Atmospheric
Measurement
Techniques
Discussions

Open Access

EGU

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

A. K. Steiner et al.

andi.steiner@uni-graz.at

Received and published: 24 February 2020

We thank the reviewer for the constructive comments and questions. We will revise the manuscript accordingly. Please find our responses to all comments below.

**Major comment

It is unclear what structural uncertainty encompasses, and how a standard deviation computed from 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 processing techniques, one would expect to see a collapse in the product spread be-

tween the various products; would this give us absolute confidence in the real stability of the instruments? (one is allowed to doubt)

*Response

GPS RO is regarded a stable technique providing benchmark observations (Goody et al. 1998; GCOS 2011). Traceability to standards of the international system of units (SI) is the foundation for benchmark measurements (e.g., Leroy et al., 2006; Ohring, 2007), i.e., they are related to absolute standards and may be repeated at any subsequent time with precise comparability (Goody et al., 1998; 2002). As GPS RO measurements are based on highly precise and stable atomic clocks, long-term frequency stability is ensured. The fundamental measurement is the GNSS signal phase change as function of time, measured with high precision, solely in terms of timing, and is an absolute measurement and SI-traceable (e.g., Ohring, 2007). RO bending angle is given as one example of a fundamental climate data record by GCOS (2016). In this respect, we investigate the consistency of RO time series from different data centers. We assess if final data products from different processing systems (based on the same raw measurements) are consistent and if they give consistent trend information. Therefore, we estimate the structural uncertainty, which arises due to different choices in processing and methodological approaches for constructing a data set from the same raw data (Thorne, 2005). As a measure for structural uncertainty, we analyze the spread in difference time series and the spread in trends (Wigley, 2006). Analyzing difference time series removes the variability that is common to the data sets and isolates those differences that may be due to differences in data set production methods, i.e., structural uncertainty of the data records. Correspondingly, we chose the standard deviation of trends as a measure of the structural uncertainty in trends. The challenge is to quantify the true spread given a limited number of data sets. Thus, it is important to have several independent data records, and at least a minimum of three independently processed data sets (Thorne, 2005).

As these points are also addressed in the detailed comments below, please find our

further detailed responses there, at Comment 2 on structural uncertainty, Comment 8 on standard deviation, and Comment 11 on stability.

**Detailed comments

*Comment 1

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.

*Response 1

Yes, all current RO data processing centers that are processing several or all available RO missions and that provide RO data for long-term records (from CHAMP to current RO missions) took part in this study. As explained in the introduction section of the manuscript, this is an RO community effort ongoing since 2006 (http://irowg.org/projects/rotrends/). All international RO processing centers collaborate on the intercomparison of RO data with the common objective of a continuous improvement of the maturity of RO data products for use in atmosphere and climate studies. We specify this in the introduction section of the revised manuscript: "We systematically intercompare RO data products provided by those five international RO processing centers that are processing several or all available RO missions and that provide RO data for long-term records (from CHAMP to current RO missions)."

*Comment 2

"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.

*Response 2

As in Steiner et al. (2013), we use the concept of structural uncertainty discussed by Thorne (2005) and Wigley (2006) with respect to uncertainties in climate trends. Structural uncertainty of a specific observational record arises due to different choices in processing and methodological approaches for constructing a data set from the same raw data (Thorne, 2005). The challenge is thus to quantify the true spread of physically possible solutions, given a limited number of data sets. The more independent data records the better, but at least three independently derived datasets are regarded necessary for an estimate of the structural uncertainty (Thorne, 2005). We included a paragraph in the introduction section of the revised manuscript on explaining structural uncertainty: "Structural uncertainty of an observational record arises due to different choices in processing and methodological approaches for constructing a data set from the same raw data (Thorne, 2005). The challenge is thus to quantify the true spread of physically possible solutions from a limited number of data sets. At least three independently derived datasets are regarded necessary for an estimate of the structural uncertainty, but the more data sets the better. Thus, multiple independent efforts should be undertaken to create climate records."

*Comment 3

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 . . .).

*Response 3

Thanks to the reviewer for these suggestions, which is actually close to the already implemented structure. Section 2 gives a description of RO data and of the data processing at the centers (done outside this study). Section 3 gives a description of the study set up and method used in this study. In Section 4, we present and discuss the results. In the revised manuscript, we further improved the structure. We structured section 2 into three subsections as follows. 2.1 Radio occultation missions and data In this subsection, we describe the occultation missions and data used in this study. 2.2 General RO data processing description In this subsection, we describe the basic retrieval steps, which are common for each center, starting from phase measurements to atmospheric bending angle and to atmospheric variables for dry and moist atmospheric conditions. 2.3 Center-specific RO processing steps and comparison In this subsection, we explicitly discuss the center-specific processing steps and the main differences between the centers. A complete and concise overview is given in Table 1, which shows commonalities and differences in the processing steps of each center. For section 3, we adjusted the title to make it clearer. It now reads: 3 Study setup and analysis method

*Comment 4

"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?)

*Response 4

All centers use the same receiver data sets (i.e. data from missions CHAMP, F3C, GRACE, METOP). The structural uncertainty involved in the raw data (e.g, firmware coding) is hard to address. In this study, we focus on analyzing the structural uncertainty due to different center processings. The centers start their individual data

processing chain at different processing levels. Some centers have a full processing chain, which means that they start at the raw data level, i.e., processing phase data and orbit data, and subsequently bending angle to temperature data (centers GFZ, JPL, UCAR). Some centers start with their processing only at the level of phase data. They use phase data and orbit data processed by another center as input for their processing of atmospheric variables from bending angle to temperature. The centers DMI and WEGC use phase and orbit data from UCAR/CDAAC in this study. We revised the respective paragraph in order to make this clearer. It now reads: "Three of the RO centers (GFZ, JPL, UCAR) have the full processing chain implemented, going from raw data level to atmospheric variables. Two centers (DMI and WEGC) start their processing at phase data level in this study, by using phase data and orbit data from UCAR/CDAAC (COSMIC Data Analysis and Archive Center). Thus, as some centers start with the same phase and orbit data (from UCAR), the products from raw data to atmospheric parameters are not strictly independent for these centers."

*Comment 5

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

*Response 5

The subscript refers to the satellite mission (not to a single receiver). In the revised manuscript, we denote it as "satellite mission (s)".

*Comment 6

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

*Response 6

Yes, we added a sentence on this in section 3 of the revised manuscript text at the beginning of paragraph 3: "The common subset of data was analyzed further."

AMTD

*Comment 7

"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?

*Response 7

The number of profiles plotted in Fig.1 is the number delivered by each center for this study. As noted in the manuscript, quality control is handled differently at each center, which means that not always the same set of profiles is delivered by each center. The profiles were collocated based on a unique event identifier (ID) including information on receiver ID, GPS satellite ID, date, and time of the observation. The JPL profiles are matched correctly to the other IDs. The fact that JPL produces less COSMIC data is likely due to a combination of retrieval processing loss and some data not fully processed.

*Comment 8

"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?

*Response 8

We investigate the consistency of RO time series. We assess if final data products from different processing systems (based on the same raw measurements) are consistent and if they give consistent trend information. Therefore, we estimate the structural uncertainty of RO time series and trends. As a measure for structural uncertainty, we analyze the spread in difference time series and the spread in trends (see Wigley, 2006). Analyzing difference time series removes the variability that is common to the data sets and isolates those differences that may be due to differences in data set production methods, i.e., structural uncertainty of the data records (Fig. 4 to 7). Correspondingly, we chose the spread of trends (the standard deviation) as a measure of the structural uncertainty in trends (Fig. 8 to 11). As discussed in our response to comment 2 above, the challenge is to quantify the true spread given a limited number of data sets. Thus, it is important to have several independent data records, but at least a minimum of three independently derived datasets are regarded necessary for an estimate of the structural uncertainty (Thorne, 2005). In section 3 of the revised manuscript, we state this now more explicitly: "The spread of the anomaly difference trends and the spread of the center trends were used for estimating the structural uncertainty (Wigley, 2006) of RO records. For each atmospheric variable and satellite mission, we computed the linear trend over the respective time period for the all-center mean, and for each center. The standard deviation of the center trends was finally used as a measure of the spread."

*Comment 9

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?

*Response 9

Actually, the spread in physical temperature is similar to the spread in dry temperature or even a bit larger for physical temperature (see Fig. 12). As explained in section 2 and in section 4.1, the different moist air retrieval implementations include a priori information, which introduces uncertainties in physical temperature. Throughout the manuscript text, we state that the spread in physical temperature is similar to that in dry temperature, except for one sentence at the end of section 4.3. We revised that sentence accordingly. It now reads: "Physical temperature shows similar uncertainty at lower altitudes, however, above about 30 km it can be larger than for dry temperature due to a priori information in moist air retrievals."

*Comment 10

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 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?

*Response 10

Thank you for addressing this point. The main objective of this work is on the long-term stability of RO products, and to inform about strengths and limitations of the products. Analyzing the center-specific potentials for improvement are within the responsibility of the respective centers. RO processing systems are under continuous development and are therefore expected to improve. This work shall support these efforts but further detailed analyses are out of scope here.

*Comment 11

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).

*Response 11

We disagree on this point as GNSS RO measurements are regarded as benchmark observations (Goody et al. 1998; GCOS 2011). Benchmark measurements are defined as measurements that can be related to absolute standards and may be repeated at any subsequent time with precise comparability (Goody et al., 1998; 2002), such that the accuracy of the record archived today can be verified by future generations. Traceability to standards of the international system of units (SI) is the foundation for benchmark measurements (e.g., Leroy et al., 2006; Ohring, 2007). The GNSS RO technique is such a stable measurement technique because it is based on highly precise and stable atomic clocks, which ensure long-term frequency stability (e.g., Steiner et al. 2011). The fundamental measurement is the GNSS signal phase change as function of time, measured with great precision, solely in terms of timing, and is an absolute measurement and SI-traceable (e.g., Ohring, 2007). Therefore, separate data sets from different RO missions can be combined without intercalibration to a give long-term stable fundamental climate data record (FCDR). RO bending angle is given as one example of a FCDR by GCOS (2016). All characteristics of RO are described in the introduction section of the manuscript. We revised paragraph 3 and included a definition of long-term stability and a further reference. The paragraph now reads: "For climate observations, the accuracy requirement is much more stringent than for weather observations (Trenberth et al., 2013). However, the key attribute is long-term stability, defined as the extent to which uncertainty of measurement remains constant with time (GCOS, 2016). The uncertainty of the measurement must be smaller than the signal expected for decadal change (Ohring et al., 2005; Bojinski et al., 2014). Accordingly, ECV product requirements for air temperature include global coverage, a vertical resolution of 1–2 km in the troposphere and the stratosphere, a horizontal resolution of 100 km, a measurement uncertainty of 0.5 K, and a stability of 0.05 K per decade

(GCOS, 2016). For a definition of the metrological quantities we refer to Annex B of GCOS (2016) and to JCGM (2012)."

*Comment 12

Typo: 'on exemplary' -> one exemplary

*Response 12

Corrected.

*Comment 13

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'?

*Response 13

There is a consistent common definition of vertical geolocation for RO parameters (but a definition was only included in the UCAR description in Table 1). We revised Table 1. We changed the description for UCAR to have it consistent with the description for the other centers. The revised sentence reads: "JGM2 geoid undulation is used to calculate MSL altitude." We included the common definition for impact altitude used by all centers, which reads: "Bending angle is given as a function of impact altitude, i.e. impact parameter minus radius of curvature minus the undulation of the geoid. The impact parameter is defined as the perpendicular distance between the local center of curvature and the ray path from the GPS satellite."

*Comment 14

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 between products. Using different reference frames can only create artificial 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.

*Response 14

Thank you for raising this point. We addressed the topic of vertical geolocation and related uncertainties in the work of Scherllin-Pirscher et al. (2017). They investigated the implication of different geoid choices, e.g., between EGM-96 and JGM-3, on geopotential height. Differences between the results from these two models were found smaller than $1-2$ m up to 35 km geopotential height, which is considered very small to negligible. For further details we refer to Scherllin-Pirscher et al. (2017).

*Comment 15

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?

*Response 15

We assume that this comment refers to Fig. 2. This is not caused by different QC, which is assigned per profile. The different number of points can occur if for one center invalid values appear at certain altitude levels.

*Comment 16

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.

*Response 16

We revised Fig. 4 to 7 and enlarged the fonts. We improved and enlarged the legends. We removed the inserts in the subpanels except for the global differences as the values of difference trends are described in the manuscript text anyway.

*Comment 17

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.

*Response 17

Figure 12 actually summarizes the final information from Fig. 8 to 11 in a condensed format, but Fig. 12 does not give the full vertical information. Fig. 8 to 11 summarize relevant information at the full vertical resolution (one of the main characteristics of RO) for 6 parameters, 4 satellite missions, 5 centers, 6 latitude zones, and up to 600 altitude levels. The figures graphically convey (i) differences between processing centers, (ii) that uncertainties in trends becoming larger at higher altitudes, (iii) that uncertainties propagate to lower altitudes for more derived parameters in the processing chain, (iv) that uncertainties are smaller at higher altitudes for the RO missions GRACE and METOP. We think that we convey all relevant information in an effective way for a comprehensive and overall picture on the consistency of RO data sets.

**References:

GCOS: Systematic observation requirements for satellite-based data products for climate, GCOS-154, World Meteorological Organization. [online] Available from: https://library.wmo.int/doc_num.php?explnum_id=3710 (Accessed 3 September 2019), 2011.

GCOS: The global observing system for climate: implementation needs, GCOS-200, World Meteorological Organization. [online] Available from: https://library.wmo.int/doc_num.php?explnum_id=3417 (Accessed 3 September 2019), 2016.

Goody, R., Anderson, J., Karl, T., Balstad Miller, R., North, G., Simpson, J., Stephens, G. and Washington, W.: Why We Should Monitor the Climate, Bull. Amer. Meteor. Soc, 83(6), 873–878, doi:10.1175/1520-0477(2002)083<0873:WWSMTC>2.3.CO;2, 2002.

Goody, R., Anderson, J. and North, G.: Testing Climate Models: An Approach, Bull. Amer. Meteor. Soc., 79(11), 2541–2549, doi:10.1175/1520-0477(1998)079<2541:TCMAA>2.0.CO;2, 1998.

JCGM: International vocabulary of metrologyâĂŤBasic and general concepts and associated terms (VIM 3rd edition), Tech. Rep. JCGM 200:2012, Joint Committee for Guides in Metrology, Office BIPM, Paris. [online] Available from: https://www.bipm.org/utils/common/documents/jcgm/JCGM_200_2012.pdf, 2012.

Leroy, S. S., Anderson, J. G. and Dykema, J. A.: Climate benchmarking using GNSS occultation, in Atmosphere and Climate: Studies by Occultation Methods, edited by U. Foelsche, G. Kirchengast, and A. Steiner, pp. 287–301, Springer-Verlag Berlin Heidelberg., 2006.

Ohring, G. (Ed): Achieving satellite instrument calibration for climate change, ASIC3 Workshop Report, NOAA/NESDIS., 2007.

Scherllin-Pirscher, B., Steiner, A. K., Kirchengast, G., Schwärz, M. and Leroy, S. S.: The power of vertical geolocation of atmospheric profiles from GNSS radio occultation, J. Geophys. Res. Atmos., 2016JD025902, doi:10.1002/2016JD025902, 2017.

Thorne, P. W.: Revisiting radiosonde upper air temperatures from 1958 to 2002, J. Geophys. Res, 110(D18), doi:10.1029/2004JD005753, 2005.

Wigley, T. M. L.: Statistical Issues Regarding Trends, in Temperature Trends in the

Lower Atmosphere: Steps for Understanding and Reconciling Differences, edited by T. R. Karl, S. J. Hassol, C. D. Miller, and W. L. Murray, Washington, D. C., 2006.

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