Applying FP_ILM to the retrieval of geometry-dependent effective Lambertian equivalent reflectivity (GE_LER) to account for BRDF effects on UVN satellite measurements of trace gases, clouds and aerosols

5 Diego G. Loyola[1], Jian Xu[1], Klaus-Peter Heue[1], Walter Zimmer[1]

[1]German Aerospace Centre (DLR), Remote Sensing Technology Institute, Oberpfaffenhofen, 82234 Wessling, Germany

Correspondence to: Diego Loyola (Diego.Loyola@dlr.de)

Abstract. The retrieval of trace gas, cloud and aerosol measurements from ultraviolet, visible and near-infrared (UVN)
10 sensors requires precise information on the surface properties that are traditionally obtained from Lambertian equivalent reflectivity (LER) climatologies. The main drawbacks of using such LER climatologies for new satellite missions are (a) climatologies are typically based on previous missions with a significant lower spatial resolution, (b) they usually do not fully take into account the satellite viewing dependencies characterized by the bidirectional reflectance distribution function (BRDF) effects, and (c) climatologies may differ considerably from the actual surface conditions especially under snow/ice
15 situations.

In this paper we present a novel algorithm for the retrieval of geometry-dependent effective Lambertian equivalent reflectivity (GE_LER) from UVN sensors based on the full-physics inverse learning machine (FP_ILM) retrieval. The radiances are simulated using a radiative transfer model that takes into account the satellite viewing geometry and the inverse problem is solved using machine learning techniques to obtain the GE_LER from satellite measurements.

20 The GE_LER retrieval is optimized for the trace gas retrievals using the DOAS algorithm and the large amount of data of the new atmospheric Sentinel satellite missions. The GE_LER can either be used directly for the computation of AMFs using the effective scene approximation or a global gapless geometry-dependent LER (G3_LER) daily map can be easily created from the GE_LER under clear-sky conditions for the computation of AMFs using the independent pixel approximation.

The FP_ILM GE_LER algorithm is applied to measurements of TROPOMI launched in October 2017 on board the EU/ESA
25 Sentinel-5 Precursor (S5P) mission. The TROPOMI GE_LER/G3_LER results are compared with climatological OMI LER data and the advantages of using GE_LER/G3_LER are demonstrated for the retrieval of total ozone from TROPOMI.

1. Introduction

Atmospheric
Measurement
Techniques
Discussions

Uncertainties about the surface reflectance and not accounting their anisotropic properties are mayor error sources for the retrieval of trace gas, cloud and aerosol information from ultraviolet, visible and near-infrared (UVN) satellites measurements (Vasilkov et al., 2018; Lorente et al., 2018; Lin et al., 2014; Seidel et al., 2012; Zhou et al., 2010). For example errors of 0.02 in the surface reflectivity may induce errors of 10%–20% in SO_2 column (Lee et al., 2009) and

5 seasonal snow cover could change the retrieved NO_2 column by 20%–50% (O'Byrne et al., 2010) and the retrieved O_3 column by 5%–35% (Lerot et al., 2014).

Traditionally, surface properties are obtained from Lambertian equivalent reflectivity (LER) climatologies and in the case of new missions like TROPOMI launched in October 2017 on board the EU/ESA Sentinel-5 Precursor (S5P) mission, the climatologies used at the beginning of the mission are based on LER data from previous missions like TOMS (Herman and

10 Celarier, 1997), GOME (Koelemeijer et al., 2003), OMI (Kleipool et al., 2008), SCIAMACHY(Tilstra et al., 2017), and GOME-2 (Pflug et al., 2008).

The unprecedented spatial resolution of TROPOMI (3.5x7 km^2) clearly showed the disadvantages of using LER climatologies based on previous missions with a significant lower spatial resolution. The initial version of the TROPOMI trace gas products using climatologies show flawed patterns related to the coarse resolution of the OMI LER climatology. A

15 LER climatology based on TROPOMI measurements could solve this particular problem, but creating such new TROPOMI LER climatology will probably require at least two years of data. Furthermore, there are two fundamental problems with typical LER climatologies: (a) the actual surface conditions of a satellite measurement may differ considerably from climatological values like for example under snow/ice situations, and (b) the effect of surface reflectance anisotropy are usually not properly covered by the climatology.

20 Retrieval of effective scene albedo has been used in total ozone algorithms from nadir and limb satellite sensors. The WFDOAS (Coldewey-Egbers et al., 2005) approach retrieves the effective LER at 377 nm, the GODFIT (Lerot et al., 2010) and SAGE III (Raul and Taha, 2007) approaches retrieve simultaneously with ozone the effective LER and other parameters.

Another approach used for NO_2 and cloud retrievals is the computation of LER from bidirectional reflectance distribution function (BRDF) data obtained from other satellite sensors. In a recent work (Vasilkov et al., 2017), the BRDF data from

25 MODIS is first resampled to the lower resolution of the OMI and then a geometry-dependent LER is computed using radiative transfer model simulations. Unfortunately MODIS BRDF data is available only from VIS wavelengths and rescaling the VIS BRDF (or LER) to UV is not straightforward. Furthermore, the radiative transfer model assumptions needed for computing LER from BRDF may not be fully compatible with the assumptions made in the trace gas retrieval.

In this paper we present a novel algorithm for the retrieval of geometry-dependent effective Lambertian equivalent

30 reflectivity (GE_LER) from UVN measurements and the creation of global gapless geometry-dependent LER (G3_LER) daily map using GE_LER data under clear-sky conditions. The GE_LER/G3_LER retrieval solves the problems of using LER climatologies and accounts for surface anisotropy effects in cloud, aerosol and trace gas retrievals in a similar way as

the effective LER (Coldewey-Egbers et al., 2005) and the geometry-dependent LER (Qin et al., 2019). But in contrast to these approaches, the GE_LER retrieval is performed in exactly the same fitting windows used for the trace gas, cloud and aerosol retrievals; furthermore our algorithm does not require data from other sensors like BRDF (land surfaces) or Chlorophyll and wind parameters (water surfaces).

5 First we describe in section 2 the full-physics inverse learning machine (FP_ILM) technique used for the retrieval of GE_LER from UVN measurements and how it is optimized for the DOAS trace gas retrievals. Section 3 describes the creation of global gapless geometry-dependent LER (G3_LER) daily map using the retrieved GE_LER under clear-sky conditions. In section 4 we apply the GE_LER algorithms to S5P measurements and then we compare the TROPOMI G3_LER results with climatological OMI LER data. Finally in Section 5 we demonstrate the advantages of using

10 GE_LER/G3_LER for the retrieval of total ozone from TROPOMI and in Section 6 we discuss future work.

2.   The FP_ILM algorithm for the GE_LER retrieval

Trace gas, cloud and aerosol retrievals from UVN measurements rely on complex radiative transfer model (RTM) simulations. The RTM are computationally expensive and therefore not well suited for processing the big data from the new generation of atmospheric composition Sentinel missions. A classical approach for speeding up the RTM simulations is to

15 use look-up tables, but they require significant amount of memory and what is more important the interpolation/extrapolation errors could be large and time consuming. To solve these issues, the DLR team developed during the last two decades machine learning techniques for the optimal generation of RTM samples (Loyola et al., 2016) and the accurate parameterizing of RTM simulations using artificial neural networks (NN). These algorithms are being used for the operational processing of GOME-2 (Loyola et al., 2010) and now TROPOMI (Loyola et al., 2018) data.

20 Machine learning can be used not only for forward problems (like the parameterization of RTM simulations), but also for solving inverse problems, see for example (Loyola et al., 2016). During the last years we developed an approached called full-physics inverse learning machine (FP_ILM) technique that was successfully applied for retrieving profile shapes from GOME-2 (Xu et al., 2017) and retrieving SO_2 layer height from GOME-2 (Efremenko et al., 2017) and TROPOMI (Hedelt et al., 2019).

25 Figure 1 shows a flow diagram of the different steps of the FP_ILM algorithm and the following subsections describe in more detail how FP_ILM is applied for the retrieval of GE_LER.

2.1.   Forward Model

The forward model has two components: first a radiative transfer model (RTM) that computes the spectral intensity as a function of the viewing geometry, atmospheric components and surface properties; and second a sensor model that

30 transforms the RTM spectra to simulated spectra using sensor information like the instrument spectral response function and the instrument signal to noise ratio.

3

The forward model F can be used to compute simulated spectra radiances R_{sim} for a given wavelength λ as

$$R_{sim}(\lambda) \pm \varepsilon_R = F(\lambda, \Theta, \Omega, A_e, Z_e) \tag{1}$$

where ε_R denotes the expected instrument error, Θ is the light path geometry (solar and satellite zenith and azimuth angles), Ω are the atmospheric composition components, and the surface properties A_e for the geometry-dependent effective Lambertian equivalent reflectivity (GE_LER) and Z_e for the effective surface pressure.

2.2. Smart Sampling

A key element of FP_ILM is creating a training data set that extensively covers the multidimensional space of the forward problem and at the same time minimizes the computational expensive calls to the radiative transfer model. We use the smart sampling techniques (Loyola et al., 2016) for creating a dataset of samples $\{\Theta, \Omega, A_e, Z_e\}$ that fully represent the expected viewing and geophysical conditions of the problem at hand.

As shown in Figure 1, the smart sampling and forward module calls are iterated in a loop until the multi-dimensional integral of the output samples dataset $\{R_{sim}(\lambda) \pm \varepsilon_R\}$ converge; see (Loyola et al., 2016) for more details.

2.3. Feature Extraction

Retrieval of trace gas, cloud and aerosol concentrations from UVN sensors requires spectrometers with sufficient spectral resolution to resolve features in the electromagnetic spectrum; therefore the fitting-window used for the retrieval of a trace gas usually contains radiances at a high-dimensional space (tens to hundreds of wavelengths). Machine learning techniques perform best with low-dimensional datasets by avoiding the effects of the curse of dimensionality.

Feature extraction is a mapping function that transforms a dataset from a high- to a low-dimensional space removing redundant information and noise. In previous FP_ILM applications (Loyola et al., 2006; Xu et al., 2017) we used principal component analysis for the feature extraction, however for the GE_LER retrieval we take advantage of the DOAS fitting results

$$R_{sim}(\lambda) = -\sum_g N_{s,g}(\Theta) \cdot \sigma_g(\lambda) - P(\lambda) \tag{2}$$

with $N_{s,g}(\Theta)$ the effective slant column density of gas g for the light path geometry Θ, $\sigma_g(\lambda)$ the associated trace gas absorption cross-section for wavelength λ, and $P(\lambda)$ the external closure polynomial.

The feature extraction step consists in applying the DOAS fit to the simulated radiances. Combining (1) and (2) for a given fitting window Λ we obtain the following approximation with simulated datasets that representing the forward problem

$$\{N_{s,g}(\Theta), P(\Lambda)\} \cong \{F(\Theta, A_e(\Lambda), Z_e)\} \tag{3}$$

2.4. Machine Learning

Machine learning approximates a function represented by input/output datasets using either linear or non-linear regression algorithms. In this paper we use artificial neural networks (NN) to learn the non-linear inverse problem by reorganizing the datasets from (3) to represent the inverse problem

$$\{A_e(\Lambda)\} \cong \{F_{NN}^{-1}(P(\Lambda), N_{s,g}, \Theta, Z_e)\} \tag{4}$$

5   In other words, a neural network solves the inverse problem and retrieves the GE_LER as function of the DOAS closure polynomial, the DOAS fitted effective slant column density, the viewing geometry and the effective surface pressure. The inverse operator are the weights and biases of the neural network approximating F_{NN}^{-1}.

2.5.   GE_LER Retrieval

Obtaining the inverse operator is very time consuming mainly due to the relative large amount of RTM simulations needed
10   to properly represent the forward problem. Finding a NN topology that learns the inverse function with a small error is also computational intensive. But all these steps are done offline and only once for a given sensor and trace gas fitting window.

Figure 2 shows the flow diagram for applying the FP_ILM to satellite measurements. There is no extra computational needed for the feature extraction part as we are using the results from the DOAS fitting and the application of the NN to retrieved GE_LER is extremely fast as it only involves simple matrix multiplications.

15   The extremely fast retrieval using the FP_ILM is a crucial advantage for the operational near-real-time processing of the Big Data from the current and future atmospheric composition Sentinel missions.

3.   Global Gapless Geometry-dependent (G3) LER Daily Map

The conversion of the DOAS effective slant column to a geometry independent total column requires the calculation of air mas factors (AMF) using either the effective scene approximation (Coldewey-Egbers et al., 2005) or the independent pixel
20   approximation (e.g. Loyola et al., 2011). The GE_LER can be used directly for the computation of AMFs using the effective scene approximation, whereas a LER is needed for the computation of AMFs using the independent pixel approximation.

A global gapless geometry-dependent LER (G3_LER) daily map can be easily created from GE_LER retrieved under clear-sky conditions. The G3_LER map for a given day is created by merging the clear-sky LER data from the same day with the G3_LER map based on the LER data from previous days, see Figure 3.

25   It is important to note that the GE_LER takes into account the bidirectional reflectance distribution function (BRDF) effects as it is based on radiative transfer model simulations using the actual viewing geometry. But when combining GE_LER data their BRDF dependencies $\rho(\Lambda, \theta, \psi)$ as function of the wavelength in the fitting window Λ, the viewing zenith angle θ, and the surface types ψ must be considered. The function can be easily obtained separately for different fitting windows Λ (in the

UV, VIS and NIR spectral region), different surface types ψ (land, water, snow/ice) and time periods (e.g. monthly) by fitting a polynomial of clear-sky LERs averaged as function of θ.

The G3_LER daily map contains normalized LER, i.e. GE_LER retrieved under clear-sky conditions divided by the fitted BRDF dependency, as well as the multiplicative factors $\rho(\theta)$ to compute the geometry-dependent LER as a function of the
5   actual satellite viewing zenith angle θ.

It is necessary to aggregate normalized LER retrievals over several days (between one to four weeks depending on cloudiness) in order to obtain a global gapless map. In contrast to LER climatologies, the G3_LER map represents the actual surface properties as it is updated on a daily basis.

4.   GE_LER and G3_LER from TROPOMI/S5P 325-335 nm

10   The GE_LER and G3_LER algorithms described in the previous sections are applied to measurements of TROPOMI/S5P in the total ozone wavelength region. The S5P operational near-real-time total ozone products (Loyola et al., 2019) are based on the DOAS algorithm using the fitting window of 325-335 nm.

4.1.   FP_ILM GE_LER Training

The training dataset is based on spectra simulated by the Vector LInearized Discrete Ordinate Radiative Transfer
15   (VLIDORT) model (Spurr, 2016). The RTM inputs are ozone concentration profiles, surface albedo, surface pressure and the viewing geometry solar and viewing angles. The smart-sampling technique (Loyola et al., 2016) was used to create more than 2×10^5 synthetic UV spectra using ozone profile, viewing geometry and surface parameters in the range listed in Table 1. We use the Bodeker et al., (2013) ozone database merged with the McPeters/Labow (Labow et al., 2015) ozone climatology for an optimal representation of the ozone vertical distribution in the stratosphere and troposphere.

20   TROPOMI/S5P-like measurements are created by applying the instrument slit function to the RTM simulated radiances and adding a Gaussian instrument noise with a signal-to-noise ratio of 300 representative of TROPOMI band 3, see Kleipool et al., 2018.

The DOAS fitting is applied to the simulated S5P radiances using a cubic polynomial resulting in a dataset of ozone slant columns and the polynomial coefficients. Figure 4 shows the optical densities difference for three scenarios: (a) with respect
25   to four typical values of surface albedo of 0.05, 0.3, 0.6, and 0.9 correspond to water, land, melted snow/ice-covered and fresh snow/ice-covered regions. The largest absolute value of the optical density corresponds to the largest surface albedo; the optical densities for four albedos do not differ significantly at the lower wavelength, while the differences increase at the higher wavelength. (b) with respect to three total ozone columns of 150 DU, 300 DU, and 500 DU; the optical density increases gradually along the selected wavelength region, the absolute value of the optical density increases when the total
30   ozone column increases. And (c) with respect to three viewing zenith angles of 50°, 30°, 10°; the absolute value of the

optical density increases when the viewing zenith angle decreases. For all cases, the optical density increases along the wavelength region.

The input and output of the simulations is reorganized according to (3) and a neural network is trained to learn the inverse function using 70% of the simulations for training, 15% for testing and 15% for validation. The best results are obtained

5 using a NN with a topology of 9-20-8-2-1, which is 9 neurons in the input layer, three hidden layers with the given number of neurons, and one neuron on the output layer.

The GE_LER retrieval errors as function of different input parameters calculated using the validation dataset (i.e. the dataset not used for the NN training) are depicted in Figure 5. The differences between the *true* and retrieved GE_LER are very small with a mean and standard deviation of only 0.0016 ± 0.0018. These results demonstrate that the NN represents the

10 inverse function in a very precise way.

4.2. FP_ILM GE_LER Retrieval

The neural network trained with the inverse function is applied to TROPOMI/S5P measurements. The inputs are the DOAS fitted polynomial coefficients and ozone slant column, the solar and viewing zenith angles, the relative azimuth angle, and the effective surface pressure Z_e computed as

15
$$Z_e = (1 - f_c)Z_s + f_c Z_c \tag{5}$$

where f_c is the cloud fraction, Z_s the surface pressure, and Z_c the cloud pressure. The S5P cloud properties are obtained from the operational TROPOMI cloud products using the OCRA and ROCINN (Lutz et al., 2016; Loyola et al., 2018) algorithms.

The TROPOMI/S5P GE_LER results for April 10th, 2018 are shown in Figure 6, as expected the GE_LER shows the same

20 patterns as the clouds for that day. In the case of clear-sky ($f_c \leq 0.05$) the GE_LER represents the surface albedo and for the cloudy cases ($f_c \geq 0.95$) the GE_LER represents the cloud albedo. Figure 7 shows the histograms of the differences between the TROPOMI clear-sky GE_LER and OMI LER climatology (Kleipool et al., 2008) and the differences between the cloudy TROPOMI GE_LER and the cloud albedo from the operational cloud product retrieved with ROCINN_CRB (Loyola et al., 2018). The second mode around 0.5 in the histogram for the snow/ice cases indicates snow conditions in TROPOMI data

25 that are not well represented in the OMI LER climatology.

The mean differences for the clear-sky and cloudy cases as function of the surface type are summarized in Figure 7, the relative larger offsets and spreads for the cloudy cases are mainly due to the different spectral regions covered by GE_LER for the total ozone fitting window in the UV (325–335 nm) and the cloud properties retrieved with ROCINN_CRB from the oxygen A-Band in the NIR (758–771 nm).

30 ### 4.3. G3_LER Daily Map

The TROPOMI G3_LER map for a given day is created by regridding (using a 0.1° x 0.1° resolution) and aggregating normalized LER from the couple of days. The FP_ILM LERs are obtained from the S5P GE_LER retrievals under clear-sky conditions. In this version of the TROPOMI G3_LER map we use the OCRA cloud fraction f_c for identifying clear-sky measurements, more concretely, we use the measurements with $f_c \leq 0.05$. In the future we plan to additionally use the S5P

5   aerosol product and the regridded VIIRS/SNPP (flying in constellation with S5P) for a more stringent cloud/aerosol screening.

The ground pixels affected by sun glint as well as the pixels influenced by solar eclipse are removed using the corresponding flags available in the S5P total ozone product (Pedergnana et al., 2018). The remaining FP_ILM LERs from a given day replace the corresponding grid points of the G3_LER map from the previous day.

10   The BRDF dependencies $\rho(\theta)$ are calculated by fitting a polynomial to the TROPOMI LER data averaged as function of the viewing zenith angle. Three different surface types are considered: land, water and snow/ice. Figure 8 shows the BRDF dependencies calculated with TROPOMI/S5P data from January, April, July and October 2018. For the surface classification we use the Land/Water mask and the snow/ice flag available in the S5P total ozone product (Pedergnana et al., 2018).

Figure 9 shows the TROPOMI/S5P G3_LER daily map corresponding to April 30th, 2018 and a comparison to the OMI LER
15   climatology for the month of April. The OMI LER is based on 3 years of data (2004 to 2007) whereas the TROPOMI G3_LER contains data of only a few weeks. The main advantages of the TROPOMI G3_LER daily map compared to climatology are first that it represents the current surface conditions like snow/ice contamination, second it takes into account the BRDF effects and third it has a better spatial resolution (0.1°).

4.4.   Usage of TROPOMI/S5P G3_LER for the Total Ozone retrieval

20   The near-real-time S5P total ozone product is based on an iterative DOAS/AMF algorithm (Loyola et al., 2019) and the current operational version (1.1.5) uses the OMI LER climatology (Kleipool et al., 2008). The median bias between near-real-time total ozone from S5P and reference data from Brewer, Dobson, and SAOZ sites is of the order of +1% (Verhoelst et al., 2018; Garane et al., 2019).

S5P near-real-time ozone agrees well with the Copernicus Atmosphere Monitoring Service (CAMS) analysis with the
25   exception of some anomalies at high latitudes (Inness et al., 2019). Those anomalies are associated to the coarse resolution of the OMI LER climatology and most important, the differences between the climatological LER values and the actual surface conditions like snow/ice.

We replace the OMI LER climatology with the TROPOMI G3_LER daily maps and the resulting total ozone field is significantly smother and with far less outliers. Figure 10 shows the TROPOMI/S5P surface albedo and total ozone retrievals
30   from April 1st, 2018 around the Bering Strait which separates Russia and Alaska. The TROPOMI G3_LER daily map agrees very well with the surface types visible in the corresponding VIIRS/SNPP images (S5P flies only 3-5 minutes behind SNPP)

8

including the water surface along the coasts of the shores of the Chukchi Sea in Russia and the Sarichef Island in the north of Alaska and the Seward Peninsula in south of Alaska. These water surfaces along the coast as well as the water of the Bering Sea are not properly represented in the OMI LER climatology that shows snow/ice over these regions. Likewise, the OMI LER climatology erroneously shows no snow/ice in the Yukon–Koyukuk Census Area in Alaska. The coarse spatial

5   resolution of the OMI LER climatology is clearly visible in the total ozone field and what is even worst the wrong snow/ice values in the OMI LER climatology induce large errors on the retrieved total ozone with differences between -10% and $+15\%$.

Moreover, the agreement of the S5P total ozone with the CAMS assimilation at high latitudes is significantly better, see Figure 11. The mean differences between total ozone from S5P and CAMS for the complete month of April 2018 are

10   summarized in Table 3. The agreement with CAMS improves considerably in all latitudinal regions: the differences in the total ozone in the region [80°S-60°S] is reduced from $-2.53 \pm 2.46\%$ using OMI LER to $0.78 \pm 3.49\%$ using TROPOMI G3_LER, in the region [60°S-50°N] is reduced from $0.25 \pm 1.17\%$ to $0.12 \pm 1.21\%$, in the region [50°N-70°N] is reduced from $1.21 \pm 2.46\%$ to $0.01 \pm 2.02\%$ and finally in the region [70°N -90°N] is reduced from $-1.004 \pm 2.58\%$ to $-0.15 \pm 2.64\%$.

15   **5.   Conclusions**

We have developed a novel algorithm for the accurate and fast retrieval of geometry-dependent effective Lambertian equivalent reflectivity (GE_LER) from UVN sensors based on the full-physics inverse learning machine (FP_ILM) technique. The main inputs to the GE_LER retrieval are the DOAS fitting polynomial and fitted trace gas slant column as well as the satellite viewing geometry. The inversion problem is solved using neuronal networks trained with radiative

20   transfer model simulations based on the same kind of RTM and settings used for the AMF calculations.

A global gapless geometry-dependent LER (G3_LER) daily map can be easily created from the GE_LER retrievals under clear-sky conditions. Both GE_LER and G3_LER take into account the satellite viewing dependencies characterized by the bidirectional reflectance distribution function (BRDF) effects.

GE_LER is retrieved from each single ground pixel using the same spectrum and DOAS/AMF settings as the trace gas

25   retrieval and therefore it is fully consistent with the trace gas retrieval in contrast to LER products based on data from other satellites or LER data from the same satellite but using different fitting or RTM settings. G3_LER maps are updated on a daily basis using the GE_LER under clear-sky conditions from that day and therefore it is clearly superior to LER climatologies that fail to represent the actual surface conditions like snow/ice.

We have applied the FP-ILM GE_LER/G3_LER to S5P and showed that the total ozone retrieval using this novel product is

30   substantially superior to the one created using the OMI_LER climatology. The ozone fields are not only much more smooth, but also the differences compared to the total ozone from CAMS is reduced from $-2.53 \pm 2.46\%$ to $0.78 \pm 3.49\%$ in the

9

latitudinal region [80°S-60°S]. Large errors on the S5P total ozone between -10% and $+15\%$ induced by snow/ice miss-representations in the OMI_LER climatology are removed using the FP-ILM GE_LER/G3_LER TROPOMI products.

FP_ILM GE_LER can be applied to any trace gas, cloud and aerosol product retrieved in the UVN and is fully compatible with the DOAS/AMF settings used for the trace gas retrievals. GE_LER and G3_LER can be used for computing AMFs
5 based on the effective scene approximation or the independent pixel approximation respectively. In this paper we demonstrated their effectiveness for improving the quality of the total ozone from TROPOMI; in the near future we will extend GE_LER/G3_LER to the fitting windows of the S5P operational UVN cloud product (Loyola et al., 2018) and UV/VIS trace gases NO_2 (van Geffen et al., 2018), SO_2 (Theys et al., 2017), HCHO (De Smedt et al., 2018) as well as S5P research product like CHOCHO and aerosol optical depth.

10 The GE_LER retrieval is accurate and extremely fast and therefore well suited for the (near-real-time) processing of the huge amount of data of the atmospheric Sentinel satellite missions. We plan to apply the FP_ILM GE_LER/G3_LER retrieval to the future Copernicus Sentinel-5 mission that like Sentinle-5P will follow a sun-synchronous polar orbit. Furthermore, we plan to assess the suitability of FP_ILM GE_LER to capture the diurnal LER dependencies on the sun-satellite geometry of the future UVN geostationary missions Sentinel-4, TEMPO and GMES.

15

Acknowledgements

20

References

Bodeker, G. E., Hassler, B., Young, P. J., and Portmann, R.W.,: A vertically resolved, global, gap-free ozone database for assessing or constraining global climate model simulations, Earth Syst. Sci. Data, 5/1, 31–43, 2013.

Coldewey-Egbers, M., Weber, M., Lamsal, L. N., de Beek, R., Buchwitz, M., and Burrows, J. P.: Total ozone retrieval from
25 GOME UV spectral data using the weighting function DOAS approach, Atmos. Chem. Phys., 5, 1015-1025, https://doi.org/10.5194/acp-5-1015-2005, 2005.

Efremenko, D. S., Loyola R., D. G., Hedelt, P., and Spurr, R. J. D.: Volcanic SO2 plume height retrieval from UV sensors using a full-physics inverse learning machine algorithm, International Journal of Remote Sensing, 38, 1–27, https://doi.org/10.1080/01431161.2017.1348644, 2017.

Garane et al.: TROPOMI/S5p total ozone column data: global ground-based validation, Atmos. Meas. Tech. Discuss., in preparation, 2019

van Geffen, J.H.G.M., Eskes, H.J., Boersma, K.F., Maasakkers, J.D. and Veefkind, J.P.: TROPOMI ATBD of the total and tropospheric NO2 data products, S5P-KNMI-L2-0005-RP available at: https://sentinel.esa.int/web/sentinel/technical-guides/sentinel-5p/products-algorithms and http://www.tropomi.eu/documents/level-2-products, 2018.

Hedelt, P., Efremenko D.S., Loyola D.G., Spurr, R., and Lieven .: SO2 Layer Height retrieval from Sentinel-5 Precursor/TROPOMI using FP_ILM, Atmos. Meas. Tech. Discuss., submited, 2019.

Herman, J. R., and Celarier E. A.: Earth surface reflectivity climatology at 340−380nm from TOMS data, J. Geophys. Res., 102(D23), 28,003−28,011, doi:10.1029/97JD02074, 1997.

Inness, A., Flemming, J., Heue, K.-P., Lerot, C., Loyola, D., Ribas, R., Valks, P., van Roozendael, M., Xu, J., and Zimmer, W.: Monitoring and assimilation tests with TROPOMI data in the CAMS system: near-real-time total column ozone, Atmos. Chem. Phys., 19, 3939-3962, https://doi.org/10.5194/acp-19-3939-2019, 2019.

Kleipool, Q., Ludewig, A., Babić, L., Bartstra, R., Braak, R., Dierssen, W., Dewitte, P.-J., Kenter, P., Landzaat, R., Leloux, J., Loots, E., Meijering, P., van der Plas, E., Rozemeijer, N., Schepers, D., Schiavini, D., Smeets, J., Vacanti, G., Vonk, F., and Veefkind, P.: Pre-launch calibration results of the TROPOMI payload on-board the Sentinel-5 Precursor satellite, Atmos. Meas. Tech., 11, 6439-6479, https://doi.org/10.5194/amt-11-6439-2018, 2018.

Kleipool, Q. L., Dobber M. R., de Haan J. F., and Levelt P. F.: Earth surface reflectance climatology from 3 years of OMI data, J. Geophys. Res., 113, D18308, doi:10.1029/2008JD010290, 2008.

Koelemeijer, R. B. A., de Haan J. F., and Stammes P.: A database of spectral surface reflectivity in the range 335−772nm derived from 5.5 years of GOME observations, J. Geophys. Res., 108(D2), 4070, doi:10.1029/2002JD002429, 2003.

Labow, G. J., Ziemke, J. R., McPeters, R. D., Haffner, D. P., and Bhartia, P. K.: A total ozone-dependent ozone profile climatology based on ozonesondes and Aura MLS data, J. Geophys. Res. Atmos., 120/6, 2537−2545, 2015.

Lee, C., R. V. Martin, A. van Donkelaar, G. O'Byrne, A. Richter, G. Huey, and J. S. Holloway, Retrieval of vertical columns of sulfur dioxide from SCIAMACHY and OMI: Air mass factor algorithm development and validation, J. Geophys. Res., 114, D22303, doi:10.1029/2009JD012123, 2009.

Lin, J.-T., Martin, R. V., Boersma, K. F., Sneep, M., Stammes, P., Spurr, R., Wang, P., Van Roozendael, M., Clémer, K., and Irie, H.: Retrieving tropospheric nitrogen dioxide from the Ozone Monitoring Instrument: effects of aerosols, surface reflectance anisotropy, and vertical profile of nitrogen dioxide, Atmos. Chem. Phys., 14, 1441−1461, https://doi.org/10.5194/acp-14-1441-2014, 2014.

Lerot C., Van Roozendael M., Spurr R., Loyola D., Coldewey-Egbers M., Kochenova S., van Gent J., Koukouli M., Balis D., Lambert J.-C., Granville J., Zehner C., Homogenized total ozone data records from the European sensors GOME/ERS-2, SCIAMACHY/Envisat, and GOME-2/MetOp-A, Journal of Geophysical Research, 119, 1639–1662, doi:10.1002/2013JD020831, 2014.

5   Lerot C., Van Roozendael M., Lambert J.-C., Granville J., Van Gent J., Loyola D., Spurr R. J. D., The GODFIT algorithm: a direct fitting approach to improve the accuracy of total ozone measurements from GOME, International Journal of Remote Sensing, 31:2, 543–550, doi:10.1080/01431160902893576, 2010.

Lorente, A., Boersma, K. F., Stammes, P., Tilstra, L. G., Richter, A., Yu, H., Kharbouche, S., and Muller, J.-P.: The importance of surface reflectance anisotropy for cloud and NO2 retrievals from GOME-2 and OMI, Atmos. Meas. Tech., 11, 10   4509-4529, https://doi.org/10.5194/amt-11-4509-2018, 2018.

Loyola, D., et al.: The near-real-time total ozone retrieval algorithm from TROPOMI onboard Sentinel-5 Precursor, Atmos. Meas. Tech. Discuss., in preparation, 2019.

Loyola, D. G., Gimeno García, S., Lutz, R., Argyrouli, A., Romahn, F., Spurr, R. J. D., Pedergnana, M., Doicu, A., Molina García, V., and Schüssler, O.: The operational cloud retrieval algorithms from TROPOMI on board Sentinel-5 Precursor, 15   Atmos. Meas. Tech., 11, 409-427, https://doi.org/10.5194/amt-11-409-2018, 2018.

Loyola, D., Pedergnana, M., and Gimeno García, S., Smart sampling and incremental function learning for very large high dimensional data, Neural Networks, 78, 75–87, https://doi.org/10.1016/j.neunet.2015.09.001, 2016.

Loyola, D., Thomas, W., Spurr, R., and Mayer, B.: Global patterns in daytime cloud properties derived from GOME backscatter UV-VIS measurements", International Journal of Remote Sensing, 31:16, 4295-4318, 2010.

20   Loyola, D., M. Koukouli, P. Valks, D. Balis, N. Hao, M. Van Roozendael, R. Spurr, W. Zimmer, S. Kiemle, C. Lerot, and J-C. Lambert, The GOME-2 Total Column Ozone Product: Retrieval Algorithm and Ground-Based Validation, J. Geophys. Res., 116, D07302, doi:10.1029/2010JD014675, 2011.

Loyola D.: Applications of Neural Network Methods to the Processing of Earth Observation Satellite Data, Neural Networks, 19/2, 168-177, https://doi.org/10.1016/j.neunet.2006.01.010, 2006.

25   Lutz, R., Loyola, D., Gimeno García, S., and Romahn, F.: OCRA radiometric cloud fractions for GOME-2 on MetOp-A/B, Atmos. Meas. Tech., 9, 2357-2379, https://doi.org/10.5194/amt-9-2357-2016, 2016.

Seidel, F. C. and Popp, C.: Critical surface albedo and its implications to aerosol remote sensing, Atmos. Meas. Tech., 5, 1653-1665, https://doi.org/10.5194/amt-5-1653-2012, 2012.

Atmospheric
Measurement
Techniques
Discussions

O'Byrne, G., R. V. Martin, A. van Donkelaar, J. Joiner, and E. A. Celarier, Surface reflectivity from the Ozone Monitoring Instrument using the Moderate Resolution Imaging Spectroradiometer to eliminate clouds: Effects of snow on ultraviolet and visible trace gas retrievals, *J. Geophys. Res.*, 115, D17305, doi:10.1029/2009JD013079, 2010.

Pedergnana, M., Loyola, D., Apituley, A., Sneep, M., Veefkind, J. P.: Sentinel-5 precursor/TROPOMI – Level 2 Product

5 User Manual – Ozone Total Column, S5P-L2-DLR-PUM-400A available at: https://sentinel.esa.int/web/sentinel/technical-guides/sentinel-5p/products-algorithms and http://www.tropomi.eu/documents/level-2-products, 2018.

Pflug B., Aberle B., Loyola D., Valks P., Near-Real-Time Estimation of Spectral Surface Albedo from GOME-2/MetOp Measurements, EUMETSAT Meteorological Satellite Conference, Darmstadt, September, 2008.

Qin, W., Fasnacht, Z., Haffner, D., Vasilkov, A., Joiner, J., Krotkov, N., Fisher, B., and Spurr, R.: A geometry-dependent

10 surface Lambertian-equivalent reflectivity product at 466 nm for UV/Vis retrievals: Part I. Evaluation over land surfaces using measurements from OMI, Atmos. Meas. Tech. Discuss., https://doi.org/10.5194/amt-2018-327, in review, 2019.

De Smedt, I., Theys, N., Yu, H., Danckaert, T., Lerot, C., Compernolle, S., Van Roozendael, M., Richter, A., Hilboll, A., Peters, E., Pedergnana, M., Loyola, D., Beirle, S., Wagner, T., Eskes, H., van Geffen, J., Boersma, K. F., and Veefkind, P.: Algorithm theoretical baseline for formaldehyde retrievals from S5P TROPOMI and from the QA4ECV project, Atmos.

15 Meas. Tech., 11, 2395-2426, https://doi.org/10.5194/amt-11-2395-2018, 2018.

Spurr, R. J. D. VLIDORT: A linearized pseudo-spherical vector discrete ordinate radiative transfer code for forward model and retrieval studies in multilayer multiple scattering media, J. Quant. Spectrosc. Radiat. Transf., 102, 316-42, 10.1016/j.jqsrt.2006.05.005, 2006.

Rault, D. F., and G. Taha, Validation of ozone profiles retrieved from Stratospheric Aerosol and Gas Experiment III limb

20 scatter measurements, J. Geophys. Res., 112, D13309, doi:10.1029/2006JD007679, 2007.

Theys, N., De Smedt, I., Yu, H., Danckaert, T., van Gent, J., Hörmann, C., Wagner, T., Hedelt, P., Bauer, H., Romahn, F., Pedergnana, M., Loyola, D., and Van Roozendael, M.: Sulfur dioxide retrievals from TROPOMI onboard Sentinel-5 Precursor: algorithm theoretical basis, Atmos. Meas. Tech., 10, 119-153, https://doi.org/10.5194/amt-10-119-2017, 2017.

Tilstra, L. G., Tuinder O. N. E., Wang P., and Stammes P.: Surface reflectivity climatologies from UV to NIR determined

25 from Earth observations by GOME-2 and SCIAMACHY, J. Geophys. Res. Atmos., 122, 4084–4111, doi:10.1002/2016JD025940, 2017.

Vasilkov, A., Yang, E.-S., Marchenko, S., Qin, W., Lamsal, L., Joiner, J., Krotkov, N., Haffner, D., Bhartia, P. K., and Spurr, R.: A cloud algorithm based on the O_2-O_2 477 nm absorption band featuring an advanced spectral fitting method and the use of surface geometry-dependent Lambertian-equivalent reflectivity, Atmos. Meas. Tech., 11, 4093-4107,

30 https://doi.org/10.5194/amt-11-4093-2018, 2018.

13

Vasilkov, A., Qin, W., Krotkov, N., Lamsal, L., Spurr, R., Haffner, D., Joiner, J., Yang, E.-S., and Marchenko, S.: Accounting for the effects of surface BRDF on satellite cloud and trace-gas retrievals: a new approach based on geometry-dependent Lambertian equivalent reflectivity applied to OMI algorithms, Atmos. Meas. Tech., 10, 333-349, https://doi.org/10.5194/amt-10-333-2017, 2017.

5   Verhoelst, T., Granville, J., Lambert, J.-C., Heue, K.-P., S5P MPC VDAF Validation Web Article: Total Column of Ozone, S5P-MPC-VDAF-WVA-L2_O3_20180720 available at: http://mpc-vdaf.tropomi.eu/index.php/ozone, 2018.

Xu, J., Schüssler, O., Loyola R, D., Romahn, F. and Doicu, A.: A novel ozone profile shape retrieval using Full-Physics Inverse Learning Machine (FP_ILM). IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens. 10(12), pp. 5442–5457, doi: 10.1109/JSTARS.2017.2740168, 2017.

10 Zhou, Y., Brunner, D., Spurr, R. J. D., Boersma, K. F., Sneep, M., Popp, C., and Buchmann, B.: Accounting for surface reflectance anisotropy in satellite retrievals of tropospheric NO_2, Atmos. Meas. Tech., 3, 1185-1203, https://doi.org/10.5194/amt-3-1185-2010, 2010.

15

14

Table 1: Range of the input parameters used for radiance simulations in the total ozone fitting window; the ozone profiles are classified as function of the total column. Smart sampling is used to generate node points optimally covering all input dimensions and more than 2×10^5 synthetic UV spectra are generated.

Parameter	Minimun	Maximum
Ozone Profile	125 DU	575 DU
Solar Zenith Angle	0°	90°
Viewing Zenith Angle	0°	70°
Relative Azimuth Angle	0°	180°
Surface Albedo	0	1
Surface Pressure	125 hPa	1013 hPa

Atmospheric
Measurement
Techniques
Open Access
Discussions
EGU

Table 2: Summary of the comparison between TROPOMI GE_LER clear-sky and OMI LER as well as for TROPOMI GE_LER cloudy and ROCINN_CRB cloud albedo. There are more than 4.5 million clear-sky and more than 1.4 million cloudy cases out of the around 15 million S5P measurements from April 10th, 2018.

	Number	Mean	Std. Dev.
Clear-sky Land	866 907	0.0014	0.0624
Clear-sky Water	1 837 686	-0.0144	0.0762
Clear-sky Snow/Ice	1 852 222	-0.0048	0.2573
Cloudy Land	254 645	0.0834	0.1865
Cloudy Water	1 084 985	0.0487	0.1464
Cloudy Snow/Ice	127 636	-0.0343	0.5432

5

Atmospheric
Measurement
Techniques
Open Access
Discussions

Table 3: Latitudinal differences between total ozone from CAMS and S5P using TROPOMI G3_LER and OMI LER for the complete month of April 2018. The values represent the total number of measurements for each latitudinal range and the mean difference ± standard deviation in percentage. Latitude bands with less than 100000 data points/degree were skipped, due to the polar winter there are hardly any data south of 81°S. The number of measurements increases in the north because of the overlapping orbits.

Latitude Range	Number	TROPOMI G3_LER	OMI LER
80°S-70°S	11297206	0.274±3.440	-2.041±2.114
70°S-60°S	29018428	0.983±3.515	-2.727±2.300
60°S-50°S	32351377	1.147±1.963	0.808±1.815
50°S-40°S	31580917	0.060±1.264	0.048±1.224
40°S-30°S	31154717	-0.373±0.962	-0.336±0.930
30°S-20°S	30948143	-0.302±0.843	-0.252±0.807
20°S-10°S	30814933	0.408±0.778	0.537±0.745
10°S-0°S	30744238	0.316±0.806	0.517±0.720
0°N-10°N	30732173	0.364±0.843	0.607±0.738
10°N-20°N	30779225	-0.034±0.799	0.142±0.728
20°N-30°N	30894360	-0.271±0.960	-0.097±0.901
30°N-40°N	31091907	-0.204±1.375	0.173±1.336
40°N-50°N	31469922	0.120±1.883	0.584±1.880
50°N-60°N	32250750	0.150±1.720	1.287±1.920
60°N-70°N	39590441	0.099±2.240	1.155±2.798
70°N-80°N	56545121	-0.049±2.719	-0.730±2.701
80°N-90°N	26178029	-0.353±2.446	-1.595±2.317

17
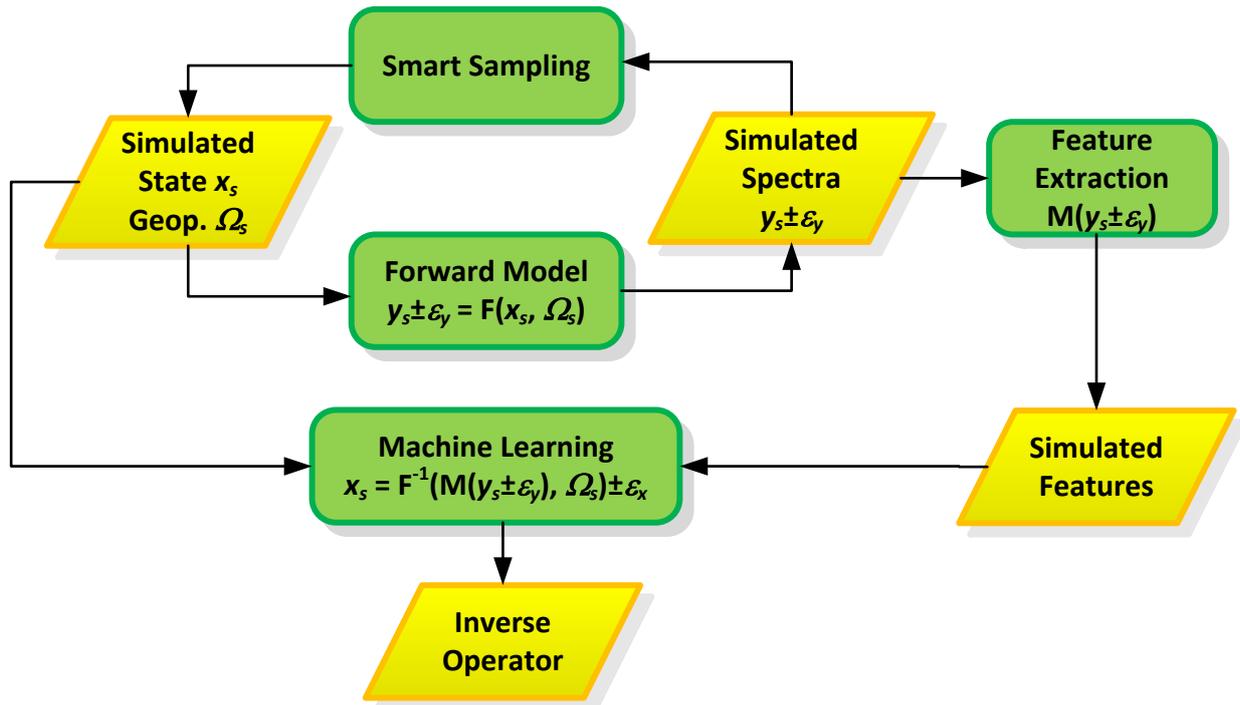
Figure 1: Data flow diagram of the FP_ILM training phase. The smart sampling techniques is used to create simulated state vector x_s and geophysical conditions Ω_s that are used as input to a forward model for the creation of simulated spectra with their expected errors $y_s + e_y$. Machine learning techniques are used for computing the inverse operator that is trained using as input the features extracted from the simulated spectra $M(y_s)$ and the geophysical conditions Ω_s as an output the state vector and the errors $x_s + e_x$.

Atmospheric
Measurement
Techniques
Discussions

Open Access

5
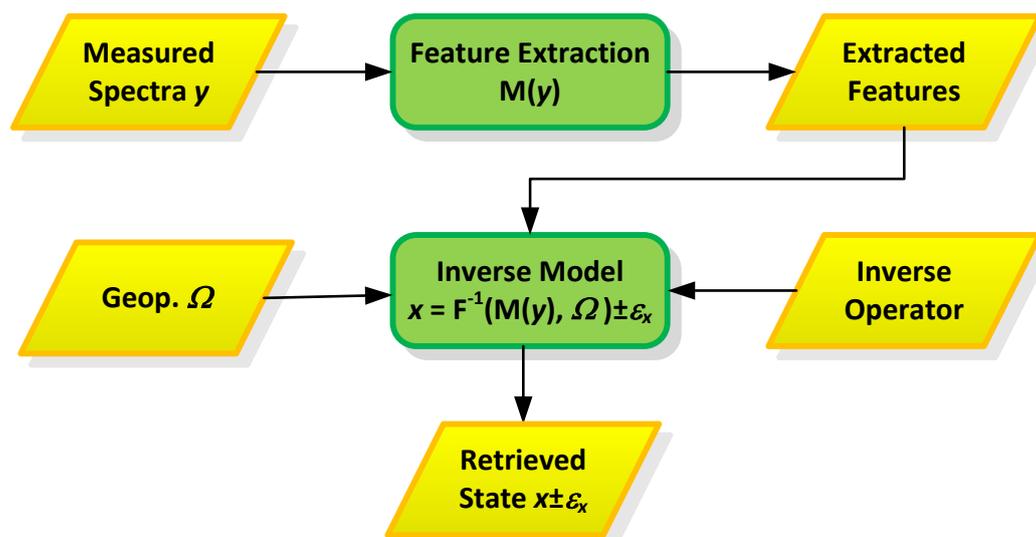
**Figure 2: Data flow diagram of the FP_ILM retrieval phase. The inverse operator computed during the FP_ILM training phase is
10 used to solve the inverse problem and retrieve the state vector *x* taking as input the features *M()* extracted from the measured
spectra *y* and the geophysical conditions *Ω*.**
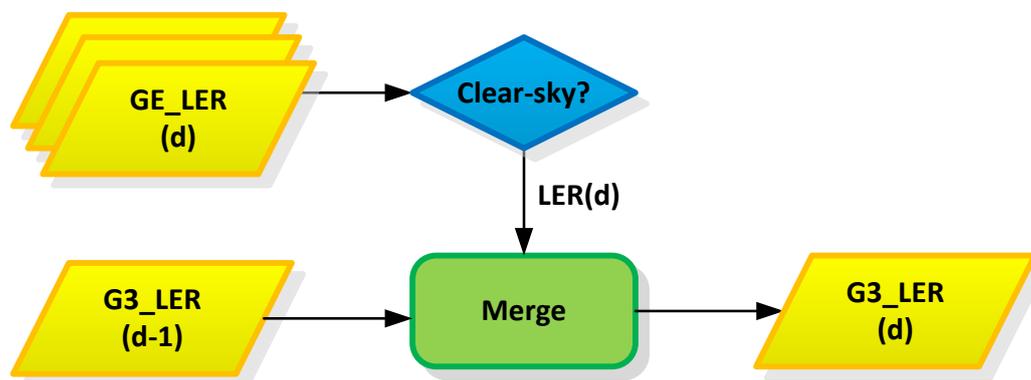
Figure 3: Data flow diagram of the creation of global gapless geometry-dependent LER (G3_LER) map for day *d* by merging the
5 **clear-sky LER data from the same day with the G3_LER map from the previous day.**

(a)   (b)   (c)
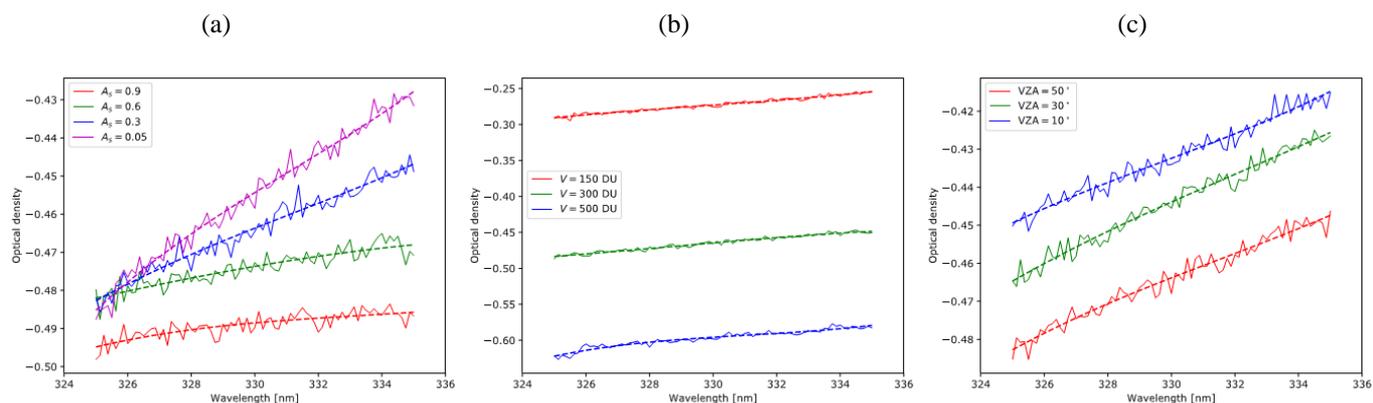
Figure 4: Optical densities difference as function of wavelength with respect to (a) surface albedo, (b) total ozone, and (c) viewing zenith angle. The doted-lines represent the DOAS fitted polynomial.

5
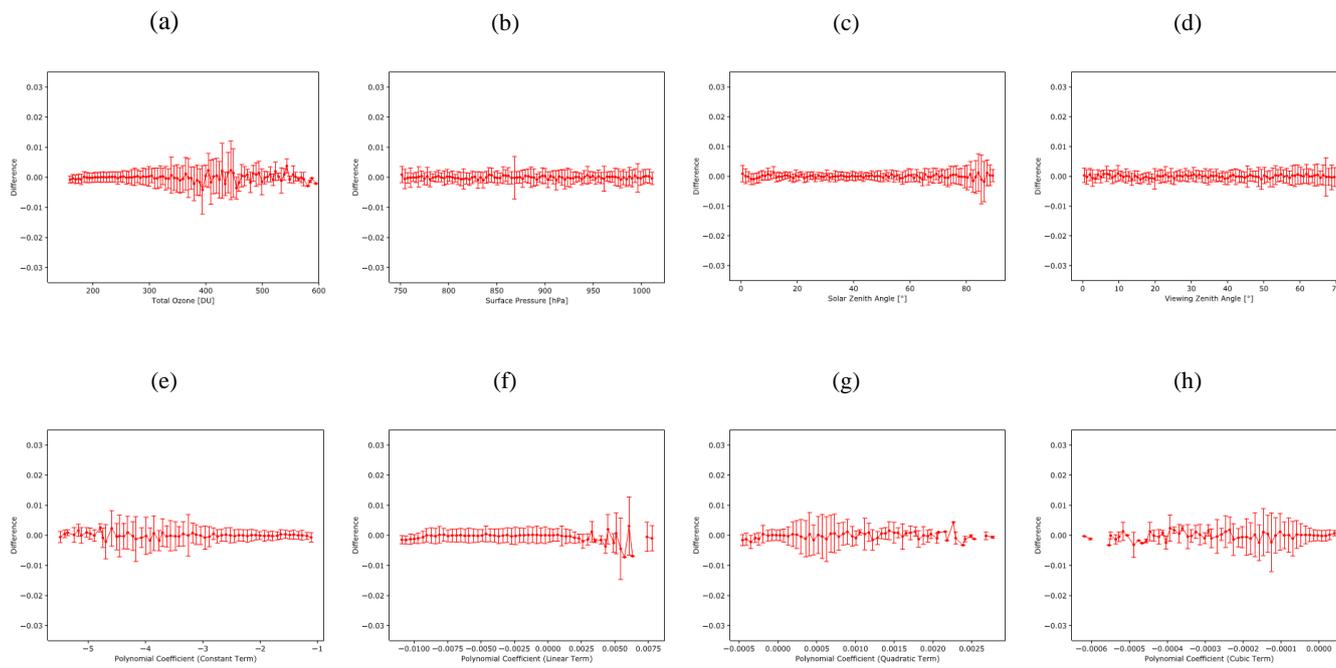
Figure 5: GE_LER retrieval error as function of (a) total ozone, (b) surface pressure, (c) solar zenith angle, (d) viewing zenith angle, and (e to h) the four DOAS polynomial fit coefficients.

5

(a)

(b)
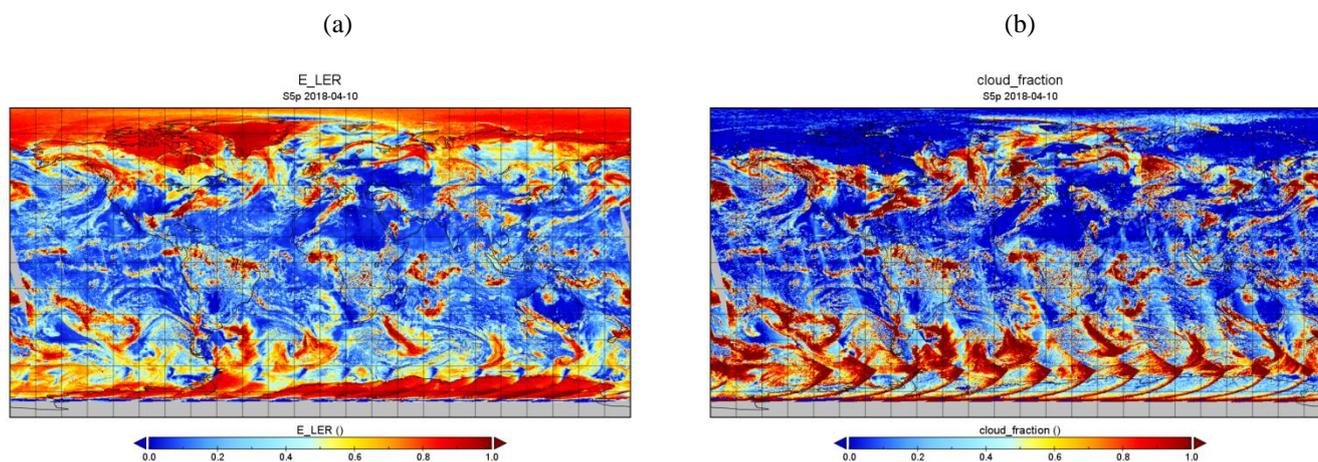
Figure 6: (a) GE_LER in the total ozone fitting windows [325-335 nm] retrieved from TROPOMI/S5P data from April 10ᵗʰ, 2018 and (b) the corresponding cloud fraction for this day.
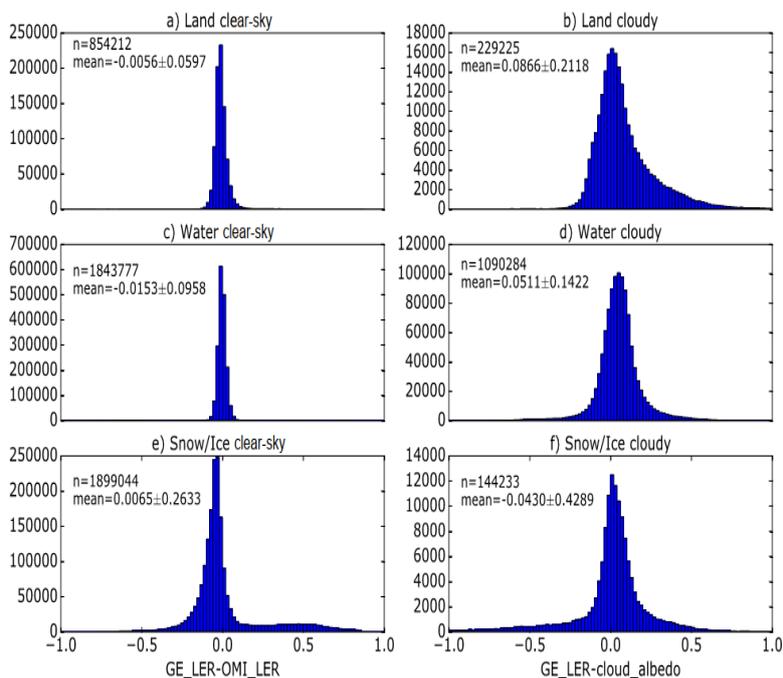
Figure 7: Histograms of the differences (left) between clear-sky TROPOMI GE_LER and OMI LER climatology and (right) between the cloudy TROPOMI GE_LER and the ROCINN_CRB cloud albedo from the operational S5P cloud product. The comparisons are performed separately per surface types (land, water, and snow/ice) and using S5P data from April 10th, 2018.

5

10
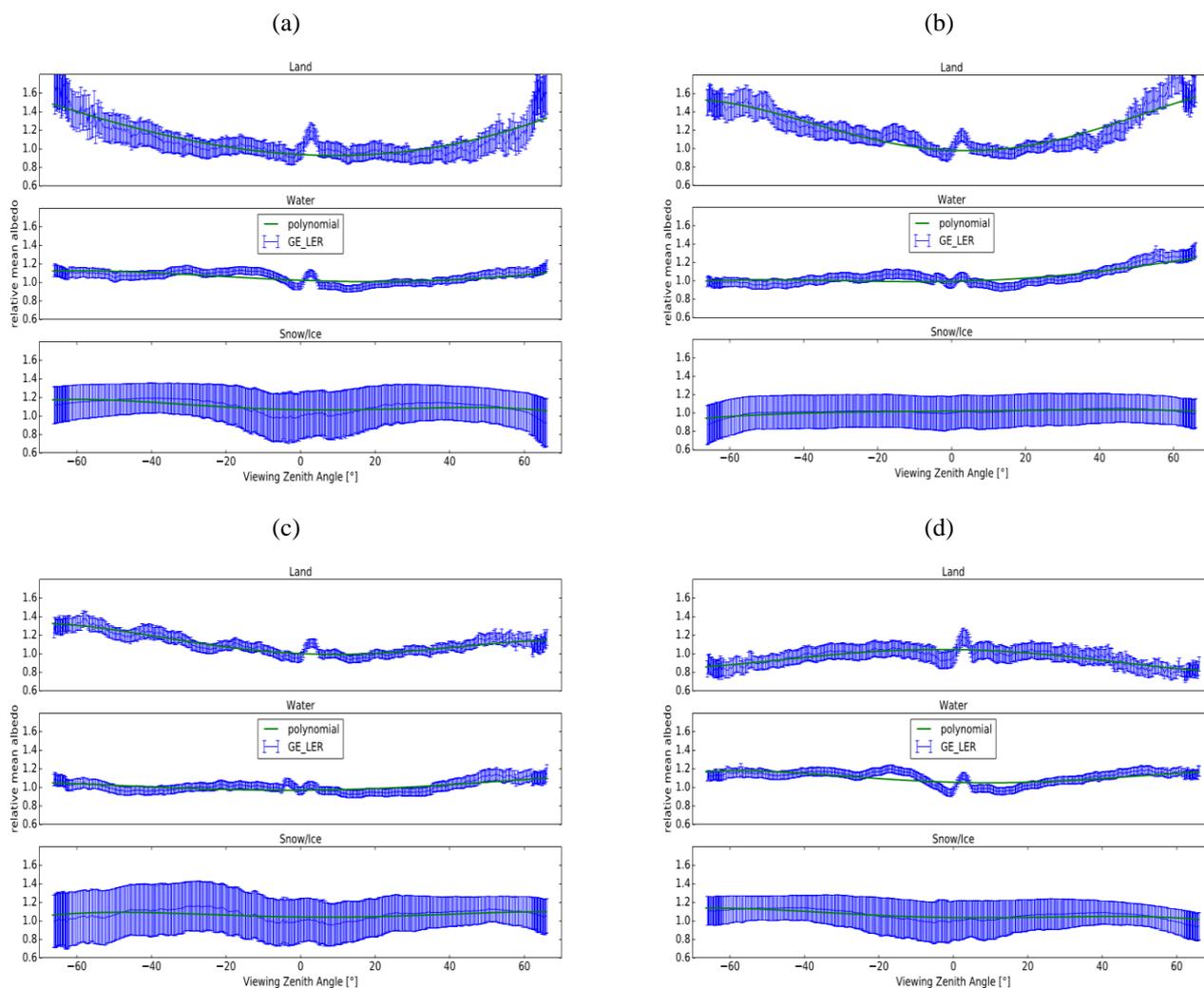
Figure 8: BRDF dependencies ρ(θ) as function of the viewing zenith angle for land, water, and snow/ice calculated with TROPOMI/S5P data from (a) January, (b) April, (c) July, and (d) October 2018. The negative viewing zenith angles correspond to the first 225 detector pixels. The discontinuity at nadir is due to numerical issues in the radiative transfer model calculations with very small relative azimuth angles.

Atmospheric
Measurement
Techniques
Discussions

(a)   (b)   (c)
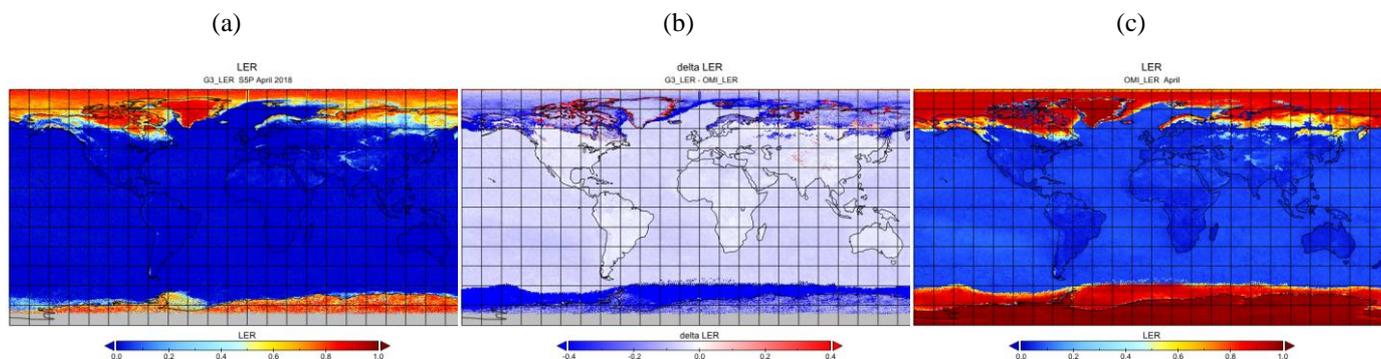
Figure 9: (a) TROPOMI G3_LER daily map corresponding to April 30th, 2018, (c) OMI LER climatology for the month of April, and (b) the difference between these two datasets. There is a very good agreement over land and water surfaces, the mayor differences are due to snow/ice regions in the OMI LER climatology from 2004-2007 that do not match with the actual surface conditions in 2018.

5

Atmospheric
Measurement
Techniques
Open Access

EGU

Discussions

(a)   (b)   (c)

(d)   (e)   (f)
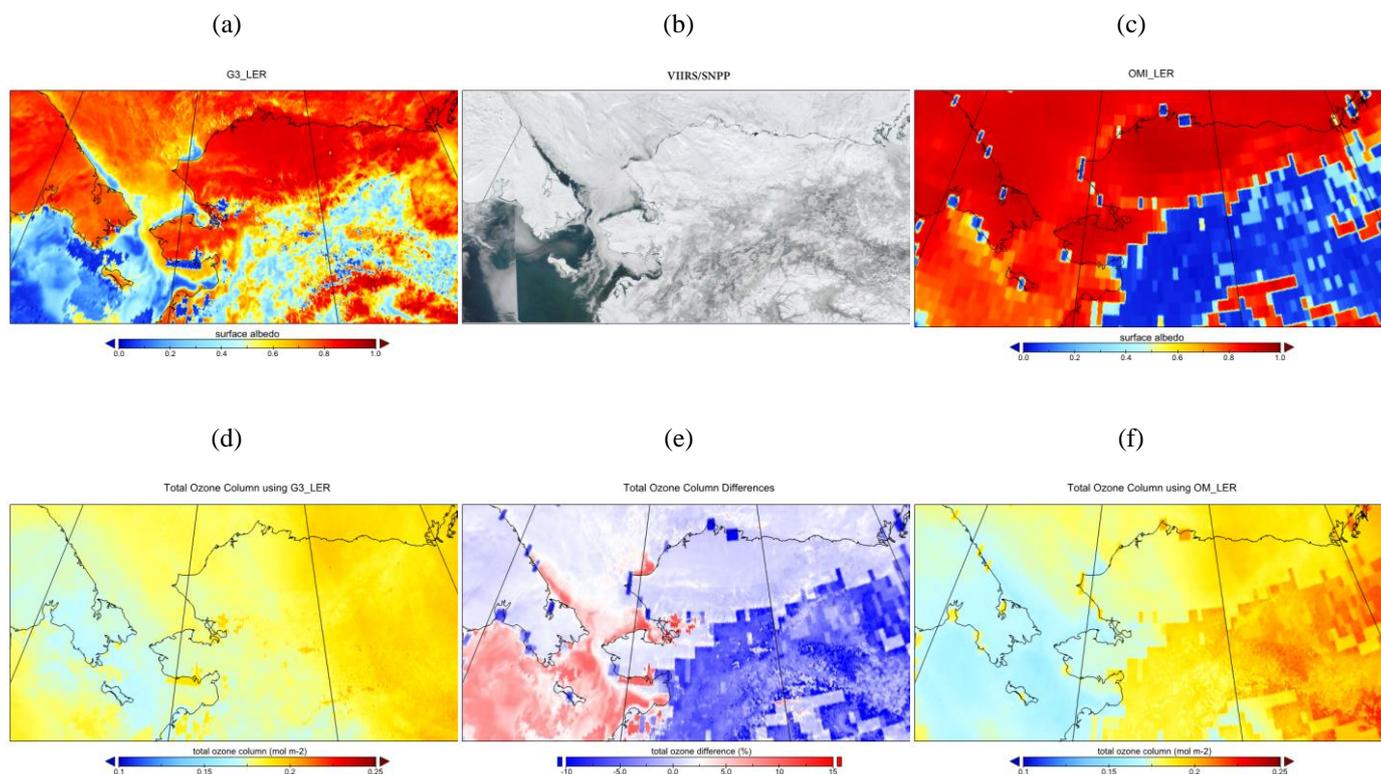
Figure 10: TROPOMI/S5P (top) surface and (bottom) ozone measurements from April 1st, 2018 around the Bering Strait. The (a) TROPOMI/S5P G3_LER daily map agrees very well with the surface types observed in the (b) VIIRS/SNPP image including the water surface along the coasts of Russia and Alaska. These water surfaces along the coast as well as the water of the Bering Sea are not properly represented in the (c) OMI LER climatology that shows snow/ice over these regions. Likewise, the OMI LER climatology erroneously shows no snow/ice in Alaska. The total ozone using the (d) TROPOMI G3_LER daily map is significantly smoother than the corresponding one using the (f) OMI LER climatology. The coarse spatial resolution of the OMI LER climatology is clearly visible in the total ozone field and what is even worst the wrong snow/ice values in the OMI LER climatology induce large errors on the retrieved total ozone (e) with differences between -10% and $+15$%.

5
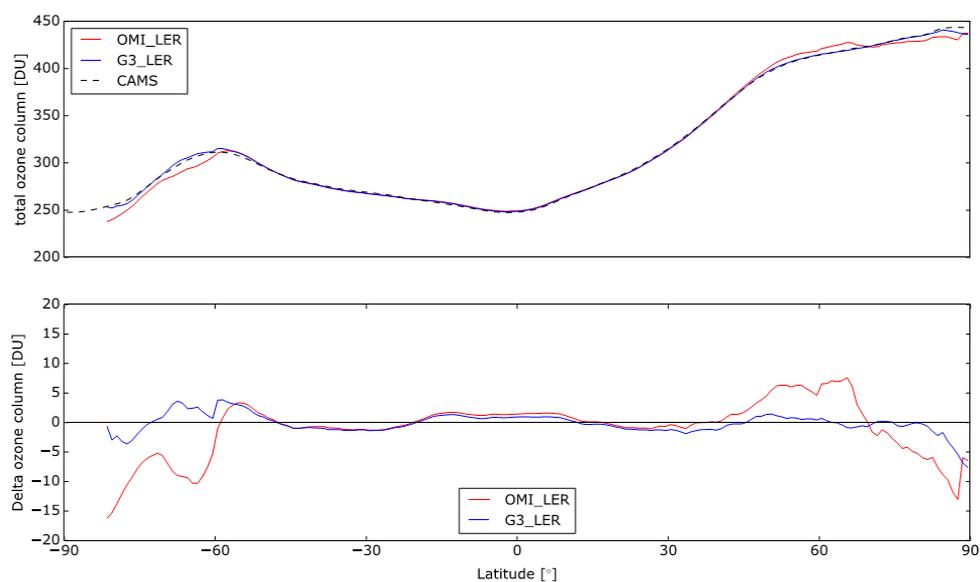
Figure 11: Comparison of total ozone from CAMS and the S5P retrieved ozone using the OMI LER climatology and the daily TROPOMI G3_LER maps for April 2018. The total ozone based on daily G3_LER maps is significantly closer to CAMS especially for the high latitude regions.

10