Atmospheric Measurement Techniques Discussions

Atmos. Meas. Tech. Discuss., referee comment RC2
https://doi.org/10.5194/amt-2021-262-RC2, 2021

Comment on amt-2021-262

Anonymous Referee #2

Referee comment on "Deep-learning-based post-process correction of the aerosol parameters in the high-resolution Sentinel-3 Level-2 Synergy product" by Antti Lipponen et al., Atmos. Meas. Tech. Discuss., https://doi.org/10.5194/amt-2021-262-RC2, 2021

This article applied a previous developed concept of using machine learning (ML) to bias-correct aerosol optical depth (AOD) and other aerosol data from conventional aerosol product. Original concept of ML post-processing of satellite data against ground truth is introduced in author's previous journal articles. This time a feed forward neural network is used on Sentinel-3 data to produce two aerosol products: machine learning generated aerosol data and bias-corrected Level-2 Synergy Product. The article claims that the post-process corrected the Sentinel-3 synergy product is a high resolution, better accuracy data products than the original aerosol product and the aerosol product generated from pure FFNN model. Within resent decade, machine learning has been rapidly applied to Earth Science field. One of doubtfulness of relying on ML is that the approach is not based on physics. The idea of machine learning post-process include both the state of art machine learning technique and traditional algorithm-based approach, which maintain the physics within the retrieval process. It is a conservative way of using ML and if successful, can be applied to many fields. However, the statement of the post-process corrected aerosol data has higher accuracy than full ML predicted aerosol data is not convincing, especially in terms of AOD. Figure 4, 5, and 6 all show comparisons between these two products. There is no significant improvement from post-process corrected product to full machine learning output. Although the error statistics against AERONET are slightly better in post-process corrected data, when investigate details in Figure 4 we can see that the overestimation of AOD especially at AOD < 0.2, is amplified in post-process corrected data than fully learned regressor model output. The smaller bias statistics in post-processed product is balanced by the overestimation in low AOD regime (AOD < 0.2) and underestimation in high AOD regime (AOD > 0.5). If we look at other evaluation plots, such as error histogram or error diagnostic plot. We may have much better look at the error distribution of two data sets. Similarly for AE comparisons, it is hard to say that the accuracy of AE prediction is improved between the two ML-involved products.

Other specific comments are:

Line 27, atmospheric spelled wrong.

Line 67 remove “accurate”

Line 107 In section ? missing a number.

Line 190 please specific list the time/spatial criteria for collocation.

Line 197-198 Can random split for each region result in data from a few sites dominate the results for one region?

Line 211-212 Regarding normalization method. If we use all data mean/std to do the z-score standardization, all the data is converted equally still within the same scale as they are originally. What is the point of normalization? For fill data, what average is used? and how much missing data is there?

Section 3.5 What is the accuracy for the two-folds testing results for training/testing/validation datasets?