

Review of " A Global Terrestrial Precipitable Water Vapor Dataset from 2012 to 2020 Based on Microwave Radiation Imager Measurements from Three Fengyun Satellites" by Xia et al.

The authors developed a global terrestrial precipitable water volume (PWV) dataset from 2012 to 2020 by applying a machine learning model using Microwave Radiation Imager (MWRI) observations on board the Fengyun satellites series. The accuracy of dataset is evaluated by comparing with the products of SuomiNet GPS and Integrated Global Radiosonde Archive Version 2 (IGRA2) PWV. This work contributes to representing spatial and temporal PWV variations and providing valuable resource for atmospheric research. The manuscript may be considered for publication after being major revised in accordance with the following comments:

General:

1. The introduction could benefit from a more comprehensive discussion of the significance of PWV dataset in the context. This could include a brief overview of existing challenges and gaps in PWV dataset construction, and how this dataset addresses them. Additionally, the literature review should be expanded to include more recent studies on PWV retrieval employing machine learning techniques. This may help establish the novelty and contribution of approach proposed in this study.
2. In the method section, the authors choose Light Gradient Boosting Machine (LightGBM), Extreme Gradient Boosting (XGBoost) and Random Forest to train the model. The reasons for selecting these models should be supplemented.
3. In the conclusion, it is essential to articulate not only the strengths of the dataset but also to elucidate its constraints and limitations.
4. Please check the grammar in the manuscript to improve the text quality. For example, the subject of the sentence that "With the development of computer science, and in particular the proliferation of machine learning (ML), has led to the widespread adoption of ML by the remote sensing community" is missing.

Specific:

1. Line 216-217, The full names of "WAT, WET, ENF, EBF, DNF, DBF, and MF" should be provide when they first appear in the manuscript.
2. Figure 5: It is clear that the amount of data when MWRI PWV is compared to IGRA2 PWV is much smaller than when it is compared to SuomiNet GPS PWV and enGPS PWV. This should be supplemented in the manuscript as well as

giving possible reasons for this discrepancy.

3. Figure 8: The different colors of the solid dots in the figure should be clearly explained. Please include a color bar to indicate what each color represents for better clarity.
4. The manuscript states that "the MWRI PWV exhibits a wet bias at low PWV values and a slight PWV underestimation at high PWV values." Could you provide possible explanations for these observed biases? Discussing potential reasons, such as instrument limitations, atmospheric conditions, or retrieval algorithm issues, would help clarify this point.