Fast Cloud Property Retrieval from TROPOMI O_2-A Band Observations Using a DISAMAR-Based Neural Network Framework

Tao Xie[1], Xiaoyun Zhang[2,3], Wenmei Li[1], Yixiang Chen[1], Ping Wang[2], Piet Stammes[2], Gijsbert Tilstra[2], Olaf Tuinder[2], Maarten Sneep[2], Feng Lu[4]

1. School of Internet of Things, Nanjing University of Posts and Telecommunications, Jiangsu, China - xtplqk@126.com
2. R&D Satellite Observations (RDSW), Royal Netherlands Meteorological Institute (KNMI)
3. Nanjing University of Information Science and Technology (NUIST), Jiangsu, China
4. Key Laboratory of Radiometric Calibration and Validation for Environmental Satellites, National Satellite Meteorological Center (National Center for Space Weather), Innovation Centre for Feng Yun Meteorological Satellite (FYSIC), China Meteorological Administration, Beijing 100049, China

Keywords: TROPOMI, O_2-A band, DISAMAR, neural-network emulators, cloud optical thickness, cloud optical centroid pressure

Abstract

With the improved spatial resolution of satellite spectrometers such as TROPOMI, more homogeneous cloudy scenes can be identified at the pixel scale, making scattering-cloud retrievals increasingly worthwhile for satellite cloud processing. DISAMAR provides physically rigorous and spectrally accurate simulations for cloud retrieval from oxygen absorption bands, but its line-by-line radiative transfer calculations are computationally too expensive for efficient large-scale application. In this study, we develop a fast cloud retrieval framework for TROPOMI O_2-A band observations by combining DISAMAR with neural-network emulators. The method supports the joint retrieval of cloud optical thickness (COT) and cloud optical centroid pressure (CLP) within an optimal-estimation framework. The neural networks are trained offline using a synthetic dataset generated by DISAMAR under representative cloud, surface, meteorological, and viewing conditions. The dataset contains about 400,000 cases, with meteorological inputs from ERA5 and observation geometries sampled from a TROPOMI orbit; about 70% of the cases correspond to land and 30% to water surfaces. The networks emulate both top-of-atmosphere reflectance and its derivatives with respect to the retrieval state variables, thereby replacing the most time-consuming part of the forward model. Initial experiments for fully cloudy O_2-A cases show that the proposed framework reproduces the main spectral and retrieval behaviour of the DISAMAR-based scheme while reducing the runtime from several hours per case to a few seconds. The proposed framework provides a practical pathway toward fast and physically consistent retrievals of cloud optical properties from TROPOMI O_2-A observations.

1. Introduction

Clouds strongly affect photon path length, scene radiance, and the shielding of lower atmospheric layers, making them a major source of uncertainty in satellite trace-gas retrievals. Reliable cloud information is therefore essential for cloud correction and quality control in UV–visible–near-infrared remote sensing. In oxygen absorption bands, cloud parameters can be inferred from O_2 spectral features and have long been used in operational cloud algorithms (Schuessler et al., 2014; Stammes et al., 2008).

The FRESCO family has been widely used for cloud retrieval from the O_2-A band (Wang et al., 2008). For TROPOMI (Tropospheric Monitoring Instrument), the operational FRESCO-S product provides effective cloud information from selected spectral windows but represents clouds as a Lambertian reflecting boundary (Van Geffen et al., 2022). Although efficient, this treatment is simplified. With the improved spatial resolution of TROPOMI, more homogeneous cloudy scenes can be identified, making scattering-cloud retrievals more meaningful at the pixel scale. DISAMAR (Determining Instrument Specifications and Analysing Methods for Atmospheric Retrieval) supports this more physical cloud representation and provides line-by-line radiative transfer simulations within an optimal-estimation framework, but its computational cost is too high for large-scale application (de Haan et al., 2022).

To address this limitation, we accelerate the DISAMAR-based retrieval using neural-network emulators. Rather than predicting the final cloud product directly, the neural networks emulate top-of-atmosphere spectra and the derivatives required by the inversion. This preserves the physics-based retrieval structure while greatly improving efficiency. In this paper, we present a DISAMAR-based neural-network framework for fast cloud retrieval from TROPOMI O_2-A measurements, focusing on the joint retrieval of cloud optical thickness (COT) and cloud optical centroid pressure (CLP), and evaluate its performance using selected fully cloudy cases.

2. Methodology

The overall workflow of the proposed framework is illustrated in Figure 1. Starting from a DISAMAR-based scattering-cloud forward model, high-resolution O_2-A spectra and their derivatives with respect to the retrieval variables are simulated for a large number of representative scenarios. These simulations are then used to train neural-network emulators of reflectance and Jacobians. In the final step, the trained emulators are coupled to an optimal-estimation scheme, so that the DISAMAR-based retrieval can be performed much more efficiently while retaining its physical structure. The present implementation focuses on the joint retrieval of cloud optical thickness and cloud optical centroid pressure from TROPOMI O_2-A observations.

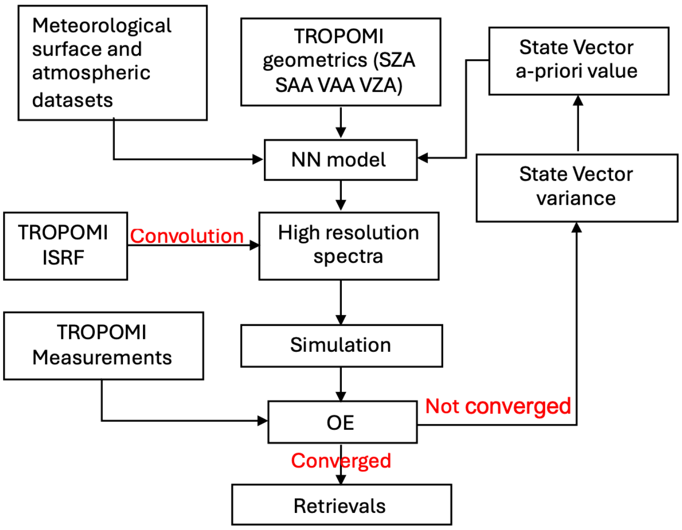

Figure 1. Overall workflow of the proposed DISAMAR-based neural-network framework for TROPOMI O_2-A cloud retrieval.

2.1 DISAMAR-based scattering-cloud forward model

DISAMAR (Determining Instrument Specifications and Analysing Methods for Atmospheric Retrieval) is a line-by-line radiative transfer model developed for atmospheric retrieval studies from passive Earth observations (de Haan et al., 2022). In the present study, the cloud is represented as a homogeneous layer of scattering particles. The scattering phase function is approximated using the Henyey–Greenstein formulation with an asymmetry factor of $g = 0.85$. In DISAMAR, the atmosphere is vertically divided into pressure intervals, and each interval is further subdivided into homogeneous layers for radiative transfer calculations. This configuration allows accurate simulations of reflectance and its derivatives, but repeated forward-model evaluations remain computationally expensive. The O_2-A fitting window used here is 757–770 nm, and the absorption database is HITRAN2020 (High-resolution Transmission Molecular Absorption Database 2020 edition) (Gordon et al., 2022).

2.2 Neural-network emulators and training dataset

To accelerate the retrieval, neural-network emulators are trained offline using synthetic spectra generated by the DISAMAR forward model. The emulators are designed to replace the most time-consuming part of the forward simulation while preserving the physical inversion framework. Instead of retrieving cloud properties directly, the neural networks predict top-of-atmosphere reflectance and the derivatives of reflectance with respect to the state variables. These predicted quantities are then passed to the optimal-estimation retrieval.

In the O_2-A retrieval setup considered in this study, the state vector consists of cloud optical thickness (COT) and cloud optical centroid pressure (CLP). Additional forward-model settings are prescribed according to the DISAMAR configuration used to generate the training dataset. The emulator architecture is a fully connected sequential neural network. In the current development tests, reflectance and derivative networks were trained separately using a common input-feature design, and the neural networks typically use three hidden layers with representative configurations of 400 neurons per layer, depending on the target variable and experiment.

2.3 Training dataset design

The training dataset is generated with DISAMAR and contains about 400,000 synthetic cases. The distributions of representative scenario parameters in the DISAMAR-generated O_2-A training dataset are shown in Figure 2. About 70% of the samples correspond to land surfaces and about 30% to water surfaces. The observation geometry, including solar zenith angle, solar azimuth angle, view zenith angle, and view azimuth angle, is selected randomly from one TROPOMI orbit. Meteorological variables are taken from ERA5, including surface pressure and temperature profile information. Surface reflectance information is also included as part of the scenario design. The training targets include reflectance and derivatives of reflectance with respect to the state variables. The purpose of this dataset design is to cover representative cloud, surface, meteorological, and viewing conditions while maintaining consistency with the DISAMAR forward model. The main settings of the O_2-A forward model, training dataset, and validation configuration are summarized in Table 1.

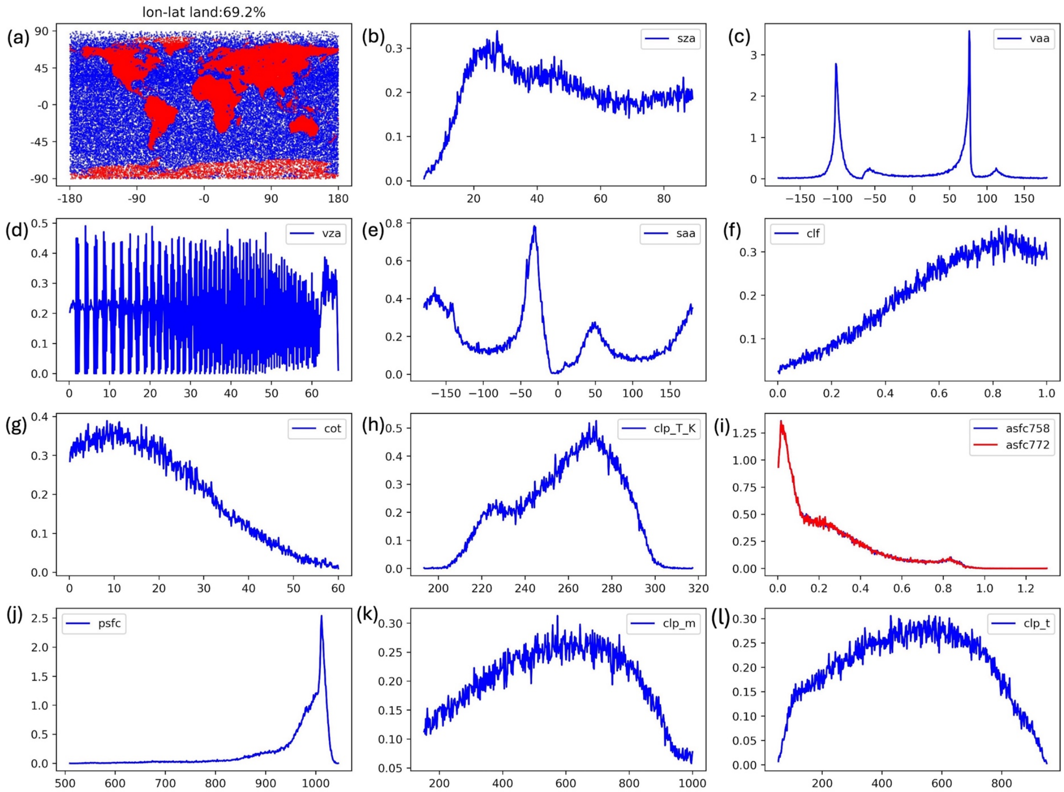

Figure 2. Histograms of representative scenario parameters in the DISAMAR-generated O_2-A training dataset.

Table 1. Main settings of the O_2-A forward model, training dataset, and validation configuration.

Group	**Setting**	Description
O_2-A forward model	Spectral window	757–770 nm
O_2-A forward model	Absorption database	HITRAN2020
O_2-A forward model	Cloud phase function	Henyey–Greenstein
O_2-A forward model	Asymmetry factor	g = 0.85
Training dataset	Total number of cases	400,000
Training dataset	Surface type distribution	70% land 30% water
Training dataset	Observation geometry	randomly selected from one TROPOMI orbit
Training dataset	Meteorological input	ERA5 surface pressure ERA5 temperature profile
Training dataset	Surface albedo input	scenario-based surface albedos from DLER
Training dataset	Training targets	reflectance derivatives of reflectance with respect to the state variables
Neural-network emulator	Network type	fully connected sequential neural network
Neural-network emulator	Representative NN architecture	3 hidden layers 400 neurons per layer
Validation setup	Validation cases	40 fully cloudy cases
Validation setup	Case screening	skip cloud edge SZA < 75
Validation setup	Inputs for retrieval validation	TROPOMI radiance and irradiance, per-pixel geometry DLER surface albedos ERA5 surface pressure and temperature profile ISRF
Validation setup	Cloud-interval pressure thickness	100 hPa
Validation setup	A-priori COT & variance	30 & 5

2.4 Retrieval framework

The trained neural-network emulators are embedded into an optimal-estimation retrieval framework. In this study, we focus on jointly retrieving COT and CLP from TROPOMI O_2-A band spectra. The neural-network forward model supplies both the simulated reflectance and the required derivatives, so the inversion can proceed in the same general way as in the DISAMAR-based retrieval, but with greatly reduced computational cost. The framework therefore combines the physical basis of radiative transfer modelling with the efficiency of machine-learning emulation.

3. Data and Experimental Setup

3.1 TROPOMI O_2-A observations

TROPOMI (Tropospheric Monitoring Instrument) on board Sentinel-5P is a nadir-viewing spectrometer with spectral coverage from the ultraviolet to the shortwave infrared. In this study, only measurements in the O_2-A band are considered. The fitting window is 757–770 nm. TROPOMI Level-1B radiance and irradiance spectra are used together with per-pixel observation geometry and instrument spectral response information (Veefkind et al., 2012). These measurements provide the observational basis for the O_2-A cloud retrieval experiments presented in this study.

3.2 Auxiliary data

Several auxiliary datasets are used to support the forward-model configuration and retrieval setup. ERA5, the fifth-generation atmospheric reanalysis of the European Centre for Medium-Range Weather Forecasts (ECMWF), is used to provide meteorological information, including surface pressure and atmospheric temperature profiles (Hersbach et al., 2020). Surface reflectance information is taken from the TROPOMI DLER product, where DLER denotes directionally dependent Lambertian-equivalent reflectivity (Tilstra et al., 2021). These auxiliary datasets are used to construct realistic retrieval scenarios and to maintain consistency between the DISAMAR simulations and the TROPOMI observations. In addition, the O_2-A spectral simulations are based on HITRAN2020 (High-resolution Transmission Molecular Absorption Database 2020 edition).

3.3 Validation datasets

Several comparison datasets are used to evaluate the retrieval performance of the proposed framework. For cloud optical thickness (COT), the retrieval results are compared with the corresponding DISAMAR retrievals and with the COT product from VIIRS (Visible Infrared Imaging Radiometer Suite) on board the Suomi National Polar-orbiting Partnership (Suomi-NPP) satellite (Platnick et al., 2021). Because the VIIRS cloud product has a much finer spatial resolution than TROPOMI, the VIIRS pixels within each TROPOMI footprint are aggregated to obtain TROPOMI-scale comparison values. For cloud optical centroid pressure (CLP), the results are compared with the TROPOMI operational FRESCO-S product and the corresponding DISAMAR retrievals. In this study, the DISAMAR retrievals are used as the full-physics baseline for evaluating the neural-network-accelerated retrieval.

3.4 Selection of validation cases

The initial evaluation presented in this study is based on 40 selected fully cloudy TROPOMI O_2-A cases used for retrieval validation. Cloud-edge pixels are excluded to reduce the influence of sub-pixel heterogeneity, and scenes with large solar zenith angles are omitted because retrieval convergence becomes more difficult under such conditions. The analysis is further restricted to optically thick cloudy scenes so that the performance of the proposed framework can be assessed under relatively controlled conditions. For these selected cases, the retrieval results are evaluated against the corresponding DISAMAR retrievals as the full-physics baseline, together with the external comparison datasets described in Section 3.3. This restricted setup is intended to demonstrate the feasibility of the proposed framework before extending the method to more diverse and more challenging cloud conditions.

4. Results

4.1 Emulator performance

The neural-network emulators reproduce the main spectral behaviour of the DISAMAR simulations in the O_2-A fitting window. The training and validation performance of the neural-network emulators is summarized in Figure 3. Development results show stable reflectance-learning performance and demonstrate that derivative networks can also be trained for the required Jacobian information. Example predicted high-resolution spectra indicate that the neural-network outputs capture the major reflectance structures in the O_2-A band and can therefore serve as efficient surrogates for the DISAMAR forward calculations. At the same time, the development summary also shows that the training loss may fluctuate, indicating that additional work is still needed to improve model robustness.

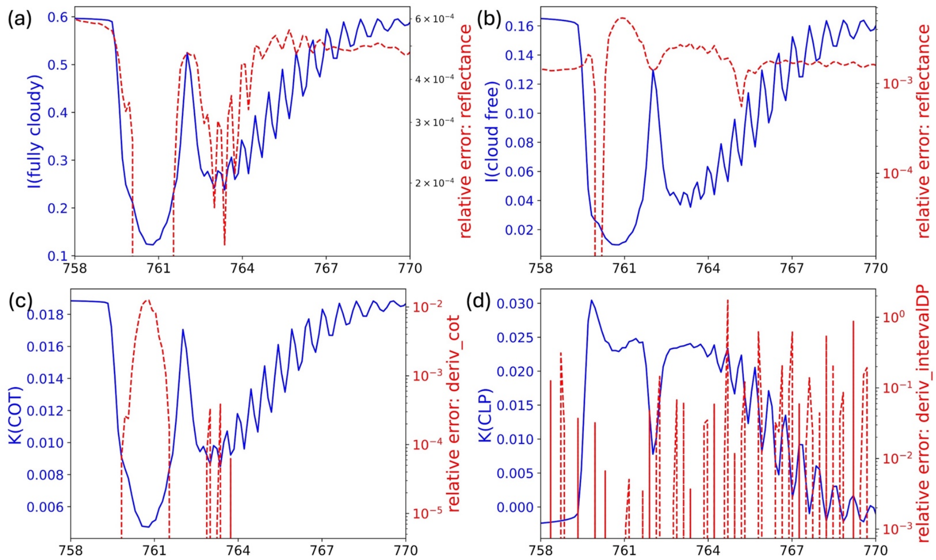

Figure 3. Training and validation performance of the neural-network emulators for O_2-A reflectance and derivative predictions.

4.2 Retrieval performance for fully cloudy O_2-A cases

For the selected fully cloudy O_2-A cases, the retrieval results from the proposed framework are evaluated against several comparison datasets. For cloud optical thickness (COT), the retrieved values are compared with the corresponding DISAMAR retrievals and with the COT product from Suomi-NPP VIIRS (labelled as “NPPC” in Figure 4). For cloud optical centroid pressure (CLP), the results are compared with the FRESCO-S product (labelled as “FRESCO” in Figure 4), the corresponding DISAMAR retrievals, and the Suomi-NPP VIIRS CTP product. Overall, these comparisons indicate that the proposed framework reproduces the main behaviour of the DISAMAR-based retrievals and shows good consistency with the external cloud products.

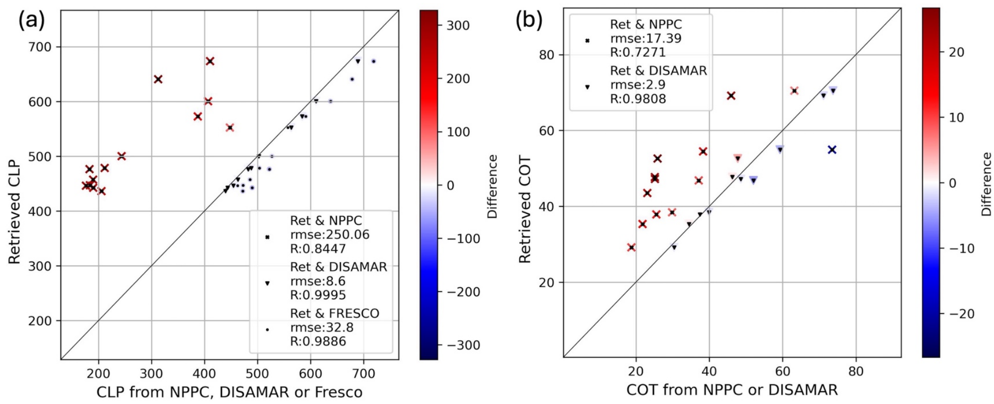

Figure 4. Comparison of retrieved cloud optical thickness (COT) and cloud optical centroid pressure (CLP) for selected fully cloudy O_2-A cases against FRESCO-S (labelled as “FRESCO” in the figure), DISAMAR, and the Suomi-NPP VIIRS cloud product (labelled as “NPPC” in the figure).

According to the current development results, the retrieved COT shows strong consistency with the comparison datasets, while the retrieved CLP agrees well with the values from FRESCO-S, DISAMAR, and the Suomi-NPP VIIRS CTP product. The fully cloudy O_2-A experiments show correlation coefficients of about 0.98 and 0.73 for COT relative to DISAMAR COT and the Suomi-NPP VIIRS COT product, respectively, and about 0.99, approximately 1.0, and 0.84 for CLP relative to FRESCO-S, DISAMAR, and the Suomi-NPP VIIRS CTP product, respectively. Overall, these results indicate that the proposed framework is capable of supporting physically meaningful retrievals of both COT and CLP under fully cloudy O_2-A conditions.

4.3 Computational efficiency

A major advantage of the proposed framework is the reduction in computational cost. The experiments indicate that replacing the line-by-line DISAMAR forward model with a neural-network emulator reduces the runtime for one case from several hours to only a few seconds. This improvement is essential for any future large-scale or near-real-time application of scattering-cloud retrievals. At the same time, the current development notes suggest that faster training data generation using fewer Gaussian points for the polar angle may affect convergence, so further optimization of both training data and model configuration remains necessary.

5. Discussion

The present results demonstrate the feasibility of combining DISAMAR with neural-network emulators for fast cloud retrieval from TROPOMI O_2-A observations. The framework is physically attractive because the neural network is used to accelerate the forward model rather than to replace the retrieval logic itself. In this sense, the method preserves the radiative-transfer-based structure of the inversion while greatly improving speed.

Several issues still need further study. First, the representativeness of the synthetic training dataset is important for generalization to real observations. Although the current dataset includes a wide range of cloud, surface, meteorological, and geometric conditions, additional validation against more independent cases is still needed. Second, retrieval accuracy depends not only on the quality of the emulated reflectance but also on the fidelity of the emulated derivatives. Since the Jacobians directly affect the optimal-estimation iterations, any systematic emulator error may propagate into the retrieved COT and CLP. Third, the scattering-cloud model used here remains an idealized description of real clouds. Horizontally inhomogeneous or multilayer cloud scenes are likely to be more challenging than the fully cloudy test cases presented in this paper.

For the conference-paper version, our goal is to demonstrate the feasibility of the DISAMAR-based neural-network framework and its initial performance for O_2-A fully cloudy retrievals. More comprehensive uncertainty analysis, independent validation, and broader scene testing will be pursued in future work.

6. Conclusions

We developed a DISAMAR-based neural-network framework for fast cloud retrieval from TROPOMI O_2-A band observations. By emulating both reflectance and its derivatives within an optimal-estimation framework, the proposed method supports joint retrievals of cloud optical thickness and cloud optical centroid pressure while preserving the physical structure of the DISAMAR-based retrieval. Initial results for fully cloudy O_2-A cases show that the framework reproduces the main retrieval behavior of the DISAMAR-based scheme and reduces runtime from several hours per case to a few seconds. These results indicate that the proposed approach is a promising pathway toward fast and physically consistent scattering-cloud retrievals for TROPOMI and future atmospheric satellite missions.

https://doi.org/10.5194/isprs-archives-XLIX-B3-2026-1487-2026

Acknowledgements

This work was supported in part by the National Natural Science Foundation of China Project (42176180) *(Corresponding author: Tao Xie)* and in part by the China Scholarship Council (CSC). We gratefully acknowledge the use of the European Weather Cloud (EWC) in the context of this research. We appreciate Olaf Tuinder (KNMI) for helping with the EWC environment. The authors would like to thank Johan de Haan for discussions and guidance.

References

de Haan, J.F., Wang, P., Sneep, M., Veefkind, J.P., Stammes, P., 2022. Introduction of the DISAMAR radiative transfer model: Determining instrument specifications and analysing methods for atmospheric retrieval (version 4.1.5). Geosci. Model Dev. 15, 7031–7050. https://doi.org/10.5194/gmd-15-7031-2022

Gordon, I.E., Rothman, L.S., Hargreaves, R.J., Hashemi, R., Karlovets, E. V., Skinner, F.M., Conway, E.K., Hill, C., Kochanov, R. V., Tan, Y., Wcisło, P., Finenko, A.A., Nelson, K., Bernath, P.F., Birk, M., Boudon, V., Campargue, A., Chance, K. V., Coustenis, A., Drouin, B.J., Flaud, J.M., Gamache, R.R., Hodges, J.T., Jacquemart, D., Mlawer, E.J., Nikitin, A. V., Perevalov, V.I., Rotger, M., Tennyson, J., Toon, G.C., Tran, H., Tyuterev, V.G., Adkins, E.M., Baker, A., Barbe, A., Canè, E., Császár, A.G., Dudaryonok, A., Egorov, O., Fleisher, A.J., Fleurbaey, H., Foltynowicz, A., Furtenbacher, T., Harrison, J.J., Hartmann, J.M., Horneman, V.M., Huang, X., Karman, T., Karns, J., Kassi, S., Kleiner, I., Kofman, V., Kwabia-Tchana, F., Lavrentieva, N.N., Lee, T.J., Long, D.A., Lukashevskaya, A.A., Lyulin, O.M., Makhnev, V.Y., Matt, W., Massie, S.T., Melosso, M., Mikhailenko, S.N., Mondelain, D., Müller, H.S.P., Naumenko, O. V., Perrin, A., Polyansky, O.L., Raddaoui, E., Raston, P.L., Reed, Z.D., Rey, M., Richard, C., Tóbiás, R., Sadiek, I., Schwenke, D.W., Starikova, E., Sung, K., Tamassia, F., Tashkun, S.A., Vander Auwera, J., Vasilenko, I.A., Vigasin, A.A., Villanueva, G.L., Vispoel, B., Wagner, G., Yachmenev, A., Yurchenko, S.N., 2022. The HITRAN2020 molecular spectroscopic database. J. Quant. Spectrosc. Radiat. Transf. 277. https://doi.org/10.1016/j.jqsrt.2021.107949

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., De Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R.J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., Thépaut, J.N., 2020. The ERA5 global reanalysis. Quarterly Journal of the Royal Meteorological Society 146, 1999–2049. https://doi.org/10.1002/qj.3803

Platnick, S., Meyer, K., Wind, G., Holz, R.E., Amarasinghe, N., Hubanks, P.A., Marchant, B., Dutcher, S., Veglio, P., 2021. The nasa modis-viirs continuity cloud optical properties products. Remote Sens. (Basel). 13, 1–30. https://doi.org/10.3390/RS13010002

Schuessler, O., Loyola Rodriguez, D.G., Doicu, A., Spurr, R., 2014. Information content in the oxygen A-band for the retrieval of macrophysical cloud parameters. IEEE Transactions on Geoscience and Remote Sensing 52, 3246–3255. https://doi.org/10.1109/TGRS.2013.2271986

Stammes, P., Sneep, M., de Haan, J.F., Veefkind, J.P., Wang, P., Levelt, P.F., 2008. Effective cloud fractions from the Ozone Monitoring Instrument: Theoretical framework and validation. Journal of Geophysical Research Atmospheres 113. https://doi.org/10.1029/2007JD008820

Tilstra, L.G., Tuinder, O.N.E., Wang, P., Stammes, P., 2021. Directionally dependent Lambertian-equivalent reflectivity (DLER) of the Earth’s surface measured by the GOME-2 satellite instruments. Atmos. Meas. Tech. 14, 4219–4238. https://doi.org/10.5194/AMT-14-4219-2021

Van Geffen, J., Eskes, H., Compernolle, S., Pinardi, G., Verhoelst, T., Lambert, J.C., Sneep, M., Linden, M. Ter, Ludewig, A., Folkert Boersma, K., Pepijn Veefkind, J., 2022. Sentinel-5P TROPOMI NO2retrieval: impact of version v2.2 improvements and comparisons with OMI and ground-based data. Atmos. Meas. Tech. 15, 2037–2060. https://doi.org/10.5194/amt-15-2037-2022
Veefkind, J.P., Aben, I., McMullan, K., Förster, H., de Vries, J., Otter, G., Claas, J., Eskes, H.J., de Haan, J.F., Kleipool, Q., van Weele, M., Hasekamp, O., Hoogeveen, R., Landgraf, J., Snel, R., Tol, P., Ingmann, P., Voors, R., Kruizinga, B., Vink, R., Visser, H., Levelt, P.F., 2012. TROPOMI on the ESA Sentinel-5 Precursor: A GMES mission for global observations of the atmospheric composition for climate, air quality and ozone layer applications. Remote Sens. Environ. 120, 70–83. https://doi.org/10.1016/j.rse.2011.09.027

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