Response to Authors for the manuscript "*A neural network radiative transfer model approach applied to TROPOMI's aerosol height algorithm*" (amt-2019-143).

1 Response to amt-2019-143-RC1

Reviewer comment (general): This work is novel and interesting. The proposed algorithm produces results that have comparable accuracy to line-by-line models while achieving three orders of magnitude speed-up in computational efficiency. This makes it possible to increase the throughput of TROPOMI aerosol retrievals and possibly enable operational retrievals for all pixels in each TROPOMI orbit. The computational speed-up also opens up the possibility of including more physics in the forward model. The usage of three neural network models is an interesting idea. It takes advantage of the fact that correlations between input parameters and different forward model outputs are different. The paper should be published, but only after the comments (see below) are dealt with.

Author's response: Thank you for taking the time to review this manuscript. The goal of this research project was to develop a faster aerosol layer height retrieval algorithm, while keeping in mind the possibility of including more details into the model in the future.

Reviewer comment (specific 1): Line 14, page 1: "eligible for retrieving aerosol layer height". Is this because of clouds? If this is the case, say so.

Author's response: The rough estimate of 3% of total eligible pixels for retrieving aerosol layer height needs to be updated following new analyses of 56 TROPOMI files over Europe containing 7.3 million pixels. The selection criterion was a UV Aerosol Indices above 0.0, of which 6.1% of all pixels considered over Europe. TROPOMI's UV Aerosol Index values are one index point lower than UV Aerosol Indices from other indices.

Changes to the manuscript: The new sentence now reads the following:

'With TROPOMI recording approximately 1.4 million pixels within a single orbit, a rough estimate based on a minimum UV Aerosol Index of 0 indicates that at least six percent of all pixels over an area as large as Europe will be eligible for retrieving aerosol layer height. This number can go beyond 50,000 pixels per orbit in many cases, placing a steep requirement on the computational infrastructure to process all possible pixels from a single orbit.'

Reviewer comment (specific 2): Lines 26-28, page 1: The previous sentence suggests that the method utilized line-by-line calculations to generate training data set. Do the authors mean that line-by-line calculations are not used in the "operational" retrieval that utilizes neural networks? The authors also need to say more to distinguish their method from that used by Chimot et al. and Loyola et al.

Reviewer comment (specific 3): Line 33, page 1 – Line 1, page 2: It is not clear what the difference is from the existing neural network approaches. Is it that Optimal Estimation is used? "using artificial neural networks to improve the computational speed of RT calculations" is very vague and general; isn't that common to all neural network approaches?

Author's response: The work by Chimot et al. utilise the DISAMAR radiative transfer model to compute synthetic OMI measurements of the slant column density in the O_2-O2 absorption band, which is a part of the feature vector in the neural network models in their implementation of retrieving aerosol layer height. Chimot et al. do not use any line-by-line calculations in their operational retrievals.

The aerosol layer height algorithm that is the subject of this paper follows a similar philosophies to Chimot et al. as well as Loyola et al, with important differences.

- With respect to Chimot et al. the paper discusses using DISAMAR to generate synthetic spectra for training a neural network model, the difference being that while Chimot et al. prefer to retrieve aerosol layer height as the output of their trained neural network model, whereas the neural network model in the paper outputs

the top-of-atmosphere oxygen A-band spectra in the forward model. These neural-network-model-calculated top-of-atmosphere oxygen A-band spectra are then utilised by an optimal estimation scheme, which outputs a retrieved aerosol layer height value.

- With respect to Loyola et al, the neural network models both compute top-of-atmosphere spectra, with the difference being that the KNMI aerosol layer height neural network retrieval algorithm has two other neural network models for the derivatives of the spectra with respect to the state vector parameters aerosol optical thickness and aerosol layer height, whereas Loyola et al. train their neural network models only for the sun-normalised radiances (section 4.4, `https://www.atmos-meas-tech.net/11/409/2018/amt-11-409-2018.pdf`)

Changes to the manuscript:

- Adjusted text to 'Chimot et al. (2017) describe an approach using a radiative transfer model to generate OMI slant column densities of the O_2-O2 band at 477 nm for different aerosol optical depths (among other input parameters) to train several artificial neural network models that directly retrieve aerosol layer height. Operationally, their neural network models use the MODIS aerosol optical depth at 550 nm product and retrieved OMI slant column densities, thereby entirely foregoing line-by-line calculations and significantly speeding up the retrieval algorithm.'.
- Amended the final paragraph of section 1 to 'The work of Chimot et al. (2017) and Loyola et al. (2018) bring to light the efficacy of artificial neural networks in satellite remote sensing of oxygen absorption bands for retrieving properties of scattering species in the atmosphere. This paper discusses a method inspired by Chimot et al. and Loyola et al. to retrieve aerosol layer height from oxygen A-band measurements by TROPOMI. While Chimot et al. directly retrieve aerosol layer heights from their neural network models, the operational algorithm in this paper utilises neural networks to calculate top-of-atmosphere radiances in the forward model. This is subsequently used by an optimal estimation scheme to retrieve aerosol layer heights. Similarly while Loyola et al. derive top-of-atmosphere sun-normalised radiances only for their cloud property retrieval algorithm, the method in this paper has dedicated neural network models that calculate the Jacobian as well as the top-of-atmosphere sun-normalised radiances.'.

Reviewer comment (specific 4): Line 8, page 2: Add the following references:

Timofeyev et al., 1995; Sanghavi et al., 2012; Geddes Bösch, 2015; Colosimo et al., 2016; Davis et al., 2017; Xu, et al., 2017; Zeng et al., 2018

Colosimo, S. F., V. Natraj, S. P. Sander, and J. Stutz (2016), A sensitivity study on the retrieval of aerosol vertical profiles using the oxygen A-band, Atmos. Meas. Tech., 9(4), 1889–1905, doi:10.5194/amt-9-1889-2016.

Davis, A. B., O. V. Kalashnikova, and D. J. Diner (2017), Aerosol layer height over water from O2 A-band: mono-angle hyperspectral and/or bispectral multi-angle observations, Preprint, doi:10.20944/preprints201710.0055.v1.

Geddes, A., and H. Bösch (2015), Tropospheric aerosol profile information from high resolution oxygen A-band measurements from space, Atmos. Meas. Tech., 8(2), 859–874, doi:10.5194/amt-8-859-2015.

Sanghavi, S., J. V. Martonchik, J. Landgraf, and U. Platt (2012), Retrieval of aerosol optical depth and vertical distribution using O2 A- and B-band SCIAMACHY observations over Kanpur: A case study, Atmos. Meas. Tech., 5(5), 1099–1119, doi:10.5194/amt-5-1099-2012

Timofeyev, Y. M., A. V. Vasilyev, and V. V. Rozanov (1995), Information content of the spectral measurements of the 0.76 m O2 outgoing radiation with respect to the vertical aerosol optical properties, Adv. Space Res., 16(10), 91–94, doi:10.1016/0273-1177(95)00385-R.

Xu, X., Wang, J., Wang, Y., Zeng, J., Torres, O., Yang, Y., et al. (2017). Passive remote sensing of altitude and optical depth of dust plumes using the oxygen A and B bands: First results from EPIC/DSCOVR at Lagrange-1 point, Geophys. Res. Lett., 44, 7544–7554, doi:10.1002/2017GL073939.

Zeng, Z.-C., V. Natraj, F. Xu, T. J. Pongetti, R.-L. Shia, E. A. Kort, et al. (2018), Constraining aerosol vertical profile in the boundary layer using hyperspectral measurements of oxygen absorption, Geophys. Res. Lett., 45, doi:10.1029/2018GL079286.

Author's response: Agreed.

Changes to the manuscript: Amended the first sentence of section 2 that discusses previous work done for retrieving vertical information of aerosols using passive spaceborne measurements of the oxygen A-band to 'The TROPOMI aerosol layer height is one of the many algorithms that exploit vertical information of scattering aerosol species in the oxygen A-band (Timofeyev et al., 1995; Gabella et al., 1999; Corradini and Cervino, 2006; Pelletier et al., 2008; Dubuisson et al., 2009; Frankenberg et al., 2012; Wang et al., 2012, Sanghavi et al., 2012; Sanders and de Haan, 2013; Hollstein and Fischer, 2014; Geddes Bösch, 2015; Sanders et al., 2015; Colosimo et al., 2016; Sanders and de Haan, 2016; Davis et al., 2017; Xu, et al., 2017; Zeng et al., 2018; Nanda et al., 2018b)'

Reviewer comment (specific 5): Line 24-26, page 4: "The polarized . . . 760 nm".

- First order scattering also has a polarization effect. Presumably, the authors mean that they ONLY compute the first order polarization. If so, please state that.
- What is "small"? $< 5\%$? $< 1\%$? This statement is potentially untrue. If true, values for how small is small must be given, with proof. In the continuum and in weak lines, the second order effects might be large.

Reviewer comment (specific 6): Lines 27-28, page 4: There is a contradiction here. If the exclusion is not advised, the effect cannot be small. The authors should simply state that they ignored this for computational reasons. Besides, the whole point of using neural networks is to speed up calculations. Why not use them for speeding up Raman calculations too, or at least use lookup tables for Raman effects?

Reviewer comment (specific 7): Lines 29-30, page 4: "From preliminary . . . significantly".

Quantify this statement, and ideally provide a figure to illustrate the effects.

Author's response:

- Polarisation is ignored in the sense that for retrieving aerosol layer heights DISAMAR only computes the first element of the Stoke's vector in the radiation fields. The exclusion of higher order Stoke's vector elements has not shown to be a significant source of error.
- The affect of ignoring Rotational Raman Scattering (RRS) in the forward model results in errors in the final retrieved aerosol layer heights. However, as clarified by Sanders and de Haan (2016) who have retrieved aerosol layer height using the same radiative transfer model while including and excluding RRS, this error is significantly less in comparison to other model errors. Because the inclusion of RRS has resulted in a significant increase in time required by the line-by-line radiative transfer model and ignoring it does not yield large errors (from synthetic experiments), it has been historically excluded in the KNMI aerosol layer height retrieval algorithm.

With regards to the reviewer's suggestions to use lookup tables for the Raman effects or potentially incorporating an artificial neural network solution for including RRS into the forward model calculations, the authors appreciate these ideas very much. However, since the goal of this paper is to create a model that replicates the existing TROPOMI aerosol layer height algorithm, RRS is ignored for the sake of comparison and benchmarking. In the future, this may be a serious consideration by the TROPOMI Level-2 algorithm development team.

Finally, the authors acknowledge the confusion in this paragraph. The sentence 'RRS can alter the line depths in the O2 A-band, but this effect is small', is not complete. The authors meant to state that the effect of excluding RRS on the retrieved aerosol layer height is small. The changes to the manuscript will reflect this. With regards to the reviewer's comments on quantifying the statement, the authors have chosen to include a citation to Sanders

and de Haan 2016, which is the Algorithm Theoretical Basis Document of the TROPOMI ALH retreival algorithm. This discusses the rationale behind excluding RRS from computations.

Changes to the manuscript: To address the questions raised by the reviewer, the following amendments have been done to Section 2.2, paragraph 3 of the manuscript.

- 'As the Rayleigh optical thickness is low at 760 nm, DISAMAR only computes the monochromatic component of light by calculating the first element of the Stoke's vector. The exclusion of higher order Stoke's vector elements of the radiation fields has not shown to be a significant source of error (Sanders and de Haan, 2016).'
- '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.'

Reviewer comment (specific 8): Lines 30-31, page 4: Although it is true that retrievals are typically performed under "cloud free" conditions, optically thin cirrus clouds need to be accounted for since they are almost always present.

Author's response: It is indeed correct that optically thin cirrus clouds need to be accounted for as they are almost always present in the scene. Currently however, there are no implementations in the algorithm to incorporate cirrus cloud properties into the radiative transfer calculations. The operational TROPOMI algorithm utilises a VIIRS cloud mask to flag potential pixels with clouds.

Changes to the manuscript: Added the following sentence: 'While optically thin cirrus layers are a known source of error in the retrieved aerosol layer height, currently there are no implementations to tackle this problem. Instead, TROPOMI incorporates information from the VIIRS instrument to detect the presence of clouds in the measured scene.'

Reviewer comment (specific 9): Lines 1-8, page 5: What are the effects of these approximations on the retrieved results? It seems that many of these simplifications are not needed because of the use of neural networks. Also, if only single scattering is used, calculation of ANY phase function is trivial and not time consuming. Considering the fact that the authors aim to produce an operational retrieval algorithm, these simplifications seem unwarranted and restrictive.

Author's response: It is indeed correct that the simplifications are unwarranted and restrictive for a retrieval algorithm that incorporates a fast neural network approach to replace a radiative transfer model. However the goal of the paper is to replicate (as much as possible) the operational algorithm that uses online line-by-line calculations, which incorporates these approximations to reduce computational time. Finally, the paper compares the retrieved aerosol layer heights from both operational algorithm implementations in order to establish an acceptable agreement between the neural network approach to the online line-by-line approach. This is the first benchmark of the neural-network-augmented retrieval algorithm, subsequently leading to further improvements in the future in line with the reviewer's recommendations.

The affects of these approximations are discussed in detail by Sanders and de Haan (2016), which is the Algorithm Theoretical Basis Document of the TROPOMI aerosol layer height algorithm. The amendment in the manuscript will reflect their work.

Changes to the manuscript: Added the following final paragraph after the mentioned simplifications.

'These simplifications in the DISAMAR forward model are a necessity for the line-by-line aerosol layer height algorithm, owing to its slow computational speed. In contrast, a neural network model is significantly faster. While the speed of the neural network model encourages increasing the complexity of the model, for a comparative study

the neural network models are trained to replicate, as best as possible, the line-by-line version. Once this is achieved, the improvement of the algorithm will be an iterative endeavour.'

Reviewer comment (specific 10): Table 2, page 8: What does "varied" mean for the aerosol layer thickness? Is the aerosol layer thickness part of the feature vector? If not, how is it handled?

Author's response: Aerosol layer thickness is not a part of the feature vector. It is a part of the training data set, and the aerosol layer pressure thickness varies between 50 hPa and 200 hPa. Currently, there is no call by the neural network model to the aerosol layer thickness. This shall be implemented into a future release.

Changes to the manuscript: Amended the table entries in Table 2 for aerosol layer thickness. The remark column now reads 'varied but excluded from feature vector', whereas the limits now read '50 hPa - 200 hPa'.

Reviewer comment (specific 11): Line 8, page 8: "a choice of 500,000 Disamar generated spectra"

How are these spectra generated? It is not clear how the choice is made.

Author's response: The amendment will reflect the clarification of spectra generation.

Changes to the manuscript: Changed the sentence 'Following testing and scrutinizing forward model calculation accuracy, a choice of 500,000 Disamar generated spectra is finalised as the size of the training data set.' to the following.

'The number of spectra generated for the training set was determined by training different models with different number of spectra in the training set ranging from 1,000 to 600,000. In general it was observed that incorporating more data resulted in a better neural network model. In order to test the trained neural network model, a choice of 500,000 spectra were selected, and 100,000 spectra were set aside for the test set. These spectra were generated using Disamar with model parameter ranges described in Table 2 and Figure 1.'

To that extent, the following line is removed from Page 6, line 3-5, as there is an incorrect reference to the correct table.

'Finding the most optimal neural network configuration requires a test data set which in this case contains 100,000 scenes outside the training data set. These test data follow the same input model 5 parameter distributions as described in Figure 1 and Table 1.'

Reviewer comment (specific 12): Lines 20-22, page 9: Need more quantitative error information, like for the derivative with respect to tau. Also, what does continuum (3d) mean?

Reviewer comment (specific 13): Lines 23-24, page 9: "these parts . . . cross sections"

Why do low oxygen absorption cross sections lead to low aerosol information content?

Author's response: The following amendment to the text clarifies the role of oxygen absorption cross sections and aerosol information content.

Changes to the manuscript: The following change has been added to the text in the final lines of the final paragraph of section 3.2.

'The neural network model for the derivative of the reflectance with respect to τ and z_{aer} perform well in general for parts of the spectrum with large oxygen absorption cross sections, where the value of the derivatives are high (indicating a higher amount of information content from those specific wavelength regions). Errors in the deepest part of the R-branch between 759 nm and 762 nm and the P-branch between 752.50 nm and 765 nm, do not exceed more than 3% for $\text{NN}_{\text{K}_{\text{zaer}}}$. The same can be said for NN_{K_τ}, which displays errors in the range of 1% in the same wavelength region. For wavelengths outside of the deepest parts of the R and P-branch, the relative errors are large, and exceed 10% easily. However, the relative errors are calculated as the absolute value of the difference between the true spectrum and the neural network calculated spectrum, divided by the true spectrum. These values can

be very large when the value of the true spectrum is very small, which is the case for the derivatives outside the deepest part of the R and P branches. The consequence of these errors in a retrieval scenario from synthetic and real spectra are discussed in the following section.'

Reviewer comment (technical):

Line 19, page 1: correlative → correlated

Line 20, page 1: Hasekamp and Butz (2008) → (Hasekamp and Butz, 2008)

Line 19, page 4: in → on

Line 25, page 5: an → a

Line 17, page 7: differentiation which → differentiation, which

Line 2, page 8: selected TROPOMI → selected to represent TROPOMI

Line 6, page 8: an → and

Line 7, page 8: isn't a → is no

Line 14, page 8: legible → physical

Line 12, page 9: were trained → was trained

Line 18, page 9: to → of

Line 20, page 9: remove "more than"

Line 23, page 9: at → in

Line 24, page 9: with respect → compared

Line 16, page 10: less than → by less than; remove "approximately"

Line 23, page 10: less than → less by

Line 29, page 10: were → was

Table 3 caption, page 11: an → and

Line 13, page 11: agreements, they primarily departed in the → agreement, they primarily differed for the

Line 16, page 11: departure, different → bias, differing

Line 17, page 11: departure → bias

Table 4 caption, page 11: disamar → Disamar

Figure 1 caption, page 14: available → shown

Figure 2 caption, page 15: A schematic → Schematic

Figure 4, page 17: need x axis label for (a), x and y axis labels for (b), correct x and y axis labels for (c), y axis label for (d), correct y axis label for (d)

Figure 5 caption, page 18: A histogram → Histogram; plotting → plotted

Figure 6 caption, page 19: A histogram → Histogram

Figure 7 caption, page 20: A MODIS → MODIS; remove "the" before "ocean"; remove "either"; cloud mask, or by a land-sea mask → cloud mask or land-sea mask

Figure 8 caption, page 21: represents the difference → Difference

Figure 9 caption, page 21: Figre (a) directly compares retrieved aerosol layer heights from the two methods. Figure (b) provides a histogram of the difference between these retrieved heights from Disamar and NN. The difference is defined as zaer(Disamar) - zaer(NN). Figure (c) compares these differences with TROPOMI's operational absorbing aerosol index product (x axis). → (a) Retrieved aerosol layer heights from the two methods; (b) Histogram of the

difference between retrieved heights from Disamar and NN. The difference is defined as zaer(Disamar) - zaer(NN); (c) Differences compared to TROPOMI's operational absorbing aerosol index product (x axis).

Author's response: Agreed.

Changes to the manuscript: Amended the document as requested by the reviewer.

2 Response to amt-2019-143-RC2

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 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 temperature 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 as follows.

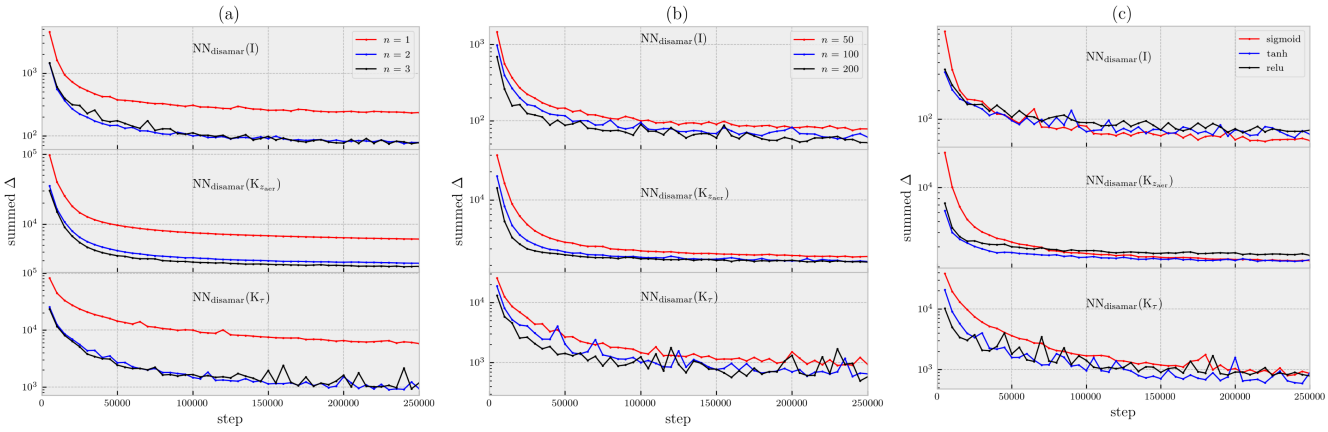

Figure 2: Summed loss as a function of training step for different neural network model configurations. **(a)** The neural network models have 50 nodes per each layer with a sigmoid activation function. **(b)** The neural network models have two hidden layers with each node activated by the sigmoid function. **(c)** The neural network models have two hidden layers with a 100 nodes for each layer.

The figure numbers of the subsequent figures are automatically adjusted. 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 compared 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 $\mathrm{NN_{disamar}(K_\tau)}$ 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.

A neural network radiative transfer model approach applied to TROPOMI's aerosol height algorithm

Swadhin Nanda[1,2], Martin de Graaf[1], J. Pepijn Veefkind[1,2], Mark ter Linden[3], Maarten Sneep[1], Johan de Haan[1], and Pieternel F. Levelt[1,2]

[1]Royal Netherlands Meteorological Institute (KNMI), Utrechtseweg 297, 3731 GA De Bilt, The Netherlands
[2]Delft university of Technology (TU Delft), Mekelweg 2, 2628 CD Delft, The Netherlands
[3]S&T Corp, Delft, The Netherlands

Correspondence to: Swadhin Nanda (nanda@knmi.nl)

Abstract. To retrieve aerosol properties from satellite measurements of the oxygen A-band in the near infrared, a line-by-line radiative transfer model implementation requires a large number of calculations. These calculations severely restrict a retrieval algorithm's operational capability as it can take several minutes to retrieve aerosol layer height for a single ground pixel. This paper proposes a forward modeling approach using artificial neural networks to speed up the retrieval algorithm. The forward model outputs are trained into a set of neural network models to completely replace line-by-line calculations in the operational processor. Results of comparing the forward model to the neural network alternative show encouraging results with good agreements between the two when applied to retrieval scenarios using both synthetic and real measured spectra from TROPOMI (TROPOspheric Monitoring Instrument) on board the ESA Sentinel-5 Precursor mission. With an enhancement of the computational speed by three orders of magnitude, TROPOMI's operational aerosol layer height processor is now able to retrieve aerosol layer heights well within operational capacity.

1 Introduction

Launched in October 13, 2017, The TROPOsperic Monitoring Instrument (Veefkind et al., 2012) on board the Sentinel-5 Precursor mission is the first of the satellite-based atmospheric composition monitoring instruments in the Sentinel mission of the European Space Agency. The aerosol layer height (ALH) retrieval algorithm (Sanders and de Haan, 2013; Sanders et al., 2015; Nanda et al., 2018a, b) is a part of TROPOMI's operational product suite, expected to be delivered near real time. The ALH (symbolised as $z_{\rm aer}$) retrieval algorithm, operating within the near infrared region in the oxygen A-band between 758 nm - 770 nm, exploits information about heights of scattering layers derived from absorption of photons by molecular oxygen — the amount of absorption indicates whether the scattering layer is closer or farther from the surface; if the number of photons absorbed by oxygen is higher, it suggests a longer photon path length due to an aerosol layer present closer to the surface. This principle has been applied to cloud height algorithms such as FRESCO (Fast Retrieval Scheme for Clouds from the Oxygen A-band) by Wang et al. (2008), which use look up tables for generating top of atmosphere (TOA) reflectances to compute cloud parameters. Since clouds are such efficient scatterers of light, FRESCO can approximate scattering by cloud using a Lambertian model — this simplification works for optically thick cloud layers quite well. For aerosol layers, however, such calculations

need to be done in much greater detail due to their weaker scattering properties. TROPOMI's ALH algorithm employs the science code Disamar (Determining Instrument Specifications and Methods for Atmospheric Retrievals) that uses the Layer-Based Orders of Scattering (LABOS) radiative transfer model based on the doubling-adding method (de Haan et al., 1987) that calculates reflectances at the TOA and its derivatives with respect to aerosol layer height and aerosol optical thickness (τ). These calculations are done line-by-line, requiring calculations at 3980 wavelengths to generate these TOA reflectances within the oxygen A-band. Having computed the TOA reflectance spectra, aerosol layer heights are retrieved with Optimal Estimation (OE), an iterative retrieval scheme developed by Rodgers (2000) that incorporates a priori knowledge of retrieval parameters into their estimation. Such a retrieval scheme also provides a posteriori error estimations, which are important for assimilation models and diagnosing the retrieval results.

The ALH retrieval algorithm is computationally expensive, requiring several minutes to compute z_{aer} for a single ground pixel (Sanders et al., 2015). As near-real time processors need to consistently go through large volumes of data recorded by the satellite for the mission lifetime, operational retrievals are time restricted. With TROPOMI recording approximately 1.4 million pixels within a single orbit, a rough estimate ~~of an average of three~~ based on a minimum UV Aerosol Index of 0 indicates that at least six percent of all ~~TROPOMI pixels in an orbit~~ pixels over an area as ~~big as Europe may~~ large as Europe will be eligible for retrieving aerosol layer height. This number can go ~~up to as much as~~ beyond 50,000 pixels per orbit ~~. This places~~ in many cases, placing a steep requirement on the computational infrastructure to process all possible pixels from a single orbit. The online radiative transfer model severely limits the ALH data product, processing only a small fraction of the total possible pixels within a single orbit while compromising the timeliness of the data delivery.

The bottleneck identified here is the large number of calculations that the forward model has to compute to retrieve information on weak scatterers such as aerosols. Several steps to circumvent this bottleneck exist, such as using ~~correlative~~ correlated k-distribution method to reduce the number of calculations ~~Hasekamp and Butz (2008)~~(Hasekamp and Butz, 2008), using a look up table for calculating forward model outputs, or entirely foregoing the forward model and directly retrieving z_{aer} from observed spectra using neural networks (Chimot et al., 2017, 2018). Studies by Sanders and de Haan (2016) have shown that the look up table for reflectance alone measure up to 46 GB in size, and perhaps similar or larger sizes for the derivatives. Chimot et al. (2017) describe an ~~artificial neural network approach using the same~~ approach using a radiative transfer model ~~as for TROPOMI to generate training data, in combination with the NASA~~ to generate OMI slant column densities of the O_2-O2 band at 477 nm for different aerosol optical depths (among other input parameters) to train several artificial neural network models that directly retrieve aerosol layer height. Operationally, their neural network models use the MODIS aerosol optical depth ~~product, and successfully retrieve aerosol layer heights directly from the O_2-O_2 bands in the visible spectral region at 477 nm~~at 550 nm product and retrieved OMI slant column densities, thereby entirely foregoing line-by-line calculations and significantly speeding up the retrieval algorithm. They demonstrated ~~this~~ their algorithm by retrieving aerosol layer heights from spectra measured by the Ozone Monitoring Instrument (OMI) on board the NASA Aura mission, without using line-by-line calculations or an iterative estimation step such as OE (Chimot et al., 2018). A similar example of retrievals is the ROCINN (Retrieval of Cloud Information using Neural Networks) cloud algorithm developed by Loyola (2004) which uses

neural networks to compute convolved reflectance spectra to retrieve cloud properties. These retrievals show the exploitable capabilities of artificial neural networks in the context of retrieving atmospheric properties from oxygen absorption bands.

The work of ~~Chimot et al. (2017) brings to light an interesting use case~~ Chimot et al. (2018) and Loyola et al. (2018) bring to light the efficacy of artificial neural networks ~~for retrieving aerosol information from~~ in satellite remote sensing of oxygen absorption bands for retrieving properties of scattering species in the atmosphere. This paper ~~approaches the problem from a different direction by using artificial neural networks to improve the computational speed of the radiative transfer calculations of the reflectance and its derivatives with respect to retrieval parameters, and keeping intact the OE approach as the a posteriori statistics generated act as diagnostic tools for analysing retrieval behaviour~~discusses a method inspired by Chimot et al. and Loyola et al. to retrieve aerosol layer height from oxygen A-band measurements by TROPOMI. While Chimot et al. directly retrieve aerosol layer heights from their neural network models, the operational algorithm in this paper utilises neural networks to calculate top-of-atmosphere radiances in the forward model. This is subsequently used by an optimal estimation scheme to retrieve aerosol layer heights. Similarly while Loyola et al. derive top-of-atmosphere sun-normalised radiances only for their cloud property retrieval algorithm, the method in this paper has dedicated neural network models that calculate the Jacobian as well as the top-of-atmosphere sun-normalised radiances. By reducing the time consumed for calculating forward model outputs, computational efficiency of TROPOMI's aerosol layer height retrieval algorithm can be significantly improved.

Section 2 introduces the operational aerosol layer height algorithm and discusses the line-by-line forward model. The neural network forward model approach is detailed in section 3, and its verification on a test data set is discussed in same section. This approach is then applied to various test cases using synthetic and real TROPOMI spectra (section 4) before concluding in section 5.

2 The TROPOMI aerosol layer height retrieval algorithm

The TROPOMI aerosol layer height is one of the many algorithms that exploit vertical information of scattering aerosol species in the oxygen A-band ~~(Gabella et al., 1999; Corradini and Cervino, 2006; Pelletier et al., 2008; Dubuisson et al., 2009; Frankenberg et al.,~~ (Timofeyev et al., 1995; Gabella et al., 1999; Corradini and Cervino, 2006; Pelletier et al., 2008; Dubuisson et al., 2009; Frankenberg et al. . These methods invert a forward model that describes the atmosphere, to compute the height of the scattering layer. This section discusses the setup of the TROPOMI ALH retrieval algorithm, which consists of the inversion of a forward model representing the atmosphere using optimal estimation as the retrieval method, and a description of the forward model.

2.1 The retrieval method

The cost function χ^2 represents the departure of the modeled reflectance $\boldsymbol{F}(\boldsymbol{x})$ from the observed reflectance $\boldsymbol{y}$ scaled by the measurement error covariance matrix $\mathbf{S}_\epsilon$, and is defined as

$$\chi^2 = [\boldsymbol{y} - \boldsymbol{F}(\boldsymbol{x})]^T \mathbf{S}_\epsilon{}^{-1} [\boldsymbol{y} - \boldsymbol{F}(\boldsymbol{x})] + (\boldsymbol{x} - \boldsymbol{x_a})^T \mathbf{S_a}{}^{-1} (\boldsymbol{x} - \boldsymbol{x_a}). \quad (1)$$

Minimising this cost function for a particular z_{aer} and τ (the elements of the state vector $\boldsymbol{x}$ to be retrieved and fitted) gives us the final retrieval product. This definition of the cost function is unique to OE, as it constrains its minimisation with a priori knowledge of the state vector $\boldsymbol{x}$, contained in $\boldsymbol{x_a}$ and the a priori error covariance matrix $\mathbf{S_a}$. In the TROPOMI ALH processor's OE framework, the a priori state vector is fixed at specific values, usually 200 hPa above the surface for z_{aer} and 1.0 for τ at 760 nm. The a priori error of the z_{aer} is fixed at 500 hPa, and the same for τ is 1.0, to allow freedom for the variables in the estimation (this also reduces the impact of the a priori on the retrieval). The modeled measured reflectance spectrum is calculated using the forward model (denoted as F) for model parameters $\boldsymbol{x}$ following,

$$F(\mathbf{x})(\lambda) = \frac{\pi I(\lambda)}{\mu_0 E_0(\lambda)}, \tag{2}$$

where μ_0 is the cosine of the solar zenith angle θ_0, $I(\lambda)$ for wavelength λ) is the Earth radiance and $E_0(\lambda)$ is the solar irradiance. Since the forward model is non-linear, a Gauss-Newton iteration is employed and the updated state vector is calculated as,

$$\boldsymbol{x_{i+1}} = \boldsymbol{x_a} + [\mathbf{K_i}^T \mathbf{S_\epsilon}^{-1} \mathbf{K_i} + \mathbf{S_a}^{-1}]^{-1} \mathbf{K_i}^{-1} \mathbf{S_\epsilon}^{-1} [\boldsymbol{y} - \boldsymbol{F(x)} + \mathbf{K_i}(\boldsymbol{x_i} - \boldsymbol{x_a})], \tag{3}$$

where i is the current iteration and $\mathbf{K_i}$ is the matrix of derivatives (Jacobian) of the reflectance with respect to state vector parameters at the current iteration. The derivatives are calculated semi-analytically similar to the method described by Landgraf et al. (2001). The retrieval is said to converge to a solution if the state vector's update is less than the expected precision (usually fixed at a certain value). The retrieval fails to converge if the number of iterations exceeds the maximum number of iterations (usually set at 12), or if the state vector parameters are projected outside their respective boundary conditions by OE. Retrieval errors are derived from the a posteriori error covariance matrix $\hat{\mathbf{S}}$, computed as

$$\hat{\mathbf{S}} = [\mathbf{K}^T \mathbf{S_\epsilon}^{-1} \mathbf{K} + \mathbf{S_a}^{-1}]^{-1}. \tag{4}$$

2.2 The Disamar forward model and its many simplifications of atmospheric properties

The forward model generates synthetic observed TOA radiance spectra by an instrument for a specific solar-satellite geometry, which is required for minimising χ^2 (Equation 1). For this, a high resolution reference solar spectrum adopted from Chance and Kurucz (2010) is used to obtain the TOA Earth radiance spectrum, which is further convolved with the instrument's slit function and combined with the solar irradiance to compute reflectances following Equation 2.

Radiances are calculated by accounting for scattering and absorption of photons from their interactions with aerosols, the surface and molecular species. Molecular scattering of photons in the oxygen A-band is described by Rayleigh scattering, and absorption is described by photon-induced magnetic dipole transition between $b^1\Sigma_g^+ \leftarrow X^3\Sigma_g^-(0,0)$ electric potential levels of molecular oxygen, and collision-induced absorption between O_2-O_2 and O_2-N_2. The total influence of the O_2 A-band ~~in~~ on the TOA reflectance is described by its extinction cross-section, which is a sum of the three aforementioned contributions. As

the vertical distribution of oxygen is exactly known, the extinction cross-section can be exploited to retrieve z_{aer} from satellite measurements of the oxygen A-band. For this, Disamar calculates absorption (or extinction) cross sections at 3980 wavelengths within the range 758 nm - 770 nm.

To reduce the number of calculations, various atmospheric properties are simplified. ~~The polarised component of light need not be calculated because second order scattering by air molecules is small compared to first order scattering, as~~ As the Rayleigh optical thickness is ~~small around~~ low at 760 nm~~.~~, Disamar only computes the monochromatic component of light by calculating the first element of the Stoke's vector. The exclusion of higher order Stoke's vector elements of the radiation fields has not shown to be a significant source of error (Sanders and de Haan, 2016).

Calculating the influence of Rotational Raman Scattering (RRS) is also ignored, as it is a computationally expensive step. ~~This exclusion of calculations~~ While this exclusion of RRS is not advised by literature (Vasilkov et al., 2013; Sioris and Evans, 2000), ~~as RRS can alter the line depths in the O_2~~ 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 ~~, but this effect is small. The choice of ignoring RRS is borne out of computational burden it puts on the overall retrieval algorithm. From preliminary tests, the exclusion of RRS seems to not affect z_{aer} retrievals significantly~~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. The atmosphere is assumed cloud-free, which is a required simplification as the retrieval of z_{aer} in the presence of clouds becomes challenging. ~~The aerosol fraction is assumed as 1.0~~While optically thin cirrus layers are a known source of error in the retrieved aerosol layer height, currently there are no implementations to tackle this problem. Instead, TROPOMI incorporates information from the VIIRS instrument to detect the presence of clouds in the measured scene, which are further on mentioned in the ouput product flags. The fraction of the pixel containing aerosols is assumed to be 100%, which further simplifies the representation of aerosols within the atmosphere. Perhaps the largest simplification of the atmosphere lies in model's description of aerosols, assumed to be distributed in a homogeneous layer at a height z_{aer} with a 50 hPa thickness, a fixed aerosol optical thickness (τ) and a single scattering albedo of 0.95 (so, scattering aerosols). The aerosol scattering phase function assumed is a Henyey-Greenstein model (Henyey and Greenstein, 1941), instead of alternatives such as Mie-scattering models which require significantly more computations. ~~Finally, the~~ 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. 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~~ spherically corrected for incoming solar radiation and remains plane-parallel for outgoing Earth radiance.

These simplifications in the Disamar forward model are a necessity for the line-by-line aerosol layer height algorithm, owing to its slow computational speed. In contrast, a neural network model is significantly faster. While the speed of the neural network model encourages increasing the complexity of the model, for a comparative study the neural network models are

trained to replicate, as best as possible, the line-by-line version. Once this is achieved, the improvement of the algorithm will be an iterative endeavour.

2.3 Application to TROPOMI

TROPOMI's near infrared (NIR) spectrometer records data between 675 nm - 775 nm, spread across two bands — band 5 contains the oxygen B-band and band 6 the oxygen A-band. The spectral resolution, which is described by the full width at half maximum (FWHM) of the instrument spectral response function (ISRF), is 0.38 nm with a spectral sampling interval of 0.12 nm. The spatial resolution is around 7 km $\times$ 3.5 km for band 5 and 6. Initial observations from the TROPOMI NIR spectrometer show a signal to noise ratio (SNR) of 3000 in the continuum before the oxygen A-band. The instrument polarization sensitivity is reduced to below 0.5% by adopting the technology of the polarization scrambler of the ozone monitoring instrument (OMI) (Veefkind et al., 2012; Levelt et al., 2006). Disamar utilizes TROPOMI's swath-dependent ISRFs to convolve $I(\lambda)$ and $E_0(\lambda)$ into $I(\lambda_i)$ and $E_0(\lambda_i)$ in the instrument's spectral wavelength grid, after which the modeled measured reflectance is calculated using Equation 2.

Input parameters required by the TROPOMI ALH retrieval algorithm encompass satellite observations of the radiance and the irradiance, solar-satellite geometry, and a host of atmospheric and surface parameters required for modeling the interactions of photons within the Earth's atmosphere (see Table 1). Meteorological parameters are derived from ECMWF (European Centre for Medium-range Weather Forecast), which provide the temperature-pressure profile at 91 atmospheric levels. The various databases supplying meteorological and surface parameters are interpolated to TROPOMI's ground pixels using nearest neighbour interpolation.

Table 1. Input parameters required for retrieving aerosol layer height using TROPOMI measured spectra.

Parameter	Source	Remarks
Radiance and irradiance	TROPOMI Level-1b product	
SNR measured spectrum	TROPOMI Level-1b product	
Geolocation parameters	TROPOMI Level-1b product	
Surface albedo	GOME-2 LER database	Tilstra et al. (2017)
Meteorological parameters	ECMWF	17km horizontal resolution
Cloud fraction	TROPOMI Level-2 FRESCO product	
Absorbing aerosol index	TROPOMI Level-2 AAI product	
Land-sea mask	NASA Toolkit	
Surface altitude	GMTED 2010	pre-averaged

Calculation of TOA reflectance and its derivatives with respect to z_{aer}, and τ in ~~an~~ a line-by-line fashion requires approximately 40-60 seconds to complete on a computer equipped with Intel(R) Xeon(R) CPU E3-1275 v5 at a clock speed of 3.60 GHz. In an iterative framework such as the Gauss-Newton method, the retrieval of z_{aer} can take between 3-6 iterations depending on the amount of aerosol information available in the observed spectra, requiring several minutes to compute retrieval

outputs for a specific scene. If these retrievals fail by not converging within the maximum number of iterations, the processor can waste up to 10 minutes on a pixel without retrieving a product. In order to compute Disamar's outputs quicker, a neural network implementation is discussed in the next section.

3 The neural network (NN) forward model

Artificial neural networks consist of connected processing units, each individually producing an output value given a certain input value. The interaction of these individual processing units, also known as nodes (or neurons), enable the connecting network to map a set of inputs (also known as the input layer) to a set of outputs (or, the output layer). The connections are known as weights whose value symbolises the strength of a connection between two nodes. Since the nodes connect inputs to the outputs, higher values in a set of connecting weights represent a stronger influence of a particular parameter in the input layer over a particular parameter in the output layer. These weights are determined after training the neural network.

The training (or optimisation) of a neural network begins with a training data set containing many instances of input and output layer elements. As true values of the output layer for a given set of inputs are exactly known in the training data set, the biased output of the neural network calculated after using randomised, non-optimised weights can be easily calculated. These biases are called prediction errors, an essential element in the optimization of the neural network weights. The mean squared error (MSE) between the true output and the calculated output is also called the loss function (henceforth annotated as Δ), which is synonymous to a cost function (Equation 5),

$$\Delta = \frac{1}{n_\lambda} \sum_{\forall \lambda} (nn_\lambda - o_\lambda)^2 \tag{5}$$

where λ is the wavelength, n_λ represents the number of elements in the output layer, nn_λ represents the calculated output for wavelength via forward propagation, and o_λ are the outputs in the training data set. The weights are updated using optimisers such as the ADAM optimiser (Adaptive Moment Estimation, Kingma and Ba (2014)) to minimise Δ, within set number of iterations.

3.1 The TROPOMI NN forward model for the ALH retrieval algorithm

The standard architecture of the NN-augmented operational aerosol layer height processor includes three neural network models for estimating top of atmosphere sun-normalised radiance, the derivative of the reflectance with respect to z_{aer}, and the same for τ. It is also possible to assign the neural network to compute the reflectance instead of the sun-normalized radiance — the results will not change. The definition of sun-normalised radiance used in this paper is the ratio of Earth radiance to solar irradiance. Disamar calculates derivatives with respect to reflectance, which is the sun-normalised radiance multiplied by the ratio of π and cosine of solar zenith angle. All three neural network models share the same input model parameters. Optimising a single neural network model for all three forward model outputs is not necessary; the correlations between the input parameters and the different forward model outputs are different, which can complicate the optimisation of a general-purpose

neural network. This paper, however, acknowledges modern developments in neural network optimisation techniques that now afford selectively optimising a neural network for different tasks (Kirkpatrick et al., 2016; Wen and Itti, 2018).

The models are trained using the python Tensorflow module (Abadi et al., 2015), and further implemented into an operational processor using C++ interface to Tensorflow. These neural network models require training data containing Disamar input and output parameters and a connecting architecture that encompasses the input feature vector containing scene-varying model parameters, the number of hidden layers, number of nodes in each hidden layer, and an activation function that maps the input to the final output layer containing Disamar outputs. In Tensorflow, the derivative of Δ with respect to the weights are computed using reverse-mode automatic differentiation~~which is a powerful algorithm that~~ , which computes numerical values of derivatives without the use of analytical expressions (Wengert, 1964).

The inputs for NN are referred together as the feature vector. The choice of the parameters included into the feature vector is a very important factor deciding the performance of the neural network. The primary classes of model parameters (relevant to retrieving z_{aer}) varying from scene to scene are solar-satellite geometry, aerosol parameters, meteorological parameters and surface parameters (Table 2). The various aerosol parameters that are fixed from scene to scene are the aerosol single scattering albedo (ω), the asymmetry factor of the phase function, and the angstrom exponent, as they are also fixed in the line-by-line operational aerosol layer height processor. The scattering phase function of aerosols is currently limited to a Henyey-Greenstein model with a fixed g value of 0.7 to mimic Disamar. Surface pressure as well as the temperature-pressure profile are two important meteorological parameters relevant to retrieving z_{aer}. A difference between Disamar and NN models is the definition of this temperature information in the input. Disamar requires the entire temperature-pressure profile of the atmosphere, whereas NN only uses the temperature at z_{aer}. Surface albedo is specified at 758 nm as well as 772 nm in Disamar, whereas it is only specified at 758 nm in the feature vector of NN. In general there is a greater scope to add detailed information in Disamar~~, whereas the goal of NN is to optimally limit input model parameters while accurately calculating forward model outputs~~. 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.

3.2 Training the neural networks

Since the NN forward model is specifically designed for TROPOMI, the solar-satellite geometry is selected to represent TROPOMI orbits for the training data. Meteorological parameters for the locations associated with these solar-satellite geometries are derived from the 2017 60-layer ERA-Interim Reanalysis data (Dee et al., 2011), and aerosol and surface parameters are randomly generated within their physical boundaries. This training data generation strategy spans the entire set of TROPOMI solar and viewing angles as well as meteorological parameters.

Generally, the required training data size increases with increasing non-linearity between input ~~an~~ and output layers in a neural network — there ~~isn't a~~ is no specific method to accurately determine the required sample size before training. ~~Following testing and scrutinizing forward modelcalculation accuracy~~The number of spectra generated for the training set was determined by training different models with different number of spectra in the training set ranging from 1,000 to 600,000. In general it

Table 2. Scene-dependent input model parameters for the NN model. See also Figure 1 for a histogram of the input parameters. The solar-satellite geometry parameters are generated in combinations conforming to the ones encountered by TROPOMI's orbits.

Parameter class	Model Parameters	Remarks	limits
Geometry	Solar zenith angle (θ_0)	in feature vector	8.20° - 80.0°
	Viewing zenith angle (θ)	in feature vector	0.0° - 66.60°
	Solar azimuth angle (ϕ_0)	in feature vector	-180.0° - 180.0°
	Viewing azimuth angle (ϕ)	in feature vector	-180.0° - 180.0°
Aerosol parameters	Aerosol pixel fraction	fixed	1.0
	Single scattering albedo (ω)	fixed	0.95
	Aerosol optical thickness (τ)	in feature vector	0.05 - 5.0
	Aerosol layer height (z_{aer})	in feature vector	75 hPa - 1000.0 hPa
	Aerosol layer thickness (p_{thick})	varied but excluded from feature vector	50 hPa - 200 hPa
	Scattering phase function	fixed	Henyey-Greenstein
	asymmetry factor (g)	fixed	0.7
	Angstrom exponent (Å)	fixed	0.0
Meteorological parameters	Temperature	in feature vector	temperature at z_{aer}
Surface parameters	Surface pressure (p_s)	in feature vector	520 hPa - 1048.50 hPa
	Surface reflectance model	LER	
	Surface albedo (A_s)	in feature vector	2.08E-7 - 0.70

was observed that incorporating more data resulted in a better neural network model. In order to test the trained neural network model, a choice of 500,000 ~~Disamar generated spectra is finalised as the size~~ spectra were selected. 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. These spectra were generated using Disamar with model parameter ranges described in Table 2 and Figure 1. The generation of this training data set is by far the most time consuming step since each Disamar run requires between 50-60 seconds to generate the synthetic spectra. Once the data has been generated, it is prepared for training the neural network models in NN. This is done by data normalisation, achieved by subtracting the mean of each of the training input and output parameters and dividing the difference by its standard deviation, which makes the learning process quicker by reducing the search space for the optimizer. The offset and scaling parameters are important, as the neural network computes outputs within this scaled range, which needs to be re-scaled back to ~~legible~~ physical values. This training requires a few hours on an Intel(R) Xeon(R) CPU E3-1275 v5 at a clock speed of 3.60 GHz.

The most optimal configurations for each of the three NN models are determined by the number of hidden layers, the number of nodes on each layer and the chosen activation function for which the discrepancy between the modeled output for specific inputs and the truth (derived from Disamar) is minimal. ~~Finding the most optimal neural network configuration requires a test~~

~~data set which in this case contains 100,000 scenes outside the training data set. These test data follow the same input model parameter distributions as described in Figure 1 and Table 1.~~ The difference between the outputs calculated by Disamar and NN for these three models provide insight on their performance. ~~The sigmoid function is chosen as the activation function for the NN processor,as it performs the best (lowest loss function value) over other alternatives.~~

~~For each of the neural network models , five configurations were tested. The first three configurations comprise of a single hidden layer,~~ 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 compared 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 ~~and three hidden layers, all consisting of 50 nodes each. Depending onthe best performing configuration of the number of hidden layers, two other configurations are added containing~~ (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 $\mathrm{NN}_{\mathrm{disamar}}(\mathrm{K}_\tau)$ 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 ~~in~~ 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 ~~. For instance, if the neural network configuration comprising of two hidden layers performs best, the last two configurations will consist of two hidden layers with 100 and 200 nodes on each layer . Each configurationwere trained for a total of 25, 000 iterations. Of every configuration~~ has a compromise between summed training loss and simplicity (Figure 2b), especially for $\mathrm{NN}_{\mathrm{disamar}}(\mathrm{K}_\tau)$. 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,~~the most optimal configuration was found to be two hidden layers containing 100 nodes each~~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.

The finalised configurations were then trained for one million iterations after which they were applied to the test data set to study prediction errors. An error analysis revealed that the trained neural networks were ~~generally~~ capable of calculating Disamar outputs with low errors, generally within 1-3% ~~to~~ of Disamar calculations. Averaged convolved errors of the neural network model for the sun normalised radiance (NN_I) did not exceed 1%. The neural network model for the derivative of the reflectance with respect to τ ~~(NN_{K_τ}) performed very well with errors not exceeding more than~~ and z_{aer} perform well in general for parts of the spectrum with large oxygen absorption cross sections, where the value of the derivatives are high (indicating a higher amount of information content from those specific wavelength regions). Errors in the deepest part of the R-branch between 759 nm and 762 nm and the P-branch between 752.50 nm and 765 nm, do not exceed 3~~%. Averaged convolved errors for the neural network model for the derivative of the reflectance with respect to z_{aer} ($\mathrm{NN}_{K_{z_{\mathrm{aer}}}}$)also show good agreements, with errors in~~ % for $\mathrm{NN}_{\mathrm{K}_{\mathrm{zaer}}}$. The same can be said for NN_{K_τ}, which displays errors in the range of 1% in the same wavelength region. For wavelengths outside of the deepest parts of the ~~spectrum with very low z_{aer} information, e.g.~~

~~the continuum (4d). It is important to note that although~~ R and P-branch, the relative errors are large, and exceed 10% easily. However, the relative errors ~~for the derivatives appear quite large at parts of the oxygen A-band spectrum , these parts have low aerosol information content due to low oxygen absorption cross sections (with respect to parts of the wavelength band with stronger oxygen absorption, i.e. the R-branch between 759 nm and 762 nm).~~ are calculated as the absolute value of the difference between the true spectrum and the neural network calculated spectrum, divided by the true spectrum. These values can be very large when the value of the true spectrum is very small, which is the case for the derivatives outside the deepest part of the R and P branches. The consequence of these errors in a retrieval scenario from synthetic and real spectra are discussed in the following section.

4 Comparison between Disamar and NN aerosol layer height retrieval algorithms

To test the NN augmented retrieval algorithm, we apply the generated NN models to synthetic test data and real data from TROPOMI, and compare its retrieval capabilities to those of Disamar. The synthetic data were produced using the Disamar radiative transfer model because of which we expect the online radiative transfer retrievals to be generally better than the NN-based retrievals. The aerosol model used in the retrieval is as in Section 2.2, using fixed parameters for aerosol single scattering albedo, aerosol layer thickness and aerosol scattering phase function.

4.1 Performance of NN versus Disamar in retrieving aerosol layer height in the presence of model errors

A comparison of biases (in the presence of model errors) in the final retrieved solution is indicative of the efficacy of NN in replacing Disamar to retrieve ALH. To directly compare z_{aer} retrieval capabilities of Disamar and NN, radiance and irradiance spectra convolved with a TROPOMI slit function were generated to replicate TROPOMI-measured spectra. Bias is defined as the difference between retrieved and true aerosol layer height (i.e., retrieved - true). A total of 2000 scenes for four synthetic experiments were generated from the test data set containing TROPOMI geometries, with randomly varied model errors in aerosol single scattering albedo, Henyey-Greenstein phase function asymmetery parameter, and surface albedo (described in Table 3).

The retrieved aerosol layer heights from Disamar and NN in the presence of model errors in aerosol layer thickness were found to be almost similar (Figure 5a), with a Pearson correlation coefficient close to 1.0. Introducing model errors in other aerosol properties such as single scattering albedo (Figure 5b) and scattering phase function (Figure 5c) also resulted in a similar agreement between Disamar and NN retrieved aerosol layer heights. Furthermore, both methods retrieved similar aerosol layer heights in the presence of model errors in surface albedo as well (Figure 5d).

A total of 5558 retrievals out of the 8000 difference cases converged to a final solution. On average, z_{aer} retrieved using NN differed by approximately 5.0 hPa from the same using Disamar (Figure 6), with a median of approximately 2.0 hPa. The spread of the retrieval differences were minimal, with a majority of the retrievals differing by less than 13.0 hPa~~approximately~~. Differences close to and above 100.0 hPa did exist, but such retrievals were very uncommon.

Out of the 8000 scenes within the synthetic experiment, NN retrieved aerosol layer heights for 546 scenes where Disamar did not. Contrariwise, 586 scenes converged for Disamar and not for NN. A comparison of the biases from these odd retrieval results indicate that retrievals from NN in cases where Disamar fails are realistic, as the distribution of the biases is very similar to those cases when Disamar succeeds and NN does not (Figure 7). Retrievals using the NN forward model on average required three more iterations to reach a solution when compared to the same by Disamar. Similarly, retrievals from Disamar had a significantly lower minimised cost function (less ~~than~~ by four orders of magnitude on average) at the end of the retrieval when compared to NN. This is within expectation as NN cannot truly replicate Disamar. Having tested the NN augmented retrieval algorithm in a synthetic environment, the retrieval algorithm was installed into the operational TROPOMI processor for testing with real data.

Table 3. A count of converged ~~an~~ and non-converged results from synthetic experiments comparing retrieved aerosol layer heights between Disamar and NN.

	experiment		Disamar		NN	
model parameter	value in sim	value in ret	converged	non converged	converged	non converged
p_{thick}	200 hPa	50 ha	1641	359	1550	450
ω	0.93 - 0.96	0.95	1396	604	1412	588
g	0.67 - 0.73	0.7	1571	429	1567	433
A_s	$0.95A_s$ - $1.05A_s$	A_s	1536	464	1575	425

4.2 Application to December 2017 Californian forest fires observed by TROPOMI

The December 2017 Southern California wildfires have been attributed to very low humidity levels, following delayed autumn precipitation and severe multi-annual drought (Nauslar et al., 2018). Particularly on December 12, the region of the fires ~~were~~ was cloud-free, owing to high-pressure conditions. The biomass burning plume extended well beyond the coastline and over the ocean, which provides a roughly cloud-free and low surface brightness test case for implementing the aerosol layer height retrieval algorithm (Figure 8a). The absorbing aerosol index values were above 5.0 in the bulk of the plume, indicating a very high concentration of elevated absorbing aerosols. Pixels with an AAI value less than 1.0 were excluded from the retrieval experiment. Pixels that were cloud contaminated were removed from the processing chain using the FRESCO cloud mask product from TROPOMI (maximum cloud fraction of 0.2), but parts of the biomass burning plume that did not contain any clouds (Figure 8b) were also removed, as the cloud fraction values for these pixels were higher than the threshold. The retrieval algorithms did not process pixels in the coastline, as the surface albedo values could be incorrect in these regions.

The operational line-by-line algorithm was applied to ground pixels within a bounding box around the plume. A total of 7418 pixels within this bounding box converged to a solution (Figure 9a). The neural network augmented operational processor retrieved 7370 pixels out of the 7418 pixels that had converged for the operational line-by-line processor (Figure 9b). Although visually discernable in the difference map in Figure 9c, the retrieved z_{aer} from both algorithms were quite similar (Figure 10a).

The neural network augmented processor retrieved aerosol layer heights which were (on average) less than 50.0 meters apart from the same by the line-by-line counterpart (Figure 10b). While the standard deviation of approximately 160 meters indicates the presence of outliers, the 15^{th} and the 85^{th} percentile values of -115.0 meters and 40.0 meters, respectively, indicate that the significant majority of retrieved pixels were only off by less than 100.0 meters. Although the retrieval algorithms have good ~~agreements~~agreement, they primarily ~~departed in~~ differed for the lower aerosol loading scenes (Table 4). The majority of the pixels where the neural network algorithm differed from the line-by-line counterpart by more than 200 meters were for absorbing aerosol index values less than 2.0 (Figure 10c). Most of these biases were due to over-estimation by the neural network retrieval algorithm. Pixels with AAI values larger than 5.0 also showed a consistent ~~departure, different~~ bias, differing on average by 60 meters with a standard deviation of 30 meters. This ~~departure~~ bias is not well understood.

Table 4. Statistics of difference between retrieved z_{aer} from ~~disamar~~ Disamar and NN, as defined in figure 9c.

AAI [-]	number of samples	mean [m]	median [m]	standard deviation [m]	15^{th} percentile [m]	85^{th} percentile [m]
<2.0	3227	-50.74	-62.10	206.44	-228.65	108.31
2.0 - 3.0	2723	-54.96	-43.20	110.75	-184.85	67.10
3.0 - 5.0	1167	10.32	19.42	63.65	-61.63	65.26
>5.0	253	61.35	61.00	30.954	26.56	95.22

The time required by the line-by-line operational processor was 184.01 ± 0.50 seconds per pixel, whereas the same for the neural network processor was 0.167 ± 0.0003 seconds per pixel. The neural network algorithm shows an improvement in the computational speed by three orders of magnitude over the line-by-line retrieval algorithm. The computational speed gained from implementing NN enables retrieval of aerosol layer heights from all potential scenes in the entire orbit within the stipulated operational processing time slot.

5 Conclusions

Of the algorithms that currently retrieve TROPOMI's suite of level-2 products, the aerosol layer height processor requires online radiative transfer calculations. These online calculations have traditionally been tackled with KNMI's radiative transfer code Disamar, which calculates sun-normalised radiances in the oxygen A-band. There are, in total, 3980 line-by-line calculations per iteration in the optimal estimation scheme, requiring several minutes to retrieve aerosol layer height estimates from a single scene. This limits the yield of the aerosol layer height processor significantly.

The bottleneck is identified to be the number of calculations Disamar needs to do at every iteration of the Gauss-Newton scheme of the estimation process. As a replacement, this paper proposes using artificial neural networks in the forward model step. Three neural networks are trained, for the sun-normalised radiance and the derivative of the reflectance with respect to aerosol layer height and aerosol optical thickness, the two state vector elements. As the goal is to replicate and replace Disamar, line-by-line forward model calculations from Disamar were used to train these neural networks. A total of 500,000 spectra were generated using Disamar, and each of the neural network models were trained for a total of 1 million iterations with the mean

squared error between the training data output and the neural network output being the cost function to be minimised in the optimisation process.

Over a test data set with 100,000 different scenes unique from the training data set, the neural network models performed well, with errors not exceeding 1-3% in general in the predicted spectra and derivatives. Having tested the neural network models for prediction errors in the forward model output spectra, they were implemented into the aerosol layer height breadboard algorithm and further tested for retrieval accuracy. In order to do so, experiments with synthetic as well as real data were conducted. The synthetic scenes included 2000 spectra with different model errors in aerosol and surface properties. In these cases, the neural network algorithm showed very good compatibility with the aerosol layer height algorithm, since it was able to replicate the biases satisfactorily.

For a real test case, TROPOMI spectra over the December 12, 2017 forest fires in Southern California were chosen. On this day, the biomass burning plume extended from land to the ocean over a dry and almost cloudless scene. Operational retrievals using both Disamar and the neural network forward models showed very similar results, with a few outliers around 500 meters for pixels containing low aerosol loads. These biases were outweighed by the upgrade in the computational speed of the retrieval algorithm, as the neural network augmented processor observed a speedup of three orders of magnitude, making the aerosol layer height processor operationally feasible. Having achieved this improvement in its computational performance, the aerosol layer height algorithm is planned to be operationally retrieving the product for the all possible pixels in each orbit of TROPOMI. Such a boost in processor output allows for better analyses of retrievals and opens the possibility to remove some of the forward model simplifications mentioned in Section 2.2, which paves the way for further developing the TROPOMI aerosol layer height algorithm.

Competing interests. The author declares no conflict of interests in the work expressed in this publication.

Acknowledgements. This publication contains modified Copernicus Sentinel data. This research is partly funded by the European Space Agency (ESA) within the EU Copernicus programme.

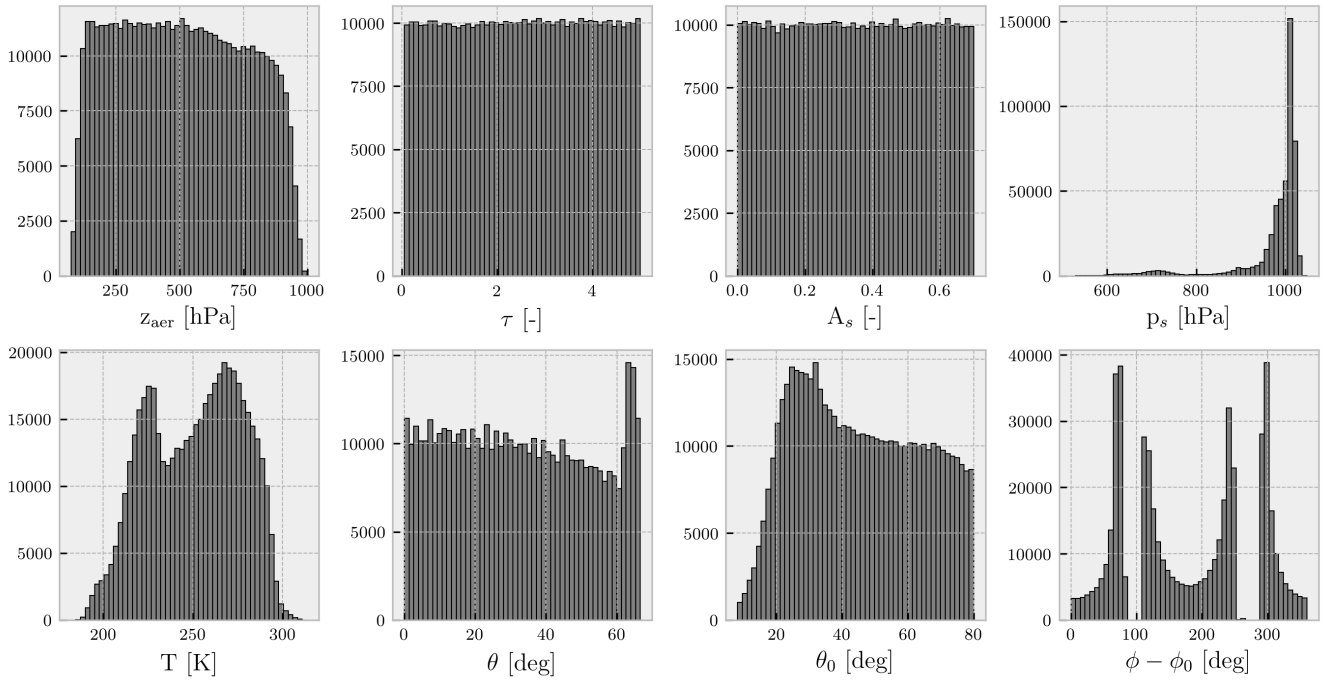

Figure 1. Histograms of the various input parameters for each of the neural network models in NN. Minimum and maximum values for each of the parameters are ~~available~~ shown in Table 2.

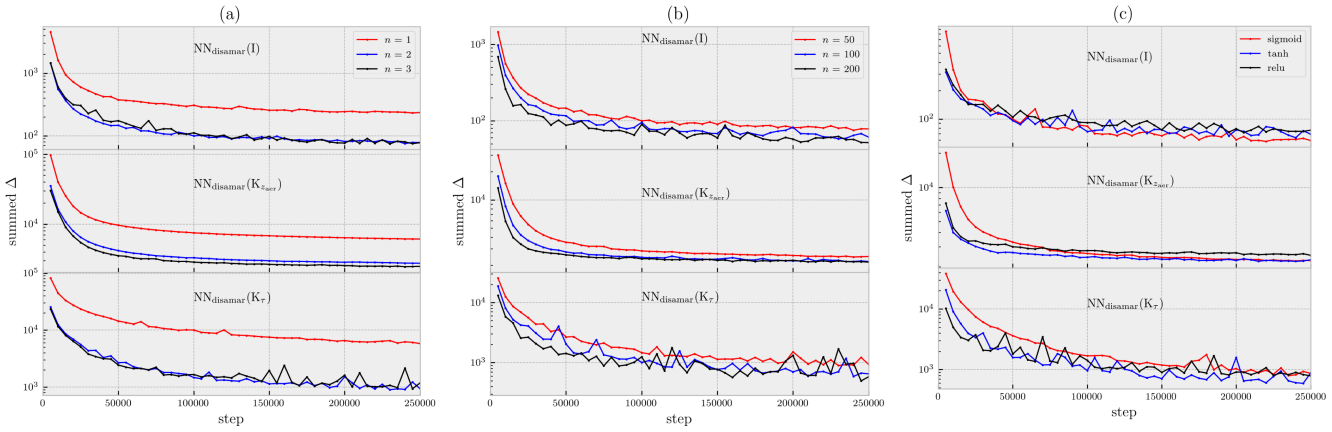

Figure 2. Summed loss as a function of training step for different neural network model configurations. **(a)** The neural network models have 50 nodes per each layer with a sigmoid activation function. **(b)** The neural network models have two hidden layers with each node activated by the sigmoid function. **(c)** The neural network models have two hidden layers with a 100 nodes for each layer.

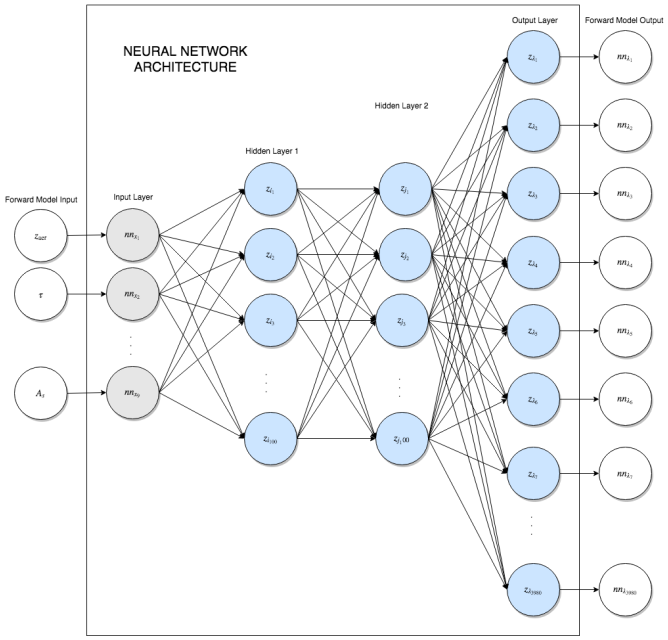

Figure 3. ~~A schematic~~ Schematic of each of the three neural networks in NN. There are two hidden layers, each containing 100 nodes. z represents inputs for each of the nodes, whereas nn represents the inputs and outputs of the neural network.

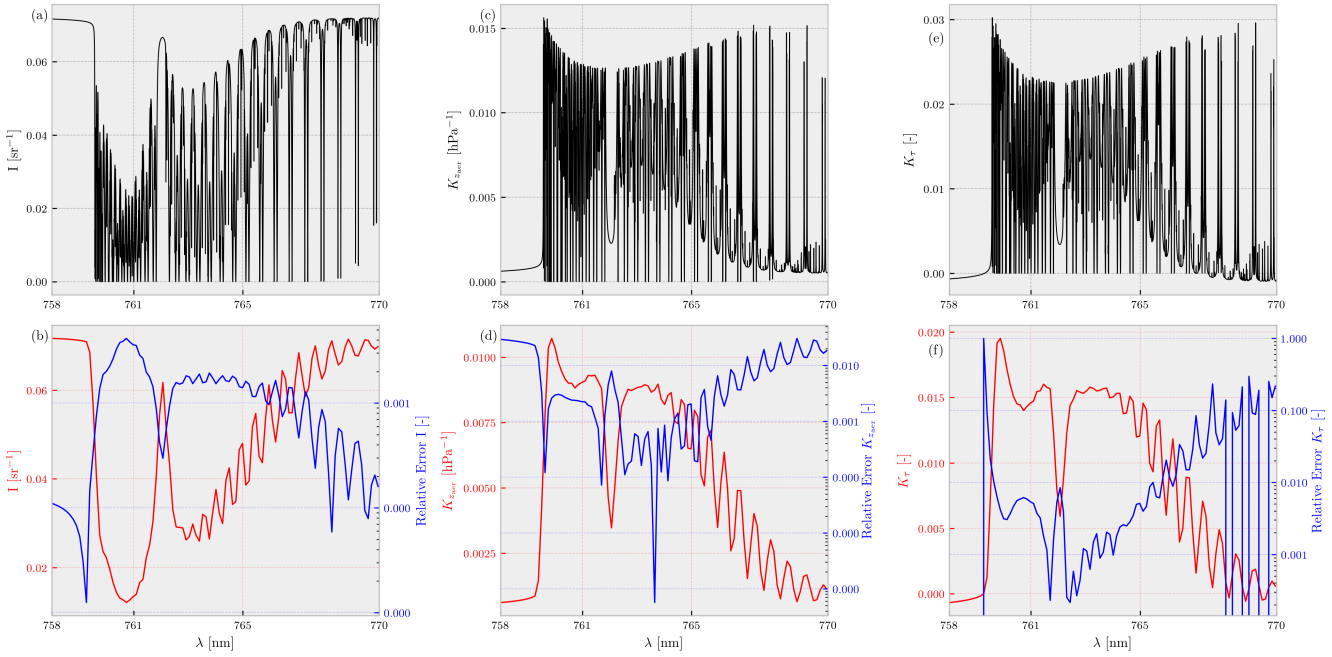

Figure 4. Performance of the finalised neural network. The top row represents the averaged output of each of the neural networks for surface albedo less than 0.4. The bottom row represents the convolved version of the top row (plotted as the red line with the left-handed y-axis) and the convolved relative error (plotted in log scale) with the truth (plotted in blue with the right-handed y-axis). The relative errors are computed as the absolute value of the difference (post-convolution) between the averaged true and averaged predicted spectra, divided by the averaged true spectra. (a,b) represent the neural network computed sun-normalised radiances, (c,d) represent the same for the derivative of reflectance with respect to aerosol layer height, and (e,f) the same with respect to aerosol optical thickness.

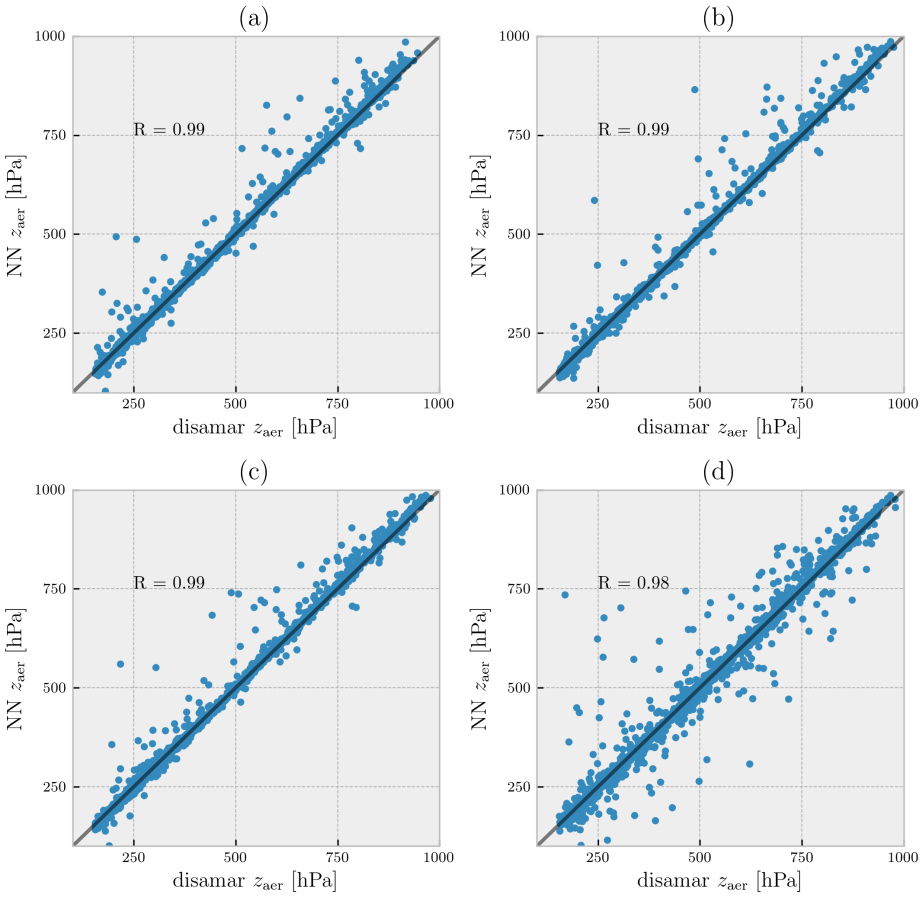

Figure 5. Retrieved layer heights compared between Disamar and NN for 2000 synthetic spectra in the presence of model errors. The dots represent converged scenes only, with the x axis representing retrievals from Disamar and the y-axis representing the same from NN. The model errors represented in this figure are (a) aerosol layer pressure thickness, (b) aerosol single scattering albedo, (c) aerosol scattering phase function asymmetry factor, and (d) surface albedo. These results as well as the introduced model errors are summarised in Table 3. The Pearson correlation coefficient (R) between the retrieved z_{aer} from different methods is mentioned in each of the plots.

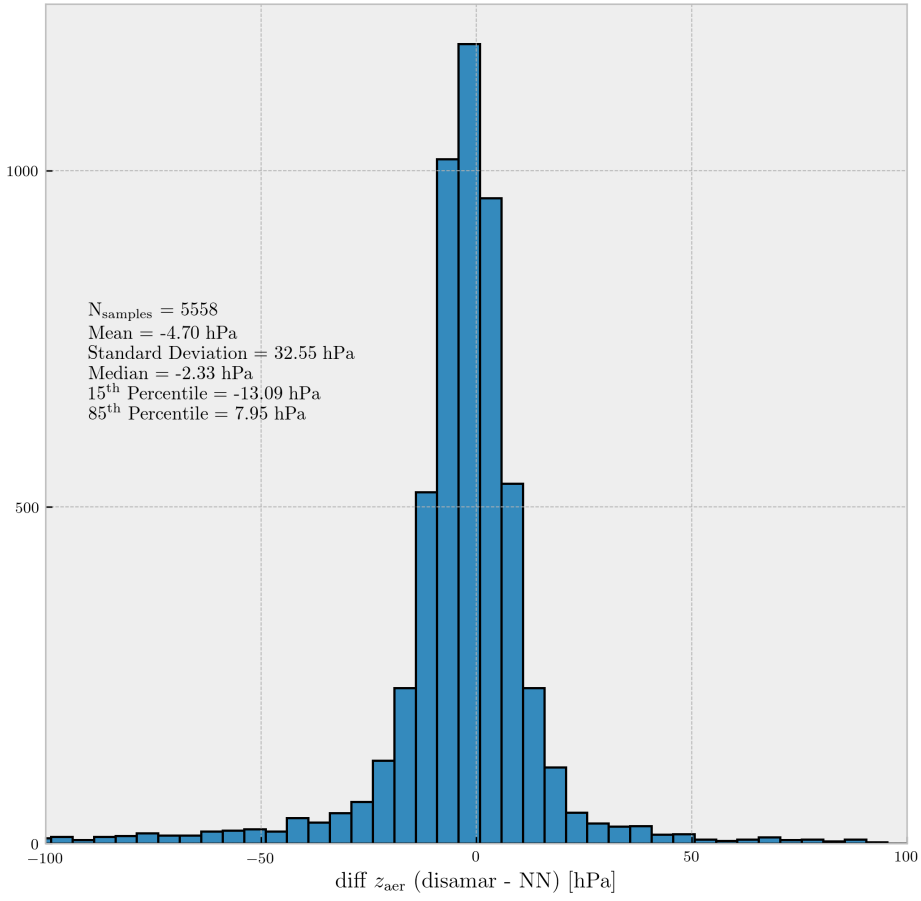

Figure 6. ~~A histogram~~ Histogram of differences between the retrieved z_{aer} values using Disamar and NN retrieval methods for synthetic spectra generated by Disamar. Total number of cases is 8000, whereas the plot contains 5558 retrieved samples for both Disamar and NN; non-converged cases are not included. A map of these differences are ~~plotting~~ plotted in Figure 9c.

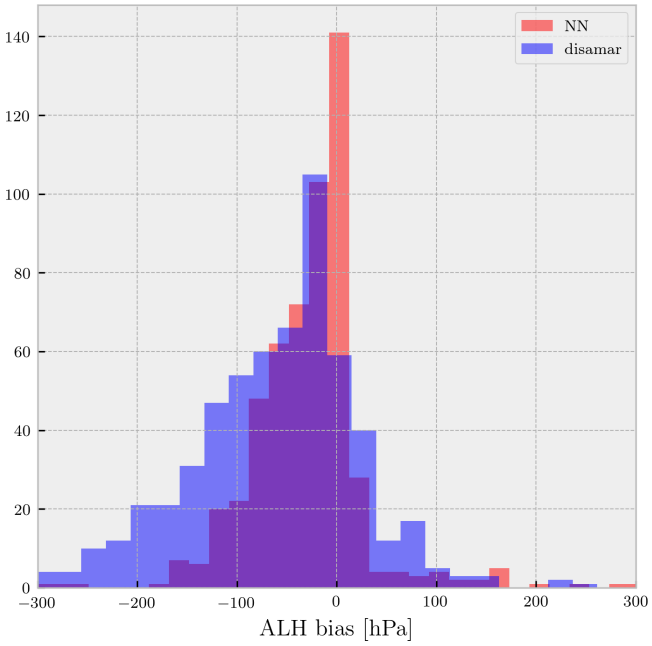

Figure 7. ~~A histogram~~ Histogram of biases (retrieved - true) for scenes in the synthetic experiment for which either NN converges to a solution (red bar plot) and Disamar does not, or Disamar converges to a solution (blue bar plot) whereas NN does not.

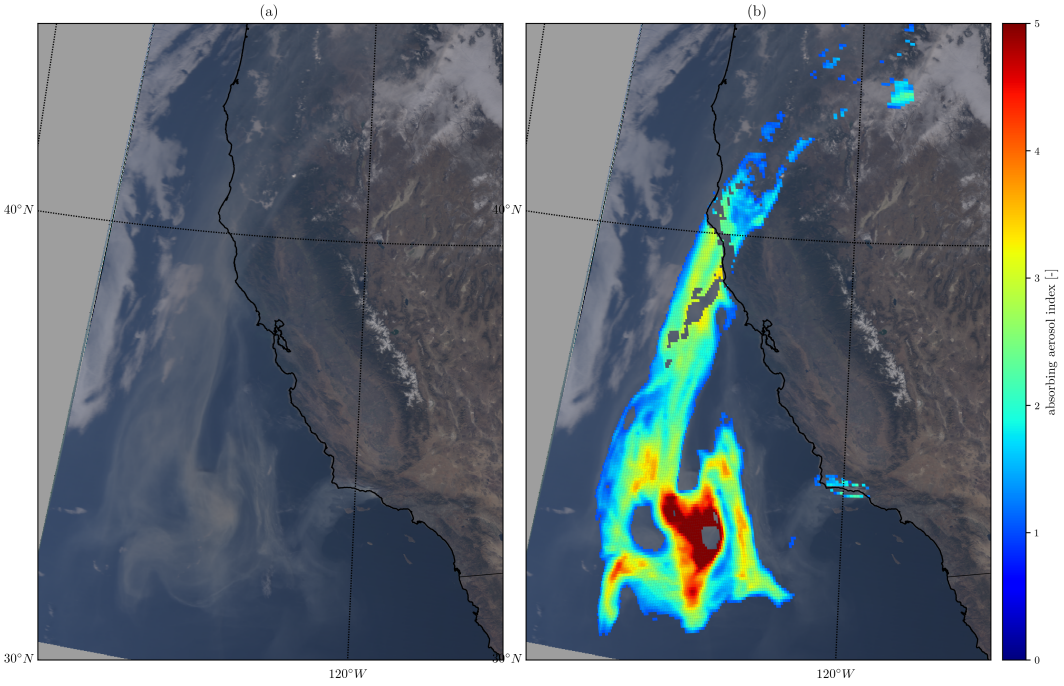

Figure 8. (a) ~~A~~ MODIS Terra image of the December 12, 2017 Southern Californian wildfire plume, extending from land to ~~the~~ ocean. (b) Calculated aerosol absorbing index from the TROPOMI level-2 processor. Missing pixels ~~either~~ are flagged by a cloud mask ~~,~~ or ~~by a~~ land-sea mask, or have an absorbing aerosol index less than 1.0.

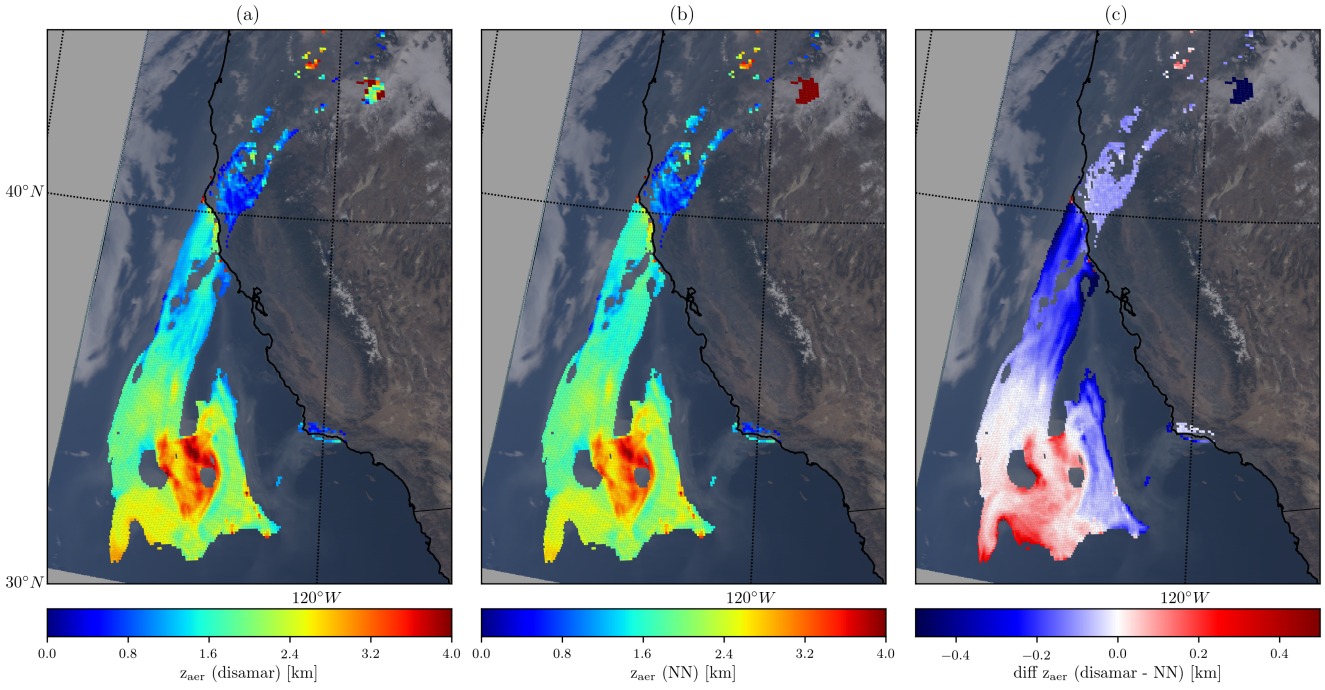

Figure 9. (a) Aerosol layer height retrieved using Disamar as the forward model. (b) The same, but with NN replacing Disamar in the operational processor. (c) ~~represents the~~ difference between Disamar and NN retrieved aerosol layer heights.

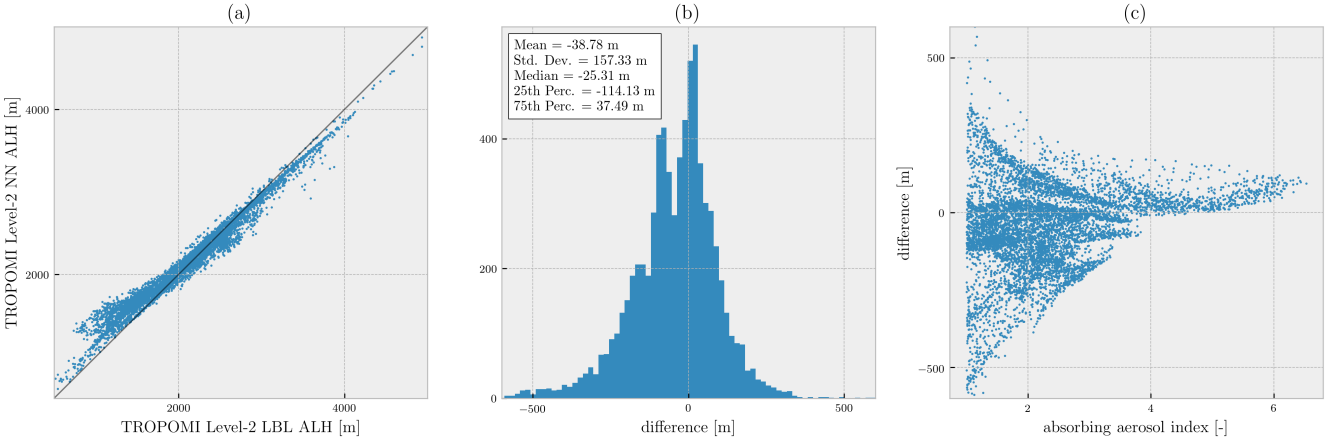

Figure 10. Comparison of retrieved aerosol layer heights from TROPOMI-measured spectra (orbit number 858) for the 12th December, 2017 Southern California fires using Disamar and NN. ~~Figre~~ (a) ~~directly compares retrieved~~ Retrieved aerosol layer heights from the two methods~~. Figure~~ ; (b) ~~provides a histogram~~ Histogram of the difference between ~~these~~ retrieved heights from Disamar and NN. The difference is defined as $z_{\mathrm{aer}}(\mathrm{Disamar})$ - $z_{\mathrm{aer}}(\mathrm{NN})$. ~~Figure~~ (c) ~~compares these differences with~~ Differences compared to TROPOMI~~'~~'s operational absorbing aerosol index product (x axis).

References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J., Mané, D., Monga, R., Moore, S., Murray, D., Olah, C., Schuster, M., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viégas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y., and Zheng, X.: TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems, https://www.tensorflow.org/, software available from tensorflow.org, 2015.

Chance, K. and Kurucz, R.: An improved high-resolution solar reference spectrum for earth's atmosphere measurements in the ultraviolet, visible, and near infrared, Journal of Quantitative Spectroscopy and Radiative Transfer, 111, 1289–1295, https://doi.org/10.1016/j.jqsrt.2010.01.036, http://linkinghub.elsevier.com/retrieve/pii/S0022407310000610, 2010.

Chimot, J., Veefkind, J. P., Vlemmix, T., de Haan, J. F., Amiridis, V., Proestakis, E., Marinou, E., and Levelt, P. F.: An exploratory study on the aerosol layer height retrieval from OMI measurements of the 477 nm O2-O2 spectral band using a neural network approach, Atmos. Meas. Tech., 10, 783–809, https://doi.org/10.5194/amt-10-783-2017, https://www.atmos-meas-tech.net/10/783/2017/, 2017.

Chimot, J., Veefkind, J. P., Vlemmix, T., and Levelt, P. F.: Spatial distribution analysis of the OMI aerosol layer height: a pixel-by-pixel comparison to CALIOP observations, Atmos. Meas. Tech., 11, 2257–2277, https://doi.org/10.5194/amt-11-2257-2018, https://www.atmos-meas-tech.net/11/2257/2018/, 2018.

Colosimo, S. F., Natraj, V., Sander, S. P., and Stutz, J.: A sensitivity study on the retrieval of aerosol vertical profiles using the oxygen A-band, Atmospheric Measurement Techniques, 9, 1889–1905, https://doi.org/10.5194/amt-9-1889-2016, https://www.atmos-meas-tech.net/9/1889/2016/, 2016.

Corradini, S. and Cervino, M.: Aerosol extinction coefficient profile retrieval in the oxygen A-band considering multiple scattering atmosphere. Test case: SCIAMACHY nadir simulated measurements, Journal of Quantitative Spectroscopy and Radiative Transfer, 97, 354–380, https://doi.org/10.1016/j.jqsrt.2005.05.061, http://www.sciencedirect.com/science/article/pii/S0022407305002207, 2006.

Davis, A. B., Kalashnikova, O. V., and Diner, D. J.: Aerosol Layer Height over Water from O2 A-Band: Mono-Angle Hyperspectral and/or Bi-Spectral Multi-Angle Observations, 2017.

de Haan, J. F., Bosma, P. B., and Hovenier, J. W.: The adding method for multiple scattering calculations of polarized light, Astronomy and Astrophysics, 183, 1987.

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N., and Vitart, F.: The ERA-Interim reanalysis: configuration and performance of the data assimilation system, Quarterly Journal of the Royal Meteorological Society, 137, 553–597, https://doi.org/10.1002/qj.828, http://doi.wiley.com/10.1002/qj.828, 2011.

Dubuisson, P., Frouin, R., Dessailly, D., Duforêt, L., Léon, J.-F., Voss, K., and Antoine, D.: Estimating the altitude of aerosol plumes over the ocean from reflectance ratio measurements in the O2 A-band, Remote Sensing of Environment, 113, 1899–1911, https://doi.org/10.1016/j.rse.2009.04.018, http://www.sciencedirect.com/science/article/pii/S0034425709001333, 2009.

Frankenberg, C., Hasekamp, O., O'Dell, C., Sanghavi, S., Butz, A., and Worden, J.: Aerosol information content analysis of multi-angle high spectral resolution measurements and its benefit for high accuracy greenhouse gas retrievals, Atmos. Meas. Tech., 5, 1809–1821, https://doi.org/10.5194/amt-5-1809-2012, https://www.atmos-meas-tech.net/5/1809/2012/, 2012.

Gabella, M., Kisselev, V., and Perona, G.: Retrieval of aerosol profile variations from reflected radiation in the oxygen absorption A band, Applied Optics, 38, 3190–3195, https://doi.org/10.1364/AO.38.003190, https://www.osapublishing.org/abstract.cfm?uri=ao-38-15-3190, 1999.

Geddes, A. and Bösch, H.: Tropospheric aerosol profile information from high-resolution oxygen A-band measurements from space, Atmospheric Measurement Techniques, 8, 859–874, https://doi.org/10.5194/amt-8-859-2015, https://www.atmos-meas-tech.net/8/859/2015/, 2015.

Hasekamp, O. P. and Butz, A.: Efficient calculation of intensity and polarization spectra in vertically inhomogeneous scattering and absorbing atmospheres, Journal of Geophysical Research: Atmospheres, 113, D20 309, https://doi.org/10.1029/2008JD010379, http://onlinelibrary.wiley.com/doi/10.1029/2008JD010379/abstract, 2008.

Henyey, L. C. and Greenstein, J. L.: Diffuse radiation in the Galaxy, The Astrophysical Journal, 93, 70, https://doi.org/10.1086/144246, http://adsabs.harvard.edu/doi/10.1086/144246, 1941.

Hollstein, A. and Fischer, J.: Retrieving aerosol height from the oxygen A band: a fast forward operator and sensitivity study concerning spectral resolution, instrumental noise, and surface inhomogeneity, Atmospheric Measurement Techniques, 7, 1429–1441, https://doi.org/10.5194/amt-7-1429-2014, http://www.atmos-meas-tech.net/7/1429/2014/, 2014.

Kingma, D. P. and Ba, J.: Adam: A Method for Stochastic Optimization, arXiv:1412.6980 [cs], http://arxiv.org/abs/1412.6980, arXiv: 1412.6980, 2014.

Kirkpatrick, J., Pascanu, R., Rabinowitz, N., Veness, J., Desjardins, G., Rusu, A. A., Milan, K., Quan, J., Ramalho, T., Grabska-Barwinska, A., Hassabis, D., Clopath, C., Kumaran, D., and Hadsell, R.: Overcoming catastrophic forgetting in neural networks, arXiv:1612.00796 [cs, stat], http://arxiv.org/abs/1612.00796, arXiv: 1612.00796, 2016.

Landgraf, J., Hasekamp, O. P., Box, M. A., and Trautmann, T.: A linearized radiative transfer model for ozone profile retrieval using the analytical forward-adjoint perturbation theory approach, Journal of Geophysical Research: Atmospheres, 106, 27 291–27 305, https://doi.org/10.1029/2001JD000636, http://doi.wiley.com/10.1029/2001JD000636, 2001.

Levelt, P. F., Oord, G. H. J. v. d., Dobber, M. R., Malkki, A., Visser, H., Vries, J. d., Stammes, P., Lundell, J. O. V., and Saari, H.: The ozone monitoring instrument, IEEE Transactions on Geoscience and Remote Sensing, 44, 1093–1101, https://doi.org/10.1109/TGRS.2006.872333, 2006.

Loyola, D. G., Gimeno García, S., Lutz, R., Argyrouli, A., Romahn, F., Spurr, R. J. D., Pedergnana, M., Doicu, A., Molina García, V., and Schüssler, O.: The operational cloud retrieval algorithms from TROPOMI on board Sentinel-5 Precursor, Atmospheric Measurement Techniques, 11, 409–427, https://doi.org/10.5194/amt-11-409-2018, https://www.atmos-meas-tech.net/11/409/2018/, 2018.

Loyola, D. G. R.: Automatic cloud analysis from polar-orbiting satellites using neural network and data fusion techniques, in: IGARSS 2004. 2004 IEEE International Geoscience and Remote Sensing Symposium, vol. 4, pp. 2530–2533 vol.4, https://doi.org/10.1109/IGARSS.2004.1369811, 2004.

Nanda, S., de Graaf, M., Sneep, M., de Haan, J. F., Stammes, P., Sanders, A. F. J., Tuinder, O., Veefkind, J. P., and Levelt, P. F.: Error sources in the retrieval of aerosol information over bright surfaces from satellite measurements in the oxygen A band, Atmos. Meas. Tech., 11, 161–175, https://doi.org/10.5194/amt-11-161-2018, https://www.atmos-meas-tech.net/11/161/2018/, 2018a.

Nanda, S., Veefkind, J. P., de Graaf, M., Sneep, M., Stammes, P., de Haan, J. F., Sanders, A. F. J., Apituley, A., Tuinder, O., and Levelt, P. F.: A weighted least squares approach to retrieve aerosol layer height over bright surfaces applied to GOME-2 measurements of the oxygen A band for forest fire cases over Europe, Atmos. Meas. Tech., 11, 3263–3280, https://doi.org/10.5194/amt-11-3263-2018, https://www.atmos-meas-tech.net/11/3263/2018/, 2018b.

Nauslar, N. J., Abatzoglou, J. T., and Marsh, P. T.: The 2017 North Bay and Southern California Fires: A Case Study, Fire, 1, 18, https://doi.org/10.3390/fire1010018, https://www.mdpi.com/2571-6255/1/1/18, 2018.

Pelletier, B., Frouin, R., and Dubuisson, P.: Retrieval of the aerosol vertical distribution from atmospheric radiance, vol. 7150, p. 71501R, International Society for Optics and Photonics, https://doi.org/10.1117/12.806527, https://www.spiedigitallibrary.org/conference-proceedings-of-spie/7150/71501R/Retrieval-of-the-aerosol-vertical-distribution-from-atmospheric-radiance/10.1117/12.806527.short, 2008.

Rodgers, C. D.: Inverse methods for atmospheric sounding: theory and practice, vol. 2, World Scientific, 2000.

Sanders, A. F. J. and de Haan, J. F.: Retrieval of aerosol parameters from the oxygen A band in the presence of chlorophyll fluorescence, Atmospheric Measurement Techniques, 6, 2725–2740, https://doi.org/10.5194/amt-6-2725-2013, http://www.atmos-meas-tech.net/6/2725/2013/, 2013.

Sanders, A. F. J. and de Haan, J. F.: TROPOMI ATBD of the Aerosol Layer Height product, http://www.tropomi.eu/sites/default/files/files/S5P-KNMI-L2-0006-RP-TROPOMI_ATBD_Aerosol_Height-v1p0p0-20160129.pdf, 2016.

Sanders, A. F. J., de Haan, J. F., Sneep, M., Apituley, A., Stammes, P., Vieitez, M. O., Tilstra, L. G., Tuinder, O. N. E., Koning, C. E., and Veefkind, J. P.: Evaluation of the operational Aerosol Layer Height retrieval algorithm for Sentinel-5 Precursor: application to Oxygen A band observations from GOME-2A, Atmospheric Measurement Techniques, 8, 4947–4977, https://doi.org/10.5194/amt-8-4947-2015, http://www.atmos-meas-tech.net/8/4947/2015/, 2015.

Sanghavi, S., Martonchik, J. V., Landgraf, J., and Platt, U.: Retrieval of the optical depth and vertical distribution of particulate scatterers in the atmosphere using O_2 A- and B-band SCIAMACHY observations over Kanpur: a case study, Atmospheric Measurement Techniques, 5, 1099–1119, https://doi.org/10.5194/amt-5-1099-2012, http://www.atmos-meas-tech.net/5/1099/2012/, 2012.

Sioris, C. E. and Evans, W. F. J.: Impact of rotational Raman scattering in the O2A band, Geophysical Research Letters, 27, 4085–4088, https://doi.org/10.1029/2000GL012231, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2000GL012231, 2000.

Tilstra, L. G., Tuinder, O. N. E., Wang, P., and Stammes, P.: Surface reflectivity climatologies from UV to NIR determined from Earth observations by GOME-2 and SCIAMACHY: GOME-2 and SCIAMACHY surface reflectivity climatologies, Journal of Geophysical Research: Atmospheres, https://doi.org/10.1002/2016JD025940, http://doi.wiley.com/10.1002/2016JD025940, 2017.

Timofeyev, Y., Vasilyev, A., and Rozanov, V.: Information content of the spectral measurements of the 0.76 m O2 outgoing radiation with respect to the vertical aerosol optical properties, Advances in Space Research, 16, 91–94, https://doi.org/10.1016/0273-1177(95)00385-R, 1995.

Vasilkov, A., Joiner, J., and Spurr, R.: Note on rotational-Raman scattering in the O_2 A- and B-bands, Atmospheric Measurement Techniques, 6, 981–990, https://doi.org/https://doi.org/10.5194/amt-6-981-2013, https://www.atmos-meas-tech.net/6/981/2013/amt-6-981-2013.html, 2013.

Veefkind, J. P., Aben, I., McMullan, K., Förster, H., de Vries, J., Otter, G., Claas, J., Eskes, H. J., de Haan, J. F., Kleipool, Q., van Weele, M., Hasekamp, O., Hoogeveen, R., Landgraf, J., Snel, R., Tol, P., Ingmann, P., Voors, R., Kruizinga, B., Vink, R., Visser, H., and Levelt, P. F.: TROPOMI on the ESA Sentinel-5 Precursor: A GMES mission for global observations of the atmospheric composition for climate, air quality and ozone layer applications, Remote Sensing of Environment, 120, 70–83, https://doi.org/10.1016/j.rse.2011.09.027, http://www.sciencedirect.com/science/article/pii/S0034425712000661, 2012.

Wang, P., Stammes, P., van der A, R., Pinardi, G., and van Roozendael, M.: FRESCO+: an improved O2 A-band cloud retrieval algorithm for tropospheric trace gas retrievals, Atmos. Chem. Phys., 8, 6565–6576, https://doi.org/10.5194/acp-8-6565-2008, https://www.atmos-chem-phys.net/8/6565/2008/, 2008.

Wang, P., Tuinder, O. N. E., Tilstra, L. G., de Graaf, M., and Stammes, P.: Interpretation of FRESCO cloud retrievals in case of absorbing aerosol events, Atmospheric Chemistry and Physics, 12, 9057–9077, https://doi.org/10.5194/acp-12-9057-2012, http://www.atmos-chem-phys.net/12/9057/2012/, 2012.

Wen, S. and Itti, L.: Overcoming catastrophic forgetting problem by weight consolidation and long-term memory, arXiv:1805.07441 [cs, stat], http://arxiv.org/abs/1805.07441, arXiv: 1805.07441, 2018.

Wengert, R. E.: A Simple Automatic Derivative Evaluation Program, Commun. ACM, 7, 463–464, https://doi.org/10.1145/355586.364791, http://doi.acm.org/10.1145/355586.364791, 1964.

Xu, X., Wang, J., Wang, Y., Zeng, J., Torres, O., Yang, Y., Marshak, A., Reid, J., and Miller, S.: Passive remote sensing of altitude and optical depth of dust plumes using the oxygen A and B bands: First results from EPIC/DSCOVR at Lagrange-1 point, Geophysical Research Letters, 44, 2017GL073 939, https://doi.org/10.1002/2017GL073939, http://onlinelibrary.wiley.com/doi/10.1002/2017GL073939/abstract, 2017.

Zeng, Z.-C., Natraj, V., Xu, F., J. Pongetti, T., Shia, R.-L., A. Kort, E., C. Toon, G., P. Sander, S., and L. Yung, Y.: Constraining Aerosol Vertical Profile in the Boundary Layer Using Hyperspectral Measurements of Oxygen Absorption, Geophysical Research Letters, https://doi.org/10.1029/2018GL079286, 2018.