Review of “A Grid Model for Vertical Correction of Precipitable Water Vapor over the Chinese Mainland and Surrounding Areas Using Random Forest”

Summary

This paper develops a machine learning based model, RF-PWV, which predicts the PWV difference between two height levels based on their height difference and the time information. RF-PWV is trained based on a 10-year-long ERA5 dataset. This study shows when only given the bottom-level ERA5-based PWV data, RF-PWV can well capture the time-dependent vertical distribution of PWV in ERA5, outperforming the earlier model C-PWVC1. When verified against the radiosondes profiles, RF-PWV also shows a marginal improvement in terms of RMSE compared to C-PWVC1.

The authors have effectively summarized the results from RF-PWV, and presented a comprehensive comparison between RF-PWV and C-PWVC1. While this paper is in general easy to follow, there are a few concerns regarding the RF-PWV model and its applicability, insufficient details of C-PWVC1, unclear motivations for comparing RF-PWV and C-PWVC1, and the ambiguities in the text and notations. Therefore, I recommend a major revision at this stage. I believe resolving these concerns can enhance the impact of this paper.

Major comments

1. About the RF-PWV model:
   (1) Why is ‘year’ included as an input variable to RF-PWV? Are there justifications for its relevance to the vertical distribution of PWV? Considering the model is trained on a 10-year dataset, which is relatively short, how reliable and generalizable is this relationship, even if one exists?

   (2) How is the ‘day of the year’ information obtained in the training data if the monthly averaged hourly dataset is used? My understanding is the monthly averaged hourly dataset for one year is a 24 (hr)*12(month) dataset, which should be something like:

   average PWV at 00:00 in January 2020
   average PWV at 01:00 in January 2020
   …
   average PWV at 23:00 in January 2020
   average PWV at 00:00 in February 2020

…
average PWV at 00:00 in December 2020
…

Is my understanding correct, or what am I missing? Please clarify this in the manuscript.

(3) I suggest expand on the practical applications of RF-PWV. For example, one could also use a data-assimilated product (like ERA5 and other reanalysis products) to obtain a vertical distribution of RF-PWV that has comparable accuracy with the RF-PWV output. Is there a current practical demand for precise real-time information of the vertical distribution of PWV that existing products are unable to provide? Please elaborate if applicable.

(4) Since RF-PWV is trained on ERA5, its accuracy is 'likely' (not necessarily, since ERA5 is not the truth) not superior than ERA5 itself when verified against accurate observations (e.g., radiosonde). Have you compared the error of RF-PWV and ERA5? Have you considered training a machine learning model purely based on accurate observations?

2. What is C-PWVC1 model? A summary of how this model works is required. For example, what are the key properties that make RF-PWV better than C-PWVC1? How does C-PWVC1 use the time information in predicting the vertical distribution of PWV? Why do you choose to compare RF-PWV with C-PWVC1 (e.g., why not C-PWVC2)?

3. There are a few ambiguities within the notations and nomenclature that might confuse and mislead readers. These ambiguities should be addressed and clarified to improve understanding. Please refer to the minor comments.

Minor comments

L13: PWV differences -> This is ambiguous. It should be something like the differences between …

L56: I suggest introduce the full name of RF-PWV here

L83-85: In the text, it is stated that "The PWV for each pressure level is determined …" while the PWV in Equation (1) is the integral of PWV of the whole column. Please

clarify the notations, e.g., using PWV_i to denote the PWV for the i-th pressure level.

L125: 'test set' -> 'validation set'. I suggest only use test set to refer to the 2018 dataset.

L138-139: It would be helpful to use better notation here such as $\Delta PWV_i = PWV_i - PWV_0$, where 0 stands for the bottom level.

L140: "$22 \times 24 \times 12 \times 10$" Please be specific what each number refers to here. I am assuming 22 levels, 24 hours, 12 months, 10 years?

L143: "the reference PWV". The word 'reference' has also been used to refer to the 'true' dataset in the validation later in the paper (e.g., L164). It is somewhat confusing.

L140-145 & Figure 3: I suggest make it very clear of not only the input/output of the RF-PWV, but also the input/output variables of the ML part

L149: 'Subsequently, the target point'. This sentence needs to be rewritten. It's incomplete and unclear.

L160: What does 'C-PWVC1' stand for? A description of the C-PWVC1 model is missing (see major comment #2)

Figure 4&7: I suggest the colorbar for RMSE to only display positive values.

Figure 5: I suggest change the x-axis from "doy" to month

L238-248: I suggest elaborate in detail on how the validation is conducted in Section 3.2. It seems that the input to RF-PWV model is no longer the PWV difference relative to the bottom level (like what was done in Section 3.1). What are the input/output of RF-PWV, and what are the output verified against? Using a few simple equations with precise notations could be beneficial. Please clarify.

L292: What does it mean when the height difference is less than 0?

A general comment for Section 3: While the authors provide a thorough description of the plot/table results, I suggest trim some unnecessary details and focus more on explaining a few key findings and their implications, which can enhance the overall presentation.