Dear Editor,

Thank you very much for your comments and suggestions. We have carefully revised the manuscript accordingly. The revisions and responses are listed below.

Q1: Regarding reviewer #2, Question 6: I think what might be meant here is that one could for example assimilate wind. Water vapor could then still be used for evaluation.
Response: The purpose of this manuscript is to conduct an interesting and fair comparison between the WRFDA model and the tomography method in retrieving wet refractivity. For the sake of fairness, both techniques use only the GPS troposphere delay products. At current stage, we only want to know how the much improvement can be achieved by assimilating the GPS tropospheric delay products into the WRFDA model.

Q2: I also do not understand why 3DVAR was applied only once at the beginning of the period of interest.
Response: Because we just need the data assimilating results at certain epochs (0:00 and 12:00 UTC) and these epochs are also the beginning of the interested periods. We have the validating data (radiosonde data) only at 0:00 and 12:00 UTC, therefore we just need the assimilating results and tomography results at these epochs. We start running the WRFDA model at 0:00 and 12:00 UTC, so we only need to run 3DVAR once.

Q3: Regarding reviewer #2, Question 7: since not only WRFDA is used, but also WRF is run, I do not understand why you "don't need the model to develop its state in time". If you are using WRF output and the domain is too small to represent the relevant scales, in my opinion, this could very well be a major problem.
Response: We find that we caused a lot of confusion by running the WRF model. Previously, we first run WRFDA 3DVAR to assimilate ZTDs and generate the output, labeled as output1. Then, we run the WRF using output1 as the background value to generate output2 at the same epoch. Actually, we just want to do data assimilating other than hindcasting, we don't need to run WRF to generate output2. I think running the WRF after 3DVAR caused a lot of confusion, we have removed the WRF part from the manuscript and redone the validation using output1 instead of output2. By these revisions, we have focused on the data assimilation and the WRFDA model.

Q4: In principle, I think that it would be good to find out how sensitive the results are to the domain layout (size, position, and resolution). Furthermore, WRF provides a large choice of physics parameterizations, and choosing specific parameterizations could also impact the results.
Response: Now, we only run WRFDA in the manuscript. The WRFDA has many options for different physical parameterizations. In order to find the best choice for the data assimilation experiment, we follow Chien et al. (2006) to set the schemes as listed in Table 1. We carry out the sensitivity test at 00:00 UTC 22nd July in 2015. The initial

domain size is set to 30 × 24 grids which just cover the study area. The grid size is 3 km × 3km. Then, we run WRFDA using the different setting schemes. The radiosonde data are used to validate the wet refractivity derived by the WRFDA output. Table 1 shows that all schemes for WRFDA has the same bias, standard deviation (STD), and Root Mean Square (RMS). It appears that the output wet refractivity is not affected by the physical parameterization settings in WRFDA.

Table R1. Physical parameterization schemes and statistics of bias, RMS and STD of wet refractivity validated by radiosonde data. Unit is mm/km.

	PBL physics	cumulus physics	microphysics	bias	STD	RMS
1	Yonsei University	Kain-Fritsch	WSM 5-class	-3.95	6.55	7.51
2	Yonsei University	Betts-Miller-Janjic	WSM 5-class	-3.95	6.55	7.51
3	Yonsei University	Grell-Freitas ensemble	WSM 5-class	-3.95	6.55	7.51
4	Yonsei University	Kain-Fritsch	Ferrier	-3.95	6.55	7.51
5	Yonsei University	Betts-Miller-Janjic	Ferrier	-3.95	6.55	7.51
6	Yonsei University	Grell-Freitas ensemble	Ferrier	-3.95	6.55	7.51
7	Mellor-Yamada-Janjic	Kain-Fritsch	WSM 5-class	-3.95	6.55	7.51
8	Mellor-Yamada-Janjic	Betts-Miller-Janjic	WSM 5-class	-3.95	6.55	7.51
9	Mellor-Yamada-Janjic	Grell-Freitas ensemble	WSM 5-class	-3.95	6.55	7.51
10	Mellor-Yamada-Janjic	Kain-Fritsch	Ferrier	-3.95	6.55	7.51
11	Mellor-Yamada-Janjic	Betts-Miller-Janjic	Ferrier	-3.95	6.55	7.51
12	Mellor-Yamada-Janjic	Grell-Freitas ensemble	Ferrier	-3.95	6.55	7.51

Reference:

Chien F C, Hong J S, Chang W J, et al. A sensitivity study of the WRF model. Part II: verification of quantitative precipitation forecasts[J]. Atmos. Sci, 2006, 34(3): 261-276.

In order to figure out how sensitive the wet refractivity output is to the domain size, we carry out a sensitivity test at 00:00 UTC 22nd July in 2015. And we increase the domain size gradually from 30 × 24 grids to 190 × 184 grids. In each run, we validate the wet refractivity derived by the WRFDA output using the radiosonde data. The statistical results of the sensitivity test are shown in Figure 1. It shows that the smaller domain size has the smaller bias, STD, and RMS. So, the domain size of the data assimilation experiment is set to 30 × 24 grids which just cover the study area. This has been discussed in lines 115-138 in the manuscript.

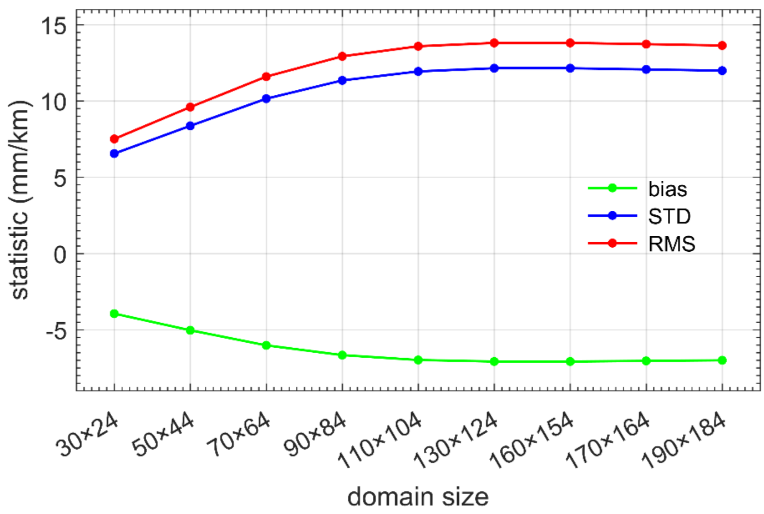

Figure R1. Statistics of sensitivity test with different domain size.

Q5: l. 3, l. 34, l. 58, and elsewhere : filed -> field (and also fileds -> fields)
Response: Thank you. We revised it in the manuscript. Lines: 3, 31, 55, 66, 70, 71, 77, 79, 293.

Q6: l. 100: And the start time of the WRF model is epoch of interest. -> please be more specific
Response: We have revised it to "And we run the WRFDA model at 0:00 UTC and 12:00 UTC, corresponding to the radiosonde observation time." See lines 102-103.

Q7: l. 171: NCL: NCAR Command Language, I suggest to add a reference (see e.g. https://www.ncl.ucar.edu/citation.shtml)
Response: We added it in the manuscript. Lines: 183 and 410-411.

Q8: l. 180: I think WRF output1 and WRF output2 would be more accurate than reanalysis1, reanalysis2.
Response: Thank you, and we revised Reanalysis1 to Output1 and Reanalysis2 to Output2 in lines 109-110 in the manuscript.