Reviewer #1
Response in italics

Paper Summary:

This paper analyses the random and bias error characteristics of data that was made available as part of the Radio Occultation Modeling Experiment (ROMEX). ROMEX is an agency and scientific community effort to quantify the benefits of assimilating increasing numbers of RO profiles for numerical weather prediction. In the three- month ROMEX experimental period, in excess of 35,000 profiles per day were made available to ROMEX participants who agreed to the terms and conditions of using ROMEX data. (This compares to less than 10,000 RO profiles per day that are available operationally.) The paper focuses on the three largest datasets from ROMEX: COSMIC-2 (C2), Spire, and Yunyao. The three-cornered-hat (3CH) method is used to determine the uncertainties of the C2, Spire, and Yunyao bending angles and refractivities. The analysis covers height dependence and geographic variations. Reanalyses are used as two corners of the 3CH method to estimate the uncertainty of the combined C2-Spire-Yunyao data set.

Review Summary:

The paper provides valuable scientific information on the ROMEX data, which is an unprecedented data set for RO and also involves new commercial sources of data such as Yunyao which has not been extensively analyzed in the past. This is the first time that such a large data set is evaluated for its error characteristics.

Understanding the error characteristics is very important for data assimilation and numerical weather prediction. Before publication, the paper should clarify certain aspects of the analysis and address questions as detailed below. After suitably addressing these aspects, the paper should be ready for publication.

Citation of the literature is generally appropriate, but with reliance on an unpublished document that could be given a DOI. The reference Aparicio 2024 is to a presentation that could be put online with a DOI.

The material is presented well in a logical and clear manner.

General response: We thank the reviewer for these perceptive comments, which have led to improvements in the paper.

Detailed Comments:

Line 107: a brief explanation of "excess phase data" should be provided so that the paper is accessible to a less specialized audience.

Response: *A radio wave is slowed as it passes through the atmosphere, which creates a doppler shift in the measured frequency of the wave and a delay in the phase of the wave as measured by the receiver. This delay is called the excess phase and is a fundamental observable in radio occultation (Kursinski et al. 1997). For details, please see The Radio Occultation Processing Package (ROPP) Pre-processor Module User Guide Version 11.3 (Section 3) https://rom-saf.eumetsat.int/romsaf_ropp_ug_pp.pdf . The text has been modified in Section 1.1 of the revised paper, directing the reader to this reference.*

Lines 150-166: while the theory of the 3CH method is explained elsewhere, this brief summary does not fully serve the paper. The paper compares different data sets to one another, and no data set is claimed to be "truth", so why is "truth" referred to here? Rather than truth, the authors might be referring to a reference data set for which biases and variances are determined relative to that reference. Wouldn't these equations still hold if one of the data sets is viewed as "reference" instead of truth? Or are these equations only valid if one of the data sets is actually "truth", which has zero bias and zero random error?

The term "bias" in this paper appears to refer to a bias between two data sets, and not between one data set and truth. What is meant by "bias" in the paper should be clarified.

Another way this brief introduction is not serving the paper is that the paper analyzes various subsets of data for which it becomes clear there is not a single bias applicable to all subsets. For example, bias appears to vary geographically, and the authors apply the analysis to the global data set which is not characterized by a single bias. Is the global bias expected to be the mean of the regional biases? Is equation (1) valid for a dataset that is characterized by multiple biases? The authors should clarify how the equations 1-3 apply to the data sets being analyzed in the paper.

Response: In a discussion of the error model that is the basis for the derivation of the 3CH equations for error variances (Eq. 1 in our paper), O'Carroll et al. (2008) point out that the concept of "truth" is non-trivial. They indicate that for the derivation of the 3CH method error variances we only need to assume that "we have a consistent definition of the true value for comparison with each observation." In general, "truth" is a reference value for each triplet of data in the 3CH method rather than necessarily corresponding to the true value of the observation. Thus in Eq. 1 the use of the word "truth" is conceptual in nature. "Truth" does not need to be constant in space or time; it is only necessary that it be the same for each individual triplet because it cancels immediately upon taking the differences between the observations in each triplet. The validity of the 3CH estimates of error variances does not depend on any of the three observations in the triplet being "truth." A more complete discussion of the 3CH method, the meaning of "truth," the factors limiting the accuracy of the method, and comparison with other widely used methods for estimating uncertainties are described in Sjoberg et al. 2021. In summary, the 3CH method provides estimates of the random error statistics, not the biases, and hence does not depend on knowledge of "truth."

The term "bias" is used in two different ways in the paper. When discussing biases in general, we are referring to biases with respect to the true (always unknown) value at the scale (footprint) of the observation. When we refer to specific quantitative examples of biases we are referring to mean differences between one or more data sets with respect to a reference dataset, which may be a model or another observation dataset. We do not assume either of the two datasets is "truth." The difference in meanings should be clear from the context. We have clarified this in Section 1.3 of the revised paper.

Both the true (unknown) and estimated biases of a dataset vary with time and geographic location.

References
O'Carroll, A. G., J. R. Eyre, and R. W. Saunders, 2008: Three-way error analysis between AATSR, AMSR-E, and in situ sea surface temperature observations. J. Atmos. Oceanic Technol., 25, 1197–1207, https://doi.org/10.1175/2007JTECHO542.1.

Sjoberg, J. P., R.A. Anthes, and T. Rieckh, 2021: The three-cornered hat method for estimating error variances of three or more data sets-Part I: Overview and Evaluation. J. Atmos. and Ocean. Technol., 38, 555-572, https://doi.org/10.1175/JTECH-D-19-0217.1

The following phrase is used on line 211: "but at the expense of fewer pairs in the sample and greater noise in the statistics." The description in lines 150-161 does not contain terms corresponding to this "statistical noise". For such a term to exist would require the concept of a sample mean as an estimate of the mean of a theoretical parent distribution. (Similarly for the parent standard deviation, etc.). The authors apparently are not concerned with statistical noise in their analyses, relying on large enough sample sizes to render the statistical noise negligible. The authors should make some reference to this implicit assumption in the paper.

Response: The reviewer is correct. For our analyses of collocated datasets, the sample sizes far exceed the sample size of order 1000 suggested by Sjoberg et al. (2021) where statistical noise in 3CH estimates may be considered negligible. We have added this statement to the revised paper in Section 1.1.

Line 270: provide some sense of what "impact height" is and how it relates to geometric height, for the less specialized reader.

Response: impact height is defined as the difference between the impact parameter and the Earth's local radius of curvature (Sokolovskiy et al. 2010). The impact parameter is a fundamental quantity used in RO technique, which refers to the distance between the ray asymptote to the local Earth's center. The impact parameter is constant along the ray path in a spherically symmetric atmosphere. Impact height is used in the retrieval of radio occultation observations and is related to the geometric height by the refractivity and the local radius of curvature of the Earth. Near the Earth's surface the impact height is approximately 2 km greater than the mean sea level altitude. The difference diminishes with height; by 10 km the difference is approximately 600 m and at 20 km the difference is approximately 125 m. The Radio Occultation Processing Package (ROPP) Pre-processor Module User Guide Version 11.3 (Section 3) https://rom-saf.eumetsat.int/romsaf_ropp_ug_pp.pdf describes the impact parameter. We have added a reference defining impact height in Section 2 of the revised paper.

Line 276-279: please clarify this sentence. It is hard to understand.

Response: We have revised the sentence as follows and inserted the following text at the end of the preceding paragraph: "The magnitude of these differences (less than 0.15%) are much smaller than the 3CH uncertainty estimates, which are 1.5% or higher. However, they may have an impact on the comparison of bending angle biases, which are of the same order of magnitude between 10 and 30 km."

Line 313-317: this paragraph and example should be better defined. There are numerous references that suggest RO can be used to advantage to improve predictions related to tropical cyclones (TC). What is meant by "resolve" here is not clear. RO inherently has relatively poor horizontal resolution per observation, which will not change with increasing numbers of observations. Even if RO cannot spatially "resolve" TC, increasing numbers of RO could improve predictions related to TC, such as intensification and track. This paragraph is not convincing regarding whether increasing numbers of RO have no benefit for TC.

Response: We agree that this paragraph could be misleading. We did not mean to imply that RO observations have no benefit for TC prediction; to the contrary, RO observations have been shown in many studies to improve the predictions of TC (Chen et al. 2022 and references therein), as noted by the reviewer. We have revised this paragraph accordingly.

Chen, Y.-J.; Hong, J.-S.; Chen, W.-J., 2022: Impact of Assimilating FORMOSAT 7/COSMIC-2 Radio Occultation Data on Typhoon Prediction Using a Regional Model. Atmosphere, 13, 1879. https://doi.org/10.3390/atmos13111879

Lines 324-325: if C2 does not exhibit the same count variation as Spire near the equator, it is worth commenting on why this might be case, if true. Both data sets sample the equatorial anomaly.

Response: The Spire and COSMIC-2 data are based on different receivers, come from satellites with different orbits and have different signal-to-noise (SNR) ratios. These differences may be the reason why quality of the Spire observations is affected more by the Equatorial anomaly than the C2 observations, and hence relatively more Spire observations are rejected by the quality control. We added a short comment hypothesizing reasons for why the pattern does not exist in the C2 observation counts in the revised paper.

Line 452: please provide the quantity (approximate) of operational data assimilated so that this statement can be provided in better context.

Response: The number of RO profiles per day assimilated in the operational ECMWF model during the ROMEX period was between 7,000 to 7,500, or about 20-21% of the total ROMEX profiles. This number has been added to the revised paper.

Lines 459-460: weren't the Yunyao data adjusted after the initial processing, so why does this artifact remain?

Response: As stated in lines 137-139, we used the original data provided by the provider, not adjusted data that were provided later. Changes made in the processing of other ROMEX missions were likely made after the original data were submitted to EUMETSAT, but it would be difficult to obtain these data or decide on which revised datasets to use and which not to use. Thus, we evaluated data processed from the original data that were submitted to EUMETSAT.

Line 502: Figure 7 is indeed impressive, but also somewhat puzzling. Whereas RO-RO comparisons have consistently shown growing uncertainties below 10 km, such uncertainty growth for the models is unexpected. For example, Figure 11 of Hersbach et al. (DOI 10.1002/qj.3803) for ERA5 temperature uncertainty does not appear to match what would occur with ERA5 bending angle uncertainty as indicated in Figure 7. While there is a modest increase of temperature uncertainty below ~8 km in the ERA5 paper, it does not increase dramatically towards the surface and is not much larger than uncertainty near 10 km. Please reconcile Figure 7 with Figure 11 of Hersbach et al.

Response: While the ERA5 EDA spread can be interpreted as a measure for uncertainty, it is not the same thing as the random error variance. The main reason for the qualitative differences in the vertical variation of temperature spread in ERA5 in the troposphere from the vertical variation of refractivity or bending angle in Fig. 7 is related to the high variability and uncertainties in water vapor in the troposphere, which is an important part of the refractivity and bending angle observations. Thus, the fractional uncertainty of refractivity and bending angle is greater than that of temperature in the moist troposphere.

Line 525: please clarify what BFRPRF in Figure 8 refers to. What are the dashed blue and green lines?

Response: BFRPRF refers to the RO data that were processed by UCAR in its level-2 BUFR product, (described in lines 137-139); it refers to the three RO missions plotted in Fig. 8. We added this to the Fig. 8 caption. The blue and green lines are defined in the caption. We added a note in the caption that there are three estimates for the error variances of ERA5 and JRA-3Q, one for each RO mission. The differences are small (a good thing since they should not depend very much on the RO mission) and are not easily visible in this plot.

Line 531: processing provenance is somewhat unclear. Were the Yunyao data used here reprocessed by EUMETSAT or UCAR?

Response: As noted in lines 137-141, all the data analyzed in this paper were processed by UCAR. The excess phase data from Yunyao were first sent to EUMETSAT and then to UCAR, where they were processed to bending angle and refractivity.

Line 583: please clarify what is the location of the ERA5 model and how is that determined. Is it a nearby grid point value? Isn't it straightforward to interpolate ERA5 values on a grid to an RO location, thus eliminating collocation error for all ERA5 comparisons?

Response: We added more detail into how we calculated the geographic maps of the bias estimates between the RO datasets and ERA5 model data in Section 4.2 of the revised paper. An alternative way would be to interpolate ERA5 model values to each RO location, but we wanted to use the same method that we used to compute RO biases vs. another RO dataset as reference,, where collocation to each RO observation was not possible.

Line 594: what is meant by short-range forecast of a reanalysis and why use that rather than the reanalysis value that is based on all contemporaneous data?

Response: Short-range ERA5 forecasts refer to forecasts made by the ERA5 model initialized 6-18 hours before an observation time in order to minimize the dependence of the model data on the observations being evaluated and thus eliminate or reduce error correlations that affect the estimates of biases and uncertainties.

Line 620: how is it known that above 30 km ERA5 biases are dominant?
Line 627: same comment – how is it known that ERA5 biases are dominant?

Response: We cannot be sure whether the large mean differences we see in the ROMEX data vs. ERA5 are observation or model biases because there is no Truth at these levels. However, above 30 km the biases of all three independent RO datasets (C2, Spire, and Yunyao) show very similar bending angle biases compared to ERA5 (Fig. 10a). The mean refractivities of C2, Spire, and

Metop B/C all show very similar differences from the mean ERA5 refractivities (not shown in the paper). We have looked at several other estimates of different RO missions against different models, and the RO data from independent RO missions with different instruments, orbits, processing details, etc. all exhibit similar biases against the same model, but the models exhibit different biases with respect to each other. The consistency of all the RO data and the inconsistency in the models indicate likely biases in the models' bending angles and refractivities. These biases could be caused by biases in the model data or the forward model used to compute refractivities and bending angles from the model data (e.g. small errors in the k_1 coefficient in the refractivity equation). We have modified the text in Section 4.2 of the revised paper..

Line 718: missing close of parentheses.

Response: Corrected.

Line 747: this suggests that the bias caused by Rc should not create model biases after assimilation because N does not exhibit the bias. The forward operator applied to bending angle should remove the bias. Please reconcile the small N bias with a bias in the model that assimilates BA. Note that Zhou et al. (DOI 10.1029/2024JD041295) detected temperature biases in C2 data. How do such biases arise if N is not biased due to Rc variations?

Response: The azimuth effect, as defined in Section 5.2.1 of the paper, is a representativeness difference associated with variations in R_c in different missions. We estimate that the BA difference of two occultations observed at the same location and time but along different azimuth angles, can be as large as 0.3%. This effect contributes to differences in direct comparisons of BA from different RO missions but won't raise an issue for BA data assimilation since the forward operator in the models accounts for the differences of R_c in different missions.

While the azimuth effect does not cause the N bias, there are other factors, such as the one discussed in Section 5.2.2., that do, and these result in the small N biases found in our estimates. The relationship between N and dry temperature is complex and non-local, so there is no simple relationship between N biases and temperature biases. Furthermore, the temperature biases estimated in Zhou et al. and the N biases estimated in this paper are computed in different ways and have different QC applied to the data being evaluated.

Lines 760-773: the DD examples used here relate to reducing collocation error. It's not immediately clear how DD reduces Rc error. Please provide more details regarding the algorithm that reduces Rc error. Include a discussion of how bending angle is computed from the model.

Response: The DD algorithm is given by Eq. (4) in the original paper (now Eq. 3 in the revised paper). It corrects for all representativeness differences, and the azimuth effect (related to Rc error) is a representativeness difference, not an error. The DD method is described in detail by

Tradowsky et al. 2017 and Gilpin et al. 2019, as referenced in the paper. The model bending angles are computed using a 1D forward model that accounts for variations in R_c *in different missions, thereby accounting for the different azimuth angles and the azimuth effect. We added this information to Sections 1.1 and 5.2 of the revised paper.*

Lines 800-802: so the claim is being made here that the local radius of curvature computed for each occultation is only computed at one altitude, and not repeatedly as the ray path drifts?

Response: Yes. The local radius of curvature is calculated at a single altitude for each occultation and does not change with the drift of the tangent points. Technically, it's possible to calculate the local radius of curvature at different altitudes, accounting for its variation with the drift of tangent points. However, this would introduce significant challenges in retrieving the refractivity profile. As far as we know, using a single local radius of curvature for the entire RO profile representation is the standard approach employed by all RO data processing centers.

Line 808: What of negative azimuths?

Response: The definition of azimuth may vary depending on the format of the RO data (e.g. BUFR or NetCDF) or data center from which it is downloaded. Herein, we use the atmPrf data processed by the CDAAC, where the azimuth angle of the occultation plane at tangent point is measured relative to North. Negative values indicate angles to the west of North.
Negative and positive values have the same effect, so only the absolute value of the azimuth angle is shown in Figs. 16 and 18. We have noted this in Section 5.2.1 of the revised paper.

Line 848: this section should explicitly note whether bending angles or refractivity are being referred to.

Response: These general conclusions refer to both bending angles and refractivity. We added this detail in the first paragraph of the Summary and Conclusions.

End of responses to Reviewer #1