

## *Interactive comment on* "Consistency and structural uncertainty of multi-mission GPS radio occultation records" *by* A K. Steiner et al.

## Anonymous Referee #2

Received and published: 8 December 2019

I thank the authors for a manuscript that has obviously been polished by thorough proof-reading before submission. This paper represents the tip of the iceberg of what seems to be a large community undertaking. It is quite nice to see such synchronization and coordination to advance the science. I have one major comment on the paper. Other detailed comments are indicated thereafter.

## Major comment

It is unclear what structural uncertainty encompasses, and how a standard deviation computed from 5 different products can say everything about the confidence one should place in these products for climate monitoring. Furthermore, it is unclear how using this metric is applicable, alone and by itself, to assert compliance to the GCOS requirements for stability. For example, if all data producers used exactly the same pro-

C1

cessing techniques, one would expect to see a collapse in the product spread between the various products; would this give us absolute confidence in the real stability of the instruments? (one is allowed to doubt)

## Detailed comments

Did all current RO data processing centers take part in this exercise, especially those centers processing long time-series and third-party missions? If not, it may be useful to indicate if future work will strive to include those other centers.

"Structural uncertainty" is neither defined nor referenced in this paper. It is a central concept to this paper, and one not defined in textbooks or standards such as the BIPM Guide to the Expression of Uncertainty in Measurement (GUM). It cannot be expected for readers to guess what this particular concept of uncertainty means, or what it represents.

Section 2 is highly informative and packs a lot of information. My feeling is that the presentation of the data (as provided by the various centers) and the presentation of the methodology (for comparing these data) need to be separated. It would be fitting to split the section accordingly (e.g., Section 2: Data, Section 3: Analysis methodology, or equivalent). This would avoid a potential confusion between radio occultation data processing (done outside of this study), and the analysis of the results (as conducted and presented in the paper).

The section on data could benefit from being reorganized as follows: - Starting from the raw data, - Proceeding to the higher-level products /retrievals (without going back at the end to discuss clock errors etc.), - Presenting, at each step, the commonalities for all centers, before indicating the differences (e.g., center X did not produce ...).

"as three centers start with the same phase and orbit data, RO products are not independent": I do not understand why this statement only applies to these 3 centers, and not all of them. (The various RO processing centers, for each given mission, do start from the same receiver data?)

The subscript s refers to the satellite receiver (not transmitter). It may be useful to indicate 'receiver'.

Were equations (1) to (6) applied to the subset of common profiles processed by all centers?

"only JPL provided a smaller amount": Looking at figure 1, one sees that JPL did provide a smaller amount indeed, of about 10,000 profiles per month, compared to the pack of other producers. However, quite interestingly, the common subset of profiles, between all producers, is also 10,000 lower than the JPL count. This near-match in the differences (from other producers to JPL, and then from JPL to common subset) is quite puzzling. Could it come from an unexpected issue in individual ID assignment (e.g., a shift by a minute or so), which would make many JPL profiles not match the other IDs?

"we are interested in the structural uncertainty of trends represented by the standard deviation of the n\_center individual center trends": Unless I misunderstood something, given that n\_centers is (at most) 5, this means that the central metric of the paper is a standard deviation based on a population of 5 members. How reliable is a standard deviation based on so few members? This needs to be discussed.

In the future, wouldn't there be a more robust estimate that can computed, to characterize this spread, or inconsistencies, given a such small sample?

The fact that the spread in physical temperatures is reduced so much, from the spread in dry temperatures, needs to be explained. Does it point to the fact that the products use similar (background) constraints in the retrieval, and then correspondingly that all products are probably quite representative of these constraints?

The difference found at high latitudes is one very interesting result of this paper. This is mentioned as being related to Arctic SSW (something which correlates well by looking

СЗ

at the 60N-90N timeseries and the occurrences of peaks in winter). I would think this deserves a separate sub-section in discussion, with additional results to go a bit further. Is it possible to illustrate the influence of the different strategies for high-altitude initialization in these situations, e.g., by picking a particular SSW event, and showing individual profiles?

Throughout the paper, all statements making the link to the GCOS stability requirement need to be revised, as they all fail to include the other sources of uncertainty affecting stability (other than differences in processing).

Typo: 'on exemplary' -> one exemplary

Table 1, impact height is only defined for UCAR. Do the other centers use different definitions? Shouldn't this have been the same definition for all? In other terms, isn't there a RO community-approved definition of 'impact height'?

Table 1 indicates several vertical reference frames, not always WGS-84 ellipsoid and EGM-96 geoid. As a reminder, WMO Executive Council 59 (in 2007) adopted a draft resolution proposed by the Commission for Basic Systems of these two elements (WGS-84 and EGM-96) as the fundamental bases for vertically referencing all station observations. This choice was also relayed in the coordinated satellite community for geostationary products, by CGMS in 2011 ("LRIT/HRIT Global Specification"). It is not so much a matter of choosing the 'best' reference frame for each observing system, but one that is fit for purpose and a unique standard in a community, so as to avoid introducing artificial discrepancies/differences. It could hence be useful to make a note that some of the differences between data producers presented in Table 1, such as this one, will eventually need resolving.

Figure 1, it is unclear why the number of points differ by altitude, even though the list of profiles is supposedly common for all centers. Is this caused by different QC at each vertical level?

Figures 4 to 7, the equations (shown in the legend inside each plot) are too small to be legible; they could go into a new table, or, better yet, be summarized in a graphic, in a similar form as Fig. 12.

Figure 8 to 11 pack, in total, over 700 vertical profiles. Surely, there must be a way to summarize this into a manageable amount of information, for readers to grasp the message. These plots are surely of value in a supplement, though.

Interactive comment on Atmos. Meas. Tech. Discuss., doi:10.5194/amt-2019-358, 2019.

C5