

## **2<sup>nd</sup> review of “Processing multiple GNSS RO data using FSI and ROPP: Results from the ROMEX” by Chen et al.**

I appreciate the authors' 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.

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.
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?

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 favorably 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.
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.
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 EUMETSAT 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.

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

### References:

Liu, H., Y.-H. Kuo, S. Sokolovskiy, et al., A quality control procedure based on bending angle measurement uncertainty for radio occultation data assimilation in the tropical lower troposphere, *J. Atmos. Oceanic Tech.*, 35(10), 2117-2131, 2018, doi: 10.1075/JTECH-D-17.0224.1.