Atmospheric Measurement Techniques

Discussions

Open Access

AMTD

Interactive comment

Interactive comment on "Machine learning as an inversion algorithm for aerosol profile and property retrieval from multi-axis differential absorption spectroscopy measurements: a feasibility study" *by* Yun Dong et al.

Anonymous Referee #1

Received and published: 26 November 2019

The manuscript "Machine learning as an inversion algorithm for aerosol profile and property retrieval from multi-axis differential absorption spectroscopy measurements: a feasibility study" presents a machine learning (ML) approach for aerosol retrieval from multi axis differential optical absorption spectroscopy (MAXDOAS). This is a new and innovative approach for aerosol retrieval from MAXDOAS measurements an therefore deserves certainly publication.

The paper gives a brief introduction to inverse modelling in general (sufficient or maybe even too much), to the MAXDOAS technique (sufficient), aerosol characterization (too

much, this is in the introduction and in the description of training data creation) and describes to some (sufficient) extend the forward model (VLIDORT 2.7) used to create the training dataset. However, it only describes very briefly the ML technique in general (not sufficient) and includes hardly any details about the actual neural network (architecture+hyper parameters) that was trained, nor does it include details about the training. The authors claim that the manuscript describes the new ML approach, but in fact, more space is used for the description of the training data set creation and about aerosol phase functions than on the ML itself. This does not reflect the title.

I certainly recommend this manuscript for publishing because it contains an innovative and very promising approach. I also think that the general approach to the work is good and that the results are of great interest and importance. Since there is no "moderate revision", I select "mayor revision", however, it is really just a bit more than "minor revision".

I strongly recommend two mayor changes:

(I) The weight of the manuscript needs to be on the ML approach, this is currently not the case. A general short description of neural networks is needed. Some details about the software package used, explanation of the choice of network architecture, why CNN why LSTM, the detailed settings etc (see SPECIFIC COMMENTS for suggestions and questions about the settings)...

(II) While the first point is likely just a matter of adding some more content to the manuscript, the second point weights a bit heavier and might include some more work: Normally, for ML, the data is split into three sets: (1) a training dataset (2) an evaluation dataset used during training to identify when the training results in overfitting (3) a completely new set of data for testing. I do not see that this has happened in this study, this is a huge weakness of the approach. The authors seem to have only used a validation data set (25% of the total data set) for the testing but no proper testing with parameters outside the training range (so not only "not this specific combination") was

C2

performed.

GENERAL COMMENTS:

(1) Introduction

There is a lengthy (and maybe not super accurate, see below) description on the aerosol phase function and the asymmetry parameter, both in the introduction and in Sect.4. However, there is no information on ML in the intro. This does not at all reflect the title. After all, this manuscript claims to be about the ML as inversion algorithm. Suggestion: Bundle the aerosol information from here and from Sect. 4 in the section about training/validation/test data creation (which is currently Sect. 4) and include some paragraph or two on ML use in inverse modelling and general ML.

(2) Ordering of content

I suggest a different ordering: (1) general introduction including advances in ML, aerosol importance in general, current retrieval techniques and why ML should be applied to aerosol retrieval (2) MAXODAS method description (3) Aerosol properties and modelling and forward modelling with VLIDORT (4) Overview of the methodology of the 3 necessary steps (instead of selling it as two steps as done in Sect. 3, where the first of the two has itself two steps) and a detailed description of the specific ML setup and choice of hyperparameters. (5...) as before

(3) Sect.2:

While what is written about OEM and parametrized methods is true, most of it is true for ML as well (i.e regarding e.g. the T/P profile). This section paints an overly dark image of OEM and parametrized codes. I think that the main problem with "traditional" methods is indeed the time they take, and this should be clearly (even more clearly) stated, since this is the one huge huge advantage of using ML. Also, especially around line 136, it gives the impression of full profile retrieval of asy and ssa, while in fact, it is "only" the aod profile and single scalar values asy and ssa valid for all layers.

(4) content of Sect. 3 + 5:

Judging from the title, this should be the heart of the manuscript. It is not. Also, why this strange split of the topics? Does it make sense? As mentioned under (2), I would keep a very rough overview of the three steps in the introduciton and put all the remaining info on ML in Sect. 4. Some comments what is missing:

(a) which backend was used? Tensorflow, Theano? Some other? Why the mentioning of the yupiler notebook? Why was it used at all? Certainly no web-based interactivity is needed? Why wasn't it simply put in a plain python script?

(b) CNN is normally used in ML for image recognition, why is it used here? Is the CNN here to correctly identify the fixed parameters RAA and SZA (which have in fact a 1:1 mapping between input and output)? Why was it decided to split for profile and ssa/asy retrieval?

(c) Why is LSTM used? Maybe some intro on recurrent neural networks in general is needed. This seems to indicate, also from the bit of explanation that comes with it, that scans are not considered separately, but as a function of time.... (so a scan from now and then from 10 minutes, not one from here and now and the next one from tomorrow and somewhere else). However, this seems not to fit your introduction and abstract where you very specifically write about a single scan. This is very confusing and needs explanation. Also maybe, you can start with explaining what a SimpleRNN layer is and why this was not chosen?

(d) What were the choices of the hyperparameters? Which batch size was used? Which lr was used for the RMSprop? Where there any drop out layers? Which activation function was used? There is no information on any of the parameters. How many nodes do the layers have?

(5) what happens if the network gets data that is by no means covered by the training data (i.e. completely outside the range in one or more parameters?) What is the

effect of measurement noise (also including "noise" from situations that are not 1 dimensional)?

SPECIFIC COMMENTS:

(1) page 1, line 23 "... and have relative short lifetimes..." –> relative to what? Also, few minutes to few weeks spans about 5 order of magnitude in time, while one end of this span can be considered as short, the other cannot really. Please specify "relative".

(2) page 1, line 26: apart from all the properties already listed, what else do you mean with "physical properties" as opposed to optical? This is very unclear

(3) page 1, line 28 "The spatial and temporal distribution of aerosols ... is greatly affected by ... the type of aerosols". I think this is incorrect, the correct verb here is "depends on".

(4) page 2, line 39–40: If you put this statement, then you need to explain more. I also cannot see any connection of this statement to the rest of the paper. The minimum that should be added is how it depends on the surface albedo.

(5) page 2, line 41–42: "escpecially of anthropogenic origin" "of"? or "for"? This sentence does not make too much sense like it is, reformulation needed.

(6) page 2, eq2 and eq3: I would think that the range of the asymmetry parameter as such depends on the normalization of the phase function, so you need to have integral(phasefunction) over 3D angle = 1. If so, then the first moment <cos theta> is the asymmetry btw. forward and backward scattering. So with this, would you not have a factor of 1/4pi missing in the HG phase function? Maybe you could check the normalization factors for consistence btw. g and P.

(7) page 2, eq. 4: You seem to use tensor notation to make a difference btw. covariant and contravariant tensors and apply Einstein summation convention. However, you still put the summation sign, but without indicating what you are actually summing over.

C5

(8) page 4, line 101: "approximately known"? Please clarify.

(9) page 5, around line 136: Since it was highlighted before that

(10) page 5, line 153..154: both input and output states run to N, one of them should have a different limit, maybe... M? Otherwise it is confusing, especially because it is written that x has 67 layers, but y has "only" 16 angles.

(11) page 7, line 196: Although VLIDORT has as direct input the viewing zenith angle, most people in the MAXDOAS community are more familiar with the elevation angle. Maybe it is an idea to change this to make it easier to connect to.

(12) page 8, line 199, 201, and other listings in the text of parameters that are summarized in the Table 1: I do not think that they need to be repeated, I think it is enough if they are in the table.

(13) page 8, table 1: Can you comment on how the direct sun cases for raa=0, sza=vza are handled?

(14) page 9, line 223: why do you need ozone absorption?

(15) page 9, line 230: maybe a small sketch to explain the aerosol profile parameters (with the two components of the profile) would be helpful

(16) page 9, line 37: The 25% were fixed between the 20 realizations, or not? Also, did you monitor the evaluation loss? It would be really good to see some plots here of the evaluation loss as a function of epoch. Also, please comment on how over-fitting was mitigated.

(17) page 9, line 236: this height is the middle height or the height of the upper boundary? This is not clear.

(18) page 9, line 247ff: I would certainly describe the architecture of the network here, not only the Fig. 3. Also, dense layers are not explained. Also, how many nodes in the layers? Do you use maxpooling layers btw. your conv1d layers? What is the size

of your convolution window? And again, how was the architecure chosen? Why does it make sense to separate the SSA and the ASY the way you do? Do you extract the SZA and RAA as well? They should certainly be == the input? Is there a test on this?

(19) page 9, line 259: While you do use 25% for test (or do you actually use this for evaluation? Not really, because you use it to test the network. What was used for the evaluation then?), because you use the same type of parametrization, this is not a good test. A different, unseen set of data should be used. How do features that were not included in the training dataset at all (by all means outside the parameter range) affect the result? What about thin cloud layers above 4 km, do they affect the result? The tests included here are not very useful.

(20) page 10, line 275 & 285: given the range of parameters, using eq. 270, the maximum error is about 20%, not 100%. This puts these low numbers in context.

(21) page 11, Fig. 5: When you wrote earlier that the mean error is "-0.14", you really just took the mean over all angles? What is the significance of this? If it were in have the parameter space +50% and in the other -50%, its mean is still 0... and the model really not good, so what is the significance of the mean error here?

(22) page 12, Fig. 7: please explain the box whisker plot. Is the line mean or median? The box is how much percentage? The whisker? There are different conventions...

(23) page 16, line 351: This paper does not present the ML based algorithm. It presents some of the results and that's it. There is not enough information on the ML model. This sentence is not summarizing the paper.

(24) page 16, line 363: maybe this is because of the choice of training paramters as a linear distributed AOD?

TECHNICAL CORRECTIONS AND SUGGESTIONS

Since I expect some mayor change in the text, only a few comments here:

(1) Many times, there are definite articles missing (e.g. page 3 line 65 "The MAX-DOAS..", page 3 line 84: "The DOAS technique", page 4 line 91 "The offset term...")

(2) Eq. 5 on page 4 is not referred to in the text.

(3) page 7, line 204: I highly doubt that Clemer et al 2010 is the only code here. I would add a "e.g.".

(4) page 9, line 229: I think you miss the AOD=0 case in this list

(5) page 10ff: I suggest to use an equi-distant grid for the raa-sza plots, as they are now, it gives a biased impression to the eye.

(6) page 12, line 311 f: I cannot quite understand the sentence "This error contribution..." maybe you can reformulate

Interactive comment on Atmos. Meas. Tech. Discuss., doi:10.5194/amt-2019-368, 2019.