10/24/16 6:00 AM
Comments on *Characterization of the Long-term Radiosondes Temperature Biases….* by
Ho et al. (ACP-2016-801)
Rick Anthes

Line numbers refer to ACP-2016-801.pdf

Overall comments

The authors have done a lot of careful work and there are interesting results in their paper
on the different accuracies, stabilities and trends of various types of radiosonde data. The
authors use radio occultation (RO) data in the upper troposphere and lower stratosphere
as a reference data set for the period June 2006 to April 2014. Since radiosonde data are
often used in climate studies, it is important to document the accuracies and uncertainties
of different types of radiosondes in different countries.

1. However, the paper is too long (58 pages) and the sections on results (Sections 3-5) are
tedious and difficult to read because of too much detail in the text that merely duplicates
what is in the tables and figures as well as too many symbols in the text (e.g. ΔT
($RS92_{200701\text{-}201012}$)). Furthermore, much of the detail is describing statistics that are small
and probably not statistically significant or of general interest. The reader is
overwhelmed with the reporting of many numbers without a focus on what is really
important and what is of little or no interest. The paper would be greatly improved and
have more impact if it were shortened significantly and only the major results included in
the text.

$\Rightarrow$ As suggested by the reviewer, we shorten this paper significantly. We rewrote Section
3-5, combining section 4.2, section 4.3, and section 5 into a new section (new section
4.2). Many symbols (e.g. ΔT ($RS92_{200701\text{-}201012}$)) are removed in the text. In addition,
we removed results for 150 hPa (i.e., Tables 4 and 6) because the results were similar
to those at 50 hPa. Appendix A is also removed, yet a part of the Appendix A is
inserted in the Introduction section. To demonstrate whether the results are
statistically significant or not, we performed statistical significance tests for RAOB –
RO trend difference. We only mention that the statistical significance results in the
revised paper. The revised manuscript is now reduced from 58 pages to 47 pages. All
changes are tracked and the tracked manuscript is also submitted. In addition, since
the heights from 150 hPa to 50 hPa are in part of upper troposphere, we added "
"Upper Troposphere and" to the title.

2. There are many statistics of radiosonde minus RO temperatures from various types of
radiosondes at different levels of the atmosphere between 200 and 20 hPa over six
different regions of the world. It is not clear which of these results are statistically
significant and which ones we should be concerned about. This makes interpretation of
the results difficult as we could be looking at small, quasi-random differences that have
no physical meaning, nor even meaning relative to the specific types of radiosonde data.
Differences are often 0.1K or less, which are well below the accuracy of radiosonde

sensors. When the different atmospheric sampling volumes of the radiosondes and RO are considered, sampling errors alone can be much larger than 0.1K. It would be helpful if the authors could do statistical significance tests and describe in the text only the results that are significant at the 95% or higher level.

⇨ To demonstrate if the computed de-seasonalized trend differences (RAOB – RO) are statistically significant or not, we performed statistical significance tests for the trend difference. In Figs. 12 and 13 we list the 95% confidence intervals for trend difference (ROAB – RO) in the parentheses.

⇨ The 95% confidence intervals for trend differences for global Vaisala (RS80, RS90, and RS92) and other sensor types during the daytime and night time in the North hemisphere mid-latitude (60°N-20°N) are summarized in the Table 4. We also discuss what the trends in radiosonde minus RO temperatures and RO temperatures means is the text (see the reply for comment 5).

3. The authors compute trends of the differences between individual types of radiosonde and RO over a 7-year period. Most of the trends are small (of order 0.2 K per five years) and quite different, with some being positive and some being negative. It is not clear what these trends mean, except as an indication of the uncertainty of the radiosonde minus RO temperatures over this short period. Again an estimate of the statistical significance of these trends would be useful. What would the magnitude of trends computed from a similar time series of random data with the same standard deviation as these differences be? A comparison of what sort of trends in temperature at these levels due to long-term climate change would be useful as well. For example, from climate models we might expect a temperature trend in the lower stratosphere to be something like 5 K per 100 years or 0.25 K per five years. Trends reported in this paper for the Vaisala RS92 radiosonde at 50 hPa (Table 3) range from -0.211 K/5 years (U.S., night) to 0.264 K/5 years (England, day), so they are comparable or slightly smaller than what one would expect for a long-term climate trend signal.

⇨ To compare trends in temperature at these levels due to long-term climate change with RO trends in this paper, we refer to stratospheric temperature trends over 1979–2015 computed by Randel et al., (2016) in line 466. Randel et al., (2016) is also added in the references. Randel et al., (2016) indicated that the linear trends over 1979–2015 show that cooling in the lower stratosphere is about -0.1 K to -0.2 K/decade. In line 464 we added, "A long-term (de-seasonalized) trend in temperature at this level associated with global warming (stratospheric cooling) might be approximately -0.1 to -0.2 K/decade or -0.05 to -0.1 K/5 years (Randel et al., 2016). Trends reported in this paper for the Vaisala RS92 radiosonde at 50 hPa (Table 3) range from -0.211 K/5 years (U.S., night) to 0.264 K/5 years (United Kingdom, day), which are comparable to those reported by Randel et al., (2016)."

4. It would also be interesting to compare these trends in radiosonde-RO temperature differences to the corresponding trends in the RO temperatures over this period. Indeed, Tables 3 and 4 give the RO trends, but they are never mentioned in the text!

⇨ As suggested by the reviewer, we rewrote sections 4 and 5 and compared de-seasonalized trends of radiosonde-RO temperature differences to the corresponding trends in RO temperatures over this period.

⇨ In line 455 we added "The de-seasonalized trends in RO temperatures are generally larger than those for the radiosonde-RO differences. A maximum de-seasonalized trend of 1.143 K/5 yrs is found for nighttime temperatures over the United Kingdom. A minimum de-seasonalized trend of -0.69 K/5 yrs is found for daytime temperatures over Canada. Trends with magnitude greater than 0.5 K/5 yrs are found over the United States, Germany, Canada and the United Kingdom. The fact that these de-seasonalized trends in RO are significantly greater than the de-seasonalized trends in the differences suggests that they represent a physical signal in these regions. However, the time series is too short to represent a long-term climate signal; instead these likely represent real but short-term trends associated with natural variability."

⇨ To shorten the paper, we also removed results for RAOB and RO temperature comparisons at 150 hPa (old Tables 4 and 6). Now the new Tables 3 and 4 are specifically mentioned in the new text.

5. The RO temperature trends at 50 hPa (Table 3) range from -0.69 (Canada, day) to 1.143 (England, night). Quite different values are found at 150 hPa (Table 4), with the 5-year trends ranging from -0.797 (Canada, day) to 1.508 (U.S. day). In general, the magnitudes of the trends of radiosonde-RO temperature differences are smaller than the trends in RO temperatures, which is an indication of the consistency between the radiosonde and RO temperatures. The large differences in RO temperature trends between regions (much larger than expected for a long-term climate change signal) probably indicates natural variability in the six different regions. The fact that they are larger than the trends in the differences indicates to me that they are a real signal in the different regions over this 7-year time period. Presumably, since the radiosonde-RO trends are smaller, the radiosondes (at least the good ones) would pick up similar trends to the RO trends. A discussion of what the trends in radiosonde minus RO temperatures and RO temperatures means is needed.

⇨ We specifically added several discussions of what the trends in radiosonde minus RO temperatures and RO temperatures means in the new Section 4.2.

As mentioned in the reply for comment 4, we added a discussion of what the trends in radiosonde minus RO temperatures and RO temperatures means in Line 455. In line 460, we specifically discuss what the trends in radiosonde minus RO temperatures and RO temperatures means: "The fact that these de-seasonalized trends in RO are significantly greater than the de-seasonalized trends in the

differences suggests that they represent a physical signal in these regions. However, the time series is too short to represent a long-term climate signal; instead these likely represent real but short-term trends associated with natural variability. "

⇨ In lines 470-480, we discuss what trends in radiosonde minus RO temperatures and RO temperatures means by stating "We compare the global trend of radiosonde – RO temperature differences for the Vaisala and other radiosondes at 50 hPa in Table 4. The Vaisala RS92 biases are 0.22 K (day) and 0.12 K (night). The global de-seasonalized temperature differences for Vaisala RS92 for daytime and nighttime are equal to 0.074 K/5yrs and -0.094 K/5yrs, respectively. The 95% confidence intervals for slopes are shown in the parentheses in Table 4. This indicates that although there might be a small residual radiation error for RS92, the trend in RS92 and RO temperature differences from June 2006 to April 2014 is within +/-0.09 K/5yrs globally. These values are just above the 1-sigma calibration uncertainty estimated by Dirksen et al. (2014). This means that probably the stability of the calibration alone could explain most of this very small trend. It is also consistent with the change in radiation correction."

⇨ We discuss the mean bias in the last two paragraphs of Section 4.2. In line 481 we state "Figure 13 depicts the de-seasonalized temperature differences for Sippican MARK IIA, VIZ-B2, AVK-MRZ, and Shanghai in North hemisphere mid-latitude (60°N-20°N) at 50 hPa and the results are summarized in Table 4. The 95% confidence intervals for slopes are shown in the parentheses in Table 4. The de-seasonalized trend of the daytime differences varies from -0.137 K/5 years (Russia) to 0.468 K/5 years (VIZ-B2). The magnitudes of the daytime trends are less than 0.2 K/5 yrs for all sensor types except VIZ-B2 and Sippican, which both exceed 0.4 K/5 yrs. These are much larger than those of the Vaisala RS92 (0.074 K/5 yrs)."

⇨ In line 489, we state, "The corresponding nighttime de-seasonalized trends in the biases vary from -0.348 K/5 yrs (VIZ-B2) to 0.244 K/5 yrs (Sippican). Again, these are much larger than those of Vaisala RS92 (-0.094 K/5 yrs). Thus the VIZ-B2 sensor stands out as having larger biases and trends than do the other sensors."

In summary, the paper contains some interesting and important results and should be published, but it requires significant rewriting, editing, and shortening with greater emphasis on what the important results are and less detail on all the individual numbers.

<u>Detailed comments</u>

1. The papers use three terms to describe the radiosonde-RO temperature differences: differences, biases, and anomalies. I suggest using only differences and biases, and eliminate all references to anomalies.

⇨ We replace all "anomalies" with "differences" in this paper.

2. Do you mean United Kingdom rather than England?

⇨ Yes, it shall be "United Kingdom". We replace all "England" with "United Kingdom" in this paper and Figures. The revised Figs. 3, 4, 5, 9, 10, and 12 are inserted.

3. An example of how a difficult to read paragraph containing a repetition of data in a table can be simplified, shortened, and made more readable is lines 264-267:

"In general, the radiosonde temperature biases vary for different sensor types. The mean ΔT for RS92 (0.16 K), RS80 (0.10 K), RS90 (0.13 K), Sippican MarkIIA (-0.08 K), Shangai (0.05 K) and Meisei (0.11 K) are smaller than those for AVK (0.33 K) and VIZ-B2 (0.22 K) (see Table 2)" (50 words) may be replaced with

"The radiosonde temperature biases vary for different sensor types. All biases are less than 0.25 K, except for AVK and VIZ-B2, which reach 0.66 and 0.71 K respectively during the day." (31 words).

⇨ The sentence in lines 264-267 was revised as suggested by the reviewer. Similar sentences in old Sections 3-5 were also revised and are not specifically mentioned. The tracked manuscript is submitted.

4. An example of unnecessary use of symbols in a sentence which makes reading difficult is: "The mean temperature biases in this region for $\Delta_{Sippican}^{Time}$, Δ_{VIZ-B2}^{Time}, Δ_{AVK}^{Time}, and $\Delta_{Shanghai}^{Time}$ for 50 hPa are summarized in Table 5." This can be written in much more readable form as "The mean temperature biases in this region for the Sippican, VIZ-B2, AVK, and Shanghai radiosondes at 50 hPa are summarized in Table 5."

⇨ We removed many symbols (for example, $\Delta_{Sippican}^{Time}$) in this paper and many sentences are revised as suggested. For example, in line 426, we now state "All daytime biases are below 0.25 K in magnitude, except for Russia (0.8 K) and VIS-B2 (0.87 K). The magnitudes of the mean nighttime biases are all less 0.25 K except for VIS-B2, which is -0.56 K. The daytime biases for Russia and VIS-B2 contain obvious inter-seasonal variation." No symbols are used in the sentence.

5. Lines 439-442. I don't understand this sentence. If the U.S. did not use RS92 radiosondes before 2012, there would be no data for comparison with RO before 2012 (i.e. none in the period 2007-2010). However, this section talks about RS92 from Jan 2007 to Dec 2010 for the U.S. and Fig. 10 shows RS92 vs. RO going back to 2007. Also, a small number of pairs in the comparison does not necessarily imply small differences—in fact, a small number of pairs could lead to large differences due to an inadequate sample size.

⇨ The US National Weather Service (NWS) did not use Vaisala RS92 radiosondes before 2012. The Vaisala RS92 radiosondes before 2012 were mainly launched by research groups (for example, at the ARM site and during individual field experiments from universities, etc).

⇨ To avoid confusion by the readers, we deleted the statement "since the US National Weather Service (NWS) did not use Vaisala RS92 radiosondes before 2012."

6. It seems strange that Table 3 is not mentioned until line 563, long after Tables 4, 5 and 6 are mentioned and discussed.

⇨ In the revised paper, we refer to Table 3 before Table 4. The Table 3 is now referred in line 415 whereas Table 4 is first referred in line 426.