





Aerosol Parameters Retrieval from TROPOMI/S5P Using Physics-Based Neural Networks

Lanalan Rao , Jian Xu , *Member, IEEE*, Dmitry S. Efremenko , Diego G. Loyola , *Senior Member, IEEE*, and Adrian Doicu

Abstract—In this paper, we present three algorithms for aerosol parameters retrieval from TROPOMI measurements in the O₂ A-band. These algorithms use neural networks (i) to emulate the radiative transfer model and a Bayesian approach to solve the inverse problem, (ii) to learn the inverse model from the synthetic radiances, and (iii) to learn the inverse model from the principal-component transform of synthetic radiances. The training process is based on full-physics radiative transfer simulations. The accuracy and efficiency of the neural network based retrieval algorithms are analyzed with synthetic and real data.

Index Terms—Aerosol information retrieval; neural networks; TROPOMI/S5P.

I. INTRODUCTION

AEROSOLS affect Earth's radiation budget by scattering and absorbing solar radiation (direct effect) and by influencing the cloud formation processes (indirect effect). Highly absorbing aerosols also have a warming effect on the atmosphere leading to the evaporation of cloud particles, which results in a reduction of the cloud cover (semi-direct effect). Accurate assessments of aerosol properties, such as optical depth and layer height, are important for the global monitoring of air pollution in the lower atmosphere.

A number of passive satellite sensors enable to monitor aerosol properties on both regional and global scale using spectral information at various wavelengths. For instance, measurements in the O₂ A-band from the Global Ozone Mapping Experiment (GOME) [1] and GOME-2 [2], the Scanning Imaging Absorption Spectrometer for Atmospheric CHartography (SCIAMACHY) [3–5], the Greenhouse Gases Observing Satellite (GOSAT) [6], and the TROPospheric Monitoring Instrument (TROPOMI) onboard the Sentinel-5 Precursor (S5P) [7, 8] are used to retrieve aerosol optical depth and height information.

The inversion methods used in atmospheric remote sensing aim to recover atmospheric parameters by minimizing the residual between the measurements and the radiative transfer model simulations. The solution of the minimization problem can be found by using deterministic (e.g., Tikhonov-type

regularization methods [9, 10]) or stochastic approaches (e.g., Bayesian methods [11]). In both cases, the computations of the forward model and the Jacobian matrix impose the performance bottleneck in the whole processing chain. Therefore, it would be problematic to adopt these approaches for the operational processing of remote sensing data from new-generation sensors. To tackle this problem, artificial neural networks, which are able to approximate very quickly any continuous function with a sufficiently high accuracy [12, 13] and to estimate the derivatives of the function with respect to the model inputs, can be used. Actually, a trained neural network may provide accurate estimates of the forward model and its Jacobian, in a fraction of time compared to classical retrieval algorithms. In atmospheric remote sensing, neural networks have already been widely applied. These techniques have been used (i) to approximate a radiative transfer model (or a part of it) [14–19], (ii) to learn the inverse mappings [20–28], and (iii) to recover some atmospheric retrieval parameters, which are then taken as initial guesses in an optimization approach [29, 30]. In this context, it should be pointed out that the two TROPOMI operational retrieval algorithms based on the O₂ A-band measurements, use neural network based forward models together with a Bayesian approach for the retrieval of cloud properties [31, 32] and the aerosol layer height [33–35].

In this study, we present three types of neural networks for aerosol retrieval from TROPOMI measurements. The first one uses a neural network to emulate the radiative transfer model and a Bayesian approach to solve the inverse problem, the second one uses a neural network to learn the inverse model from the synthetic radiances, and the third one uses a neural network to learn the inverse model from the principal-component transform of synthetic radiances following the full-physics inverse learning machine method [18]. The major goal of this study is to incorporate the three neural-network algorithms into a common tool, and to analyze and compare their retrieval performances. To the best of our knowledge, such a comparison study had not been done before. The paper is organized as follows. In Section II, we summarize the main features of the adopted radiative transfer model, while in Section III, we provide a detailed description of three physics-based algorithms using neural networks. In Section IV, the corresponding retrieval performances of the neural networks are analyzed using synthetic and real TROPOMI data.

II. RADIATIVE TRANSFER MODEL

Any physics-based retrieval algorithm uses a model for computing the radiative transfer in a planetary atmosphere.

Manuscript created May, 2022; This work was supported by the German Aerospace Center (DLR) programmatic [Nachwuchsgruppe "Retrieval der nächsten Generation", 2 472 469] and the Chinese Academy of Sciences (CAS) "Pioneering Initiative Talents Program" under Grant E1RC2WB2. The work of L. Rao was partly funded by the Chinese Scholarship Council. (Corresponding author: Jian Xu.)

Lanlan Rao, Dmitry S. Efremenko, Diego G. Loyola, and Adrian Doicu are with the Remote Sensing Technology Institute, German Aerospace Center (DLR), 82234 Oberpfaffenhofen, Germany (email: lanlan.rao@dlr.de; dmitry.efremenko@dlr.de; diego.loyola@dlr.de; adrian.doicu@dlr.de).

Jian Xu is with the National Space Science Center, Chinese Academy of Sciences, Beijing 100190, China (email: xujian@nssc.ac.cn).

We begin our analysis by summarizing the peculiarities of the radiative transfer model used in this study.

Regarding the TROPOMI instrument, each swath row (angle) r with $r = 1, \dots, R$, is characterized by its own measurement wavelength grid $\{\lambda_{mk}^r\}_{k=1}^{N_{m\lambda}}$ and slit function g_r , where R is the number of swath rows and $N_{m\lambda}$ the number of spectral grid points. This is due to the optics of the spectrometer (point spread function, aberrations, and defocusing) and small changes in the width of the slit. The noisy and synthetic radiances measured by the instrument at a wavelength λ_{mk}^r are given respectively, by

$$I^\delta(\lambda_{mk}^r) = I(\lambda_{mk}^r) + \delta_{mk}, \quad (1)$$

$$I(\lambda_{mk}^r) = \int_{-\infty}^{\infty} g_r(\lambda_{mk}^r - \lambda) I(\lambda) d\lambda, \quad (2)$$

where $I(\lambda)$ is the radiance computed by a radiative transfer model at a forward wavelength λ (before convolution) and δ_{mk} the measurement noise. Note that because $\{\lambda_{mk}^r\}_{k=1}^{N_{m\lambda}}$ changes slightly with r , we assumed that the noise is row independent.

The synthetic radiances $I(\lambda)$ are computed on the forward wavelength grid by a radiative transfer model relying on the discrete ordinate method with matrix exponential [36, 37]. To speed up the computation,

- standard acceleration methods, i.e., the telescoping technique [38, 39] and the method of false discrete ordinate [40], as well as
- hyperspectral acceleration methods, i.e., the correlated k-distribution method [41] and principal component analysis [42–46],

are employed. Line-by-line calculations [47] with optimized rational approximations for the Voigt line profile [48] (taken from HITRAN database [49]) are used to compute the absorption cross sections of gas molecules, the methodology describe in [50] is used to compute the Rayleigh cross-section and depolarization ratios, and the delta-M approximation [51] and TMS (Truncated Multiple and Single scattering approximation) correction [52, 53] are used in the radiative transfer calculations. The radiative transfer model includes several types of bidirectional reflectance distribution functions over land and water. However, in order to simplify the analysis, we use the geometry-dependent effective Lambertian equivalent reflectivity (GE_LER) product that accounts for satellite viewing dependencies and improves the accuracy for actual snow/ice conditions [28].

The aerosol optical depth τ and layer height H are retrieved in the oxygen absorption A-band (758–771 nm). To simplify the analysis, the aerosol layer is assumed to be homogeneous with a fixed thickness of 0.5 km, spreading evenly from $H - 0.25$ km to the $H + 0.25$ km. As the considered spectral range is narrow, the aerosol optical depth is assumed to be constant within this range. There are five sets of aerosol microphysical models [54, 55] included in the radiative transfer model under the assumption of spherical aerosol particles. For example, the set of aerosol models taken from the MODIS Dark-Target (DT) algorithm [56] includes: non-absorbing, moderately absorbing, and absorbing aerosols, as well as, desert dust. Each aerosol model is characterized by a bi-mode log-normal volume size

distribution comprising a fine and a coarse mode. Specifically, in the case of moderately absorbing aerosols, the median radius of the fine mode r_{vf} and the coarse mode r_{vc} are given respectively, by

$$r_{vf} = 0.145 + 0.0203\tau \text{ (}\mu\text{m)}, \quad (3)$$

$$r_{vc} = 3.101 + 0.3364\tau \text{ (}\mu\text{m)}, \quad (4)$$

the standard deviations of the fine mode s_f and the coarse mode s_c by

$$s_f = 0.374 + 0.1365\tau, \quad (5)$$

$$s_c = 0.729 + 0.098\tau, \quad (6)$$

the volumes of the fine mode particles V_{of} and the coarse mode particles V_{oc} by

$$V_{of} = 0.1642\tau^{0.775} \text{ (}\mu\text{m}^3/\mu\text{m}^2), \quad (7)$$

$$V_{oc} = 0.1482\tau^{0.684} \text{ (}\mu\text{m}^3/\mu\text{m}^2), \quad (8)$$

and the complex refractive index of the aerosol particles by $m = (1.43, 0.008 - 0.002\tau)$. Note that the parameters of the size distribution and the refractive index depend on the aerosol optical depth τ . In addition, the radiative transfer model can deal with several types of aerosol profiles, as for example, elevated box, exponential decay, a combination of exponential decay and ground box, and Gaussian.

III. NEURAL NETWORK ALGORITHMS

In this study, neural networks are employed (i) to emulate the radiative transfer model and (ii) to learn the inverse model. The first one is referred to as a neural network for the forward operator, whereas for the second one a neural network for the inverse operator.

The neural network approach can be briefly summarized as follows. Let us consider the model $\mathbf{y} = \mathbf{F}(\mathbf{x})$, where $\mathbf{x} \in \mathbb{R}^{N_x}$ and $\mathbf{y} \in \mathbb{R}^{N_y}$ are the input and output vectors, respectively, and \mathbf{F} is a deterministic function. In order to approximate $\mathbf{F}(\mathbf{x})$ by a neural network model $\mathbf{f}(\mathbf{x}, \boldsymbol{\omega})$ with parameters $\boldsymbol{\omega}$, we consider a deep neural network consisting of units (nodes) arranged in an input, output, and several hidden layers. For a neural network with $L + 1$ layers and N_l units in layer l , the feed-forward operations read as

$$\mathbf{u}_l = \mathbf{W}_l \mathbf{y}_{l-1} + \mathbf{b}_l, \quad (9)$$

$$\mathbf{y}_l = \phi_l(\mathbf{u}_l), \quad l = 1, \dots, L, \quad (10)$$

where $l = 0$ and $l = L$ are the input and output layers, respectively, ϕ_l is the layer activation function, $\mathbf{W}_l \in \mathbb{R}^{N_l \times N_{l-1}}$ the matrix of weights connecting the layers $l - 1$ and l , $\mathbf{b}_l \in \mathbb{R}^{N_l}$ the vector of biases corresponding to layer l , and $\boldsymbol{\omega} = \{\mathbf{W}_l, \mathbf{b}_l\}_{l=1}^L$ the set of network parameters. In the input and output layers, we have $\mathbf{y}_0 = \mathbf{x}$ and $\mathbf{y}_L = \mathbf{f}(\mathbf{x}, \boldsymbol{\omega})$, respectively, so that, $N_0 = N_x$ and $N_L = N_y$. Deep learning is the process of computing the network parameters $\boldsymbol{\omega}$ on a data set $\mathcal{D} = \{(\mathbf{x}^{(n)}, \mathbf{y}^{(n)})\}_{n=1}^N$, where $\mathbf{y}^{(n)} = \mathbf{F}(\mathbf{x}^{(n)})$ and N is the number of samples. This is done by computing a point estimate $\hat{\boldsymbol{\omega}}$ as the minimizer of a loss function with a penalty term controlling the amplitudes of the network parameters, i.e.,

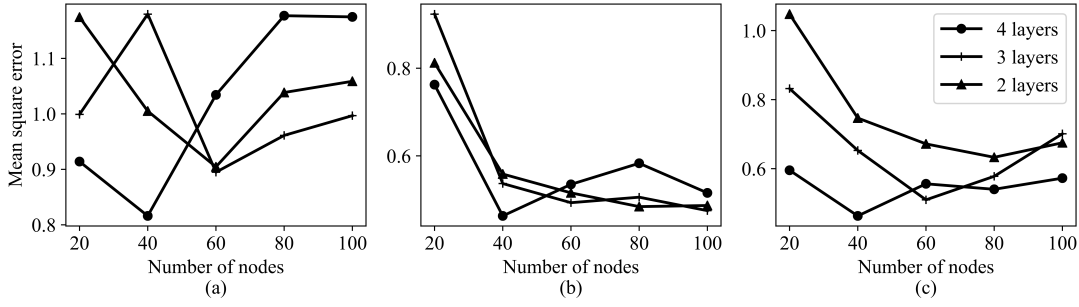


Fig. 1. Mean square error for the validation data set. The results correspond to all neural networks considered in this study, i.e., the forward-operator neural network (a), the inverse-operator neural network with synthetic radiances (b), and the inverse-operator neural network with the principal-component transform of synthetic radiances (c).

TABLE I
VALUE RANGES OF THE OPTICAL AND GEOMETRICAL PARAMETERS FOR GENERATING THE DATA SET.

Parameter	Value range
τ	0.05 – 5
H	0.1 – 15.75 km
θ_0	0 – 75°
θ	0 – 70°
$\Delta\varphi$	0 – 180°
H_s	0 – 2.61 km
A_s	0 – 0.4

$E(\omega) = \frac{1}{2} \sum_{n=1}^N \|\mathbf{y}^{(n)} - \mathbf{f}(\mathbf{x}^{(n)}, \omega)\|^2 + \gamma \|\omega\|^2$, where γ is the regularization parameter.

To formulate the retrieval problem, we group the optical and geometrical parameters into

- retrieval parameters, which include the aerosol optical depth τ at 760 nm and aerosol layer height H defined as the middle height of an aerosol layer with a fixed thickness of 0.5 km, and
- forward model parameters, which include the solar zenith angle θ_0 , viewing zenith angle θ , relative azimuth angle $\Delta\varphi$, surface height H_s , and surface albedo A_s (note that the forward model parameters are not part of the retrieval).

For generating the data set, samples of optical and geometrical parameters are produced by means of a smart sampling technique [57] based on Halton sequences [58]; their ranges of variations are shown in Table I. The neural networks are trained for the moderately absorbing aerosol model from the MODIS DT algorithm.

The hyperparameters of the neural network, i.e., the number of hidden layers and the number of units in each layer, are optimized by using 10% of the samples from the training set for validation. In the validation stage, the holdout cross-validation and a grid search procedure are used; the grid search is performed over the sets $\{2, 3, 4\}$ of hidden layers and $\{20, 40, 60, 80, 100\}$ of layer units. For all neural network considered in this study, a network architecture with 4 hidden layers and 40 units in each layer yields the lowest mean-square error on the validation data set (see Fig. 1), and no over-fitting has been observed. A hyperbolic tangent activation function is taken, and as optimization tool, the mini-batch gradient descent with ADaptive Moment Estimation (ADAM) [59] is

utilized.

A. Neural network for the forward operator

For emulating the radiative transfer model, we consider a neural network in which, the input \mathbf{x} is the set of optical and geometrical parameters, while the output \mathbf{y} is the set of synthetic radiances $I(\lambda_k)$ computed on the forward wavelength grid $\{\lambda_k\}_{k=1}^{N_\lambda}$, i.e.,

$$\mathbf{x} = \begin{bmatrix} [\tau, H]^T \\ [\theta_0, \theta, \Delta\varphi, H_s, A_s]^T \end{bmatrix} \mapsto \mathbf{y} = [I(\lambda_k)]_{k=1}^{N_\lambda}. \quad (11)$$

Thus, the dimensions of the input and output vectors are $N_x = 7$ and $N_y = N_\lambda$, respectively.

The forward wavelength grid consists of $N_\lambda = 465$ equidistant spectral points ranging from 757.4 to 771.6 nm, while the number of samples in training set is $N_t = 151\,423$ based on the number of combinations of optical and geometrical parameters defined by Table I. After the radiative transfer model is learned, the synthetic radiances computed at a high spectral resolution on the forward wavelength grid $I(\lambda_k)$ are convolved with a slit function g_r to obtain the synthetic radiances on the measurement wavelength grid $I(\lambda_{mk}^r)$ (cf. Eq. (2)).

The retrieval of aerosol parameters encapsulated now in the state vector $\mathbf{x} = [\tau, H]^T$ requires the solution of the non-linear equation

$$\mathbf{y}^\delta = \mathbf{F}(\mathbf{x}) + \delta_m, \quad (12)$$

where, for a given swath row r , $\mathbf{y}^\delta = [I_{\text{mes}}(\lambda_{mk}^r)]_{k=1}^{N_{m\lambda}}$ is the measurement vector, $\mathbf{F}(\mathbf{x}) = [I(\lambda_{mk}^r)]_{k=1}^{N_{m\lambda}}$ the forward model, and $\delta_m = [\delta_{mk}]_{k=1}^{N_{m\lambda}}$ the measurement noise vector. The non-linear equation (12) is solved by using a Bayesian approach [11]. In this approach, the a posteriori density $p(\mathbf{x} | \mathbf{y}^\delta)$ representing the conditional probability density of the state vector \mathbf{x} given the data \mathbf{y}^δ is the quantity of interest. Assuming that

- 1) the measurement noise vector δ_m is a Gaussian random vector with zero mean and noise covariance matrix $\mathbf{C}_m = \text{diag}[\sigma_{mk}^2]_{k=1}^{N_{m\lambda}} = \sigma_m^2 \mathbf{C}_m$, i.e., $\delta_m \sim \mathcal{N}(\mathbf{0}, \mathbf{C}_m)$, where $\sigma_m^2 = \sum_{k=1}^{N_{m\lambda}} \sigma_{mk}^2$ is the noise variance, and
- 2) the state vector \mathbf{x} is a Gaussian random vector with mean \mathbf{x}_a and a priori covariance matrix $\mathbf{C}_x = \text{diag}[\sigma_{xk}^2]_{k=1}^{N_x} = \sigma_x^2 \mathbf{C}_x$, i.e., $\mathbf{x} \sim \mathcal{N}(\mathbf{x}_a, \mathbf{C}_x)$, where \mathbf{x}_a is the a priori

state vector, $\sigma_x^2 = \sum_{k=1}^{N_x} \sigma_{xk}^2$ the a priori state variance, and the notation $\mathcal{N}(\mathbf{x}_{\text{mean}}, \mathbf{C}_x)$ symbolizes a normal distribution with mean \mathbf{x}_{mean} and covariance matrix \mathbf{C}_x , we find

$$p(\mathbf{x} | \mathbf{y}^\delta) \propto \exp\left[-\frac{1}{2}V_\alpha(\mathbf{x} | \mathbf{y}^\delta)\right], \quad (13)$$

where

$$V_\alpha(\mathbf{x} | \mathbf{y}^\delta) = \frac{1}{\sigma_m^2} \left\{ [\mathbf{y}^\delta - \mathbf{F}(\mathbf{x})] \mathbf{C}_m^{-1} [\mathbf{y}^\delta - \mathbf{F}(\mathbf{x})]^T + \alpha (\mathbf{x} - \mathbf{x}_a) \mathbf{C}_x^{-1} (\mathbf{x} - \mathbf{x}_a)^T \right\} \quad (14)$$

is the a posteriori potential and $\alpha = \sigma_m^2 / \sigma_x^2$ the regularization parameter. Here, the notation $\mathcal{N}(\mathbf{x}_{\text{mean}}, \mathbf{C}_x)$ stands for a normal distribution with mean \mathbf{x}_{mean} and covariance matrix \mathbf{C}_x . The maximum a posteriori estimate $\hat{\mathbf{x}}_\alpha^\delta$ is then computed as

$$\hat{\mathbf{x}}_\alpha^\delta = \arg \min_{\mathbf{x}} V_\alpha(\mathbf{x} | \mathbf{y}^\delta). \quad (15)$$

After scaling the data model (12) with the matrix $\mathbf{P} = \mathbf{C}_m^{-1/2}$, i.e., after transforming the data model into a model with white noise, and introducing the regularization matrix \mathbf{L} through the Cholesky factorization $\mathbf{C}_x^{-1} = \mathbf{L}^T \mathbf{L}$, we are led to a least-squares problem which is solved by using the iteratively regularized Gauss-Newton method [60]. This method supplies the optimal value of the regularization parameter and the corresponding regularized solution.

In the inversion step, the noise covariance matrix is chosen as $\mathbf{C}_m = \text{diag}[\sigma_{mk}^2]_{k=1}^{N_{m\lambda}}$ with $\sigma_{mk} = 0.02 \times \bar{I}(\lambda_{mk})$ and $\bar{I}(\lambda_{mk}) = (1/R) \sum_{r=1}^R I(\lambda_{mk}^r)$ for all $k = 1, \dots, N_{m\lambda}$, and the a priori covariance matrix as $\mathbf{C}_x = \text{diag}[\sigma_{xk}^2]_{k=1}^2$ with $\sigma_{xk} = 0.2 \times x_k$ and x_k standing for τ and H .

B. Neural network for the inverse operator

For solving the inverse problem, we designed two types of neural networks following the full-physics inverse learning machine method. The first one uses as input, the synthetic radiances computed on the measurement wavelength grid, while the second one uses as input, the principal-component transform of synthetic radiances.

1) *Neural network for the inverse operator with synthetic radiances:* In principle, for emulating the inverse model, we may use a neural network in which, the input \mathbf{x} includes the noisy radiances on a measurement wavelength grid and the forward model parameters, while the output \mathbf{y} includes the set of parameters to be retrieved, i.e.,

$$\mathbf{x} = \begin{bmatrix} [I(\lambda_{mk}^r) + \delta_{mk}]_{k=1}^{N_{m\lambda}} \\ [\theta_0, \theta, \Delta\varphi, H_s, A_s]^T \end{bmatrix} \mapsto \mathbf{y} = [\tau, H]^T. \quad (16)$$

In this case, the dimensions of the input and output vectors are $N_x = N_{m\lambda} + 5$ and $N_y = 2$, respectively.

The problem which appears is that because we are dealing with a random measurement noise and a set of measurement wavelength grids, the same output corresponds to different realization of the random noise, as well as, to different wavelength grids. To reduce the dimension of the data set, we use the jitter approach under the assumption that the measurement wavelength grid $\{\lambda_{mk}^r\}_{k=1}^{N_{m\lambda}}$ is a discrete random

variable which can take the values $\{\lambda_{mk}^1\}_{k=1}^{N_{m\lambda}}, \dots, \{\lambda_{mk}^R\}_{k=1}^{N_{m\lambda}}$. According to this approach, at each forward pass through the network, a measurement wavelength grid $\{\lambda_{mk}^r\}_{k=1}^{N_{m\lambda}}$ is randomly selected from the R wavelength grids, and a new random noise $\delta_{mk} \sim \mathcal{N}(0, \sigma_{mk}^2)$ is added to the synthetic radiance $I(\lambda_{mk}^r)$. In other words, the input sample is different every time it is passed through the network.

In the training stage, the number of swath rows is $R = 448$, the number of points in each measurement wavelength grid is $N_{m\lambda} = 131$, and the measurement wavelength grids are chosen from the TROPOMI Level-1 product, e.g.,

$$\begin{aligned} \{\lambda_{mk}^1\}_{k=1}^{N_{m\lambda}} &= \{755.120, \dots, 770.929 \text{ nm}\}, \\ \{\lambda_{mk}^2\}_{k=1}^{N_{m\lambda}} &= \{755.133, \dots, 770.942 \text{ nm}\} \\ &\vdots \\ \{\lambda_{mk}^R\}_{k=1}^{N_{m\lambda}} &= \{755.264, \dots, 771.071 \text{ nm}\}, \end{aligned} \quad (17)$$

As before, the noisy spectra are generated with the noise standard deviation $\sigma_{mk} = 0.02 \times \bar{I}(\lambda_{mk})$, where $\bar{I}(\lambda_{mk}) = (1/R) \sum_{r=1}^R I(\lambda_{mk}^r)$. The number of samples in the training set is $N_t = 404901$, where each sample consists of a set of optical and geometrical parameters and the corresponding synthetic radiances computed on all measurement wavelength grids $\{\lambda_{mk}^r\}_{k=1}^{N_{m\lambda}}$, $r = 1, \dots, R$.

2) *Neural network for the inverse operator with the principal-component transform of synthetic radiances:* To reduce the dimension of the synthetic radiance vector $\mathbf{i}_m = [I(\lambda_{mk}^r)]_{k=1}^{N_{m\lambda}} \in \mathbb{R}^{N_{m\lambda}}$, the principal component analysis is applied. Here, the dependency of \mathbf{i}_m on the swath row r is implicitly assumed. For the N_t -dimensional data set $\{\mathbf{i}_m^{(n)}\}_{n=1}^{N_t}$, let $\langle \mathbf{i}_m \rangle = (1/N_t) \sum_{n=1}^{N_t} \mathbf{i}_m^{(n)}$ be the sample mean of the data. The goal of a linear embedding method is to find an M -dimensional subspace ($M < N_{m\lambda}$) spanned by a set of linear independent vectors $\{\mathbf{u}_k\}_{k=1}^M$, such that the centered data $\mathbf{i}_m^{(n)} - \langle \mathbf{i}_m \rangle$ belongs mainly to this subspace, i.e., $\mathbf{i}_m^{(n)} \approx \langle \mathbf{i}_m \rangle + \mathbf{U}_M \hat{\mathbf{i}}_m^{(n)}$, where $\mathbf{U}_M = [\mathbf{u}_1, \dots, \mathbf{u}_M] \in \mathbb{R}^{N_{m\lambda} \times M}$ is the (inverse) mapping from the low- to the high-dimensional space. The dimensionality-reduced input is then $\hat{\mathbf{i}}_m^{(n)} = \mathbf{U}_M^\dagger (\mathbf{i}_m^{(n)} - \langle \mathbf{i}_m \rangle) \in \mathbb{R}^M$, where the (forward) mapping from the high- to the low-dimensional space $\mathbf{U}_M^\dagger \in \mathbb{R}^{M \times N_{m\lambda}}$ is the pseudo-inverse of \mathbf{U}_M , i.e., $\mathbf{U}_M^\dagger \mathbf{U}_M = \mathbf{I}_M$. In the principal component analysis, the transformation matrix \mathbf{U}_M is computed as follows: (i) stack all centered data $\mathbf{i}_m^{(n)} - \langle \mathbf{i}_m \rangle$ into the columns of the matrix \mathcal{I} , i.e., $\mathcal{I} = [\mathbf{i}_m^{(1)} - \langle \mathbf{i}_m \rangle, \dots, \mathbf{i}_m^{(N_t)} - \langle \mathbf{i}_m \rangle] \in \mathbb{R}^{N_{m\lambda} \times N_t}$, (ii) compute the covariance matrix $\mathbf{C} = (1/N) \mathcal{I} \mathcal{I}^T \in \mathbb{R}^{N_{m\lambda} \times N_{m\lambda}}$ and a singular value decomposition of \mathbf{C} , i.e., $\mathbf{C} = \mathbf{U} \mathbf{\Sigma} \mathbf{U}^T$, where $\mathbf{\Sigma} = \text{diag}[\sigma_k]_{k=1}^{N_{m\lambda}}$ is the diagonal matrix of the singular values $\sigma_1 > \sigma_2 > \dots > \sigma_{N_{m\lambda}} > 0$ and $\mathbf{U} = [\mathbf{u}_1, \dots, \mathbf{u}_{N_{m\lambda}}] \in \mathbb{R}^{N_{m\lambda} \times N_{m\lambda}}$ is the orthogonal matrix of the singular vectors, and (iii) take the inverse transformation matrix as $\mathbf{U}_M = [\mathbf{u}_1, \dots, \mathbf{u}_M] \in \mathbb{R}^{N_{m\lambda} \times M}$. In this case, the forward transformation matrix is $\mathbf{U}_M^\dagger = \mathbf{U}_M^T \in \mathbb{R}^{M \times N_{m\lambda}}$. The number of principal components

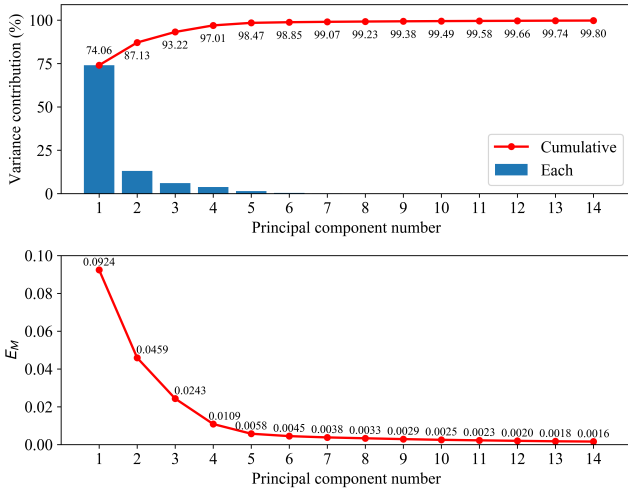


Fig. 2. Contribution of the first 14 principal components to the total variance (top) and reconstruction error with respect to the first 14 principal components (bottom).

M is determined by monitoring the reconstruction error

$$E_M = \sum_{n=1}^{N_t} \|(\mathbf{i}_m^{(n)} - \langle \mathbf{i}_m \rangle) - \mathbf{U}_M \hat{\mathbf{i}}_m^{(n)}\|_2^2 \quad (18)$$

$$= \sum_{n=1}^{N_t} \|(\mathbf{i}_m^{(n)} - \langle \mathbf{i}_m \rangle) - \mathbf{U}_M \mathbf{U}_M^\dagger (\mathbf{i}_m^{(n)} - \langle \mathbf{i}_m \rangle)\|_2^2$$

as function of M , and the M , for which E_M is below a prescribed tolerance, is chosen. For the noisy radiance vector $\mathbf{i}_m^\delta = \mathbf{i}_m + \delta_m$, where $\delta_m \sim \mathcal{N}(\mathbf{0}, \mathbf{C}_m)$ is the measurement noise vector, we find $\langle \mathbf{i}_m^\delta \rangle = \langle \mathbf{i}_m \rangle$, yielding $\hat{\mathbf{i}}_m^\delta = \mathbf{U}_M^T (\mathbf{i}_m^\delta - \langle \mathbf{i}_m \rangle) = \hat{\mathbf{i}}_m + \hat{\delta}_m$ with $\hat{\delta}_m = \mathbf{U}_M^T \delta_m \sim \mathcal{N}(\mathbf{0}, \hat{\mathbf{C}}_m)$ and $\hat{\mathbf{C}}_m = \mathbf{U}_M^T \mathbf{C}_m \mathbf{U}_M \in \mathbb{R}^{M \times M}$.

Thus, instead of the synthetic radiances $\mathbf{i}_m \in \mathbb{R}^{N_{m\lambda}}$, the input of the neural network is the principal-component transform of synthetic radiances $\hat{\mathbf{i}}_m \in \mathbb{R}^M$, and during each forward pass through the network, the random noise $\hat{\delta}_m \sim \mathcal{N}(\mathbf{0}, \hat{\mathbf{C}}_m)$ is added to $\hat{\mathbf{i}}_m$. To simplify the computation, we approximate the dimensionality-reduced noise covariance matrix $\hat{\mathbf{C}}_m$ by a diagonal matrix, i.e., $\hat{\mathbf{C}}_m \approx \text{diag}[\hat{C}_{mkk}]_{k=1}^M$, where \hat{C}_{mij} are the entries of $\hat{\mathbf{C}}_m$. Through a numerical analysis we found that for $M = 14$, $E_M < 1.6 \times 10^{-3}$; thus, 14 principal components appears to be sufficient for aerosol retrieval. The corresponding analysis results for the first 14 principal components are shown in Fig. 2.

It should be pointed out that the number of principal components M can be also determined by reducing the measurement noise [29]. In this case, the reconstruction error $E_M = \sum_{n=1}^{N_t} \|(\mathbf{i}_m^{(n)} - \langle \mathbf{i}_m \rangle) - \mathbf{U}_M \mathbf{U}_M^\dagger (\mathbf{i}_m^{(n)} - \langle \mathbf{i}_m \rangle)\|_2^2$ is monitored, and the M that minimizes E_M is chosen.

IV. RESULTS AND DISCUSSION

In this section we analyze the retrieval performances of the neural network retrieval algorithms using the synthetic and real TROPOMI data.

A. Synthetic retrieval

To test the performances of the retrieval algorithms on synthetic data, we consider a prediction or a test set consisting of $N_p = 11\,868$ samples

$$(\tau^{(n)}, H^{(n)}, \theta_0^{(n)}, \theta^{(n)}, \Delta\varphi^{(n)}, H_s^{(n)}, A_s^{(n)}), \quad (19)$$

chosen randomly in their assumed intervals of variation. For the forward-operator neural network equipped with a Bayesian approach, the initial and a priori values for the aerosol optical depth and layer height are 2 and 2 km, respectively.

To interpret the results, we split the interval of variation of x , $[x_{\min}, x_{\max}]$, where x stands for τ and H , into $N_b = 40$ equidistant bins, i.e., $[x_{\min}, x_{\max}] = \cup_{j=1}^{N_b} B_{xj}$, and compute the (bin) mean

$$\mathbb{E}_j(x_{\text{pred}}) = \frac{1}{N_{xj}} \sum_{n, \text{ s.t. } x^{(n)} \in B_{xj}} x_{\text{pred}}^{(n)} \quad (20)$$

and standard deviation

$$\sqrt{\mathbb{E}_j([x_{\text{pred}} - \mathbb{E}_j(x_{\text{pred}})]^2)} = \sqrt{\frac{1}{N_{xj}} \sum_{n, \text{ s.t. } x^{(n)} \in B_{xj}} [x_{\text{pred}}^{(n)} - \mathbb{E}_j(x_{\text{pred}})]^2}, \quad (21)$$

over all N_{xj} samples $x_{\text{pred}}^{(n)}$, whose corresponding $x^{(n)}$ are in B_{xj} . To quantify the retrieval accuracy we use the first two moments of the absolute error over the prediction set $\Delta_x = x_{\text{pred}} - x$, where x_{pred} and x are the predicted and true values, respectively. These are the mean absolute error

$$\mathbb{E}(|\Delta_x|) = \frac{1}{N_p} \sum_{n=1}^{N_p} |\Delta_x^{(n)}| \quad (22)$$

and the standard deviation of the absolute error

$$\sqrt{\mathbb{E}([\Delta_x - \mathbb{E}(\Delta_x)]^2)} = \sqrt{\frac{1}{N_p} \sum_{n=1}^{N_p} [\Delta_x^{(n)} - \mathbb{E}(\Delta_x)]^2}. \quad (23)$$

In Fig. 3, the mean $\mathbb{E}_j(x_{\text{pred}})$ and standard deviation $\sqrt{\mathbb{E}_j([x_{\text{pred}} - \mathbb{E}_j(x_{\text{pred}})]^2)}$ are plotted versus the midpoint \bar{x}_j of the j th bin, while in Table II we show the mean absolute error $\mathbb{E}(|\Delta_x|)$ and the standard deviation of the absolute error $\sqrt{\mathbb{E}([\Delta_x - \mathbb{E}(\Delta_x)]^2)}$ over the prediction set. Note that (i) $\sqrt{\mathbb{E}([\Delta_x - \mathbb{E}(\Delta_x)]^2)}$ reproduces the square root of the diagonal elements of the so-called epistemic covariance matrix of all network errors over the prediction set, and (ii) the epistemic uncertainties are large if there are large variations around the mean, e.g., if $\sqrt{\mathbb{E}_j([x_{\text{pred}} - \mathbb{E}_j(x_{\text{pred}})]^2)}$ are large. Also note that non-optimal hyper- and training parameters, as well as, a non-optimal optimization algorithm are the main sources of epistemic or model uncertainty [61]. The following conclusions can be drawn.

- 1) In general, the accuracy is low for small values of the aerosol optical depth τ and layer height H .
- 2) The inverse-operator neural networks with synthetic radiances and the principal components of synthetic radiances have comparable accuracies; these are higher than that of the forward-operator neural network. The

low accuracy of the forward-operator neural network, especially for small value of the aerosol optical depth and layer height, is due to the fact that in this domain, the residual function has several local minima and the global minimum cannot be found by the iteratively regularized Gaussian method (which is a local optimization method).

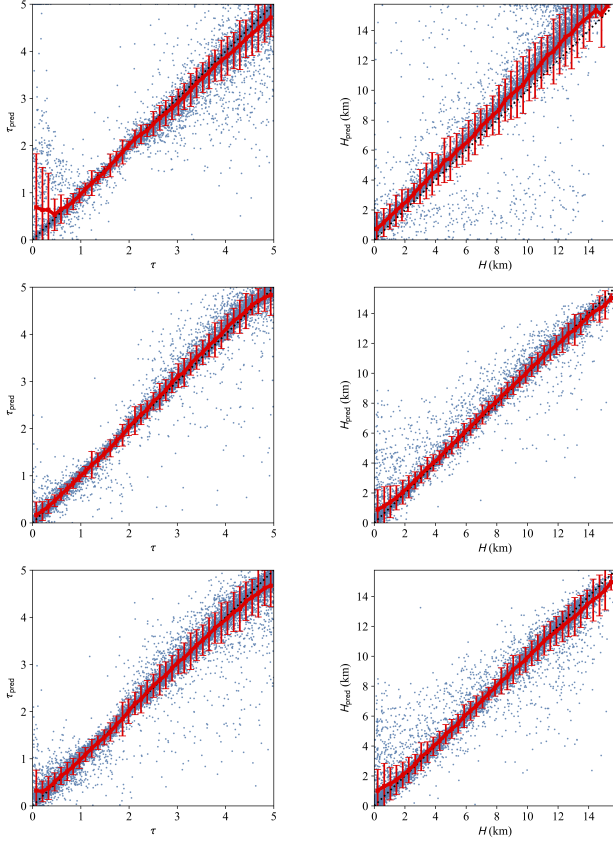


Fig. 3. Predictions of the forward-operator neural network (top row), the inverse-operator neural networks with synthetic radiances (middle row) and the principal-component transform of synthetic radiances (bottom row), respectively. The plots show the predicted values $x_{\text{pred}}^{(n)}$ (blue points) together with the mean $\mathbb{E}_j(x_{\text{pred}})$ (red points) and standard deviation $\sqrt{\mathbb{E}_j(x_{\text{pred}} - \mathbb{E}_j(x_{\text{pred}}))^2}$ (red error bars) over all samples $x_{\text{pred}}^{(n)}$ whose corresponding $x^{(n)}$ are in the j th bin. The value range of x is split into $N_b = 40$ bins, where x stands for the aerosol optical depth τ and aerosol layer height H .

The variations of the absolute error $\Delta_x = x_{\text{pred}} - x$ with respect to the optical and geometrical parameters are illustrated in Figs. 4, 5, and 6. As before, each interval of variation of a parameter b , $[b_{\min}, b_{\max}]$ is split into $N_b = 40$ equidistant bins, i.e., $[b_{\min}, b_{\max}] = \bigcup_{j=1}^{N_b} B_{bj}$, and the mean $\mathbb{E}_j(\Delta_x)$ and standard deviation $\sqrt{\mathbb{E}_j([\Delta_x - \mathbb{E}_j(\Delta_x)]^2)}$ over all samples $\Delta_x^{(n)}$, whose corresponding $b^{(n)}$ are in the j th bin B_{bj} , are plotted versus the midpoint \bar{b}_j of the bin. The plots show that

- 1) the standard deviation of the absolute error in the aerosol optical depth τ is large when τ is small, and the solar zenith angle θ_0 , viewing angle θ and surface albedo A_s are large;
- 2) the standard deviation of the absolute error in the aerosol layer height H is large for small values of the aerosol

TABLE II
MEAN ABSOLUTE ERROR $\mathbb{E}(|\Delta_x|)$, AND THE STANDARD DEVIATION OF THE ABSOLUTE ERROR $\sqrt{\mathbb{E}([\Delta_x - \mathbb{E}(\Delta_x)]^2)}$ OVER THE PREDICTION SET. THE RESULTS CORRESPOND TO THE FORWARD-OPERATOR NEURAL NETWORK (METHOD 1), AND THE INVERSE-OPERATOR NEURAL NETWORKS WITH SYNTHETIC RADIANCES (METHOD 2) AND THE PRINCIPAL-COMPONENT TRANSFORM OF SYNTHETIC RADIANCES (METHOD 3).

Method	x	$\mathbb{E}(\Delta_x)$	$\sqrt{\mathbb{E}([\Delta_x - \mathbb{E}(\Delta_x)]^2)}$
1	τ	0.169	0.410
	H	0.879	1.749
2	τ	0.115	0.243
	H	0.336	0.740
3	τ	0.136	0.316
	H	0.437	0.951

optical thickness τ and large values of the surface albedo A_s ;

- 3) the smallest standard deviations correspond to the inverse-operator neural network with synthetic radiances, while the largest correspond to the forward-operator neural network.

For instance in Fig. 4 (the forward-operator neural network), the standard deviation in Δ_τ can be of 1.14 for small values of τ , and of 0.69, 0.64, 0.92 for large values of θ_0 , θ , and A_s , respectively, whereas the standard deviation in Δ_H can be of 5.32 km for small values of τ and of 2.36 km for large values of A_s .

B. Retrieval from real data

To investigate the performances of the retrieval algorithms on real TROPOMI data, we first choose a wild fire scene in California on 12 December 2017. In this case, the surface albedo is given by the GE_LER product [28], and pixels with

- 1) a cloud fraction (taken from the operational TROPOMI cloud product (OCRA/ROCINN) [31]) greater than 0.25, or
- 2) an aerosol absorbing index (taken from the TROPOMI Level-2 AAI product) lower than 1,

are not considered in the retrieval.

The retrieval results for the aerosol optical depth and layer height are illustrated in Fig. 7. With the results by the inverse-operator neural network with synthetic radiances as a reference, Fig. 8 shows the absolute differences in the retrieved aerosol optical depth $\delta_\tau = \tau_{\text{pred}} - \tau_{\text{pred}}^{\text{ref}}$ and aerosol layer height $\delta_H = H_{\text{pred}} - H_{\text{pred}}^{\text{ref}}$ corresponding to the forward-operator neural network and the inverse-operator neural network with the principal-component transform of synthetic radiances. The plots demonstrate that the differences in the retrieved aerosol optical depth δ_τ are smaller than 0.1 over the entire scene, while the absolute differences in the retrieved aerosol layer height δ_H are smaller than 0.4 km.

Finally, we compare the retrieval results between the three neural network based algorithms and the official operational algorithm for the entire year of 2019. From Fig. 9, the following features are apparent.

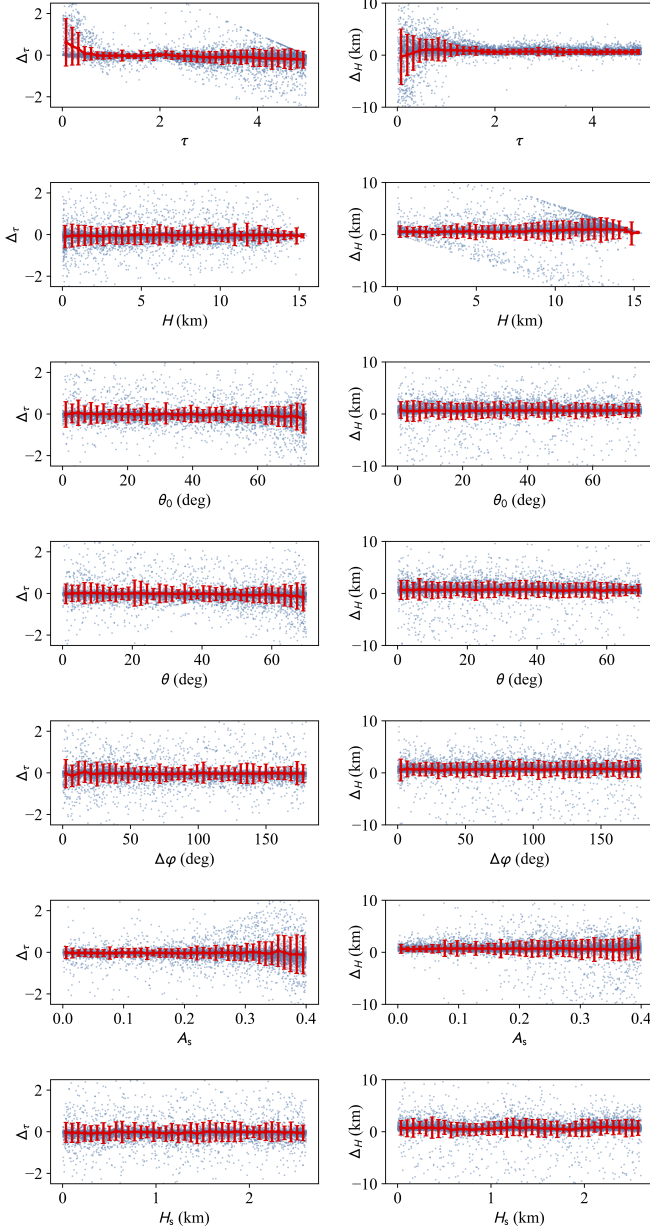


Fig. 4. Absolute error in the retrieved aerosol optical depth $\Delta\tau$ and aerosol layer height ΔH versus the optical and geometrical parameters b , where b stands for the aerosol optical depth τ , aerosol layer height H , solar zenith angle θ_0 , viewing zenith angle θ , relative azimuth angle $\Delta\varphi$, surface albedo A_s , and surface height H_s . The results correspond to the forward-operator neural network. The plots show the absolute error $\Delta_x^{(n)}$ (blue points) together with the mean $\mathbb{E}_j(\Delta_x)$ (red points) and standard deviation $\sqrt{\mathbb{E}_j([\Delta_x - \mathbb{E}_j(\Delta_x)]^2)}$ (red error bars) over all samples $\Delta_x^{(n)}$, whose corresponding $b^{(n)}$ are in the j th bin. The interval of variation of each parameter b is split into $N_b = 40$ bins.

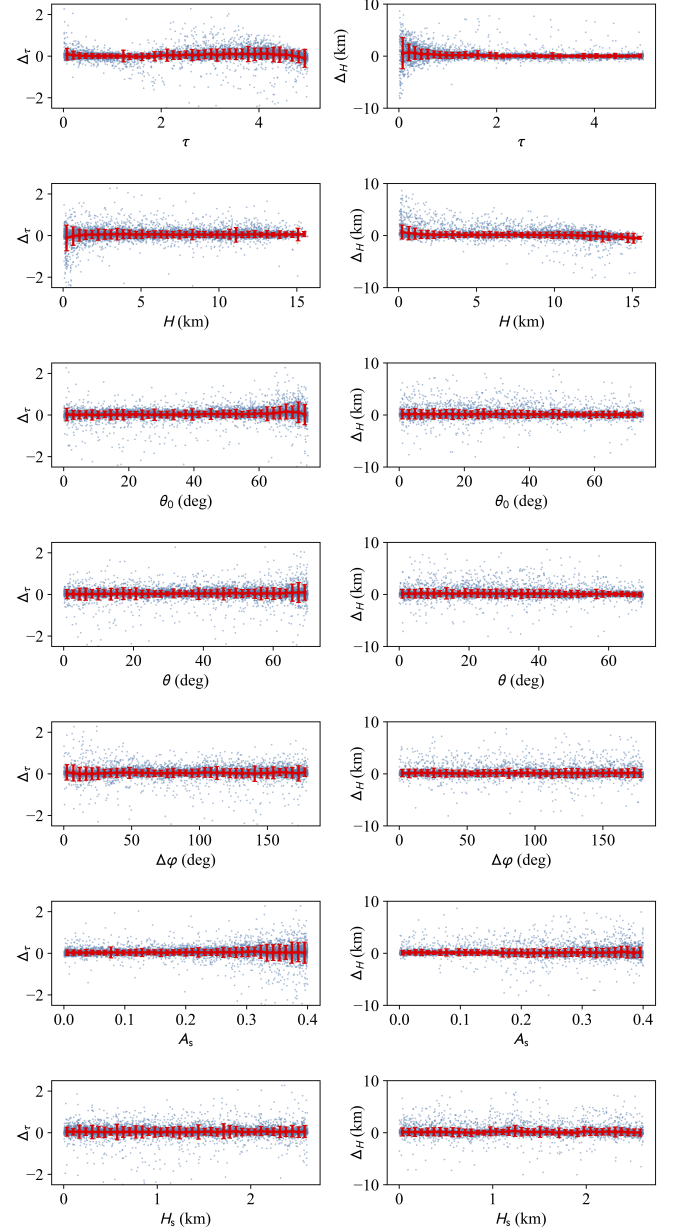


Fig. 5. The same as in Fig. 4 but for the inverse-operator neural network with synthetic radiances.

- 1) The mean values of the aerosol optical depth delivered by the three neural network algorithms are in general underestimated. The reason for this discrepancy is that the official retrieval algorithm uses different aerosol microphysical properties (a fact also seen in the previous study [8]).
- 2) The mean values of the aerosol layer height delivered by the three neural network algorithms agree well with those by the operational retrieval algorithm. However, as expected, a better agreement can be seen between the forward-operator neural network and the operational algorithm whose forward model adopts the neural network approach [35].

It should be pointed out that the computational time of a

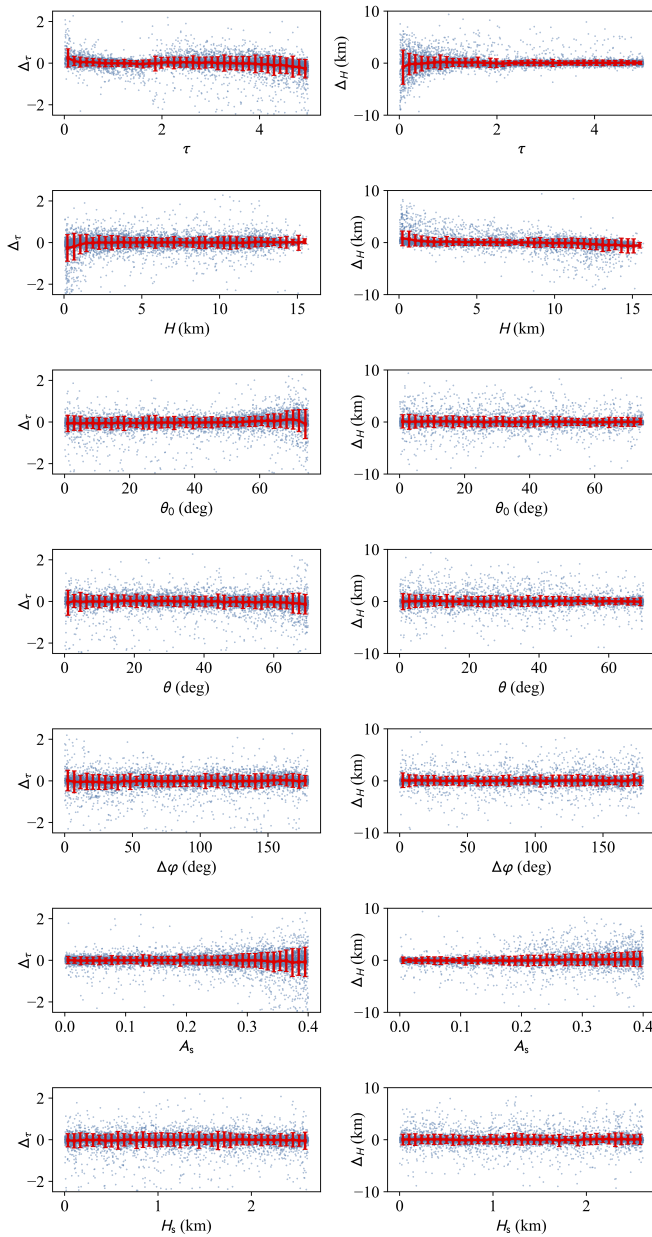


Fig. 6. The same as in Fig. 4 but for the inverse-operator neural network with the principal-component transform of synthetic radiances.

physic-based retrieval algorithm using online radiative transfer calculations is approximately 3 minutes for one ground pixel on a computer Intel Core i7-4770 CPU 3.40GHz with 16 GB RAM, while the computational time of a forward-and an inverse-operator neural networks are 0.4 s and 0.003 s, respectively. For these calculations, the Bayesian approach typically converges in less than five iterations.

Likewise, we perform a comparison with the weighted aerosol heights derived from the CALIPSO (Cloud-Aerosol LIDAR Infrared Pathfinder Satellite Observations) Level-2 aerosol extinction profile product for the entire year of 2019. The weighted aerosol height is computed as a linear combination of the heights above the sea level weighted by the corresponding extinction coefficients. Fig. 10 depicts

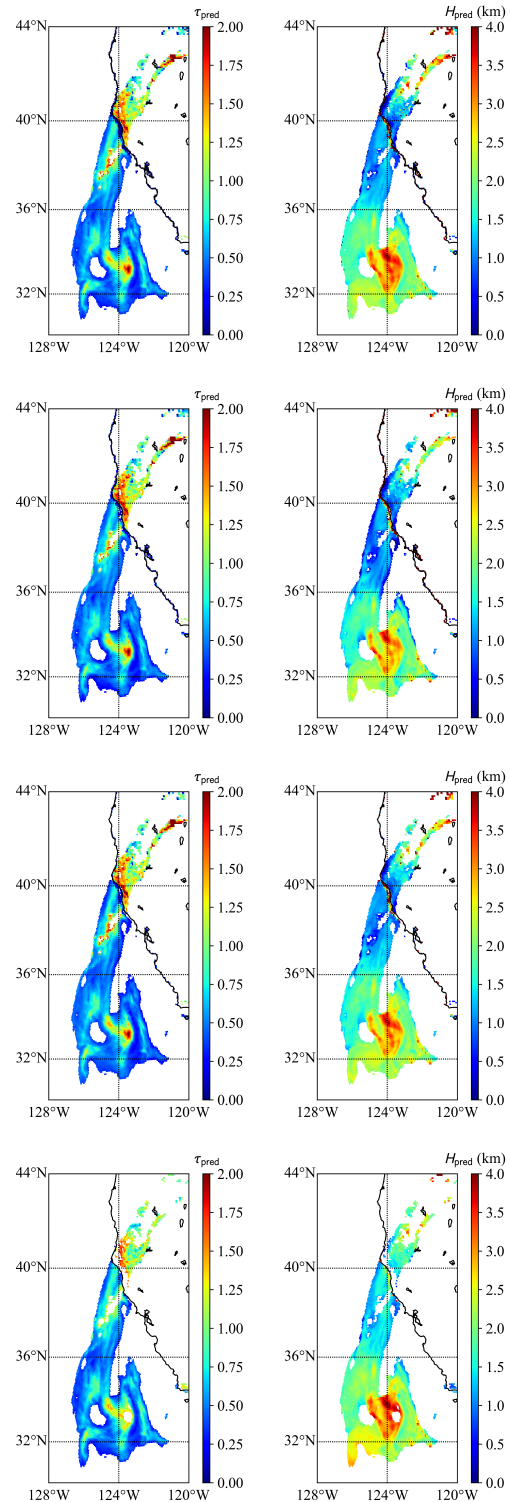


Fig. 7. Retrieved aerosol optical depth τ_{pred} and aerosol layer height H_{pred} from TROPOMI measurements recorded on 12 December 2017 in California. The results correspond to the forward-operator neural network (the first row), the inverse-operator neural network with synthetic radiances (the second row), the inverse-operator neural network with the principal-component transform of synthetic radiances (the third row), and the TROPOMI official operational product (the fourth row).

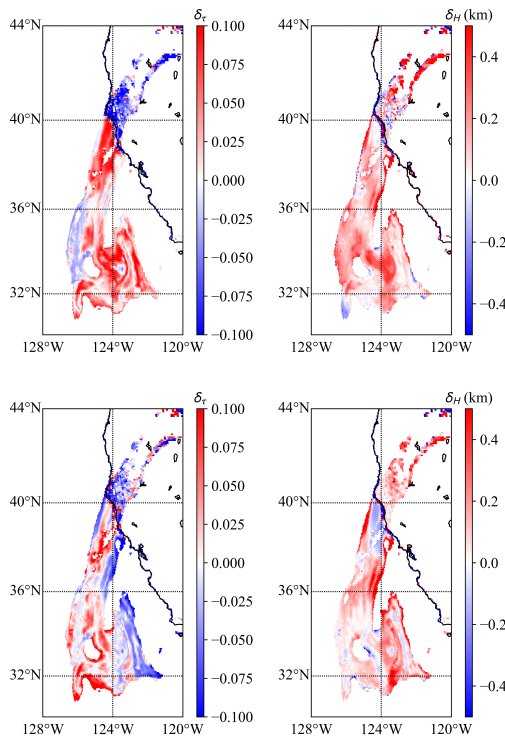


Fig. 8. Absolute differences in the retrieved aerosol optical depth δ_τ and aerosol layer height δ_H corresponding to the forward-operator neural network (upper panels) and the inverse-operator neural network with the principal-component transform of synthetic radiances (lower panels). The results provided by the inverse-operator neural network with synthetic radiances are taken as a reference.

the histograms of the differences in aerosol layer height (TROPOMI minus CALIPSO), where a TROPOMI value is the mean value over all TROPOMI pixels within a distance of 100 km to a single CALIPSO pixel. The difference $E(\Delta_H) \pm \sqrt{E([\Delta_H - E(\Delta_H)]^2)}$ is highest for the forward operator neural network (1.54 ± 1.62 km), whereas a better agreement with the CALIPSO product is apparently achieved by the inverse operator with synthetic radiances (1.38 ± 1.47 km) and the inverse operator with the principal-component transform of synthetic radiances (1.48 ± 1.43 km).

V. CONCLUSIONS

In this paper, we have developed three neural network algorithms for aerosol retrieval from TROPOMI measurements in the O_2 A-band.

- 1) The first algorithm uses a neural network to emulate the radiative transfer model and a Bayesian approach to solve the inverse problem. To speed up the computation, the radiative transfer model combines standard acceleration methods (the telescoping technique and the method of false discrete ordinate) with hyperspectral acceleration methods (the correlated k-distribution method and the principal component analysis). The inverse problem is solved by using the iteratively regularized Gauss-Newton method, which provides at the same time the optimal value of the regularization parameter and the corresponding regularized solution.

- 2) The second and third algorithms employ a neural network to learn the inverse model and use as input either the synthetic radiances computed on the measurement wavelength grid or the principal-component transform of synthetic radiances. The design of an inverse-operator neural network for TROPOMI/S5P is not a trivial task because, on the one hand, we are dealing with random measurement noise and, on the other hand, there are a large number of measurement wavelength grids corresponding to each swath row. To solve this problem, we used the jitter approach. More precisely, in the training stage and at each forward pass through the network, a measurement wavelength grid is randomly selected from a set of possible wavelength grids, and a new random noise is added to the synthetic radiance. Note that in the algorithm relying on the principal-component analysis, the random noise is described through an analytic dimensionality-reduced noise covariance matrix.

The neural networks are incorporated into a common tool and their performances are analyzed and compared with synthetic and real data. Our numerical analysis has shown that the inverse-operator neural networks are more accurate and efficient than a forward-operator neural network. The reason is that for small values of the aerosol optical depth and layer height, the residual function has several local minima, and in the latter case, the inversion method which is a local optimization method, cannot determine the global minimum. These results may suggest that a TROPOMI operational retrieval algorithm can be built on an inverse-operator neural network rather than on a forward-operator neural network.

The design and refinement of neural networks for atmospheric retrieval is a very complicated research field that requires more developments that consist of

- 1) application of the inverse-operator neural networks to the remaining aerosol models considered in the MODIS algorithm, i.e., non-absorbing, absorbing, and desert dust (the selection of an appropriate aerosol model is then based on a combination of spectral and geographic information);
- 2) training the neural networks to learn the relative evidences of different aerosol models, so that, a mean solution estimate, representing a linear combination of candidate solutions weighted by their evidences, can be computed [62];
- 3) redesign of the neural networks in a Bayesian deep learning framework in order to predict input aleatoric and model uncertainties [61].

ACKNOWLEDGMENTS

The authors are grateful to DLR and KNMI colleagues for sharing the relevant data and information of TROPOMI/S5P. Thanks to EU/ESA/KNMI/DLR for providing the TROPOMI/S5P Level-1b data.

REFERENCES

- [1] G. A. A. Koppers and D. P. Murtagh, "Retrieval of height resolved aerosol optical thickness in the atmospheric

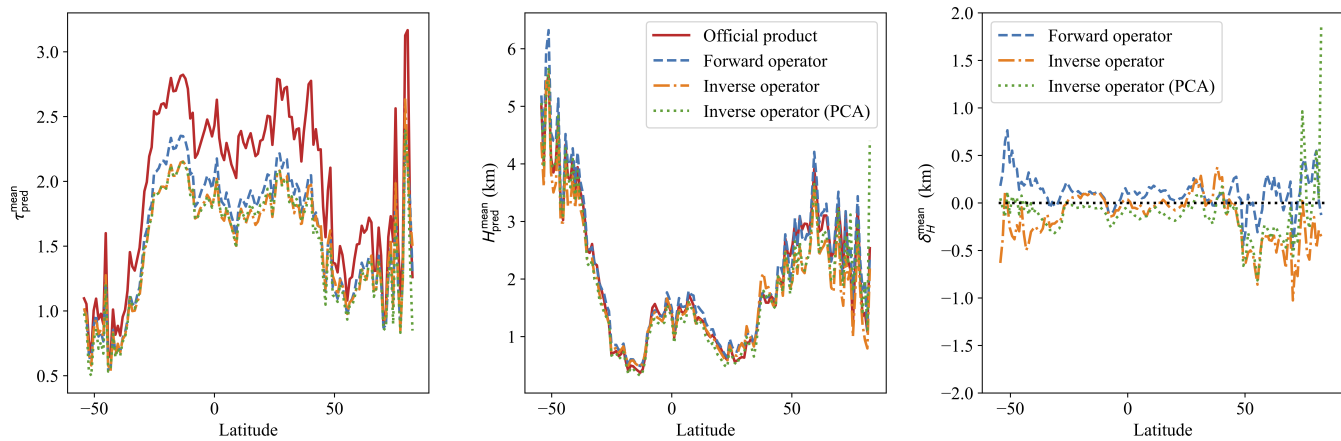


Fig. 9. The mean values of aerosol optical depth $\tau_{\text{pred}}^{\text{mean}}$ and aerosol layer height $H_{\text{pred}}^{\text{mean}}$ for the three neural network based algorithms and the official operational product. The mean differences in aerosol layer height against the official operational product are plotted as δ_H^{mean} . The comparison is done for the entire year of 2019.

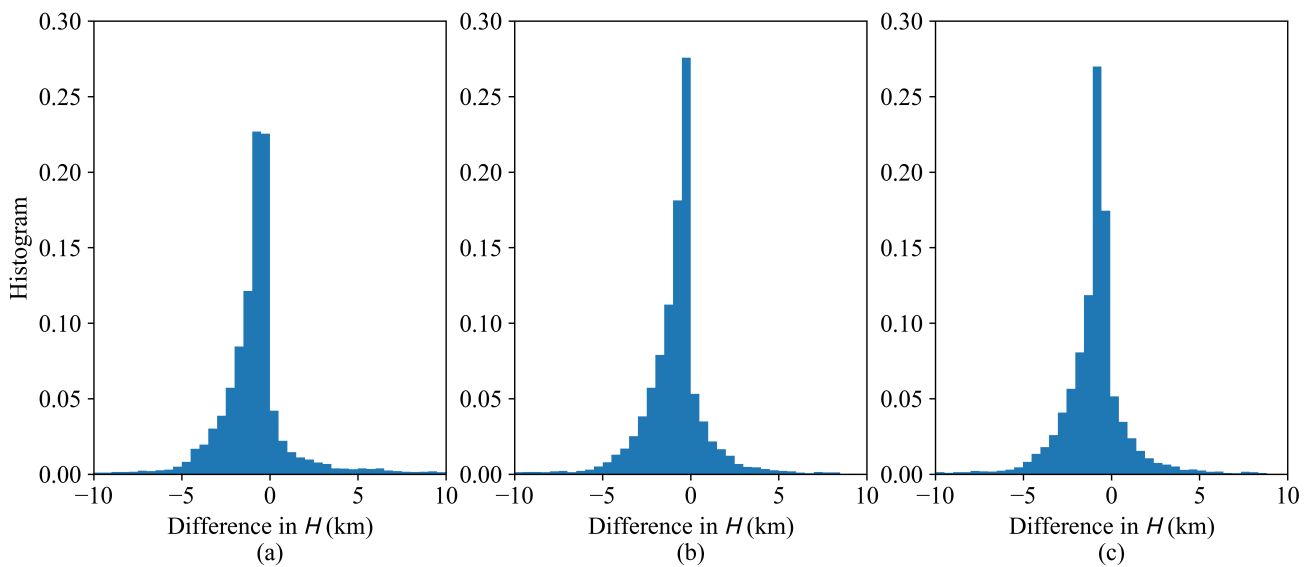


Fig. 10. Histogram of the differences in aerosol layer height H between the TROPOMI and CALIPSO products. The TROPOMI retrievals are generated by the forward-operator neural network (a), the inverse-operator neural networks with synthetic radiances (b) and the principal-component transform of synthetic radiances (c).

- band.” Stockholm, Sweden: Stockholm University, 1997.
- [2] L. Tilstra, O. Tuinder, P. Wang, and P. Stammes, “GOME-2 absorbing aerosol height algorithm theoretical basis document,” Royal Netherlands Meteorological Institute (KNMI), Tech. Rep., 2019.
 - [3] S. Corradini and M. Cervino, “Aerosol extinction coefficient profile retrieval in the oxygen A-band considering multiple scattering atmosphere. Test case: SCIAMACHY nadir simulated measurements,” *J. Quant. Spectrosc. Radiat. Transf.*, vol. 97, no. 3, pp. 354–380, 2006.
 - [4] A. A. Kokhanovsky and V. V. Rozanov, “The determination of dust cloud altitudes from a satellite using hyperspectral measurements in the gaseous absorption band,” *Int. J. Remote Sensing*, vol. 31, no. 10, pp. 2729–2744, 2010.
 - [5] S. Sanghavi, J. V. Martonchik, J. Landgraf, and U. Platt, “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,” *Atmos. Meas. Tech.*, vol. 5, no. 5, pp. 1099–1119, 2012.
 - [6] C. Frankenberg, O. Hasekamp, C. O’Dell, S. Sanghavi, A. Butz, and J. Worden, “Aerosol information content analysis of multi-angle high spectral resolution measurements and its benefit for high accuracy greenhouse gas retrievals,” *Atmos. Meas. Tech.*, vol. 5, no. 7, pp. 1809–1821, 2012.
 - [7] S. Nanda, M. d. Graaf, J. P. Veefkind, M. Sneep, M. t. Linden, J. Sun, and P. F. Levelt, “A first comparison of TROPOMI aerosol layer height (ALH) to CALIOP data,” *Atmospheric Measurement Techniques*, vol. 13, no. 6, pp.

- 3043–3059, 2020.
- [8] L. Rao, J. Xu, D. S. Efremenko, D. G. Loyola, and A. Doicu, “Hyperspectral satellite remote sensing of aerosol parameters: Sensitivity analysis and application to TROPOMI/SSP,” *Frontiers in Environmental Science*, p. 744, 2022.
- [9] J. Xu, F. Schreier, A. Doicu, and T. Trautmann, “Assessment of Tikhonov-type regularization methods for solving atmospheric inverse problems,” *J. Quant. Spectrosc. Radiat. Transf.*, vol. 184, pp. 274–286, 2016.
- [10] J. Xu, L. Rao, F. Schreier, D. S. Efremenko, A. Doicu, and T. Trautmann, “Insight into construction of Tikhonov-type regularization for atmospheric retrievals,” *Atmosphere*, vol. 11, no. 10, 2020, 1052.
- [11] C. D. Rodgers, *Inverse methods for atmospheric sounding: Theory and practice*. World scientific, 2000, vol. 2.
- [12] V. Arnold, “On the representation of functions of several variables as a superposition of functions of a smaller number of variables,” *Mathematics, its teaching, applications and history, Matem. prosv.*, ser. 2, vol. 3, pp. 41–61, 1958.
- [13] G. Cybenko, “Approximation by superpositions of a sigmoidal function,” *Mathematics of Control, Signals, and Systems*, vol. 2, no. 4, pp. 303–314, 1989.
- [14] Y. Fan, W. Li, C. K. Gatebe, C. Jamet, G. Zibordi, T. Schroeder, and K. Stamnes, “Atmospheric correction over coastal waters using multilayer neural networks,” *Remote Sensing of Environment*, vol. 199, pp. 218–240, 2017.
- [15] C. Fan, G. Fu, A. Di Noia, M. Smit, J. H. H. Rietjens, R. A. Ferrare, S. Burton, Z. Li, and O. P. Hasekamp, “Use of a neural network-based ocean body radiative transfer model for aerosol retrievals from multi-angle polarimetric measurements,” *Remote Sensing*, vol. 11, no. 23, p. 2877, 2019.
- [16] C. Shi, M. Hashimoto, K. Shiomi, and T. Nakajima, “Development of an algorithm to retrieve aerosol optical properties over water using an artificial neural network radiative transfer scheme: First result from GOSAT-2/CAI-2,” *IEEE Transactions on Geoscience and Remote Sensing*, 2020.
- [17] M. Gao, B. A. Franz, K. Knobelspiesse, P.-W. Zhai, V. Martins, S. Burton, B. Cairns, R. Ferrare, J. Gales, O. Hasekamp *et al.*, “Efficient multi-angle polarimetric inversion of aerosols and ocean color powered by a deep neural network forward model,” *Atmospheric Measurement Techniques*, vol. 14, no. 6, pp. 4083–4110, 2021.
- [18] D. G. Loyola R, “Applications of neural network methods to the processing of earth observation satellite data,” *Neural networks*, vol. 19, no. 2, pp. 168–177, 2006.
- [19] D. S. Efremenko, “Discrete ordinate radiative transfer model with the neural network based eigenvalue solver: Proof of concept,” *Light & Engineering*, vol. 01, pp. 56–62, 2021.
- [20] C. Jiménez, P. Eriksson, and D. Murtagh, “Inversion of Odin limb sounding submillimeter observations by a neural network technique,” *Radio Science*, vol. 38, no. 4, pp. 27–1, 2003.
- [21] G. Holl, S. Eliasson, J. Mendrok, and S. Buehler, “SPARE-ICE: Synergistic ice water path from passive operational sensors,” *Journal of Geophysical Research: Atmospheres*, vol. 119, no. 3, pp. 1504–1523, 2014.
- [22] J. Strandgren, L. Bugliaro, F. Sehnke, and L. Schröder, “Cirrus cloud retrieval with MSG/SEVIRIs using artificial neural networks,” *Atmospheric Measurement Techniques*, vol. 10, no. 9, pp. 3547–3573, 2017.
- [23] D. Wang, C. Prigent, F. Aires, and C. Jimenez, “A statistical retrieval of cloud parameters for the millimeter wave Ice Cloud Imager on board MetOp-SGs,” *IEEE Access*, vol. 5, pp. 4057–4076, 2017s.
- [24] M. Brath, S. Fox, P. Eriksson, R. C. Harlow, M. Burgdorf, and S. A. Buehler, “Retrieval of an ice water path over the ocean from ISMAR and MARSS millimeter and submillimeter brightness temperatures,” *Atmospheric Measurement Techniques*, vol. 11, no. 1, pp. 611–632, 2018.
- [25] N. Håkansson, C. Adok, A. Thoss, R. Scheirer, and S. Hörnquist, “Neural network cloud top pressure and height for MODIS,” *Atmospheric Measurement Techniques*, vol. 11, no. 5, pp. 3177–3196, 2018.
- [26] D. S. Efremenko, D. G. Loyola R., P. Hedelt, and R. J. D. Spurr, “Volcanic SO₂ plume height retrieval from UV sensors using a full-physics inverse learning machine algorithm,” *Int. J. Remote Sensing*, vol. 38, no. sup1, pp. 1–27, 2017.
- [27] J. Xu, O. Schüssler, D. Loyola R, F. Romahn, and A. Doicu, “A novel ozone profile shape retrieval using Full-Physics Inverse Learning Machine (FP-ILM),” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 12, pp. 5442–5457, 2017.
- [28] D. G. Loyola, J. Xu, K.-P. Heue, and W. Zimmer, “Applying FP_ILM to the retrieval of Geometry-dependent Effective Lambertian Equivalent Reflectivity (GE_LER) daily maps from UVN satellite measurements,” *Atmos. Meas. Tech.*, vol. 13, no. 2, pp. 985–999, 2020.
- [29] A. Di Noia, O. Hasekamp, G. Van Harten, J. Rietjens, J. Smit, F. Snik, J. Henzing, J. De Boer, C. Keller, and H. Volten, “Use of neural networks in ground-based aerosol retrievals from multi-angle spectropolarimetric observations,” *Atmospheric Measurement Techniques*, vol. 8, no. 1, p. 281, 2015.
- [30] A. D. Noia, O. P. Hasekamp, L. Wu, B. v. Diedenhoven, B. Cairns, and J. E. Yorks, “Combined neural network/Phillips–Tikhonov approach to aerosol retrievals over land from the NASA Research Scanning Polarimeter,” *Atmospheric Measurement Techniques*, vol. 10, no. 11, pp. 4235–4252, 2017.
- [31] D. G. Loyola, S. Gimeno García, R. Lutz, A. Argyrouli, F. Romahn, R. J. D. Spurr, M. Pedernana, A. Doicu, V. Molina García, and O. Schüssler, “The operational cloud retrieval algorithms from TROPOMI on board Sentinel-5 Precursor,” *Atmos. Meas. Tech.*, vol. 11, no. 1, pp. 409–427, 2018.
- [32] S. Compennolle, A. Argyrouli, R. Lutz, M. Sneep, J.-C. Lambert, A. M. Fjæraa, D. Hubert, A. Keppens, D. Loyola, E. O’Connor *et al.*, “Validation of the Sentinel-5

- Precursor TROPOMI cloud data with Cloudnet, Aura OMI O2–O2, MODIS, and Suomi-NPP VIIR,” *Atmos. Meas. Tech.*, vol. 14, pp. 2451–2476, 2021.
- [33] A. F. J. Sanders and J. F. de Haan, “Retrieval of aerosol parameters from the oxygen A band in the presence of chlorophyll fluorescence,” *Atmos. Meas. Tech.*, vol. 6, no. 10, pp. 2725–2740, 2013.
- [34] A. F. J. Sanders, J. F. de Haan, M. Sneep, A. Apituley, P. Stammes, M. O. Vieitez, L. G. Tilstra, O. N. E. Tuinder, C. E. Koning, and J. P. Veefkind, “Evaluation of the operational Aerosol Layer Height retrieval algorithm for Sentinel-5 Precursor: application to O_2 A band observations from GOME-2A,” *Atmos. Meas. Tech.*, vol. 8, no. 11, pp. 4947–4977, 2015.
- [35] S. Nanda, M. de Graaf, J. P. Veefkind, M. ter Linden, M. Sneep, J. de Haan, and P. F. Levelt, “A neural network radiative transfer model approach applied to the TROPospheric Monitoring Instrument aerosol height algorithm,” *Atmospheric Measurement Techniques*, vol. 12, no. 12, pp. 6619–6634, 2019.
- [36] A. Doicu and T. Trautmann, “Discrete-ordinate method with matrix exponential for a pseudo-spherical atmosphere: Scalar case,” *J. Quant. Spectrosc. Radiat. Transf.*, vol. 110, pp. 146–158, 2009.
- [37] D. S. Efremenko, V. Molina García, S. Gimeno García, and A. Doicu, “A review of the matrix-exponential formalism in radiative transfer,” *Journal of Quantitative Spectroscopy and Radiative Transfer*, vol. 196, pp. 17–45, 2017.
- [38] R. Spurr, “LIDORT and VLIDORT: Linearized pseudo-spherical scalar and vector discrete ordinate radiative transfer models for use in remote sensing retrieval problems,” A. Kokhanovsky, Ed. Berlin: Springer Verlag, 2008, vol. 3, pp. 229–271.
- [39] D. Efremenko, A. Doicu, D. Loyola, and T. Trautmann, “Acceleration techniques for the discrete ordinate method,” *J. Quant. Spectrosc. Radiat. Transf.*, vol. 114, pp. 73–81, 2013.
- [40] E. Chalhoub and R. Garcia, “The equivalence between two techniques of angular interpolation for the discrete-ordinates method,” *Journal of Quantitative Spectroscopy and Radiative Transfer*, vol. 64, no. 5, pp. 517–535, 2000.
- [41] R. Goody, R. West, L. Chen, and D. Crisp, “The correlated-k method for radiation calculations in nonhomogeneous atmospheres,” *J. Quant. Spectrosc. Radiat. Transf.*, vol. 42, no. 6, pp. 437–651, 1989.
- [42] V. Natraj, R.-L. Shia, X. Huang, J. S. Margolis, and Y. L. Yung, “Application of principal component analysis to high spectral resolution radiative transfer: A case study of the O_2 A band,” *J. Quant. Spectrosc. Radiat. Transf.*, vol. 95, no. 4, pp. 539–556, 2005.
- [43] V. Natraj, R.-L. Shia, and Y. L. Yung, “On the use of principal component analysis to speed up radiative transfer calculations,” *J. Quant. Spectrosc. Radiat. Transf.*, vol. 111, no. 3, pp. 810–816, 2010.
- [44] D. Efremenko, A. Doicu, D. Loyola, and T. Trautmann, “Optical property dimensionality reduction techniques for accelerated radiative transfer performance: Application to remote sensing total ozone retrievals,” *Journal of Quantitative Spectroscopy and Radiative Transfer*, vol. 133, pp. 128–135, 2014.
- [45] W. Hou, J. Wang, X. Xu, J. S. Reid, and D. Han, “An algorithm for hyperspectral remote sensing of aerosols: 1. development of theoretical framework,” *J. Quant. Spectrosc. Radiat. Transf.*, vol. 178, pp. 400–415, 2016.
- [46] V. Molina García, S. Sasi, D. S. Efremenko, A. Doicu, and D. Loyola, “Radiative transfer models for retrieval of cloud parameters from EPIC/DSCOVR measurements,” *J. Quant. Spectrosc. Radiat. Transf.*, vol. 213, pp. 228–240, 2018.
- [47] F. Schreier, S. G. Garcia, P. Hedelt, M. Hess, J. Mendrok, M. Vasquez, and J. Xu, “GARLIC—A general purpose atmospheric radiative transfer line-by-line infrared-microwave code: Implementation and evaluation,” *Journal of Quantitative Spectroscopy and Radiative Transfer*, vol. 137, pp. 29–50, 2014.
- [48] F. Schreier, “Optimized implementations of rational approximations for the voigt and complex error function,” *Journal of Quantitative Spectroscopy and Radiative Transfer*, vol. 112, no. 6, pp. 1010–1025, 2011.
- [49] L. S. Rothman, I. E. Gordon, A. Barbe, D. C. Benner, P. F. Bernath, M. Birk, V. Boudon, L. R. Brown, A. Campargue, J.-P. Champion *et al.*, “The HITRAN 2008 molecular spectroscopic database,” *Journal of Quantitative Spectroscopy and Radiative Transfer*, vol. 110, no. 9–10, pp. 533–572, 2009.
- [50] B. A. Bodhaine, N. B. Wood, E. G. Dutton, and J. R. Slusser, “On rayleigh optical depth calculations,” *Journal of Atmospheric and Oceanic Technology*, vol. 16, no. 11, pp. 1854–1861, 1999.
- [51] W. Wiscombe, “The delta-M method: Rapid yet accurate radiative flux calculations for strongly asymmetric phase functions,” *Journal of Atmospheric Sciences*, vol. 34, no. 9, pp. 1408–1422, 1977.
- [52] T. Nakajima and M. Tanaka, “Algorithms for radiative intensity calculations in moderately thick atmospheres using a truncation approximation,” *Journal of Quantitative Spectroscopy and Radiative Transfer*, vol. 40, no. 1, pp. 51–69, 1988.
- [53] M. Momoi, H. Irie, T. Nakajima, and M. Sekiguchi, “Efficient calculation of radiative intensity including the polarization effect in moderately thick atmospheres using a truncation approximation,” *J. Quant. Spectrosc. Radiat. Transf.*, vol. 277, p. 107976, 2022.
- [54] S. Sasi, V. Natraj, V. Molina García, D. S. Efremenko, D. Loyola, and A. Doicu, “Model selection in atmospheric remote sensing with an application to aerosol retrieval from DSCOVR/EPIC, Part 1: Theory,” *Remote Sens.*, vol. 12, no. 22, 2020, 3724.
- [55] S. Sasi, V. Natraj, V. Molina García, D. S. Efremenko, D. Loyola, and A. Doicu, “Model selection in atmospheric remote sensing with application to aerosol retrieval from DSCOVR/EPIC. Part 2: Numerical analysis,” *Remote Sens.*, vol. 12, no. 21, 2020, 3656.
- [56] R. C. Levy, L. A. Remer, and O. Dubovik, “Global

aerosol optical properties and application to Moderate Resolution Imaging Spectroradiometer aerosol retrieval over land,” *J. Geophys. Res. Atmos.*, vol. 112, no. D13, pp. n/a–n/a, 2007.

- [57] D. G. Loyola, M. Pedernana, and S. Gimeno Garcia, “Smart sampling and incremental function learning for very large high dimensional data,” *Neural Networks*, vol. 78, pp. 75–87, 2016.
- [58] J. H. Halton, “Algorithm 247: Radical-inverse quasi-random point sequence,” *Communications of the ACM*, vol. 7, no. 12, pp. 701–702, 1964.
- [59] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” 2017.
- [60] A. Doicu, T. Trautmann, and F. Schreier, *Numerical regularization for atmospheric inverse problems*. Springer Science & Business Media, 2010.
- [61] A. Doicu, A. Doicu, D. S. Efremenko, D. Loyola, and T. Trautmann, “An overview of neural network methods for predicting uncertainty in atmospheric remote sensing,” *Remote Sensing*, vol. 13, no. 24, p. 5061, 2021.
- [62] L. Rao, J. Xu, D. S. Efremenko, D. G. Loyola, and A. Doicu, “Optimization of aerosol model selection for TROPOMI/S5P,” *Remote Sensing*, vol. 13, no. 13, p. 2489, 2021.



Lanlan Rao received the Master degree in cartography and geographic information engineering from China University of Mining and Technology, Xuzhou, China, in 2017. She is pursuing the Ph.D. degree with the German Aerospace Center (DLR), Remote Sensing Technology Institute (IMF), Weßling, Germany and the Technische Universität München (TUM), Munich, Germany. Her research interests focus on remote sensing of aerosol properties from satellite data.



Jian Xu (Member, IEEE) received the B.E. degree in geographic information systems from Hohai University, Nanjing, Jiangsu, China, in 2004, and the M.S. degree in Earth-oriented space science and technology and the Dr.-Ing. degree in atmospheric remote sensing from the Technische Universität München, Munich, Germany, in 2009 and 2015, respectively.

From 2010 to 2021, he was with the Remote Sensing Technology Institute (IMF), German Aerospace Center (DLR), Oberpfaffenhofen, Weßling, Germany.

His was involved in the development/refinement of atmospheric retrieval algorithms for ESA's Sentinel-5P and Sentinel-4 satellite missions. Since 2021 he has been with the National Space Science Center (NSSC), Chinese Academy of Sciences (CAS), Beijing, China. His research interests include remote sensing of planetary atmosphere, radiative transfer modeling, and ill-posed inverse problems.



deconvolution problems, and machine learning. Since 2018 he leads a junior research group “Atmospheric Retrieval of the Next Generation”.

Dmitry S. Efremenko graduated from the Moscow Power Engineering Institute (MPEI) in 2009. He received the PhD degree from the Moscow State University in 2011 and the Habilitation degree from MPEI in 2017. Since 2011 he works as a research scientist at the German Aerospace Center (DLR). He gives academic courses “Radiative Transfer” and “Non-linear Optimization” at the Technische Universität München, Munich, Germany. He received the Elsevier/IQSRT Goody Award in 2017. His scientific interests include radiative transfer, remote sensing,



Diego G. Loyola (Senior Member, IEEE) received the Lic. Inf. degree in computer science from ES-LAI, Universidad de Lujan in Argentina and the Dr.-Ing. degree in remote sensing and machine learning from the Technische Universität München, Munich, Germany.

Since 1990 he has been a Scientist and Project Manager with the German Aerospace Center (Deutsches Zentrum für Luft- und Raumfahrt - DLR) in Oberpfaffenhofen, Germany, and a Visiting Scientist with the NASA Goddard Space Flight Center in Maryland, USA in 2000, 2016/2017 and 2022/2023. Since 2000, he has been with DLR's Remote Sensing Technology Institute, where he leads the passive sensing group responsible for the operational atmospheric composition and cloud products from the European missions GOME on ERS-2, GOME-2 on MetOp-A/-B and -C, TROPOMI on Sentinel-5 Precursor, and Sentinel-4 on MTG-S. Additionally, he is a Visiting Professor at the School of Physics, Aristotle University of Thessaloniki in Greece. His research interests are with remote sensing of atmospheric trace gases, cloud and aerosol properties, as well as computational intelligence and machine learning techniques.

Dr.-Ing. Loyola holds the DLR title Senior Scientist since 2015, he is winner of the DLR Scientific Award 2011, he is senior member of the IEEE Geoscience and Remote Sensing Society and IEEE Computational Intelligence Society, and senior member of the International Neural Network Society.



Adrian Doicu received the Eng. and Ph.D. degrees in mechanical engineering and optics from the Technical University of Bucharest, Bucharest, Romania, in 1986 and 1996, respectively.

He is currently a Senior Scientist with the Remote Sensing Technology Institute (IMF), the German Aerospace Center (DLR), Oberpfaffenhofen, Germany. Since 2012, he has been appointed as an Associate Professor (German: Privatdozent) in atmospheric remote sensing at the Technische Universität München, Munich, Germany. His research interests

include electromagnetic scattering theory, inverse problems, and radiative transfer.