Atmos. Meas. Tech. Discuss.,
doi:10.5194/amt-2019-143-AC2, 2019

Atmospheric Measurement Techniques Discussions

Interactive comment on "A neural network radiative transfer model approach applied to TROPOMI's aerosol height algorithm" *by* S. Nanda et al.

S. Nanda et al.

nanda@knmi.nl

Received and published: 11 July 2019

Reviewer comment (general):

Nanda et al. present a method to accelerate radiative transfer calculations based on neural networks (NN), to speed up an optimal-estimation based retrieval of aerosol layer height (ALH) from TROPOMI O2-A band observations. The neural network is trained/validated on a set of simulated TROPOMI spectra. The ALH retrieval using the NN-based forward model is compared to results using an explicit line-by-line forward model. The NN version of the retrieval produces results consistent with the explicit model for a synthetic and real test case, whilst improving the speed by three orders of

Printer-friendly version

Discussion paper

magnitude.

Overall, I think this paper is interesting and well within the scope of AMT. In Its present form, it is missing some key details (see comments). Once these are addressed I will recommend publishing.

Author's response: Thank you for the constructive criticism and for taking the time to review this manuscript. The response to reviewer comments are addressed in the following.

Reviewer comment (specific 1): Page 2 Line 18: "The bottleneck identified. . ."

There are a few other commonly used methods for accelerating RT simulations e.g. optical property PCA and low-streams interpolation (see cited review below). It may be worth mentioning why NN is being chosen over these.

Natraj V. (2013) A review of fast radiative transfer techniques. In: Kokhanovsky A. (eds) Light Scattering Reviews 8. Springer Praxis Books. Springer, Berlin, Heidelberg

Author's response: We have not tried methods such as PCA or low-streams interpolation to check whether or not artificial neural networks are better at approximating the full physics radiative transfer model.

This is a valid suggestion and related to the remarks by Reviewer 1 on different methods (**amt-2019-143-RC1** Reviewer specific comments 2 and 3) where, in our response and our manuscript amendments, we explain the differences between their approaches and ours. Our main goal was to use a method that can very quickly estimate the reflectance as well as the Jacobian, which is used in the optimal estimation retrieval method. NN turned out to be very efficient for this. It is not known whether other methods, PCA or low-stream interpolations, are capable of doing this. As far as we know, this has never been attempted. We have also looked into k-distribution method, but this turned out to be extremely complex when the Jacobians also had to be included

and therefore abandoned.

Reviewer comment (specific 2): Page 4 Line 30: From preliminary tests, the exclusion of RRS. . ..

Since there were preliminary tests for the sensitivity to RRS, it may be worth mentioning what these were and quantitatively how these impacted zaer. Solar-induced chlorophyll fluorescence in the A-Band may also have a similar effect spectrally, and likely has a greater impact on the spectra. See

Frankenberg et al. (2011) Disentangling chlorophyll fluorescence from atmospheric scattering effects in O2 A-Band spectra of reflected sunlight, Geophysical Research Letters, vol 38, L03801 https://doi.org/10.1029/2010GL045896

Author's response: The issue on quantifying the sensitivity to RRS has been addressed in the Author's response to Reviewer comment (specific 7) in **amt-2019-143-RC1**.

With regards to the influence of solar-induced chlorophyll fluorescence in the A-band, this is a feature that will be considered in the future release of the TROPOMI aerosol layer height algorithm, as the current focus is to incorporate a neural network model as a replacement for DISAMAR (which is the main line-by-line model). Due to this, the neural network models mimic the operational DISAMAR model (which currently excludes fluorescence) used for retrieving aerosol layer height from TROPOMI measurements as much as possible.

Changes to the manuscript:

- With regards to the reviewer's comment on including preliminary tests for the sensitivity of the retrieval to the exclusion of RRS, Section 2.2, paragraph 3 of the manuscript is amended as follows.

'While this exclusion of RRS is not advised by literature (Sioris and Evans, 2000; Vasilkov et al. 2013), preliminary experiments by Sanders and de Haan (2016) have ascertained that the errors in the retrieved aerosol layer height resulting from ignoring RRS of the oxygen A-band in the forward model are significantly smaller than the effect of other model errors. Due to this, the KNMI aerosol layer height retrieval algorithm has historically ignored calculating RRS cross sections.'

- With regards to the reviewer's comment on the impact of chlorophyll, the following amendment has been added to the final paragraph of section 2.2.

  'The surface is assumed to be an isotropic reflector with a brightness described by its Lambertian Equivalent Reflectivity (LER). This is also an important simplification, requiring less computations over other surface models such as a Bi-directional Reflectance Model. Although the forward model is capable of including sun-induced chlorophyll fluorescence into the retrieval, it is currently being considered for a future implementation of TROPOMI's operational ALH retrieval algorithm. Lastly, the atmosphere is spherically corrected for incoming solar radiation and remains plane-parallel for outgoing Earth radiance.'

Reviewer comment (specific 3): Page 4 Line 31: The aerosol fraction is assumed as 1.0

Could you define what you mean by aerosol fraction?

Author's response: Agreed.

Changes to the manuscript: With respect to the reviewer's comment, the line in contention has been amended to:

'The fraction of the pixel containing aerosols is assumed to be 100%, which further simplifies the representation of aerosols within the atmosphere.'

Reviewer comment (specific 4): Page 5, Line 1-5: Perhaps the largest simplification...

There are many assumptions here - The aerosol optical properties are fixed (0.95 SSA, g=0.9 using a simplified HG phase function). Either literature justifying why these assumptions are ok must be cited, or the authors need to test these how these assumptions impact the retrieval results e.g. by testing against synthetic data with realistic optical properties. I would be curious to see how the retrievals perform for cases with very different optical properties e.g. dust

Author's response: These fixed aerosol optical properties have been derived from AERONET data, and the consequences of fixing them are discussed by Sanders et al. (2015), who use GOME-2 and SCIAMACHY spectra to show that these model assumptions aren't the main source of error.

Changes to the manuscript: The following sentence and citation has been added to the end of line 5 in page 5.

'These fixed aerosol optical properties have been derived from AERONET data and the consequences of fixing them are discussed by Sanders et al. (2015), who used GOME-2 spectra to show that the algorithm is robust against these model assumptions.'

Reviewer comment (specific 5): Page 7, Line Line 19: The inputs for the NN are referred together as the feature vector...

The fixture of the aerosol optical properties in the optimal estimation approach seems quite restrictive. Have you performed an information content analysis of the TROPOMI O2-A band to check if retrieving some of these can reduce the uncertainty in aerosol height? E.g. allowing the SSA to vary may reduce the potential influence on its parameter error inducing a corresponding ALH error.

Reviewer comment (specific 6): Page 7, Line 28: "whereas NN only uses the tem-

perature at zaer"

For the meteorological parameters, is there any rationale for excluding other potentially important predictors from being included e.g. PBL height or surface heating fluxes and wind speeds that may also provide prior information about the ALH.

Author's response: Yes, these analyses have been conducted as a part of TROPOMI's as well as Sentinel-4 and Sentinel-5 ALH algorithm development activities, and published their respective ATBDs (some of which are public and the others not) and in Nanda et al. (2018a). In theory, varying SSA makes sense, especially over land where its uncertainty is the largest source of error in the retrieved ALH (Nanda et al., 2018a). However, in practice it leads to reduced convergence rates. Because it is not well understood, and also because the goal of the paper is not in fact to introduce new variables into the state vector and instead to try and replicate the existing ALH retrieval setup as much as possible, a discussion on incorporating SSA into the feature vector has been excluded from this paper.

The meteorological parameters mentioned by the reviewer are not incorporated into the current ALH algorithm's forward model, simply because their exclusion is not as important a source of error as the biases in the retrieved product from model errors in the surface reflectance as well as the interaction of photons reflected from the aerosols and the surface.

Changes to the manuscript: The following sentence is amended to the end of the paragraph.

'In general there is a greater scope to add detailed information in Disamar. However, Disamar has historically incorporated many simplifications in order to reduce computational time. The current NN model is developed with the aim to mimic Disamar as much as possible, without including additional state vector elements into the retrieval, such as chlorophyll fluorescence, aerosol optical properties, cloud properties, and so on.'

Reviewer comment (specific 7): Page 8 Section 3.2 First Paragraph From my reading of this, the profiles are generated randomly after selecting a random set of tropomi solar-viewing geometry combinations. Naively, I would expect that the most representative way to create the training set would be to select the profiles from the ERA reanalysis corresponding to the randomly selected orbit geometries - the model probably doesn't need to reproduce a spectrum of Saharan dust for a typical Antarctic viewing geometry. Perhaps there is a heuristic argument that the way you are doing it could more reliably span the entire set of profile/viewing combinations, but if this is so you should state it in the manuscript. Ideally you would want to test the performance by comparing multiple approaches of generating the training data, though I am not sure of your computational resources

Author's response: Generating training data can take weeks to months with the available resources at the KNMI. The authors of the paper agree with your statement, and will add the following to the manuscript.

Changes to the manuscript: Added the following to the end of the first paragraph of section 3.2.

'This training data generation strategy spans the entire set of TROPOMI solar and viewing angles as well as meteorological parameters.'

Reviewer comment (specific 8): Page 9, Line 1: Finding the most optimal neural network configuration requires...

Using a single set for validation leads to questions about robustness in the validation. Typically cross validation e.g. k-fold methods are used to derive more accurate estimates of model performance. Have you looked into how robust the single training set is?

Author's response: With regards to the reviewer's comment on whether the paper

utilises neural networks trained and validated with a 'k-fold' cross-validation approach, the answer is yes. The training dataset was first shuffled and then split into a basic train-test split, which is equivalent to a 2-fold cross-validation approach. The manuscript does not mention this, and will do so in the amendment discussed in the following.

Changes to the manuscript: The following is added to the first paragraph in Page 9 of the manuscript.

'Finding the most optimal neural network configuration requires testing the trained neural network model. To that extent, the training data set was split into a training-testing split, where the model was trained on a majority of the training data set and tested on the remaining minority. Once trained, the model was tested again on a test data set with 100,000 scenes outside of the training data set.'

Reviewer comment (specific 9): Page 9, Line 6: The sigmoid function is chosen for activation... Can you show evidence for the sigmoid function outperforming the other forms?

Reviewer comment (specific 10): Page 9, Line 8: For each of the neural network models. . .

It would be useful showing a comparison of performance for the different models tested - I don't have an intuitive sense from the description of how different the actual results were for the different layer/node combinations, or how robust one instance of a 25000 iteration training is. For instance if I retrained the two layer model with 100+100 nodes again with a different initialization, would it still be the optimum configuration?

Author's response: When it comes to deciding which neural network model configuration is the most optimal depends on many factors, one of which is simply the neural network model architecture. Sometimes, adding an extra layer can result in an improvement of the neural network model, i.e. the mean squared error, summed for all

wavelengths into a single number, between the predicted and the true output spectra is low. But if this improvement is a very small number, a simpler architecture is a better alternative as it takes less time to train and compute outputs.

This is the basis of deciding whether or not a certain configuration outperforms the other. The amendment to the manuscript will include a set of plots (three plots into a single figure) and a paragraph that explain the rationale of choosing the sigmoid function, the two-layered and 100 nodes per layer choices.

If the most optimal neural network model was retrained with a different, randomly chosen initialisation of weights, the neural network model would still remain as the most optimum configuration.

Finally, there is a typographical error in this paragraph. The number of iterations for testing the configurations are 250,000 and not 25,000 (which is too few training iterations to make any concrete statement for this specific case).

Changes to the manuscript: A new figure, as per the recommendation of the reviewer is added. The figure numbers of the subsequent figures are automatically adjusted. The following caption is used for the figure:

'Summed loss as a function of training step for different neural network model configurations. **(a)** The neural network models have 50 nodes per each layer with a sigmoid activation function. **(b)** The neural network models have two hidden layers with each node activated by the sigmoid function. **(c)** The neural network models have two hidden layers with a 100 nodes for each layer.'

The last line of the 3rd paragraph is removed and the 4th paragraph in section 3.2 is amended as follows

'In order to test the most optimal number of layers, the most optimal number of nodes per each layer and the activation function, several neural network configurations were trained for 250,000 iterations and their summed losses (defined as $\Delta \times n_\lambda$) were com-

pared to find out which was the best configuration. To begin, with 50 nodes per each hidden layer, three neural networks for each of the three models were trained — one-layered, two-layered and three-layered. The neural network models performed best with at least two hidden layers (Figure 2a). For all three models, their two-layered versions show a similar summed loss to their three-layered alternatives, with the summed loss for the two-layered $NN_{disamar}(K_\tau)$ showing more stability with training epoch. Because of this, a simpler two-layered architecture is chosen for all three models. Continuing on, three other architectures for each of the three models were chosen with 50, 100, and 200 nodes for each of the two hidden layers. The results that with more training steps, the choice of 100 nodes for each of the two layers has a compromise between summed training loss and simplicity (Figure 2b), especially for $NN_{disamar}(K_\tau)$. Finally, going ahead with a two-layered and 100 nodes for each layer configuration, three activation functions namely the sigmoid function, the hyperbolic tangent function (tanh) and the rectified linear unit (relu) function were tested for each of the neural network models (Figure 2c). In this case, while all functions converge to similar summed loss values by 250,000 iterations, the sigmoid function has a good compromise between training loss and stability. Figure 3 gives a graphic representation of the neural network model.'

Reviewer comment (specific 11): Page 11, Line 12: Although the retrieval algorithms have good agreements...

For the low aerosol loading scenes, what happens when you include both surface albedo values in the NN model?

Author's response: This is not known, as the trained neural network models do not include both surface albedo values, and they are also not stored in the training data set.

Reviewer comment (Minor Corrections): Equation 2: Bold the x in the forward model to be consistent with notation for vectors.

Page 7, Line 17: Change “automatic differentiation which is a powerful algorithm that computes” to “automatic differentiation which computes”

Author’s response: Agreed.

Changes to the manuscript: Amended the manuscript as per the reviewer’s comment.

Interactive comment on Atmos. Meas. Tech. Discuss., doi:10.5194/amt-2019-143, 2019.

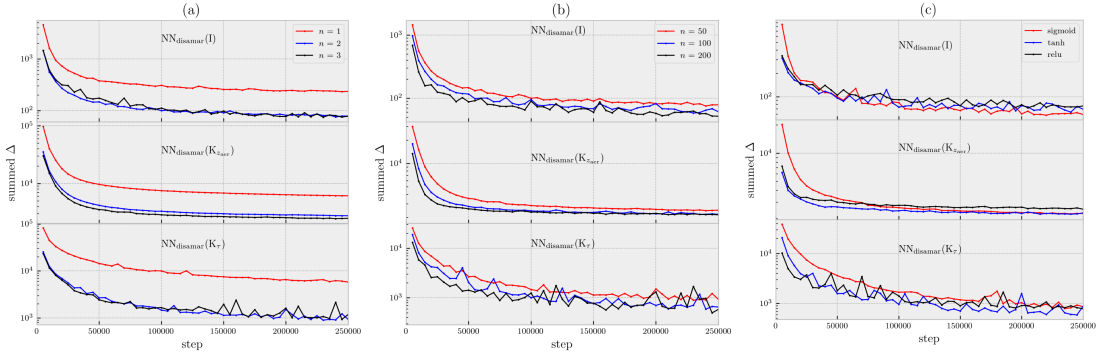

Fig. 1.

Printer-friendly version

Discussion paper