May 21, 2026

Dr. Richard Anthes
Editor

Dear Dr. Anthes,

We sincerely appreciate the editor and the reviewers for their constructive comments and insightful suggestions provided during the second round of reviews, which have significantly enhanced the clarity and quality of our manuscript. This document provides a detailed point-by-point response to the concerns raised.

Response to Editor (Richard Anthes)

- **Minor Revisions and Evidence:** We agree with the reviewer's point that the claim regarding the superior fine-scale resolution of FSI is not universally valid and depends on implementation details. Providing definitive quantitative evidence for this claim is challenging. To address this, we have modified the text in the manuscript to represent a more balanced view and removed the better vertical resolution claim in other places.
- **Figure 4 Deletion:** Per your suggestion and the recommendation of Reviewer 3, Figure 4 has been removed to streamline the manuscript.
- **Formatting of Figures 11-15:** The x-axis for Figures 11-15 have been standardized to ensure consistency, but we keep the y-axis since the right panels in Figures 11-15 show the number of common (matched) profiles used for the structural uncertainty analysis, not the penetration depth profiles.

Response to Reviewer #2

Most of my comments were addressed satisfactorily. Thank you!

One point that remains a strong concern to me is the use of the MSIS model to remove the navigation data bits. I believe this could introduce a great deal of uncertainty in the lower troposphere. Was this done for all the missions processed by STAR? Did you compare with processing done by other centers (like UCAR) where the navigation data bits were removed internally based on knowledge of the data bits? At the very least, this uncertainty needs to be clearly acknowledged in the paper.

Reply:

Although the MSIS model is used to calculate the reference phase for these corrections, it introduces a known source of uncertainty in the lower troposphere. This uncertainty highlights the critical need for an external navigation bit stream, since standard navigation-bit-free cycle slip correction routines (such as those in ROPP) often struggle with atmospheric multipath in this region. To mitigate this issue, processing is often terminated above a 7-10 km impact height. However, this challenge can be circumvented by using pilot signals (Galileo E1C and E5BQ, GPS L2C, and BeiDou B1CP and B2AP), which are not modulated with navigation bits and can therefore yield reliable results in the lower troposphere even without external bit stream data (Jonathan Brandmeyer, personal communication, June 2025).

Changes in manuscript:
The changes are in lines 357-366

A more minor point, but I am also not convinced that FSI can resolve “high-frequency, fine-scale structures” better. Could you point me to specific examples in the cited papers that demonstrated this?

Reply:
We agree with the reviewer’s point that the claim about FSI’s superior fine-scale resolution is not universally valid and depends on implementation details. Providing definitive quantitative evidence for this claim is challenging. To address this, we have modified the manuscript text to represent a more balanced view and removed the claim of better vertical resolution elsewhere.

The revised statement reads: “While FSI is designed to leverage full-spectrum information for potentially enhanced sensitivity to fine-scale atmospheric structures, it is also generally more noise-sensitive. CT2, in contrast, provides more stable retrievals in strong multipath conditions, sometimes at the expense of reduced small-scale resolution. Consequently, the effective vertical resolution of both methods is highly dependent on specific implementation and tuning choices, such as filter settings and truncation strategies.”

Changes in manuscript:
The changes are in lines 60-66, and 96-100.

Response to Reviewer #3

I appreciate the author’s efforts in revising the manuscript and responding to prior comments. However, the current version and the responses further reinforce my concern that this work is more suitable as a technical report rather than a journal publication, primarily due to a lack of clear novelty.

The authors explicitly state that “the core FSI algorithm remains consistent with that described in Adhikari et al. (2021)”, while also claiming that “the novelty of the present work lies in the development and presentation of the complete STAR RFSI end-to-end processing framework”. I find this claim unconvincing. The complete STAR RFSI processing framework appears to have already been well documented in Adhikari et al. (2021), as evidenced by the strong similarity between Fig. 4 in that paper and Fig. 1 in the current manuscript. Aside from differences in QC schemes and the discussion of BA uncertainty in Section 3.2.4, the manuscript doesn’t demonstrate substantial new methodological or conceptual contributions. Relatedly, the authors should cite Liu et al. (2018), which already discussed the methodology used for BA uncertainty estimation. More broadly, given that FSI has been established for over two decades, reiterating the method in detail without clear advancement reduces the scientific contribution of the manuscript.

The authors also state that “the value of this study lies in demonstrating the structural uncertainty among different processing methods”. However, the manuscript does not clearly identify the sources of these structural differences. It remains unclear whether they arise from differences in WO methods, QC procedures, data truncation strategies, or other aspects of the processing chain. This lack of clarity weakens the interpretability of the results.

Overall, while the presented results may still be useful, they are more appropriate for a technical report or conference proceeding than for a scientific journal.

Reply:
We appreciate the reviewer’s detailed and critical assessment of the manuscript's claims regarding novelty and the analysis of structural uncertainty. We offer the following clarifications and planned revisions:

1. On Novelty and Scientific Contribution
We agree that the core FSI algorithm is consistent with the methodology detailed in Adhikari et al. (2021). The novelty of the present work is not in the invention of the FSI algorithm, but in the successful development and systematic application of the complete STAR RFSI end-to-end processing framework and its subsequent inclusion in the ROMEX intercomparison as an independent data source.

The scientific contribution is distinct from previous FSI work in three key ways. (1) Systematic Multi-Mission Application: This study provides the first comprehensive, cross-mission validation of the STAR RFSI system using a large and diverse ROMEX dataset (including commercial and government-funded missions) against other major community processing centers (such as EUMETSAT and UCAR). (2) Quantification of Structural Uncertainty: The primary objective and scientific value of this study is the quantification of the structural uncertainty among different processing approaches, FSI, STAR CT2, and the EUMETSAT ROM SAF ROPP CT2, under a ROMEX framework. This analysis provides critical metrics for interpreting ROMEX forecast impact studies. (3) Operational Documentation: The manuscript documents the complete STAR RFSI processing chain (including data preprocessing, quality control, and statistical optimization), which is essential for transparency and future operational use within NOAA STAR.

We revised the introductory statement in Section 3 to more clearly define the novelty as the systematic validation and structural uncertainty analysis of the STAR RFSI system within the ROMEX framework, rather than the core FSI algorithm itself.

Changes in manuscript:
The changes are in lines 202-219.

2. On Citation of Bending Angle Uncertainty

We appreciate the reviewer pointing out the relevant reference. We updated Section 3.2.4 to include a proper citation for Liu et al. (2018), which discusses the methodology used for estimating bending angle uncertainty. The existing text states that the methodology is “similar to Liu et al. (2018)”.

Changes in manuscript:
The change is in line 477.

3. On Reiterating FSI Methodology

The detailed description of the STAR RFSI processing (Section 3 and Figure 1) is included to ensure full transparency and reproducibility, particularly for the ROMEX community. Since the structural uncertainty analysis is a core component of this study, it is crucial to clearly define the specific implementation details that differentiate the STAR RFSI method from the STAR ROPP CT2 and EUMETSAT methods. These details include: (1) SNR-Based Truncation: The SNR-based thresholds and cut-off procedures in Section 3.1. (2) Filtering Choices: The customized smoothing windows (e.g., 0.05 seconds for 100 Hz data below 10 km) that are key differentiators from other algorithms. (3) GO/WO Transition: The application of a sliding polynomial filter below 25 km and an optimal estimation filter above 25 km, which can introduce systematic offsets.

Detailing these steps is essential for the interpretation of the results, as these choices are the precise cause of the structural differences identified in Section 4.3.

4. On Sources of Structural Differences

We acknowledge that a clearer, consolidated summary of the sources of structural differences will greatly enhance the interpretability of our results. The manuscript does discuss these sources throughout Section 4.

We revised Section 4 to explicitly summarize the penetration depths of RO profiles regarding Metop and TanDEM-X: "The comparatively shallower penetration depths observed for Metop (in Fig. 5) and TanDEM-X (in Fig. 6) relative to missions like COSMIC-2 and PlanetiQ are linked to a combination of mission-specific characteristics (specifically SNR) and the data truncation strategies employed by the processing algorithm: 1) Impact of lower SNR: Penetration depth is fundamentally dependent on the signal cutoff criteria applied during processing, and higher SNR is known to improve tropospheric penetration. Metop-B/C and TanDEM-X exhibit lower mean SNRs (e.g., Metop-B GPS SNR is 730.2, TanDEM-X is 549.3) compared to the missions with the deepest penetration (COSMIC-2 at 1295.1 and PlanetiQ at 1440.5). This lower SNR makes them more susceptible to noise and multipath effects in the lower troposphere, leading to a shallower signal cutoff. 2) Truncation strategy: The systematic signal truncation used in the STAR RFSI method is based on SNR thresholds to prevent noise-contaminated signals from propagating. For lower-SNR missions like Metop and TanDEM-X, these thresholds are reached at higher altitudes, resulting in shallower penetration. 3) Processing algorithm variation: The choice of processing algorithm also contributes to variation. The CT2 method, used in both ROPP systems (STAR ROPP and ROM SAF ROPP), is less sensitive to SNR and resulted in generally narrower penetration depth profiles below 1 km compared to RFSI. This is further reflected by the fact that Metop-B/C penetration is substantially improved when processed by EUMETSAT's CT2 compared to STAR's algorithms, suggesting differences in the specific configuration parameters used by the processing centers."

Regarding identifying the sources of these structural differences, we added: “This divergence in STD differences near 25 km for Metop-B/C, compared to COSMIC-2, PlanetiQ, and Spire, suggests that the upper-atmospheric retrievals for Metop-B/C are more sensitive to processing methodology. This is supported by the data showing that in the upper stratosphere (above 25 km), the EUMETSAT algorithm exhibits the lowest variability. Given that Metop-B/C are EUMETSAT missions and their data are routinely processed by them, this superior performance suggests that the EUMETSAT processing methodology is likely fine-tuned and highly stable for their mature sensor platforms, which allows for effective mitigation of noise and uncertainties at high altitudes, resulting in the lowest variability observed across all datasets.”

Changes in manuscript:
The changes are in lines 697-715, and 880-889.

Additional comments:
1. The manuscript repeatedly claims that FSI is more suitable for resolving fine-scale atmospheric structures than other WO methods. However, no quantitative evidence is provided. For example, in Lines 60-63, can the authors provide an estimate of the effective vertical resolution for FSI compared to CT2? I do not believe this claim is universally valid, as it strongly depends on implementation details. A well-tuned CT2 can preserve fine structure very well, whereas a conservatively configured FSI can look smoother than CT2.

Reply:
We agree with the reviewer’s point that the claim about FSI’s superior fine-scale resolution is not universally valid and depends on implementation details. Providing definitive quantitative evidence for this claim is challenging. To address this, we have modified the manuscript text to represent a more balanced view and removed the claim of better vertical resolution elsewhere.

The revised statement reads: “While FSI is designed to leverage full-spectrum information for potentially enhanced sensitivity to fine-scale atmospheric structures, it is also generally more noise-sensitive. CT2, in contrast, provides more stable retrievals in strong multipath conditions, sometimes at the expense of reduced small-scale resolution. Consequently, the effective vertical resolution of both methods is highly dependent on specific implementation and tuning choices, such as filter settings and truncation strategies.”

Changes in manuscript:
The changes are in lines 60-66, and 96-100.

2. The authors have not adequately addressed my previous comment regarding the WO method used by EUMESAT. Specifically, is the method CT2, phase matching or another approach? If EUMESAT also uses CT2, it would be particularly valuable to compare retrievals using ROPP CT2 vs. STAR ROPP CT2. In addition, when comparing the penetration depths of RO profiles (Fig. 6 vs. Fig. 7), certain missions stand out, Metop in Fig. 6 and TanDEM-X in Fig. 7 appear to have worse penetration than others. The manuscript does not provide an explanation for this. Is this behavior related to differences in WO methods, QC procedures, data truncation strategies, or mission-specific characteristics?

Reply:
The EUMETSAT-processed data (excluding COSMIC-2) utilized ROM SAF ROPP as the internal processing system, employing the CT2 method in the troposphere. Note that while STAR ROPP and EUMETSAT ROPP both use the CT2 method in the troposphere, their specific configuration parameters may differ.

The comparatively shallower penetration depths observed for Metop (in Fig. 5) and TanDEM-X (in Fig. 6) relative to missions like COSMIC-2 and PlanetiQ are linked to a combination of mission-specific characteristics (specifically SNR) and the data truncation strategies employed by the processing algorithm: 1) Impact of lower SNR: Penetration depth is fundamentally dependent on the signal cutoff criteria applied during processing, and higher SNR is known to improve tropospheric penetration. Metop-B/C and TanDEM-X exhibit lower mean SNRs (e.g., Metop-B GPS SNR is 730.2, TanDEM-X is 549.3) compared to the missions with the deepest penetration (COSMIC-2 at 1295.1 and PlanetiQ at 1440.5). This lower SNR makes them more susceptible to noise and multipath effects in the lower troposphere, leading to a shallower signal cutoff. 2) Truncation strategy: The systematic signal truncation used in the STAR RFSI method is based on SNR thresholds to prevent noise-contaminated signals from propagating. For lower-SNR missions like Metop and TanDEM-X, these thresholds are reached at higher altitudes, resulting in shallower penetration. 3) Processing algorithm variation: The choice of processing algorithm also contributes to variation. The CT2 method, used in both ROPP systems (STAR ROPP and ROM SAF ROPP), is less sensitive to SNR and resulted in generally narrower penetration depth profiles below 1 km compared to RFSI. This is further reflected by the fact that Metop-B/C penetration is substantially improved when processed by EUMETSAT's CT2 compared to STAR's algorithms, suggesting differences in the specific configuration parameters used by the processing centers.

Changes in manuscript:
The changes are in lines 666-668, and lines 697-715.

3. I also find the response to my previous comment on ERA5 dependence unsatisfactory. My concern is that RFSI applies QC using ERA5 forecast, which are not fully independent from ERA5 reanalysis. Since the reanalysis is used as the reference for evaluation, this raises the possibility of a systematic bias whereby RFSI retrievals may appear to agree more favorable with ERA5. The statement that "the impact of small-scale ERA5 errors on the QC is minimized by employing a very high threshold" requires clarification. What exactly is meant by "small-scale errors in ERA5" in this context? The QC is applied to RO observations, not the other way. The explanation as written is not sufficiently clear to me. Furthermore, since Spire and PlanetiQ data are processed by EUMESAT, it would be more appropriate to compare RFSI-processed QC rates against EUMESAT-processed ones rather than CDAAC-processed ones.

Reply:
We appreciate the reviewer's insistence on clarifying the potential for systematic bias related to our QC procedure's reliance on ERA5. We acknowledge the reviewer's concern regarding the non-independence of the ERA5 forecast fields (used in the STAR RFSI/CT2 QC) and the ERA5 reanalysis (used as the reference for evaluation). While the two fields are intrinsically linked, the design of our QC procedure minimizes the possibility of systematic bias in the evaluation results:

1) High threshold for outlier removal: Our QC flags a profile as "bad" only if the fractional bending angle difference exceeds a very high threshold of $7\sigma_{year}$ between 10 and 40 km. This threshold is intentionally conservative, designed to remove only gross errors (e.g., severe processing failures, cycle slips, or physically unrealistic profiles) rather than subtle atmospheric features or small systematic biases. A profile passing this QC threshold primarily indicates successful processing, not necessarily high-level agreement with ERA5.

2) Limited altitude range: The QC is strictly applied only in the 10-40 km range, where RO data quality is highest and ERA5-observation agreement is strongest. The key statistical evaluations in this study (mean bias, standard deviation) are performed across the entire vertical profile (0-40 km), so the QC does not directly bias the lower troposphere results (below 10 km), where the largest O-B discrepancies are observed and structural uncertainty is most pronounced.

3) Focus on evaluation, not correction: The QC is used to filter out corrupted data before evaluation, ensuring a fair intercomparison of processing algorithms. It does not apply any model-dependent correction or statistical optimization (which uses a background model) that would force the RO retrieval towards the ERA5 state.

In summary, the highly restrictive nature of the QC threshold ensures that the selection process is largely immune to the minimal differences between the ERA5 forecast and the ERA5 reanalysis, preventing the QC from systematically introducing an artificial bias that favors agreement with the ERA5 reference.

I want to clarify on “small-scale ERA5 errors” in the previous response. The phrase “small-scale ERA5 errors” refers to errors in atmospheric variability that the global ERA5 model does not fully resolve at the specific time and location of the RO observation.

RO observation data are known to be highly sensitive to fine-scale atmospheric structures, while the global model fields, such as ERA5, represent the atmosphere with relatively smooth fields at a finite resolution. The difference between the actual high-resolution atmospheric structure at the tangent point from RO data and the smoother, model-resolved structure constitutes the “representativeness error”, or what we termed “small-scale ERA5 errors”. By setting the QC threshold so high ($7\sigma_{year}$), we ensure that the QC procedure is not triggered by these small-scale atmospheric features or model-observation mismatches that fall within the expected noise or representativeness error. Instead, the QC focuses on identifying outliers caused by instrumental or processing failures.

Since the QC is an external check applied to RO observations to identify gross failures, the high threshold allows us to retain profiles containing valid small-scale features that may slightly deviate from the ERA5 forecast (due to model smoothing) yet remain physically sound.

Regarding the QC comparison baseline

Comparison of STAR RFSI QC rates with CDAAC QC rates was necessary given the specific data available in the ROMEX distribution. The ROMEX dataset provided by EUMETSAT did not include quality control applied to the Spire and PlanetiQ data. Consequently, a direct comparison of STAR RFSI QC rates against EUMETSAT-processed QC rates for these missions was not possible. CDAAC-processed data, widely recognized and used by the community, provided the best available baseline for contextualizing the STAR RFSI QC rates.

To ensure a consistent inter-algorithm evaluation for the remainder of the study (Section 4), we applied our standard QC procedure to the EUMETSAT dataset, where QC was absent, thereby allowing a valid comparison across the different processing centers' results.

Changes in manuscript:
The changes are in lines 555-558

4. I still find Fig. 4 difficult to interpret, particularly in distinguishing differences between months. If the monthly variability in STD is not statistically significant, the figure may not be necessary or should be redesigned for improved clarity.

Reply:
Figure 4 was removed

5. Regarding Fig. 11-15, it is unclear whether EUMETSAT applies consistent processing algorithms across all ROMEX missions? The results tend to suggest that processing may be fine-tuned for Metop data. For example, kink structure near 25 km appear in the STD differences for Spire and PlanetiQ, but not for Metop. This raises questions about whether EUMESAT consistently applies WO processing for the whole profiles for Metop but uses a combination of GO and WO methods for other missions. Can the authors verify this? In terms of figure quality, I recommend:

- Making the x-axis presentation (relative BA difference, left panels) consistent across figures;
- Adjusting the y-axis range (e.g. 0-10 km) in the right panels to better highlight penetration depth differences.

Reply:
We appreciate the reviewer's detailed observations regarding Figures 11-15 and their suggestions for improving figure quality. This divergence in STD differences near 25 km for Metop-B/C, compared to COSMIC-2, PlanetiQ, and Spire, suggests that the upper-atmospheric retrievals for Metop-B/C are more sensitive to processing methodology. This is supported by the data showing that in the upper stratosphere (above 25 km), the EUMETSAT algorithm exhibits the lowest variability. Given that Metop-B/C are EUMETSAT missions and their data are routinely processed by EUMETSAT, this superior performance suggests that the EUMETSAT processing methodology is likely fine-tuned and highly stable for their mature sensor platforms, enabling effective mitigation of noise and uncertainties at high altitudes, and resulting in the lowest variability observed across all datasets.

Changes in manuscript:
The changes are in lines 880-889

We agree with the reviewer’s recommendations for improving the consistency and clarity of Figures 11-15.

X-axis Consistency: The x-axis presentation for the relative bending angle (BA) difference in the left panels was made consistent across all figures, using a fixed range (e.g., from -1.5% to +1.5%).

Y-axis Range (Right Panels): The suggestion to adjust the y-axis range in the right panels to better highlight differences in penetration depth requires clarification. The right panels in Figures 11-15 show the number of common (matched) profiles used in the structural uncertainty analysis, not the penetration-depth profiles (which are shown in the right panels of Figures 5-7). Since the profile count must be identical across all three processing methods in a matched-profile comparison, the current y-axis range will be retained to accurately reflect the same number of matched observations.

Changes in manuscript:
Figures 11-15 (now Figures 10-14) were updated.

L879: ROPP or RFSI? as no ROPP results are discussed here.

Reply:
We replace ROPP with RFSI to make it clearer. Since the FSI algorithm was implemented into the customized STAR ROPP version 10.0, both the STAR RFSI and STAR ROPP (CT2) processing systems share the same consistent occultation point definition.

Changes in manuscript:
The change is in line 955.