Referee 2：

General comments:

The paper "Impact of processing strategies on Long-term GPS ZTD" concerns investigating the influence of the selected GNSS observation processing strategies on the reliability of position and ZTD. In general, the Authors compared the GPS processing approach, which differs in mapping function (GMF, GPT2, GPT3, VMF1, VMF2) and elevation cut-off angle (3°, 7°, 10°,15°, 20°, 25°, 30°). They have mainly focused on the ZTD time series but also provided some basic results regarding position repeatability. Although GNSS meteorology is a well-known concept, it still requires improving existing algorithms and validating possible/new solutions, including new mapping functions. In light of this, the general idea of the paper is justified. However, the complexity of the impact of individual observation processing elements on the reliability of the final solution is very high. Hence it requires a very detailed analysis, which in my opinion, has not been done by the Authors.

Response: We appreciate all the valuable comments and suggestions. All the comments are responded point by point as shown below. We have added more detailed analysis accordingly in the revised manuscript which can hopefully make the reviewer more satisfied.

Specific comments:

Firstly, there is no information about trend estimation uncertainties, which are significant when assessing various solutions. Some differences between different observation processing strategies are expected, but assessing their significance is the most important.

Response: In fact, we have given the trend uncertainties in Figure 4 and Figure 7 (now Figure 6), but we did not discuss it accordingly. Following the reviewer's suggestions, we have added analysis about trend estimation uncertainties in the revised manuscript. Please see L195-197 and L215-216.

The Authors have analysed 46 IGS stations, while only 19 have presented results in Figures 4 and 7. There is no appendix to see what is happening with the rest of the stations. Figure 8 presents results for all stations?, but it is unclear. Additionally – we do not have any information about data quality. The data completeness probably varies for different stations and may affect final solution. ?

Response: 44 IGS stations were selected in this study, considering two factors: 1) having first GPS observations before 1999, and 2) having collocated radiosonde stations within 100 km in horizontal and 150 m in vertical. Among the 44 IGS stations, 19 stations with common ZTD time series between GPS and radiosonde longer than 15 years were selected in the trend analysis. In the revised manuscript, we also excluded months in which GPS observations were less than half the time, so that two more stations (BRMU and JOZE) were rejected. Finally, 17 stations were used for comparison and analysis of trends. We have explained this in L156-158.

Figure 6 (now Figure 5) and Figure 8 (now Figure 7) presents average results for the 19 stations (now 17 in the revised manuscript), we have made it clear in the figure captions in the revised manuscript.

We have added Table 9 to show the data length in the revision. The length of the data is calculated based on the number of the months.

I am also wondering why the Authors have used 1995-2014? Before 2000, quite a poor quality of orbits and SA negatively affect GPS solutions. The station selection is also questionable – 100 km is a lot and may result in different troposphere conditions. Here a table with exact differences between GPS and RS sites is necessary. Moreover, Dai et al. (2011) presented a homogenised dataset until 2011 (or at least that's what the text says). But if the GPS data were processed until 2014, what was the reference for the last three years?

Response: In this study, we reprocessed the GPS ZTD using the IGS repro2 orbits products in order to avoid the inconsistency in the data processing models and strategies. The IGS repro2 covers the period from 1995 to 2014, and we therefore used 1995-2014. SA mainly affects the broadcast ephemeris and should has little effect on the result of reprocessing.

We have added Table 4 to show the distances between GPS and RS stations.

In our study, we only used the homogenized RS data from 1995 to 2012 and the data from 2013 to 2014 were not used. The homogenized radiosonde dataset from 1995 to 2012 was provided by Junhong Wang from University at Albany, SUNY. She homogenized the dataset by using method proposed in Dai (2011). We have added this in the Acknowledgments, please see L281-282.

Why did the Authors decide to verify different mapping functions using 7° cut-off angle?

Response: We need to fix an elevation cut-off angle when comparing different mapping functions. The reason why 7° cut-off angle was used is that some analysis centers (i.e., GFZ, JPL and WHU) also use 7° cut-off angle in data reprocessing, and the position accuracy derived from GMF is slightly worse when using elevation cut-off angles lower than 7° (Dousa et al., 2017 and Qiu et al., 2020).

Dousa, J., Vaclavovic, and P., Elias, M.: Tropospheric products of the second GOP European GNSS reprocessing (1996-2014). Atmos. Meas. Tech., 10, 3589–3607. https://doi.org/10.5194/amt-10-3589-2017, 2017.

*Qiu, C.; Wang, X.; Li, Z.; Zhang, S.; Li, H.; Zhang, J.; Yuan, H: The Performance of Different Mapping Functions and Gradient Models in the Determination of Slant Tropospheric Delay. Remote Sens. 2020, 12, 130. https://doi.org/10.3390/rs12010*130.

The Authors wrote, "The method for calculating ZTD from ERA5 can be referred to Haase et al. (2003)" – please be more specific about whether used by the Authors method is the same as in Haase, or not. Additionally, what is a temporal and (even more important) spatial (vertical) interpolation between ERA5 and GPS site – there is no information about this.

Response: To make the method clearer, we have replaced "The method for calculating ZTD from ERA5 can be referred to Haase et al. (2003)" with "The method described in Haase et al. (2003) was used for calculating ZTD from ERA5". ERA5 products have improved temporal resolution of 1 h and we also used GPS ZTD products with temporal resolution of 1 h, so there is no need to conduct a temporal interpolation. Regarding the spatial interpolation, we used the method described in Zhang et al. (2017). We have added this information in the revised manuscript. Please see L98-99.

Zhang, W. X., Lou, Y. D., Haase, J., Zhang, R., Zheng, G., Huang, J., et al.: The use of ground-based GPS precipitable water measurements over China to assess radiosonde and ERA-Interim *moisture rends and errors from 1999 to 2015. Journal of Climate, 30, 7643-7667. https://doi.org/10.1175/JCLI-D-16-0591.1, 2017.*

The ERA5 homogenised dataset (according to Dai et al. 2011) should cover a time span until 2011. Please be more specific about the exact source of radiosonde data (the link given in Dai et al. 2011 does not exist at this moment). This also concerns ‘raw Radiosonde’. That would be helpful for the readers. Additionally, please add the info on whether the Authors used exactly the Haase (2003) method for calculating ZTD from RS.

Response: You mean the radiosonde homogenized dataset? The homogenized radiosonde dataset from 1995 to 2012 was provided by Junhong Wang from University at Albany, SUNY. She homogenized the dataset by using method proposed in Dai (2011). We have added this in the Acknowledgments. In the manuscript, we used the homogenized RS data from 1995 to 2012 and the ‘raw Radiosonde’ products from 1995 to 2014.

To make the method clearer, we have replaced “The method for calculating ZTD from radiosonde observation can be referred to Haase et al. (2003)” with “The method described in Haase et al. (2003) was used for calculating ZTD from radiosonde data”. Please see L107-108.

I’m not sure why the Authors have focused on analysing position accuracy since there are no conclusions (just a description of the results) and, more importantly, the results from this part of the manuscript were not considered in any other part. The small variability of position is rather excepted and obtained differences are very small (hundredths of a millimetre).

Response: In this work, we reprocessed the 44 IGS station data from 1995 to 2014 by using different strategies. The main purpose of this work is to access the impact of these strategies on GPS ZTD. Since ZTD and coordinate up component are strongly correlated, we decided to analyze the coordinate accuracy first. Regarding the position accuracy, mapping functions have small impact than cut-off angle, but when the interested product is station position, we do not recommend using cut-off angles higher than 15° in data processing. We have made the conclusions clear in the revised manuscript. Please see L131-132.

There is also no specific conclusion from analysing bias, STD and RMS from differences between GPS ZTD and ERA5 ZTD

Response: Following your suggestion, we have added specific conclusion about GPS ZTD accuracy. Please see L146-147.

There should be more discussion regarding the impact of homogenisation on long-term trends. It is clear that adopting various homogenisation approach influence the final solution the most (since homogenisation may 'fix' even distinct inhomogeneities resulting from adopting various processing strategies). Several papers concern different methods of GNSS time-series homogenisation. It is unclear from what the Authors wrote whether the changepoints they found are correct, better/worse than changepoints that may be found with other approaches.

Response: We agree with the reviewer that different homogenization methods do make a significant difference to trends. The PMTred method has been widely used in long-term trend studies, such as Xu et al. (2013), Ning et al. (2016), and Li et al. (2017). In the revised manuscript, our results also show that the influence of different strategies on ZTD trends is weakened after homogenization, which may illustrate that our homogenization method is effective to some extent. We have added more discussions on the impact of homogenization on long-term trends, please L14, L212-214 and L220-221.

Xu W., Li Q., Wang X., Yang S., Cao L., and Feng Y.: Homogenization of Chinese daily surface air temperatures and analysis of trends in the extreme temperature indices. Journal of Geophysical Research, 118. https://doi.org/10.1002/jgrd.50791, 2013.

Ning T., Wickert J., Deng Z., Heise S., Dick G., Vey S., and Schöne T.: Homogenized time series of the atmospheric water vapor content obtained from the GNSS reprocessed data. J Clim 29:2443–2456. https://doi.org/10.1175/JCLI-D-15-0158.1, 2016.

Li Y., Wang G., Han X., Li H., Fan W., Liu K., and Wang H.: Homogenization of Sea Surface Temperature at Xiao Changshan marine station in the east of the Bohai Sea using the PMT method. IOP Conference Series: Earth and Environmental Science, 52, 012055. DOI: 10.1088/1742-6596/52/1/012055, 2017.

Figures 4 and 5 make me worry about the reliability of the homogenisation process. After taking a closer look at e.g. JOZE station, we can see that the trends are similar before homogenisation, while after homogenisation there is a distinct difference between VMF1 and VMF3. These mapping functions rely on the numerical weather model and are very similar regarding the a and b coefficients. Therefore such differences are unexpected. BRMU station also looks interesting – before homogenisation all trends are similar, after homogenisation, there is a distinct division between climatological and discrete mapping functions. At this point, I would not worry about the comparison to the RS since it may even be 100 km away (there is no info about that).

Figure 5, in turn, makes me worry about the reliability of the GPS observation processing. Presented by the Authors monthly ZTD anomalies present a distinct shift in the case of GPT3 mapping function, while using VMF3 there is no such situation. The main problem is that GPT3 is a climatological mapping function and is therefore continuous. Therefore presented by the Authors shifts in this particular solution are not a problem of GPT3, but of the processing itself.

Response: Following your comments, we carefully checked our estimated results. We found the strange phenomenon was caused by our mistake. We forgot to exclude the ZTD estimates with the number of GNSS observations being zero. The data processing software will use the a-priori ZTD as output when observations missing. For different mapping functions, different ZTD priori models were used. For example, GMF used the GPT model, and VMF1 used the ZTD prior values provided by VMF1. This leads to the obvious difference between ZTD from using GPT3 and VMF3 when the observations are missing, resulting in the difference in the detected changepoints.

In the revised manuscript, we have fixed this problem by excluding the ZTD outputs when observations missing. Now different mapping functions have little influence on the accuracy of the results and the changepoints detection.

I am also not sure why the Authors focus on ‘Raw radiosonde’ as a reference since they stated in the introduction that RS homogenisation is important. I am also not sure why the Authors focus on the un-homogenized GPS ZTD time-series and, based on them, assess various cut-off angles. “However, for other situations, i.e., taking Dai- or ERA5-derived ZTD trends as references for un-homogenized GPS ZTD evaluation….”. If we already know that GPS time series may be affected by various factors (e.g. antenna/receiver changes), why should we focus on un-homogenized ZTD, while comparing it to the reference set?

Response: We have removed the Raw radiosonde comparison in Figure 4 and 6 in the revised manuscript, but still retained the Raw radiosonde results in Figure 5 and 7 for comparison with conclusions from previous studies, and we found that taking Raw data as a reference, we came to a similar conclusion to Ning et al. (2012).

In the study, we want to analyze not only the imapcts of mapping function and elevation cut-off angle on the long-term GPS ZTD, but also the impact of homogenization on it. Thus, we still retained the un-homogenized GPS ZTD for comparison with the homogenized GPS ZTD to analyze the impact of homogenization, especially the impact of homogenization on the trends of GPS ZTD estimated from different strategies.

To all figures and tables – please change their description to make it possible to correctly understand the presented in them results, without looking for information in the manuscript's main body. Figures are often not well readable.

Response: Following reviewer's suggestions, we have modified all figures and tables to make them well readable. Please see our responses in 'More detailed comments'.

Overall it seems that the presented paper covers too many issues that are too briefly analysed. A proper analysis of each of its elements (i.e. the impact of processing strategy on ZTD, the impact of processing strategy on position and homogenisation on long-term ZTD reliability) is a big task. Therefore it is hard to find reliable outcomes from the conducted analysis.

Response: We agree with the reviewer that the long-term ZTD analysis is a big and complex task and we are trying our best to do some contribution. Specifically, we have added more analyses about ZTD trend uncertainties (L195-197 and L215-216), and have also supplemented each section with clear conclusions based on your suggestions. Please see L131-132 and L146-147.

More detailed comments:

Figure 1 - there is no 'BOGO' station in IGS

Response: Yes, you are right. We checked the list of IGS sites and removed ‘BOGO’ and ‘CASC’ stations. Please see Figure 1.

Page 6, Table 4 – please add info that all cut-off angles were tested using VMF3 (I know it was pointed out, but the table should be read correctly, without looking for further information in the manuscript body

Response: corrected. Please see Table 5, L133 and Table 6, L134.

Page 7 Tables 5 and 6 – same as above, but regarding cut-off angle, and mapping function

Response: corrected. Please see Table 7, L148-149 and Table 8, L150-151.

Page 7, Tables 5 and 6 – please add info that this is a difference

Response: We have added this information. Please see Table 7, L148-149 and Table 8, L150-151.

Page 10, Figure 5 – add y-axis description to the figure

Response: We have removed the Figure 5.

Page 9, figure 2 – the colours are way too similar. Instead of the legend, I suggest you add the solution name to the axis

Response: We have add the solution name to the axis. Please see Figure2.

Page 11, line 208 – shouldn’t it be Baldysz et al.2016?

Response: It should be Baldysz et al. (2018) and we have corrected it. Please see L218.

Page 12, Figure 8 – The figure description should be corrected (left/right instead of top/bottom)

Response: corrected. Please see Figure 5, L206 and Figure 7, L228.

Page 13, lines 233-235 – this is rather expected. Since we estimate differences between GPS ZTD and ERA5 ZTD and then use these differences to correct GPS ZTD time series, the final GPS ZTD solution will be similar to the ERA5

Response: Yes, you are right. That's why we focused on the ZTD time series homogenized by 'ABS' method rather than by 'REL' method.