Author's final comments and changes in the manuscript *amt-2019-417*

Reviewer #1, specific comments

(1)
Reviewer: It would be helpful if the paper would say a little more about the ERA-I data sets. On P2 line 35 to P3 line 1 it says "reanalysis data." Later, e.g. P4 line 2 it says "short-term forecasts." According to Berrisford et al. (2011 ECMWF Report), the ERA-I data sets include four analyses per day and two ten-day forecasts per day, initialized at 00 and 12 UTC. Exactly which of these data sets are being used to interpolate to the RO locations in space and time? Please define "short-term" and say why forecasts were used instead of analyses. Since the analysis fields are given only every six hours and the forecast fields are given every three hours, the higher temporal resolution of the forecast fields likely give better temporal interpolation to the RO times and this may be the reason why forecasts are used. A sentence or two would clarify exactly what is being done and why. A brief description of the temporal and horizontal and vertical interpolation schemes of ERA-I to RO locations would also be useful.

Author response: A problem with using reanalysis data as a reference for comparison with radio occultation (RO) data is that RO data itself have been assimilated by the reanalysis system. If we interpolate in an ERA-I *analysis* field to the times and locations of RO events, the interpolated model data will be strongly influenced by the very same RO data that we wish to compare with. Such comparisons would have an obvious problem with circularity. Using ERA-I *forecasts* instead of *analyses* removes most of this circularity. ERA-I provides forecasts at 3-hour intervals, initialized at 00 and 12 UTC. The shortest possible forecast times thus vary from 3 hours up to 12 hours. This is what we mean by "short-term" forecasts.

The model resolution is 3 hours and data are interpolated from a 1.0x1.0 degree latitude-longitude grid. In the vertical, as a part of the forward-modelling from model space to observation space, the data are first interpolated from the original ECMWF model levels to the vertical coordinates of the observed data. All data – both observed and model data - are then interpolated to a regular 200-meter grid that is used in this study.

Manuscript changes: We have extended Section 2.2, which now reads: "*We used ERA-Interim reanalysis \citep{DeeEtal2011} data as a reference in the evaluation. To avoid the direct impact of the observed data on our comparison reference (RO data are assimilated by ERA-Interim), we used the reanalysis forecasts rather than analyses. ERA-Interim provides forecasts at three-hour intervals, intialized at 00 and 12 UTC. Hence, the shortest possible forecast time vary from 3 hours to 12 hours. For each RO event, a co-located vertical profile of model data was obtained by interpolation in the global forecast fields representing the atmospheric state at three-hour intervals (UTC 00, 03, \ldots) on a $1.0^{\circ}\times1.0^{\circ}$ latitude-longitude grid. The model data are forward-modelled to the set of geophysical variables used in this study. Dry temperature profiles are computed from the ERA-Interim refractivities using the same method as for the observed profiles (see Section~3.1). This is followed by monthly averaging in latitude bins and interpolation onto an equidistant 200 meter height grid, using the methods described in Section~3.4*".

(2)

Reviewer: P5 lines 12-15. Thepaper mentions 1D-Var retrievals of water vapor, yet no water vapor comparisons of the RO data sets with themselves or with ERA-I are presented. Presumably this is outside the scope of this paper, but perhaps this should be stated here because I was expecting to see some water vapor comparisons later on.

Author response: The 1D-Var retrievals are mentioned for completeness. From a climate perspective, the 1D-Var retrievals come with their own set of problems that are only partly related to the more fundamental errors and uncertainties of the bending-angle, refractivity, and dry temperature data. A discussion of the problems and errors with the 1D-Var retrievals is out of the scope of this paper. We will mention this in Section 3.1 of the updated manuscript by adding a sentence.

Manuscript changes: Added a sentence to Section 3.1: "*The tropospheric variables retrieved through a 1D-Var algorithm are not discussed further in this paper*".

(3)

Reviewer: The authors should state how ERA-I dry temperatures are computed for comparison with the RO dry temperatures, as in Fig. 6. Presumably the ERA dry temperature is computed by first computing the ERA-I refractivity N using the ERA-I temperature, pressure and water vapor pressure, and then computing ERA-I dry temperature from the Smith-Weintraub equation with e=0 (i.e., Tdry=77.6P/N). At first I thought this was RO dry temperatures compared with ERA-I temperatures, but this is clearly not the case.

Author response: The ERA-I dry temperatures are computed by first forward modelling ERA-I pressure, temperature, and humidity to ERA-I refractivity. Then we use similar processing as for the observed refractivity: we assume hydrostatic equilibrium and do a downward vertical integration of the hydrostatic integral from the top of the atmosphere to get the dry pressure. The refractivity and the dry pressure then give us the dry temperature under the assumption of zero water vapour pressure. We will update the text in Section 2.2 to be more specific on this – see suggested text update under point 1 above.

Manuscript changes: see under (1) above.

(4)

Reviewer: P20 lines 6-7: The variation of the results with latitude bands compared to the global average is important. Most of the results in this paper are global means. The results are quite likely to be different in the tropics where the environment is more challenging for RO. Perhaps this should be mentioned somewhere toward the front?

Author response: Yes, many of the results in the paper are global means, as we had to limit the discussions somehow. However, Figure 8, which is a key figure in this paper, gives a quite detailed view of the latitudinal differences. These are discussed in some detail in Section 5.1. There are indications both in the Abstract and in Section 5.2 that the consistency amongst RO missions is lower for limited latitude bands than for global means.

It is true that the environment is challenging for RO in the tropics, particularly in the lower and middle troposphere where humidity has a large impact on the refractivity. In general, the lower troposphere is challenging, and this is also where we find the largest biases between the RO missions. The main limitation of the study is actually in the vertical. The large differences between the RO missions below 8 km are only mentioned briefly while the differences above 8 km are discussed in some detail. We will add a sentence in the Introduction to point out this limitation.

Manuscript changes: Sentence added in the Introduction: "*The evaluation is largely limited to the stratosphere and the upper troposphere, above about 6-8 km*".

Reviewer #1, editorial/technical comments

(5) P2 line 4: "develops" should be "develop"
(6) P2 line 24 and P26 line 3: "Rieck" should be "Rieckh".
(7) P3 line 6: Please define "UCAR" (University Corporation for Atmospheric Research).
(8) P9 line 8-"accuracy" should be "bias".
(9) P13 line 6: better wording might be "..profiles deviate somewhat. . ."
(10) P13 line 10: "lead" should be "leads"
(11) P17 line 7: Suggested rewording rather than starting sentence with "Believed. . .": "This bias is believed to be. . .." or "We attribute this bias to the undersampling... . ."
(12) P18 line 2: Suggested rewording rather than starting sentence with "Believed. . ..":"This effect is believed to be. . ." or "We attribute this effect to the undersampling of the. . ."
(13) P18 line 7: So far not Sofar

Manuscript changes: Text updated accordingly.

Reviewer #2, specific comments

(1)
Reviewer: I would like to see more of the results broken down in different latitude bands, not just global averages. To minimize the number of figures, such information can perhaps be summarized in a table. After all, the CDR is a gridded dataset that covers all latitudes, so users of this dataset would want to know how the results vary as a function of latitudes. (There are some mentions that the differences are up to a factor 2 larger for certain latitude bands but this is too vague.)

Author response: Note that Figure 8 is actually broken down into latitudes. The latitude-resolved results in Figure 8 are discussed in Section 5.1. It is true that the vertically averaged time-series in Figs. 9, 10, and 11 are only provided as global averages. The idea is that Figure 8 provides a detailed view across height and latitude for a limited set of RO missions and for bending angle only, while Figs. 9, 10, and 11 provide a more complete view across RO missions and for all three geophysical variables. Anyway, we have now added a table showing latitude-resolved time-averaged RO mission differences as suggested by the reviewer.

Manuscript changes: Added a Table 2 and references to it in the end of Sections 5.2.1, 5.2.2, and 5.2.3.

(2)
Reviewer: Is the gridded dataset a 2D (latitude-height) or 3D (lat-lon-height)? Please specify up front.

Author response: The monthly mean data are provided on a 2D latitude-height grid. This is now mentioned in the Introduction.

Manuscript changes: Added to the Introduction: "…. *we present results from an evaluation of the ROM SAF monthly-mean climate data records (CDRs) provided on a two-dimensional latitude-height grid, ….*".

(3)
Reviewer: To provide a broader context for the readers, it would be useful to compare the multi-mission differences among the missions to the structural uncertainty estimated from the multi-center comparisons (e.g., Ho et al. 2009; Steiner et al. 2013).

Author response: The multi-center inter-comparisons (Ho et al, 2009; Steiner et al., 2013) primarily resulted in quantifications of the trends in the differences between processing centers for RO variables retrieved from CHAMP data. Those results gave a first view of the uncertainties in the trends to be expected from the differences between processing centers (limited to CHAMP data only). The RO mission differences presented here constitute an additional source of uncertainty for the trends of long-term time series of RO data. However, we have only presented the differences themselves. We have not described the consequences for the long-term trends. We will come back to this issue in a future study, and by then we will hopefully be able to compare our results with an extended study of multi-center differences including several RO missions.

(4)

Reviewer: The paper mentions in multiple occasions that sampling error correction is necessary for combining the data from various RO missions. I don't necessarily agree. Wouldn't it be possible to combine the data and then perform the sampling error correction on the combined dataset?

Author response: Yes, it is possible to first combine the data and then do sampling error correction. This is actually our preferred method for constructing long multi-mission data records. However, it does not avoid the need to do sampling error correction as a remedy for differences in the sampling characteristics between RO missions. Either way (first do sampling error correction then combine data sets, or first combine data sets and then do sampling error correction) it is necessary to do sampling error correction if we want to avoid spurious long-term variability in the combined time series as new satellite missions replace older ones.

(5)

Reviewer: Section 2.1: Table 1 listed the input data used for the study. It would be helpful to provide some key information up front about the changes in versions since they can lead to abrupt changes in the time series. For example, we learned later on that the GRACE processing had changed from single-differencing to zero-differencing. Please also mention rationale for using the UCAR Metop input data even though it is not officially part of the CDR. Regarding the low-level input data, was there quality control done by UCAR and EUMETSAT at this point?

Author response: We have updated the first paragraph of Section 2.1 accordingly. The statement about the change from single- to zero-differencing was removed. This statement was based on a) the paper *Angerer et al.* (2017) where it was stated that UCAR used single-differencing for GRACE data version 2010.2640, and b) it was consistent with what we see in our time series (e.g., Figure 2). However, from personal communication with D. Hunt at UCAR we have been informed that all GRACE data from UCAR are based on zero-differencing. Concerning the question of whether there is quality control done on the low-level data, we know that EUMETSAT does it and it is actually taken into account in our own QC. Regarding UCAR data, we would assume there is a certain degree of low-level quality control and data rejection. However, it is difficult to find detailed documentation on that, and it is also a bit out of the scope of this article to go into such detail.

Manuscript changes: The first paragraph of Section 2.1 now reads: "*The ROM SAF CDR v1.0 includes data from four RO missions: CHAMP, GRACE, COSMIC, and Metop. The processing of data from the first three missions was based on low-level input data from the University Corporation for Atmospheric Research (UCAR), while the Metop data were processed with input data from EUMETSAT. In addition, we also processed Metop data using input data from UCAR, to allow for investigation of differences related to the low-level input data. The low-level input data consist of amplitude and excess phase data, together with positions and velocities for Global Positioning System (GPS) and low-Earth orbit (LEO) satellites. The input data versions are shown in Table 1. During the time period, COSMIC and GRACE experienced version updates that involved low-level processing software changes at UCAR*".

(6)

Reviewer: Section 3.2: The results on the bending angle quality from different missions are interesting. However, for most readers not familiar to RO, it is not clear how they are relevant to the retrieved physical parameters shown later. Please provide some discussions of that.

Author response: We have added a new second paragraph to Section 3.2.

Manuscript changes: A new, second, paragraph was added to Section 3.2: "*High noise levels and other errors may lead to occultations being rejected by the quality screening. Any errors in the retained bending angles may propagate further down the processing chain, since bending angles are the starting point for the retrieval of the other geophysical variables. In particular, bending angle data in the upper stratosphere is affected by residual ionospheric errors resulting in errors in refractivity and dry temperature lower down in the stratosphere*". In addition, the last sentence of the fourth paragraph was slightly changed. It now reads: "*This stepwise change is related to a switch of input data version (Table 1) involving several changes in UCAR's processing software (D. Hunt, pers. comm.)*".

(7)

Reviewer: Figure 5: There is a significant jump between 2009-2010 between 20-30 km. Is that related to the "update of the COSMIC NRT data processing in October 2009"? Please provide more details.

Author response: The jump that we see in the RO-ERAI differences (Figure 5) in October 2009 is related to a bias shift in ERA-Interim. It is our understanding that this bias shift in ERA-Interim was caused by an update of the COSMIC data processing at UCAR, leading to a change in the data being assimilated by the ERA-Interim reanalysis system. The reference provided (Healy, 2013) states that the change in the operational processing of COSMIC data took place on October 12, 2009, which is consistent with what we see in our data records.

Manuscript changes: Minor change of a sentence in the second paragraph of Section 4 to "*… an update of the COSMIC data processing in October 2009 (Healy, 2013), …*".

(8)

Reviewer: Figure 5: In 4-8 km height range, the inter-mission differences in bending angles seem big, but the refractivity differences appear to be much smaller (except for CHAMP). Why? Is it a matter of vertical scale in the plotting?

Author response: The inter-mission differences in the 4-8 km height range are actually larger for bending angle than for refractivity. The effect can also be seen in Figure 6 for time-averaged data. It is partly a matter of the vertical scale of the plots: the bending angle is plotted as a function of impact altitude while the refractivity is plotted versus altitude. The two vertical height variables can differ by up to 1-2 km near the surface. The downward propagation of information through the Abel transform also means that at a particular height (altitude or impact altitude) in the 4-8 km height range, less biased data from higher atmospheric layers contribute a larger fraction of the data for the refractivity compared to bending angle.

(9)

Reviewer: Section 3.3: The authors describe the quality control criteria used to remove bad data. However, the descriptions were rather generic. For example, it's not clear how criteria a, b, and c are distinct from each other. I understand this is a complex subject and not the main focus of this paper, and the authors did reference a document on the ROM SAF web site. However, a little more technical information, along with some statistics on the percentage of occultations removed from each criterion, would be useful.

Author response: We have now updated the text in Section 3.3 in order to provide some more information.

Manuscript changes: The text in Section 3.3 has been updated and now reads: *"Before processing the atmospheric profiles to gridded monthly-mean data, all profiles are checked against a set of quality criteria. The quality criteria include tests to identify occultations that a) do not provide any meaningful bending angles or only provide bending angles in a limited height range, b) have degraded bending angles indicated by increased noise in the L2 signal or by certain types of deviation from a bending-angle climatology, c) could be regarded as outliers as evidenced by comparison with ECMWF reanalysis data, or d) encounter problems in the 1D-Var processing. More detailed descriptions of the data quality screening are found in the series of validation reports available at the ROM SAF web site*
 (http://www.romsaf.org/product_documents.php).
    If an occultation does not pass one or several of the tests in a, b, or c, the bending angle, refractivity, and dry variables are marked as non-nominal. Otherwise, they are regarded as nominal and the refractivity profiles are passed on to the 1D-Var processing. The fraction of data rejected in the quality screening varies with time (Fig. 1). On average, around 10-20 % of the occultations are rejected, although with large differences between the RO missions. Metop and GRACE show the highest throughput of data; almost no data are rejected by criterion a and about 5-10 % are rejected by criteria b and c. COSMIC and CHAMP have roughly similar overall rejection rates. However, for COSMIC about 5-10 % of data are rejected by criterion a, while for CHAMP that criterion removes about 15 % or even more of the data".

(10)

Reviewer: Section 3.4: The authors describe the binning and averaging technique used to produce the gridded monthly means. It introduces a technique where each latitude bin is divided into two sub-bins to obtain latitude weighted average within each bin. What is the advantage of doing it this way vs. something like Gaussian weighting (for example)?

Author response: The purpose of the division into sub-bins is to compute a bin mean that as closely as possible approximate an area-weighted mean:

$$\bar{X} = \frac{1}{A} \int X dA$$

where *X* is a geophysical variable and *A* is the bin area. The main problem that is addressed by the weighting is non-uniform sampling of *X* across the bin. As stated in the text, we only consider the non-uniform distribution of observations in latitude. The division of the 5-degree latitude bins into two sub-bins can be regarded as a very coarse discretization of the above mentioned integral. During time periods with a lot of RO data, and for monthly bins, a finer discretization would be possible. However, we want to use the same binning across the whole time series, from CHAMP to COSMIC and Metop, and, hence, we chose two sub-bins for the 5-degree main bins.

(11)

Reviewer: Section 3.6: The authors stated that "In the generation of anomaly time series, the same seasonal cycle should be used for all missions and throughout the time series." I would argue that if you want to look at differences in seasonal cycles and anomalies separately, it would be better to derive the seasonal cycle from each time series and obtain the anomalies by removing corresponding seasonal cycles.

Author response: First a clarification: as stated in Section 3.6, by the "mean seasonal cycle" we mean the long-term average as a function of latitude, height, and season. For each latitude, height, and January month, we average across many years. Similarly for all February months, etc. This is what Eq. 10 expresses. Our "mean seasonal cycle" is a sum of the long-term all-season means and the seasonal cycle on top of that long-term mean. We do not separate the two.

   Let's say that we have two overlapping missions, A and B. Should we compute anomalies for A using data from A as reference, and anomalies for B using data from B as reference, we wouldn't be able to detect constant biases between the missions. The same argument could be made even if we separate long-term all-season means from the seasonal cycle: we wouldn't be able to detect differences in the seasonal cycle as measured by A and by B.

   In addition, assume that we generate long-term climatologies from several non-overlapping RO missions. If we construct the anomalies using mission-specific anomaly references, we would not only remove any biases (systematic errors) between the missions, but also true climatological time variations.

(12)

Reviewer: Figure 5: There is a small constant bias (after 2006) between 8-20 km between RO and ERA-interim in both bending angle and refractivity. Can you comment on that?

Author response: In Section 4, we now mention that even though the magnitude of the RO-ERAI difference suddenly decreased as a result of the start of assimilation of COSMIC data in 2006, the difference remains slightly negative in the 12-20 km height interval throughout the data record.

Manuscript changes: The updated sentence in the second paragraph of Section 4 now reads: *"In December 2006 the magnitude of the bias suddenly decreased in the 12-20 km and 30-40 km intervals, even though it remained slightly negative in the 12-20 km range throughout the time period"*.

(13)

Reviewer: Section 5.1, p. 17: "Large-scale hemispherically asymmetric (north-south) Metop-COSMIC bias on the order of 0.1% above 35-40 km, and increasing upward . . . is believed to be related to differences in LEO satellite orbits from the two sources of input data." Can you provide some evidence of this claim?

Author response: In our input data we find differences between the EUMETSAT and UCAR LEO orbits with the right magnitude, and with a periodicity of one orbit, that potentially could explain the hemispheric differences that we observe (see Figure A below). The source, and the ultimate cause, of these orbital differences is still a matter of ongoing work. More detailed information can also be found in the series of validation reports available at http://www.romsaf.org/product_documents.php. It is, however, still premature and out of the scope of this paper to include a detailed description of what we believe is the cause of the differences.
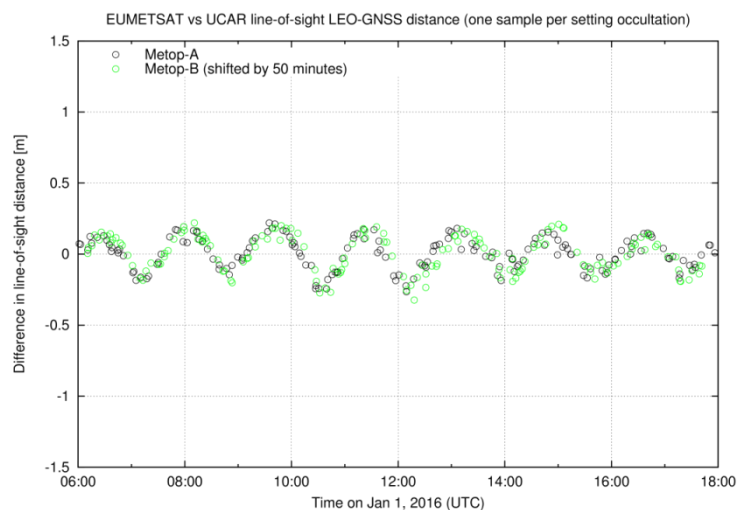
Figure A. *Differences between Metop orbits from EUMETSAT and UCAR, as quantified by the differences in the Metop-GPS distances.*

(14)

Reviewer: In general, for the bullets in pp. 15-17, the authors should use references when possible to substantiate the arguments.

Author response: The purpose of the bullet list is to summarize our own observations of inter-mission biases as shown in Figure 8 (and to some extent Figure 9). In some bullets we also point out possible causal relations, a few of them well-known while other should be regarded as preliminary conclusions from the present study itself. We have added references to the first bullet (signal tracking issues) and the third bullet (Metop software upgrades).

Manuscript changes: First bullet in Section 5.1 now ends with: "… the atmosphere (e.g., *Sokolovskiy et al.,* 2010)". Third bullet ends with: "… appeared (described in the series of ROM SAF validation reports at *http://www.romsaf.org/product_documents.php*)".

(15)

Reviewer: Section 5.2.1, p. 18: The difference between Metop and COSMIC bending angle over 8-30 km is large compared with other missions. Would that change if the Metop-UCAR input data were used instead?

Author response: No, the Metop-COSMIC globally averaged differences that we see in Figure 9 would be very similar for Metop(UCAR) as for Metop(EUM). You can see this in Figure 8, by comparing the middle column (Metop(UCAR) minus COSMIC) with the rightmost column (Metop(EUM) minus COSMIC). The dominating differences between Metop(UCAR) and Metop(EUM) are a large-scale hemispheric asymmetry (see the top panels in Figure 8) and some differences in the lower troposphere.

(16)

Reviewer: Also: "Metop shows . . . a stepwise decrease of the bias in mid-2013. . ." Any idea why?

Author response: The Metop-COSMIC bias shift in 2013 can also be identified in Figure 8, where you also can see the latitude distribution of the shift. It is related to firmware upgrades in the Metop RO instruments described by the third bullet in Section 5.1.

(17)

Reviewer: P. 2, line 4: Mission acronyms were never spelled out.

Author response: Mission acronyms are now spelled out in the Introduction.

Manuscript changes: The second paragraph of Section 1 now begins: "The RO Meteorology Satellite Application Facility (ROM SAF), which is a decentralized operational RO processing center under EUMETSAT, has recently undertaken a reprocessing of RO data from four satellite missions: CHAMP (CHAllenging Minisatellite Payload; *Wickert et al.*, 2001), GRACE (Gravity Recovery and Climate Experiment; *Beyerle et al.*, 2005), COSMIC (Constellation Observing System for Meteorology, Ionosphere, and Climate; *Anthes et al.*, 2008), and Metop (*Luntama et al.*, 2008)".

(18)
Reviewer: P. 4, line 5: "The model data profiles are forward-modelled to the set of observed geophysical variables." Please provide more information on the forward modelling (e.g., from what variable to what variable, do you account for tangent point drifts, upper boundary for Abel integration, etc.).

Author response: We have now expanded Section 2.2 somewhat, based on the comments of both reviewers. We also added a new reference (*Healy and Thépaut*, 2006) that provides a detailed description of the forward-modelling to refractivity and bending angle.

Manuscript changes: The updated Section 2.2 now reads: *"We used ERA-Interim reanalysis (Dee et al., 2011) data as a reference in the evaluation. To avoid the direct impact of the observed data on our comparison reference (RO data are assimilated by ERA-Interim), we used the reanalysis forecasts rather than analyses. ERA-Interim provides forecasts at three-hour intervals, intialized at 00 and 12 UTC. Hence, the shortest possible forecast time vary from 3 hours to 12 hours. For each RO event, a co-located vertical profile of model data was obtained by interpolation in the global forecast fields representing the atmospheric state at three-hour intervals (UTC 00, 03, …) on a 1.0^O x 1.0^O latitude-longitude grid.*

The vertical profiles of model data (pressure, temperature, and humidity as function of geopotential height) are forward-modelled to the set of geophysical variables used in this study. The model refractivity is calculated from the Smith-Weintraub equation, and the bending angles are obtained by an Abel integral over the refractivity profile assuming an exponential decay above the model top (Healy and Thépaut, 2006). Dry temperature profiles are computed from the model refractivities using the same method as for the observed profiles (see Section 3.1). This is followed by monthly averaging in latitude bins and interpolation onto an equidistant 200 meter height grid, using the methods described in Section 3.4".

(19)
Reviewer: P. 18, line 6: Change "Sofar" to "So far"

Author response: Done.

Manuscript changes: Section 5.2 now begin: "*The RO mission differences have so far been described …."*.

(20)
Reviewer: In the abstract and elsewhere, the authors differentiate the upper troposphere from lower by saying "6-8 km". I find this a little awkward (is it 6 or 8 km?). I think it's better to just use 8 km as the boundary.

Author response: The notation "6-8 km" was used to indicate that this is not an exact altitude. We have now changed four occurrences of "6-8 km" in the manuscript to "8 km" (in Abstract and in Sections 1 and 5.1).

Manuscript changes: Changed four occurrences of "6-8 km" to "8 km" in Abstract and in Sections 1 and 5.1.

Evaluation of the 15-year ROM SAF monthly mean GPS radio occultation climate data record

Hans Gleisner[1], Kent B. Lauritsen[1], Johannes K. Nielsen[1], and Stig Syndergaard[1]

[1]Danish Meteorological Institute, Lyngbyvej 100, Copenhagen, Denmark

Correspondence: Hans Gleisner (hgl@dmi.dk)

Abstract. We here present results from an evaluation of the ROM SAF gridded monthly-mean climate data record (CDR v1.0), based on GPS radio occultation (RO) data from the CHAMP, GRACE, COSMIC, and Metop satellite missions. Systematic differences between RO missions, as well as differences of RO data relative to ERA-Interim reanalysis data, are quantified. The methods used to generate gridded monthly mean data are described, and the correction of monthly-mean RO climatologies for

5 sampling errors, which is essential for combining data from RO missions with different sampling characteristics, is evaluated.

We find a good overall agreement between the ROM SAF gridded monthly-mean CDR and the ERA-Interim reanalysis, particularly in the 8-30 km height interval. Here, the differences largely reflect time-varying biases in ERA-Interim, suggesting that the RO data record has a better long-term stability than ERA-Interim. Above 30-40 km altitude, the differences are larger, particularly for the pre-COSMIC era.

10 In the 8–30 km altitude region, the observational data record exhibits a high degree of internal consistency between the RO satellite missions, allowing us to combine data into multi-mission records. For global mean bending angle the consistency is better than 0.04%, for refractivity 0.05%, and for global mean dry temperature the consistency is better than 0.15 K in this height interval. At altitudes up to 40 km, these numbers increase to 0.08%, 0.11%, and 0.50 K, respectively. The numbers can be up to a factor of 2 larger for certain latitude bands compared to global means. Below about ~~6-8~~ 8 km the RO mission

15 differences are larger, reducing the possibilities to generate multi-mission data records. We also find that the residual sampling errors are about one third of the original and that they include a component most likely related to diurnal or semi-diurnal cycles.

1 Introduction

Radio Occultation (RO) measurements, exploiting radio signals emitted by Global Navigation Satellite System (GNSS) satel-

20 lites, are increasingly making important contributions to the global observing system. RO data now have a significant impact in weather forecasting ~~(e.g., ?Cardinali and Healy, 2014)~~ (e.g., *Healy et al.*, 2005; *Cardinali and Healy, 2014*) and in atmospheric reanalysis (*Poli et al*, 2010; *Simmons et al.*, 2017), and as the RO data records become longer, they are also increasingly useful for climate monitoring and climate studies (e.g., *Steiner et al.*, 2011; *Anthes*, 2011). The RO measurement technique has

a number of attractive features: it provides geophysical information with high vertical resolution throughout the troposphere and stratosphere, it is insensitive to clouds and the underlying surface, and it has an intrinsic long-term stability that does not rely on inter-calibration between satellites or instruments (*Kursinski et al.*, 1997; *Leroy et al.*, 2006). The latter feature is particularly important for climate applications where small differences in atmospheric properties that ~~develops~~ develop over

5   decades are monitored.

The RO Meteorology Satellite Application Facility (ROM SAF), which is a decentralized operational RO processing center under EUMETSAT, has recently undertaken a reprocessing of RO data from four satellite missions: CHAMP ~~(*Wickert et al.*, 2001), GRACE (*Beyerle et al.*, 2005), COSMIC (*Anthes et al.*, 2008), and~~ (CHAllenging Minisatellite Payload; *Wickert et al.*, 2001), GRACE (Gravity Recovery and Climate Experiment; *Beyerle et al.*, 2005), COSMIC (Constellation Observing System for

10   Meteorology, Ionosphere, and Climate; *Anthes et al.*, 2008), and Metop (*Luntama et al.*, 2008). The reprocessed RO data cover the time period from late 2001 to end of 2016. Over the last few years, reprocessing activities including these four RO missions, or subsets of them, have been undertaken by several processing centers in a joint effort to quantify the structural uncertainty of RO data. The results have been described in *Ho et al.* (2009, 2012) and *Steiner et al.* (2013), where the impacts of using different processing schemes were investigated. At low- and mid-latitudes between 8 and 25 km the associated structural un-

15   certainties were found to be small enough for RO data to be used in climate change detection studies. For higher altitudes and latitudes, there are mission-specific limitations that need to be considered.

RO data from the four satellite missions included in the ROM SAF reprocessing have somewhat different characteristics related to instrumental noise, signal tracking methods and accuracies, data numbers, and spatio-temporal sampling characteristics. The low-level data (excess phase, amplitude, and satellite orbit data) may also exhibit subtle differences de-

20   pending on the source of those data. If such differences propagate to the retrieved geophysical monthly-mean data, and are large enough, the evolution of the global RO constellation may lead to spurious long-term variability as new satellite missions replace older ones. However, despite differences between the RO missions there is an expectation that measurements from different missions can be combined without any adjustments or inter-calibrations to form long time series of RO data, provided that they use the same processing scheme (e.g., *Foelsche et al.*, 2011; *Angerer et al.*, 2017). Multi-mission

25   RO time series have been used in several studies, implicitly assuming inter-mission consistency, e.g., in studies of atmospheric temperature trends (*Ladstädter et al.*, 2011; *Khaykin et al.*, 2017; *Leroy et al.*, 2018), in climate model evaluation studies (*Lackner et al.*, 2011; *Ao et al.*, 2015; *Schmidt et al.*, 2016), and in studies of atmospheric structure and dynamics ~~(*Scherllin-Pirscher et al.*, 2012, 2014; **?**; *Wilhelmsen et al.*, 2018)~~(*Scherllin-Pirscher et al.*, 2012, 2014; *Rieckh et al.*, 2014; *Wilhelmsen et al.*, ...)
.

30   In this paper we present results from an evaluation of the ROM SAF ~~gridded~~ monthly-mean climate data records (CDRs) provided on a two-dimensional latitude-height grid, with a focus on the temporal stability of the data series and on the differences between the RO missions. The evaluation is largely limited to the stratosphere and the upper troposphere, above about 8 km. The methods used to generate the gridded monthly mean data and the de-seasonalized anomalies are described, including the sampling-error correction method. The observational RO data time series are compared to the ERA-Interim reanalysis. The

35   consistency of climatologies obtained from different RO missions during mission overlap periods are studied, with a view to

Table 1. Low-level input data to ROM SAF CDR v1.0.

Mission	Data provider	Version	Time period
CHAMP	UCAR	2014.0140	2001-09 to 2008-09
GRACE	UCAR	2010.2640	2007-03 to 2014-03
		2014.2760	2014-04 to 2016-12
COSMIC	UCAR	2013.3520	2006-07 to 2014-04
		2014.2860	2014-05 to 2016-12
Metop	EUMETSAT	1.4	2006-12 to 2016-12
Metop[1]	UCAR	2016.0120	2008-03 to 2016-12

[1]Not a formal part of CDR v1.0.
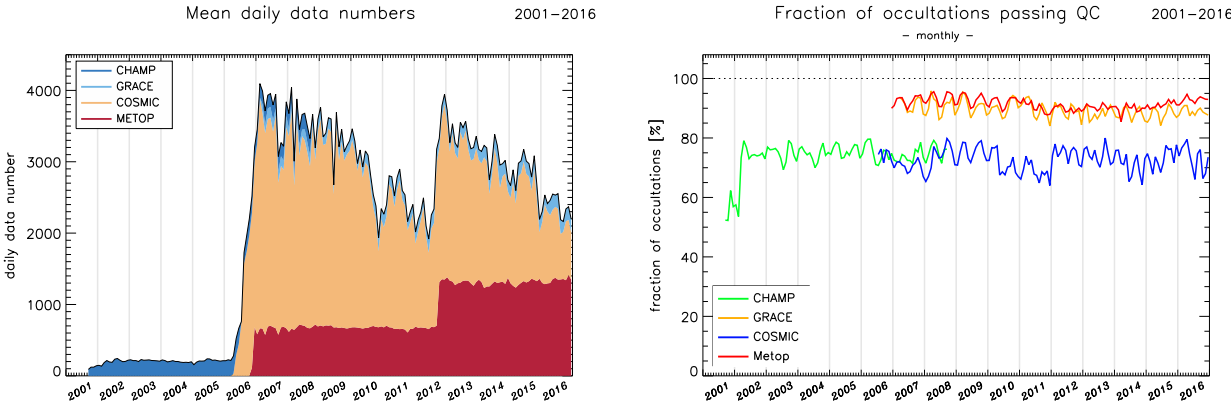
Figure 1. Left panel: Mean daily number of occultation events used in the generation of the ROM SAF climate data record, based on the four RO missions CHAMP , GRACE, COSMIC, and Metop. Right panel: Fraction of occultations available for the generation of gridded monthly mean data, after quality screening of the input profile data.

identify systematic differences that may have an impact on data series constructed from multiple RO missions. We also evaluate the sampling-error correction, which is essential for combining data from RO missions with different sampling characteristics.

Section 2 provides an overview of the data that are being evaluated, and ~~of~~ the data used as a reference for the evaluation. In Section 3, the processing of the data to gridded monthly mean climatologies is described, including a discussion ~~on~~ of the time evolution of bending-angle quality and ~~on~~ the quality screening of the data. Section 4 describes a comparison with ERA-Interim reanalysis data, while in Section 5 the consistency of climatologies obtained from different RO missions are analyzed. The study results are discussed and the main conclusions are summarized in Section 6.

3

2 Data

2.1 GPS radio occultation measurements

The ROM SAF CDR v1.0 includes data from four RO missions: CHAMP, GRACE, COSMIC, and Metop. The processing of data from the first three missions ~~were~~ was based on low-level input data from ~~UCAR~~the University Corporation for Atmospheric Research (UCAR), while the Metop data were processed with input data from EUMETSAT. In addition, we also processed Metop data using input data from UCAR, to allow for investigation of differences related to the low-level input data. The low-level input data consist of amplitude and excess phase data, together with positions and velocities for Global Positioning System (GPS) and low-Earth orbit (LEO) satellites. The input data versions are shown in Table 1. During the time period, COSMIC and GRACE experienced version updates that involved low-level processing software changes at UCAR.

The input data were processed to geophysical data using the ROM SAF processing system GPAC v2.3.0, with the Radio Occultation Processing Package (ROPP) v8.1 as an integral part. The variables discussed in the present article are bending angle, refractivity, and dry temperature (the concept of "dry" variables is described in Section 3.1). The ROM SAF CDRs also contains dry pressure and dry geopotential height (the geopotential heights of dry-pressure surfaces) as well as temperature and humidity in atmospheric regions where humidity has a significant influence on the refractivity. The CDRs also include tropopause height derived from the dry-temperature profiles, as well as from bending angle and refractivity profiles.

In total, the four RO missions include nearly 12 million occultations collected from September 2001 to December 2016. Fig. 1 shows that the mean daily number of occultations peaked at well above 3000 during 2007–2009. The launch of the second Metop satellite, in combination with an update of the operational mode of the COSMIC mission, led to a second peak in the daily data numbers in 2013. After removal of data based on quality screening, about 10 million atmospheric profiles remain for generation of the CDRs.

~~Low-level input data to ROM SAF CDR v1.0. Mission Data provider Version Time period CHAMP UCAR 2014.0140 2001-09 to 2008-09 GRACE UCAR 2010.2640 2007-03 to 2014-03 2014.2760 2014-04 to 2016-12 COSMIC UCAR 2013.3520 2006-07 to 2014-04 2014.2860 2014-05 to 2016-12 Metop EUMETSAT 1.4 2006-12 to 2016-12 Metop[2] UCAR 2016.0120 2008-03 to 2016-12~~

~~Left panel: Mean daily number of occultation events used in the generation of the ROM SAF climate data record, based on the four RO missions CHAMP , GRACE, COSMIC, and Metop. Right panel: Fraction of occultations available for the generation of gridded monthly mean data, after quality screening of the input profile data.~~

2.2 ERA-Interim reanalysis data

We used ERA-Interim reanalysis (*Dee et al.*, 2011) ~~short-term forecasts~~ data as a reference in the evaluation. To avoid the direct impact of the observed data on our comparison reference (RO data are assimilated by ERA-Interim), we used the reanalysis forecasts rather than analyses. ERA-Interim provides forecasts at three-hour intervals, intialized at 00 and 12 UTC. Hence, the

---

[2] ~~fnote1~~
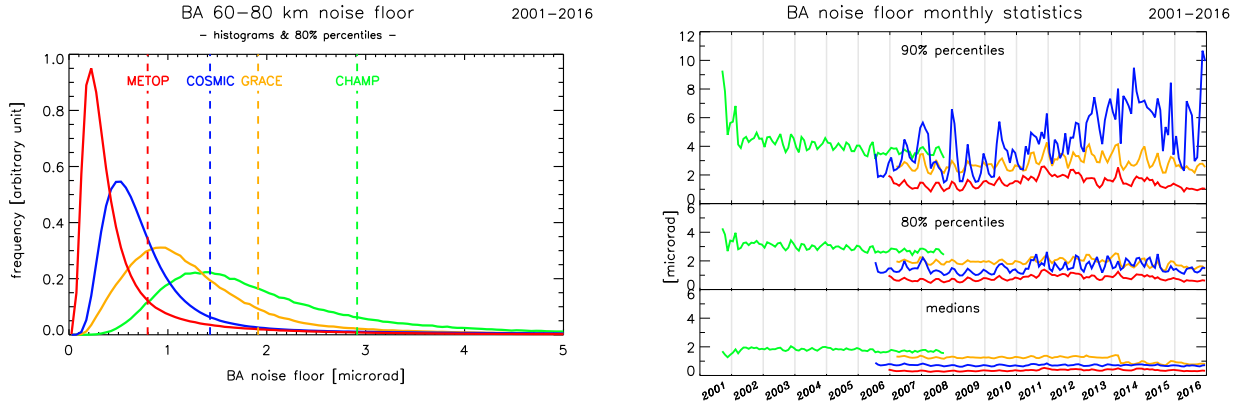
Figure 2. Bending angle (BA) noise floor for the RO missions CHAMP, GRACE, COSMIC, and Metop: overall distributions in the left panel, and the time evolution of the associated medians, 80% and 90% percentiles in the panel to the right. The statistics include all data, also those that are rejected in the quality screening.

shortest possible forecast time vary from 3 hours to 12 hours. For each RO event, a co-located vertical profile of model data was obtained by interpolation in the global forecast fields representing the atmospheric state at three-hour intervals (UTC 00, 03, . . .) on a $1.0° \times 1.0°$ latitude-longitude grid. The model data profiles

The vertical profiles of model data (pressure, temperature, and humidity as function of geopotential height) are forward-

5 modelled to the set of observed geophysical variables . geophysical variables used in this study. The model refractivity is calculated from the Smith-Weintraub equation, and the bending angles are obtained by an Abel integral over the refractivity profile assuming an exponential decay above the model top (*Healy and Thépaut*, 2006). Dry temperature profiles are computed from the model refractivities using the same method as for the observed profiles (see Section 3.1). This is followed by monthly averaging in latitude bins using the same methods as for the observed profiles (and interpolation onto an equidistant 200 meter

10 height grid, using the methods described in Section 3.4).

3 ROM SAF processing of RO data

This section provides a short description of the processing of RO measurements to atmospheric profiles of bending angles and associated geophysical variables, and further on to the gridded monthly mean data. The quality of the bending angles, and the quality screening are also briefly discussed.

15 ### 3.1 Processing to atmospheric profiles

The input data to the ROM SAF processing consist of amplitude and excess phase time series collected during the satellite occultation events, together with precise orbits for the GPS and LEO satellites. The input data were obtained from EUMETSAT (for the Metop mission) and from UCAR (for the CHAMP, GRACE, COSMIC, and Metop missions).

5

Bending angles at the two GPS frequencies L1 (1575.42 MHz) and L2 (1227.60 MHz) are calculated from the excess phase and amplitude data through a geometrical optics approach (*Kursinski et al.*, 1997) above 25 km, a wave optics approach (*Gorbunov and Lauritsen*, 2004) below 20 km, and a gradual transition in between. After correction for ionospheric effects through a linear combination of the L1 and L2 bending angles, we obtained the so called "raw" ionospheric corrected bending

5   angle. With optimal linear combination (*Gorbunov*, 2002), using a bending angle climatology (*Scherllin-Pirscher et al.*, 2015), we obtain statistically optimized bending angle profiles that can be used to further retrieve geophysical information. Under the assumption of local spherical symmetry in the vicinity of the occultation point, we use the Abel transform (*Fjeldbo et al.*, 1971) to compute a vertical refractivity profile from the bending angles. Details of the processing steps can be found in algorithm technical baseline documents (ATBDs) at the ROM SAF web site (*http://www.romsaf.org/product_documents.php*).

10   For a dry atmosphere the refractivity is directly proportional to air density. Dry pressure is retrieved from refractivity under the assumption of hydrostatic equilibrium and by ignoring the presence of water vapour. The corresponding dry temperatures are obtained by applying the ideal gas law. The "dry" approximation is a valid assumption in the upper troposphere and stratosphere, where the dry variables are accurate approximations for the corresponding physical variables (*Danzer et al.*, 2014).

15   Under moister conditions, the dry-wet ambiguity can be resolved by a one-dimensional variational (1D-Var) retrieval (*Healy and Eyre*, 2000), using additional information from co-located ERA-Interim short-term forecasts. This gives estimates of the "wet" (physical) temperature and humidity, appropriate for atmospheric regions where humidity has a significant influence on the refractivity. The tropospheric variables retrieved through a 1D-Var algorithm are not discussed further in this paper.

3.2   Bending angle quality

20   The quality of the retrieved bending angles differ between RO missions. In addition to the effects of residual ionospheric noise, the quality depends on RO instrument characteristics as well as on the data processing; the use of single- or double-differencing of excess phases (e.g., *Schreiner et al., 2010*, *von Engeln et al., 2011*), and filtering of the data applied at different steps in the processing (*Schreiner et al.*, 2011). The bending angle noise between 60 and 80 km, an altitude range where bending due to the neutral atmosphere is small, provides and indication of the bending angle quality (*Schreiner et al.*, 2011; *Li et al.*, 2015;

25   *Angerer et al.*, 2017). For each occultation, we compute the standard deviation of the bending angle difference with respect to a fitted background. The smallest standard deviation over any 7.5 km interval between 60 and 80 km is referred to as the bending angle noise floor for an occultation.

High noise levels and other errors may lead to occultations being rejected by the quality screening. Any errors in the retained bending angles may propagate further down the processing chain, as these are the starting point for the retrieval of the other

30   geophysical variables. In particular, bending angle data in the upper stratosphere is affected by residual ionospheric errors resulting in errors in refractivity and dry temperature lower down in the stratosphere.

Fig. 2 shows the noise floor distributions for the four RO missions, and the time evolution of the associated medians and percentiles. The lower panels on the right show that 50% (80%) of the bending angle profiles have noise floors smaller than about 1.8 (2.9) µrad for CHAMP, 1.2 (1.9) µrad for GRACE, 0.7 (1.4) µrad for COSMIC, and 0.4 (0.8) µrad for Metop. These

6
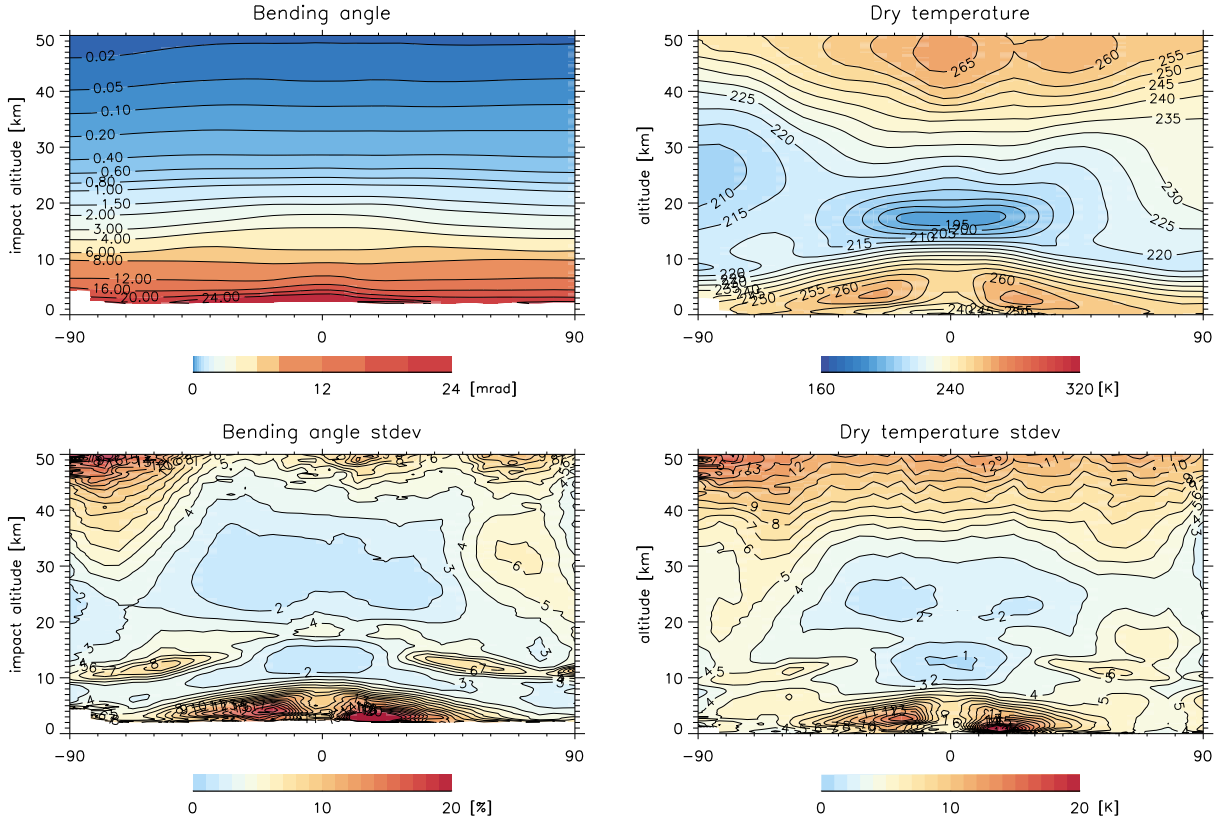
Figure 3. ~~Bending~~ Zonally gridded monthly means and standard deviations for bending angle (~~BA~~left panels) ~~noise floor for the RO missions~~ ~~CHAMP, GRACE, COSMIC,~~ and ~~Metop: overall distributions~~ dry temperature (right panels) in ~~the left panel~~April 2014. The monthly means were sampling-error corrected, and ~~the time evolution of the associated medians, 80% and 90% percentiles in the panel to the right. The statistics~~ include ~~all~~ data ~~, also those that are rejected in~~ from the ~~quality screening~~Metop-A and Metop-B satellites.

numbers are somewhat smaller than those found by *Schreiner et al.* (2011) and *Angerer et al.* (2017) due to their use of standard deviations computed over the whole 60-80 km and 65-80 km height intervals, respectively.

The bulk of the noise floor distributions are relatively constant, as shown by the time evolution of the medians. However, as indicated by the 80% and 90% percentiles time series in Fig. 2, the number of high-noise profiles is more variable. CHAMP exhibits increased noise levels during the first months of the data record. Following an instrument software update in March 2002, the bending angle noise settles at an almost constant level, but with the number of high-noise profiles slowly declining. The bending angle noise in the GRACE data is relatively constant, except for a sudden decrease in April 2014 which affects the bulk of the profiles, not only the number of noisy profiles. This stepwise change is ~~due~~ related to a switch ~~to zero-differencing in the generation of the excess phases~~ of input data version (Table 1) involving several changes in UCAR's ~~version 2014.2760 of GRACE data.~~ processing software (*D. Hunt, pers. comm.*).

For COSMIC, there is a substantial increase with time of a the number of profiles with very high bending angle noise, mainly attributed to rising occultations (not shown here). Metop exhibits a somewhat larger number of high-noise profiles for rising than for setting occultations, and also shows an interesting pattern which may be related to the solar cycle. A similar pattern may be discernible in the GRACE data.

3.3 Data quality screening

Before processing the atmospheric profiles to gridded monthly-mean data, all profiles are checked against a set of quality criteria. The quality criteria include tests to identify occultations that a) ~~are obviously corrupt or show signs of major problems~~ do not provide any meaningful bending angles or only provide bending angles in a limited height range, b) have degraded bending angles indicated by increased noise in the L2 signal or by certain types of deviation from a bending-angle climatology, c) could be regarded as outliers as quantified by comparison with ECMWF reanalysis data, or d) encounter problems in the 1D-Var processing. More detailed descriptions of the data quality screening are found in the series of validation reports available at the ROM SAF web site (*http://www.romsaf.org/product_documents.php*).

If an occultation does not pass one or several of the tests in a, b, or c, the bending angle, refractivity, and dry variables are marked as non-nominal. Otherwise, they are regarded as nominal and the refractivity profiles are passed on to the 1D-Var processing, ~~followed by the associated quality tests~~. The fraction of data rejected in the quality screening varies ~~over time and between the RO satellite missions~~ with time (Fig. 1). On average, around 10-20% of the occultations are rejected, although with large differences between the RO missions. Metop and GRACE show the highest throughput of data; almost no data are rejected by criterion a and about 5-10% are rejected by criteria b and c. COSMIC and CHAMP have roughly similar overall rejection rates. However, for COSMIC about 5-10% of data are rejected by criterion a, while for CHAMP that criterion removes about 15% or even more of the data.

3.4 Monthly averaging in latitude bins

The gridded monthly mean data are obtained by a simple binning-and-averaging technique. Each occultation is assigned to a 5-degree latitude band and calendar month. The RO profiles that pass the quality screening are interpolated onto an equidistant 200 meter height grid (the height variable is impact altitude for bending angles, pressure height for dry geopotential height, and mean-sea level altitude for the other geophysical variables). At each height, and within each bin, the data undergo a weighted averaging. The purpose of the weighting is to reduce the effects of a non-uniform spatial sampling density across a grid box, in order to better approximate an area-weighted mean. The distribution of observations in longitude is nearly uniform and is not explicitly addressed. The distribution of observations in latitude, on the other hand, can be highly non-uniform. This is addressed by sub-dividing each 5-degree latitude bin into two sub-bins, and giving each data point, i, a weight, w_i, according to which sub-bin, s, it belongs to:

$$w_i = \frac{A_s}{A} \frac{n}{n_s} \tag{1}$$
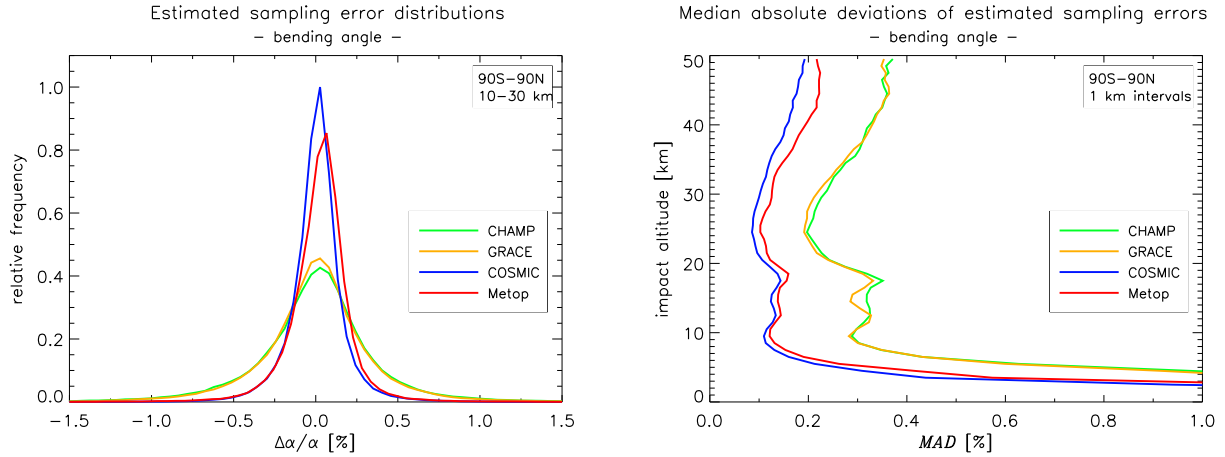
Figure 4. ~~Zonally gridded monthly means and standard deviations for~~ The panel to the left shows distributions of estimated bending angle ~~(left panels)~~ sampling errors based on all monthly bins in the 10-30 km height interval, for the four RO missions CHAMP, GRACE, COSMIC, and ~~dry temperature~~ Metop. The spread of the distributions, as quantified by the median absolute deviation (~~right panels~~MAD)~~in April 2014.~~ ~~The monthly means were sampling-error corrected~~, is about 0.3% for CHAMP and ~~include data from the Metop-A~~ GRACE and ~~Metop-B satellites~~0.1% for COSMIC and Metop. The right-hand panel shows the MADs for height resolved sampling-error distributions.

where A and n are the total area and data number for the bin, and A_s and n_s are area and data number for sub-bin s. Within each latitude bin and calendar month, a weighted arithmetic average is computed as

$$\bar{X}(h) = \frac{\sum w_i X_i(h)}{\sum w_i} \tag{2}$$

where X_i is a geophysical quantity, and $\bar{X}$ is the corresponding monthly mean for the latitude bin. The corresponding weighted

5 standard deviation is given by

$$s(h) = \sqrt{\frac{\sum w_i (X_i(h) - \bar{X}(h))^2}{((n-1)/n) \sum w_i}} \tag{3}$$

using the same weights as in Eq. 2. The dependency of the weights, w_i, and the data numbers, n, on height is not shown explicitly in the above equations. Fig. 3 shows an example of bending angle and dry-temperature means and standard deviations for Metop data from April 2014.

10 ~~The panel to the left shows distributions of estimated bending angle sampling errors based on all monthly bins in the 10-30 km height interval, for the four RO missions CHAMP, GRACE, COSMIC, and Metop. The spread of the distributions, as quantified by the median absolute deviation (MAD), is about 0.3% for CHAMP and GRACE and 0.1% for COSMIC and Metop. The right-hand panel shows the MADs for height resolved sampling-error distributions.~~

3.5 Sampling errors and sampling-error correction

The finite number of observations is not enough to fully account for all variability within the time-latitude bins, leading to a sampling error in the monthly means. The sampling error, $\varepsilon_{\mathrm{samp}}$, can be estimated by sampling a model atmosphere at the same times and locations as the observations, and then subtract the true model monthly mean from the monthly mean based on the sampled model data,

$$\varepsilon_{\mathrm{samp}} = \bar{X}_{\mathrm{model}}^{\mathrm{samp}} - \bar{X}_{\mathrm{model}}^{\mathrm{true}} \tag{4}$$

The sampled monthly mean in Eq. 4 is constructed similarly to the observed monthly mean, using the methods described in Section 3.4. The true model mean for a monthly bin is computed from the full 4-dimensional reanalysis model field,

$$\bar{X}_{\mathrm{model}}^{\mathrm{true}} = \frac{1}{n_t n_\varphi n_\lambda} \cdot \frac{1}{\sum_{k=1}^{n_\varphi} \cos\varphi_k} \cdot \sum_{t=1}^{n_t} \sum_{k=1}^{n_\varphi} \sum_{l=1}^{n_\lambda} X_{tkl} \cos\varphi_k \tag{5}$$

where φ_k is the latitude at a model grid point, and the summation loops over all model grid points located within the 5-degree latitude band for that calendar month. Similar techniques for sampling-error estimation have been described by, e.g., *Foelsche et al.* (2003, 2008), *Scherllin-Pirscher et al.* (2011), and *Ho et al.* (2009). In the ROM SAF CDR, sampling errors are estimated from ECMWF reanalysis short-term forecast fields (currently, ERA-Interim) at a $2.5° \times 2.5°$ latitude-longitude grid and a 6 hour time step.

The accuracy of the estimated sampling error, $\varepsilon_{\mathrm{samp}}$, depends on the ability of the model to describe the true atmospheric variability within the monthly bins, at the spatio-temporal resolution of the observations. This includes both synoptic-scale variability as well as various modes of cyclic variability, e.g., the atmospheric diurnal and semi-diurnal cycles. The accuracy of the mean state within the bins is less important as it is largely removed by the subtraction in Eq. 4.

The magnitude of the sampling error depend on how well the dominating modes of variability are sampled. The errors are smaller for COSMIC than for CHAMP or GRACE, mainly due to larger data numbers, but also due to a better sampling of the diurnal cycle. The left-hand panel of Fig. 4 shows the distributions of estimated sampling errors for the four RO missions, computed from all monthly bins within a core region of 10-30 km. The bending angle spread, as measured by the median absolute deviation (MAD), is about 0.3% for CHAMP and GRACE and 0.1% for COSMIC and Metop. The corresponding numbers for dry temperature are 0.3 K and 0.1 K, respectively (not shown here). However, a small number of bins have much larger sampling errors, mainly associated with wintertime mid- and high-latitude variability. The panel to the right in Fig. 4 shows the deviations for height resolved sampling error distributions. The estimated sampling errors are smallest in a region between 10 and 35 km.

The estimation of sampling errors by means of a model provides an opportunity to do a partial correction of this important class of error. Such a correction, or adjustment, can potentially reduce systematic biases between climatologies obtained from different RO missions with different sampling characteristics and reduce systematic bias changes as the global RO constellation changes with time. Sampling-error corrected means are computed by subtracting the estimated sampling error from the

observed mean

$$\bar{X}^{\mathrm{corr}} = \bar{X} - \varepsilon_{\mathrm{samp}} \tag{6}$$

The consequence of the correction is clearly seen when comparing gridded monthly means computed from disjoint sets of RO observations, e.g., monthly means computed from different RO missions during overlap periods. This is further discussed in Section 5.3 where it is shown that sampling-error correction significantly decreases inter-mission differences, leaving a residual sampling error, $\varepsilon_{\mathrm{resamp}}$, that may be handled as a quasi-random, statistical error.

3.6 Anomaly data time series

The gridded monthly mean RO data records discussed in this paper can be described as time series of variables on a two-dimensional latitude-height grid

$$\bar{X}_{ijm} = f(\varphi_i, h_j, m) \tag{7}$$

where $\bar{X}_{ijm}$ is a monthly-mean climate variable (e.g., refractivity or dry temperature). Indices i and j denote the latitude and height bins (with reference latitude, φ_i, and height, h_j, respectively) and m denotes the time (a running month number). The anomalies are defined as the deviations from a climatological seasonal cycle, $\bar{X}_{ijs}^{\mathrm{clim}}$, where $s = 1, \ldots, 12$ is the season (month of the year). Hence, the anomalies are given by

$$\bar{X}_{ijm}^{\mathrm{anom}} = \bar{X}_{ijm} - \bar{X}_{ijs}^{\mathrm{clim}} \tag{8}$$

and the fractional anomalies are given by

$$\bar{X}_{ijm}^{\mathrm{anom}} = (\bar{X}_{ijm} - \bar{X}_{ijs}^{\mathrm{clim}}) / \bar{X}_{ijs}^{\mathrm{clim}} \tag{9}$$

where the latter is the preferred expression for variables that have a dominating exponential altitude dependence.

The mean seasonal cycle, i.e. the long-term mean state as a function of latitude, height, and season, is constructed from RO data itself, although it may be based on a different combination of RO missions:

$$\bar{X}_{ijs}^{\mathrm{clim}} = \frac{1}{N_{\mathrm{yr}}} \sum_{k=1}^{N_{\mathrm{yr}}} \bar{X}_{ijm}' \quad, \quad m = 12 \cdot (k-1) + s \tag{10}$$

where N_{yr} is the number of years used in the generation of the climatology. In the generation of anomaly time series, the same seasonal cycle should be used for all missions and throughout the time series. This is particularly important for investigations of differences between RO missions.

The anomalies depend on latitude, altitude, and time. Averaging over a latitude band, properly weighting the latitude bins proportional to their areas, gives a two-dimensional time-altitude data set (see Fig. 7), while averaging over both a latitude band and a height layer gives a one-dimensional time series (see Figs. 9 and 10).
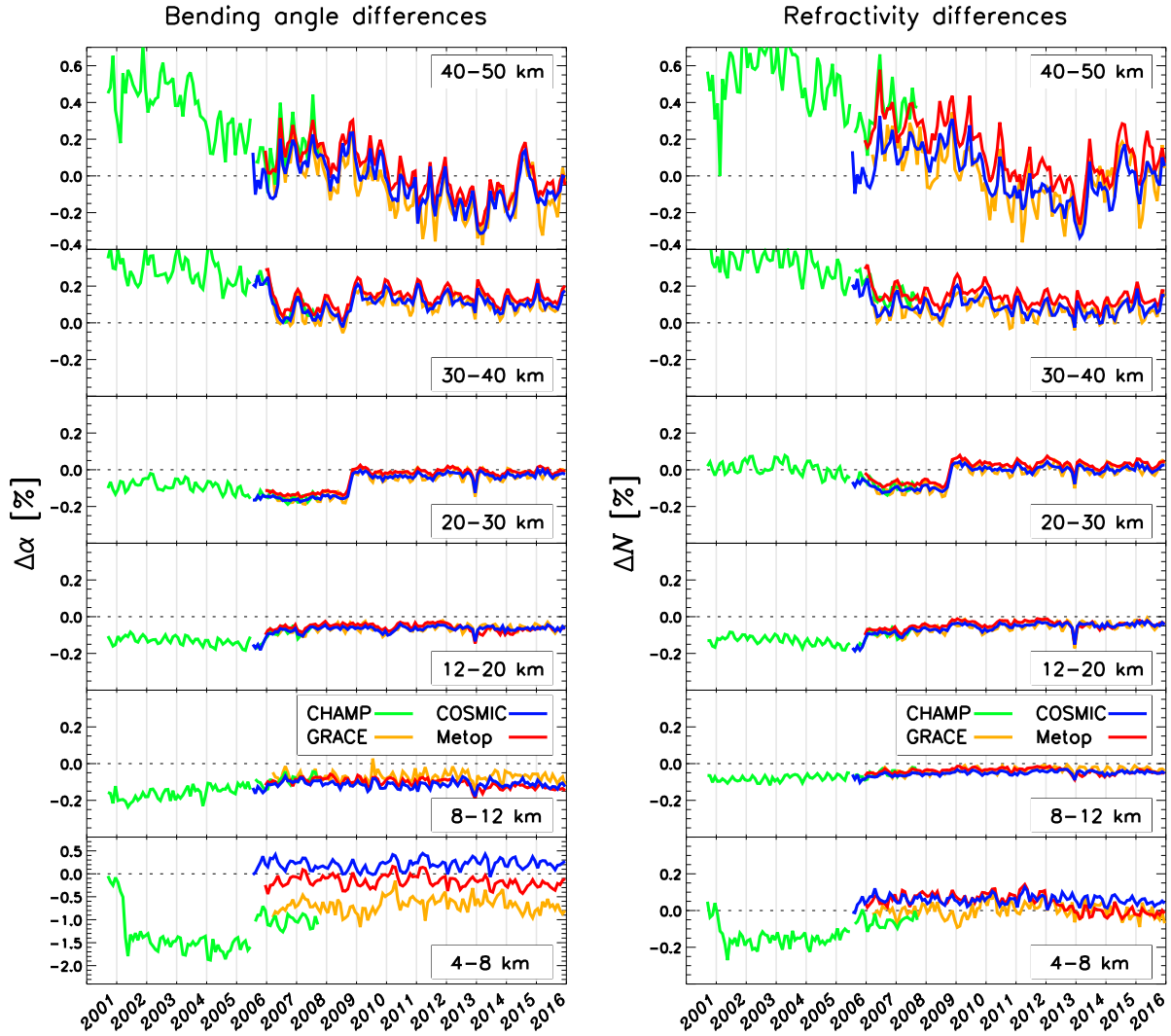
Figure 5. Globally averaged monthly mean bending angle (left panels) and refractivity (right panels) differences with respect to ERA-Interim, for the four RO missions included in the ROM SAF CDR. From bottom to top, the panels show vertically averaged differences in the 4-8 km, 8-12 km, 12-20 km, 20-30 km, 30-40 km, and 40-50 km intervals, where the height coordinate is impact altitude for bending angle and altitude for refractivity. Note that for the lowest bending angle panel (left column), the vertical axis has been compressed to accommodate the larger differences between the missions.

4 Comparison with ERA Interim reanalyses

The ROM SAF CDR is evaluated using the ERA-Interim reanalysis as a reference, with the purpose to provide a better understanding of the time evolution of the RO data and the stability in time. As a side-effect, time-varying biases and sudden bias
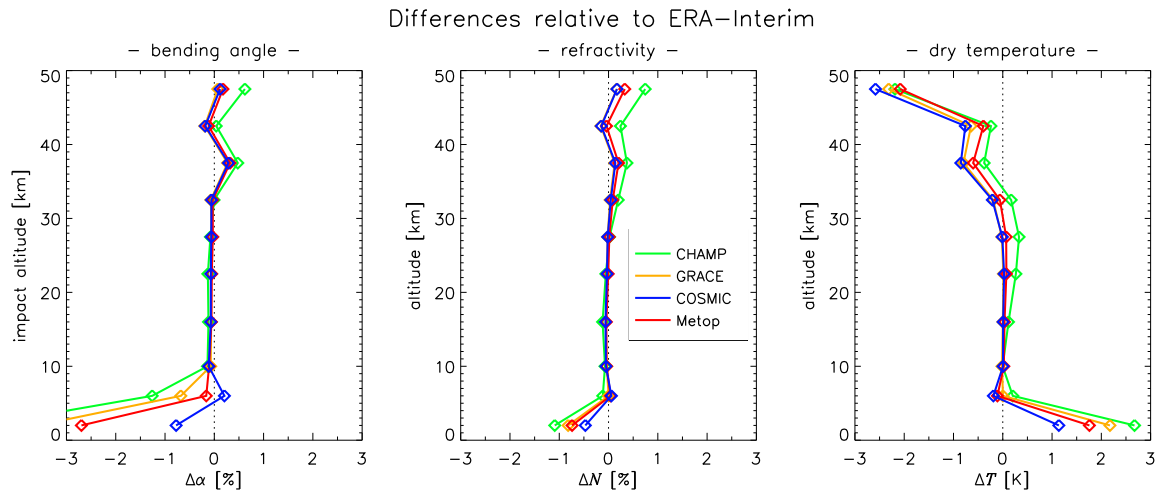
Figure 6. Globally averaged monthly mean bending angle (left panel), refractivity (middle panel), and dry temperature (right panel) differences relative to ERA-Interim, for the four RO missions included in the ROM SAF CDR. The profiles show time averages, with the averaging period chosen as the full length of the respective satellite mission.

shifts in the ERA-Interim reanalysis data are identified. In addition, the comparison of all four RO missions against the same reference provides an indication of the consistency of RO climatologies generated from different missions.

Fig. 5 shows globally averaged relative differences of observed bending angle and refractivity with respect to ERA-Interim, for the four RO missions and for six altitude layers from 4 to 50 kilometers. The smallest differences are found at altitudes
5　between about 8 and 30 km. In this altitude region, the dominating features in the difference time series reflect bias shifts in the ERA-Interim data itself. In December 2006 the magnitude of the bias suddenly decreased in the 12-20 km and 30-40 km intervals, even though it remained slightly negative in the 12-20 km range throughout the time period. In November 2009 there was a shift from a small negative bias to a near zero difference in the 20-30 km interval. Then in late 2013 the difference suddenly dropped back to a small negative bias level during a few weeks. These events are most likely related to the start of
10　assimilation of COSMIC data in 2006 ~~(Dee et al. (2011))~~(Dee et al., 2011), an update of the COSMIC ~~NRT~~ data processing in October 2009 ~~(Healy (2013))~~(Healy, 2013), and a temporary drop-out of RO data from the ERA-Interim assimilation system in late 2013 (*S. Healy, pers. comm.*). Above 30 km, the differences between RO and ERA-Interim are larger, particularly for the earlier pre-COSMIC time period, when the impact of RO on the reanalysis was weaker due to lower data numbers.

In the 8 to 40 km altitude interval, the spread amongst the RO missions is generally smaller than the differences between
15　RO and ERA-Interim. In combination with the fact that the dominating shifts in the difference time series can be attributed to ERA-Interim, this suggests that the RO data have better long-term stability than the ERA-Interim data. At the highest altitudes, above 40 km, the larger differences in the earlier time period and the smaller differences later on, lead to a long-term trend in the differences. Below 8 km, the spread amongst the RO missions are larger, with the COSMIC and Metop data showing a better match to ERA-Interim than the CHAMP and GRACE data. The CHAMP data series exhibits bias shifts in March 2002

and in July 2006. The former shift coincides with a firmware update of the GPS-RO instrument onboard the CHAMP satellite (*J. Wickert, pers. comm.*).

Fig. 6 shows the RO versus ERA-Interim differences in the form of global time-averaged vertical profiles for bending angle, refractivity, and dry temperature. The profiles for the different missions do not represent identical time periods, as the time averaging is done over the full length of the respective satellite mission. In line with the findings in Fig. 5, the CHAMP profiles ~~are somewhat deviating~~ deviate somewhat from the profiles of the other satellite missions, particularly above 35 km for bending angle and refractivity and also at lower altitudes for dry temperature. Of the four RO missions, CHAMP also exhibits the largest lower-tropospheric biases relative ERA-Interim.

Regarding the stability in time, it should be noted that even though the ERA-Interim reanalysis system in itself does not change with time, the evolving global observing system ~~lead~~ leads to time-varying biases (*Dee et al.*, 2011). ERA-Interim does not provide a stable enough reference against which to accurately measure temporal stability of the RO data. Between about 8 and 30 km, the RO data records are likely to have a higher temporal stability than ERA-Interim. At higher and lower altitudes, the long-term temporal stability of the multi-mission RO time series is limited by the evolving global RO constellation, and depends on the magnitude and character of the differences between the RO satellite missions. This is discussed Section 5.

5 Differences between RO missions

Differences in the monthly means obtained from RO missions that overlap in time are due to a combination of random profile errors, sampling errors, and systematic errors of instrumental or data-processing origin. While random errors contribute to a general degradation of the quality of the climatologies, they do not prevent us from combining data from different missions. Systematic errors on the other hand can, potentially, introduce time-evolving biases in combined multi-mission data records. In this section some of the RO mission differences, detected from mission overlaps, are identified. The influence of these differences on time series of bending angle and dry-temperature anomalies is assessed. The sampling-error correction method, described in Section 3.5, is also evaluated and its efficiency in reducing differences between the RO missions is investigated.

5.1 Time-altitude bending angle plots

Fig. 7 shows global monthly mean bending angle anomalies for the four RO missions. The four data records are structurally very similar above about ~~6-8~~ 8 km and below about 35-40 km impact altitude. This altitude span encompasses the core region from the middle troposphere to the lower/middle stratosphere where RO measurements are known to have the highest quality (e.g., *Kuo et al.* (2005), *Scherllin-Pirscher et al.* (2011)). In the lowest few kilometers, there are known biases in bending angle observations obtained in moist low-latitude regions, leading to substantial differences between the missions. Throughout the stratosphere, Metop and COSMIC are qualitatively very similar, while GRACE and CHAMP exhibit quasi-random, noise-like, structures above about 35 km.

Fig. 8 shows bending angle differences between GRACE and COSMIC (left column), between Metop and COSMIC with Metop based on input data from UCAR (middle column), and between Metop and COSMIC with Metop based on input data
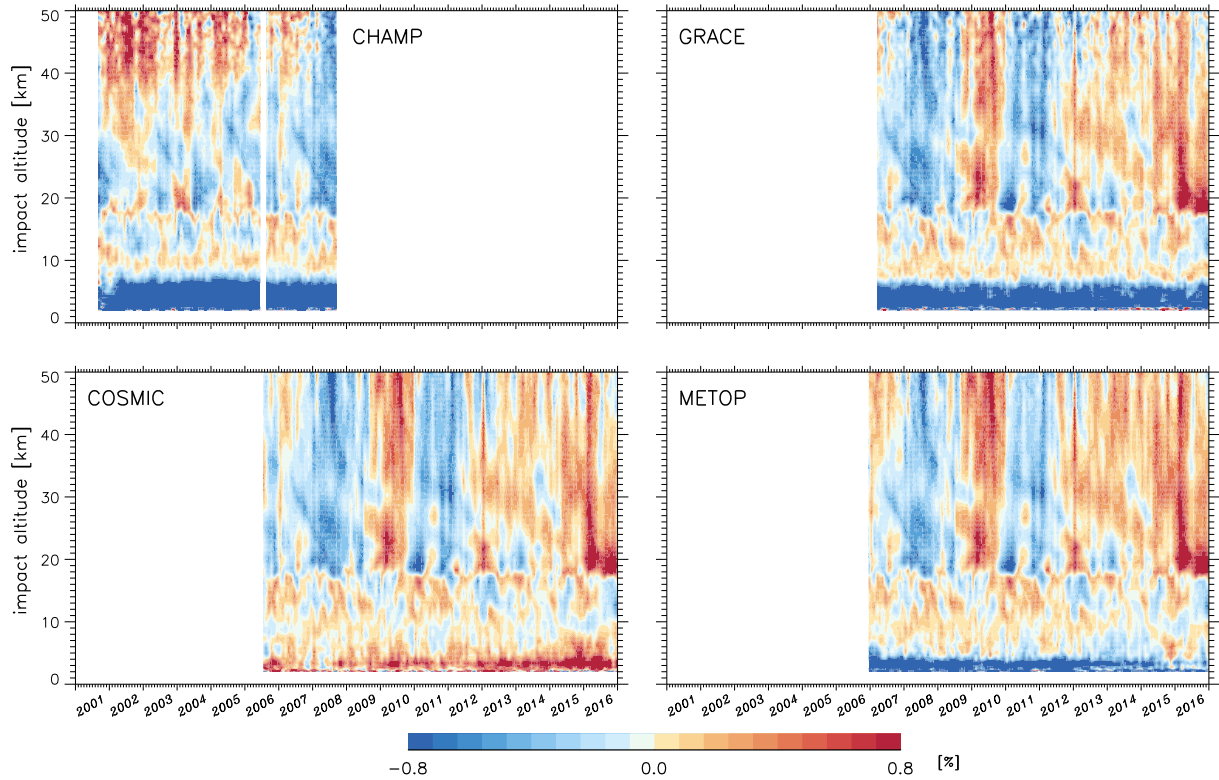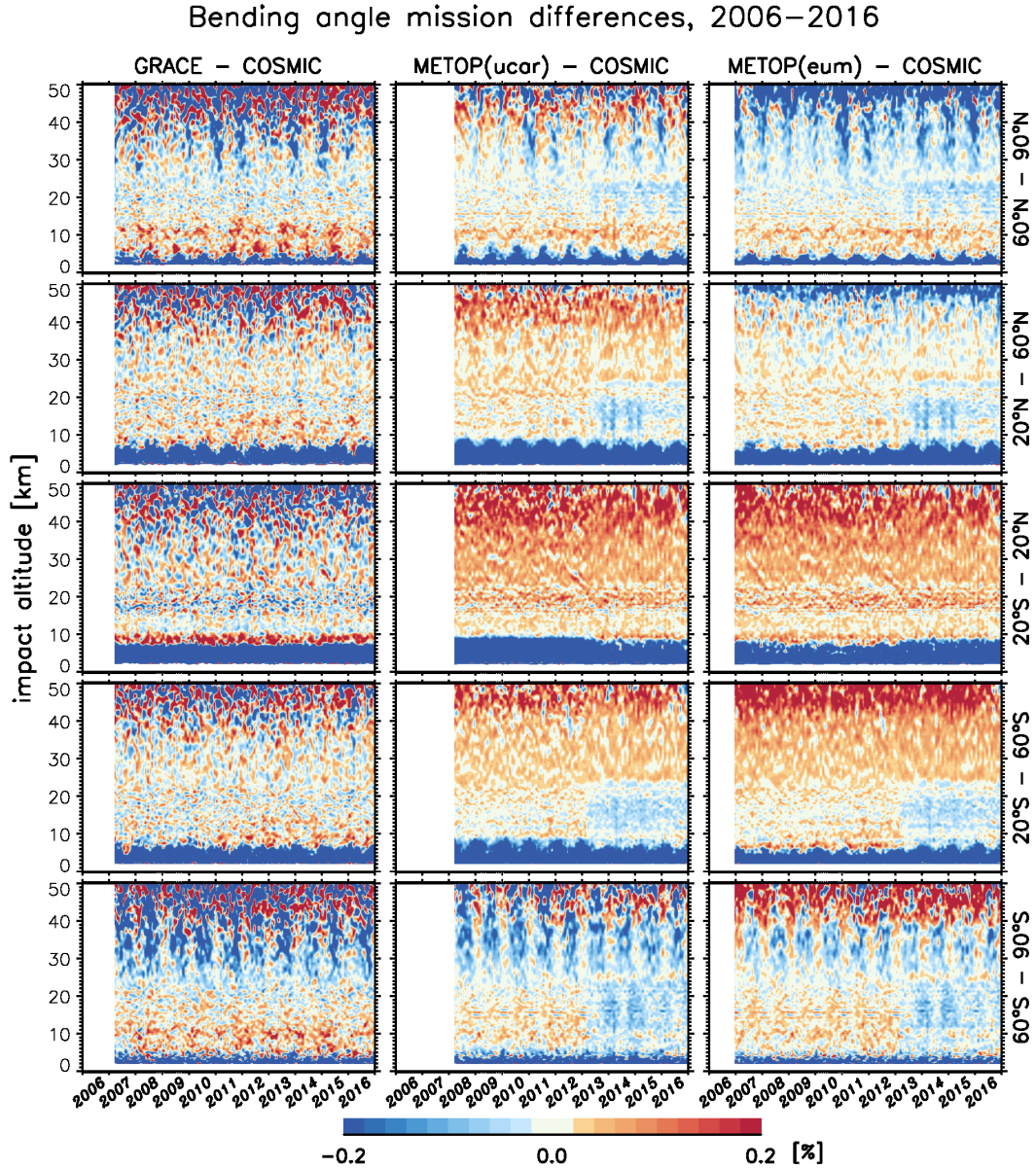
Figure 7. Global monthly mean bending angle anomalies, for the four RO missions included in the ROM SAF Climate Data Record v1.0. There is overlap between CHAMP and COSMIC from August 2006 to September 2008, between GRACE and COSMIC from March 2007 to December 2016, and between Metop and COSMIC from December 2006 to December 2016. Data are given on a 5 degrees by 200 meter latitude-altitude grid and have been sampling-error corrected. The anomalies for the four satellite missions are computed based on the same reference climatology, obtained from RO data itself.

from EUMETSAT (right column). COSMIC is chosen as comparison reference because it provides the longest record of the four missions, and because it has a good local-time coverage. Above ~~6-8~~ 8 km impact altitude, the differences between the RO missions are small (note that the bending angle color range in the plots only spans ±0.2%). A large fraction of the variability in the difference plots consists of a quasi-random, noise-like pattern, with a broad minimum between 10 and 25 km altitude. This pattern is most evident in the GRACE-COSMIC plots (Fig. 8, left column). The quasi-random pattern is also present in the Metop-COSMIC plots (Fig. 8, middle and right columns), but is less visible as it is superposed on an almost uniform positive bias level (red colors) at low- and mid-latitudes.

The difference plots in Fig. 8 reveal a range of systematic differences between RO missions that cannot be explained by random profile errors or by quasi-random sampling effects. We identify the following systematic bending angle biases:

Figure 8. Differences between monthly-mean bending angle climatologies retrieved from different missions; between GRACE and COS-MIC (left column), between Metop(ucar) and COSMIC (middle column), and between Metop(eum) and COSMIC (right column), where Metop(ucar) is processed by ROM SAF based on input data from UCAR and Metop(eum) is based on input data from EUMETSAT. Differences are shown in five latitude bands, south to north from lowest panel to top panel.

- Biases in the lower troposphere up to a few percent (out of scale). The biases are stronger and have a larger vertical extent at low latitudes, and are believed to be linked to signal tracking issues in moist regions of the atmosphere (e.g., *Sokolovskiy et al.*, 2010).

16
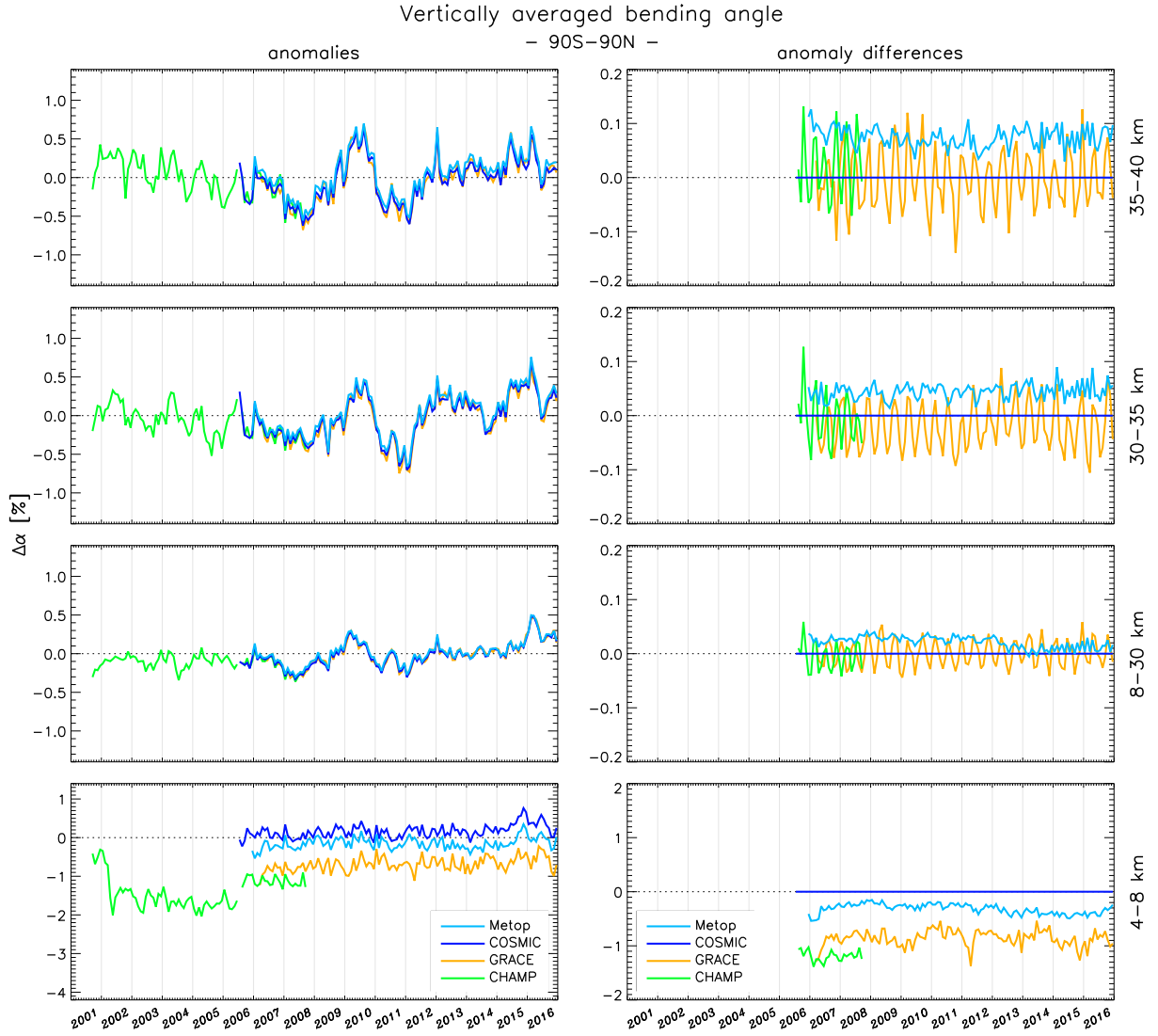
Figure 9. Monthly mean bending angle anomalies for the four RO missions (left panels) and differences of CHAMP, GRACE, and Metop with respect to COSMIC (right panels). From bottom to top, the panels show the 4-8 km, 8-30 km, 30-35 km, and 35-40 km height layers. Note that for the lowest plots, the vertical axes have been compressed to accommodate the larger differences between the missions.

- Seasonally varying biases up to 0.1-0.2% at high altitudes (>30 km) and high latitudes (>60 degrees latitude).

- Bias shifts on the order of 0.1% below about 20 km in the Metop-COSMIC differences in 2013. These shifts are related to firmware upgrades of the RO instruments onboard Metop-B (in April 2013) and Metop-A (July 2013). The tracking of the GPS signals for the Metop rising occultations was changed, which had the effect that a minor bias due to the L2
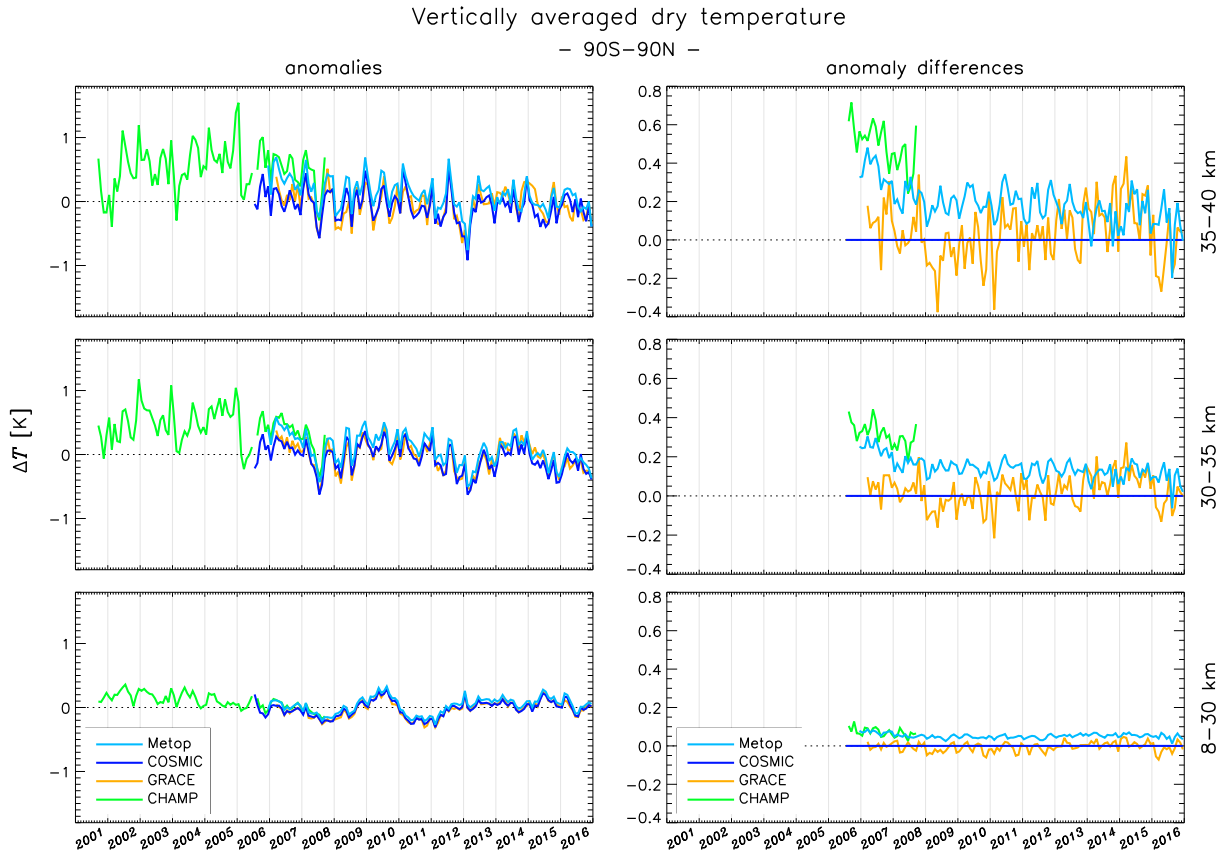
Figure 10. Monthly mean dry temperature anomalies for the four RO missions (left panels) and differences of CHAMP, GRACE, and Metop with respect to COSMIC (right panels). From bottom to top, the panels show the 8-30 km, 30-35 km, and 35-40 km height layers.

extrapolation in the ROM SAF processing suddenly appeared (described in the series of ROM SAF validation reports at *http://www.romsaf.org/product_documents.php*).

- Large-scale hemispherically asymmetric (north-south) Metop-COSMIC bias on the order of 0.1% above 35-40 km, and increasing upward. Only seen in the plots with Metop data based on input from EUMETSAT (Fig. 8, rightmost column). This difference is believed to be related to differences in LEO satellite orbits from the two sources of input data.

- Relatively uniform ~~bias in~~ Metop-COSMIC ~~differences~~ difference at low- and mid-latitudes on the order of 0.03% at 20 km and increasing upward (0.1% at 40 km). ~~Believed~~ This difference is believed to be related to under-sampling of the diurnal cycle, in combination with imperfect sampling-error correction of the Metop data.

- GRACE-COSMIC and CHAMP-COSMIC cyclic differences (the latter not shown here) at low- and mid-latitudes on the order of 0.03% at 20 km and increasing upward. This is a weak effect and is just barely seen in Fig. 8 (vertical averaging

makes this effect more easily detected, see Fig. 9). The cycle period is around 5 months for GRACE-COSMIC and 4 months for CHAMP-COSMIC. ~~Believed~~ These differences are believed to be related to under-sampling of the diurnal cycle, in combination with imperfect sampling-error correction.

Most of these RO mission differences are caused by systematic errors in the underlying profile data, that are propagated to the gridded monthly means. The exception is the sampling errors that are intrinsic to the gridded data.

5.2 Anomaly time series

~~Sofar, the~~ The RO mission differences have so far been described in terms of bending angles. However, the identified differences are also relevant for the geophysical variables retrieved from bending angle, e.g., refractivity and temperature. Generally, errors in bending angle propagate downward to lower altitudes in the retrieval chain. This becomes evident in the anomaly time series discussed below.

5.2.1 Bending angle

The left column of Fig. 9 shows globally averaged monthly-mean anomalies of bending angle, vertically averaged in four height layers: 4-8 km, 8-30 km, 30-35 km, and 35-40 km. Each plot includes data for the four RO missions. The corresponding panels in the left column of Fig. 9 show the differences of CHAMP, GRACE, and Metop with respect to COSMIC. During the years 2008 to 2014, the six-satellite COSMIC constellation also had a nearly complete local-time coverage, which is important for detecting the impact of under-sampling the diurnal cycle in the other missions.

In 8-30 km vertically averaged data (Fig. 9, third row from top), the four time series show a very close match. How well the overlapping time series match must be evaluated in relation to the variability of the time series itself. There is variability on a broad range of time scales, from short-range intra-seasonal variations to inter-annual and decadal variability, and long-term climatological trends. We find that for the 8-30 km time series, the mean (time averaged) differences between the missions are -0.005%, 0.001%, and 0.02%, respectively, for CHAMP, GRACE, and Metop relative to COSMIC. This is much smaller than the intrinsic variability of the time series – the total range of global monthly mean bending angle anomalies in this height range is about 1%. Metop shows a relatively steady bias relative to COSMIC, with a stepwise decrease of the bias in mid-2013, and with a tendency to oscillations after 2014. The CHAMP and GRACE differences with respect to COSMIC exhibit strong oscillating behaviour with peaks that reach about the same magnitude as the Metop-COSMIC bias. The cycle periods for the oscillating difference time series are about 4 months and 5 months, respectively, closely corresponding to the precession rate of the respective satellite orbit. This could be explained as a consequence of the sampling-error correction not being able to fully compensate for the effects of under-sampling the diurnal and semi-diurnal cycles. It would also be consistent with the near-constant Metop-COSMIC biases because of the Sun-synchronous Metop orbit.

Above the middle troposphere, the mean differences in Fig. 9 increase with altitude, and for Metop-COSMIC the differences are about 0.04% in the 30-35 km interval and 0.08% in the 35-40 km interval. The mean GRACE-COSMIC differences are small, while the mean CHAMP-COSMIC differences increase to 0.02% in the 35-40 km interval. The magnitude of the
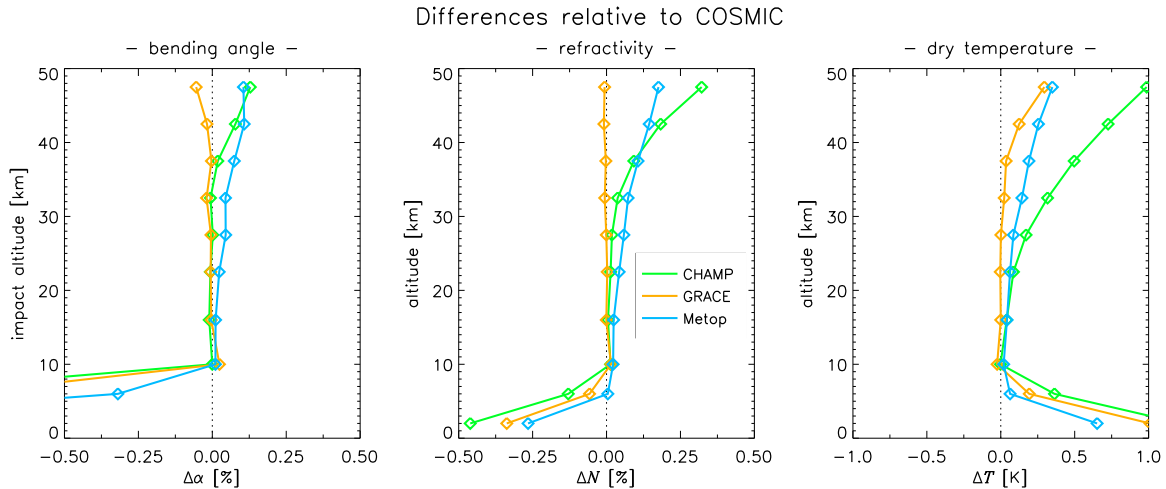
Figure 11. Summary of global CHAMP, GRACE, and Metop differences relative to COSMIC, quantified by the time averages of the mission differences computed over the respective overlap period. The RO data have been binned, averaged, sampling-error corrected, and converted to anomalies according to the methods described in Section 3.

CHAMP-COSMIC and GRACE-COSMIC oscillations increases with altitude, such that the peak biases of these two mission differences reach about the same values as the Metop-COSMIC biases.

In the tropospheric 4-8 km impact altitude interval (Fig. 9, left column), we find relatively large biases between the missions. Monthly global averages of CHAMP and GRACE bending angles are about 1% smaller than the corresponding COSMIC data, while Metop bending angles are about 0.3% smaller. It should also be noted that the CHAMP data record itself shows substantial bias shifts in 2002 and 2006.

The left panel of Fig. 11 summarizes the mission differences for global mean bending angle anomalies. The summary is based on the time-averaged differences computed over the respective overlap period. The RO mission consistency, defined as the largest time-averaged difference between any two missions, is about 0.04% above 8 km and below 30 km, increasing to 0.08% below 40 km, and about 0.18% below 50 km. These numbers can be up to a factor of two larger for 30-degree latitude bands compared to global means (Table 2).

5.2.2 Refractivity

Similarly to the bending angle anomalies, the global refractivity anomalies for the four RO missions show a very close match in the 8-30 km vertically averaged data (not shown here). The differences of the refractivity anomalies for Metop relative to COSMIC are nearly constant (about 0.04%) over the 10 years of overlap, with just a small decrease of the differences after mid-2013, and with a tendency to short time-scale oscillations towards the end of the time period. The differences of CHAMP and GRACE relative to COSMIC show similar, steady oscillation patterns as the corresponding bending angle time series.

Median absolute deviations of GRACE−COSMIC differences
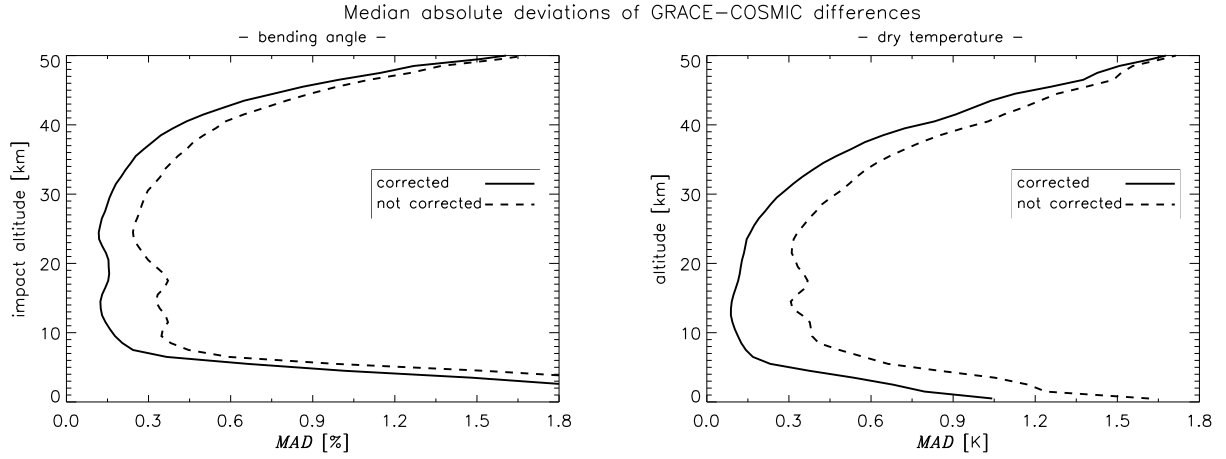
Figure 12. Median absolute deviations of GRACE-COSMIC differences computed from all monthly bins in 1-km height intervals during the mission overlap period March 2007 to December 2016; deviations for bending angles to the left and for dry temperature to the right. The solid lines show deviations with sampling-error correction applied, and the dashed lines show deviations without correction.

Vertically averaged global refractivity anomalies for 5-kilometer layers from 30 to 50 km show a systematic increase of the mission differences with height (not shown here). This is similar to the bending angles, although the differences are larger for refractivity due to the downward propagation of errors in the Abel transform (Section 3.1).

The middle panel of Fig. 11 summarizes the mission differences for global mean refractivity anomalies. The consistency is about 0.05% between 8 km and 30 km, increasing to 0.11% below 40 km, and about 0.32% below 50 km. These numbers can be up to a factor of two larger for 30-degree latitude bands compared to global means (Table 2).

Table 2. ~~Summary of global CHAMP, GRACE, and Metop differences relative to COSMIC, quantified by the time averages of the mission differences computed over the respective overlap period. The~~ RO ~~data have been binned, averaged, sampling-error corrected,~~ mission consistency above 8 km and ~~converted to anomalies according to~~ below 30 km. Consistency is here defined as the ~~methods described in Section 3.~~ largest time-averaged difference between any two RO satellite missions.

latitude range	bending angle	refractivity	dry temperature
globally	0.04 %	0.05 %	0.15 K
60°N–90°N	0.01 %	0.03 %	0.25 K
20°N–60°N	0.02 %	0.02 %	0.25 K
20°S–20°N	0.08 %	0.09 %	0.30 K
20°S–60°S	0.03 %	0.04 %	0.15 K
60°S–90°S	0.02 %	0.04 %	0.20 K

5.2.3 Dry temperature

The globally averaged dry-temperature anomalies are shown in Fig. 10. The RO differences in the 8-30 km vertical averages are smaller than 0.10 K. We note that the oscillating behaviour seen in the bending angle and refractivity differences for CHAMP and GRACE relative to COSMIC is there also for dry temperature, but is less obvious as it has a much more irregular appearance. At higher altitudes, the dry-temperature anomalies in 5-kilometer layers from 30 to 50 km show increasingly larger differences between the RO missions. In addition to the errors propagated from bending angle to refractivity, there is also a downward propagation of errors due to the hydrostatic integration used in the retrieval of dry temperature.

The right-hand panel of Fig. 11 summarizes the mission differences for global mean dry temperature anomalies. The RO mission consistency is about 0.15 K between 8 km and 30 km, increasing to 0.30 K up to 40 km , and 0.50 K up to 40 km. These numbers can be up to a factor of two larger for 30-degree latitude bands compared to global means (Table 2).

Median absolute deviations of GRACE-COSMIC differences computed from all monthly bins in 1-km height intervals during the mission overlap period March 2007 to December 2016; deviations for bending angles to the left and for dry temperature to the right. The solid lines show deviations with sampling-error correction applied, and the dashed lines show deviations without correction.

5.3 Evaluation of the sampling-error correction

Mission differences during overlap periods allow us to investigate some of the consequences of sampling-error correction, and to assess the magnitude of the residual sampling errors remaining after correction. Fig. 12 shows the median absolute deviations of GRACE-COSMIC differences based on all monthly bins in 1-km height intervals during the mission overlap period March 2007 to December 2016. The solid lines are computed from GRACE and COSMIC data with sampling-error correction applied, while the dashed lines are computed from data without correction. The application of sampling-error correction substantially reduces the GRACE-COSMIC differences, both for bending angle and dry temperature, as well as for other geophysical variables (not shown).

The deviations remaining after sampling-error correction (indicated by the solid lines in Fig. 12) are due to a combination of GRACE and COSMIC random profile errors, residual sampling errors, and any systematic differences between the RO missions that have a sufficiently strong variation with time and/or latitude. In the core region 8-30 km, random profile errors can at most explain a part of the 0.1-0.2% deviations for the bending angles and the 0.10-0.15 K deviations for dry temperature. Assuming, conservatively, that these remaining errors are due solely to residual sampling errors, we find that around one third of the original sampling error remain after sampling error correction. This is roughly in line with the findings of *Scherllin-Pirscher et al.* (2011).

Fig. 13 shows an example of the consequences of sampling-error correction for the mission differences. The left-hand panel of Fig. 13 shows mission difference anomaly time series without correction, and the right-hand panel show the same differences with the correction applied. The CHAMP and GRACE differences relative to COSMIC are dominated by a periodic oscillation, presumably due to aliasing between the LEO satellite orbital precession and diurnal or semi-diurnal cycles in the atmosphere.
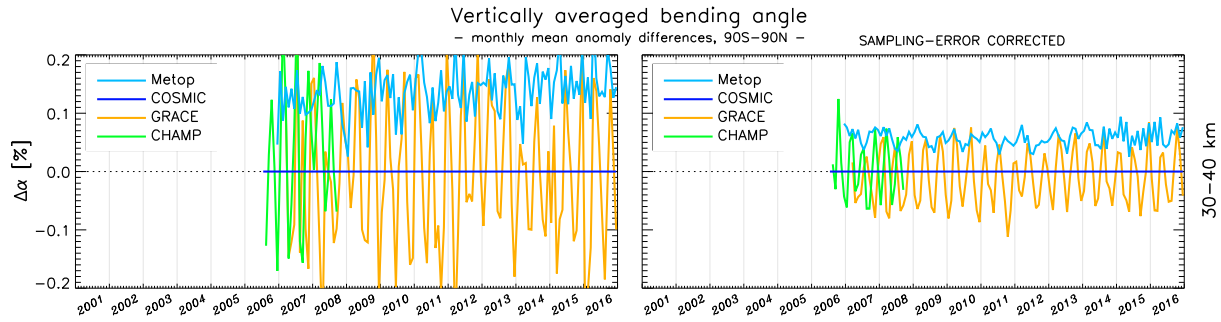
Figure 13. Differences between monthly mean CHAMP, GRACE, and Metop refractivity anomalies relative to COSMIC, globally averaged in the 30-40 km altitude layer. Left-hand panel without sampling-error correction and to the right with sampling-error correction applied.

The magnitude of these oscillations are substantially reduced by the sampling-error correction. However, the fact that they are not entirely removed indicates that the estimated sampling errors, based on sampling ERA-Interim reanalysis fields, are not able to fully capture the diurnal and/or semi-diurnal cycles. Unlike the case with CHAMP and GRACE, the differences between Metop and COSMIC are not periodic. Given the Sun-synchronous orbit of Metop, the errors resulting from under-sampling of the diurnal cycle are expected to be near-constant in time, which is roughly in line with the findings.

6   Summary and conclusions

In this study, we present results from an evaluation of the 15-year ROM SAF CDR v1.0 consisting of separate data records from four different RO satellite missions: CHAMP, GRACE, COSMIC, and Metop. The processing of the RO data to gridded monthly means is described, including the sampling-error correction of the monthly mean data. The observed bending angle, refractivity, and dry-temperature records are compared to the ERA-Interim short-term forecasts. The four RO data records are also inter-compared during mission overlap periods and the impact of the sampling-error correction, applied to the gridded monthly mean data, is evaluated.

In general, there is good overall agreement between the ROM SAF gridded monthly mean CDR and the ERA-Interim reanalysis, particularly in the 8-30 km height interval. Here, the differences appear to mainly reflect time-varying biases in ERA-Interim, as indicated from the timing of the bias shifts and the fact that the spread amongst the RO missions is smaller than the differences between RO and ERA-Interim. We interpret this as a better temporal stability in the RO data records than in the ERA-Interim time series. At high altitudes, above 30-40 km, we find larger differences between RO and reanalysis, and also a long-term trend in the difference time series. At altitudes below 8 km the differences are again larger, particularly for bending angle, with a relatively large spread amongst the RO missions.

To fully exploit the RO data records scientifically requires that we can combine the data records from several RO missions into multi-mission data records. There is an expectation that this can be done without any adjustments or inter-calibrations. However, any differences between the missions in the retrieved geophysical data may lead to time-varying biases in the multi-

mission data record as new satellite missions replace older ones. We investigated the presence of such differences during mission overlap periods and found that there is a high degree of consistency between the RO satellite missions in the 8-30 km altitude region. The remaining differences in this altitude interval are predominantly oscillatory or highly variable for CHAMP and GRACE relative to COSMIC, while for METOP the differences relative to COSMIC largely consist of small, but stable, offsets. These differences should be considered in the generation of multi-mission data records. At higher altitudes the differences between the RO missions become increasingly larger, and at altitudes below 8 km we find biases and bias shifts that substantially reduces the inter-mission consistency.

The cause of the inter-mission biases can in many cases be identified from difference plots during the mission overlap periods. In this study, we have identified the most dominating bending angle biases that are propagated from the input data or from the geophysical profile data to the gridded monthly means: lower-tropospheric biases linked to moist regions of the atmosphere, seasonally varying biases at high altitudes and high latitudes, Metop-COSMIC bias shifts related to firmware upgrades, and a high-altitude hemispherically asymmetric bias related to small differences between the UCAR and EUMETSAT low-level input data. We also find systematic residual sampling errors that appear to be caused by the under-sampling of diurnal or semi-diurnal cycles not being fully corrected for by the sampling-error correction method.

The results presented here also affect the other geophysical variables retrieved from RO measurements, which are not explicitly discussed in the present study: dry pressure, dry geopotential heights, temperature, and humidity. For the latter two variables, obtained through a 1D-Var retrieval using additional information from a model background (see Section 3.1), the relatively large inter-mission biases in the lower troposphere will have an impact on the temperature and humidity data records, which was not investigated here.

This study shows that above the lower troposphere and below about 30 km, data records from different RO satellite missions exhibit only small systematic differences. Further reduction of these differences most likely requires an improved sampling-error correction. Reducing the inter-mission differences at higher altitudes also requires reduced impacts from subtle differences in the input data, and from the statistical optimization of the bending angles, as well as an understanding of the cause of the high-altitude, high-latitude seasonally varying differences. A continued reduction of the relatively small, but systematic, inter-mission biases, is important for the generation of long-term stable, homogeneous RO-based CDRs extending to higher altitudes.

References

Angerer, B., Ladstädter, F., Scherllin-Pirscher, B., Schwärz, M., Steiner, A.K., Foelsche, U., and Kirchengast, G.: Quality aspects of the Wegener Center multi-satellite GPS radio occultation record OPSv5.6, *Atmos. Meas. Tech.*, *10*, 4845–4863, doi:10.5194/amt-10-4845-2017, 2017.

5 Anthes, R.A., Bernhardt, P.A., Chen, Y., Cucurull, L., Dymond, K.F., Ector, D., Healy, S.B., Ho, S.-P., Hunt, D.C., Kuo, Y.-H., Liu, H., Manning, K., McCormick, C., Meehan, T.K., Randel, W.J., Rocken, C., Schreiner, W.S., Sokolovskiy, S.V., Syndergaard, S., Thompson, D.C., Trenberth, K.E., Wee, T.-K., Yen, N.L., and Zeng, Z.: The COSMIC/FORMOSAT-3 mission: Early results, *Bull. Am. Meteorol. Soc.*, *89*, 313–333, doi:10.1175/BAMS-89-3-313, 2008.

Anthes, R.A.: Exploring Earth's atmosphere with radio occultation: contributions to weather, climate and space weather, *Atmos. Meas. Tech.*, 10 *4*, 1077–1103, doi:10.5194/amt-4-1077-2011, 2011.

Ao, C.O., Jiang, J.H., Mannucci, A.J., Su, H., Verkhoglyadova, O., Zhai, C., Cole, J., Donner, L., Iversen, T., Morcrette, C., Rotstayn, L., Watanabe, M., and Yukimoto, S.: Evaluation of CMIP5 upper troposphere and lower stratosphere geopotential height with GPS radio occultation observations, *J. Geophys. Res.*, *NN*, xxxx–yyyy, doi:10.1002/2014JD022239, 2015.

Beyerle, G., Schmidt, T., Michalak, G., Heise, S., Wickert, J., and Reigber, C.: GPS radio occultation with GRACE: Atmospheric profiling 15 utilizing the zero difference technique, *Geophys. Res. Lett.*, *32*, L13806, doi:10.1029/2005GL023109, 2005.

Cardinali, C., and Healy, S.: Impact of GPS radio occultation measurements in the ECMWF system using adjoint-based diagnostics, *Q. J. R. Meteorol. Soc.*, *140*, 2315-2320, doi:10.1002/qj.2300, 2014.

Danzer, J., Foelsche, U., Scherllin-Pirscher, B., and Schwärz, M.: Influence of changes in humidity on dry temperature in GPS RO climatologies, *Atmos. Meas. Tech.*, *7*, 2883–2896, doi:10.5194/amt-7-2883-2014, 2014.

20 Dee, D.P., Uppala, S.M., Simmons, A.J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M.A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A.C.M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A.J., Haimberger, L., Healy, S.B., Hersbach, H., Hólm, E.V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A.P., Monge-Sanz, B.M., Morcrette, J.J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N., Vitart, F.: The ERA-Interim reanalysis: configuration and performance of the data assimilation system, *Q. J. R. Meteorol. Soc.*, *137*, 553-597, 2011.

25 Fjeldbo, G., Kliore, A.J., and Eshleman, V.R.: The neutral atmosphere of Venus as studied with the Mariner V radio occultation experiments, *Astron. J.*, *76*, 123–140, 1971.

Foelsche, U., Kirchengast, G., and Steiner, A.K.: Global climate monitoring based on CHAMP/GPS radio occultation data, in: First CHAMP Mission Results for Gravity, Magnetic and Atmospheric studies, edited by: Reigber, C., Luhr, H., and Schwintzer, P., Springer, 397–407, 2003.

30 Foelsche, U., Borsche, M., Steiner, A.K., Gobiet, A., Pirscher, B., Kirchengast, G., Wickert. J., and Schmidt, T.: Observing upper troposphere-lower stratosphere climate with radio occultation data from the CHAMP satellite, *Clim. Dynam.*, *31*, 49–65, doi:10.1007/s00382-007-0337-7, 2008.

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*, 2007–2018, doi:10.5194/amt-4-2007-2011, 2011.

35 Gorbunov, M.E.: Ionospheric correction and statistical optimization of radio occultation data, *Radio Sci.*, *37*, 1084, doi:10.1029/2000RS002370, 2002.

Gorbunov, M.E., and Lauritsen, K.B.: Analysis of wave fields by Fourier integral operators and their application for radio occultations, *Radio Sci.*, *39*, 4010, doi:10.1029/2003RS002971, 2004.

Healy, S.B., and Eyre, J.R.: Retrieving temperature, water vapor and surface pressure information from refractive–index profiles derived by radio occultation: A simulation study, *Q. J. R. Meteorol. Soc.*, *126*, 1661-1683, 2000.

5   Healy, S.B., Jupp, A.M., and Marquardt, C.: Forecast impact experiment with GPS radio occultation measurements, *Geophys. Res. Lett.*, *32*, doi:10.1029/2004GL020806, 2005.

Healy, S.B., and Thépaut, J.-N.: Assimilation experiments with CHAMP GPS radio occultation measurements, *Q. J. R. Meteorol. Soc., 132,* 605-623, 2006.

Healy, S.B.: Surface pressure information retrieved from GPS radio occultation measurements, *Q. J. R. Meteorol. Soc.*, *139*, 2108-2118,
10   2013.

Ho, S.-P., Kirchengast, G., Leroy, S., Wickert, J., Mannucci, A.J., Steiner, A.K., Hunt, D., Schreiner, W., Sokolovskiy, S., Ao, C., Borsche, M., von Engeln, A., Foelsche, U., Heise, S., Iijima, B., Kuo, Y.-H., Kursinski, R., Pirscher, B., Ringer, M., Rocken, C., and Schmidt, T.: Estimating the uncertainty of using GPS radio occultation data for climate monitoring: Intercomparison of CHAMP refractivity climate records for 2002 to 2006 from different data centers, *J. Geophys. Res.*, *114*, D23107, doi:10.1029/2009JD011969, 2009.

15   Ho, S.-P., Hunt, D., Steiner, A.K., Mannucci, A.J., Kirchengast, G., Gleisner, H., Heise, S., von Engeln, A., Marquardt, C., Sokolovskiy, S., Schreiner, W., Scherllin-Pirscher, B., Ao, C., Wickert, J., Syndergaard, S., Lauritsen, K.B., Leroy, S., Kursinski, E.R., Kuo, Y.-H., Foelsche, U., Schmidt, T., Gorbunov, M.: Reproducibility of GPS radio occultation data for climate monitoring: Profile-to-profile inter-comparison of CHAMP climate records 2002 to 2008 from six data centers, *J. Geophys. Res.*, *117*, D18111, doi:10.1029/2012JD017665, 2012.

20   S.M. Khaykin, Funatsu, B.M., Hauchecorne, A., Godin-Beekmann, S., Claud, C., Keckhut, P., Pazmino, A., Gleisner, H., Nielsen, J.K., Syndergaard, S., and Lauritsen, K.B.: Post-millennium changes in stratospheric temperature consistently resolved by GPS radio occultation and AMSU observations, *Geophys. Res. Lett.*, *44*, 7510–7518, doi:10.1002/2017GL074353, 2017.

Kuo, Y.-H., Schreiner, W.S., Wang, J., Rossiter, D.L., and Zhang, Y.: Comparison of GPS radio occultation soundings with radiosondes, *Geophys. Res. Lett.*, *32*, L05817, doi:10.1029/2004GL021443, 2005.

25   Kursinski, E.R., Hajj, G., Schofield, J., Linfield, R., and Hardy, K.: Observing Earth's atmosphere with radio occultation measurements using the Global Positioning System, *J. Geophys. Res.*, *102*, 23429–23465, 1997.

Lackner, B.C., Steiner, A.K., Hegerl, G.C., and Kirchengast, G.: Atmospheric climate change detection by radio occultation data using a fingerprinting method, *J. Clim.*, *24*, 5275–5291, doi:10.1175/2011JCLI3966.1, 2011.

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

Leroy, S.S., Dykema, J.A.,, and Anderson, J.G.: Climate benchmarking using GNSS occultation, *Atmosphere and Climate*, Springer, Berlin Heidelberg, 287–301, 2006.

Leroy, S.S., Ao, C., and Verkhoglyadova, O.P.: Temperature trends and anomalies in modern satellite data: infrared sounding and GPS radio
35   occultation, *J. Geophys. Res.*, *123*, 11431–11444, doi:10.1029/2018JD028990, 2018.

Li, Y., Fritzer, J., Schwärz, M., and Zhang, K.: Dynamic statistical optimization of GNSS radio occultation bending angles: advanced algorithm and performance analysis, *Atmos. Meas. Tech.*, *8*, 3447–3465, doi:10.5194/amt-8-3447-2015, 2015.

Luntama, J.-P., Kirchengast, G., Borsche, M., Foelsche, U., Steiner, A., Healy, S.B., von Engeln, A., OClerigh, E., and Marquardt, C.: Prospects for the EPS GRAS mission for operational atmospheric applications, *Bull. Am. Meteorol. Soc.*, *89*, 1863–1875, doi:10.1175/BAMS2399.1, 2008.

Poli, P., Healy, S., and Dee, D.P.: Assimilation of Global Positioning System Radio Occultation data in the ECMWF ERA-Interim reanalysis,
5   *Q. J. R. Meteorol. Soc.*, *136*, 1972-1990, doi:10.1002/qj.722, 2010.

~~Rieck~~

Rieckh, T., Scherllin-Pirscher, B., Ladstädter, F., and Foelsche, U.: Characteristics of tropopause parameters as observed with GPS radio occultation, *Atmos. Meas. Tech.*, *7*, 3947–3958, doi:10.5194/amt-7-3947-2014, 2014.

Scherllin-Pirscher, B., Kirchengast, G., Steiner, A.K., Kuo, Y.-H., and Foelsche, U.: Quantifying uncertainty in climatological fields from
10   GPS radio occultation: an empirical-analytical error model, *Atmos. Meas. Tech.*, *4*, 2019–2034, doi:10.5194/amt-4-2019-2011, 2011.

Scherllin-Pirscher, B., Steiner, A.K., Kirchengast, G., Kuo, Y.-H., and Foelsche, U.: Empirical analysis and modeling of errors of atmospheric profiles from GPS radio occultation, *Atmos. Meas. Tech.*, *4*, 1875–1890, doi:10.5194/amt-4-1875-2011, 2011.

Scherllin-Pirscher, B., Deser, C., Ho, S.-P., Chou, C., Randel, W., and Kuo, Y.-H.: The vertical and spatial structure of ENSO in the upper troposphere and lower stratosphere from GPS radio occultation measurements, *Geophys. Res. Lett.*, *39*, L20801, doi:10.1029/2012GL053071,
15   2012.

Scherllin-Pirscher, B., Steiner, A.K., and Kirchengast, G.: Deriving dynamics from GPS radio occultation: Three-dimensional wind fields for monitoring the climate, *Geophys. Res. Lett.*, *41*, 7367-–7374, doi:10.1002/2014GL061524, 2014.

Scherllin-Pirscher, B., Syndergaard, S., Foelsche, U., and Lauritsen, K.B.: Generation of a bending angle radio occultation climatology(BAROCLIM) and its use in radio occultation retrievals, *Atmos. Meas. Tech.*, *8*, 109–124, doi:10.5194/amt-8-109-2015, 2015.

20   Schmidt, T., Schoon, L., Dobslaw, H., Matthes, K., and Thomas, M.: UTLS temperature validation of MPI-ESM decadal hindcast experiments with GPS radio occultations, *Meteorologische Zeitschrift*, *25(6)*, 673–683, doi:10.1127/metz/2015/0601, 2016.

Schreiner, W., Rocken, C., Sokolovskiy, S., and Hunt, D.: Quality assessment of COSMIC/FORMOSAT-3 GPS radio occultation data derived from single- and double-difference atmospheric excess phase processing, *GPS Solut.*, *14*, 13–22, doi:10.1007/s10291-009-0132-5, 2010.

Schreiner, W., Sokolovskiy, S., Hunt, D., Rocken, C., and Kuo, Y.-H.: Analysis of GPS radio occultationdata from the FORMOSAT-
25   3/COSMIC and Metop/GRAS missions at CDAAC, *Atmos. Meas. Tech.*, *4*, 2255–2272, doi:10.5194/amt-4-2255-2011, 2011.

Simmons, A.J., Berrisford, P., Dee, D.P., Hersbach, H., Hirahara, S., and Thépaut, J.-N.: A reassessment of temperature variations and trends from global reanalyses and monthly surface climatology datasets, *Q. J. R. Meteorol. Soc.*, *143*, 101-119, doi:10.1002/qj.2949, 2017.

Sokolovskiy, S., Rocken, C., Schreiner, W., and Hunt, D.: On the uncertainty of radio occultation inversions in the lower troposphere, *J. Geophys. Res.*, *102*, D22111, doi:10.1029/2010JD014058, 2010.

30   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*, RS0D24, doi:10.1029/2010RS004614, 2011.

Steiner, A.K., Hunt, D., Ho, S.-P., Kirchengast, G., Mannucci, A.J., Scherllin-Pirscher, B., Gleisner, H., von Engeln, A., Schmidt, T., Ao, C., Leroy, S.S., Kursinski, E.R., Foelsche, U., Gorbunov, M., Heise, S., Kuo, Y.-H., Lauritsen, K.B., Marquardt, C., Rocken, C., Schreiner, W, Sokolovskiy, S., Syndergaard, S., and Wickert, J.: Quantification of structural uncertainty in climate data records from GPS radio
35   occultation, *Atmos. Chem. Phys.*, *3*, 1469–1484, doi:10.5194/acp-13-1469-2013, 2013.

von Engeln, A., Andres, Y., Marquardt, C., and Sancho, F.: Radio occultation on-board of Metop, *Adv. Space Res.*, *47*, 336–347, doi:10.1016/j.asr.2010.07.028, 2011.

von Engeln, A., Healy, S., Marquardt, C., Andres, Y., and Sancho, F.: Validation of operational GRAS radio occultation data, *Geophys. Res. Lett.*, *36*, L17809, doi:10.1029/2009GL039968, 2009.

Wilhelmsen, H., Ladstädter, F., Scherllin-Pirscher, B., and Steiner, A.K.: Atmospheric QBO and ENSO indices with high vertical resolution from GNSS radio occultation temperature measurements, *Atmos. Meas. Tech.*, *11*, 1333–1346, doi:10.5194/amt-11-1333-2018, 2018.

5   Wickert, J., Reigber, C., Beyerle, G., König, R., Marquardt, C., Schmidt, T., Grunwaldt, L., Galas, R., Meehan, T., Melbourne, W., and Hocke, K.: Atmosphere sounding by GPS radio occultation: First results from CHAMP, *Geophys. Res. Lett.*, *28*, 3263–3266, 2001.