

Response to Reviewer 2

The authors thank Anonymous Reviewer #2 for his or her time in reviewing our manuscript, and for their helpful comments. Addressing them has enabled us to improve this manuscript. Our point-by-point responses are indicated in blue below.

This manuscript presents a summary of Hurricane Analysis and Forecast System (HAFS) experiments to test the impact of assimilating different radio occultation (RO) missions – and thereby more/less RO observations – into the 12x12-degree inner mesh of the model. These experiments, forecasting four Atlantic-basin tropical cyclones (TCs) from 2022, consist of a control, a data-omission experiment, an enhanced data experiment, and a data-degraded experiment. In the control, all previously-available government-owned RO data were assimilated; COSMIC-2 was omitted from this set for the omission experiment; ROMEX commercial RO data were included with the control data for the enhanced experiment; and all RO data below 5 km was omitted for the degraded experiment. The authors find that, for this sample of TCs, the extra commercial RO observations, especially those below 5 km, improve the short- to mid-range forecast performance.

One of the key questions raised in the justification of ROMEX was what the forecast impact would be of a sizeable increase in the number of available RO profiles, particularly for the lower troposphere. This study adds valuable information toward answering that question. The authors nicely show that, in one TC (Fiona), the increased sample density nudges the model to dry the near eyewall environment and thereby inhibit vortex enhancement that the control simulation resolves. Further, by eliminating RO data below 5 km, the authors show evidence that RO observations in the lowermost troposphere provide positive impact to the forecasts.

I overall find the work to be of high quality and significance, but have concerns about some of the ways the authors pose their experiments, the discussion or omission of assimilation counts, the completeness of the analysis, and the clarity of the discussion.

Major comments:

1. The two denial experiments need more justification for why the particulars of each setup were chosen.

Why would one want to withhold only COSMIC-2 and not, say COSMIC-2 and MetOp or perhaps all RO? Are there particular questions in the community about the value of COSMIC-2 that need to be answered? If so, the authors should introduce and discuss these questions.

Yes, given uncertainty about whether a COSMIC-2 follow-up mission will be launched and our limited computational resources, we thought that it would be most beneficial to the scientific community to evaluate impacts from withholding COSMIC-2 rather than run a “no-RO” experiment.

We also revised the first sentence of Section 2d when introducing the HAFS experiment design, and it now read as (lines 279-281):

“To demonstrate the impacts of assimilating the ROMEX commercial observations, COSMIC-2, and all lower-tropospheric RO data, four independent HAFS-A observing system experiments (OSEs) are run as follows:”

We also added this sentence to the Conclusions (lines 789-793):

“Additionally, by showing Control_noC2’s short-range intensity and medium-range track forecast degradations relative to Control, this study supports previous work showing positive impacts of COSMIC-2 assimilation on regional model TC forecasts (Miller et al. 2023; Teng et al. 2023) and the need to sustain COSMIC-2-like coverage over the tropics in future RO satellite missions.”

We expect the other government-owned RO missions to have limited impact on HAFS tropospheric water vapor, temperature, and TC intensity forecasts, due to GSI rejecting all Metop bending angles below 8 km (described in the paragraph below), as well as the very small profile counts for the other federal missions, i.e., *TerraSAR-X*, *TanDEM-X*, and *KOMPSAT-5*.

Our HAFS-GSI version, like the GSI implementations used for many years with the GFS, imposes a hard reject of all Metop bending angles below the 8-km impact height because previous work found these observations to have large background departures against the GFS (Dutta et al. 2019). Therefore, since the Metop-B/C profiles assimilated in our study have their the lower-to-middle tropospheric portions below ~ 650 hPa rejected by GSI, we expect the Metop-B/C profiles to have very limited impact on the HAFS lower-to-middle tropospheric temperature and water vapor analysis fields.

On Dec 4, 2024, HAFS-GSI removed the QC check for rejecting Metop data (satellite platform ID 3, 4, or 5) below the 8-km impact height, as shown in the program history log for the latest version of the RO bending angle forward operator code:

[https://github.com/hafs-](https://github.com/hafs-community/GSI/blob/e7c12d84066ddbfaceaca7bec69450d1d21d4937/src/gsi/setupbend.f90)

[community/GSI/blob/e7c12d84066ddbfaceaca7bec69450d1d21d4937/src/gsi/setupbend.f90](https://github.com/hafs-community/GSI/blob/e7c12d84066ddbfaceaca7bec69450d1d21d4937/src/gsi/setupbend.f90)

We are happy to see that NCEP is now allowing assimilation of sub-8 km Metop data that passes the other QC checks, given that the prior work showing poor Metop O-B fit was done years ago and may no longer be a problem today. In future HAFS RO OSE studies, we plan to remove the sub-8km Metop hard reject from the forward operator code (or simply use NCEP’s most recent public release of HAFS-GSI code) and evaluate the O-B, O-A, and forecast impact statistics of the Metop observations.

The expected limited forecast impacts of Metop profiles should be mentioned in the text, and so we added these sentences to Section 2c when introducing the ROMEX dataset (lines 267-272):

“Prior to a December 2024 code upgrade, GSI implementations with NCEP’s forecasting models rejected all Metop bending angles below the 8-km impact height (~ 650 hPa) based on prior work finding larger Metop-A/B departures against the GFS background (Dutta et al. 2019). Since our HAFS-GSI build predates removal of this QC check, the assimilated Metop-B/C profiles are expected to have little impact on HAFS-A forecast lower-to-middle tropospheric temperature and water vapor fields.”

We added the Dutta et al. (2019) citation below to the References section.

Dutta, S., Vandenberghe, F., Shao, H., Zhang, H., and J. Doe, 2019: Assessment of radio-occultation from multiple GNSS platforms: Operational and non-operational. *JCSDA Quarterly*, **62**, 33-37, <https://doi.org/10.25923/w2dh-ep66>.

Given the repeated points the authors make comparing low signal-to-noise (SNR) observations from Spire with those from COSMIC-2, is it because they are trying to make statements about the value of high SNR observations? This needs clear discussion if so. As well, as I point out in later a comment, PlanetiQ produce high SNR observations and are selectively omitted when the authors discuss the SNR.

We decided to remove discussion relating differences between the Spire and COSMIC-2 outlier rejection statistics and diagnosed observation errors to SNR from this manuscript. Our reasons are as follows. First, SNR information is not available in the ROMEX RO observation BUFR files, so we cannot directly quantify the contribution of individual observations' SNR to the data assimilation statistics reported in Section 3. Second, this is a complex topic that deserves more detailed treatment in a different study. For example, COSMIC-2's antenna geometry causes its SNR to have latitudinal dependence (Ho et al. 2023), and SNR is only one of many contributors to RO retrieval uncertainty in the lower troposphere.

The sentence in Section 3a beginning with "This is an interesting result..." was removed, and the discussion of Spire's diagnosed observation error statistics in the Conclusions now reads as (lines 732-735):

"Spire provided the bulk of the ROMEX observations – about 16750 per day globally, and their observation error standard deviation diagnosed from HAFS-A O-B and O-A statistics was found to be smaller than that of COSMIC-2 in most layers of the lower-to-middle troposphere."

Reference:

Ho, S.-P., X. Zhou, X. Shao, Y. Chen, X. Jing, and W. Miller, 2023: Using the commercial GNSS RO Spire data in the neutral Atmosphere for climate and weather prediction studies. *Remote Sens.*, **15(19)**, 4836, <https://doi.org/10.3390/rs15194836>.

If this is a means of producing a "no RO" experiment, this is obviously an imprecise setup for doing that. Though COSMIC-2 dominates the Control sampling, there are a non-trivial quantity of other government-provided data. Consider Figure 2: there are O(100) unique COSMIC-2 occultations assimilated for Hurricane Ian (2022). Given the global sampling values given in Section 2c, we would expect there to be O(10) unique occultations from remaining missions in the Control_noC2 experiment for this storm. Healy et al. (2005) demonstrated that even 160 CHAMP profiles/day could improve a global forecast; O(10) in a 12x12-degree mesh will not have a trivial impact.

As we mention above, we are more interested in showing what happens in HAFS when COSMIC-2 data are removed, given ongoing questions regarding whether a "COSMIC-3" follow-up mission will be funded. We agree that the other non-COSMIC-2 government missions

including Metop observations above 8 km could have some impact on HAFS track and intensity forecasts but given our limited computational resources we decided not to run an additional sensitivity experiment isolating the impacts of these observations.

For the ROMEX_noLowlevRO experiment, what motivated the choice of 5 km impact height as the threshold below which RO data are withheld?

HAFS-GSI uses the 5-km impact height threshold in its super-refractivity (SR) QC check, whereby bending angles with impact heights below 5 km are checked for background SR conditions (defined as the vertical N-gradient exceeding 0.75 times the -157 N-units/km threshold) and rejected if they occur within or near a background SR layer. Therefore, given that the rejection rate ramps up for bending angles with impact heights below 5 km, we can use the ROMEX_noLowlevRO experiment as a way to assess the usefulness of the observations in this layer that survive HAFS QC screens and motivate future work to optimize parameters in the GSI RO data QC algorithm. Ideally, if we had more computational resources, we could have run several HAFS sensitivity experiments that varied the hard-reject height threshold over a range such as 3 km, 5 km, and 7 km. Since the 5-km impact height tends to be near the 3-km geometric height, we consider it to be a reasonable way to define the top of the moist lower troposphere, where many kinematic and thermodynamic processes important to TC intensity changes occur but where RO observation uncertainty is also higher due to water vapor effects.

We added this sentence to Section 2d explaining our reasons for choosing 5-km as the lower tropospheric height cutoff threshold (lines 288-291):

“The 5-km height threshold is chosen for ROMEX_noLowlevRO because it serves as a reasonable proxy for the top of the moist lower troposphere where many processes important to TC intensification occur (Montgomery and Smith 2014), and because HAFS-GSI checks RO bending angles below this height for background SR conditions (Section 2b).”

2. The assimilation counts for each experiment and each storm should be presented.

One can get a sense of the RO data counts for Hurricane Ian (2022) from Figure 2, but this is only one of the storms, and a graphical presentation is not the easiest to interpret for understanding, e.g., the actual increase of observations between the Control and ROMEX experiments. A table of values would be quite valuable for the reader, especially for appreciating how the data quantity has changed in the ROMEX experiment for Hurricane Fiona, the sole subject of section 5.

Related, in the Summary, it is appropriate to give the total global counts of the government-owned and commercial datasets, but it does not give the correct context for this work that considers a moving 12x12-degree grid in the Atlantic basin. It would be appropriate to also give the total occultation counts that were actually assimilated in this work.

Thank you for this excellent suggestion – and we agree. We added a new Table 1 to the manuscript (copied below) that lists the number of profiles assimilated in Control_noC2,

Control, and ROMEX for each TC case. This provides additional details on assimilation counts that could be helpful when interpreting our results.

N _{eye} Date (n _{eye})	Earl			Fiona			Ian			Julia		
	noC2	Control	ROMEX	noC2	Control	ROMEX	noC2	Control	ROMEX	noC2	Control	ROMEX
03 Sep (3)	1	19	73	—	—	—	—	—	—	—	—	—
04 Sep (4)	2	24	92	—	—	—	—	—	—	—	—	—
05 Sep (4)	4	22	83	—	—	—	—	—	—	—	—	—
06 Sep (4)	4	22	86	—	—	—	—	—	—	—	—	—
07 Sep (4)	4	17	69	—	—	—	—	—	—	—	—	—
08 Sep (4)	3	17	101	—	—	—	—	—	—	—	—	—
09 Sep (4)	3	27	97	—	—	—	—	—	—	—	—	—
10 Sep (4)	4	13	101	—	—	—	—	—	—	—	—	—
15 Sep (4)	—	—	—	4	17	87	—	—	—	—	—	—
16 Sep (4)	—	—	—	6	30	104	—	—	—	—	—	—
17 Sep (4)	—	—	—	4	14	80	—	—	—	—	—	—
18 Sep (4)	—	—	—	1	24	97	—	—	—	—	—	—
19 Sep (4)	—	—	—	6	40	85	—	—	—	—	—	—
20 Sep (4)	—	—	—	5	27	101	—	—	—	—	—	—
21 Sep (4)	—	—	—	3	29	100	—	—	—	—	—	—
22 Sep (4)	—	—	—	5	28	102	—	—	—	—	—	—
23 Sep (4,3)	—	—	—	5	30	82	4	29	58	—	—	—
24 Sep (4,4)	—	—	—	6	10	96	9	42	94	—	—	—
25 Sep (4)	—	—	—	—	—	—	4	31	79	—	—	—
26 Sep (4)	—	—	—	—	—	—	4	19	58	—	—	—
27 Sep (4)	—	—	—	—	—	—	6	18	84	—	—	—
28 Sep (4)	—	—	—	—	—	—	5	25	105	—	—	—
29 Sep (4)	—	—	—	—	—	—	4	21	103	—	—	—
30 Sep (4)	—	—	—	—	—	—	5	25	84	—	—	—
01 Oct (2)	—	—	—	—	—	—	5	12	61	—	—	—
06 Oct (2)	—	—	—	—	—	—	—	—	—	3	15	33
07 Oct (4)	—	—	—	—	—	—	—	—	—	6	38	56
08 Oct (4)	—	—	—	—	—	—	—	—	—	3	24	40
09 Oct (2)	—	—	—	—	—	—	—	—	—	1	8	22
Total	25	161	702	45	249	934	46	222	726	13	85	151

Table 1. Number of GNSS RO profiles assimilated for the Control_noC2 (noC2), Control, and ROMEX experiments in the $12^\circ \times 12^\circ$ HAFS-A moving nest for each day that overlaps a TC case cycling period. Dashes indicate that no HAFS DA cycles occurred for that storm on that day, as identified by the column and row respectively. Numbers in parentheses list the number of DA cycles for a given day, with the maximum value of four indicating that observations were assimilated in the 00, 06, 12, and 18 UTC cycles. The comma-delimited values for 23 and 24 September indicate that the HAFS-A nest assimilated profiles for both the Fiona and Ian cases, with the Fiona cycling period ending at 18 UTC 24 Sep and the Ian cycling period beginning at 06 UTC 23 Sep. The sets of profiles assimilated for Fiona and Ian on 23-24 Sep were different and non-overlapping due to differences in the HAFS-A nest position for the two storms.

We added this discussion of Table 1 to the text (lines 320-327):

“Table 1 lists the number of RO profiles assimilated per day in the Control_noC2, Control, and ROMEX experiments for each TC case. Over the full evaluation period, the addition of Spire and PlanetiQ data in ROMEX experiment increases the assimilated profile count 4.4-fold for Earl, 3.8-fold for Fiona, 3.3-fold for Ian, and 1.8-fold for Julia, relative to Control. The larger gain for Earl and Fiona in RO data volume from experiment Control to ROMEX is consistent with those storms spending more time outside the tropics (Fig. 1), where COSMIC-2 data have sparser coverage (Ho et al. 2020b). A total of 129, 717, and 2513 RO profiles were assimilated in the Control_noC2, Control, and ROMEX experiments respectively.”

We revised our statement later in the Summary/Conclusions that you mention to describe the amount of RO data assimilated in the HAFS nest, rather than what was available globally (lines 775-779):

“In conclusion, this study demonstrates that a 3.5-fold increase in the number of assimilated GNSS RO profiles with the addition of commercial Spire and PlanetiQ observations has an overall beneficial impact on HAFS-A track, P_{MIN} , lower-tropospheric wind speed, and middle-to-upper-tropospheric temperature and water vapor forecasts for the four 2022 Atlantic TCs evaluated.”

3. Are the results, especially comparisons with Control, dependent on latitude?

If “the statistical forecast error analysis...was heavily influenced by...cases which spent most of their time...outside the tropics” (lines 697-699), then the impact of removing COSMIC-2 from the assimilation – one of the authors’ primary experiments – is not so clear. The authors acknowledge that COSMIC-2 sampling decreases in the extratropics, and it is apparent from Figure 2 that the COSMIC-2 sampling within the mesh decreases as Hurricane Ian (2022) moves poleward. How then does the statistical analysis change when only considering those forecasts that are for TC locations equatorward of some latitude? As a reviewer, it’s not obvious what latitude this should be as the assimilated data count for each assimilation window in each experiment and storm is not given (though I encourage them to consider its inclusion in the manuscript). The authors would, however, have access to this information and could select an appropriate latitude.

Thank you for raising this important question. A clean assessment of the latitudinal dependence of our HAFS forecast statistics is challenging for two primary reasons. First, the three TCs that moved northward (Earl, Fiona, and Ian) also underwent substantial intensification during this period. It is possible that stronger, more mature TCs and weaker, early-developing TCs could show different sensitivities to assimilation of additional GNSS RO data, and this could be explored in a future study with a larger set of cases. For example, the inertial stability attained by a strong TC’s tangential wind circulation has been shown in idealized modeling studies to increase the vortex’s resilience to dry air intrusions (Riemer and Montgomery 2011; Finocchio and Rios-Berrios 2021). Second, as these three TCs moved northward, a greater number of RO observations were cumulatively assimilated into the background vortex over multiple cycles. To properly stratify forecast statistics by latitude, while controlling for both TC intensity and the

“spin-up” period since the HAFS cold start, would require a much larger set of TC cases than was available from this study. This would allow for separate evaluation of tropical storm/tropical depression and hurricane intensity TC initializations poleward and equatorward of a specified cutoff latitude in future work. We revised a sentence in the Summary and Conclusions to emphasize this point, with the inserted words highlighted in bold (lines 784-789):

“Considering the marked decrease in COSMIC-2 profile density with latitude in the extratropics (Ho et al. 2020b), the incremental benefits of assimilating additional RO data may be proportionally greater **for TCs moving** outside the tropics, given the more uniform latitudinal distribution of sun-synchronous polar orbiting Spire and PlanetQ satellites (Anthes et al. 2024). **Future studies evaluating larger TC samples could test this hypothesis.**”

However, following your suggestion we examined our new Table 1 and chose the 24-degree N latitude as a cutoff separating equatorward-TC-centered HAFS analysis cycles that, despite some day-to-day variability, tended to have a smaller increase in RO profile count between Control and ROMEX, compared to the analyses with the TC initialized poleward of 24 degrees N. This subset of 42 HAFS “equatorward” analyses consists of those from Earl through 18 UTC 06 September, Fiona through 12 UTC 21 September, and Ian through 18 UTC 27 September, and all Julia analyses (with the first 48 hours of cycling since each storm’s cold start excluded from forecast statistics, as was done for the results shown in Figures 5 and 6). Interestingly, when considering only the equatorward 42-analysis subset, COSMIC-2 positive impacts on TC mean position error are increased somewhat during the 48-96 h forecast period, as shown by Control_noC2’s greater loss of relative skill in Figure R2.1, compared to the full sample set shown in Figure 5. When comparing the P_{MIN} mean absolute error statistics for the equatorward-analysis subset in Figure R2.2 with the full sample shown in Figure 6, we find little difference in COSMIC-2 impacts, although the forecasts from the equatorward analyses show more relative skill degradation from Control to ROMEX between the 54- and 96-hour lead times.

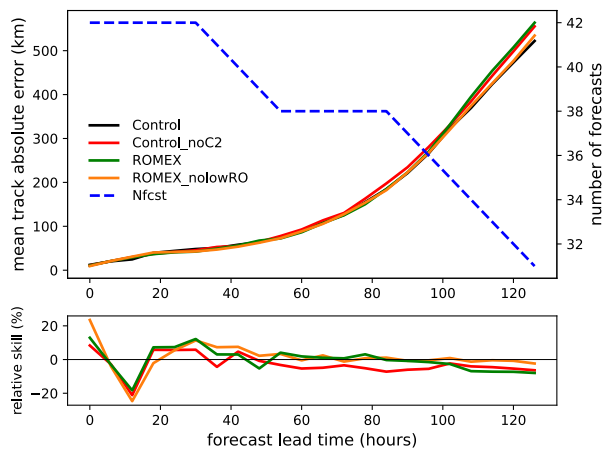


Figure R2.1 As in the manuscript’s Figure 5 but limiting the TC position error sample to forecasts initialized with a TC center south of 24 degrees N. Note the larger y-axis scale for relative skill shown in the lower panel, compared to Figure 5.

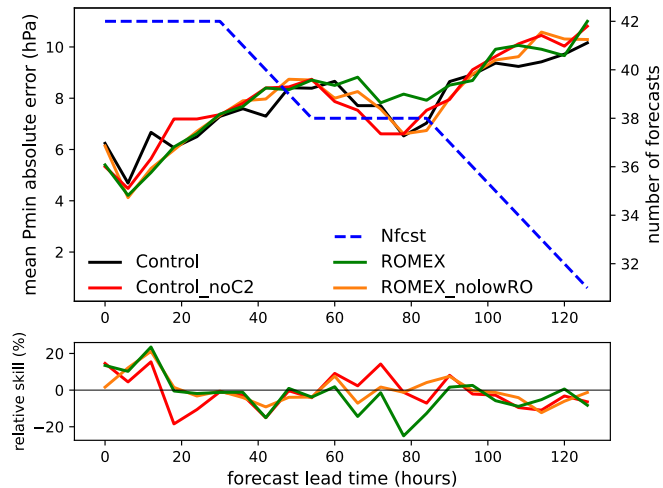


Figure R2.2 As in the manuscript’s Figure 6, but limiting the P_{MIN} mean absolute error sample to forecasts initialized with a TC center south of 24 degrees N.

We added a new sentence to Section 4a describing COSMIC-2 impacts on our HAFS TC track forecasts, highlighted in bold (lines 448-452):

“Control_noC2’s negative relative skill during the same period (Fig. 5b) indicates that COSMIC-2 observations are also beneficial for TC track forecasts. **This is particularly true for the subset of analyses that initialize TCs equatorward of 24° N, for which removing COSMIC-2 from Control yields negative 5-10% relative skill during $t = 54-96$ h (not shown).**”

References:

Riemer, M., and M. T. Montgomery, 2011: Simple kinematic models for the environmental interaction of tropical cyclones in vertical wind shear. *Atmos. Chem. Phys.*, **11**, 9395–9414.

Finocchio, P. M., and Rios-Berrios, 2021: The intensity and size-dependent response of tropical cyclones to increasing vertical wind shear. *J. Atmos. Sci.*, **78**, 3673-3690.

4. Omission of PlanetiQ in discussion of low vs. high SNR datasets should be corrected.

Consider Figure 3 and some of the discussion around it (e.g., line 313; this also applies to lines 647-650). The separation in the figure and the omission in the discussion of PlanetiQ helps the authors raise what they find to be an “interesting result” about COSMIC-2. But isn’t it more accurate to say that the result is that the low SNR dataset (Spire) has a notably different rejection rate in the lowermost atmosphere than both two other high SNR datasets (COSMIC-2 and PlanetiQ)?

The authors presented results for each of the three selected missions side-by-side in Figure 2. I don’t think there’s justification for not doing so in Figure 3. And the discussion about low vs. high SNR datasets should not omit PlanetiQ and should clarify what the authors find interesting about the results.

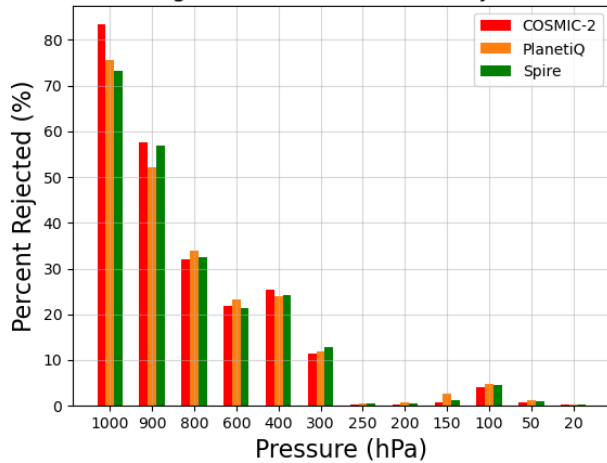
We decided to remove discussion relating the rejection statistics of Spire and COSMIC-2 to their SNR characteristics for several reasons. First, the topic is complex and would require more in-depth analysis that is beyond this study's focus on hurricane weather forecast impacts. Our main purpose in presenting Figures 3 and 4 and their related discussion in Section 3 is to show that both PlanetiQ and Spire provide RO data with generally comparable-to-better retention and diagnosed error characteristics in the lower troposphere, relative to COSMIC-2. For example, the COSMIC-2 SNR statistics are known to display strong latitudinal dependence, and this has been attributed to their receiver antenna geometry and orbital characteristics (Ho et al. 2023 Fig. 5). SNR information for individual profiles is also not available in BUFR files containing RO bending angle data for assimilation in weather models. Additionally, although you correctly point out that PlanetiQ's satellite receivers record RO signals with SNR comparable or better than COSMIC-2's TGRS receiver, there is no guarantee that future delivery contracts between PlanetiQ (or any other company) and federal customers would include the company's highest-SNR data at a mutually agreed-upon price. Therefore, we think it's best to not comment or speculate on how SNR may affect differences among COSMIC-2, Spire, and PlanetiQ observation uncertainty and extreme outlier occurrence in HAFS in this study.

Also, we re-generated Figure 3 (copied below) to show the three missions on the same bar chart, and to keep the bar coloring scheme consistent with Figures 2 and 4 (i.e., red for COSMIC-2, orange for PlanetiQ, and green for Spire). This should make it easier for the reader to compare the missions in terms of their QC rejection rate.

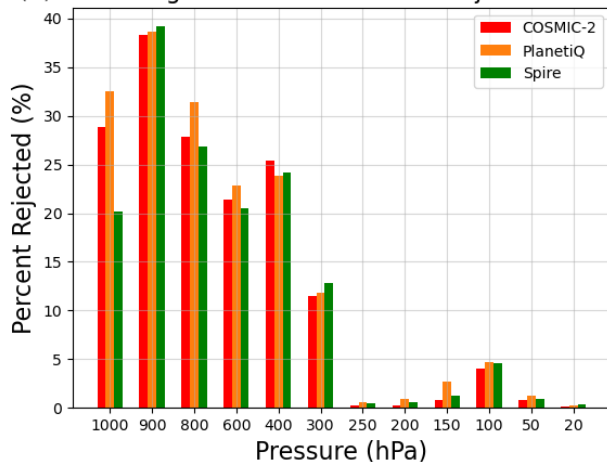
Reference:

Ho, S.-P., X. Zhou, X. Shao, Y. Chen, X. Jing, and W. Miller, 2023: Using the commercial GNSS RO Spire data in the neutral atmosphere for climate and weather prediction studies. *Remote Sensing*, **15**, 4836.

(a) Percentage of RO Observations Rejected: ALL QC



(b) Percentage of RO Observations Rejected: OMB QC



5. The analysis and storytelling of the results in Section 5 is quite nice.

[Thank you for this feedback.](#)

Other comments:

1. Section 3a: why was any discussion about MetOp excluded from this subsection? As it is an assimilated RO dataset in all four experiments, it should be introduced and included in figures here alongside COSMIC-2, Spire, and PlanetiQ. With only ~1100 global profiles/day it likely has a smaller impact than COSMIC-2 and Spire, but from a readers' perspective, its numbers are not so few as to be justifiably ignored.

In our opinion, Metop can be left out of Figures 2-4 and their discussion in Section 3 because the version of GSI used in our HAFS-A experiments rejects all Metop bending angles below the 8-km impact height. Therefore, although we agree that Metop's ~1100 daily global profiles should have an appreciable impact on TC forecasts by any model system that permits assimilation of sub-8 km Metop bending angles, such as GSI versions implemented since the Dec 4, 2024 removal of this hard rejection criterion (see our response to Major Comment 1), we expect Metop assimilation to have a relatively small impact on our HAFS model's forecast lower-to-

middle tropospheric temperature and water vapor fields that affect TC intensity. Section 3 focuses on comparing differences in the DA statistics of RO datasets assimilated in the lower-to-middle troposphere, and so we don't think that Metop results need to be shown in Figures 2-4.

The expected limited forecast impacts of Metop profiles should be mentioned in the text, and so we added these sentences to Section 2c when introducing the ROMEX dataset (lines 267-272):

“Prior to a December 2024 code upgrade, GSI implementations with NCEP’s forecasting models rejected all Metop bending angles below the 8-km impact height (~ 650 hPa) based on prior work finding larger Metop-A/B departures against the GFS background (Dutta et al. 2019). Since our HAFS-GSI build predates removal of this QC check, the assimilated Metop-B/C profiles are expected to have little impact on forecast lower- to middle-tropospheric temperature and water vapor fields.”

We added the Dutta et al. (2019) citation below to the References section.

Dutta, S., Vandenberghe, F., Shao, H., Zhang, H., and J. Doe, 2019: Assessment of radio-occultation from multiple GNSS platforms: Operational and non-operational. *JCSDA Quarterly*, **62**, 33-37, <https://doi.org/10.25923/w2dh-ep66>.

2. In various parts (lines 438, 512, 514, 532, others?), when discussing the ROMEX_noLowlevRO experiment, the authors state that it shows how the assimilation of sub-5-km RO observations changes the forecast quality relative to the ROMEX experiment. It reads a bit awkward, however, as it is possible to interpret the wording as not contrasting the authors’ ROMEX experiment that they perform but rather with the ROMEX data itself. Since all RO below 5 km is being rejected, this is more than just the impact of ROMEX (the dataset); the authors do make this clear in line 534. This is a subtlety, but clarity will ensure the reader does not misinterpret. Perhaps something like “our ROMEX experiment” in these spots would help?

We agree that our text in the places you mention could benefit from further clarification, so that the reader understands that the ROMEX_noLowlevRO experiment is removing *all* RO data below 5 km, not just the commercial ROMEX observations. We revised these portions of the text as follows, with the edit insertions highlighted in bold:

- Lines 526-531 (formerly line 438): “Finally, removing **all (commercial- and government-mission)** RO data below 5 km in ROMEX_noLowlevRO causes nearly all water vapor RMSD improvement below 400 hPa in ROMEX to be lost (cf. 9a, 9e). This implies that lower-tropospheric RO observations can also improve HAFS-A water vapor estimates in the middle troposphere via the background-error covariance matrix.”

- Lines 600-602 (formerly lines 512-514): “Removing the sub-5 km RO data from **our ROMEX experiment** reduces the cool-temperature bias in this layer, bringing it closer to that of Control, suggesting that sub-5 km RO observations drive **the ROMEX experiment’s** degradation there.”

- Lines 619-621 (formerly line 532): “Removing ~~the~~ all sub-5 km RO data more than reverses all of ROMEX’s wind RMSD improvement, and additionally, removing the COSMIC-2 observations degrades the Control wind RMSD.”

When we refer to ROMEX_noLowlevRO results later in the Conclusions, the text summarizing the experiment’s design should be reasonably clear, and we think that we can leave it as-is (lines 767-769):

“The ROMEX_noLowlevRO experiment, which rejected all RO data below the 5-km impact height, assessed the value of assimilating lower tropospheric RO data.”

3. Section 2d, conclusions: I would be fine with retaining the naming, but is “data denial” the best classification for the experiments? One of the big questions this paper seeks to answer is what the impact of the ROMEX commercial data has on TC forecasting. One of the three (plus a control) experiments is a data enhancement experiment; the other two are data denial. I suppose one could consider Control to be a data denial from the ROMEX experiment, but that doesn’t quite represent the experimental setup.

In our opinion, “data denial” would be an acceptable term to use for describing our experiments, if we consider the ROMEX experiment to be representative of a future observing system that utilizes a similar volume of commercial data and the other experiments therefore show the degradation with data denial.

But we agree that “data denial” is not the best terminology to use for describing the full set of our experiments, since we are assessing both enhancement of the currently operationally assimilated RO observation dataset in ROMEX and removal of routinely assimilated observations in noC2 and noLowLevRO. We changed “data denial experiment” to “observing system experiment (OSE)” in Section 2d (lines 280-281) and the Conclusions (line 796).

4. Please include all data (e.g., RO, GFS) in the data availability statement. It is appreciated that the datasets are available upon request, but the source of the data are sure to be public-facing repositories, even if, like the ROMEX dataset, they have restrictions.

We expanded our data availability statement to add mention of other data sources, and it now reads as:

“Large input and output datasets from the OSEs described in this study are stored on the NOAA Research and Development High-Performance Computing System (RDHPCS) High Performance Storage System (HPSS; <https://docs.rdhpcs.noaa.gov/data/storage.html>). They include the HAFS-A gridded forecasts, operational GFS forecasts providing the outer domain’s initial and boundary conditions, HAFS-GSI DA diagnostic files, and the ROMEX global dataset covering the four Atlantic TC forecasting periods. The latter consists of BUFR file-formatted Level-2 commercial and government-mission GNSS RO bending angle observations processed by EUMETSAT, except for COSMIC-2 data, which are processed by UCAR. Additionally, software used for generating the results presented in this study are stored on the University of Maryland Linux servers. These resources can be made available upon request, although

restrictions apply to sharing and redistribution of the ROMEX RO dataset (<https://irowg.org/ro-modeling-experiment-romex/>). The HAFS-A model code repository is publicly available on GitHub at <https://github.com/hafs-community/HAFS>.”

5. Please update references to include DOIs.

Done

Line-by-line comments:

Line 28, 124: “EXperiment”

Fixed

Line 40, 690: what is the relative increase within the study domain? COSMIC-2 observations are not global and thus, the global sampling does not scale to a 12x12-degree mesh that may fall outside the tropics.

The relative increase in RO data count from Control to ROMEX within the HAFS study domain shows case-to-case variability: 4.4-fold for Earl, 3.8-fold for Fiona, 3.3-fold for Ian, and 1.8-fold for Julia (as shown in our new Table 1). We added a new passage to the text describing these case-dependent differences in assimilated profile counts, as well as the total number of profiles assimilated in the HAFS nest for each experiment (lines 320-327):

“Table 1 lists the number of RO profiles assimilated per day in the Control_noC2, Control, and ROMEX experiments for each TC case. Over the full evaluation period, the addition of Spire and PlanetiQ data in ROMEX experiment increases the assimilated profile count 4.4-fold for Earl, 3.8-fold for Fiona, 3.3-fold for Ian, and 1.8-fold for Julia, relative to Control. The larger gain for Earl and Fiona in RO data volume from experiment Control to ROMEX is consistent with those storms spending more time outside the tropics (Fig. 1), where COSMIC-2 data have sparser coverage (Ho et al. 2020b). A total of 129, 717, and 2513 RO profiles were assimilated in the Control_noC2, Control, and ROMEX experiments respectively.”

Line 133: there are more than this number of occultations per day, and some NWP centers are leveraging all of them. Perhaps better to say “...approximately 27,000 global daily commercial RO...”

We agree, and we made this change as you suggested (line 147). We were restricted from evaluating certain commercial datasets in NOAA’s models.

Line 135: here too, it may help clarity to reword along the lines of “...impact of assimilating a large subset of the total commercial RO bending angle dataset...”

We added a footnote here to prevent any confusion caused by differences in global profile counts assimilated in our experiment and those used by other ROMEX participants (lines 149-151):

“This study supports ROMEX by evaluating the impact of assimilating a large commercial RO bending angle dataset, comprising about 20,000 Spire and PlanetiQ RO bending angle profiles per day globally¹,..”

¹ Our experiments do not use data from all ROMEX commercial providers.

Line 171: was that the minimum pressure at landfall? Was the record based on the minimum pressure anywhere along the hurricane track or at the time of landfall?

Yes, 931 hPa was Fiona’s pressure at landfall in Nova Scotia. The record was based on Fiona’s pressure at landfall, although this also tied the lowest pressure over water achieved by Fiona about 30 hours earlier (Pasch et al. 2023 Table 1). Although Fiona’s 85-kt maximum surface winds at landfall had weakened considerably from their 120-kt over-water intensity peak at 06 UTC 21 Sep, its minimum central sea-level pressure weakened more slowly and then fell as the storm approached Nova Scotia, which is not unusual for a TC undergoing extratropical transition, as its wind field expands and a nearby upper-level trough helps aid mass evacuation out of the eye region (Hurricane Sandy 2012 is another example). Pasch et al. (2023, p4) state that “Fiona was the deepest cyclone on record (by minimum pressure) to make landfall in Canada.”

Reference (also cited in the manuscript):

Pasch, R. J., B. J. Reinhart, and L. Alaka, 2023: Tropical cyclone report: Hurricane Fiona (AL072022). National Hurricane Center Tech. Rep., 60 pp. [Available online at https://www.nhc.noaa.gov/data/tcr/AL072022_Fiona.pdf]

Lines 185-188: you may consider including this statement earlier in this subsection, for instance, at line 159. It is fine here, but I was wondering while reading the earlier lines where the authors got the various bits of information about the storms.

We kept this statement at the end of Section 2a, but we added the NHC tropical cyclone report references earlier when introducing each storm, for example, in lines 175-178, the revised manuscript reads as:

“Category 2 Hurricane Earl (2 – 10 September) and Category 4 Hurricane Fiona (14 – 23 September) were long-tracked storms that spent most of their time at hurricane intensity, defined as having maximum surface wind speed (V_{MAX}) exceeding 65 kt, outside the tropics (Blake 2023; Pasch et al. 2023).”

Lines 201-203: what the is model lid for HAFS? Vertical resolution? How is bending angle initialized at upper levels for HAFS?

HAFS uses an 81-level sigma pressure hybrid vertical coordinate with a 2 hPa model top. We added a sentence to Section 2b that states this (line 220): “The 81-level sigma pressure hybrid vertical coordinate system has a 2-hPa model top...”

Bending angles from Commercial RO are not assimilated above the 30-km impact height based on “high altitude cutoff” parameter, whereas COSMIC-2 and other government-mission bending angles can be assimilated as high as 50 km, or the 2 hPa model top, whichever is lower.

Line 308-309: what is the total count of profiles in this 1000-950 layer for the three missions?

From the four TC cases, the total number of profiles with at least one bending angle assimilated in the 1000-950 hPa layer are 106 for COSMIC-2, 314 for Spire, and 43 for PlanetiQ.

Lines 327-336: I think this work warrants some additional explanation for the reader. Is this analysis from HAFS using those occultations that fall in the inner nest?

Yes, our RO observation error diagnostics shown in Figure 4 are computed from the HAFS inner nest only, because HAFS does not assimilate any observations in the outer domain. We added some additional information to this paragraph, and it now reads as (lines 409-420):

“Figure 4 compares profiles of COSMIC-2, Spire, and PlanetiQ observation error standard deviations estimated from the ROMEX experiment’s O-B and observation-minus-analysis (O-A) **data collected from all four TC cases. These diagnosed observation errors include contributions from instrument, retrieval, and representative sources, and we estimated them using the Desroziers et al. (2005) algorithm, which has been widely employed by operational NWP centers (Cucurull et al. 2013; Bonavita 2014; Bowler 2020; Lien et al. 2021). Some caution is warranted in drawing broader conclusions about the uncertainty characteristics of these three RO observation platforms, given our relatively small DA diagnostic data sample drawn from a TC-following regional domain. From our experiments,** we find that all three platforms have similar observation uncertainty in the upper troposphere. However, Spire observations show the smallest uncertainty below the 5-km impact height in the tropics (Fig. 4a) and below the 7-km impact height in the extratropics (Fig. 4b).”

Bolded portions indicate changes from the first submitted manuscript version.

Line 401: I have no doubts about the quality of ERA5, but I’m also unfamiliar with its skill in reanalyzing TC locations and pressure minima. Have you measured ERA5 errors in TC location/intensity relative to NHC or is there an appropriate reference?

We found an analysis of ERA5 TC position errors relative to best track databases from 1998-2019 in Slocum et al. (2022). They showed in their Fig. 2a-c that the ERA5 median position error for Atlantic TCs was around 25 km, 20 km, and 15 km for storms in the 34-63 kt, 64-95 kt, and ≥ 96 kt ranges, respectively. It makes sense that the ERA5 position errors become smaller with increasing TC intensity because stronger, axisymmetric vortices are easier for a global modeling system to resolve compared to weaker, more asymmetric circulations. Since our gridpoint-to-gridpoint HAFS temperature and water vapor verification shown in Figures 8 and 9

extends out to 300 km from the TC center, we do not expect our results to be too sensitive to ERA5 TC position errors in the 15-25 km range.

Reference:

Slocum, C. J., M. N. Razin, J. A. Knaff, and J. P. Snow, 2022: Does ERA5 mark a new era for resolving the tropical cyclone environment? *J. Climate*, **35**, 7147-7164, <https://doi.org/10.1175/JCLI-D-22-0127.1>.