Reviewer#1 Comments:

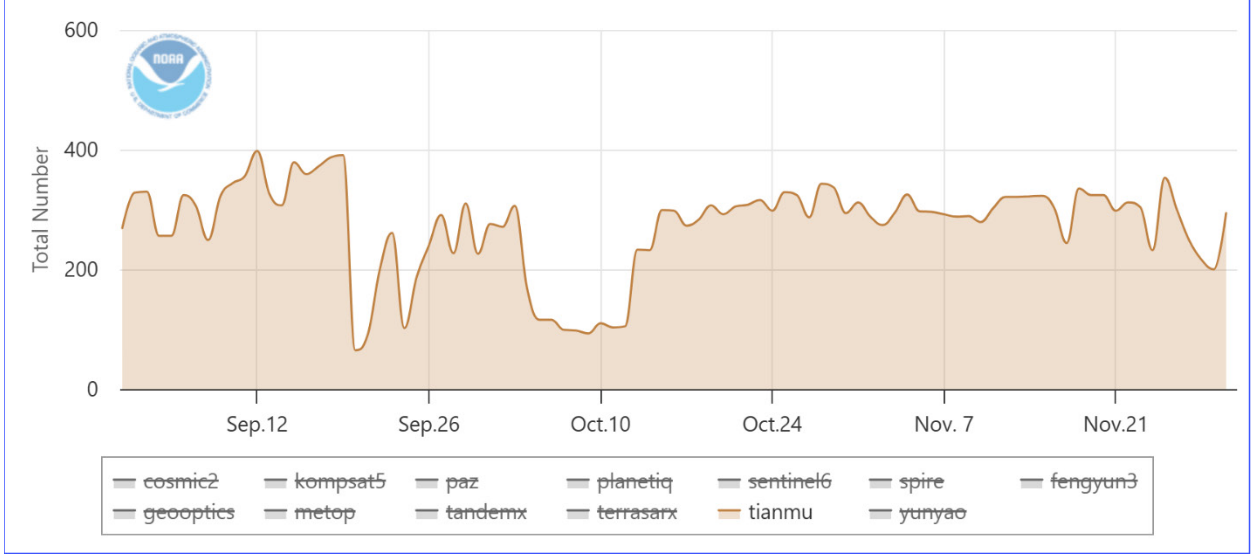
The paper is interesting as a source of homogenized structural uncertainty studies, where the ROMEX database has been used to provide a coherent sampling of different source data. It is relevant to be aware of difficult to reduce errors, sometimes systematic, and that some atmospheric circumstances lead to retrievals of better quality, while others are quite dependent on apparently minor details of the retrieval. The issue of the retrieval uncertainty is still insufficiently addressed.

The subject is thus interesting and a good application of the ROMEX database. There however are a number of corrections and clarifications that I believe are important, as detailed below.

Reply:
Thank you for your constructive suggestions, which have helped improve the quality and clarity of this manuscript.

L150 and Tab1: The volume of Tianmu is indeed small within the ROMEX ensemble. However the average I have, admittedly on a superficial count, is of the order of 300 profiles/day, rather than 100/day. Please verify.

Reply:
Thank you for pointing this out. We verified and updated the daily profile count for Tianmu (~300 profiles/day) in Table 1 and the corresponding text (Line 150) to ensure the presented numbers are accurate. You can see the daily number below.



Changes in manuscript:
The numbers are updated in Table 1, line 155 and line 174.

L213, etc but also elsewhere: The authors mention interpolation of certain quantities onto a common sampling grid. Since certain quantities such as position are non-trivial to interpolate to the required accuracy, please indicate summarily some details about the sufficiency of the interpolation. A quadratic is mentioned. Is a quadratic interpolation sufficient at 1Hz?

Reply:
We checked the code and actually the cubic spline interpolation instead of quadratic interpolation is used to align the satellite position vectors (GNSS and LEO), which are typically sampled at a low rate (e.g., 1 Hz), to the much higher excess-phase sampling rate (e.g., 50–100 Hz). For aligning 1Hz vectors to 50-100 Hz excess phase data, cubic spline interpolation (or high-order Lagrange polynomials) is the standard. The interpolation is of the satellite ephemerides, which are inherently smooth. However, a quadratic fit may introduce jumps every second, which may create artificial noise in the high-rate phase residuals. A cubic spline ensures that not only is the position smooth, but the velocity and acceleration are also continuous at the boundaries. This mimics the physics of a satellite moving through a gravity field much more accurately. We have updated this in the text.

Changes in manuscript:
We replaced quadratic interpolation with cubic spline. See line 254, Equations (1) and (3).

L234: "intersection of the line-of-sight with the WGS84 ellipsoid". The authors seem to be defining the SLTA=0 line and tangent point. This straight line between satellites is not a "line-of-sight", as the propagation trajectory is bent. Also, this line at this moment is not intersecting the ellipsoid, but is tangent to it. Please adjust the wording.

Reply:
We agree with this correction. The tangent point is defined as the location where the straight line between the transmitter and receiver touches the ellipsoid (straight line tangent altitude SLTA or height of straight line HSL equals 0). We adjusted the wording in Line 234 to be more physically accurate by replacing "intersection of the line-of-sight with the WGS84 ellipsoid" with "The tangent point location on the reference ellipsoid (WGS84) is determined where the straight line between the transmitter and receiver touches the ellipsoid"

Changes in manuscript:
The change is in lines 272-273.

L235: Presumably also the center of curvature has to be computed, and the coordinate system adjusted to be centered at that point.

Reply:
We agreed. We added "…and the center of curvature are computed at this reference tangent point…"

Changes in manuscript:
The change is in line 274.

L257: Please provide some references to some of the existing studies around this truncation. This is a significantly complex issue, and the referencing should indicate its non-trivial character.

Reply:

We agree to add relevant references to existing studies on SNR-based signal truncation and cut-off methods to strengthen this part of the methodology: "An appropriate cut-off height is essential to ensure the accuracy and reliability of tropospheric information retrieved from GNSS signals in OL tracking mode (Sokolovskiy et al., 2009; Sokolovskiy et al., 2010; Adhikari et al., 2021; Paolella et al., 2025)."

Changes in manuscript:

The change is in lines 280-282.

L272-296: The MSIS-90 is ok for the temperature and to establish the approximate range in moist refractivity. However, one of the major issues is the presence of large vertical gradients of relative humidity, which may lead to large bending angles, superrefraction, etc. This model phase does not account for this possibility, nor seems to be considering the range of moistures. Is this model only identifying bit slips? Or will it introduce (false positive) bit corrections when there are RH gradients?

Reply:

The model phase derived from MSIS-90 (Lines 271–290) is an approximate, simplified reference for the sole purpose of identifying and correcting the abrupt, unphysical $\pm\pi$ navigation bit jumps (Lines 292–296). It is not intended to model high moisture gradients or super-refraction, and the final retrieval is not based on it.

L300: There are many parameters here. Although I agree that filtering is to some extent desirable, are these parameters optimal? Have other filtering parameters been tested, with the optimal results chosen? Under which criteria are the mentioned parameters optimal?

Reply:

The 0.5-second filter window is chosen because it approximately corresponds to the vertical extent of the first Fresnel zone (Lines 304–306), which is commonly used as a benchmark for the intrinsic vertical resolution of RO observations. The shorter windows (0.05 s (100 Hz) and 0.1 s (50 Hz)) are applied below 10 km to better resolve fine-scale structures in the lower troposphere while still maintaining sufficient noise suppression. The choice of these parameters is based on systematic tuning experiments in which multiple configurations were evaluated. The selected values showed here provide the most consistent statistical agreement with model outputs, such as ERA-5, yielding reduced bending angle bias and standard deviation compared with other tested configurations. We added these statements in the text.

Changes in manuscript:

The change is in lines 352-355.

L324: With MSP there may be a number of subcomponents that are stationary. There may be from none under anomalous propagation to several under multipath, although there is often just one dominant component. Should not L327 (eq 9) be a sum of perhaps several stationary components? Or has it been assumed that we focus only in the dominant? Does the text apply only to a case where there is one dominant component? Or does eq 9 express only one of the stationary components, in case there were several? Please clarify.

Reply:
Equation (9) is not a sum because it is presented within the context of applying the MSP approximation, which assume a single stationary point for a given frequency $\widehat{\omega}_q$ to illustrate the fundamental relationship between the phase derivative and the pseudo frequency. This assumption holds true for ideal occultations where the physics ensures a unique relationship between frequency and the ray's impact parameter. However, the FSI method is designed to overcome this limitation in realistic occultations (e.g., multipath). The technique uses the entire signal spectrum by applying a single Fourier transform to the entire complex signal, and applies corrections (such as resampling the signal with respect to open angle and correcting for radial variations) to eliminate interferences caused by higher-order derivatives and non-spherical trajectories. By applying these geometric and orbital corrections, the FSI method effectively modifies the signal so that the fundamental assumption of a single stationary point, or a unique frequency-to-ray relationship, is restored, allowing the single-component MSP approximation (and thus the framework containing Eq. 9) to be successfully applied to the corrected signal.

L363: Does this projection modify only the satellite orbits? Or, since the radius and center of curvature vary during the occultation, are these somehow reprojected also?

Reply:
The radius and center of curvature are treated as fixed parameters for a given RO event; they are not dynamically reprojected. The correction at Line 363 applies specifically to the satellite orbits, transforming the actual LEO/GNSS trajectories into idealized circular paths. This projection is performed relative to the pre-computed local center of curvature (Lines 364–370) to maintain a consistent coordinate system for the inversion process. We added these statements in the text.

Changes in manuscript:
The change is in lines 419-422.

L386: Does this criterion reject the low troposphere?

Reply:
The FSI amplitude criterion is the final step in a multi-stage SNR truncation process (Lines 240–261). Its purpose is to define the lowest valid bending angle-impact parameter pair for the profile. This threshold is indeed applied to reject signal contaminated by high noise (which often occurs in the lowest troposphere), as retaining this noisy data would lead to unphysical oscillations (Lines 257–260). This is an integral part of the noise-mitigation strategy. The FSI method does provide an alternative estimation at all practical altitudes.

Changes in manuscript:
We added changes in lines 435-439.

L399, eq 17: Aren't you using alpha in a different sense wrt before? Before it was the bending angle, while here it would suggest that alpha is the instantaneous bending angle minus the average bending angle (thus something like alpha-). Please make sure that the variable naming is consistent.

Reply:
We agreed. We revised the variable naming in the text and in Eq. (17) for consistency, to clearly define the variable representing the shifted bending angle relative to the central component.

Changes in manuscript:
We added changes in lines 448-449 and updated Eq. (17).

L417: Please reference Eq 18, as it is not the definition of the neutral bending angle, but a good practical approximation to obtain it from alpha L1 and alpha L2 (Vorobev & Krasilnikova, 1994).

Reply:
We added a reference (Vorobev and Krasilnikova, 1994) for the dual-frequency linear combination (Eq. 18) and clarified its role as an approximation for the neutral atmosphere bending angle.

Changes in manuscript:
We added the reference in lines 471-473.

L441: Could this climatological data be specified?

Reply:
The bending angle profile is extended to 150 km using the MSIS-90 climatological model to ensure numerical stability during the Abel inversion's upper boundary condition.

Changes in manuscript:
We added the climatological model (MSIS-90) in lines 484-485 and line 498.

L456…: Is this procedure valid only in hindcast? Or can it be applied in Near-Real-Time? (Availability of ERA5). It is also unclear in L457 if sigma_Year means the yearly variability of O-B, or the variability of B. In L466, thus later, it appears that it is O-B. Besides, is this statistic global? It could have been regional, by latitude band… Please clarify.

Reply:
The described QC is only for the post-processed RO data and cannot be applied to Near-Real-Time (NRT) processing, since ERA5 is not available in NRT. This QC is applied to the ROMEX dataset (hindcast) using ERA5 forecasts as the background (Lines 448–451). A separate internal QC system is under development for NRT processing (Lines 193–194, 474–475). σ_{year} is the annual mean standard deviation of the fractional bending angle O-B difference (%) (Lines 456–457). This statistic is based on the annual global average of the fractional bending angle O-B differences.

L470: Is this intended to analyze only height between 8-40km (or 10-40). The most interesting part is below 5 km. Is the method inappropriate here?

Reply:
The ERA5-based QC is intentionally restricted to the 10–40 km range (Lines 457–460) because this is where RO data quality and agreement with the background model are most stable. We do not apply this specific QC below 10 km due to increased atmospheric variability and signal complexity (Lines 469–470). By bypassing this filter in the lower troposphere, we ensure that a larger volume of bending angle data is retained for the lower altitudes you mentioned.

L512: Comment here if the ionosphere was particularly active during the study period, thus if these are average circumstances, worse-case, etc.

Reply:
The ionospheric conditions during September-November 2022 were generally moderate and typical for the ascending phase of Solar Cycle 25, but intermittently disturbed (including several geomagnetic storms and enhanced irregularities). They are not representative of worst-case conditions, but also not purely quiet-average; they are better described as moderately active with episodic disturbances. We added these sentences in the text.

Changes in manuscript:
We added the ionospheric conditions in lines 576-580.

L514…: Comment here on the typical SNR of the different missions.

Reply:
Agreed. We added a table (Table 2) to summarize the typical SNR characteristics of GNSS receivers across different missions, supporting the discussion of mission-dependent performance and structural uncertainties.

Table 2 Signal-to-Noise Ratio (SNR)[#] characteristics of GNSS receivers across different missions

RO mission	GPS	GLONASS	Galileo	BeiDou
Metop B	730.2			
Metop C	789.3			
COSMIC-2	1295.1	1181.4		
SPIRE	387.9	707.2	316.4	
PlanetiQ	1440.5	1580.6	1124.8	1335.7
KOMPSAT-5	617.2			
PAZ	503.9			
TerraSAR-X	622.4			
TanDEM-X	549.3			

[#] Mean SNR between altitudes 60 km and 80 km, with unit (volt/volt)

Changes in manuscript:

We added Table2 at lines 194-196 and lines 184-186 to show the SNRs from different missions.

L581: Isn't there a GO-to-FSI switch at a given altitude? 25km? It is mentioned again below, but would it not be appropriate to remind it somewhere in this section?

Reply:
Agreed. We explicitly stated in Section 3.2.2 that "Bending angles are computed using geometric optics (GO) above 25 km and the full spectrum inversion method below this altitude" (Line 311). We pointed this out again in section 4.2.

Changes in manuscript:
We pointed this out in lines 721-722 in section 4.2.

L654…: The average and std between methods are shown, but not compared to the O-retrieval. Would it not be appropriate to indicate the fraction of the O-B differences that are explained by structural uncertainties of the retrieval? That is, are the O-B dominated by the retrieval method? Or is the exact retrieval a secondary source of uncertainty? This fraction is likely a function of altitude.

Reply:
We agree with the reviewer that explicitly framing the comparison between the total O-B uncertainty and the structural uncertainty is essential for the interpretation of our results. The comparison of the standard deviation profiles in Figs. 5-7 (total O-B Uncertainty) and Figs. 11-15 (structural uncertainty) is, in fact, the intended purpose of presenting the results in this manner, to isolate and quantify processing-related effects. The structural uncertainty is the portion of the total O-B error attributable solely to differences among retrieval algorithms.

To address the comment directly and clearly convey this key finding, we added a sentence in the Discussions and Summary (Section 5) that explicitly: "The explicit comparison of the total O-B standard deviation (Figs. 5-7) with the structural uncertainty (Figs. 11-15) quantifies the contribution of retrieval algorithm differences to the total error budget, demonstrating that structural differences account for approximately one-fourth of the total O-B standard deviation over all the altitudes, providing a critical metric for interpreting ROMEX forecast impact studies and refining GNSS RO data assimilation systems."

Changes in manuscript:
The changes are in lines 929-934 in section 5.

L840: Does the FSI being presented restrict the vertical range where it is practical to apply? Some issues mentioned above would suggest that the procedure may be inappropriate for the low troposphere. Is it the case? Or the method does provide an alternative estimation at all practical altitudes?

Reply:
The FSI technique, a core component of the STAR FSI algorithm, is specifically designed to address the challenges posed by strong vertical gradients and multipath propagation that

complicate retrievals in the lower troposphere. It is not inherently restricted in its vertical range and does provide an estimation at all practical altitudes. The reviewer's concern about the low troposphere is related to signal quality, not the FSI technique's applicability. The FSI method is explicitly chosen for its strength in resolving fine-scale atmospheric structures in the lower atmosphere. The SNR truncation criterion is a separate quality control step applied to the processed profile. Its purpose is to mitigate noise (which is often highest in the lowest troposphere) and prevent unphysical oscillations in the retrieved profiles. This quality control step ensures that only the lowest valid bending angle-impact parameter pair is retained, but the FSI method itself can produce estimates across the full vertical extent of the atmosphere.

Reviewer#2 Comments:

This paper describes the FSI-based retrieval system implemented at NOAA STAR (referred to as "RFSI") and applied to process the ROMEX multi-mission dataset. The retrieved bending angle and refractivity profiles were compared statistically with results obtained using "standard" ROPP (based on CT2), starting from the same Level 1b excess phase data, and with results processed by EUMETSAT.

Overall, the paper contains interesting results that can be worthy of publication. I agree with the general finding that "structural uncertainty depends on both the processing algorithm and the satellite mission" (Lines 808-809) and studies like this are an important step in the right direction. The difference found in the COSMIC-2 bias is very interesting (Fig. 16). However, as detailed below, I think the explanation/attribution of the differences between the different processing systems was not done very rigorously. There are also certain aspects of the processing that should be explained better. These should be revised.

Reply:
Thank you for your constructive suggestions, which have helped improve the quality and clarity of this manuscript.

Specific comments:

"Multipath effects": The paper claims throughout the article that "multipath effects" are an important factor that account for the differences among the different processing methods (and among the different RO instruments). I would like to see this elaborated or substantiated. Given that FSI and CT methods both untangle the atmospheric multipaths in a similar way, why would the multipath leads to a different result? And is there any reason to think that the atmospheric multipath would impact the open-loop signal tracking among the different RO instruments/missions?

Reply:
There is a fundamental difference between the FSI and CT methods in their approaches to wave optics inversion. Although both are Wave Optics methods, their mathematical treatment of the signal leads to different outcomes in complex multipath environments (i.e., the lower troposphere): FSI utilizes the full spectrum of the received signal, analyzes the Doppler shift and its spectral component, and is highly sensitive to small-scale atmospheric features and to SNR cut-off height.

CT methods use a canonical transformation to map phase path data onto a canonical surface, simplifying signal processing and enabling efficient retrieval, while being less sensitive to noise and other signal disturbances. This difference in mathematical treatment explains the difference in the resulting bending angle profiles, particularly where atmospheric multipath is significant. The impact of atmospheric multipath on open-loop (OL) signal tracking varies by mission because the high water vapor variability in tropical/subtropical regions (e.g., COSMIC-2's low-inclination orbit) makes OL tracking inherently more difficult for those specific observations, as further detailed in the response to Lines 502-505.

Changes in manuscript:
The changes are in lines 57-63.

FSI vs CT2: The paper claims that there is a fundamental difference between FSI and CT2 in that "The FSI method, designed to resolve fine-scale atmospheric structures by leveraging full-spectrum signal information, demonstrated improved sensitivity in the lower atmosphere…" (Lines 796-798). Also Linea 57-58 (and other places throughout the paper): "FSI has demonstrated particular strength in resolving fine-scale atmospheric structures." Is it really true that FSI has improved sensitivity over CT/CT2 or phase matching? Is there any published paper that can back up this claim?

Reply:
FSI relies on explicit signal localization to enhance vertical resolution and mitigate multipath by reducing spectral mixing, whereas CT2 achieves an implicit, physics-based localization in impact parameter space, enabling more robust separation of multipath contributions. As a result, FSI is generally more sensitive to fine-scale atmospheric structures but also more noise-sensitive, while CT2 provides more stable retrievals in strong multipath conditions at the expense of reduced small-scale resolution. We will clarify that the claim is based on the published literature that introduced and developed the FSI method (e.g., Jensen et al., 2003; Adhikari et al., 2016, 2021). These publications discuss FSI's theoretical ability to resolve high-frequency, fine-scale structures. We will ensure the manuscript clearly cites these sources whenever this claim is made, providing proper academic backing.

Changes in manuscript:
The changes are in lines 57-63, lines 94-99, and line 909.

RFSI SNR cutoff (Lines 240-261): RFSI uses a fairly complex SNR cutoff method. Unless it is described elsewhere (in which case please provide a citation), it would be important to include some examples showing how removing the low-level SNR spikes would modify the bending angle retrieval results.

Reply:
We agree that this is valuable for the reader's understanding. While full examples may be too detailed for the scope of this paper, we include a citation illustrating the problem (noise spikes) and the intended solution (clean cutoff) to aid the reader's understanding (See Adhikari et al., 2021 Figure 5). We also add relevant references to existing studies on SNR-based signal truncation and cut-off methods to strengthen this part of the methodology: "An appropriate cut-off height is essential to ensure the accuracy and reliability of tropospheric information retrieved from GNSS

signals in OL tracking mode (Sokolovskiy et al., 2009; Sokolovskiy et al., 2010; Adhikari et al., 2021; Paolella et al., 2025)."

Changes in manuscript:
The changes are in lines 280-283.

Section 3.2.1: computation of model phase. It is not entirely clear to me this model phase is used for. Is it for correcting navigation bit jumps only? Why are you not using the actual GNSS navigation bits to account for this? Such "external" navigation bit jumps correction could be unreliable when the SNR is low.

Reply:
We reiterate that the model phase derived from MSIS-90 is computed solely to serve as a high-fidelity reference for identifying and correcting the abrupt, unphysical $\pm\pi$ navigation bit jumps (Lines 292–296) that may remain in the Level 1b excess phase data. This method is necessary for robust phase continuity correction at high sampling rates. We don't have the external navigation bits information to correct these jumps.

Section 3.4: quality control. A "7 sigma" (annual global average) determined from ERA5 forecast field is used to perform QC over 10-40 km altitude (not impact height?). Why "7 sigma"? And is the rejection rate? Also, Lines 465-466: "Model simulation data is unavailable" – why would it be unavailable?

Reply:
The $7\sigma_{year}$ threshold was selected based on a sensitivity study to achieve the "optimal balance between data retention and data quality" (Lines 459–460). The QC rejection rate for the ROMEX dataset depends on the mission. For example, the QC pass rates are 85.7%, 94.8%, and 93.9% for COSMIC-2, Spire, and PlanetiQ, respectively. These values are comparable to those from CDAAC-processed data: 87.6%, 95.7%, and 90.6%, respectively. The "Model simulation data is unavailable" flag means that if the Background (B) simulation data is unavailable, the QC check cannot be performed, and the profile is conservatively flagged to ensure that only profiles fully validated against the ERA5 reference are retained. This condition has been removed to avoid potential confusion in the text.

Changes in manuscript:
The changes are in lines 523-529.

Lines 385-385: "The lowest point is defined as the location where the amplitude drops below 0.35 of the normalized amplitude." How is 0.35 justified?

Reply:
The 0.35 threshold for the normalized FSI amplitude is an empirically determined optimal value based on extensive testing within the STAR FSI system. This value was chosen to maximize vertical penetration depth while minimizing the introduction of noise-induced artifacts into the final bending angle profile.

Changes in manuscript:
The justification is in lines 436-439.

Sec 3.3.1: Please explain what background bending angle was used for the statistical optimization.

Reply:
The background bending angle used for statistical optimization is derived from the MSIS-90 background field.

Changes in manuscript:
The changes are in lines 484-485.

Lines 502-505: Figs 5-6, COSMIC-2 shows higher BA standard deviation below 10 km, especially using the RFSI method. "These discrepancies are likely attributed to increased atmospheric variability in the boundary layer, limitations in signal tracking during multipath propagation, and the sensitivity of the bending angle retrievals to SNR cut-off thresholds." First, "boundary layer" should be changed to "troposphere" since the increase in standard deviation extends up to 10 km. Second, I don't know what "limitations in signal tracking during multipath propagation" is referring to. And why should the "SNR cut-off thresholds" affect COSMIC-2 more than other missions?

Reply:
We changed "boundary layer" to "lower troposphere" (Line 503). "Limitations in signal tracking" refers to challenges faced by the onboard receiver, especially in the tropical regions (COSMIC-2's orbit), where high water vapor variability makes OL tracking difficult. COSMIC-2 is affected more because its low-inclination orbit concentrates observations in the tropics/subtropics, a region of high atmospheric variability and extreme multipath, making its retrievals inherently more sensitive to the choice of SNR cut-off.

Changes in manuscript:
The change is in lines 566-568.

I find it quite difficult to tell the different lines apart from the bending angle and refractivity figures in Figs 5-10. Can this be improved by changing the color scheme? Perhaps it makes sense to skip some missions (e.g, TSX, TDX, KS5, Paz should be quite similar, so you can probably combine them or just show one of them and state any notable difference among them in the caption or text). To interpret Figs 5-10 properly in the lowest few km, it would be important to see the profile depth penetration for the different missions/processing. Could this be included somehow without cluttering the plots too much?

Reply:
Agreed. Figures 5-10 have been revised to adopt a more distinct color scheme, that clearly distinguishes among missions and processing systems. In addition, profile penetration depths below 8 km have been included in Figures 5-7 for different missions and processing systems to aid interpretation of lower-tropospheric performance. As the penetration depth exhibits a similar behavior for refractivity, it is not shown in Figures 8-10.

Changes in manuscript:
Figures 5-10 have been revised, and penetration depths have been included in Figures 5-7.

Review#3 Comments:

1st review of "Processing Multiple GNSS RO Data Using FSI and ROPP: Results from the ROMEX" by Chen et al.

This manuscript evaluates the performance of the STAR RFSI processing system in deriving atmospheric profiles from level-1b data across multiple GNSS RO missions collected during the ROMEX project. By comparing the results against ROPP (using the CT2 method) and EUMETSAT-processed datasets, the authors further discuss the structural uncertainties inherent in RO retrievals. While the study presents noteworthy results, certain areas remain ambiguous (listed below) and would benefit from further clarification.

Reply:

Thank you for your constructive suggestions, which have helped improve the quality and clarity of this manuscript.

General comments:

1. The FSI method was introduced two decades ago and is recognized as a powerful WO method for resolving fine-scale structures and multipath effects. However, its requirement for specific geometric conditions (e.g. circular orbits) meant that it required significant/complicated correction (as described in Line 363-374) before it could be reliably used in operational RO data processing. In addition, the FSI method is highly sensitive to noise, as acknowledged by the authors. In the lower troposphere, where SNR is often low, FSI performance can degrade significantly. These limitations have likely contributed to the fact that FSI has not been widely adopted in operational RO processing centers. Given these considerations, the manuscript would benefit from a clearer explanation of the scientific or practical motivation for reprocessing data using FSI. Beyond generating an additional version of RO products, what specific advantages does the FSI offer compared to existing operational algorithms?

Reply:

We agree that a clearer motivation is needed. We added a statement to the Introduction emphasizing that "FSI is theoretically well suited for resolving fine-scale atmospheric structures and handling multipath in the lower troposphere (Jensen et al., 2003; Adhikari et al., 2021), a key source of uncertainty for NWP. The use of the STAR RFSI algorithm within ROMEX is therefore to quantify the structural uncertainty associated with this alternative, high-resolution retrieval approach relative to community-standard methods, thereby providing critical insight for optimizing multi-mission data assimilation strategies."

Changes in manuscript:

We added the changes in lines 94-99.

2. The authors devote a substantial portion of Section 3 to describing the methodology. It would be helpful if they could first clearly summarize the core methodological differences between the approach presented here and that described in Adhikari et al. [2021]. Such clarification would allow readers to better understand the novelty and specific contributions of the current work.

Reply:

The core FSI algorithm remains consistent with that described in Adhikari et al. (2021). The novelty of the present work lies in the development and presentation of the complete STAR RFSI end-to-end processing framework (including data preprocessing, quality control, and statistical optimization) and its systematic, harmonized application to the large and diverse multi-mission ROMEX dataset for comprehensive assessment of structural uncertainties against other major processing centers (such as EUMETSAT and UCAR). We added these sentences in the text.

Changes in manuscript:
We added the changes in lines 200-205.

3. The title of Section 3, "Full Spectrum Inversion Algorithm", appears somewhat misleading. Based on the content, the section describes the STAR FSI-based processing framework as a whole, including the full processing chain for retrieval Level 2 data products from Level1b data. This scope extends well beyond the pure FSI algorithm itself, which is fundamentally a WO method for resolving multipath effects in the troposphere. The title could therefore be reconsidered to better reflect the broader processing steps being presented.

Reply:
We agree. We changed the title of Section 3 to "STAR RFSI Algorithm and Processing Chain" to better reflect the broader scope, which covers the full retrieval process from Level 1b to Level 2 data products.

Changes in manuscript:
We added the change in line 198.

4. The manuscript also introduces several technical terms and abbreviations that are not sufficiently explained or referenced, making it difficult for readers to follow. For example: what're the differences between STAR RFSI and STAR ROPP? Is it correct to simply understand the STAR RFSI uses the FSI algorithm, whereas STAR ROPP applies the CT2 algorithm? What is the core difference between STAR ROPP and the well-known ROM SAF ROPP processing system? Is the STAR ROPP considered the "community standard" ROPP referred to in Lines 19-20 and in the title?

Reply:
STAR RFSI is our algorithm that uses the FSI method to retrieve bending angles. STAR ROPP is the customized ROPP v10.0 system at STAR capable of running either the CT2 or FSI (RFSI) method. The official "community standard" ROPP is the ROM SAF ROPP, from which our STAR ROPP is distinguished (Lines 77–78). In this study, the community-standard dataset refers specifically to data generated using the STAR ROPP CT2 method. We explicitly stated this in the text.

Changes in manuscript:
We added the changes in lines 82-92.

5. Does EUMETSAT use ROM SAF ROPP operationally to process RO data? If not, what kind

of WO method is used in their processing? CT2?

Reply:
For the ROMEX dataset, the EUMETSAT-processed data (except for COSMIC-2) are generated using ROM SAF ROPP as the internal processing system. The COSMIC-2 data in the EUMETSAT dataset were processed by UCAR CDAAC using the Phase-Matching (PM) method (Line 681).

6. Regarding QC (Section 3.4), the authors use ERA5 forecasts as reference for QC checks, and ERA5 reanalysis as a reference for evaluating RO retrievals. Could the authors assess the impact of using ERA5 as a QC reference on the resulting dataset? For example, how sensitive are the retained profiles to the choice of ERA5-based thresholds?

Reply:
The impact of small-scale ERA5 errors on the QC is minimized by employing a very high threshold $7\sigma_{year}$, which is intentionally conservative. This threshold is chosen to filter only the most extreme outliers, ensuring the retained profiles remain largely insensitive to fine-scale variations in the ERA5 forecast field. The resulting QC pass rates are comparable to those from CDAAC-processed data with internal QC procedures. The QC rejection rate for the ROMEX dataset depends the mission. For example, the QC pass rates are 85.7%, 94.8%, and 93.9% for COSMIC-2, Spire, and PlanetiQ, respectively, compared to 87.6%, 95.7%, and 90.6%, for the corresponding CDAAC-processed datasets.

Changes in manuscript:
We added the changes in lines 526-529.

7. In Section 4, global RO retrievals from different ROMEX missions processed by different systems are compared. This raises concerns of the impact of significant sampling differences on the statistics shown in Fig. 5-10. In particular, C2, as the only low-inclination mission, with observations concentrated in mid- or low latitudes, can give significantly different statistical results. In addition, it would be very informative if the authors could include penetration depth statistics for different missions and processing systems in Fig. 5-10. Such information would help readers better interpret differences in lower-tropospheric performance and assess the practical impact of the different algorithms.

Reply:
We agree that sampling differences across missions can influence the statistical comparisons. We have addressed this concern by explicitly discussing the distinct characteristics of COSMIC-2, including its low-inclination orbit and concentration of observations in the tropical/subtropical regions, which contribute to its higher variability and positive bias (Lines 522–527). To further account for sampling differences, we provided additional analysis in Section 5 (Fig. 16), where the comparison is restricted to a common latitude range (±45°) consistent with COSMIC-2 coverage. This allows for a more balanced inter-mission comparison.

In response to the reviewer's suggestion, Figures 5-10 have been revised to adopt a more distinct color scheme, that clearly distinguishes among missions and processing systems. In addition, profile penetration depths below 8 km have been included in Figures 5-7 for different missions and processing systems to aid interpretation of lower-tropospheric performance. As the penetration depth exhibits a similar behavior for refractivity, it is not shown in Figures 8-10.

Changes in manuscript:
We added the changes in lines 566-568. Figures 5-10 have been revised, and penetration depths have been included in Figures 5-7.

8: I found the results presented in Fig. 16b particularly interesting, as they show strong agreement between the STAR RFSI-processed C2, Spire and PlanetiQ datasets. This consistency is noteworthy given the findings in Anthes et al. (2025) (*https://doi.org/10.5194/amt-18-6997-2025*), which discuss the bending angle biases associated with the sideways sliding of the tangent point. That research suggests such sliding can introduce a slightly positive BA bias in C2 and typically negative biases in other high-inclination missions. I would appreciate the authors' interpretation of why the RFSI framework yields such high agreement across different missions.

Reply:
We thank the reviewer for this insightful comment. We agree that the strong agreement among the STAR RFSI-processed COSMIC-2, Spire, and PlanetiQ datasets shown in Fig. 16b is noteworthy, particularly in light of the findings of Anthes et al. (2025), which highlight mission-dependent bending angle differences associated with the sideways sliding of the tangent point.

While the STAR RFSI results exhibit improved cross-mission consistency, the underlying causes of this agreement are still under active investigation (Line 785). One contributing factor may be the treatment of the horizontal sliding of the tangent point within the RFSI framework. The differences in the definition of the occultation point (or georeferencing) can influence the magnitude of the sliding-related correction. As discussed in Anthes et al. (2025), UCAR defines the occultation point based on the location where the L1 excess phase exceeds 500 m, typically in the lower troposphere, whereas ROPP defines it as the location where the straight-line tangent altitude is reached (height of straight line equals 0), typically in the upper troposphere–lower stratosphere (UTLS). This sideways sliding effect can introduce a positive bias in COSMIC-2 bending angles of up to ~0.05%.

We added a brief discussion of these factors in the manuscript to provide context for the observed cross-mission consistency in the STAR RFSI results.

Changes in manuscript:
We added the changes in lines 888-896.

Specific comments:
1. L57-58: Can the authors be more specific about the "particular strength" of FSI, compared to other WO methods in resolving fine-scale structures?

Reply:
FSI relies on explicit signal localization to enhance vertical resolution and mitigate multipath by reducing spectral mixing, whereas CT2 achieves an implicit, physics-based localization in impact parameter space, enabling more robust separation of multipath contributions. As a result,

FSI is generally more sensitive to fine-scale atmospheric structures but also more noise-sensitive, while CT2 provides more stable retrievals in strong multipath conditions at the expense of reduced small-scale resolution.

Changes in manuscript:
We added the changes in lines 58-63.

2. L97: “ROEMX” -> “ROMEX”

Reply:
Agreed. Corrected.

Changes in manuscript:
The change is in line 107.

3. L113-114: As far as I know, C2, Spire and PlanetiQ are all using OL tracking, instead of the combination of OL and CL tracking. Can the authors recheck the fact and may readdress the statement?

Reply:
We have rephrased the statement to ensure accuracy regarding the latest tracking strategies of these missions. “However, recent GNSS radio occultation missions, including COSMIC-2, Spire, and PlanetiQ, primarily employ OL tracking throughout the occultation in order to maximize tracking robustness and data continuity across all atmospheric layers.” The general statement about the benefit of combining OL and CL tracking will be maintained as a conceptual point.

Changes in manuscript:
We added the changes in lines 133-138.

4. L170: “GNS” -> “GNSS”

Reply:
Agreed. Corrected.

Changes in manuscript:
The change is in line 207.

5. L171-172: "STAR RFSI algorithm has been integrated into ROPP version 10.0": Do you mean the STAR RFSI algorithm has been included in the ROMSAF ROPP package? If not, please rephrase the sentence. It's misleading.

Reply:
Agreed. We rephrased to clarify that "the STAR RFSI algorithm has been integrated into the ROPP version 10.0 customized at NOAA STAR."

Changes in manuscript:

The change is in line 209.

6. L214-217: Grammatical error. Rephase it.

Reply:
Agreed. The sentence was rephrased, the grammatical error was corrected.

Changes in manuscript:
The change is in line 252.

7. Section 3.2.1: I had thought that the navigation data modulation had been removed in the provided Level1b data, isn't it true?

Reply:
While Level 1b data providers typically remove the navigation bit, it is a best practice to correct for residual navigation bit jumps that may remain in the excess phase data. The model phase serves as a robust and necessary reference for identifying these $\pm\pi$ discontinuities (Lines 292–296).

8. The current presentation of Fig. 4 is somewhat confusing, as different months are represented by colors that appear identical.

Reply:
Agreed. Figure 4 is revised with a more distinct color scheme to clearly differentiate the monthly standard deviations.

Changes in manuscript:
Figure 4 was updated.

9. L465-466: So the RO data may be flagged as "bad" solely because "the model simulation data are unavailable"? This doesn't sound like a good QC criterion.

Reply:
The profile is not being judged as "bad" in the physical sense. The QC relies on the Observed-minus-Background (O-B) comparison. If the Background (B) simulation data is unavailable, the QC check cannot be performed, and the profile is conservatively flagged to ensure that only profiles fully validated against the ERA5 reference are retained. This condition has been removed to avoid potential confusion in the text.

Changes in manuscript:
We added the changes in lines 523-529.

10. L743-745: could the authors elaborate on the statement regarding “degraded L2 signals and reduced GNSS SNR above ~20 km”?

Reply:

We thank the reviewer for pointing this out. There is a typo in the original text: "above" should be clarified to refer specifically to the ~20–25 km altitude range, and we will revise the wording accordingly. The L2 signal is inherently weaker than L1 and therefore more susceptible to noise. In the 20–25 km region, although the neutral atmospheric signal is already relatively small, the L2 signal quality can degrade due to its lower SNR. This leads to increased uncertainty in the ionospheric correction, which relies on the combined use of L1 and L2 measurements. As the FSI technique is more sensitive to measurement noise, these effects are amplified, resulting in degraded retrieval quality in this altitude range.

Changes in manuscript:
We added the changes in lines 840-841.

EC comments:

The authors say in lines 772-774 "Anthes et al. (2025) reported that the UCAR-processed COSMIC-2 bending angles included in ROMEX exhibit a positive bias of approximately +0.1-0.15% relative to ERA5 in the lower stratosphere, larger than the biases seen for Spire and other ROMEX datasets." This statement by itself is incomplete and misleading. In the Anthes et al. (2025) preprint and in the final published version *https://doi.org/10.5194/amt-18-6997-2025* it is made clear that the COSMIC-2 "bias" of 0.1 to 0.15% in the lower stratosphere (specifically between 10 and 30 km) is mostly a representativeness difference and not a true bias, and is caused by the different orbits of COSMIC-2 and the other ROMEX missions around the non-spherical Earth and the associated varying radius of curvature. This is described in detail in Section 5.2 of the paper as the *azimuth effect* and results in most (about 0.1%) of the apparent bias. Because it is a representativeness difference and not a true bias, it does not affect data assimilation in models. The remaining small part of the apparent bias (less than 0.05%) is due to the sideways sliding of the occultation plane and can be easily corrected in the processing of the RO data by applying a correction to the impact height.

Reply:
We thank the EC for the comments on this important clarification and agree with the interpretation provided in Anthes et al. (2025). The statement in Lines 772–774 will be revised to more accurately reflect the findings of that study. Specifically, we will clarify that the reported 0.1–0.15% difference in COSMIC-2 bending angles in the lower stratosphere (10–30 km) is predominantly a representativeness difference rather than a true systematic bias. As described in Section 5.2 of Anthes et al. (2025), this effect is primarily attributed to orbit-related sampling differences over a non-spherical Earth (the azimuth effect), which accounts for the majority (~0.1%) of the apparent discrepancy.

We noted that, because this is a representativeness effect, it does not significantly impact data assimilation in numerical weather prediction (NWP) systems. The remaining small component of the difference (less than 0.05%) is attributable to the sideways displacement of the occultation plane and can be mitigated by appropriate corrections to the impact height during RO data processing.

The revised text will explicitly distinguish between representativeness differences and true biases to avoid potential misunderstanding.

Changes in manuscript:
We added the changes in lines 874-879.