Response to Reviewer #1

We thank the Reviewer for their useful comments and suggestions that have greatly improved our manuscript. The following outlines the changes made to our manuscript in response to the Reviewer's concerns. Reviewer comments are in italics, and responses are in regular font.

General Reviewer Comment: *The last part of this study is the comparison of Arctic winter PBLH between RO and MERRA-2 reanalysis. However how is the PBLH calculate from the reanalysis data? This information is vital for the readers to understand whether the method used is appropriate and the value of the results.*

Response: The version of the GEOS model used in MERRA-2, includes two PBL parameterization schemes, viz. Lock scheme which is activated for unstable PBLs and the Louis scheme which is activated for stable PBLs. The model PBL depth is defined as the model level where the eddy heat diffusivity coefficient (K_H) value falls below 2 m^2s^{-1} threshold. At a given time, only one PBL depth value is calculated by the model, either by the Lock scheme or by the Louis scheme. Both schemes use different methods of estimating eddy diffusivity coefficients, and therefore PBL heights. We added a description of the schemes and a discussion of their relevance for the Arctic Ocean is included in the revised paper. Please, see Section 2.2.

Reviewer Comment 1: *As shown in Fig. 6, the typical PBLH can be as low as 300 m over the sea ice region. The specific cut-off threshold of 500 m used for selecting RO profiles may introduce biases in the resolved PBLH. Purely visualizing the resolved shallow PBLH from commercial RO as in Fig. 6 doesn't justify the choice of this threshold. A sensitivity test about the cut-off threshold is needed.*

Response: In the revised paper, the sensitivity to cut-off altitude threshold is extensively discussed in sections 2.1.3 and 3.4. For most RO datasets, the standard 500m cut-off altitude threshold works well. It is found that Spire NASA data perform better when a lower cut-off altitude of 300 m is used instead of the standard 500 m threshold. Please see section 2.1.3, section 3.4, and Fig. 11 of the revised manuscript.

Reviewer Comment 2: *In addition to the map of the monthly averaged penetration probability in Fig. 5, please include the map of the monthly averaged minimum penetration depth.*

Response: We agree with the reviewer. This is a good suggestion. We have included a map of the monthly average minimum penetration depth, and revised Figure 3 to accommodate the inclusion. Please, see revised Fig. 3 in the main manuscript.

Reviewer Comment 3: *Explain how the PBL height of MERRA-2 reanalysis was obtained/calculated, and the vertical resolution of the MERRA-2 reanalysis data.*

Response: Thank you for the suggestion. We have included text to describe the parameterization schemes used for computing MERRA-2 PBL height and their relevance for the Arctic Ocean. Please, see Section 2.2.

The vertical grid of MERRA-2 is based on terrain-following sigma coordinate system, wherein the exact model level height is a function of the surface pressure. In general, the first model level is around 50 m above surface and the spacing is approximately 100 m within the lowest five model levels. **This information is now included in section 2.2 of the main manuscript as well**.

Reviewer Comment 4: *Expand the PBLH study to include the Arctic summer season.*

Response: The methodology used to compute the PBLH in this study is based on Ganeshan and Wu (2015) which is a validated RO technique for PBLH estimation over the Arctic Ocean during winter months (Nov-Apr). This technique works well when the specific humidity is low, and the refractivity profile is mainly sensitive to temperature gradients. In the revised paper, we have expanded on the PBLH study to all cold season months (i.e. Nov-Apr).

Specific comments:

1. L36-37: "*improved predictability over flat surfaces compared to varying slopes*".

What's the meaning of this sentence? Any references support this statement?

Response: This refers to the fact that GNSS RO has a low horizontal resolution (100-200 km) and therefore is expected to perform well over topographically homogenous surfaces, which includes flat surfaces (sea ice, open ocean) compared to sharp varying slopes (e.g. coastal land mass areas). The statement has been rephrased as follows: "improved performance over flat surfaces (sea ice, open ocean) compared to sharp varying slopes (land mass),…". Please see section 1.1, Line 42, of revised manuscript.

2. L45-46: *I don't understand the statement like "RO profiles over the Arctic Ocean dropped sharply …". Please rephrase it.*

Response: This statement has been rephrased in the revised manuscript (section 1.1, Line 53) as follows:

"From the analysis of 8 years of COSMIC-1 data, it was found that availability of RO profiles over the Arctic Ocean reduced significantly at tangent heights below 1km, which introduces a sensitivity of the retrieved PBL height to the choice of the cut-off altitude, or minimum RO penetration depth, used for profile selection".

3. In L78, *the authors mentioned that "Spire data are provided at a similar vertical grid and resolution as other GNSS RO missions", but later, when the authors tried to explain the observed lesser regional variation of PBLH from Spire compared to GeoOptics, it is mentioned that Spire* *data have coarser resolution due to smoothing (L250). Are these two statements contradictory to each other? What's the vertical resolution of Spire data?*

Response: Spire data are indeed provided at a similar vertical grid as other GNSS RO missions, however, NASA Spire bending angle profiles are excessively smoothed prior to the refractivity retrieval, thus reducing their effective vertical resolution (i.e. loss of information due to smoothing).

4. L82: *what is "the amplitude of computed phase match integral"? Any references?*

Response: The methodology for retrieving GeoOptics neutral atmosphere profiles and their quality control is described in the "GeoOptics Processor for Radio Occultation (GeoPRO) User Guide" which was provided to NASA by the vendor. Phase matching (Jensen et al., 2004) in RO processing is a wave optics technique designed to extract the full information from the received wave field. It is conceptually and practically simpler than other wave optics techniques, while producing a number of useful diagnostics and additional features. It has often been used in detailed analysis of individual occultations and produces radio holographic images of each occultation which are extremely useful in diagnosing signal or processing issues and can even reveal new information. As part of the quality control, profiles are cut-off at low altitude when the phase match amplitude falls below a certain threshold.

5. L83: *What data is considered as "at lower levels"? Below 8 km?*

Response: Yes, the levels below 8km are flagged as bad quality if the blanket criteria check is failed. This is clarified in the revised manuscript (section 2.1.1, Line 86).

6. L86-89: *The part is not clear to me. Has the GeoOptics data below what is called "sharp" layer been discarded by QC check? If so, does it mean the resolved PBLH later would be equal to the minimum penetration depth?*

Response: No, if the QC check is passed at any altitude below "sharp layers", the data are not discarded. Each profile is evaluated individually to determine the minimum penetration depth ascertained by the lowest above-surface level with a "good" quality flag. If a "sharp" PBL inversion layer with poor QC flag exists above the minimum penetration depth, then this is not disregarded. We rephrased this part and better explained it in the revised manuscript. See, Section 2.1.

7. L90-94: *Are the NOAA Spire and GeoOptics data processed by UCAR? If so, UCAR's processing starts from which level? Any useful information can we derive from the comparison between NASA and NOAA purchased commercial data?*

Response: Yes, the NOAA Spire data are processed by UCAR from purchased L1b data. The NASA-purchased Spire data are processed to level 2 by the vendor.

8. L184-185: *Any explanations for "missing seasonal variation in NASA Spire data, but presented in NOAA Spire data"?*

Response: Differences in seasonal variability between NASA Spire and NOAA Spire data are evident because the two datasets are processed by different methodologies.

9. L194: *What is the vertical resolution of radiosonde observations?*

Response: Radiosonde observations are no longer used in this study.

10. L200-202: *Please provide the results similar to Fig. 3 and 4 at 300 m, 500 m and 700 m?*

Response: Figures 3 and 4 are no longer part of the analysis. The water vapor and RO penetration probability relationship will be explored in a follow-on study.

11. L252-254: *Clearly the cut-off height of 500 m is not sufficient to derive the shallow PBL height. I don't understand the logic of the cut-off threshold used in data analysis being allowed to be mission dependent.*

Response: The cut-off altitude or minimum required RO penetration depth, in some sense, is a first guess estimate of the expected typical height of the PBL. A sampling bias may occur in the retrieved PBLH due to a sharp drop in available RO profiles, thereby necessitating the selection of an optimal cut-off altitude threshold for minimum RO penetration depth. While the standard cut-off altitude of 500m has been regarded as sufficient for deriving refractivity-based PBLH from COSMIC-1 RO observations in the Arctic, it is carefully examined in this study for use with different RO datasets. It appears that the 500 m cut-off altitude when applied to NASA GeoOptics and ROMSAF MetOp data is sufficient for obtaining a realistic representation of the shallow Arctic PBLH. However, in the case of NASA Spire data, the derived PBLH values are slightly higher compared to the other two RO datasets and MERRA-2 reanalyses (Fig. 6 in revised manuscript). This is because the percentage of available NASA Spire RO profiles drops significantly going from 400 to 300 m and then from 300 to 200 m (Figure 10 in revised manuscript), which could potentially lead to a positive bias in the retrieved PBLH values when the standard cut-off altitude of 500 m is chosen. No such sharp drop is seen for GeoOptics and MetOp datasets. Moreover, a similar comparison with the NOAA Spire product shows that this sharp rate of decline only exists in the NASA Spire data. As a result, the PBLH retrievals from NASA Spire data are recomputed using a lower cut-off altitude threshold of 300 m, and the resulting PBLH values are found to be significantly lower and in better agreement with other datasets (see Fig. 11 in revised manuscript). However, the spatial patterns and seasonality are not impacted by the choice of cut-off altitude threshold. In summary, an optimal cut-off altitude threshold for RO products can be chosen based on their rate of RO penetration decline within the PBL. However, the impact of cut-off altitude threshold is limited to simply an improvement in the magnitude of the retrieved PBLH, and no strong sensitivity is observed to spatiotemporal patterns. This is explained in section 3.4 of the revised paper.

12. *I don't see the value of Fig. 5 e-g. May consider remove them.*

Response: These figures and discussion are no longer part of the new manuscript.

13. *The observed extremely high GeoOptics PBLH over the 30E to 60E sector is presented in Fig. 6. What's the explanation for this? Is it physical-related or outlier-effected?*

Response: The extreme high values of GeoOptics based PBLH in the Atlantic Sector seems to be related to the high minimum penetration altitude in this region (seen in Fig. 3). This is noted in the revised manuscript in section 3.3, Line 234.

Author References:

Jensen, A. S., Lohmann, M. S., Nielsen, A. S., and Benzon, H.-H.: Geometrical optics phase matching of radio occultation signals, Radio Science, 39, n/a-n/a, https://doi.org/10.1029/2003rs002899, 2004.

Response to Reviewer # 2

We thank the Reviewer for their useful comments and suggestions that have greatly improved our manuscript. The following outlines the changes made to our manuscript in response to the Reviewer's concerns. Reviewer comments are in italics, and responses are in regular font.

Reviewer Comment 1: *There are multiple RO missions data set available from UCAR and ROMSAF websites, including KOMPSAT-5, PAZ, PlanetiQ, TerraSAR-X, TanDem-X, Sentinel-6a and GRACE-FO. These missions are contemporaneous with Spire and GeoOptics covering global area. They also show good penetrating capability and usefulness in PBLH application. Authors need to give explanation of the reason choosing C-1 and MetOp as the counterpart in evaluation for commercial RO.*

Response: The Reviewer makes a good point. The primary reason for not selecting other listed satellites such as, KOMPSAT-5, PAZ, TerraSAR-X, TandDem-X, Sentinel-6a, and GRACE-FO is because RO measurements is a secondary observable. They are primarily radar imaging and radar altimeter type instruments and do not offer consistent sampling of neutral atmosphere RO as their receivers are not always turned on, as opposed to COSMIC-1 and MetOp's GRAS (Global Navigation Satellite System Receiver for Atmospheric Sounding) that are dedicated GNSS RO missions, have similar orbital heights of roughly 800 km, and offer consistent sampling of neutral atmosphere at regular fixed local time. Moreover, we did not select PlanetiQ data for this study as we are focusing on NASA-purchased commercial RO datasets with at least one year of observations.

In the revised study, we have made one change to the source of MetOp dataset. Based on the Reviewer's suggestion, we have used MetOp data from ROMSAF, which is re-processed with an updated, single software version, similar to COSMIC-1 re-processed datasets from UCAR. In the revised paper, we have made this point in section 2.1.2., Line 119 (see below): "To remove ambiguity resulting from software updates - and to ensure consistency - only those RO mission products that have been re-processed with the same software version are compared against Spire and GeoOptics."

Reviewer Comment 2: *The radiosonde profiles from MOSAiC expedition ship campaign provide unique and valuable information on the PBLH detection in Arctic region as independent verification data. I expect to see the radiosonde derived PBLH and the direct comparison of the collocated RO-RAOB PBLH matching pairs. But in this paper the radiosonde data was only treated as source of water vapor for exploring the relationship* *between moisture and RO penetration. Consider the very limited geographic coverage of the ship campaign, does the conclusion cannot be achieved by using water vapor profile from climate model data, like MERRA-2 or ERA-5?*

Response: As the Reviewer points out, the radiosonde data are limited spatially and not ideal for comparing against GNSS RO-derived measurements which have a coarse horizontal resolution (100-200 km). As a result, the radiosonde profiles are not used for RO-RAOB PBLH comparisons. Instead, we use MERRA-2 reanalyses for comparing RO-derived PBLH. We agree with the Reviewer that the Arctic PBLH comparison should be the main focus of this paper, and we have therefore removed the section comprising the analysis of water vapor and RO penetration probability. MOSAiC radiosonde observations are no longer used as a dataset in the revised paper.

Reviewer Comment 3: *As far as I know, the MERRA-2 reanalysis provides two PBLH values. One is calculated based on the total eddy diffusion coefficient of heat (K_h), and the other one is estimated using the bulk Richardson number method. The PBLH calculated from K_h is usually higher than the one from bulk Richardson number in most regions. It would be interesting to check the difference over the Arctic area and validate it using RO derived PBLH. However, the paper did not provide much description on the MERRA-2 PBLH. Reader don't even know how the PBLH was extracted from MERRA-2 reanalysis.*

Response: The version of the GEOS model used in MERRA-2 includes two PBL parameterization schemes, viz. Lock scheme which is activated for unstable PBLs and the Louis scheme which is activated for stable PBLs. The model PBL depth is defined as the model level where the eddy heat diffusivity coefficient (K_H) value falls below 2 m^2 s^{-1} threshold. At a given time, only one PBL depth value is calculated by the model, either by the Lock scheme or by the Louis scheme. Both schemes use different methods of estimating eddy diffusivity coefficients, and therefore PBL heights. We have included a description of the schemes and a discussion of their relevance for the Arctic Ocean in the revised paper. Please, see Section 2.2.

Reviewer Comment 4: *In the PBLH deriving method section (2.1.3), the method of first minima of the refractivity gradient to exceed -40 N-unit km-1 and the 500 m threshold for RO cut-off height are chosen without justification. A sensitivity study for the threshold, comparison of different methods (minimum gradient, wavelet covariance transformation etc.) and variables (refractivity based, bending angle based) are recommended, according to the discussion for figure 6(b) and 6(c).*

Response: The goal of this study is to compare the PBL heights derived from commercial RO datasets using the previously established and validated methodology (Ganeshan and Wu 2015) for Arctic Ocean. This methodology is found to work well for cold season (Nov-Apr) months over the Arctic. Radiosonde observations from the SHEBA campaign were used to demonstrate that low specific humidity during the cold season months led to a heightened sensitivity of refractivity to temperature gradients (Ganeshan and Wu 2015). The radiosonde data show an empirical relationship between the height of the first local minima in refractivity gradient and the Arctic PBL temperature inversion height. Furthermore, the methodology is found to yield reasonable monthly mean PBLH values when applied to COSMIC-1 observations (Ganeshan and Wu 2015). Thus, in this study, the same methodology is adopted. The sensitivity to cut-off altitude threshold, however, is discussed in section 3.4.

Reviewer Comment 5: *The seasonal variation of RO penetration probability is displayed and discussed, whereas the more important seasonal variation of PBLH was not provided. In my opinion, a big picture of PBLH in north pole region is desirable (seasonal variation, diurnal circle if any, longitudinal variability related to the Atlantic Ocean current and sea ice distribution etc.) in the section 1.1, then a statement of how commercial RO can improve the understanding in section 3.3.*

Response: This is a good suggestion. We have included in the revised paper monthly mean RO-derived PBLH for six months of the year, showing seasonality and spatial variability due to the distribution of sea ice and open water (Figures 5-9). Since the RO-derived PBLH retrieval only works well for cold season months as described in Ganeshan and Wu (2015), we only estimate the seasonal cycle for November to April.

Reviewer Comment 6: *In figure 1, the different penetration probability of Spire NOAA and Spire NASA may contributed by the sample noncoincidence, because Spire NOAA is a small subset of Spire NASA (~3500 out of ~12000 in one day). Whereas for GeoOptics, NOAA and NASA are basically covering the same observations. Therefore the explanation of the discrepancy of orange/red lines may be completely different. Since the paper introduced NASA purchased commercial RO, which is processed by vendor, and NOAA purchased commercial RO, which is processed by UCAR, it's ideal to derive PBLH using both NASA and NOAA commercial RO, to help understanding the factors affecting RO penetration.*

Response: This is a great recommendation. In the revised paper, we have included a figure comparing the exact same subset of Spire radio occultations purchased by NASA and by NOAA. Indeed, the difference in the penetration statistics are due to differences in processing software as shown in Figure 2(a). We further drive home this point by including a figure (Fig. 2(b)) showing the similarity in penetration statistics between two different sources of radio occultations, viz. COSMIC-2 and Spire NOAA, over the tropics that are processed by the same software.