

Summary

This manuscript evaluates the impact of RO observations from ROMEX on the NWP. Using the GFS model and JEDI assimilation system, results show that assimilating additional RO data can further improve the global weather forecasts, with beneficial impacts lasting up to five days. The result also indicates that there is no saturation for the forecast improvement. The manuscript is well written, and the experimental design and results are solid. The manuscript could be a valuable contribution to the communities of RO, observations, assimilation, models, and forecasts. I have several comments below.

Thank you very much for your valuable comments. We greatly appreciate your positive comments on our work. We have reviewed and revised the manuscript accordingly in response to your comments and questions. A point-by-point response is provided below in blue for clarity.

1. Section 2 and Figure 1, it is unclear about the differences among ROMEX, ROMEX20K, ROMEX sub dataset. Based on the words, I would guess Fig. 1a + Fig. 1b = Fig. 1c, and Fig. 1a + Fig. 1d = Fig. 1e, while Fig. 1d is not included in Fig. 1b. But based on Fig. 1, it is hard to imagine Fig. 1a + Fig. 1d = Fig. 1e. It would be nice to clearly describe the dataset, which could be consistent with the assimilation experiments.

We have revised the figure caption and panel titles for clarity. The figures have also been replotted with different color bars, incorporating the comments from the other two reviewers.

2. 'Quality control' is defined as 'QC' at line 211, so that QC can be consistently used later on (e.g., line 229, 236, 351...).

Thank you so much for catching it. Revisions have been made.

How to QC the RO observations are discussed in the manuscript. Does it use the 3 times of standard deviations of the observation error, or something else?

Yes, we use the 3 times of standard deviations of the observation error. A paragraph has been added to describe specific QC procedures, as shown below.

"The first QC procedure applied in this study checks the quality flag provided by the data providers; observations labeled "non-nominal" are excluded. The second procedure, a background check QC, rejects observations if the difference between the simulated and observed values exceeds three times the specified observation error."

3. I have a curious question about the bias introduced by assimilating the RO. Since the observation error at high altitudes is already large (Fig. 2), which implicitly contains somewhat effect of observation error inflation. Is it possible to conduct

bias correction to the RO, like the commonly adopted bias correction for the satellite radiances? Are there systematic features for the bias?

Thank you for your questions and comments. We have revised the manuscript to discuss potential sources of bias, which may arise from the reference analysis, model bias, or the configuration of the DA system. Further optimization of the DA system for these data will be an important focus of our future work.

Bias of different systems (e.g., Met Office, ECMWF, NRL) appears at different levels, suggesting that such biases are model dependent. Even if some bias originates from the observations themselves, its magnitude remains relatively small. We plan to examine this further in our next study.

4. Based on Fig. 8, the authors state that “Notably, there is no clear sign of saturation, as most levels continue to show improvement with increasing numbers of RO profiles. However, the degree of this non-saturation appears to depend on both the variable and vertical level and could be influenced by the specific data assimilation configuration.” It seems an overstatement for the ‘no clear sign of saturation’, especially for the state variables at high levels (e.g., T at 400 hPa, wind speed at 250 hPa). It would be nice to discuss these results. Are the non-saturation errors due to the data assimilation algorithm, or observation type, or other potential reasons?

We agree that the original content included somewhat overstatements. We have made revisions accordingly. For example, we rephrased “Notably, there is no clear sign of saturation” to “Overall, no consistent evidence of saturation was found”.

We also mentioned in the revised manuscript that “However, the extent of this non-saturation could be influenced by the specific DA and forecast system”. However, since multiple NWP centers also showed positive impacts of increased RO number, it does confirm that RO is a key observation type for such non-saturation impact features.