Dear reviewers,

Thank you for your careful reviews and suggestions. Followings are our responses to your comments. Please note that the line numbers in our responses are the numbers in the revised manuscript using track change.

Response to Anonymous Reviewer #2:

1. GPT2w model can also provide the Tm globally for real time applications. Authors should include this model in the comparisons

Response: We have included Tm estimations from GPT2w model in all of our comparisons. And the results still indicate that our Ts-Tm model has accuracy advantages over other Tm estimation models. Detail statistics are included in the revised manuscript. (Line 452~775)

2. What do the numbers in Figure 4 stand for? Are they results from the radiosonde? If so, you can also use the circles as shown in Figure 1.

Response: The numbers in figure 4 are the contour values. In static Ts-Tm model Tm = *a**Ts + *b*, *a* stands for slope constant and b stands for intercept constant. We estimated *a* and *b* at each gird node of ERA-Interim data. In figure 4, top figure is the global color contour of *a*, middle figure is the global color contour of *b*, and bottom figure is the global color contour of Ts-Tm model RMSE.

3. Explain m1, m2, n1, n2 in equation (5).

Response: (m_1, m_2), (n_1, n_2) and (p_1, p_2) are the fitting coefficients of formula (5) items, and these coefficients can indicate amplitudes of annual, semiannual and diurnal variations in our Ts-Tm models. We have included these explains in the revised manuscript (Line 431~432).

4. When interpolating the Tm onto the GNSS sites, did you consider the impact of height differences?

Response: We have considered two types of height differences in our interpolations:

(a) The difference between geodetic height (applied by GNSS) and geopotential height (applied by NWP). A GNSS site's geodetic height is converted to altitude height using EGM2008 model, and the difference between altitude height and geopotential height is neglected as we done in our another study(Jiang et al., 2016);

(b) The height differences between GNSS sites and NWP levels. For a GNSS site, we estimated the Tm at its four neighbor gird nodes in ERA-Interim data and then horizontally interpolated them onto GNSS site's location. At each gird node, the Tm integral geopotential height range is from GNSS site's geopotential height to 1 hPa level's height. The pressure, temperature and humidity at GNSS site's geopotential height is interpolated from its upper and lower ERA-Interim levels.

Reference:

Jiang, P., Ye, S. R., Chen, D. Z., Liu, Y. Y., and Xia, P. F.: Retrieving Precipitable Water Vapor Data Using GPS Zenith Delays and Global Reanalysis Data in China, Remote Sensing, 8, 10.3390/rs8050389, 2016.

Response to Anonymous Reviewer #3:

We have turned to a language editing service when prepared the revised manuscript. And the editors tell us that they will also offer English language copy-editing for final revised accepted papers. Followings are our responses to your comments:

1. Consider using the term integrated water vapour, IWV, instead of PWV.

Response: In our opinion, IWV has the same meaning as PWV. And "GPS PWV" seems to be used together more widely in many articles. So we choose the term "PWV" rather than "IWV".

2. Section around line 65: Is it well established that global empirical Tm models without Ts components are less good? If so give some references. If not, include some of them in your comparisons of different sources of global Tm. We want to know whether your dataset is the best global set for Tm estimation around for the moment, or just better than other Ts-Tm based sets.

Response: We have included Tm estimations from GPT2w model in all of our comparisons. GPT2w model is a global empirical Tm models newly developed by Bohm in 2015 (Bohm et al., 2015). And the results still indicate that our Ts-Tm model has accuracy advantages over other Tm estimation models. Detail statistics are included in the revised manuscript. (Line 452~775)

3. Notice that in numerical weather prediction one in general uses GPS ZTD, not PWV. PWV is important for climate monitoring, and for meteorologists doing weather forecasting combining information from weather prediction models and observations.

Response: Thank you for your kind suggestion. We have modified our expressions about the use of GPS PWV in weather prediction. (Line 166)

4. around line 100: Did you take into consideration that water vapour pressure (and density) varies approximately exponentially with height when doing the integral in eq. 4. It is not likely to have a large impact, but the fewer RS levels you have access to, the larger the effect. 5 levels is not a lot.

Response: Yes, we have considered the temperature and water vapor pressure's variations with height. Between two neighbor RS observation heights, we calculated the water vapor pressure at the middle height for equation (4) by exponential interpolation of the two RS heights' water vapor pressure, while we estimated the temperature by linear interpolation. We have added some explanations in our revised manuscript. (Line 278~279)

4. around line 120: "geoid height" should be "geometric height". The reference surface doesn't

matter for the integral.

Response: Revisions have been made. (Line 312)

5. around line 150: There are several places where text is not properly separated, here E180. etc., text of figure 3 is another example.

Response: Revisions have been made. (Line 361, 394, 663, 665, and 753)

6. Figure 1. It is hard to see properly the RS circles. Consider making the figure little bit larger, and draw a thin black line around the RS circles, in order that one can see them also where they agree with ERA.

Response: Figure 1 have been plotted bigger. Actually our figures are vector graphs, so readers can zoom in on them to see the plots more clearly.

7. At many places in the text and in the figures units are missing.

Response: Revisions have been made.(Figure 2,3,4,5,6,7,9,10,11,12, and 13)

8. Figure 6: Are the colors plotted in a particular order, such that for example large rms will be plotted on top of small rms? If so, do a check that plotting in the opposite order yield almost similar plots. Otherwise enlarge.

Response: We have enlarged our plots in figure 6 to ensure that no color point is covered by others.

9. Figure 6: Regarding the RS ERA comparisons. Is anything done to handle altitude offsets between RS surface and ERA surface?

Response: In Equation (5), which is the time-varying Ts-Tm model, there are eight coefficients (a, b, m_1, m_2, n_1, n_2, p_1, p_2) which are estimated at each ERA-I grid node. In figure 6, we evaluated the performance of equation (5) at each RS site. Considering the horizontal offsets between RS sites and ERA-I grid nodes, we obtained the eight coefficients (a, b, m_1, m_2, n_1, n_2, p_1, p_2) at each RS location by horizontally interpolated the ones of RS site's four neighbor ERA-I grid nodes. However, we think that the height differences between RS sites and it neighbor ERA-I surfaces have little impact on our Ts-Tm models (not Tm estimations). Such impact on Ts-Tm model can be compensated by the Ts input, which should be changed with altitude.

10. Around line 280. At 10.82 % of the sites inclusion of the time variations in ERAI resulted in a poorer results. That indicates ERAI has particular problems at these locations. If you plot them on a map, do you see any systematics in their location?

Response: GPT2w model has been added in our comparisons and we modified our results in the revised manuscript.

Unfortunately we found that there is no obvious characteristic in the distribution of RS site with poorer results. The reasons for such poorer results seems to be complicated and need specific study in the future.

11. Around line 310. The pressure used to determine ZHD should be the pressure at the GNSS antenna level, not surface pressure. Did you do something to correct for height offsets, or is the barometer installed at the same altitude as the GNSS antenna at these locations?

Response: The air pressure for each ZHD calculation was measured by the barometer equipped together with GNSS antenna at the GNSS site. Their locations are considered to be the same. (Line 694~696)

12. Similarly the Tm integral should in principle run from the antenna level and up, not from the surface. In almost all cases that is not likely to create problems, but there will be locations where the difference between the surface altitude of ERA and the altitude of a GNSS site is huge. I'm not familiar with the location of IGS sites, but for GNSS reference sites in general the altitude difference

can be more than 1000 m between a GNSS site and an NWP model with higher horizontal resolution than ERAI.

Response: We have considered such height differences in our comparisons. We interpolated (or extrapolated) ERA-I profile to GNSS antenna's location, and then started the Tm integral from GNSS antenna's altitude.

13. Around line 385: It would nice if in the final article you could add an extra line with a link to your dataset. It seems very useful to many people.

Response: A link to our model has been added. (Line 833).

Reference:

Bohm, J., Moller, G., Schindelegger, M., Pain, G., and Weber, R.: Development of an improved empirical model for slant delays in the troposphere (GPT2w), Gps Solutions, 19, 433-441, 10.1007/s10291-014-0403-7, 2015.

Response to David Adams:

General Comments

1. Firstly, this paper is in terrible need of proofreading by a native English speaker. The paper is very poorly written and is nearly incomprehensible is some parts. Correcting grammar errors is not the responsibility of the reviewer; this paper should have been checked beforehand. After line 170, I have stopped correcting these grammar errors and I just focus on content.

Response: Thank you very much for your careful corrections. We have turned to a language editing service when prepared the revised manuscript. And the editors tell us that they will also offer English language copy-editing for final revised accepted papers.

2. The subject of the Bevis and other models has been examined many, many times, globally and

regionally. See the work of June Wang (SUNY) and for example Luiz Sapucci (Brazil). There has to be a very strong motivation to continue this type of work. Improvements of 1 or 2mm in PWV are not very impressive and hardly seem worthwhile considering all of the other inherent errors in the GPS PWV method itself, errors in radiosondes, errors in ERA Interim data and other reanalysis products (which may not be independent of radiosonde in the first place).The authors need to better justify why these tiny improvements given the work required are really advantageous. More importantly, errors due to surface pressure are much larger. The height of the antenna relative to the surface pressure measurement will introduce a much larger error than what you present here with Tm models. You should do an analysis of these errors. How large are the errors associated with a reasonable error or 5mb in surface pressure? Likewise, the assumption of ZHD representing the mass of the atmosphere (excluding water vapor) is erroneous. What are the errors associated with this assumption? Compare the Amazon to a desert region to assess this error somehow.

Response: The uncertainty of the PWV estimated from GNSS observations has been discussed comprehensively in detail by Ning's research in 2016 (Ning et al., 2016). In his study, the uncertainties in ZTD estimation, ZHD estimation and the conversion factor Q from ZWD to PWV were analyzed and their contributions to the total uncertainty of GPS PWV were evaluated. However, Ning's study assumed the Tm were obtained from NWP models. We estimated the Tm from surface air temperature Ts. So we replaced the uncertainty of Tm by our statistical results. Considering the uncertainty of Tm is related to the uncertainty of conversion factor Q, we calculated the percentages of total GPS PWV's uncertainty due to the errors in Q at radiosonde stations over the world. We found that the contribution of Q's uncertainty to the total GPS PWV's uncertainty has been dropped significantly by using more precise Ts-Tm model. The experiments and results are shown in line

613~652 of the revised manuscript.

The differences between the statistical results of various Tm models showed in table 2 seemed to be tiny. Actually the improvements due to the application of more precise Tm model were very evident at a considerable number of sites (or during wet seasons) as showed in figure 8/9. But such significant improvements were smoothed in the statistical results with respect to all sites and all seasons.

Specific Comments.

1. Check your spacing on the in-text citations.

Line 10 Write “In near real-time GPS-PWV retrievals”

Line 13 Write “ without data smoothing”,

Line 14 “Then static and time-varying global gridded T s –T m”,

Line 16 Write “have prominent advantages over other”,

Line 17 Write “Large biases in Bevis’ equation or in latitude-related linear models at a considerable number of stations”,

Line 18 Write “Multiple statistical tests at the 5% significance level”,

Line 29-30. This sentence is awkward, rewrite,

Line 31 Write “such as radiosondes and water vapor radiometers,”,

Line 44 Write “The mapping function...”

Line 54 Write “ The integration of”

Line 66 Write “ can be lost without the constraint of real data.”,

Line 68 Write “numerical weather prediction.”, Line 81 Write “oceanic regions” not sea regions,

Line 87 Write “spatial smoothing of the data”,

Line 105 Write “We employed radiosonde data from the Integrated”,

Line 107 Write “may be”,

Line 120 Write “However, water vapor is solely concentrated in the troposphere, and most of it, specifically within the first 3 kilometers above sea-level.”,

Line 137 Write “We first carried out a linear regression analysis”,

Line 145 Write “It is evident that T m varies”,

Line 169 Write “Since the T s -T m relationship has”

Response: Grammar revisions have been made in the revised manuscript.

2. Line 36. Slant water vapor(SWV) are not “widely used”

Response: We deleted SWV in the revised manuscript.

3. Line 37-38 I don´t know what you are trying to say here.

Response: We rewrite the sentences in the revised manuscript. (Line 59~60)

4. Ps also has a significant effect on PWV calculated from GPS signal delays, because it represents the mass of the dry atmosphere.

And Line 106. You should show PWV sensitivity to surface pressure. Using Bevis´ model, the sensivity if about 3 times greater to surface pressure than to Ts.

Response: As our response to general comments #2, the impact of Ps was considered in the newly added experiment.

5. Line 58-60 Rewrite this sentence, it is difficult to follow.

Response: We rewrite this sentence to "However, another data source, radiosonde data, has low spatial and temporal resolution. At most of the radiosonde sites, sounding balloons are daily cast at 00:00 UTC and 12:00 UTC. Furthermore, a large amount of GPS stations are not located close enough to the radio sounding sites." (Line 140-142)

6. Line 80. "there exist large differences between the oceanic and terrestrial atmospheric properties." This depends. At the surface layer and boundary layer, yes, but about the boundary layer it is less clear. You should specify what you mean here.

Response: We rewrite this sentence to "Significant differences exist between the oceanic and terrestrial atmospheric properties, especially at the surface layer and boundary layer. The change of Ts from land to ocean may be very different from that of Tm." （line 179~180）

7. Line 85 "however is statistic and the estimated T m residuals due to time variations are not fixed (Yao et al., 2014a)." I don´t understand what you are trying to say here.

Response: We rewrite this sentence to "However, the Ts-Tm relationship has time variations and can produce residuals in the static Tm estimations (Yao et al., 2014a). Such residuals are not fixed in Lan's model." (Line 184~185).

8. Line 101 What do you mean specifically "atmospheric top" (i.e. top of the atmosphere). Give a

value.

Response: We rewrite this sentence to "The integral intervals are from earth surface to top level. The height of the top level depends on the data sources we employed, which will be shown in section 2.2." (Line 279~281)

9. Surface observations must be available, and top profile level should not be lower than 300 hPa standard level. Oftentimes, the first level of a radiosonde has erroneous data, over influenced by surface conditions. You should quality check the first level of the sounding for bad temperature and humidity data.

Response: The IGRA data providers declare that NCEI scientists have applied a comprehensive set of quality control procedures to the data to remove gross errors (https://www.ncdc.noaa.gov/data-access/weather-balloon/integrated-global-radiosonde-archive). We also performed a quality check to remove the abnormal temperature and humidity data.

10. Much greater the 5 levels 1000mb to 300mb is necessary. Maybe a more stringent criterion is needed. Maybe 10 levels, you should check the sensitivity of your results to this assumption.

Response: We adjusted the criterion to 10 levels (line 291). Furthermore, we considered the temperature and water vapor pressure's variations with height. Between two neighbor RS observation heights, we calculated the water vapor pressure at the middle height for equation (4) by exponential interpolation of the two RS heights' water vapor pressure, while we estimated the temperature by linear interpolation. We added some explanations in our revised manuscript (line 278~279). According to our experiments, the Tm differences due to the number of radiosonde levels are smaller than 0.55 K which are significantly smaller than the Tm's uncertainty shown in table 2. So in our opinion, it is reasonable to employ these radiosonde data to evaluate the errors of Tm estimation model.

The number of radiosonde stations in our comparisons decreased to 723 due to the change of such criterion. Therefore some statistical results changed in figure 6~8 and table 2~3.

11. "Profile data including same elements are usually provided by NWP products at certain vertical levels". It is not clear what you mean by same elements. I assume you mean the same variables. You should also say what the cone of representation of the GPS is, so one can consider that with respect to the "vertical" measurements of the radiosondes.

Response: We have realized the word "same elements" is not proper for what we tried to express. We meant the temperature, relative humidity and geopotential which can be used to calculate Tm. However, there are much more other variables in ERA-Interim, so we rewrite this sentences to "Profile data are usually provided by NWP products at certain vertical levels." In the comparisons between GPS PWV and radiosonde PWV, we described the differences of the physical meaning between the different PWV results. (Line 750~753)

12. Line 125 "to the height replacement will extremely approximate to zero." This doesn´t make sense, rewrite.

Response: We rewrite this sentence to "The Tm value nearly has no change after such height replacement" (line 317~318).

13. Line 140 Correct this sentence.

Response: We rewrite this sentences to "Our analyses also indicated that the correlation coefficient between Ts and Tm is generally related to the latitude. The same conclusion has been drawn in other studies (Yao et al., 2014b)."(Line 331~332).

14. Line 148 This sentence makes no sense.

Response: We realized that there is no experiment to support such guess, therefore this sentence is deleted in the revised manuscript.

15. Remember that the radiosondes are incorporated/assimilated into reanalysis products, so the data are not independent. So you may not be correctly capturing real errors and biases in the data.

Response: As described in line 457~459 of the revised manuscript, our model was developed from the NWP data covering the period from 2009~2012. And the radiosonde data used for the assessments in section 4.3 were observed in 2016. Therefore the two datasets are independent. Of course in both NWP and RS datasets, there may exist common bias due to the errors of sounding sensors. However the sounding data is oftentimes more precise than Ts-Tm model, so we think it is reasonable to evaluate Tm estimation model using radio sounding data.

16. Remember, there is an inherent error in the assumption of the calculation of ZHD, which assumes that the water vapor is not contributed to the mass of the atmosphere. This error may not be important but should be evaluated for very wet regions (e.g, the Amazon) and very dry regions.

Response: Experiments have been added as described in the response to the general comment #2.

17. “Because the Tm from ERA-Interim is believed to be the most accurate” This is a very strong statement. You need to provide evidence considering the purpose of this paper is entirely dependent on the quality of the observations. As I said before these data are not necessarily independent of the radiosondes and their humidity data in regions with few observations can be awful.

Response: We rewrite this sentence to “The Tm from ERA-Interim is believed to be the most accurate among our Tm estimates at the selected GPS sites” (line 712). We assumed this based on some studies which claimed that the NWP datasets were good choices for global Tm estimation (Wang et al., 2005; Wang et al., 2016). We also repeated those comparisons between NWP and radiosonde datasets and concluded the similar accuracy. Although the ERA-Interim data has assimilated the data from various satellites and radiosonde, we found that the ERA-Interim Tm estimates do not agree well with the real observations in some regions. These large errors have impacts on our Ts-Tm models as we described in line 670~673 of the revised manuscript.

Reference:

Ning, T., Wang, J., Elgered, G., Dick, G., Wickert, J., Bradke, M., Sommer, M., Querel, R., and Smale, D.: The uncertainty of the atmospheric integrated water vapour estimated from GNSS observations, Atmos. Meas. Tech., 9, 79-92, 10.5194/amt-9-79-2016, 2016.

Wang, J. H., Zhang, L. Y., and Dai, A.: Global estimates of water-vapor-weighted mean temperature of the atmosphere for GPS applications, Journal of Geophysical Research-Atmospheres, 110, 10.1029/2005jd006215, 2005.

Wang, X. M., Zhang, K. F., Wu, S. Q., Fan, S. J., and Cheng, Y. Y.: Water vapor-weighted mean temperature and its impact on the determination of precipitable water vapor and its linear trend, Journal Of Geophysical Research-Atmospheres, 121, 833-852, 10.1002/2015jd024181, 2016.