

# FULL-PHYSICS INVERSE LEARNING MACHINE FOR SATELLITE REMOTE SENSING OF OZONE PROFILE SHAPES AND TROPOSPHERIC COLUMNS

J. Xu\*, K.-P. Heue, M. Coldewey-Egbers, F. Romahn, A. Doicu, D. Loyola

Remote Sensing Technology Institute (IMF), German Aerospace Center (DLR), Oberpfaffenhofen, Germany - jian.xu@dlr.de

Commission III, WG III/8

**KEY WORDS:** Atmospheric composition measurements, Satellite remote sensing, Machine learning, Ozone, UV spectroscopy

## ABSTRACT:

Characterizing vertical distributions of ozone from nadir-viewing satellite measurements is known to be challenging, particularly the ozone information in the troposphere. A novel retrieval algorithm called Full-Physics Inverse Learning Machine (FP-ILM), has been developed at DLR in order to estimate ozone profile shapes based on machine learning techniques. In contrast to traditional inversion methods, the FP-ILM algorithm formulates the profile shape retrieval as a classification problem. Its implementation comprises a training phase to derive an inverse function from synthetic measurements, and an operational phase in which the inverse function is applied to real measurements. This paper extends the ability of the FP-ILM retrieval to derive tropospheric ozone columns from GOME-2 measurements. Results of total and tropical tropospheric ozone columns are compared with the ones using the official GOME Data Processing (GDP) product and the convective-cloud-differential (CCD) method, respectively. Furthermore, the FP-ILM framework will be used for the near-real-time processing of the new European Sentinel sensors with their unprecedented spectral and spatial resolution and corresponding large increases in the amount of data.

## 1. MOTIVATION

Ozone ( $O_3$ ) plays a crucial role in the Earth's atmosphere and its chemical processes (production and destruction) are highly related to climate change and air pollution caused by anthropogenic emissions. Therefore, accurate information of global/regional  $O_3$  vertical distributions over the troposphere and stratosphere turns out to be important to atmospheric environment communities. Satellite remote sensing of  $O_3$  information using the ultraviolet (UV) radiation has been comparatively mature. A number of European satellite sensors, e.g., the Global Ozone Monitoring Experiment (GOME) series, the Scanning Imaging Absorption spectroMeter for Atmospheric CHartographY (SCIAMACHY), and the Ozone Monitoring Instrument (OMI), have mapped the global and regional  $O_3$  distributions. On October 13, 2017, the TROPOspheric Monitoring Instrument onboard the Sentinel-5 Precursor (TROPOMI/S5P), which is one of the next generation of European Copernicus atmospheric composition missions, was launched from Plesetsk, Russia. With its global coverage and open data policy, the mission will support global efforts to monitor atmospheric pollution and to improve our understanding of chemical and physical processes.

These nadir-viewing satellite sensors are preferable to retrieve total column products. Total columns of  $O_3$  can be accurately and efficiently estimated by the Differential Optical Absorption Spectroscopy (DOAS) algorithm that essentially uses the Huggins absorption band (320–360 nm). This conventional approach normally requires an ozone climatology for air mass factor (AMF) calculation, but discrepancies between the climatological profile and the actual vertical distribution could lead to a retrieval error of up to 4 % at high solar zenith angles (SZAs) in derived total column amount (Lerot et al., 2014). As most atmospheric ozone resides in the stratosphere above, tropospheric columns of  $O_3$  can be derived by subtracting an estimate of the stratospheric columns

or by differencing total columns in cloud-free pixels from those in nearby pixels with thick/high convective clouds (the so-called CCD method). The CCD method can only be applied in the tropical region where the assumption of a zonally invariant stratospheric column is valid. A number of relevant studies (Valks et al., 2014, Heue et al., 2016) have been conducted for GOME-2 measurements. Therefore, the retrieval of total and tropospheric  $O_3$  abundances can largely benefit in terms of representativity by obtaining reliable an ozone profile shape.

However, characterizing  $O_3$  profile shapes from nadir-viewing satellite measurements is still known to be challenging, particularly the ozone information in the troposphere. Direct retrieval of tropospheric information has also been investigated and applied to the Global Ozone Monitoring Experiment (GOME) class of instruments. In general, estimating atmospheric parameters of interest directly from spectral measurements is often treated as an ill-posed inverse problem that often requires an iterative inversion of large matrices and multiple calls to radiative transfer calculations. An accurate forward model is important and needs to depict the relationship between atmospheric parameters and measured intensity. Furthermore, this classical inversion method is computationally expensive and often needs additional constraints, reliable a priori knowledge (Rodgers, 2000) and regularization parameterization (Xu et al., 2016) can be decisive to the retrieval outcome. Alternatively, machine learning techniques, such as neural network (NN), Gaussian processes, support vector machines, can learn this relationship quickly through a data-driven training. Although machine learning has been widely used in many research fields, a lot of potential capabilities can be exploited in atmospheric retrieval applications.

To derive  $O_3$  profile shapes from satellite UV measurements in a very fast way, we have developed a novel retrieval algorithm called Full-Physics Inverse Learning Machine (hereafter, FP-ILM) (Xu et al., 2017) and compared the first retrievals with the RAL retrieval using the optimal estimation method (Miles et al., 2015).

\*Corresponding author

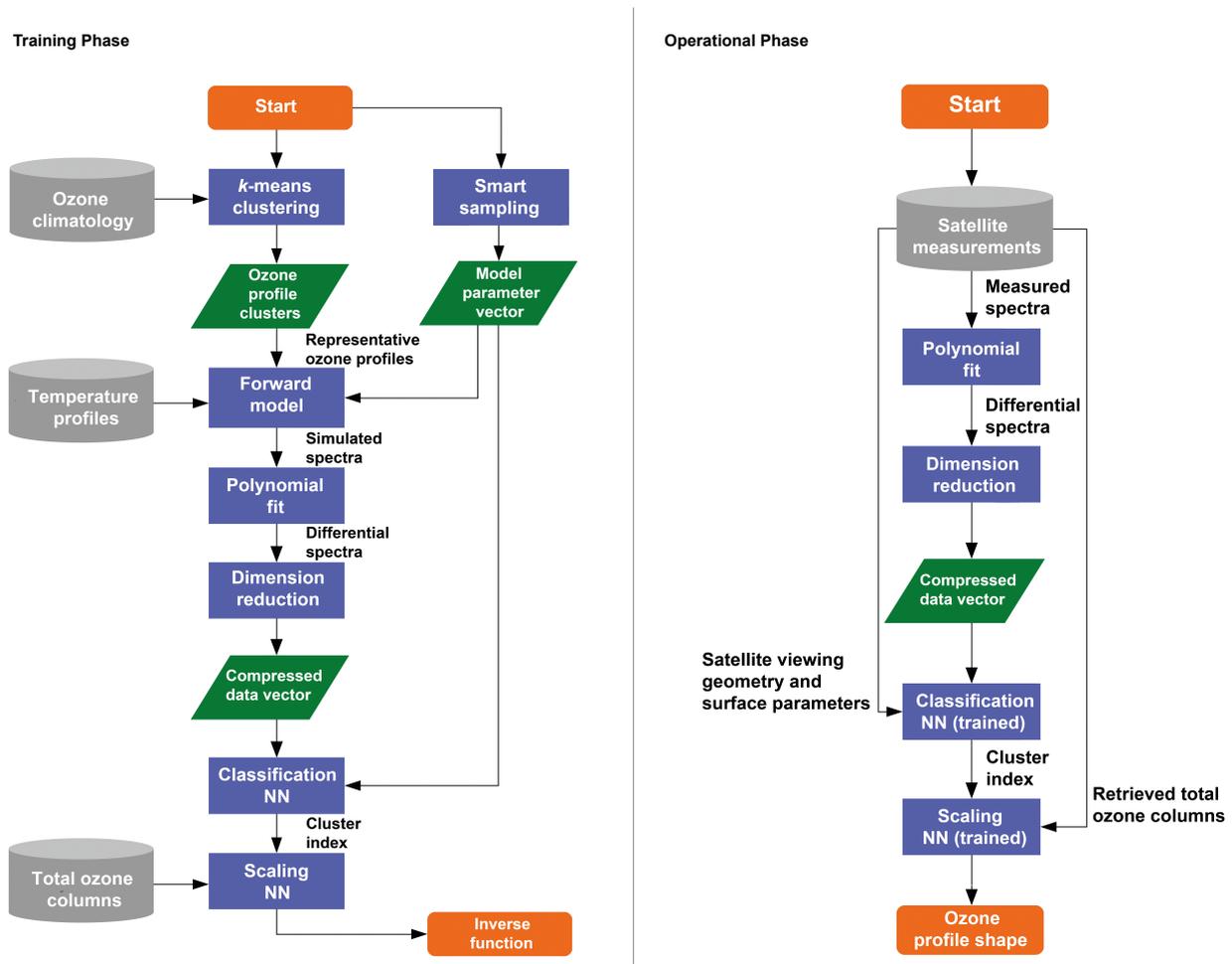


Figure 1. Diagram of FP-ILM for estimating O<sub>3</sub> profile shapes from satellite measurements (Xu et al., 2017, Fig. 1). NN stands for a neural network.

In this paper, we further apply the FP-ILM O<sub>3</sub> profile shape to total column retrieval from GOME-2 measurements and to tropospheric column estimation.

## 2. ALGORITHM DESCRIPTION

The detailed theoretical background and implementation of FP-ILM can be found in the previous work (Xu et al., 2017), and is just summarized here for completeness. Figure 1 depicts a schematic diagram FP-ILM algorithm during the training and operational phases. During the training, the FP-ILM algorithm consists the following main steps:

1. clustering O<sub>3</sub> profile shapes;
2. simulating UV spectra with representative O<sub>3</sub> profiles using “smart-sampling” and a radiative transfer model;
3. obtaining the differential spectra and computing principal components from these spectra;
4. training a NN for classifying the O<sub>3</sub> profile shape corresponding to an input;
5. developing a NN for scaling the O<sub>3</sub> profile based on the given total vertical column density (VCD).

The reference O<sub>3</sub> profiles (volume mixing ratios) were built on the Bodeker database (Bodeker et al., 2013) (“Tier 1.4” in this

study) which were globally taken from eight different spaceborne data sets merged with ozonesonde data. The McPeters/Labow climatology (McPeters and Labow, 2012) (the latest version described in (Labow et al., 2015)) was merged with the Bodeker database in order to provide the tropospheric O<sub>3</sub> concentration.

We used the radiative transfer model VLIDORT (Vector Linearized Discrete Ordinate Radiative Transfer) (Spurr, 2006) that requires the model parameters consisting of the solar zenith angle, viewing zenith angle, relative azimuth angle, surface albedo, and surface pressure. The simulations were done for the wavelength  $\lambda$  range between 290 and 335 nm. In particular, the so-called “smart sampling” (Loyola R et al., 2016) approach was used to generate a minimal number of training samples so that the multi-dimensional input space and the output space can be optimally covered.

The simulated spectra  $\mathbf{y}^\delta(\lambda)$  were converted into the differential spectra  $\mathbf{y}_c^\delta(\lambda)$  by a lower-order polynomial fit:

$$\mathbf{y}_c^\delta(\lambda) = \mathbf{y}^\delta(\lambda) - \mathbf{P}_N(\lambda, \mathbf{p}_c), \quad (1)$$

where  $\mathbf{P}_N(\lambda, \mathbf{p}_c)$  is a polynomial of degree  $N$ . A total of nine principal components were extracted from the original differential spectra with the transformed measurement vector  $\bar{\mathbf{y}}$  being

$$\bar{\mathbf{y}} = \mathbf{U}_M^T \mathbf{y}_c^\delta(\lambda), \quad (2)$$

where  $\mathbf{U}_{\overline{M}} = [\mathbf{u}_1, \dots, \mathbf{u}_{\overline{M}}] \in \mathbb{R}^{M \times \overline{M}}$  is an orthogonal unitary matrix incorporating the  $M$  singular vectors  $\mathbf{u}_k$  of the covariance matrix of the original measurement vectors.

The input vector to the classification NN comprised the five model parameters used in the forward simulations and the nine principal components. The weights and biases of each layer were initialized with the Nguyen-Widrow procedure (Nguyen and Widrow, 1990), and the corresponding training was done by the Scaled Conjugate Gradient backpropagation algorithm (Møller, 1993) which is often used for pattern recognition applications. The input vector to the scaling NN for estimating  $\text{O}_3$  profile shapes comprised only the retrieved total VCD. For each  $\text{O}_3$  profile cluster, a scaling NN was trained with the corresponding  $\text{O}_3$  VCDs using the Nguyen-Widrow initialization procedure the Levenberg-Marquardt backpropagation algorithm (Hagan and Menhaj, 1994).

During the operational phase, we implemented the inverse functions (i.e., both trained NNs) derived from the training phase in the framework of total column retrieval from satellite measurements. Since the AMF/VCD conversion is an iterative process, the profile shape estimated from the VCD at the current iteration was used to obtain the next iterate. With the newly retrieved VCD, the  $\text{O}_3$  profile shape was further adjusted.

### 3. FIRST RESULTS

In this section, first results of retrieved total columns of  $\text{O}_3$  from GOME-2 onboard the MetOp-A satellite (GOME-2A, hereafter) data using FP-ILM profile shapes are presented. For comparison, we used the operational GOME-2 product generated by the GOME Data Processor (GDP) (Van Roozendael et al., 2006, Loyola et al., 2011, Hao et al., 2014) that relies on the TOMS version 8  $\text{O}_3$  profile climatology (Bhartia, 2003, McPeters et al., 2007).

Figure 2 shows the total VCD retrieval results from GOME-2A data on November 25, 2017. It can be seen that the retrieved VCDs using the two  $\text{O}_3$  profile schemes agree well, revealing that the FP-ILM profile shape used in the total ozone retrieval seems reasonable and may reflect the actual measurement conditions.

The FP-ILM  $\text{O}_3$  profile can be used to obtain tropospheric ozone columns by integrating the partial columns at the layers in the troposphere. The computations using the FP-ILM and CCD schemes (not shown here) were done for monthly averaged values (below 200 hPa) on a  $1.25^\circ$  by  $2.5^\circ$  latitude-longitude grid for the tropical region between  $20^\circ$  N and  $20^\circ$  S. However, the tropospheric columns using the FP-ILM retrieval seem less sensitive to atmospheric variability including the identification of trends, indicating the needs for further investigations.

### 4. CONCLUSIONS

This paper exploited the ability of the FP-ILM algorithm to retrieve total and tropospheric ozone from satellite UV measurements. The retrieved total ozone from GOME-2A data using the FP-ILM profile shape seemed comparable with the one using the TOMS climatology, whereas the tropospheric ozone retrieval using the FP-ILM retrieval showed discrepancies as compared to the CCD method. Future work will focus on improving the algorithm in order to optimize the results.

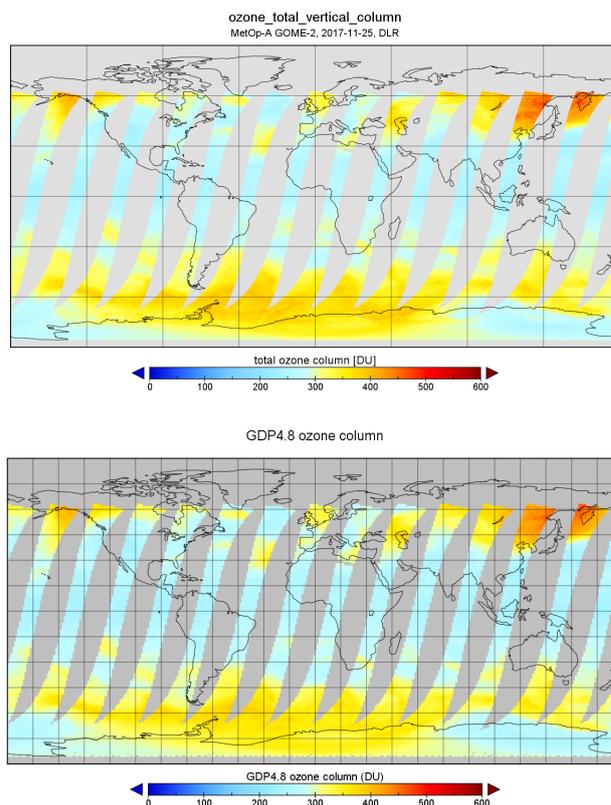


Figure 2. Comparison of retrieved total  $\text{O}_3$  columns from GOME-2A data between the DOAS retrieval using the FP-ILM profile (top) and the GDP 4.8 product (bottom) on November 25, 2017.

### REFERENCES

- Bhartia, P. K., 2003. OMI Algorithm Theoretical Basis Document Volume II: OMI Ozone Products. Algorithm theoretical basis document, NASA Goddard Space Flight Center.
- Bodeker, G. E., Hassler, B., Young, P. J. and Portmann, R. W., 2013. A vertically resolved, global, gap-free ozone database for assessing or constraining global climate model simulations. *Earth Syst. Sci. Data* 5(1), pp. 31–43.
- Hagan, M. T. and Menhaj, M. B., 1994. Training feedforward networks with the Marquardt algorithm. *IEEE Trans. Neural Netw.* 5(6), pp. 989–993.
- Hao, N., Koukouli, M. E., Inness, A., Valks, P., Loyola, D. G., Zimmer, W., Balis, D. S., Zyrichidou, I., Van Roozendael, M., Lerot, C. and Spurr, R. J. D., 2014. GOME-2 total ozone columns from MetOp-A/MetOp-B and assimilation in the MACC system. *Atmos. Meas. Tech.* 7(9), pp. 2937–2951.
- Heue, K.-P., Coldewey-Egbers, M., Delcloo, A., Lerot, C., Loyola, D., Valks, P. and Van Roozendael, M., 2016. Trends of tropical tropospheric ozone from 20 years of European satellite measurements and perspectives for the Sentinel-5 Precursor. *Atmos. Meas. Tech.* 9(10), pp. 5037–5051.
- Labow, G. J., Ziemke, J. R., McPeters, R. D., Haffner, D. P. and Bhartia, P. K., 2015. A total ozone-dependent ozone profile climatology based on ozonesondes and Aura MLS data. *J. Geophys. Res. Atmos.* 120(6), pp. 2537–2545.

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. and Zehner, C., 2014. Homogenized total ozone data records from the european sensors GOME/ERS-2, SCIAMACHY/Envisat, and GOME-2/MetOp-A. *J. Geophys. Res. Atmos.* 119(3), pp. 1639–1662.

Loyola, D. G., Koukouli, M. E., Valks, P., Balis, D. S., Hao, N., Van Roozendael, M., Spurr, R. J. D., Zimmer, W., Kiemle, S., Lerot, C. and Lambert, J.-C., 2011. The GOME-2 total column ozone product: Retrieval algorithm and ground-based validation. *J. Geophys. Res. Atmos.* 116(D7), pp. n/a–n/a.

Loyola R, D. G., Pedernana, M. and Gimeno García, S., 2016. Smart sampling and incremental function learning for very large high dimensional data. *Neural Netw.* 78, pp. 75–87. Special Issue on "Neural Network Learning in Big Data".

McPeters, R. D. and Labow, G. J., 2012. Climatology 2011: An MLS and sonde derived ozone climatology for satellite retrieval algorithms. *J. Geophys. Res. Atmos.* 117(D10), pp. n/a–n/a.

McPeters, R. D., Labow, G. J. and Logan, J. A., 2007. Ozone climatological profiles for satellite retrieval algorithms. *J. Geophys. Res. Atmos.*

Miles, G. M., Siddans, R., Kerridge, B. J., Latter, B. G. and Richards, N. A. D., 2015. Tropospheric ozone and ozone profiles retrieved from GOME-2 and their validation. *Atmos. Meas. Tech.* 8(1), pp. 385–398.

Møller, M. F., 1993. A scaled conjugate gradient algorithm for fast supervised learning. *Neural Netw.* 6(4), pp. 525–533.

Nguyen, D. and Widrow, B., 1990. Improving the learning speed of 2-layer neural networks by choosing initial values of the adaptive weights. In: *1990 IJCNN International Joint Conference on Neural Networks*, Vol. 3, pp. 21–26.

Rodgers, C., 2000. *Inverse Methods for Atmospheric Sounding: Theory and Practise*. World Scientific, Singapore.

Spurr, R., 2006. 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(2), pp. 316–342.

Valks, P., Hao, N., Gimeno Garcia, S., Loyola, D., Dameris, M., Jöckel, P. and Delcloo, A., 2014. Tropical tropospheric ozone column retrieval for GOME-2. *Atmos. Meas. Tech.* 7(8), pp. 2513–2530.

Van Roozendael, M., Loyola, D., Spurr, R., Balis, D., Lambert, J.-C., Livschitz, Y., Valks, P., Ruppert, T., Kenter, P., Fayt, C. and Zehner, C., 2006. Ten years of GOME/ERS-2 total ozone data: the new GOME data processor (GDP) version 4: 1. Algorithm description. *J. Geophys. Res. Atmos.* 111(D14), pp. n/a–n/a.

Xu, J., Schreier, F., Doicu, A. and Trautmann, T., 2016. Assessment of Tikhonov-type regularization methods for solving atmospheric inverse problems. *J. Quant. Spectrosc. Radiat. Transf.* 184, pp. 274–286.

Xu, J., Schüssler, O., Loyola R, D., Romahn, F. and Doicu, A., 2017. 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.