Author Responses to Reviewer Comments on: *GNSS Radio Occultation Soundings from Commercial Off-the-Shelf Receivers Onboard Balloon Platforms*

by Kevin J. Nelson[1], Feiqin Xie[1], Bryan C. Chan[2], Ashish Goel[2], Jonathan Kosh[2], Tyler G. R. Reid[2], Corey R. Snyder[2], and Paul M. Tarantino[2]
[1]Texas A&M University - Corpus Christi, Corpus Christi, TX
[2]Night Crew Labs, LLC, Woodside, CA

We thank all reviewers for their detailed and constructive comments and suggestions during the second round of review. We have made efforts in addressing all reviewers' additional comments and suggestions and have done our best to incorporate them into the revised version of the manuscript. Below is our point-by-point response to each of the reviewers' comments.

Comments from Reviewer #1:

General comments:

My comments below relate to either new text in the revision or to replies to comments in the first review. There are several places where there are differences between the tracked changes provided by the authors and the revised manuscript. Also, a new figure (Figure 3) was added together with a new paragraph discussing it, but the new text was not marked as new in the tracked changes. I ended up not relying on the tracked changes. Comments below refer to what is in the revised manuscript.

Specific comments:

1. L6-8: I urge the authors to remove the claim in the abstract that the data are 'high-quality' and that the variability is comparable to spaceborne RO missions. In their answer (2a) to my comment on this in the first review, the authors refer to the paper by Schreiner et al. (2020) and write: "Between 10 and 20 km, the COSMIC-2 refractivity difference standard deviation was shown to be approximately 1%, slightly smaller than our observed variability estimates." However, 1% is not 'slightly smaller' than 2.3%. The authors also write in their answer: "Given the size of the receiver and the payload being constructed of commercially available parts, we consider this to be high-quality." Whether something is 'high-quality' cannot depend on how it comes about. There is a very recent paper by Cao et al. (2022) (https://acp.copernicus.org/articles/22/15379/2022/) showing that BRO can provide refractivity profiles with a standard deviation less than 1% between 10 and 15 km, in comparison to ERA5 (their Figure 8). The fact that it is not possible to provide this kind of quality with commercially available parts (at least not in this study) is an important message to get across. In any case, the abstract should objectively state the results of the study.

a. **Thank you for this comment. We have elected to remove the phrase “high-quality” from the sentence in question from the manuscript.**

2. L39: Perhaps missing an 'and' before "by fuel range of the aircraft."
   a. **Thank you for catching this error. It has been corrected in the manuscript.**

3. L124 (and many other places): I think the word 'delay' refers to the time delay of a signal, but for the phase the word 'excess' seems more appropriate (the signal is delayed in time, but there is an excess in the phase due to the neutral atmospheric influence). Please consider to remove 'delay' when talking about the phase.
   a. **Thank you for this comment. We have removed the word “delay” from references to excess phase in the manuscript.**

4. L142: Like the other reviewer, I find it relevant to discuss the impact of the POD uncertainty for BRO in step (b). In their answer (3a, bottom of page 9) to the other reviewer the authors write that "a brief discussion was added to reflect the concerns of the POD uncertainty.". But where? I don't see it.
   a. **Thank you for this comment. The discussion referenced in the prior round of revisions was probably insufficient. We have expanded the discussion of POD impacts on *N*-bias and moved it into section 4 of the manuscript: “… One potential cause is the difference in precise orbit determination (POD) solutions for BRO missions. Generally speaking, precise orbit determination (POD) solutions for LEO missions aim to have LEO velocity accuracies of 0.5 mm s^{-1} or better, and LEO position accuracies of 10 cm or better. In contrast, BRO missions are generally capable of velocity accuracies of 30 mm s^{-1} or better, and position accuracies of 5 cm or better. The larger POD velocity errors are due to difficult-to-model disturbances such as wind gusts and other aerodynamic factors. Xie et al., (2008) showed that the addition of simulated of 5 mm s^{-1} random excess Doppler errors will not result in additional *N*-bias, but could possibly introduce less than 1% refractivity error near the receiver (~10 km) and less than 0.2% below ~ 6 km. In the case of our study, the larger BRO receiver positioning errors (if random) will not introduce significant *N*-bias (Fig. 9) for World View cases.”**

5. L172-177: It is still unclear to me how many ERA5 profiles are used. The authors answer (15a) to my comment on this in the first review, and their revised manuscript, did not help. In their answer they write: "... median profile used to run the initial ROSAP simulation, run the initial Forward Abel Integrator, and determine the time series of refractivity at the receiver during the occultation.". So one profile for all this. But in the revised manuscript it says: "... one referencing refractivity profile .... at the zero-elevation ... used to compute the time series of refractivity at the receiver.... Furthermore ... we use a median refractivity profile ... surrounding the zero-elevation TP location for input into the initial ROSAP and FAI simulations." Thus, two profiles as I read it, one at zero elevation TP to determine the time series of refractivity at the receiver,

and another one (taking the median of grid points surrounding the zero-elevation TP) for ROSAP and FAI simulations. If their answer to me is correct, then the text needs to be updated to make it clear that it is the same profile used for these three tasks. I understand that a second (or third) profile at the 5 km TP location is used for comparisons because the lower TPs are drifting away from the zero-elevation TP location.

   a. **Thank you for this comment. We have rewritten Section 2.4 to better summarize the ERA5 refractivity profiles used during the retrieval process. The relevant portion now reads: "…. To best evaluate the quality of the individual retrieved BRO refractivity profiles, the final refractivity comparison uses the ERA5 profile at the 5 km TP location determined by ROSAP. Therefore, three separate refractivity profiles from ERA5 are used during the retrieval process, the zero-elevation angle refractivity profile, the median refractivity profile surrounding the zero-elevation angle location, and the 5 km TP location refractivity profile."**

6. Figure 5: I believe the dN (d for deci) in the x-axis label is the wrong unit. In the metric system 20 dm = 0.2 m. Thus 20 dN-units = 0.2 N-units. As mentioned in my first review, I think it should be deca N-units, so that 20 deca N-units = 200 N-units.
   a. **Thank you for catching this error. The reviewer is correct. After further investigation the abbreviation in the x-axis label should be "da" instead of "d." This change has been made to the figure and in the manuscript.**

7. Figure 5: The dashed lines need to be mentioned in the Figure caption. In their answer (16a) to my comment on this in the first review the authors write: "The dashed lines indicate the boundary layer height detected with the gradient method based on temperature, specific humidity, and refractivity (Nelson et al., 2021; Ao et al., 2012, 2008; Winning et al., 2017; Xie et al., 2006).". However, these papers are not referred to in the manuscript, where it just says (L189) "... weak gradients in specific humidity, and refractivity ...", which is not very understandable. I cannot see these weak gradients at 0.9 km in particular. Please mention the gradient method and the references in the paper. It is still not clear why there are two lines if they are based on temperature, specific humidity, and refractivity. Please make that clear in the paper. In the Figure caption it could say something like "The dashed lines at approximately 0.9 km indicate the PBLH (see text for details)."
   a. **Thank you for this comment. The gradient method for determining PBLH and selected references have been added to the manuscript. The various line styles and colors have also been added to the text. There are, indeed, three individual lines showing the PBLH. However, the PBLH identified from specific humidity and refractivity is exactly the same, so the lines overlap. Each line has a different style, and this is clarified in the text as well.**

8. L191-195: Please mention the sampling rate of the observations here or earlier in the paper. Are the simulations done at the same sampling rate? In their answer (19a), the

authors tell that the simulations has now been run through the same smoothing process as the observed excess phase data to remove high-frequency variability. This needs to be told in the paper.

a. **Thank you for this comment. The explanation of the data processing and the retrieval process in Section 2.3 has been expanded to include further discussion of the sampling rates for the data and the smoothing algorithms applied.**

9. Figure 6: The ROSAP results look much better now. Not only is the high-frequency noise removed, but it also seems that the excess phase has come much closer to the calibrated observations than in the original figure (where the two curves crossed near 1000 sec). What is the reason for the latter? I don't suppose the smoothing can change the excess phase like that.
    a. **Thank you for this comment. In Figure 6, we have chosen to remove the raw, uncalibrated observations in addition to showing smoothed ROSAP excess phase and Doppler. Because the same smoothing algorithm is now applied to both the observed and simulated excess phase, the two datasets are much more similar to each other than they were previously.**

10. L199-202: I don't want to debate whether 141 and 700 is of the same order of magnitude as claimed by the authors in their answer (20a), and in the manuscript. To me the text is still misleading when it says: "The overall mean SNR from the GROOT receiver (141.79 V/V) is on the same order of magnitude (order of 100) as the mean SNR from the COSMIC-1 and SAC-C GNSS RO satellite missions (approximately 700 V/V ...)". I would rather say something like: "The overall mean SNR from the GROOT receiver is 141.79 V/V. This is relatively small compared to the mean SNR from the COSMIC-1 and SAC-C GNSS RO satellite missions, which is approximately 700 V/V." However, I think the COSMIC-1 and SAC-C SNRs are a bit smaller than 700 V/V due to defocusing at comparable tangent point altitudes (which could be relevant when you do such SNR comparisons). Please check. On the other hand, since the signal only passes through about half of the atmosphere in BRO relative to spaceborne RO, the defocusing effect in BRO will be smaller. This makes SNR comparison to spaceborne RO tricky. And how about the L2 signal? These discussions on SNR seems to focus only on L1, but the data quality also depends on the L2 signal. It should be mentioned in the paper that the discussion is about the L1 signal. It would be interesting to see also the SNR for the L2 signal in Fig. 6c. Please consider to add this.
    a. **Thank you for this comment. We have changed the discussion to focus on the L1 SNR to better reflect the reviewer's discussion on the differences between spaceborne and BRO SNR to the following: "The overall mean L1 SNR from the GROOT receiver (141.79 V V^{-1}) is smaller than the mean SNR from the COSMIC-1 and SAC-C GNSS RO satellite missions (approximately 700 V V^{-1}, Ao et al., 2009; Ho et al., 2020). Although the L1 SNR values from the Piksi receiver are approximately 5 times less than the values from spaceborne RO missions, considering the compact size of the Piksi receiver, such L1 SNR values are quite**

impressive. L2 frequency data was not as reliable as L1 frequency data and would often be interrupted or drop out earlier than L1. As such, the L2 data was only used for pre-processing data quality check (e.g., cycle slips), but not used for ionospheric correction." This has been added to the discussion in Section 2 for clarification as well.

   b. **We agree with the reviewer regarding the complexity of the SNR comparison between airborne/balloon-borne and spaceborne RO due to the potential influence of the defocusing effect resulting from RO geometry differences. We have also added the following to the discussion in Section 3 highlighted below: "Although the L1 SNR values from the Piksi receiver are approximately 5 times less than the values from spaceborne RO missions, considering the compact size of the Piksi receiver, such L1 SNR values are quite impressive, which is partially attributed to the less defocusing effect than spaceborne RO as the Piksi receiver is in the atmosphere."**

11. Figure 7: How is the ROSAP bending angle calculated? From simulated excess Doppler? Or ray traced bending angle? The authors answer (22a) to this question did not help to clarify this. Please clarify in the paper.
   a. **Thank you for this comment. The ROSAP bending angle is referring to the bending angle resulting from the ray-tracing simulation, so the total bending angle is calculated using accumulated changes in bending along each ray path at each impact parameter iteration. The text explaining the ROSAP simulation now reads: "which simulates the GNSS RO signal and calculates the associated excess phase and excess Doppler as well as the along-path accumulated bending angle at each impact parameter as it travels through a prescribed Earth's atmosphere (with either spherical or oblate Earth) by a given 1-dimensional atmospheric refractivity profile."**

12. L211-214: "Differences between the retrievals and the ROSAP simulation between ~6 and 8 km are most likely caused by differences between oblate and spherical Earth geometry assumptions. While we perform an Earth oblateness correction as part of the transmitter/receiver geometry processing, these effects may not be completely removed.". I very much doubt that to be the case. Why would the oblateness create something so distinct between 6 and 8 km, and then even after the oblateness correction? Is there a reason why you don't think this could be due to variations in the real atmosphere? Or could such variations be caused by yaw instability? Or ionospheric effects not adequately removed? Please revise the text unless you have substantial evidence (if so, please provide it) suggesting that this comes from the oblateness correction (you could just remove these new sentences if you don't have a likely explanation, and say that the variations are not understood). There are also differences around 11-12 km that should be discussed if these around 6-8 km are discussed.
   a. **Thank you for this comment. We agree with the reviewer that the oblate correction in ROSAP is not the reason for the differences in the retrieved bending angles. On the contrary, ROSAP bending angle is expected to be closer**

to the "reality", if the input ERA5 profile represents the atmosphere well, which we believe to be true in a statistical sense. We hypothesize that the bending differences between the retrievals and the ROSAP simulation are more likely due to yaw rotation errors and their subsequent self-correction aboard the balloon platform. We see this more clearly with the feature between 10 and 11 km, where the retrieval bending increases only to relax back to values in line with both the ROSAP and Forward Abel simulations. These features also qualitatively coincide with changes in SNR (Fig. 6c) which is indicative of platform yaw rotation for a balloon platform. For the WVG26 case, the relative closeness of the retrievals to the Forward Abel simulation (assuming spherical symmetric atmosphere) opposed to the ROSAP (with oblateness correction) is likely by chance, as retrievals from other individual cases (not shown) do not necessarily match the Forward Abel better than ROSAP.

b. **As far as other potential reasons for the bending angle differences, although we did not carry out ionospheric corrections using L1/L2 linear combination due to the lower quality of L2 signal, we do expect the ionosphere effects are mostly corrected resulting from the bending angle differencing to compute partial bending angle. In addition, residual ionospheric effects should be more pronounced at higher altitude and is not seen in Figures 7 and 8. Therefore, it is unlikely that the variations are the result of ionospheric effects.**

13. L225-230: In their answer (24a) the authors explain that the median and MAD values referred to in the text are calculated over certain altitude regions: "As these are individual profiles, the median 'above 10 km' is a single value calculated from all data points above 10 km as well as within other regions.". But this is not explained in the revised text. Please do that if you prefer to discuss these median and MAD values. Later on (in Section 4) the median and MAD values are based on ensembles of profiles, which is quite different. Please make that clear. Alternatively (which I think would be much better), the discussion of median and MAD values based on the results in Fig. 8b in Section 3 could be omitted. It doesn't seem to add any useful information.
    a. **Thank you for this comment. We have removed the discussion of the median/MAD for the single profiles shown in Figure 8.**

14. L249-250: "The ZPM-1 refractivity differences from both retrieval methods show much more variability across all heights in both the median profile and the individual refractivity profiles." I disagree with this statement regarding the individual profiles. I see less variability in the grey lines in Fig. 10 than in Fig. 9. The document with tracked changes, and the authors answer (28a), mention "oscillations" that are not mentioned in the revised manuscript. Still, it is not clear what "oscillations" versus "variability" mean in this context. I think it would be better to acknowledge that there is less variability in the individual ZPM-1 results. The larger variability in the median seems to be a result of the lower number of profiles. In any case, there are so few profiles so that the results are likely not statistically significant.

   a. **Thank you for this comment. We have rewritten the discussion of Figure 10 to indicate that the overall variability in the individual profiles is smaller than the World View campaign data. In the previous response to reviewer comments, we intended for the "oscillations" to refer to the 1000 m-scale features in Figure 10 where the *N*-bias increases over ~500 m then relaxes back to near its original value. Given the reviewer's comments we have elected to discuss them as "1000 m-scale vertical fluctuations in *N*-bias" rather than "oscillations."**

15. L258: "Reasons for errors ... can come from ... causes." Sentence does not make much sense. Instead, you could say: "Errors ... can come from ... causes".
    a. **Thank you for catching this grammar/syntax error. This has been updated in the manuscript to: "Errors in BRO refractivity retrievals can come from a variety of potential sources."**

16. L278: In this sentence the authors added 'and temporal': "The added benefit of using BRO platforms is the dense spatial and temporal sampling available due to the low platform velocities relative to the LEO-based RO satellites." It can be difficult to understand why the low platform velocities would increase the temporal sampling, but perhaps replacing the word 'available' with 'over targeted regions' in this sentence would more clearly describe what I think the authors mean.
    a. **Thank you for this comment. This change has been made in the manuscript.**

17. Figure A1: I still don't see any distinction between Piksi 1 and Piksi 2 in the text. In their answer (33a) the authors explain it nicely. Please consider to put that explanation in the Appendix.
    a. Thank you for this comment. We have repurposed portions of our response from the previous review iteration that discusses the differences between Piksi 1 and Piksi 2 and placed it in the appendix as the following: "Balloon-borne RO can collect high density observations around the platform, particularly compared to spaceborne RO. Figure A1 shows a Sankey plot filtering visualization of the World View predicted occultations. Onboard the GROOT payload, we had duplicate Piksi receivers. Piksi 1 was programmed to log data from GPS, Galileo, Glonass, and Beidou, whereas Piksi 2 was only programmed to log data from GPS and Galileo. As such Piksi 2 was able to log more ROs from GPS. We believe the Piksi receivers and antennae were manufactured such that they were tuned to maximize GPS performance, and as such satellites from the GPS constellation consistently showed better performance than the other constellations. Of the original 680 predicted ROs from all GNSS constellations (originally discussed in section 2.1), a total of 485 were incomplete, and therefore not suitable for retrieval processing. The remaining 195 occultations then filter further by removing those from QZSS, Galileo, BeiDou, and GLONASS with low quality due to the frequency tuning inherent to the Piksi receiver. Of the 167 from the GPS system, only 8 good quality cases were ready to use immediately. Another 7 cases required additional pre-processing in the form of

cycle-slip corrections. The same process was also applied for cases observed during the ZPM-1 flight campaign."

Comments from Reviewer #2:

Manuscript Title: "GNSS Radio Occultation Soundings from Commercial Off-the-Shelf Receivers Onboard Balloon Platforms"

Manuscript Number: amt-2022-198

Date Reviewed: 19 DECEMBER 2022

Manuscript Evaluation: This work could potentially augment space-born ROs by better sampling the lower-to-middle troposphere. The authors have done an excellent job in addressing all my comments from the previous review cycle. I am not satisfied with the status of the current manuscript, and I recommend publication in its current form. However, I would like the authors (if they keep working on balloon ROs) to work on techniques to remove the SNR drops during BROs, as well as perform additional sensitivity studies to understand the influence of velocity accuracy on derived products.

Thank you to the reviewer for their constructive suggestions and improvements to the manuscript.