We appreciate that the reviewer provides such valuable comments.

Responses to the specific comments

1. Page 2 (lines 24-30), page 3 (lines 1–2): As with the pre-existing GPS-RO sounders…, the raw observations from GNOS consist of phase and signal to noise ratio (SNR) measurements. In addition, auxiliary information provided by the International GNSS Service (IGS), such as the GPS precise orbits, clock files, Earth orientation parameters, and the coordinates and measurements of the 1 ground stations, are also needed.

What about the navigation bits? Does Beidou have navigation bits, similar to GPS/GLONASS ones? If so, are they also provided for the precise demodulation?

A: The navigation bits contain information concerning the satellite clock, the satellite orbit, the satellite health status, and various other data. Beidou have navigation bits too. IGS provides the Beidou navigation bits, but not in near real time. The timeliness can be about 7 days.

2. Page 3 (line 26): if they exceed the three sigma from a statistical point of view.

… if they exceed 3 times the standard deviation. How is the standard deviation defined?

A: GNOS data is compared to background data, e.g. ECMWF reanalysis. The standard deviation is defined as $\mathrm{std} = \frac{\sqrt{\sum(\mathrm{x_i}-\overline{\mathrm{x}})^2}}{\mathrm{n}}$, n is number, $\mathrm{x_i} = \left(\frac{\mathrm{O-B}}{\mathrm{B}}\right) * 100\%$, $\overline{\mathrm{x}}$ is the average of $\mathrm{x_i}$.

3. Page 4 (lines 4–6): Therefore in this work we developed and tested a new L2 bending angle extrapolation method for GNOS data, and implemented it in ROPP.

Once speaking about a "new" method of the L2 extrapolation, one must cite the papers describing the "old" extrapolation technique.

There may also be some other publications on this topic. These papers cited, the differences between the old and new approaches must be discussed. Is it the "old" extrapolation method that the authors call the "ROPP extrapolation"? Or does ROPP use a different method? What is the reason of the failure of the old extrapolation technique for the GNOS data?

A: We agree that the "standard" ROPP should be described in more detail in the revised manuscript.

In the context of the difficulties processing GNOS data, ROPP includes a pre-processing step in order to correct degraded L2 data. The approach is based on Gorbunov et al (2005,2006), and it is used routinely for other GPS-RO missions. Briefly, smoothed L1 and L2 bending angle and impact parameters are computed. An impact height, PC, above which the L2 data is considered reliable, is estimated using an empirical "badness score". The empirical badness score at time t, is defined as,

$$Q(t) = \left(\frac{abs(\overline{p_1(t)}-\overline{p_2(t)})}{\Delta p_a} + \frac{\delta p_2(t)}{\Delta p_b}\right)^2$$

where δp_2 is a measure of the width of the L2 spectrum, $\overline{p_1(t)}$ and $\overline{p_2(t)}$ are the L1 and L2 impact parameters, respectively, computed from smoothed timeseries, Δp_a=200 m and Δp_b=150 m (See also, Eq. 11 Gorbunov et al, 2006 for a slightly modified form). The largest $Q(t)$

value in the impact height interval between 15 km to 50 km is stored as the badness score for the occultation, potentially for quality control purposes.

The mean L1 and L2 bending angle and impact parameters are then computed in a 2 km impact parameter interval directly above PC. Simulated L2 bending angles and impact parameters are computed by adding the mean (L2-L1) differences to both the L1 bending angle and impact parameter values, using the data in the 2 km interval. Simulated L2 and L1 phase values are then computed from these bending angles. Corrected L2 excess phase values are computed by merging the observed L2 phase above PC, with the simulated values below PC, using a smooth transition over 2 km, centred on PC. The corrected L2 phase values are subsequently used in the wave optics processing of the L2 signals.

A difficulty with the GNOS processing is related to determining the impact height PC, used for both the computation of the mean L1 and L2 differences, and defining the transition between observed and modelled L2 phase values. Although the "badness score" is used to determine PC, PC also has a maximum value (20 km). This is defined as the wave optics processing height (25 km) minus a 5 km "safety border". Therefore, the mean bending angles and impact parameters used in the L2-L1 correction can only be computed in a 2 km interval up to a maximum impact height of 22 km. Unfortunately, this is not high enough for GNOS L2 signals, with the result that the mean L2-L1 bending angle and impact parameters computed in the 2 km interval above PC are corrupted.

M. E. Gorbunov, K. B. Lauritsen, A. Rodin, M. Tomassini, and L. Kornblueh (2005), Analysis of the CHAMP Experimental Data on Radio-Occultation Sounding of the Earth's Atmosphere, Izvestiya, Atmospheric and Oceanic Physics, 41, No. 6, 726–740.

Gorbunov, M. E., K. B. Lauritsen, A. Rhodin, M. Tomassini, and L. Kornblueh (2006), Radio holographic filtering, error estimation, and quality control of radio occultation data, J. Geophys. Res., 111, D10105, doi:10.1029/2005JD006427.

4. *The paper by Zou and Zeng is in the reference list, but is not discussed nor referenced in the text. Please provide a comparative analysis of the old and new QC methods with the explanation of why the old QC methods are not sufficient for your data analysis. In particular, will the "badness score" introduced by Gorbunov et al. and successfully applied for CHAMP, COSMIC, METOP and other observations, be also useful for the FY3C/GNOS data analysis? If not, why?*

A: Thanks for the comments. More references will be cited and discussed in the revised manuscript. Originally, we'd like to find out a method to identify the quality of GNOS profiles based on physical meaning and without using background data, just as the "badness score". When we look at the performance of "badness score", it is not suitable for GNOS (see fig1). The values of L2 badness score range from 15 to 1000 plus. The reason might be related to some empirical parameters. Other missions work well using "badness score" since the lowest SLTA of L2 is low enough. When discussed with scientists from EUMETSAT, GRAS can get down to 15km for more than 90%. But it is not the case for GNOS. Only 70% of L2 can be reached below 20km. So the noise_estimate parameter is used as a quality indicator, which could show the performance of L2 extrapolation.

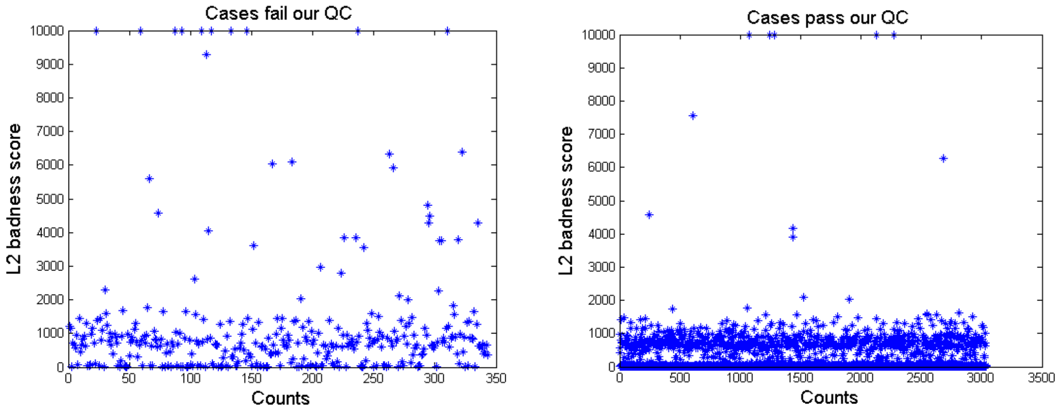

Fig 1. The cases of L2 badness score fail our QC and pass our QC

5. *Page 9 (lines 3–7): The physical meaning of noise_estimate is easy to understand.*
What is easy to understand is the fact that $\Delta\alpha$ is restricted to be close enough to its estimate obtained from a simple ionospheric model. Nevertheless, it is a good idea for the authors to explicitly mention this rather than appeal that something is "easy to understand". Still, some questions remain. Does n in formula (4.1) stay for refractivity of number of data? Number of data is definitely missing somewhere, because the sum in this formula needs to be normalized by the number of data. If n is refractivity, at what height is it taken? Provide explanations or definition regarding n.

A: Thanks for the suggestion. n in formula 4.1 is the number of data. This will be fixed in the revised manuscript.

6. *Page 11 (lines 21–22): The GRAS standard deviations are worse in the troposphere might due to sampling; essentially GRAS is able to measure more difficult cases.*
This statement needs more explanation. What are "more difficult cases"? Do they mostly occur in tropics? Can the authors provide any examples? Is it possible to evaluate a regionalized statistics (tropics, mid-, and polar latitudes)?

A: The comparison between GRAS and GNOS is not the most important part of the manuscript, thus a general remark is made. However, the work is worthy to be done. We'll see if it is possible to add statistics related to the two data.