The paper by Zhang et al. describes results from an analysis of Precipitable Water Vapor (PWV) data from the MERSI-II instrument on Fengyun-3D. The satellite data are compared to ground based radiosonde data, and a example application for the Qinghai-Tibet Plateau is described.

Although the MERSI-II PWV seems to be a good product, I think the paper needs a major revision before publication. Especially, the issues listed below need to be addressed.

General Comments:

1. The paper uses MERSI-II PWV data, but there is no detailed information (or reference) given about the retrieval algorithm, possible adaptations specific for MERSI-II data, any assumption made or possible external data used in the generation of the PWV product.

2. The paper presents a statistical analysis of collocated data without much physical interpretation. For example, it is not clear (or at least discussed) if observed differences between satellite and ground based data are significant in view of atmospheric variability and/or errors of the products.

3. Actually, the observed differences between the MERSI-II PWV and the radiosonde data are quite small (only a few percent). I would expect that this is within the expected precision/accuracy of the involved products, especially taking into account the large spatial and temporal variability of water vapor. Nevertheless, a lot of effort is made to (statistically) quantify the reasons for these differences, namely distances in space or time or surface altitude, although the findings are essentially as expected (e.g. better agreement for smaller distances). Therefore, this part of the paper can be significantly shortened.

4. The conclusions drawn for the Qinghai-Tibet Plateau application are not supported by the corresponding data/analysis. They are mainly based on the comparisons of two stations, which actually show a very similar seasonal variation (the data agree within their 25%/75% percentile ranges). So, how can there be any conclusion about transport be drawn? Of course, water vapor transport is affected by the terrain / surface elevation, but this is no new finding.

5. The English of the paper needs improvement. In may cases, 'the' is missing in the sentences (also in the title before 'Precipitable'); use of singular/plural is wrong, and wording/formulations are sometimes misleading, e.g., when using in the abstract 'distance discrepancy' instead of 'spatial distance'. Some of these issues are addressed in this review, but there are way too many to be listed completely.

Specific Comments:

1. l. 15:
   Mentioning a 'peak value of the MB' of 0.00 is misleading, especially since in the following underestimations are mentioned. It should be clarified that 'peak value' refers to the histogram analysis.

2. Introduction:
   In the Introduction only NIR satellite water vapor measurements are addressed. PWV is measured from satellites at various spectral regimes (VIS, NIR, SWIR, TIR, MW, ...); especially the MW measurements have a very long heritage. This should at least be mentioned.

3. l. 30:
   'upward radiance over the view of satellite' – unclear formulation. Probably you mean that the radiance measured by the satellite instrument is affected by water vapor. Please clarify.

4. l. 30:
   'For MERIS, the PWV retrieval algorithm employing the ratio of top of atmosphere (TOA) radiance at one water vapor absorption channel (900 nm) to TOA radiance at 885 nm, which is outside water vapor absorption region (Bennartz and Fischer, 2001).'

   This is not even a real sentence - please reformulate.

5. Section 2.1:
   This section lacks a description (or a proper reference to) the MERSI-II PWV retrieval algorithm. Some information is given in the referenced paper from Yang et al. (2019), but this is not clear from how the paper is referenced now, and the contained information is very limited. Especially, information should be given on assumptions made in the retrieval, any external data used in the retrieval, consideration of surface elevation, expected (random) error of the product, handling of aerosols, etc. — all information which actually defines the product. Furthermore, some instrument details should be given, e.g. swath width and spatial/temporal coverage.

6. l.121:
   The radiosonde product is defined as surface-to-500-hPa PWV - this means that it does not include some percents of the total column. Has this been considered in the comparisons? This may be relevant as observed differences are also in the order of a few percent.

7. Section 3.1:
   This section could possibly be shortened as all statistical indicators except for EE are quite common. However, the definition of EE is not fully clear. In the text, EE is described as a limit in percent whereas eq. 5 results in absolute values. Furthermore, eq. 5 includes an offset of 0.05 which needs to be explained - is this specific for IGRA data and derived from some validation activities? Please clarify.

8. l. 164–165:
   'In processing, only the point that has matching data in each interval is selected for the comparison reliability.'

   This sentence is unclear, please clarify / reformulate.

9. Fig. 1:
   Why is the number N of collocations the same for all time intervals?

10. l. 179:
    Please justify the 0.25 cm limit for STD used here.

11. l. 190:
    'the closest PWV retrieval ... is selected'

    Do you mean spatially closest?

12. Section 4.1:
    Instead of Fig. 3, this section should include a global map of MERSI-II PWV data (e.g. daily or monthly, depending on coverage) on which the global water vapor distribution is discussed. Doing this solely based on the collocated data (Fig. 3) is not appropriate, because as shown in Fig. 2 the sampling of each site is different. The average at one station may contain e.g. different seasons for each station and is not representative for a global value. Inclusion of a full daily or monthly map would also show the coverage of the MERSI-II data.

13. l. 219–221:
    'there are some individual points with the unnormal MB and MRB. Therefore, the top 1% and bottom 1% of MB and MRB are not present in the histogram in order to show an intuitive acknowledge of distributions of MB and MRB.'

This is unclear. What do you mean with 'unnormal' and 'intuitive acknowledge'? Do you mean that there is an outlier filter applied here? Is this part of the MERSI-II product (if not, why do you need it)? Is this filter also applied for other analyses in the paper?

14. Fig. 4 and related text:
The histograms given in Fig. 4 are produced using a certain bin width, which should be specified. This bin width should also be considered when specifying values in the text (no. of digits). For example, you cannot derive an MRB value of -2.38% (2 digits accuracy) if the bins are e.g. 1%.

15. l. 228–230:
'this result is comparable to the accuracy of MODIS NIR PWV product, which is compared with MWR PWV and with a 5%-10% error range.'

    To which MERSI-II/MODIS values do you refer here (MRB, STD)? Note that the accuracy of the MODIS product is not the same as an error.

16. Fig. 5 and related text:
The percentage values given in the text cannot be inferred from Fig. 5, they should be taken from the histograms in Fig. 4. Actually, it seems that the given numbers are not in line with Fig. 4, e.g.:

    l. 238: 'About 80% of all sites have negative MB values'

    From Fig. 4b, it should be 60%. Please check/harmonize.

    In general, because the sampling at each station is different, it is unclear how representative Fig. 5 and the conclusions on spatial distribution of deviations etc. are. This needs to be discussed.

    Please also discuss some potential (physical) reasons for the spatial distribution of differences in addition to sampling. Are these due to possible limitations of the measurements (e.g. low signal, high noise) or local features (e.g. surface reflectance, aerosols, ...).

17. Section 4.2:
This part could possibly be shortened. Is, for example, Fig. 6 (seasonal results) needed in addition to Table 1 (monthly results)?

    As this section is about temporal variations it would be good to show at least one example for a full time series at a station in a plot.

    It should also be justified why the results are considered to be significant noting that e.g. differences in MB for different months in Table 1 are small.

18. l 271–272:
'with a high positive MB in winter'

    Actually, the MB given in Fig. 6d is negative and small.

19. Section 4.3:
As already mentioned in the General Comments, it should be justified why a detailed analysis of influence factors is required at all.

    The observed deviations are small, they have to be put in relation to the (currently not specified error) of the product. In view of natural variability and sampling issues, is the analysis accurate enough to separate (sub-)percentage effects?

    Qualitatively, the results shown in this section are in most cases as expected – the closer the correlative date in time and space the better the agreement. One exception is that MRB is smallest for a temporal distance of 1–2 h instead of 0–1 h, but there is no explanation given for this.

    Actually, an analysis of temporal effects was already shown in section 3.2 using AERONET data with the conclusion that 6 h differences are sufficient. So, why investigate this again?

20. l. 341–343:
'there is no height correction that can be used to eliminate false signals...'

    Which height is assumed in the algorithm? Is varying surface elevation considered? This is an example why more information about the algorithm is needed.

21. l. 347–349:
How are the 'hazy conditions' identified?

    'the influence of haze is hardly corrected completely in the MERSI-II PWV retrieval algorithm'

    Does this mean there is a correction or not? If yes, how does it work?

22. Table 2 and related text:
Noting the different sampling of the stations and their distribution with latitude it should be justified why the presented values are significant.

    How much do the results depend on the selected month?

    Regarding altitude dependence, what is the vertical sensitivity of the MERSI-II measurements?

23. Section 5:
As mentioned in the General Comments, it is not clear how transports effects can be identified solely from time series from two single stations, especially since these time series are almost identical. The values at the stations for July agree within the percentiles - why is this a significant difference? For the identification of transport one would at least need e.g. a time series at the source region (Bay of Bengal) which shows a peak during a certain month which is then observed with some delay at one of the stations. The selection of stations should also be more motivated - why exactly these? What is the role of surface elevation / surrounding terrain?

    In general, it is very difficult to follow the argumentation in this section.

24. l. 388:
'the trends of PWV at the two sites are similar'

    There are no trends shown in Fig. 9. Do you mean 'seasonal variation'?

25. Fig. 9d and related text:
Please confirm that the PWV data shown in the time series are from MERSI-II and not radiosondes.

Technical Corrections:

1. In most cases, too many digits are given for values in the text. Only significant digits (in relation to the uncertainty of the products and the analysis methods) should be given. Two digits should be usually sufficient, in case of percentage values even less.

2. Color bars of figures:
Figures 1–8 use a discrete color bar which contains quite similar colors at the lower and higher end (light pink, orange) which can be misleading. Furthermore, the steps between colors are in most cases not equidistant, which makes a visual interpretation difficult, especially for the map plots. I suggest to use for all figures a color bar with equidistant steps and colors e.g. similar to the ones in Fig. 9, or maybe even a conceptually uniform color map.

3. Fig. 1 (and actually all scatter plots):
'=EE:' although explained in the text, this notation (with a '=' in the variable name) is a bit misleading. I suggest to rename this.

4. Fig. 1:
   Remove symbols on top of panel (a).

5. Figs. 3 and 5 (and possibly Fig. 2):
   As these show global maps I suggest to remove the labels on the latitude and longitude grid.

6. Fig. 5:
   Please describe panels a–d in the caption and label in each figure the quantity (MB, ...) shown.

7. Figs. 7 and 8:
   Suggestion: These figures are not really needed - their results could be placed in a table instead.

8. l. 334:
   'observations in April are selected in the comparison rather than the annual mean value (MEAN)'

   As I understand, MEAN is not the annual mean value but the mean over all values in April.

9. Table 2:
   Remove/adapt line breaks in left column / top line.

10. l. 366–367:
    'As we all know, water vapor can significantly affect climate change, radiation balance and hydrological cycle.'

    This sentence is essentially identical to one in the introduction and can be deleted.

11. Fig. 9:
    Please include/mark the Brahmaputra River and the Brahmaputra Grand Canyon in the maps. This would help a lot to follow the argumentation.