Atmos. Meas. Tech. Discuss.,
doi:10.5194/amt-2019-368-RC2, 2019

Atmospheric Measurement Techniques Discussions

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 #2

Received and published: 3 December 2019

Dong et al. introduce a new MAX-DOAS aerosol inversion algorithm which is based on machine learning. The algorithm itself utilizes neural networks which were trained with a pre-calculated synthetic dataset of O_4 AMF with various profiles, geometries and aerosol properties. A fraction of this dataset is used for the validation of the algorithm by analyzing results based on the difference between true and retrieved quantity.
This novel approach seems to be promising, particularly due to the advantage of being able to retrieve single scattering albedo (SSA) and asymmetry factor along with profiling information in near real-time.

General comments

The manuscript is well written but it lacks necessary information and analyses. I suggest to change the manuscript according to the following points:

1. Not enough information about the machine learning (ML) algorithm itself are given. The introduction focuses on aerosols and MAX-DOAS only without introducing ML. In section 5, there is no explanation why the individual ML steps were chosen as they are.
2. The validation section appears to be insufficient to assess the performance of the algorithm:
   (a) Why not changing the testing dataset to realistic profiles which are not included in the training data? How can you be sure that you do not over-fit your results?
   (b) Why not using larger aerosol loads?
   (c) Why did you use 16 different elevation angles for the testing dataset even though this number is much too high for most measurement locations? What happens if you just use 8 or 10 elevation angles? Does the algorithm still performs well?
   (d) The training dataset was created by using an US standard atmosphere. This is mostly a poor representation of the true atmosphere. What happens if the conditions change?
   (e) Why not testing the algorithm on real data?
3. Are there plans to extend the training dataset by more wavelengths, SSA/asymmetry factors, albedo, profiles, trace gases (as suggested in Sec. 7)?

Printer-friendly version

Discussion paper

4. Note that many articles are missing in the manuscript.

Specific comments

P2, L36-42: There is no need to show the equation of SSA and its detailed description. I suggest to change those 6 lines into one sentence only.
P2, L43-62: This part about the aerosol phase functions is much too long, especially since there is no further discussion about this topic in your paper. The lines about the Legendre expansion could be completely removed without loosing important information for the understanding of the manuscript.
P2, L58: When kept, please change the index L of $P_L(\cos\theta)$
P3, L65: "The" MAX-DOAS...
Figure 1: A single sky scan...
P3, L84: "The" DOAS technique...
P4, L91: "The" offset term...
P4, L101: "Forward model parameters that are considered approximately". I guess you are referring, among other parameters, to a priori knowledge when using "approximately" here? Please change the wording or reformulate this sentence.
P4, L114-115: "A priori information about...". I find this sentence to be rather confusing. What do you want to say here?
P5, L123-125: "None of the algorithms perform perfectly". That depends on what you understand as "perfect". I don't think that it is possible at all to retrieve the true atmosphere in a extremely high vertical resolution. However, as far as I know, the second part of this sentence is correct. I would suggest to reformulate the first part.
P5, L123-130: These 7 lines describe problems that also apply to your algorithm and only highlight on problems of existing algorithms. As long as you don't show perfectly validated profiles from real measurements, the reader does not believe that your algorithm "performs perfectly".

Note that you also used external information about the atmosphere for creating your training dataset. You directly applied a priori knowledge by using exponentially decreasing profiles including Gaussian's for your dataset. And I would not say that a priori knowledge does not exist. You look out of the window and know that it is a hazy day so you adapt the a priori. Sometimes you know about local sources or have ancillary measurements available. I would strongly suggest to change this paragraph.

P7, L176: AMF represents → An AMF represents

P7, L178: observations → observation

P7, L181: Where "the" vertical column density...

P7, L183-191: This part could be shortened as you have already introduced aerosols before.

P7, L205: "The" VLIDORT code...

Table 1: Why do your Gaussian profiles don't have center heights higher than 2km? Since the vertical sensitivity for higher altitudes is an issue for common algorithms, I am wondering if your algorithm performs better here? What is the scaling height of your exponential functions?

P9, L235: What was the reason for changing the grid step width to a coarser resolution for higher altitudes? I have the feeling that your choice of Gaussian profile center heights and retrieval grid steps might deteriorate retrieval results for higher altitudes (as indicated in Fig. 9 and 11).

Section 5: I think it would be nice to add more information to this section to explain also the in-between steps and parameters of your CNN and LSTM.

P9, L251: "RMSprop was chosen...". Please explain.

Figure 4: It would be interesting for the reader to see a similar plot describing the profile shape distribution. You could show 3 more plots for different partitioning, showing Gaussian center heights and width on x and y axis, respectively. Furthermore, the number of profiles with a certain total AOD would also be interesting (especially when looking at Figure 7). I fear that the reader might loose the connection to the actual profile shape due to the rather statistical analysis in the following paragraphs.

Figure 5, 6 and 8: It is interesting to see that mean error and standard deviation show areas with high or low values at certain geometries. I was wondering if this is a matter of the scattering angle (angle between incident and outgoing photon assuming single scattering)? Could you please create a plot showing the scattering angle versus the respective error/standard deviation? Since e.g. RAA = 30° and a low SZA is equivalent to a large scattering angle (e.g. Fig 5c) it might show issues for certain scattering geometries. In addition, you could check if there are certain profile shapes or aerosol parameters more frequent for areas with a large standard deviation or high mean errors compared to other geometries. I was also wondering about the outliers in all three histograms. Any reason for that?

P11, L292: I agree that OEM methods also struggle with data inversion measured at small RAA but I was wondering why your synthetic analysis fails?

P12, L295: "The total AOD retrieval..." or "The retrieval of total..."

P12, L297: In general, "the" ML algorithm

P12, L299: What is the reason for the second peak in the histogram?

Figure 7: Please explain all depicted quantities (mean, median, percentiles...) in the caption of this figure. Here, it would also be interesting to see if the largest underestimations correspond to certain profiles or parameters.

Figure 9: Why is the error larger for 1.5km than for 2km? Since you also included Gaussian's with Peak heights around 2km, I would expected the largest error at higher altitudes. Especially when considering the higher sensitivity of MAX-DOAS measurements for aerosol loads closer to the surface (which can be seen for altitudes lower than 1.5km).

Figure 10: It appears that there is also an underestimation of the predicted AOD for all sub-figures with true partial AOD's larger than 0.2. For example in the upper left sub-figure, but also in the second row (first figure), the third row (2nd and 3rd fig). Do these underestimations correspond to problematic scenarios/geometries/parameters?

P16, L353: Training and evaluation of "the" ML

P16-17, L365-372: Points 1 and 2 are valid but only a demand for near real-time applications. I doubt that there is a need for science to have profiles immediately after the measurement. For point 3, I was wondering if this is a major advantage. The dependence of profiling results on SSA is rather small and the Henyey-Greenstein approximation is in most cases a poor representation of the scattering distribution of aerosols. So why should the reader decide for your algorithm when an AERONET station nearby measures "real" phase functions and SSA? In point 4 you even diminish the potential of your approach by saying that it might be used as an initial guess for other algorithms. To me, this does not sound as if the authors are convinced of the capability of ML algorithms. If this is true, why not? If not, why are there no strong arguments in favor of your approach?

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