Response to Reviewers

We thank the reviewers for their comments and the time spent preparing them.

Reviewer 1

Major comments:

1. L9-13: "To test this ... the same results": These sentences show your purpose and main procedures. It is better to provide your main conclusions here. such as, how about the observed GWs depend on the viewing angles, the vertical and horizontal resolutions of SABER and GNSS-RO by using the same atmosphere data (high-resolution GEOS model). However, in the subsequent sentences, I think the GW Ep were calculated from the actual observations from SABER and GNSS-RO.

The GW Ep is calculated twice, once from observations and another from model for the purpose of comparison. However, there is no possibility to measure how the observed GW Ep would change with the angle or the resolution as we cannot change the instrument itself, that is the reason why we used the model data and sample it as similar as possible to the observations. We could then alter the viewing angle and the resolutions hypothetically and see how the results help us with the observations. The model results show that the vertical resolution is the most effective factor out of the 4 ones tested. We applied this result back to the observations and proved that it is correct.

2. L14 and L19: Does the 1.3 Jkg-1 and 0.5 Jkg-1 are different from zero in statistical sense? These two distributions are stat. sig.

We have carried out a test of statistical significance using the non-parametric two-sample Kolmogorov-Smirnoff test, which we describe the results of in the section on 'Time series and Distribution', and have mentioned in the abstract that these values are statistically significant. The choice of a non-parametric test is due to the non-Gaussian nature of the SABER and smoothed GNSS-RO potential energy distributions, as shown in Figure 12b.

3. L15-17: "...the most significant factor in determining...", How about the quotative relation between the difference (between the results of both instruments) between the vertical resolution? Does the resolution mean the vertical step of between two consecutive sampling points.

Yes, vertical resolution and sampling point spacing are different – the vertical step between sampling points is the spatial separation between the centres of two infrared (SABER) / radio (GNSS) point spread functions which define the spatial lower bound of features that can be distinguished, i.e. the PSF represents the true resolution, which in both cases is much coarser than the step spacing. This is discussed in section 5 (L435), where we define what the vertical resolution is and how it varies from one instrument to another depending on how each instrument is designed.

4. The second paragraph of the Abstract: What is the conclusion of this paragraph in supporting the "interactions between GWs and the QBO" as stated in L6-9

Our ability to **observationally** constrain the physics of interaction between GWs and the QBO is inherently and fundamentally limited by our ability to accurately measure these processes. While the bulk QBO is fairly easy to constrain accurately even in reanalysis, this is much more of a technical challenge for GWs. By assessing the effect of the observational filter on the GW-QBO interaction, we are able to draw conclusions about our fundamental ability to say anything about these interactions at all.

Our study and the conclusions we draw quantify how much we can trust these measurements in the context of this observational filter differences between the two sets of measurements, feeding back up the chain to the primary question of measurement reliability and consistency for long term study.

5. The third paragraph of the Abstract: This paragraph only shows the essential rule of studying gravity waves. The question is what are the appropriate observational methods for gravity wave research of this work?
This paragraph is to emphasis that GNSS-RO is still capturing short vertical wavelengths of gravity waves than SABER, but when we adjust the vertical resolution of GNSS-RO to match that of SABER, we get very similar results, hence we can use GNSS-RO to extend the SABER climatology. This is demonstrated in detail in the results section of the paper.

6. Please clarify the terms used in the text: resolution, sampling step, lower limit of the vertical wavelength used to derived Ep.
The term resolution is defined in section 5 (L435). The term sampling step is defined in section 3.1 L171 – 175. The lower limit of vertical wavelength for each instrument is defined in section 4 L281 -285.

7. Looking through the abstract, I still cannot get the main purpose of this manuscript. In the response: As stated in the response, the main purposes are: (1) address the observational filter differences between GNSS-RO and SABER measurements, (2) given the known observational filter differences between both techniques, (3) validate the reliability of SABER observation through comparing the GW derived from both SABER and GNSS-RO. Please fulfil these purposes in the abstract since the abstract should not provide only the purpose of the work but provide a concise summary of the research purpose, methods, main results, and conclusions. After these issues are resolved, you may find a proper title.
The main purpose of this manuscript is (1) and (2): given the known observational filter differences between both techniques, we identify the main reason behind the difference in GW Ep between both instruments. This is confirmed because, when we adjust GNSS-RO to match SABER's vertical resolution, we get good agreement between both timeseries. We have already discussed in our previous response how the title corresponds to the main results of the paper and to strengthen this title-to-content linkage, we have added more material including Fig. 12 and additional text in the Abstract, summary and conclusions that complement the title and answer the question in it. Due to these changes and in the context of our previous arguments, we believe that the title is descriptive of the work presented but are happy to reconsider this on request of the editor.

<u>Minor comments:</u>

1. L100: Please provide some references of "other datasets".
This was a misleading sentence; the choice of temporal and spatial coverage is to ensure consistency with SABER and GNSS-RO. This is fixed in the revised version 2.

2. L118: What is the mean of "vertical resolution of ~1 km". It is better to say that the temperature profiles have a uncertainty or accuracy of? K with vertical sampling step of 1 km in the stratosphere. Corresponding revisions should be made for SABER temperature (L132-133). Otherwise, if one does not consider the temperature uncertainty or accuracy, the vertical sampling stem can be arbitrary through over-sampling or down-sampling or interpolation.
Fixed.

3. L122-125: How about the consistency of the GNSS-RO measurements from different missions? i.e., the standard derivations of the concurrent temperature profiles measured from different missions.
Fig. 13 was added based on a comment from one of the reviewers in the first round and it shows how each mission of the GNSS-RO contributes to the GW Ep.

4. L148: Here, the vertical resolutions may be replaced by vertical step. Please use "resolution" properly throughout the manuscript for the purpose of more readable and accurate.
Fixed

5. L204-210: "excluding any with vertical wavelengths greater than 12 km", Does this apply on both SABER and GNSS-RO data? How about the shortest vertical wavelengths of GWs derived from SABER and GNSS-RO data?
Yes, it applies to both GNSS-RO and SABER. We set a limit for the highest vertical wavelengths to examine only the short vertical gravity waves due to their importance in interaction with the QBO. The lower limit is defined by the Nyquist sampling limit of each instrument.

6. L223: "eastward peaks of the zonal wind": How about the time step of zonal wind, daily or monthly? It seems that the zonal wind is on a monthly scale as seen from Figure 3. Please clarify this in the text.
Fixed.

7. L281-284: "The horizontal resolution" might be "the horizontal sampling step"? This leads the resolved horizontal wavelength of more than 270 km? Moreover, this is conflict to the statements in L260-264: "however, accurate measurements of the λh are difficult or impossible for GNSS-RO because the profiles are not regularly spaced". In fact, even at the early missions of GNSS-RO, the RO profiles were not regularly spaced. Please clarify these statements.
The horizontal resolution and the horizontal sampling step are two different concepts, one is a property of the instrument and the other is a choice for the analysis to sample the data. See our response to major comment 3 for further detail.
Based on radio propagation theory as derived by Kursinski et al., 1997, the horizontal line-of-sight resolution is around 270 km. Hence, gravity waves with $\lambda_h \sim 270$ km in the line of sight are less likely to be detected by GNSS-RO in the along-boresight direction but can be detected in the across-boresight direction and hence the bound is an approximation rather than a hard limit. The upper limit is not well-defined, and fundamentally limited by the minimal detectable phase change in the vertical imposed by the smoothly varying wave along an unknown angle.
Yes, the early missions were not regularly spaced but close enough to be able to estimate the horizontal wavelength of GW using Alexander 1998 method and this has been published by Alexander and Wang 2010 and Faber et al. 2013.

8. L281-288: From these sentences, I get that the upper limit of the vertical wavelength is 12 km for both SABER and GNSS-RO. I guess that the lower limit of the vertical wavelengths is, receptively, 2 km for GNSS-RO and 4 km for SABER. However, the lower limit of the vertical wavelengths shown in Figure A2 is receptively, 4 km for GNSS-RO and 7 km for SABER. I confuse that what is the lower limit of the vertical wavelength, which is used to derive Ep from GNSS-RO and SABER. This is important since the Ep is dependent on the range of the vertical wavelength even for the profiles measured by

same instrument. Please clarify the lower limit of the vertical wavelength, which is used to derive Ep.
We did not set a lower limit for the vertical wavelength calculation, but the proportion of very short vertical wavelengths captured by each instrument to the rest of the spectra is very small. The most dominant wavelengths captured by each instrument are the ones shown in figure A2.

9. L292: Please provide the method of get periodogram.
Fixed.

10. L339-346: This paragraph provides a statistical test for specific period. However, it is better to the confidence level of the statistical test. For example, the QBO is 2.4 years with 0 FAP. One can infer that the period of 2.5 years has a FAP of below 0.01, which is still a secure result. Due to the variability of QBO period from year to year, a better representation of QBO period might be a statistical distribution with center of ? months and standard deviation of ? months.
The significance of periodogram peaks is typically evaluated using the false alarm probability (FAP), which is commonly calculated under the assumption of uncorrelated noise through numerical methods such as bootstrapping or Monte Carlo simulations. FAP indicates the likelihood that a detected peak arises from noise rather than a genuine signal. In our case, the very low FAP values strongly suggest that the observed peaks are indeed due to a true signal. We selected this approach because it is widely used for assessing the statistical significance of periodogram peaks, relies on robust numerical techniques, and provides results that are straightforward for readers to interpret.

11. Section 5.1: The range of vertical wavelengths of GW, which is used to derive Ep should also be clarified.
Fixed (L426).