Below, reviewers' comments are shown in black, and our responses are provided in blue.

Reviewer #1: General comments

This paper describes experiments at NRL with the high-volume GNSS-RO data from ROMEX. Similar to other centres, forecast degradations are seen when the observations are used without any alteration. Experiments using an empirical bias correction and an adjustment of the refractivity formulation are more successful and overall forecast improvements are seen.

The paper is well written, and the conclusions drawn are supported by the evidence that is presented. The paper is a useful addition to the literature, and provides an interesting perspective on the use of large numbers of GNSS-RO observations. My specific comments are largely minor details of presentational clarity.

We sincerely thank the reviewer for their thorough and thoughtful review of our manuscript, and for their kind and encouraging assessment of our work. We have carefully addressed the reviewer's specific comments regarding presentational clarity in the revised manuscript, and we believe these revisions have further improved the overall quality and readability of the paper.

Specific comments

L30, L34 and L40: I would suggest reserving "significant" for contexts in which statistical significance is indicated. In these lines it would be better to use large, substantial or important.

Thanks for pointing that out! We have revised the wording at these lines.

L32: I would add Samrat et al (2025) to the list of references demonstrating the importance of GNSS-RO.
Samrat, N.H., Candy, B., Lewis, O., Cotton, J., Smith, F., Halloran, G., et al. (2025) Observation impact evaluation through data denial experiments in the Met Office global numerical weather prediction system. Quarterly Journal of the Royal Meteorological Society, 151(771), e5002. Available from: https://doi.org/10.1002/qj.5002

Thanks for the reference, which is added now.

L41: It should be 6000 COSMIC-2 profiles. I suggest also mentioning that this is a daily number "with the addition of 6,000 COSMIC-2 refractivity profiles per day."

The COSMIC-2 satellites assessed in Cucurull et al. (2018) consist of six equatorial (total of 6000 soundings per day) and six polar (total of 6000 soundings per day) receiver satellites. So it should be 12,000 profiles per day. We added "per day" to clarify.

L43: The CGMS is not part of the WMO. Formally, the WMO is a founding member of the CGMS. Otherwise they are unrelated.

We deleted WMO in that sentence.

L54: 30,000 to 40,000 occultations per day is no longer three times the number of RO observations used in NWP, approximately 20,000 occultations are currently available daily.

It is true that the number of RO profiles available daily has increased substantially in recent years. We have revised the sentence to clarify that the phrase “three times the number of RO observations used in NWP” refers to the number of RO profiles assimilated by operational NWP centers around 2022. At that time, approximately 10,000 RO observations were used daily, whereas the constellation discussed here is expected to provide 30,000–40,000 occultations per day.

L64: Since the data assimilation uses hybrid covariances, it would be useful to note whether the ensemble prediction system is rerun with the additional observations to provide updated covariances with the experiments. Therefore, please can this information be added to the description?

The dependence of background error covariance on the observations is only indirect through the analysis and subsequent ensemble generation from the analysis. There is no additional rerun.

L87: The formatting of this equation makes it difficult to follow how the piecewise linear function is constructed. For instance, the third line is "linearly varies from -0.01% to 0.2% 18 < h < 23 km". I would normally read that as -0.01% at 18km and 0.2% at 23km. In reality it is the other way around in order to be continuous with the adjacent layers. I would suggest a fundamental reformat of this equation to make the presentation unambiguous. I also note that an em-dash is used in place of a minus sign.

Thanks for noting the inconsistency. We have corrected the end point to make it continuous with the adjacent layers. The appearance of em-dash could be just an artifact of Word equation plugin.

L89: It would be helpful to be clear about whether y' is added to or subtracted from the observation.

The sentence right before the equation was modified to indicate that for each observation, y' was subtracted.

L111: The text indicates that the innovations are normalised by the error estimates. Figure 2 however, also normalises by the background bending angles. Therefore, I would suggest replacing "error estimates" with "the background bending angles and by the estimated observation uncertainties".

Suggested wording added.

L125 and L136: I suggest replacing "error" with "uncertainties", since they are not observation errors.

Replaced.

L135: Please remove "more physically-based". The authors have not presented any arguments for a physical basis for adjusting the refractivity coefficients.

Removed.

L170: Please state what the change is relative to (the ROMEX control presumably).

Added.

L293: It would be worth noting that the RMSE performance at 700 hPa is worse for romex_bc and (to a lesser extent) romex_nl than for romex. This is despite the specific humidity biases (relative to ECMWF) being smaller at this height.

Noted.

L330: The manuscript makes no mention of the possibility of adjusting other coefficients in the refractivity formulation, or the possibility of using more accurate refractivity equations than that of Smith and Weintraub. I feel that these topics should be mentioned, even if they are not addressed directly by the experiments.

Added. Thanks for the suggestions!

Typographical corrections

L9: OSE is normally written as "observing system experiments" rather than "observation".

Corrected.

L18: Please change "treatment of the observation" to "treatment of observations".

Changed.

L25: Please change "low-Earth orbit" to "low-Earth orbiting".

Changed.

L152: Please delete "the most".

Removed.

L199: Presumably you are using the operational analyses from ECMWF. Therefore, please insert "operational" after "ECMWF".

Added.

L229: Please insert a space between "Fig." and "6".

Inserted.

L241: Please delete "change to".

Deleted.

L277: Please change "leads to" to "demonstrates" as this reads more clearly.

Changed.

L298: Please add "analyses" after ECMWF and change "observation" to "observations".

Changed.

L313: Please add "for dry air" after "refractivity coefficient".

Changed.

Reviewer #2: This is a nice short paper summarizing the results from evaluating the assimilation of the ROMEX GNSS-RO observations in the NAVGEM system. The authors conducted two sets of OSEs covering the ROMEX period from September to November 2022. The first set evaluated the impact of the increased number of RO observations, while the second examined the influence of different refractivity coefficients and the bending angle bias correction. The results showed an overall positive impact from the ROMEX data, and a bias in the innovations was further investigated through additional experiments.

Thank you for your positive and encouraging feedback. We appreciate your summary of our work and are glad that the objectives and key findings of our study were clear. Your comments are greatly appreciated.

The manuscript is clear and easy to follow. I only have a few minor suggestions.

1. Please revise Eq. (1) to make it clearer and more consistent with Fig. 1. From Fig. 1, it appears that the bending angle bias correction decreases with height, from 2% near the surface to 0.2% at 10 km impact height, so the current expression in Eq. (1) is somewhat confusing. From 18 km to 23 km, the correction appears to decrease linearly from 0.2% to -0.01%, and from 23 km to 40 km it further decreases from -0.01% to -0.08%, please revise Eq. (1) for a better description. What is particularly confusing is that Eq. (1) indicates a fixed increase of 0.1% at 40 km and above, but this feature is not apparent in Fig. 1.

Eq. (1) is now revised as suggested.

2. In the left panel of Fig. 1, the small inset plot in the upper right seems unnecessary since the zoomed-in view is already shown in the right panel. Also, to make Fig. 1 less confusing, it may be better to plot the bias correction starting from the surface rather than from 2 km height.

We thank the reviewer for this suggestion. The inset plot in the left panel has been removed, as the enlarged view is already provided in the right panel. The bias correction profile is now plotted starting from the surface.

3. In Table 1, please add the average number of profiles assimilated for each experiment.

Added.

4. For the FSOI analysis, could you provide more details about the 24-h forecast metric? For example, what variables and vertical levels are included, and how is the forecast error evaluated.

Additional explanation on FSOI is added in the revision.

5. Why does Spire-07 have the largest total FSOI? Is it because it contains more observations than the other Spire groups, or because the data quality is better? Also, why does Spire-05 have a much larger per-observation FSOI than the other Spire groups and other satellites, except for COSMIC-2?

Spire-07 group has the largest total FSOI because it contains more observations than the other Spire groups.

Thanks for pointing out that Spire-05 (receiver range 103–117) exhibits a noticeably higher per-observation FSOI compared to other Spire groups and most other satellite systems (except COSMIC-2). While the zonal mean (Fig. S1) and standard deviation (Fig. S2) of normalized innovations for Spire-05 are qualitatively similar to those of other Spire groups, the zonal distribution for Spire-05 (Fig. S3) shows some differences, including greater coverage over the Southern Hemisphere, as well as in lower latitudes. Additionally, the global mean of normalized innovations for Spire-05 in the lower troposphere (impact height 2–18 km, Fig. S4–S5) is slightly higher than for other Spire groups, combined with the FSOI error norm having larger values in the lower latitudes. We believe these larger innovations and 24-hour and 30-hour error norms for Spire-05 contribute to its higher per-observation FSOI. We have added these discussions to the revision.

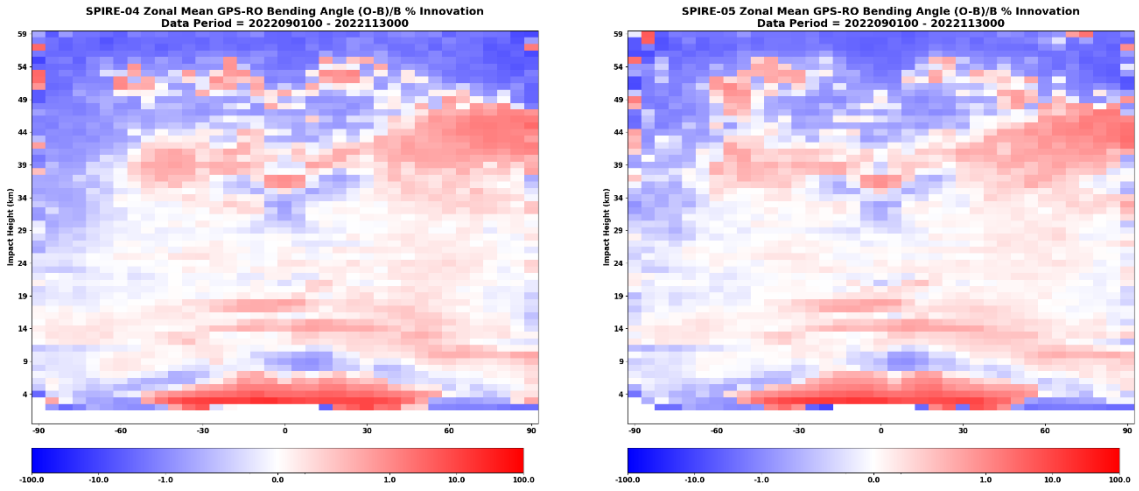

Fig. S1: Zonal mean of normalized innovations for (left) SPIRE-04 and (right) SPIRE-05 from September to November, 2022.

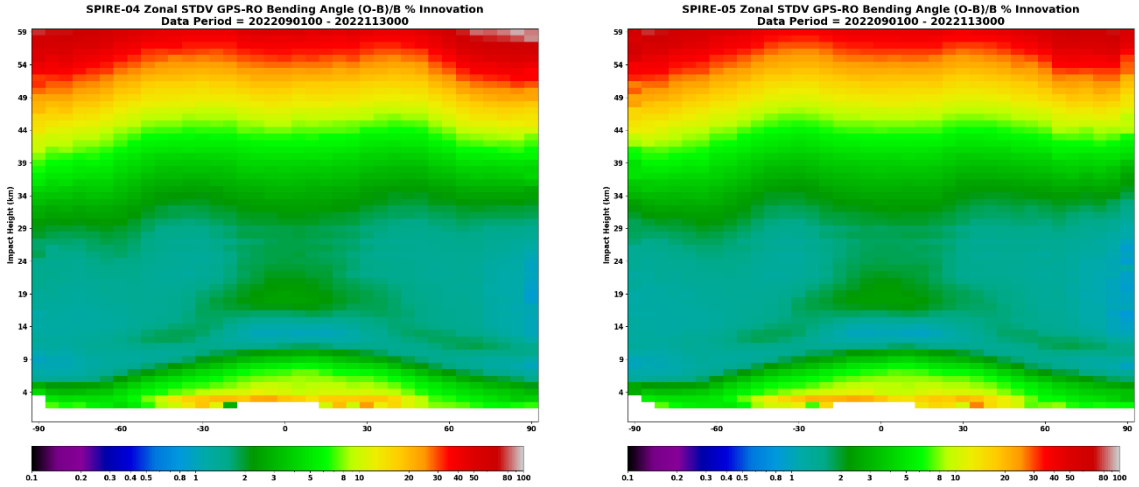

Fig. S2: Same as Fig. S1 except for zonal standard deviation of normalized innovations.

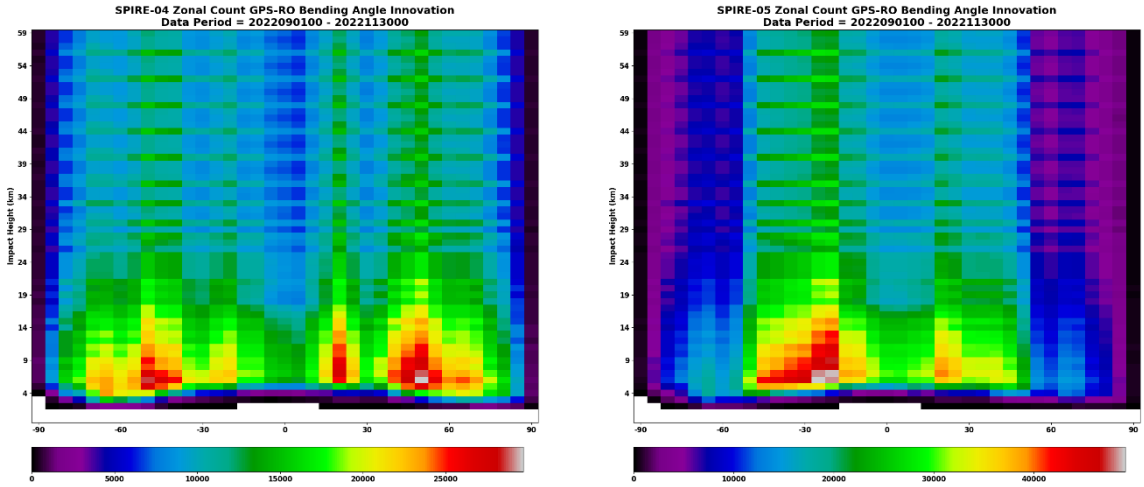

Fig. S3: Same as Fig. S1 except for zonal mean of data counts.

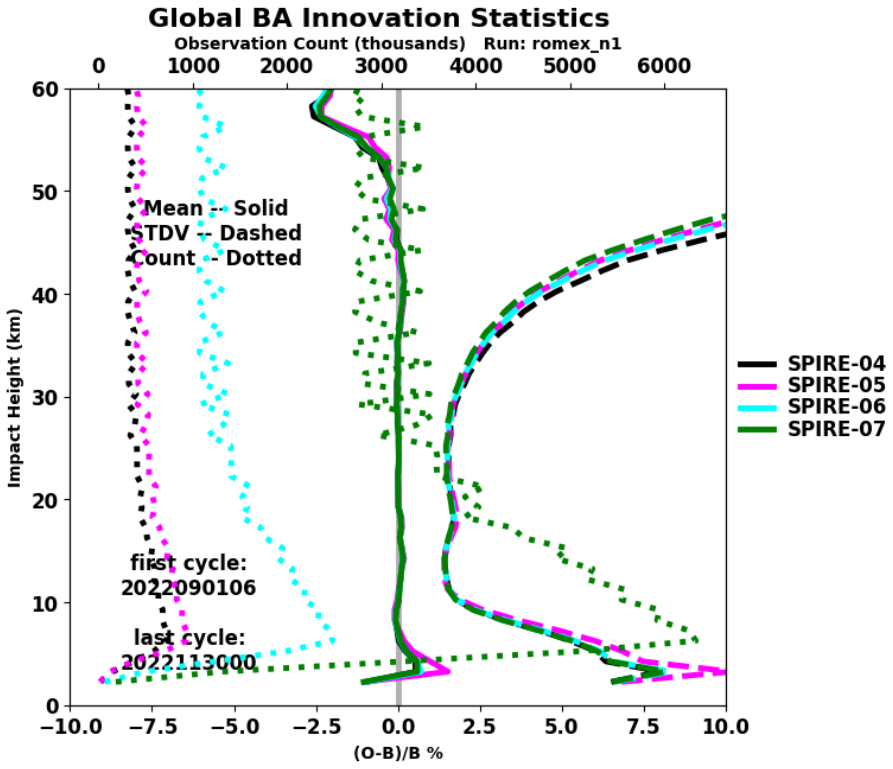

Fig. S4: Global mean, standard deviation, and observation count for the normalized innovations for SPIRE-04, -05, -06, and -07 from September to November, 2022.

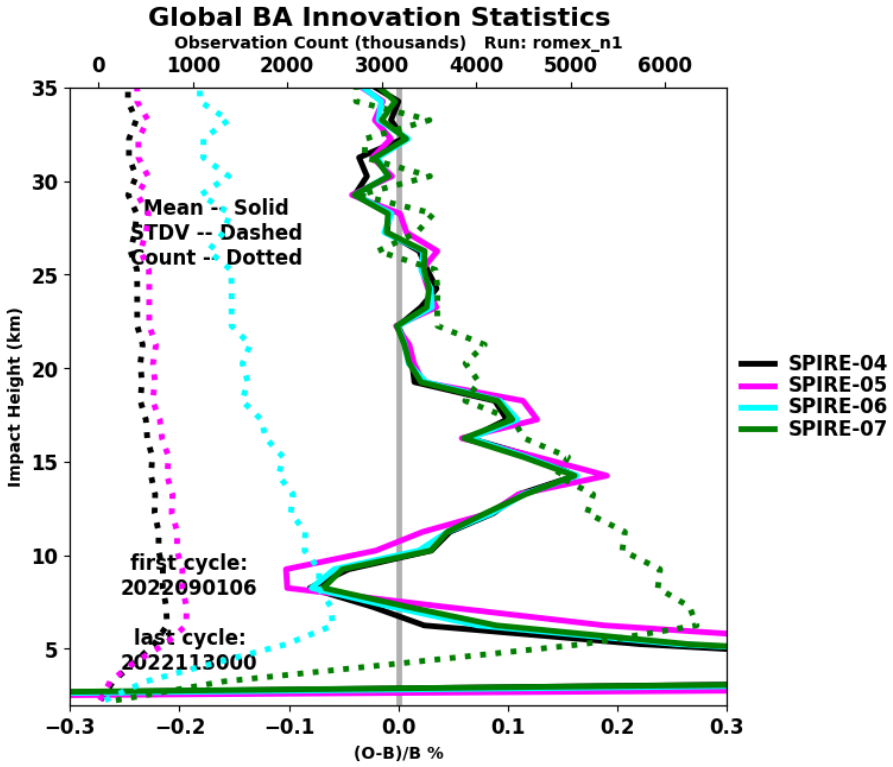

Fig. S5: A zoom-in plot for Fig. S4.

Line 73: missing a period at the end of the sentence.

Thanks for catching this.

Reviewer #3: This manuscript " Impact of high-volume GNSS Radio Occultation data on improving the accuracy of Navy's global weather prediction models." by Christophersen, Ruston and Tyndall assesses the impact of high-volume GNSS Radio Occultation data during ROMEX on improving the accuracy of Navy's global weather prediction models. I recommend to publish this work with some minor revisions, see listed below.

In general, it would be good to see that not only RMSE is used to assess forecast scores but also statistical measures as well, as e.g. std deviation or anomaly correlation.

We appreciate the reviewer's thoughtful comments and suggestions. We agree that it is generally beneficial to consider additional metrics beyond RMSE, such as standard deviation or anomaly correlation, when evaluating forecast performance. In our routine analyses, we do generate these supplementary performance metrics. However, for the purposes of this manuscript, we chose to focus on RMSE statistics, as we believe they effectively convey the key findings regarding forecast accuracy. Thank you again for your valuable feedback.

P3, eq (1); for h lower equals 10km which correction is used at 0km? Does it start with 2% and the linearly decreasing to 0.2% at 10km?

Yes, the threshold value at 0km is 2%, and at 10km is 0.2%, with a linear interpolation for 0-10km.

We added the starting value for $h \leq 10\ km$ range.

$$linearly\ ramped\ from\ 0.2\%\ to\ 2\%, \qquad h \leq 10\ km$$

P4, l.110/111: Be clearer what you mean with "normalized by error", as you normalise be background in Fig. 2a and by obs error in Fig. 2b.

We revised the sentence to differentiate the two normalizations for Fig. 2.

P5, Fig.2: Why didn't you plot romex_cntl? It would have been interesting to see how it looks for this experiment. Also colour choice is not colourblind friendly (also applies to Fig.3).

Thanks for the suggestion. We added the control experiment in Fig. 2 and updated all the colors in Fig. 2, Fig. 3, Fig. 6, Fig. 8, and Fig. 10 to be colorblind friendly.

P8, Fig 4. COSMIC-2 satellites naming by E1 to E6 needs to be clearer for the reader. Maybe COSMIC-2 E1….

Added.

P 9, l.212 Could you mention the magnitude of this bias for GP?

The bias for geopotential is about 5 meters. Added in the revision.

P 15, chapter 4. Please mention that also other studies using a more advanced forward operator - e.g. 2D ROPP version – could also be tested if this helps to improve forecast scores, especially in the troposphere.

Added.