



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

Swadhin Nanda<sup>1,2</sup>, Martin de Graaf<sup>1</sup>, J. Pepijn Veeffkind<sup>1,2</sup>, Mark ter Linden<sup>3</sup>, Maarten Sneep<sup>1</sup>, Johan de Haan<sup>1</sup>, and Pieternel F. Levelt<sup>1,2</sup>

<sup>1</sup>Royal Netherlands Meteorological Institute (KNMI), Utrechtseweg 297, 3731 GA De Bilt, The Netherlands

<sup>2</sup>Delft university of Technology (TU Delft), Mekelweg 2, 2628 CD Delft, The Netherlands

<sup>3</sup>S&T Corp, Delft, The Netherlands

*Correspondence to:* Swadhin Nanda ([nanda@knmi.nl](mailto: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 TROPOspheric Monitoring Instrument (Veeffkind 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  $z_{\text{aer}}$ ) 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 ( $\tau$ ).

5 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.

10 The ALH retrieval algorithm is computationally expensive, requiring several minutes to compute  $z_{\text{aer}}$  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 percent of all TROPOMI pixels in an orbit over an area as big as Europe may be eligible for retrieving aerosol layer height. This number can go up to as much as 50,000 pixels

15 per orbit. This places 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

20 k-distribution method to reduce the number of calculations Hasekamp and Butz (2008), using a look up table for calculating forward model outputs, or entirely foregoing the forward model and directly retrieving  $z_{\text{aer}}$  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 radiative transfer model as for TROPOMI to generate training data, in combination with the NASA MODIS aerosol optical depth product, and successfully retrieve aerosol layer heights directly from

25 the O<sub>2</sub>-O<sub>2</sub> bands in the visible spectral region at 477 nm. They demonstrated this 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

30 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 of artificial neural networks for retrieving aerosol information from oxygen absorption bands. 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

35 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. 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., 2012; Wang et al., 2012; Hollstein and Fischer, 2014; Sanders and de Haan, 2013; Sanders et al., 2015; Sanders and de Haan, 2016; Nanda et al., 2018b). 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  $\chi^2$  represents the departure of the modeled reflectance  $F(\mathbf{x})$  from the observed reflectance  $\mathbf{y}$  scaled by the measurement error covariance matrix  $\mathbf{S}_\epsilon$ , and is defined as

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

Minimising this cost function for a particular  $z_{\text{aer}}$  and  $\tau$  (the elements of the state vector  $\mathbf{x}$  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  $\mathbf{x}$ , contained in  $\mathbf{x}_a$  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  $z_{\text{aer}}$  and 1.0 for  $\tau$  at 760 nm. The a priori error of the  $z_{\text{aer}}$  is fixed at 500 hPa, and the same for  $\tau$  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  $\mathbf{x}$  following,

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



where  $\mu_0$  is the cosine of the solar zenith angle  $\theta_0$ ,  $I(\lambda)$  for wavelength  $\lambda$  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,

$$\mathbf{x}_{i+1} = \mathbf{x}_a + [\mathbf{K}_i^T \mathbf{S}_\epsilon^{-1} \mathbf{K}_i + \mathbf{S}_a^{-1}]^{-1} \mathbf{K}_i^{-1} \mathbf{S}_\epsilon^{-1} [\mathbf{y} - \mathbf{F}(\mathbf{x}) + \mathbf{K}_i(\mathbf{x}_i - \mathbf{x}_a)], \quad (3)$$

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

$$\hat{\mathbf{S}} = [\mathbf{K}^T \mathbf{S}_\epsilon^{-1} \mathbf{K} + \mathbf{S}_a^{-1}]^{-1}. \quad (4)$$

## 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  $\chi^2$  (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  $\text{O}_2\text{-O}_2$  and  $\text{O}_2\text{-N}_2$ . The total influence of the  $\text{O}_2$  A-band in 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  $z_{\text{aer}}$  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 the Rayleigh optical thickness is small around 760 nm. Calculating the influence of Rotational Raman Scattering (RRS) is also ignored, as it is a computationally expensive step. This exclusion of calculations is not advised by literature (Vasilkov et al., 2013; Sioris and Evans, 2000), as RRS can alter the line depths in the  $\text{O}_2$  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  $z_{\text{aer}}$  retrievals significantly. The atmosphere is assumed cloud-free, which is a required simplification as the retrieval of  $z_{\text{aer}}$  in the presence of clouds becomes challenging. The aerosol fraction is assumed as 1.0, 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  $z_{\text{aer}}$  with a 50 hPa thickness, a fixed aerosol optical thickness ( $\tau$ ) 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 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. Lastly, the atmosphere is spherically-corrected for incoming solar radiation and remains plane-parallel for outgoing Earth radiance.

### 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  $7 \text{ km} \times 3.5 \text{ km}$  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.

Calculation of TOA reflectance and its derivatives with respect to  $z_{\text{aer}}$ , and  $\tau$  in an 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  $z_{\text{aer}}$  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.



**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

### 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  $\Delta$ ), which is synonymous to a cost function (Equation 5),

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

where  $\lambda$  is the wavelength,  $n_\lambda$  represents the number of elements in the output layer,  $nn_\lambda$  represents the calculated output for wavelength via forward propagation, and  $o_\lambda$  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  $\Delta$ , 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  $z_{\text{aer}}$ , and the same for  $\tau$ . 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  $\pi$  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  $\Delta$  with respect to the weights are computed using reverse-mode automatic differentiation which is a powerful algorithm that 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  $z_{\text{aer}}$ ) 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 ( $\omega$ ), 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  $z_{\text{aer}}$ . 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  $z_{\text{aer}}$ . 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.



**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 ( $\theta_0$ )	in feature vector	$8.20^\circ - 80.0^\circ$
	Viewing zenith angle ( $\theta$ )	in feature vector	$0.0^\circ - 66.60^\circ$
	Solar azimuth angle ( $\phi_0$ )	in feature vector	$-180.0^\circ - 180.0^\circ$
	Viewing azimuth angle ( $\phi$ )	in feature vector	$-180.0^\circ - 180.0^\circ$
Aerosol parameters	Aerosol fraction	fixed	1.0
	Single scattering albedo ( $\omega$ )	fixed	0.95
	Aerosol optical thickness ( $\tau$ )	in feature vector	0.05 - 5.0
	Aerosol layer height ( $z_{\text{aer}}$ )	in feature vector	75 hPa - 1000.0 hPa
	Aerosol layer thickness ( $p_{\text{thick}}$ )	varied	-
	Scattering phase function	fixed	Henyey-Greenstein
	asymmetry factor ( $g$ )	fixed	0.7
	Angstrom exponent ( $\text{\AA}$ )	fixed	0.0
Meteorological parameters	Temperature	in feature vector	temperature at $z_{\text{aer}}$
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	$2.08\text{E-}7 - 0.70$

### 3.2 Training the neural networks

Since the NN forward model is specifically designed for TROPOMI, the solar-satellite geometry is selected 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.

Generally, the required training data size increases with increasing non-linearity between input and output layers in a neural network — there isn’t a specific method to accurately determine the required sample size before training. 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. 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 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, two hidden layers and three hidden layers, all consisting of 50 nodes each. Depending on the best performing configuration of the number of hidden layers, two other configurations are added containing 100 and 200 nodes in each of the 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 configuration were trained for a total of 25,000 iterations. Of every configuration tested for each of the neural network models, the most optimal configuration was found to be two hidden layers containing 100 nodes each. Figure 2 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, within 1-3% to 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  $\tau$  ( $NN_{K_\tau}$ ) performed very well with errors not exceeding more than 3%. Averaged convolved errors for the neural network model for the derivative of the reflectance with respect to  $z_{aer}$  ( $NN_{K_{z_{aer}}}$ ) also show good agreements, with errors in parts of the spectrum with very low  $z_{aer}$  information, e.g. the continuum (3d). It is important to note that although 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).

#### 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  $z_{\text{aer}}$  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 asymmetry 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 4a), with a Pearson correlation coefficient close to 1.0. Introducing model errors in other aerosol properties such as single scattering albedo (Figure 4b) and scattering phase function (Figure 4c) 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 4d).

A total of 5558 retrievals out of the 8000 difference cases converged to a final solution. On average,  $z_{\text{aer}}$  retrieved using NN differed by approximately 5.0 hPa from the same using Disamar (Figure 5), with a median of approximately 2.0 hPa. The spread of the retrieval differences were minimal, with a majority of the retrievals differing 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 6). 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 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.

#### 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 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 7a). 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



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

model parameter	experiment		Disamar		NN	
	value in sim	value in ret	converged	non converged	converged	non converged
$p_{\text{thick}}$	200 hPa	50 ha	1641	359	1550	450
$\omega$	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

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 7b) 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.

5 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 8a). 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 8b). Although visually discernable in the difference map in Figure 8c, the retrieved  $z_{\text{aer}}$  from both algorithms were quite similar (Figure 9a). 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 9b). While the standard deviation of approximately 160 meters indicates the presence of outliers, the 15<sup>th</sup> and the 85<sup>th</sup> 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, they primarily departed in 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 9c). 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 on average by 60 meters with a standard deviation of 30 meters. This departure is not well understood.

**Table 4.** Statistics of difference between retrieved  $z_{\text{aer}}$  from disamar and NN, as defined in figure 8c.

AAI [-]	number of samples	mean [m]	median [m]	standard deviation [m]	15 <sup>th</sup> percentile [m]	85 <sup>th</sup> 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  $184.01 \pm 0.50$  seconds per pixel, whereas the same for the neural network processor was  $0.167 \pm 0.0003$  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 bread-board 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.

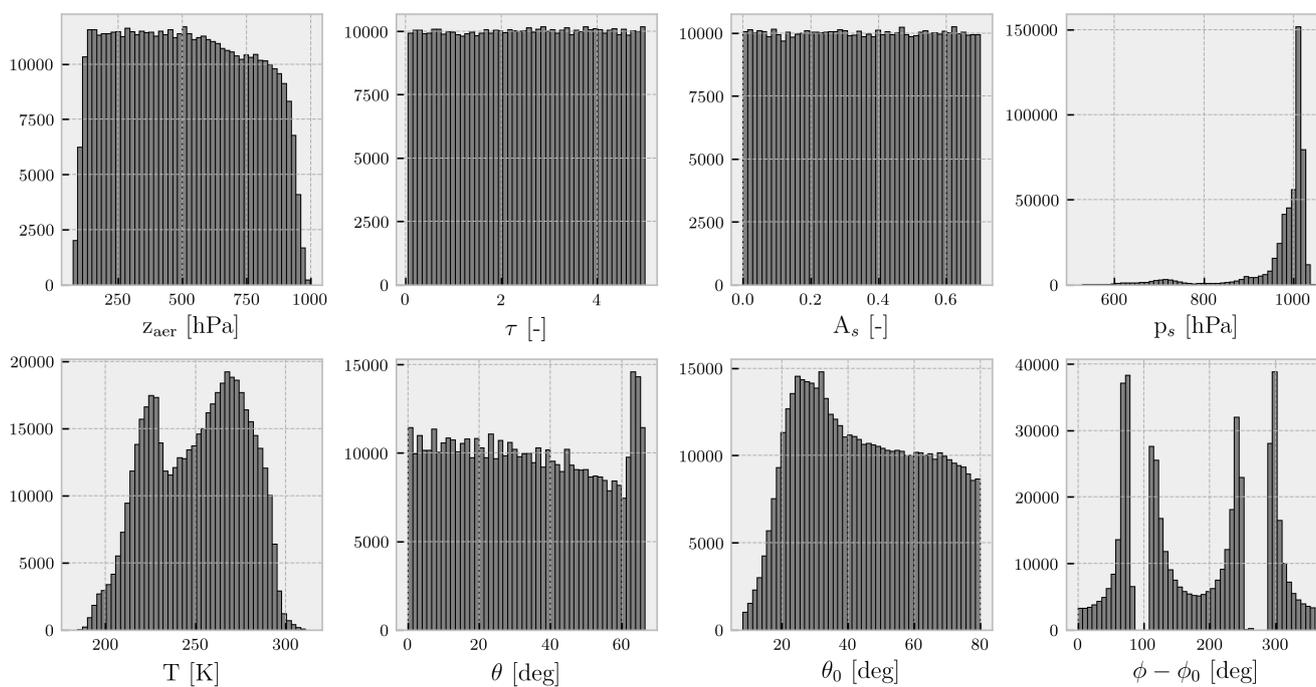
For a real test case, TROPOMI spectra over the December 12, 2017 forest fires in Southern California were chosen. On this day, the biomass burning plume extended from land to the ocean over a dry and almost cloudless scene. Operational retrievals using both Disamar and the neural network forward models 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



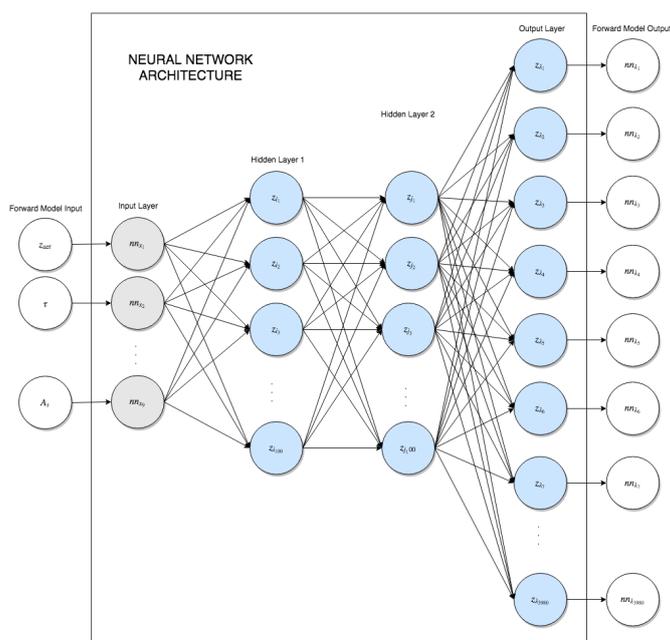
TROPOMI. Such a boost in processor output allows for better analyses of retrievals and opens the possibility to remove some of the forward model simplifications mentioned in Section 2.2, which paves the way for further developing the TROPOMI aerosol layer height algorithm.

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

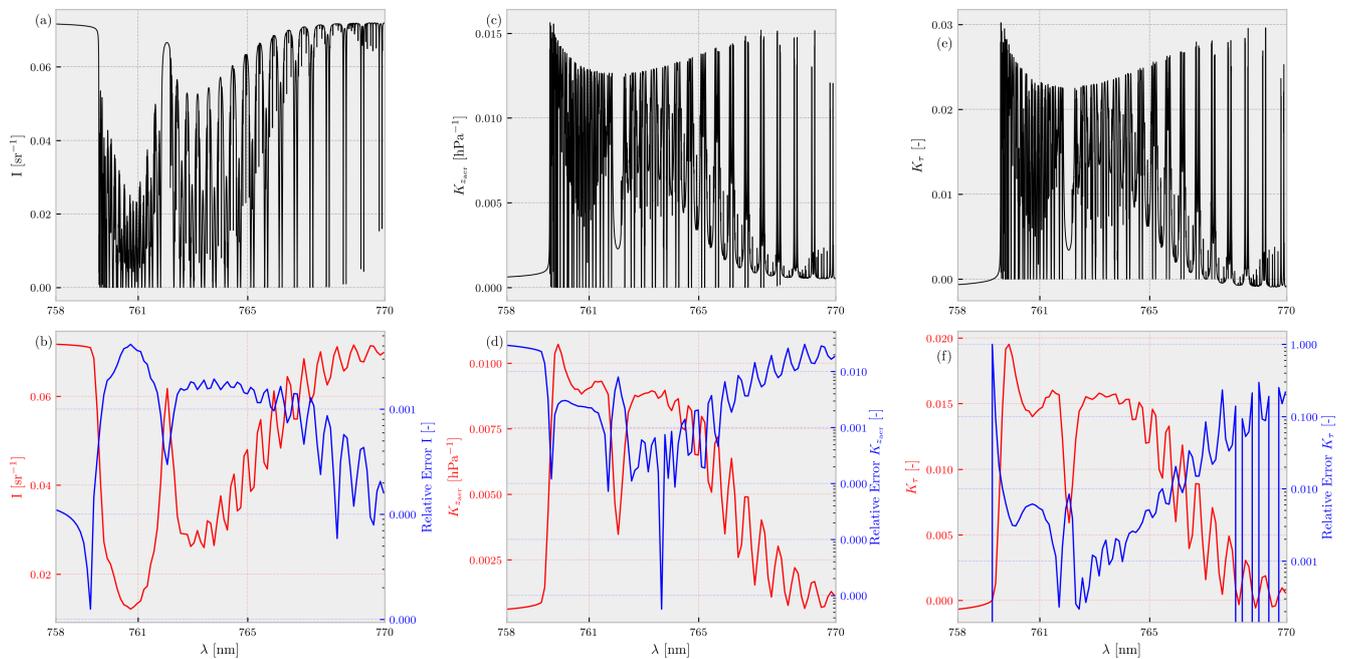
- 5 *Acknowledgements.* This publication contains modified Copernicus Sentinel data. This research is partly funded by the European Space Agency (ESA) within the EU Copernicus programme.



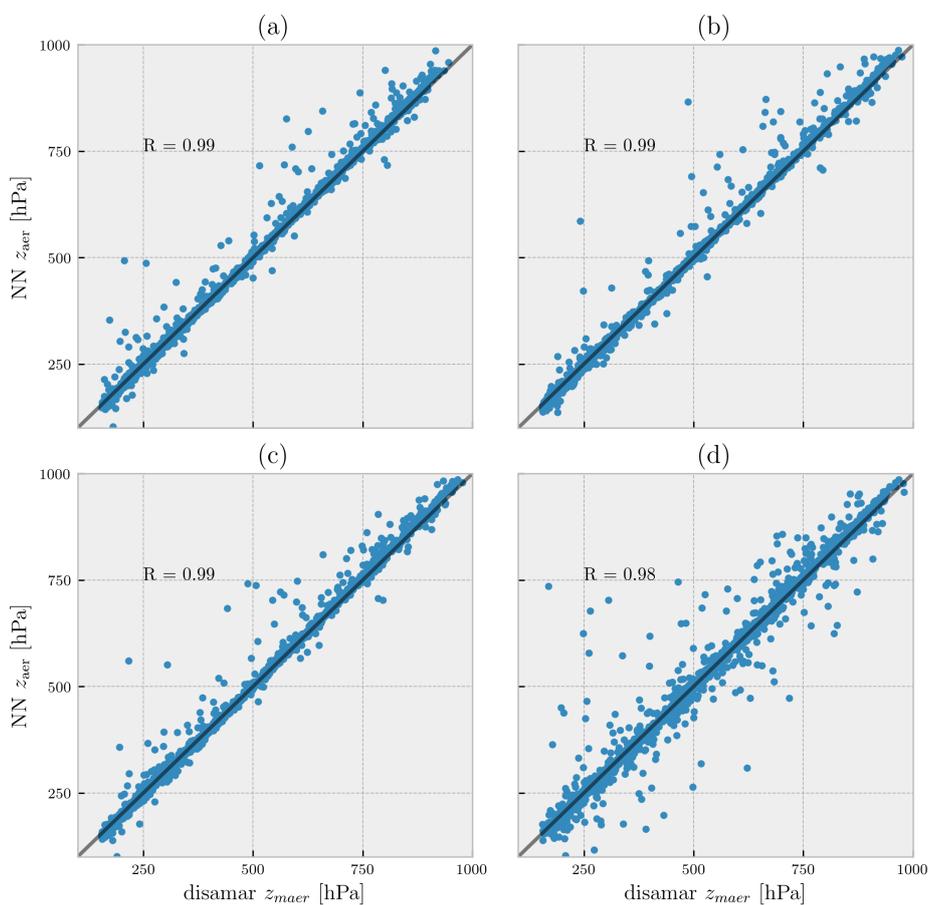
**Figure 1.** Histograms of the various input parameters for each of the neural network models in NN. Minimum and maximum values for each of the parameters are available in Table 2.



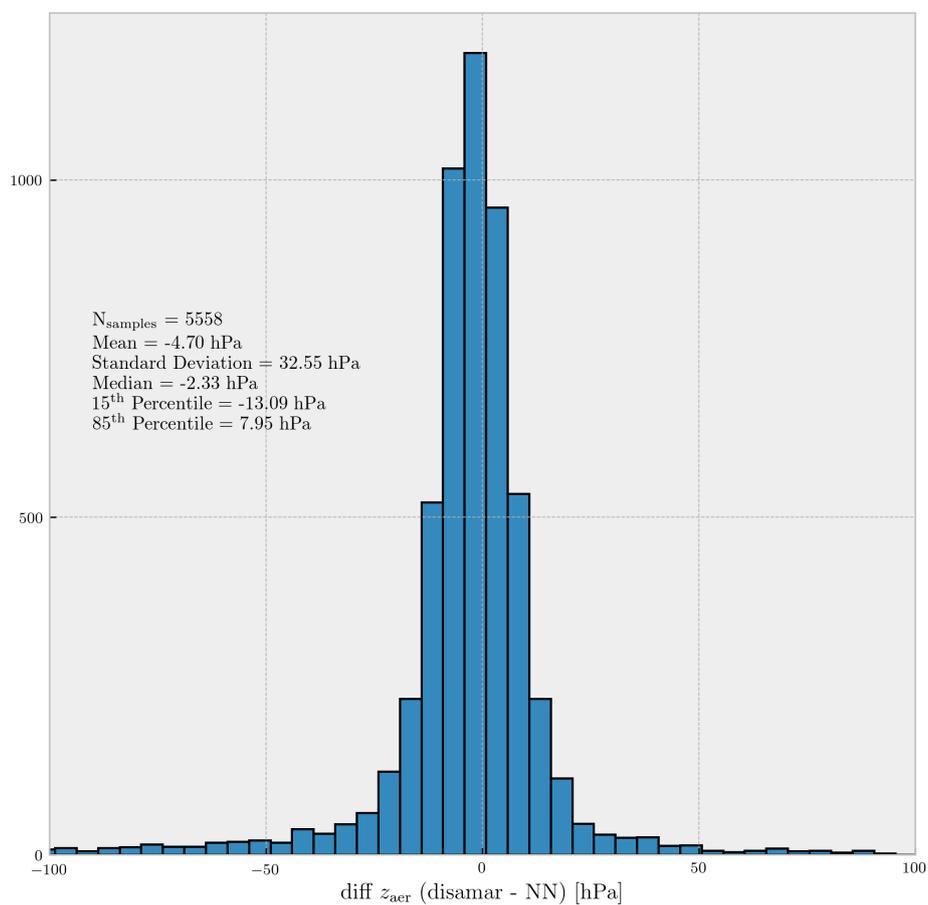
**Figure 2.** A schematic of each of the three neural networks in NN. There are two hidden layers, each containing 100 nodes.  $z$  represents inputs for each of the nodes, whereas  $nn$  represents the inputs and outputs of the neural network.



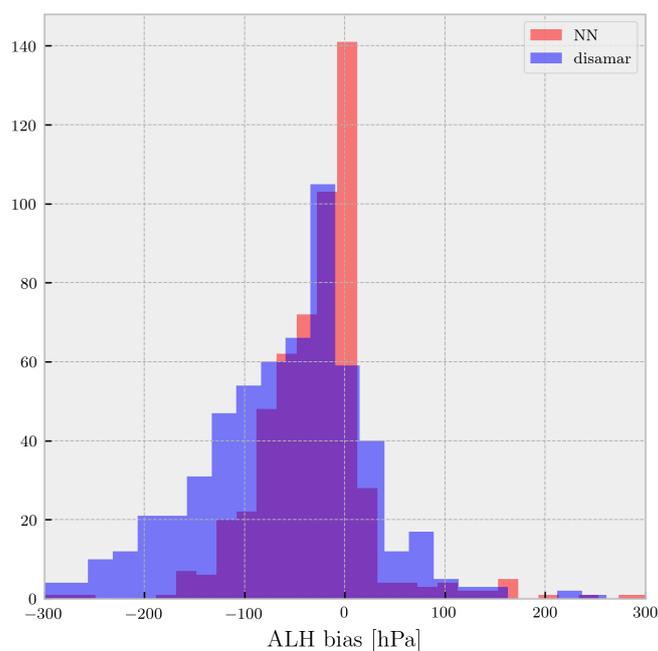
**Figure 3.** Performance of the finalised neural network. The top row represents the averaged output of each of the neural networks for surface albedo less than 0.4. The bottom row represents the convolved version of the top row (plotted as the red line with the left-handed y-axis) and the convolved 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.



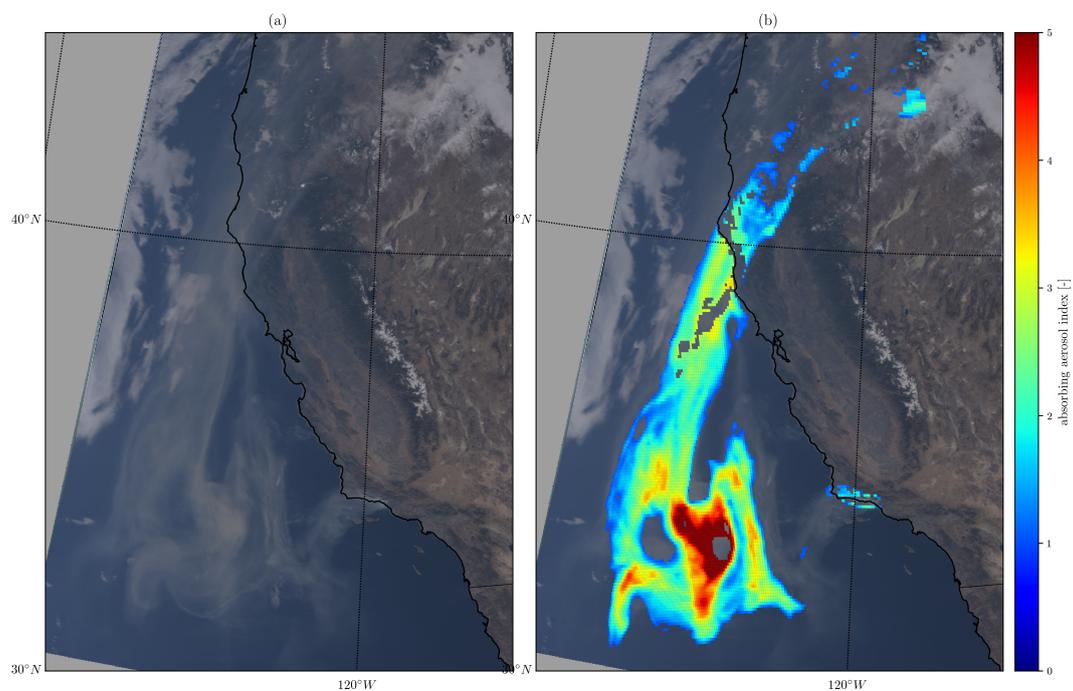
**Figure 4.** Retrieved layer heights compared between Disamar and NN for 2000 synthetic spectra in the presence of model errors. The dots represent converged scenes only, with the x axis representing retrievals from Disamar and the y-axis representing the same from NN. The model errors represented in 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  $z_{\text{aer}}$  from different methods is mentioned in each of the plots.



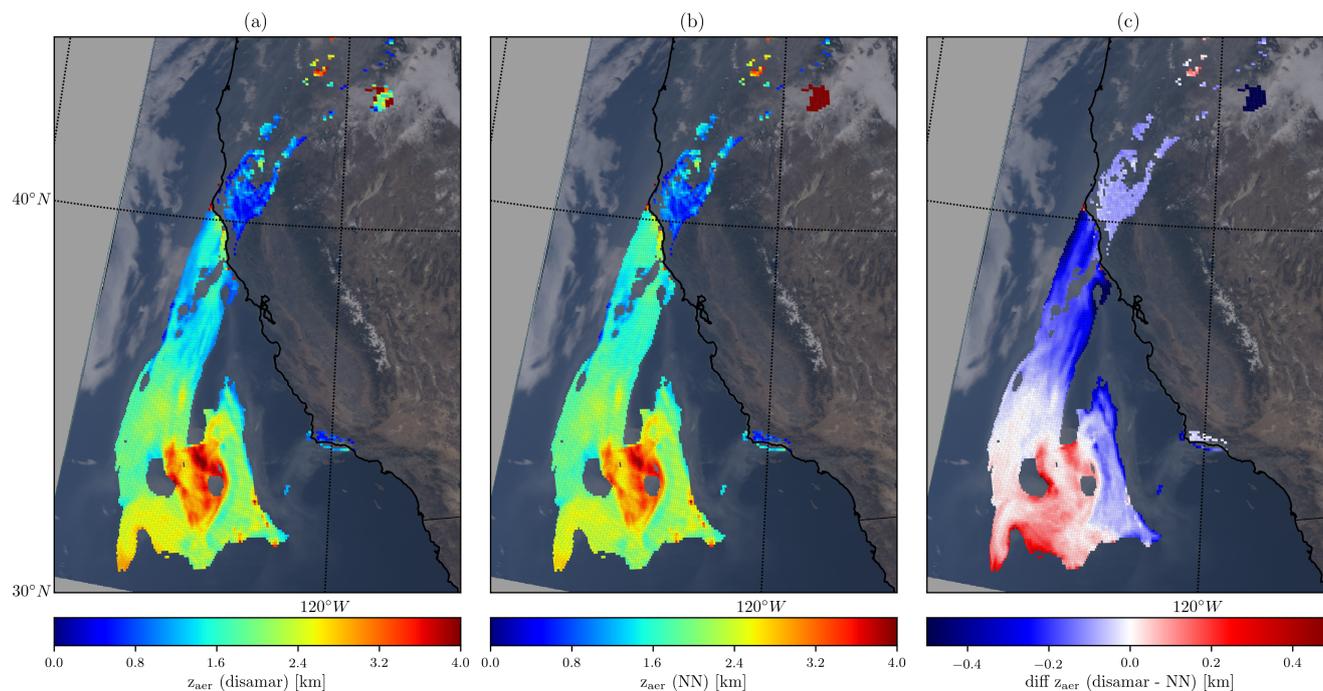
**Figure 5.** A histogram of differences between the retrieved  $z_{aer}$  values using Disamar and NN retrieval methods for synthetic spectra generated by Disamar. Total number of cases is 8000, whereas the plot contains 5558 retrieved samples for both Disamar and NN; non-converged cases are not included. A map of these differences are plotting in Figure 8c.



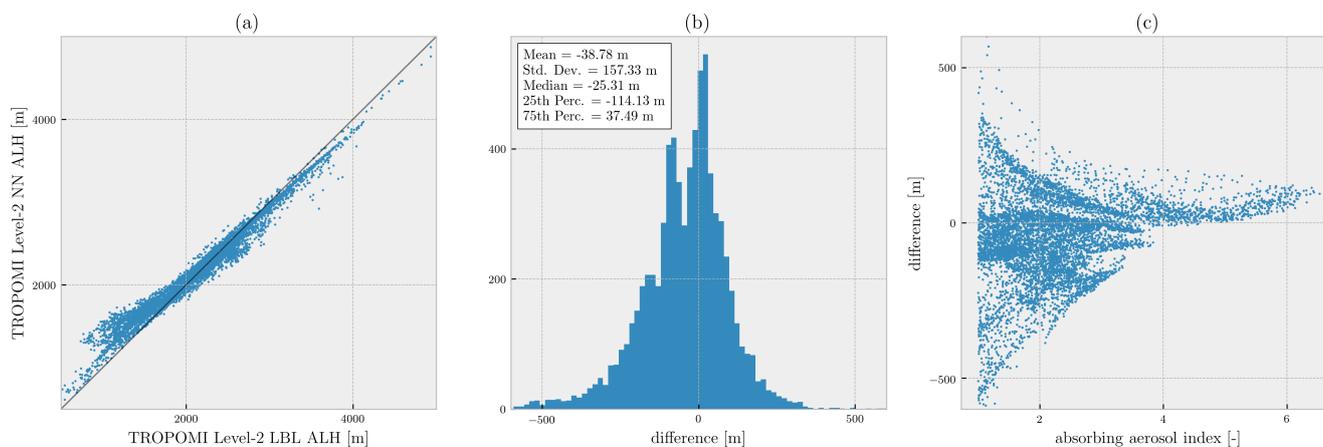
**Figure 6.** A histogram of biases (retrieved - true) for scenes in the synthetic experiment for which either NN converges to a solution (red bar plot) and Disamar does not, or Disamar converges to a solution (blue bar plot) whereas NN does not.



**Figure 7.** (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 8.** (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 9.** 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. Figure (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  $z_{\text{aer}}(\text{Disamar}) - z_{\text{aer}}(\text{NN})$ . Figure (c) compares these differences with 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.
- 10 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 O<sub>2</sub>-O<sub>2</sub> 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.
- 15 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.
- 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.
- 20 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, I., 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.
- 25 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 O<sub>2</sub> 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.
- 30 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.
- 35



- 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.
- Heney, L. C. and Greenstein, J. L.: Diffuse radiation in the Galaxy, *The Astrophysical Journal*, 93, 70, <https://doi.org/10.1086/144246>,  
5 <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:  
10 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.
- 20 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,  
25 161–175, <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.
- 30 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.
- 35 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](http://www.tropomi.eu/sites/default/files/files/S5P-KNMI-L2-0006-RP-TROPOMI_ATBD_Aerosol_Height-v1p0p0-20160129.pdf), 2016.
- 5 Sanders, A. F. J., de Haan, J. F., Sneep, M., Apituley, A., Stammes, P., Viteez, 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.
- 10 Sioris, C. E. and Evans, W. F. J.: Impact of rotational Raman scattering in the O<sub>2</sub>A 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.
- 15 Vasilkov, A., Joiner, J., and Spurr, R.: Note on rotational-Raman scattering in the O<sub>2</sub> 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.
- 20 Wang, P., Stammes, P., van der A, R., Pinardi, G., and van Roozendaal, M.: FRESKO+: an improved O<sub>2</sub> 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.
- 25 Wang, P., Tuinder, O. N. E., Tilstra, L. G., de Graaf, M., and Stammes, P.: Interpretation of FRESKO 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.
- 30 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.