

Interactive comment on “Gradient Boosting Machine Learning to Improve Satellite-Derived Column Water Vapor Measurement Error” by Allan C. Just et al.

Anonymous Referee #3

Received and published: 26 February 2020

General

The manuscript appears smartly prepared and well written. However, to my opinion the manuscript is not appropriate for publication in AMT. The manuscript deals with a machine learning concept to improve the MODIS/MAIAC column water vapour retrieval. Only machine learning aspects are discussed, and these aspects are just described. There is no chance for the reviewer to check the quality of the work. I have to believe what they write. This is rather unsatisfactory.

MAIAC needs a large bunch of surface, atmospheric, and technical input parameters to successfully retrieve water vapour information. These input parameters have partly

C1

large uncertainties. The surface and atmospheric input data change from day to day, with time (morning vs afternoon), with season, with land use changes. Nevertheless, the MAIAC methodology seems to be very robust, the accuracy of the MAIAC products is very good (without any machine learning effort)! To my opinion, it is impossible to further improve the MAIAC column water vapour values! However, the authors of the manuscript want to convince the reader that the machine learning concept overcomes this insurmountable wall of given and (unknown) uncertainties. It improves the results, and reduces the overall uncertainties although the given uncertainties are unknown! How is that possible? The paper gives not answer to this.

I recommend a rejection also because the paper does not contain any aspect of atmospheric science or technological development that would justify publication in AMT. Only machine learning aspects are given. Why not choosing again Remote Sensing where the first paper of Just et al. was published?

Some details:

The title is ‘strange’, not logical! What does it mean: ... to improve ... the error...? What does it mean: ... satellite-derived ... measurement??? The column water vapour is clearly a retrieval product. ... There is no ‘direct’ measurement.

Lines 85-90: In the introduction it is written: machine learning approaches such as XGBoost can model complex phenomena etc... The resulting prediction model can provide an algorithm to reduce the retrieval errors. I conclude: yes, the model can do that provided the complex input parameter set is free of uncertainties. But many aspects (input data) are not well known in the case of the MAIAC retrieval, uncertainties in the input data are large and that is the reason for the uncertainties in the product.

Section 2:

Line 118-120: Target modelling parameter is the difference between MAIAC and AERONET CWV... My question is: When the machine learning approach finds the

C2

best way for correction (e.g. based on all the 75 station of northeastern United States in Figure 1) can this approach then be applied to the rest of the world? I do not believe that this will work! Probably we have to find optimum ways for corrections again and again, region by region and all this for different seasons.

Section 3

Some examples that explain my general feeling with the paper: Lines 147-149: The XGBoost package is used! Ok! But the reference for this is a conference contribution, grey literature!

Lines 153-154: XGBoost is combined with DART (here the reference does not indicate any journal?). Can we believe, everything is ok with this procedure? Can we trust? Is all the material peer reviewed by machine learning experts?

Lines 159-161: Bayesian optimization for hyperparameter tuning of XGBoost models was performed using the autogxboost R package (Thomas et al, 2018). The reference points to arXiv. . . . This is a preprint archive (no peer review, nothing). So, what is this? Can we trust?

Lines 175-176. . . The contribution of each feature to cross-validated predictions was estimated by SHAP values (reference. . . arXiv). . . .again this preprint archive. . .

Lines 177-227: A lot of information and description is given by the authors, written in a smart appearing way, but it does not help. The reader is lost! He/she just has to believe that everything is ok with this way. But he/she does not trust.

Section 4: results:

I avoid to give my comments to the text. . . nobody can check what they state. . . , what is ok, what is not ok, what is trustworthy, what is not trustworthy. There is nothing to judge!

To the figures. . .

C3

Figure 1: There is no hint where we are? no city name, e.g., Boston, New York, no name of any state. . . Maryland. . . . Figure 1 is a nice 'indicator' , . . . of the feeling I have with the entire paper.

Figure 3 and the following figures tell me: MAIAC does a good job, seasonally dependent uncertainties are visible. This is ok, surface properties change and are not perfectly considered in the retrieval. One should accept that.

Machine learning procedures may purge the deviations in this specific 'learning region' of Northeast USA. But for any new region . . . ? We have to start again, I believe.

Conclusion section: just 9 lines: This is good, there is practically nothing to conclude.

Interactive comment on Atmos. Meas. Tech. Discuss., doi:10.5194/amt-2019-308, 2019.

C4