Reviewer comments on the paper "Validation of merged MSU4 and AMSU9 temperature climate records with a new 2002–2012 vertically resolved temperature record" by A. A. Penckwitt et al.

This paper is dedicated to validation of merged MSU4 and AMSU9 temperature climate records with a new 2002–2012 vertically resolved temperature record, which was created using ESA and ESA-TPM temperature profile data. The objective of paper is important for the assessment of atmospheric temperature trends.

The paper is generally well written. However, I have concerns about the methods presented in the paper. Please find below main and detailed comments.

MAIN COMMENTS

1. Motivation and novelty

Several studies dedicated to assessment of the combined (A)MSU climate temperature record with radio-occultation (RO) data have been performed in the past (Ho et al., 2009; Ladstädter et al., 2011; Steiner et al., 2011). Since RO data dominate in the new created dataset, it should be clarified why new dataset suits better for the study (see also comments below).

2. The choice of datasets

Datasets for merging

In general, to get the best climate data record, it is not necessarily to use as many data as possible. Instead, the best solution would be the measurements of the same type, which are coherent with each other and do not require calibration. For the purpose of the paper, this would be more logical to use the radio-occultation measurements, which can be used together without any correction. However, such analysis has been already performed (Ladstädter et al., 2011).

The dataset created by the authors is also "RO-dominated", but it has additional difficulties when merging measurements from other-type instruments (and thus differences in measured parameters (note that RO give "dry temperature") and uncertainties due to bias and drift correction, see also below).

I think that the collection of datasets for merging not optimal. The motivation for using so diverse data (some of them are drifting due to instrument aging) should be explained.

Datasets for validation

- Why do you prefer using the RATPAC-A dataset with relatively low sampling (85 stations) and insufficient information about the data averaging instead of larger radiosonde databases such as RAOBCORE (~1000 stations) or IGRA (~1500 stations)?
- The differences with respect to the NCEPCFSR dataset are large. Should the readers make the conclusion that NCEPCFSR is wrong? If yes, why this dataset is used for validation of VRT?

3. Merging method and merged dataset

1) For nearly all instruments used in this paper, the "native" vertical coordinate is altitude (the only exception is SMR). The effect of altitude-pressure conversion using the meteorological models should be mentioned and evaluated (long-term temperature records from meteorological models do not suit for trend analyses, as they have artificial jumps due to different assimilated data).

2) What is the reason for selection of relatively narrow latitude zones of 5°? This results in increased sampling error.

3) Applying the correction of sampling bias to radio-occultation data and not applying it to ACE-FTS and SMR looks very strange (erroneous). I think, the sampling uncertainty should be either corrected for all instruments or taken into account as additional uncertainty (also for all instruments).

4) The dataset used for merging have different vertical resolution. This (substantial) difference in vertical resolution will not affect the mean value, but it will affect the estimates of uncertainty for the monthly zonal mean profiles by the standard error of the mean: the sample variance of the high-resolution profiles will be larger due to better resolved gravity-wave fluctuations.

5) From my visual perception, the fit by Eq.(2) does not agree with the differences shown in right panels of Fig.1 (especially for GRACE, TSX, ACE-FTS). This means that the bias-drift correction by the function (2) might introduce additional uncertainties. Look at the panel for SMR in Fig.1: uncorrected data agree better with the merged time series than the corrected ones in years 2002-2007!

Have you optimized the regression model? Are all parameters statistically significant?

6) Eq.(2) assumes that the drift is linear in time, which can be not realistic in reality. In general, the comparison with COSMIC presented in section 4.2 clearly indicates problems with the merged dataset.

7) For the results, it would be interesting to see also the linear fit coefficients for the difference MSU4+AMSU9 and iVRT, as well as their comparison with analogous presented in (Ladstädter et al., 2011). Such information and comparison would be informative for both temperature trend estimates and demonstration of VRT capabilities.

DETAILED COMMENTS

1) P.242, l. 1-2: “Additionally, the RO data were screened such that temperatures below 150 K were omitted as were temperatures above 330 K.”

How often this occurs; what is the percentage of screened data? Do you remove only the value at some layer of the whole profile containing temperatures below 150K?

2) P.243, l.5-10. What is the reason of GRACE bias with respect to CHAMP and TSX? This contradicts with the study by Foelsche et al (2011), which report high consistency of the temperature climate records from multiple radio occultation satellites.

3) P.244-245: “The monthly mean data from an instrument are excluded from this merging process if there are fewer than 4 measurements in a particular month or less than 5% of measurements of the month with the highest number of measurements within that year” For good fitting by Eq.(2), all seasons should be covered. Is this satisfied in all latitude zones for ACE-FTS?

4) P. 251, l.8: Why the number of harmonics in the regression model is increased compared to Eq.(2)?

5) P. 254, l. 24-26: “It has been shown that the uncertainty on the monthly mean zonal mean temperatures decreases with an increased number of instruments used in the merging.” This is true only if the bias and drift correction is perfect. Otherwise, this statement is generally not correct (and Eq.(7) is not valid: RHS should be changed in order to account for differences in the mean profiles). Your statement on page 244 is more accurate.

References

Foelsche, U., Scherllin-Pirscher, B., Ladstädter, F., Steiner, A. K. and Kirchengast, G.: Refractivity and temperature climate records from multiple radio occultation satellites consistent within 0.05%, Atmos. Meas. Tech., 4(9), 2007–2018, doi:10.5194/amt-4-2007-2011, 2011.

Ho, S.-P., Goldberg, M., Kuo, Y.-H., Zou, C.-Z. and Schreiner, W.: Calibration of Temperature in the Lower Stratosphere from Microwave Measurements Using COSMIC Radio Occultation Data: Preliminary Results, Terr. Atmos. Ocean. Sci., 20(1), 87–100, doi:10.3319/TAO.2007.12.06.01(F3C), 2009.

Ladstädter, F., Steiner, A. K., Foelsche, U., Haimberger, L., Tavolato, C. and Kirchengast, G.: An assessment of differences in lower stratospheric temperature records from (A)MSU, radiosondes, and GPS radio occultation, Atmos. Meas. Tech., 4(9), 1965–1977, doi:10.5194/amt-4-1965-2011, 2011.

Steiner, A. K., Lackner, B. C., Ladstädter, F., Scherllin-Pirscher, B., Foelsche, U. and Kirchengast, G.: GPS radio occultation for climate monitoring and change detection, Radio Sci., 46(6), n/a–n/a, doi:10.1029/2010RS004614, 2011.