me think of reduced bias, not reduced variability. Perhaps "fit the background with less variability" is a better description of what is shown in Figure 11. Is less variability or smaller biases in O-B differences more important overall? I would think smaller biases.

Changed to "fit the variations in background wind more closely". Both small bias and small standard deviation (SD) in O-B are important. The biases are small for both ascent and descent, the main difference is in the SDs.

Line 289: What is the typical (range?) of averaging windows used to reduce pendulum effects on wind data? Do these change from flight to flight? For a given flight are the averaging windows different for ascent and descent?

The averaging window does not change from flight to flight or ascent to descent. The text has been changed to 'In both cases a time filter with a fixed window is applied to all profile data. Improvements to this are possible as is clear from Section 2.5.'

Figure 12 and Line 294: The low vertical resolution in Figure 12 makes it difficult to distinguish between pendulum motion and noise. Can you instead display some sections of the profiles at higher vertical resolution so the pendulum motion becomes more obvious?

Two sections are now shown in more detail (it is now Figure 11).

Line 297: Does "more pendulum motion" imply a higher frequency or a larger amplitude of oscillations? Which has the greater impact on the standard deviations of O-B averages?

Changed to "larger amplitude pendulum motion". Which has the greater impact – probably the amplitude.

Table 2: Are these statistics for all radiosonde sites, a subset of sites, or a single site? I assume "Sample" is the sample size for either RS-RO or RS-B observations, but don't know which. Adding the sample sizes for the observations not represented by "Sample" would be helpful (and will show your point about the smaller number of RS-RO differences.

Added clarifications to the caption: "for all stations" and "the comparisons with the background are limited to the profiles collocated with RO".

Also added sample size to the mian text: "- note that the sample size is much smaller than for the O-B statistics (137 versus 2190 at 70 hPa)."

Figures 14 and 15: I don't see the need for Figure 15 when Figure 14 clearly shows the warm temperature biases of the descent profiles. Lines 325-326 are adequate to explain why a simple time lag argument for the biases is not adequate.

Figure 15 deleted.

Line 331: "by 300 hPa" is confusing. Less confusing is "below 300 hPa".

Changed to "below"

Figures 18 and 19: I don't see the need to have a Figure each for specific and relative humidity differences when the radiosondes directly measure relative humidity. Also, specific humidity calculated from relative humidity measurements requires pressure and temperature data that can themselves produce biases in specific humidity values. It's difficult to argue that 2 separate figures are needed to support 4 sentences in the text.

Figure 19 moved to supplement.

Line 377: Why wasn't the Prague site included in Table 1 and Figure 7? Here, it just comes out of the blue as a site with many useful descent soundings.

We have added Czechia to Table 1 and explained in the caption to Table 1 that the data were not provided to ECMWF in real-time (hence the absence from Figure 7). Some NMSs were/are reluctant to provide 'experimental data' via the GTS (especially using the dropsonde template).

Line 378: What is a "comparable level"?

Changed to "there were about 528 000 points at which ascent and descent could be compared"

Line 381: Here, the change in writing style (and presumably authorship) is obvious. The lead author may want to reconstruct some of the sentences to flow as the previous text did and to help with clarity in this section.

Yes, section 4 was written by MM. It had been rewritten in places to improve the flow/English, but more has now been done. Some of the figures have been modified to show mean differences as dashed lines and SDs as solid lines for consistency with section 3.

Line 384: I have no idea what this means: "was expected warm bias for 06 and 12 terms, and cold bias for 00 term due to the diurnal variation.

Much of this paragraph has been rewritten for clarity (shortening it where possible).

Line 387: I presume "dividing the sample into groups of 1000 m" refers to the binning of differences by altitude using 1000 m-thick layers?

Yes, rewritten as "After dividing the sample into bins of 1000 m in altitude".

Line 390: Does "separating data into 00, 06 and 12 UTC groups" indicate that the Prague site launches 4x per day at 00, 06, 12 and 18 UTC? Why not simply say that" And say that the data from soundings performed at each of the four launch times were analyzed separately? If launches were performed at 18 UTC, why are those data not shown or discussed? I assume it is because of the large difference reduction in near-surface temperatures between 18 and 20 UTC, but this needs to be stated.

There are three launches a day, this is now stated explicitly.

Figure 21: Which launch time is this for?

All launch times, now stated explicitly.

Line 399: If not from a quadratic fit of the data in Figure 20, how was the "best estimate" determined? What is the uncertainty of A?

The best estimate was determined as the minimum of the mean square difference of T between ascent as descent - so similarly as it would be from Figure 20, but for all 528000 original comparable points. Uncertainty of A wasn't determined (we were not sure how to do that for the quadratic coefficient).

Line 401: "the root mean square ΔT is lowered from 1.22 °C to 1.06 °C, indicating that the correction explains 24.4 % of the variance seen". How does the reduction in RMS explain 24.4% of the variance? Isn't that determined from a correlation coefficient?

Variance between ascent and descent temperatures is given by various parameters - three main seems to be descent rate implying friction, time and space difference of measurement and uncertainty of measurement itself. Measure of the variance is (mean) square difference. As the RMS of ΔT is lowered from 1.22 °C to 1.06 °C, MS of ΔT is lowered from 1,49 to 1,12, which is 24.4% decrease. The rest 1,12 belongs to the other parameters than friction.

Table 3: Each coefficient needs an uncertainty estimate or their statistical significance remains unknown. "Best estimate" still needs to be explained since it appears to not be based on a standard parametric fit.

See explanation above (reply to comment on line 399). I think that best estimate is quite standard.

Table 4: What are "compared levels: 527 779" ?

Changed to 'sample size'.

Line 431: "the exact value of the correction coefficient is slightly uncertain." Are you suggesting that the correction should be the same for descents with and without a parachute? I don't understand how that is expected, since descent rates without a parachute are typically much higher than those with a parachute. If you divide the data into types: high descent rates typical of no parachute descents and lower descent rates (parachute) and independently fit each type, can you improve the quadratic fits and lower the uncertainties in "A" for each case?

There is also plenty of overlap in the descent rates with and without a parachute. We would prefer to use the same correction for both, unless there is strong evidence of the need for different corrections. Figure 19 (moved from section 3) supports the use of a single correction.

Line 436: "method is used for processing of the data from RS41-SGP radiosondes" - what method are you referring to and how is it used to process the data? The equation is only used to calculate geopotential height.

Reworded and simplified.

Line 438: In order to calculate the pressure profile the geometric altitudes from GPS must be converted to geopotential heights. It is very important to mention this.

Yes, changed to "The RS41-SG radiosonde measures geometric height using GPS, this is converted to geopotential height, and the pressure is calculated with the hydrostatic equation."

Line 439: A radiosonde doesn't "estimate" anything, it measures descent temperatures with a positive bias.

Reworded: "As discussed above, descent at high speeds, mostly in the stratosphere, causes the measured temperature to be too high." I have left the next sentence unchanged "This overestimation of temperature leads to underestimation of air density." as it seems to be the simplest/shortest way of describing the effect on the processing.

Line 443: "the shift of height still remains in the troposphere levels" - why would you expect it to go away? The effect is cumulative during the entire descent so the geopotential height biases continue to increase down to the surface.

Changed to "the shift of height affects the troposphere levels". This insight (from MM) is obvious once pointed out, but hadn't occurred to me earlier.

Figure 23: It would be instructive to say that the vertical offsets between ascent and descent temperatures shown are for a RS41-SG and that the offsets are due to what is described in Lines 444-445.

Now Figure 20. Text shortened and clarified.

Lines 452-453: The descriptions of curve colors in Figure 24 belong in the caption, not here in the text. The caption for Figure 24 needs some rewording, including a statement of the Figure being based on a single flight or many flights.

Colours removed from text and figure (now 21) replotted. Caption changed to clarify that it is based on many flights.

Line 454: "and the lines are almost the same". Wouldn't it be clearer to say that the biases and standard deviations in differences between ascent and descent temperature measurements are the same regardless of whether the profiles are aligned using height or pressure?

The suggested wording is much longer, we changed to: "In the stratosphere the choice of coordinate has little impact on the ΔT statistics (because ΔT comes mainly from direct heating)."

Line 474: Would the "positive effect" be the reduction in biases between ascent and corrected descent measurements?

This introductory sentence was deleted and the paragraph reworded in several places.

Figure 26: The caption for Figure 26 needs some rewording.

Reworded.

Figure 27: The "% differences" units are misleading because they are not % differences. My initial thoughts were that the 100% line (control) was perfect agreement and deviations from this line represented positive and negative biases. Then the caption indicates that values <100% indicate improved forecasts. A clearer explanation is needed here to make the results shown in this figure more understandable.

They are % differences. Text changed to 'modest improvements in the root-mean-square (rms) fit of the 12 h forecast to radiosonde ascent data (Figure 24; 100*rms_test/rms_control is shown).' Rms is a natural measure for vector winds.

Line 513: You might check with the US National Weather Service Field Support Laboratory in Sterling, VA, about publications arising from wind tunnel tests of radiosonde temperature sensors. They have decades of experience performing such tests in a wind tunnel environment.

We have checked with someone at Sterling, and also someone at INRIM in Italy. Neither was aware of any such tests, but there was some interest in performing them.

Line 519: "have a closer RMS fit" - this is far from standard terminology. What is an "RMS fit"?

Changed to "closer overall fit". (RMS - root-mean-square, quite often used in NWP, but simpler to avoid it here.)

Line 528: "Jimsphere balloons" need to be described here

Added 'One technique to obtain accurate wind profile data is the Jimsphere - a balloon with roughness elements used without a separate instrument package.' See https://spinoff.nasa.gov/spinoff1996/48.html. They were developed and used by NASA, but it is unclear to me how much they are still used.

Line 529: If you have documented evidence that supports this statement please provide it here (along with the reference) otherwise this appears to be opinion rather than fact.

Changed to "Dropsonde wind profiles suffer much less from pendulum motion than radiosonde ascent winds (Wang et al, 2008) and probably less than radiosonde descent winds." This is discussed in a bit more detail in the introduction.

Line 546: "this will reduce with improved processing/bias correction" - it is not clear what "this" is.

Changed to "the benefits of parachutes or pressure sensors will reduce with improved processing/bias correction".

Lines 552-554: The text inside the parenthesis needs to be re-worked, as it is not clear what is meant.

Changed to “(it would be better to exclude data based on the actual descent rate, but this would require more work).” Arguably data should be rejected for a while after the parachute has opened abruptly (see Fig 9, was Fig 10) but I think we’ll leave this detail out.

Line 557: Remove the space between "n" and "ow"

No space in “now” - a quirk of the pdf?