Atmos. Meas. Tech. Discuss.,
doi:10.5194/amt-2020-376-RC1, 2020

Atmospheric Measurement Techniques Discussions

Interactive comment on "Volcanic SO_2 Effective Layer Height Retrieval for OMI Using a Machine Learning Approach" *by* Nikita M. Fedkin et al.

Anonymous Referee #1

Received and published: 13 November 2020

I focus my review on the portion of the paper dealing with deep neural networks (DNNs) and data processing. Despite some analysis choices made that I cannot understand, the ideal of using neural networks to retrieve SO2 plume heights is interesting. I am disappointed that the evaluation of the performance of the DNNs is not well presented. A series of sensitivity experiments could be conducted to improve the quality of this study.

Comments

Line 194-195: How large is the sample size large enough? I think it would be good to have the number of trainable parameters in the DNNs reported here.

Line 244-246: Have you tried to increase the number of neurons in each layer of the

DNNs or increase the number of hidden layers in the architecture? A sensitivity analysis during a "fine-tuning" process is very helpful for the optimization of DNNs.

Line 233-234: The tanh activation function is more frequently used in classification problems. For regressions, the rectified linear unit (ReLU) and parametric ReLU (PReLU) are more used.

Line 217: How do you control overfitting? Do you have a validation set? A validation set should be a small subset sampled randomly from the training set, over which the performance of the DNNs is checked after each epoch of the training. Sometimes the training loss gets reduced, but validation loss is not.

Figure 3: Relative errors are much more meaningful here because the "truth" of volcanic SO2 layer heights is not a constant set. It is also clear that the predictability of volcanic SO2 layer heights has a strong correlation with SO2 column information. Since OMI is biased, have you tried to add SO2/O3 column information from other sources as predictors?

Line 229-231: If you scaled the parameters to be within the max/min range, plus the tanh activation function you used, the gradients of DNNs with respect to trainable parameters in DNNs would not be sensitive to predictors closer to the max/min values. The degraded performance of your DNNs shown in Figure 3b could be a result of this. Maybe try batch normalization?

Specific points

- Regarding the stability of DNNs, you could also consider to add skip connections. This is not technically hard. It would smooth the surface of the loss function and reduce the number of local minima. (arXiv:1712.09913)
- It is difficult to evaluate the performance of your DNNs by comparisons shown in Figures 4, 5, 7 and 9. A heat map could be very helpful here, between the predicted volcanic SO2 layer heights and those retrieved from other satellites.

- Table 2 and 3 show statistics intercompared within the synthetic data set. I think it is meaningful to also have statistics compared with other independent satellite retrievals.

- If you do have access to other retrievals of volcanic SO2 layer heights, then I would suggest a multi-stage training. In stage 1, synthetic data sets are used. In state 2, you can keep training the model from stage 1, using a subset of real retrievals as the training set and the other as the testing set. If only synthetic data sets are used and the forward model generating these sets is not perfect, then the trained model must have a degraded performance compared to the forward model (due to errors from deep learning model itself). Moreover, during the forward model calculation, if the column data (O3 and SO2) are sampled within some range, then your DNNs would have a difficult time in predicting outliers and extrapolation should be very careful. Speed would be the only advantage then. The data-driven nature of DNNs should be taken advantage of.

- Bayesian neural networks are ideal for such prediction problem, when there are uncertainties associated with the outputs. This could be future work due to the technical complexity. But you could still alter the random seed to generate an ensemble of DNNs, such that you can provide a rough estimate of uncertainties on the predictions.

Minor comments

Line 61: replace VCD by vertical column density (VCD)

Fig 4: can you change the colorbar of (c) to the same scale as (a) and (b)?

Interactive comment on Atmos. Meas. Tech. Discuss., doi:10.5194/amt-2020-376, 2020.