Royal Netherlands
Meteorological Institute
*Ministry of Infrastructure*
*and Water Management*

# Algorithm Theoretical Basis Document for the OMI L2 OMNO2A Data Processor Collection 4


document number : AURA-OMI-KNMI-L2-0300-SD
version : 4.0.0.488
date : 2026-02-23

**Contributing authors**

| | |
|---|---|
| Jos van Geffen | KNMI |
| Henk Eskes | KNMI |
| Mark ter Linden | KNMI |
| Folkert Boersma | KNMI & WUR |
| Pepijn Veefkind | KNMI & TUDelft |

# Contents

# List of Tables



# List of Figures

# 1 Introduction

## 1.1 Identification

This document, identified as AURA-OMI-KNMI-L2-0300-SD, is the Algorithm Theoretical Basis Document for the KNMI OMI L2 Nitrogen Dioxide Slant Column (OMNO2A) Data Processor Collection 4.

The subsequent Vertical Columm calculation (OMNO2B) using the DOMINO data assimilation / chemistry transport model system will be described in a separate ATBD in due time.

The OMNO2A processor produces output files that contain data variables for the OMNO2A and OMNO2B results, where the latter will be left empty (i.e. contain fill values), while the processing and quality flags contain things raised by the OMNO2A step.

The OMNO2A data product DOI is: 10.5067/AURA/OMI/DATA2433

## 1.2 Purpose and objective

The purpose of this document is to describe the algorithm that is used to create the OMNO2A Collection 4 data products.

## 1.3 Document overview

Section 2 provides the reference documents and papers.
Section 3 provides terms, definitions and abbreviated terms.
Section 4 provides an overview of OMI instrument.
Section 5 provides the description of the algorithm.

# 2 Reference documents

[1] Terms and Definitions for the OMI KNMI L2 Data Processors Collection 4.
**source:** KNMI; **ref:** AURA-OMI-KNMI-L2-0001-LI.

[2] M. R. Dobber, R. J. Dirksen, P. F. Levelt *et al.*; Ozone monitoring instrument calibration. *IEEE T GEOSCI REMOTE*; **44** (2006) (5), 1209; 10.1109/TGRS.2006.869987.

[3] Algorithm Theoretical Basis Document for the collection 4 L01b data processing of the Ozone Monitoring Instrument.
**source:** KNMI; **ref:** AURA-OMI-KNMI-L01B-0002-SD; **date:** 2021-07-16.

[4] TROPOMI ATBD of the total and tropospheric NO2 data products.
**source:** KNMI; **ref:** S5P-KNMI-L2-0005-RP; **issue:** 2.4.0; **date:** 2022-07-11.

[5] K. F. Boersma, H. J. Eskes, A. Richter *et al.*; Improving algorithms and uncertainty estimates for satellite NO2 retrievals: Results from the Quality Assurance for Essential Climate Variables (QA4ECV) project Variables (QA4ECV) project. *Atmos. Meas. Tech.*; **11** (2018), 6651; 10.5194/amt-11-6651-2018.

[6] QA4ECV - Quality Assurance for Essential Climate Variables.
**source:** KNMI; **ref:** EU-project 607405, SPA.2013.1.1-03; **date:** November 2012.

[7] QA4ECV website. URL `http://www.qa4ecv.eu/`.

[8] J. J. Moré; The Levenberg-Marquardt Algorithm: Implementation and Theory. *In Numerical Analysis* (edited by G. A. Watson); (pp. 105–116) (Springer, Berlin, 1978); 10.1007/BFb0067700.

[9] W. H. Press, S. A. Teukolsky, W. T. Vetterling *et al.*; *Numerical recipes in Fortran 77: The art of scientific computing* (Cambridge University Press, 1997); 2nd edition; ISBN 978-0521430647.

[10] W. H. Vandevender and K. H. Haskell; The SLATEC mathematical subroutine library. *ACM SIGNUM Newsletter*; **17.3** (1982), 16. URL `http://spectrolab.aeronomie.be/no2.htm`.

[11] K. W. Fong, T. H. Jefferson, T. Suyehiro *et al.*; Guide to the SLATEC Common Mathematical Library (1993). URL `https://netlib.org/slatec/guide`.

[12] C. D. Rodgers; *Inverse Methods for Atmospheric Sounding: Theory and Practice* (World Scientific Publishing, 2000); ISBN 9789810227401; 10.1142/3171.

[13] Q. Kleipool, N. Rozemeijer, M. van Hoek *et al.*; Ozone Monitoring Instrument (OMI) collection 4: establishing a 17-year-long series of detrended level-1b data. *Atmospheric Measurement Techniques*; **15** (2022) (11), 3527; 10.5194/amt-15-3527-2022. URL `https://amt.copernicus.org/articles/15/3527/2022/`.

[14] U. Platt; Differential Optical Absorption Spectroscopy (DOAS). *Air monitoring by spectroscopic techniques, Chem. Anal.*; **127** (1994), 27. Edited by M.W. Sigrist.

[15] U. Platt and Z. Stutz; *Differential Optical Absorption Spectroscopy, Principles and Applications* (Springer, Heidelberg, Germany, 2008); 10.1007/978-3-540-75776-4.

[16] K. F. Boersma, H. J. Eskes, R. J. Dirksen *et al.*; An improved retrieval of tropospheric $NO_2$ columns from the Ozone Monitoring Instrument. *Atmos. Meas. Tech.*; **4** (2011), 1905; 10.5194/amt-4-1905-2011.

[17] J. H. G. M. Van Geffen, K. F. Boersma, M. Van Roozendael *et al.*; Improved spectral fitting of nitrogen dioxide from OMI in the 405 – 465 nm window. *Atmos. Meas. Tech.*; **8** (2015), 1685.

[18] J. H. G. M. Van Geffen, K. F. Boersma, H. Eskes *et al.*; S5P/TROPOMI $NO_2$ slant column retrieval: method, stability, uncertainties, and comparisons with OMI. *Atmos. Meas. Tech.*; **13** (2020), 1315; 10.5194/amt-13-1315-2020.

[19] NO2 PGE Detailed Processing Model.
**source:** Space Sytems Finland; **ref:** TN-NO2-0200-SSF-001; **issue:** 1.2; **date:** 2010-04-21.

[20] J. H. G. M. Van Geffen, H. Eskes, S. Compernolle *et al.*; Sentinel-5P TROPOMI $NO_2$ retrieval: impact of version v2.2 improvements and comparisons with OMI and ground-based data. *Atmos. Meas. Tech.*; **15** (2022), 2037; 10.5194/amt-15-2037-2022.

[21] J. H. G. M. Van Geffen, H. Eskes, M. Sneep *et al.*; Improved $NO_2$ spectral fits for TROPOMI and OMI by removing wavelengths around 430 nm. *Atmos. Meas. Tech.*; (2026), in review; 10.5194/egusphere-2025-5836.

[22] Algorithm Theoretical Basis Document for the OMAERO Data Processor Collection 4.
**source:** KNMI; **ref:** AURA-OMI-KNMI-L2-0100-SD.

[23] M. Zara, K. F. Boersma, E. De Smedt *et al.*; Improved slant column density retrieval of nitrogen dioxide and formaldehyde for OMI and GOME-2A from QA4ECV: intercomparison, uncertainty characterization, and trends. *Atmos. Meas. Tech.*; **11** (2018), 4033; 10.5194/amt-11-4033-201.

[24] EOS-Aura/OMI NO2 slant column retrieval: stability & uncertainties. URL `https://www.temis.nl/tropomi/no2scd/omi_index.php`.

[25] A. C. Vandaele, C. Hermans, P. C. Simon *et al.*; Measurements of the $NO_2$ absorption cross-section from 42000 $cm^{-1}$ to 10000 $cm^{-1}$ (238-1000 nm) at 220 K and 294 K. *J. Quant. Spectrosc. & Radiat. Transfer*; **59** (1998), 171; 10.1016/S0022-4073(97)00168-4. URL `http://spectrolab.aeronomie.be/no2.htm`.

[26] Vandaele et al. NO2 cross sections. URL `http://spectrolab.aeronomie.be/no2.htm`.

[27] V. Gorshelev, A. Serdyuchenko, M. Weber *et al.*; High spectral resolution ozone absorption cross-sections: Part I. Measurements, data analysis and comparison around 293K. *Atmos. Meas. Tech.*; **7** (2014), 609; 10.5194/amt-7-609-2014.

[28] A. Serdyuchenko, V. Gorshelev, M. Weber *et al.*; High spectral resolution ozone absorption cross-sections: Part II. Temperature dependence. *Atmos. Meas. Tech.*; **7** (2014), 625; 10.5194/amt-7-625-2014.

[29] S5P/TROPOMI Static input for Level 2 processors.
**source:** KNMI/SRON/BIRA/DLR; **ref:** S5P-KNMI-L2CO-0004-SD; **issue:** 4.0.0; **date:** 2016-03-21.

[30] R. Thalman and R. Volkamer; Temperature dependant absorption cross-sections of O2-O2 collision pairs between 340 and 630 nm at atmospherically relevant pressure. *Phys. Chem. Chem. Phys.*; **15** (2013), 15371; 10.1039/C3CP50968K.

[31] R. M. Pope and E. S. Fry; Absorption spectrum (380-700 nm) of pure water. II. Integrating cavity measurements. *Appl. Opt.*; **36** (1997) (33), 8710; 10.1364/AO.36.008710.

[32] K. V. Chance and R. J. D. Spurr; Ring effect studies: Rayleigh scattering, including molecular parameters for rotational Raman scattering and the Fraunhofer spectrum. *Appl. Opt.*; **36** (1997) (21), 5224.

[33] K. V. Chance and R. L. Kurucz; An improved high-resolution solar reference spectrum for earth's atmosphere measurements in the ultraviolet, visible, and near infrared. *J. Quant. Spectrosc. Radiat. Transfer*; **111** (2010), 1289.

[34] M. Dobber, R. Voors, R. Dirksen *et al.*; The High-Resolution Solar Reference Spectrum between 250 and 550 nm and its Application to Measurements with the Ozone Monitoring Instrument. *Solar Physics*; **249** (2008), 281–291; 10.1007/s11207-008-9187-7.

[35] A. Richter, M. Begoin, A. Hilboll *et al.*; An improved $NO_2$ retrieval for the GOME-2 satellite instrument. *Atmos. Meas. Tech.*; **4** (2011), 1147; 10.5194/amt-4-1147-2011.

[36] J. Lampel, U. Frieß and Platt U.; The impact of vibrational Raman scattering of air on DOAS measurements of atmospheric trace gases. *Atmos. Meas. Tech.*; **8** (2015), 3767; 10.5194/amt-8-3767-2015.

[37] QA4ECV D4.2 - Recommendations on best practices for retrievals for Land and Atmosphere ECVs..
**source:** KNMI; **ref:** EU-project 607405, SPA.2013.1.1-03; **date:** April 2016.

[38] Q. Kleipool, A. Ludewig, Babić L. *et al.*; Pre-launch calibration results of the TROPOMI payload on-board the Sentinel-5 Precursor satellite. *Atmos. Meas. Tech.*; **11** (2018), 6439; 10.5194/amt-11-6439-2018.

[39] A. Ludewig, Q. Kleipool, R. Bartstra *et al.*; In-flight calibration results of the TROPOMI payload on-board the Sentinel-5 Precursor satellite. *Atmos. Meas. Techn.*; **13** (2020), 3561; 10.5194/amt-13-3561-2020.

[40] V. M. E. Schenkeveld, G. Jaross, S. Marchenko *et al.*; In-flight performance of the Ozone Monitoring Instrument. *Atmos. Meas. Tech.*; **10** (2017), 1957; 10.5194/amt-10-1957-2017.

[41] P. Wang, P. Stammes, R. Van der A *et al.*; FRESCO+: an improved $O_2$ A-band cloud retrieval algorithm for tropospheric trace gas retrievals. *Atmos. Chem. Phys.*; **8** (2008), 6565; 10.5194/acp-8-6565-2008.

[42] M. Desmons, P. Wang, P. Stammes *et al.*; FRESCO-B: a fast cloud retrieval algorithm using oxygen B-band measurements from GOME-2. *Atmos. Meas. Tech.*; **12** (2019), 2485; 10.5194/amt-12-2485-2019.

[43] J. P. Veefkind, J. F. de Haan, M. Sneep *et al.*; Improvements to the OMI $O_2$-$O_2$ operational cloud algorithm and comparisons with ground-based radar-lidar observations. *Atmospheric Measurement Techniques*; **9** (2016) (12), 6035; 10.5194/amt-9-6035-2016. URL `https://amt.copernicus.org/articles/9/6035/2016/`.

[44] Algorithm Theoretical Basis Document for the OMCLDO2 Data Processor Collection 4.
**source:** KNMI; **ref:** AURA-OMI-KNMI-L2-0200-SD.

[45] Q. L. Kleipool, M. R. Dobber, J. F. De Haan *et al.*; Earth surface reflectance climatology from 3 years of OMI data. *J. Geophys. Res.*; **113** (2008) (D18308), 22 pp.; 10.1029/2008JD010290.

[46] Q. L. Kleipool, M. R. Dobber, J. F. De Haan *et al.*; OMI/Aura Surface Reflectance Climatology L3 Global Gridded 0.5 degree x 0.5 degree V3 (2010); 10.5067/Aura/OMI/DATA3006. URL `https://disc.gsfc.nasa.gov/datasets/OMLER_003/summary`.

[47] L. G. Tilstra, M. De Graaf, V. J. H. Trees *et al.*; TROPOMI surface reflectivity climatology [Data set] (2023); 10.21944/tropomi-surface-albedo.

[48] L. G. Tilstra, M. de Graaf, V. J. H. Trees *et al.*; A directional surface reflectance climatology determined from TROPOMI observations. *Atmos. Meas. Tech.*; **17** (2024), 2235; 10.5194/amt-17-2235-2024.

[49] Primary Ancillary Data Geo-Colocated to OMI/Aura VIS 1-Orbit L2 Swath 13x24km V4 (OMVANC). 10.5067/AURA/OMI/DATA2439.

[50] P. De Rosnay, L. Isaksen and M. Dahoui; Snow data assimilation at ECMWF. *ECMWF Newsletter No. 143, Spring 2015*; (2015), 26.

[51] M. J. Cooper, R. V. Martin, A. I. Lyapustin *et al.*; Assessing snow extent data sets over North America to inform and improve trace gas retrievals from solar backscatter. *Atmos. Meas. Tech.*; **11** (2018), 2983; 10.5194/amt-11-2983-2018.

[52] M. J. Brodzik and J. S. Stewart; Near Real-Time SSM/I EASE–Grid Daily Global Ice Concentration and Snow Extent. Boulder, Colorado USA. NASA National Snow and Ice Data Center (2016); 10.5067/3KB2JPLFPK3R. Updated daily; URL `http://nsidc.org/data/NISE`.

[53] Product User Manual for the OMNO2A Data Processor Collection 4.
**source:** KNMI; **ref:** AURA-OMI-KNMI-L2-0308-MA.

[54] GEOS-5 FP-IT 2D Time-Averaged Single-Level Diagnostics Geo-Colocated to OMI/Aura VIS 1-Orbit L2 Swath 13x24km. 10.5067/AURA/OMI/DATA2437.

[55] Y. Zhou, D. Brunner, K. F. Boersma *et al.*; An improved tropospheric $NO_2$ retrieval for OMI observations in the vicinity of mountainous terrain. *Atmos. Meas. Tech.*; **2** (2009), 401; 10.5194/amt-2-401-2009.

[56] J. D. Maasakkers, K. F. Boersma, J. E. Williams *et al.*; Vital improvements to the retrieval of tropospheric $NO_2$ columns from the Ozone Monitoring Instrument. *Geophys. Res. Abstracts*; **15** (2013) (EGU2013-714), 1. EGU General Assembly 2013.

[57] TROPOMI ATBD of the total and tropospheric NO2 data products.
**source:** KNMI; **ref:** S5P-KNMI-L2-0005-RP; **issue:** 2.4.0; **date:** 2022-07-11.

[58] What are outliers in the data? URL `https://www.itl.nist.gov/div898/handbook/prc/section1/prc16.htm`.

[59] Important z-scores. URL http://www.cs.uni.edu/~campbell/stat/normfact.html.

[60] The Standard Normal Distribution. URL http://sphweb.bumc.bu.edu/otlt/MPH-Modules/BS/BS704_Probability/BS704_Probability9.html.

[61] Sentinel-5 L2 Prototype Processor – Algorithm Theoretical Baseline Document for Cloud data product. **source:** KNMI; **ref:** KNMI-ESA-S5L2PP-ATBD-005; **issue:** 3.1; **date:** 2019-05-02.

[62] Determine the effective cloud fraction for a specific wavelength. **source:** KNMI; **ref:** S5P-KNMI-L2-0115-TN; **issue:** 2.0.0; **date:** 2019-04-10.

[63] S. Chandrasekhar; *Radiative Transfer* (Dover Publications, New York, 1960); ISBN 0-486-60590-6.

# 3 Terms, definitions and abbreviated terms

Terms, definitions and abbreviated terms can be found in [1]. Terms specific to this document can be found below.

## 3.1 Terms and definitions

There are no terms and definitions specific to this document.

## 3.2 Acronyms and abbreviations

| | |
|---|---|
| AAI | Absorbing Aerosol Index |
| AMF | Air-mass factor |
| ATBD | Algorithm Theoretical Basis Document |
| CTM | Chemistry Transport Model |
| DEM | Digital Elevation Map |
| DOAS | Differential Optical Absorption Spectroscopy |
| DOMINO | Dutch OMI $NO_2$ data products of KNMI for OMI |
| ECMWF | European Centre for Medium-Range Weather Forecast |
| GOME | Global Ozone Monitoring Experiment |
| ISRF | Instrument Spectral Response Function (*aka* slit funtion) |
| KNMI | Royal Netherlands Meteorological Institute |
| LER | Lambertian equivalent reflectivity |
| LUT | Look-up table |
| NISE | Near-real-time Ice and Snow Extent |
| $NO_2$ | Nitrogen dioxide |
| NRT | near-real time (i.e. processing within 3 hours of measurement) |
| NSIDC | National Snow & Ice Data Center |
| $O_2$–$O_2$ | Oxygen collision induced absorption (oxygen-oxygen-dimer) |
| OE | Optimal Estimation |
| OMI | Ozone Monitoring Instrument |
| OMAERO | OMI L2 Aerosol Data Processor or Data Product |
| OMCLDO2 | OMI L2 Cloud Data Processor or Data Product |
| OMNO2A | OMI L2 NO2 slant column data processor or data product |
| OMNO2B | OMI L2 NO2 vertical column data processor or data product |
| OSIPS | Processing system at NASA |
| PDGS | Sentinel-5Precursor Payload Data Ground Segment (at DLR) |
| PUM | Product User Manual |
| QA4ECV | European "Quality Assurance for Essential Climate Variables" project |
| SLATEC | Sandia, Los Alamos, Air Force Weapons Laboratory Technical Exchange Committee |
| SCD | Slant column density |
| SCIAMACHY | Scanning Imaging Absorption Spectrometer for Atmospheric Cartography |
| SNR | Signal-to-Noise Ratio |
| TM5-MP | Data assimilation / chemistry transport model (version 5) |
| TOA | Top-of-atmosphere |
| TROPOMI | Tropospheric Monitoring Instrument |
| VCD | Vertical column density |

# 4 OMI Instrument Overview

Figure 1 shows the functional overview of the the overall architecture of the OMI instrument.

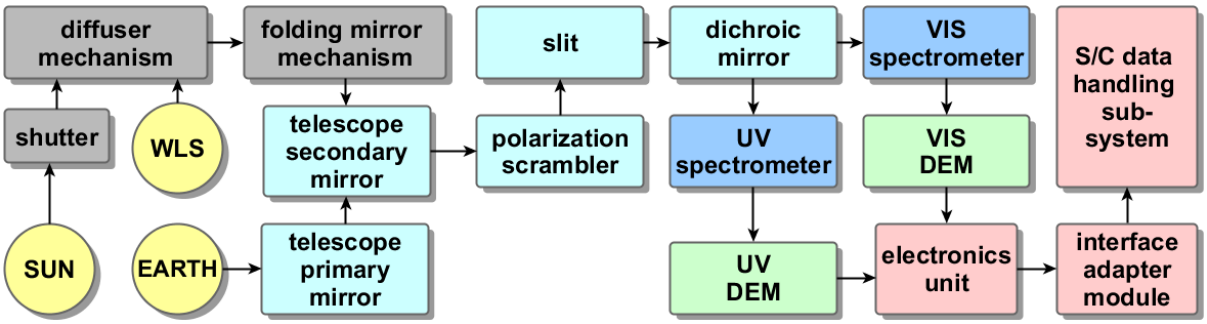


**Figure 1**: The overall architecture of OMI

The light reflected from the Earth's atmosphere and surface is imaged by two telescope mirrors onto the entrance slit. Before the slit the incoming light is depolarized by a scrambler. After the slit the light is split by a dichroic mirror into two wavelength ranges 264–383 nm (bands 1 and 2, UV) and 349–504 nm (band 3, VIS) and guided to the two spectrometers. The wavelength range, spectral resolution and sampling are listed per band in Table 1.

| Band | Total range | Average spectral resolution (FWHM) | Average spectral sampling distance |
|---|---|---|---|
| Band 1 (UV-1) | 264–311 nm | 0.63 nm | 0.33 nm / pixel |
| Band 2 (UV-2) | 307–383 nm | 0.42 nm | 0.14 nm / pixel |
| Band 3 (VIS) | 349–504 nm | 0.63 nm | 0.21 nm / pixel |

**Table 1**: Spectral range, resolution and sampling distances of the OMI instrument. Note that there is a small spectral overlap between bands 1 and 2. Bands 1 and 2 are imaged separately onto detector 1 to allow for different spatial sampling and higher signal to noise in band 1. The numbers are as reported in [2].

In the spectrometers the light is dispersed and imaged onto CCD detectors. The CCD signals are digitized and co-added in the electronics unit (ELU). Further data formatting is performed in the interface adapter module (IAM) before the data is transmitted to the S/C data handling sub-system. In addition to the Earth's radiance, light from the Sun can be measured by opening a shutter and moving the folding mirror mechanism in between the two telescope mirrors. The solar light is reflected by either one of three diffusers. A white light source (WLS) can also be coupled into the telescope via the folding mirror.

The different parts of the instrument are described in more detail in the OMI L01b ATBD [3].

# 5 Description of the $NO_2$ algorithm

## 5.1 Introduction

The KNMI OMI L2 $NO_2$ Data Processor Collection 4 essentially makes use of the same algorithm used in the TROPOMI $NO_2$ operational processor, which is described in [4].

The KNMI OMI L2 $NO_2$ processing consists of two parts, depicted in Fig. 2:

1. The computation of the $NO_2$ slant column density (SCD)
   performed on every input Level-1b (L1B) spectrum file
   – this step is done by NASA at OSIPS and called OMNO2A
2. The computation of the $NO_2$ vertical column density (VCD)
   performed on every Level-2 (L2) SCD file
   – this step is done by KNMI at ECMWF and called OMNO2B / TM5-MP-DOMINO

where "TM5-MP-DOMINO" is the name of the data assimilation / chemistry transport model used for the conversion of slant to vertical column data.

This ATBD describes the OMNO2A step; the OMNO2B / TM5-MP-DOMINO step will be described in a separate ATBD.

The OMNO2A output product only contains the slant column densities and related information, such as the wavelength calibration parameters, as well as additional information, notably geolocation and viewing data and external imput data (such as surface and cloud data). Data variables for the OMNO2B / TM5-MP-DOMINO step are present in the output product but these are left empty (i.e. contain a fill value).

Note that for Collection 3 there was both a near-real time (NRT) mode and an offline mode. Since OMI $NO_2$ NRT data is currently of little interest, the NRT data stream is no longer supported in Collection 4 and therefore not described here. A user interested in $NO_2$ NRT data is referred to TROPOMI $NO_2$ NRT data.

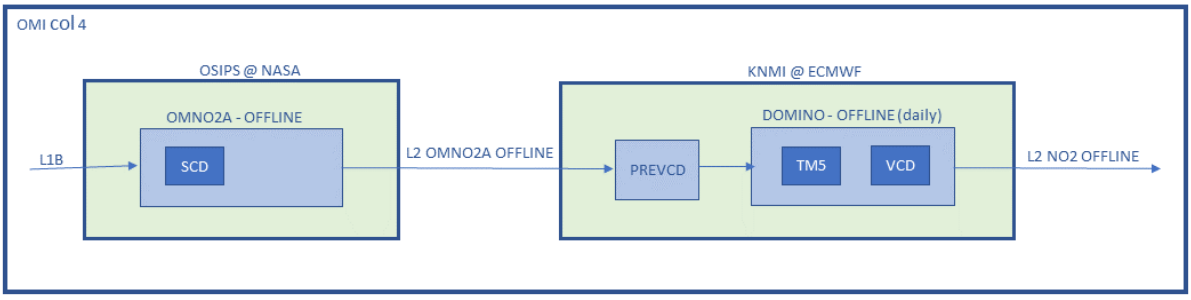


**Figure 2**: OMI $NO_2$ collection 4 data flow: the OMNO2A slant column processing step (left block) and OMNO2B / TM5-MP-DOMINO vertical column processing step (right).

### 5.1.1 Orbit data filenames

The name of an OMNO2A slant column orbit file looks, for example, like this:

`OMI-Aura_L2A-OMNO2_2005m1003t093252-o006482_v004-2022m1216t204201.nc`

with the following elements, separated by underscores:

1. `OMI-Aura` = mission name
2. `L2A-OMNO2` = data product name [1]

[1] Older OMNO2A files may also have the string "`L2-OMNO2A`" as data product name.

3. `2005m1003t093252-o006482` = start date & time of the orbit and orbit number
4. `v004-2022m1216t204201` = collection number and processing date & time [2]

Previously, the name of a vertical column orbit file often containted the string "`L2-OMNO2`" as data product name, which could refer to a VCD file made by KNMI with TM5-MP-DOMINO or by NASA with their own processor, while also the string "`L2-OMNO2B`" has been seen around. What will be the final name of the OMNO2B / TM5-MP-DOMINO product is to be decided.

### 5.1.2 Structure of the chapter

This chapter is has the following structure:

- Section 5.2 describes the $NO_2$ slant column fit, i.e. the OMNO2A part.
- Section 5.3 describes the external data inserted by the OMNO2A processor.
- Section 5.4 describes necessary and preferred updates needed to the current (that is: January 2026) OMNO2A processor and output file
- Section 5.5 describes the contents of the OMI $NO_2$ data product.

### 5.1.3 Some general remarks

Different versions of DOMINO for the $NO_2$ VCD calculation have been and are in use. For the OMNO2B processing of OMI collection 3 data DOMINO-v2 was used. For the QA4ECV-v1.1 ([5], [6], [7]) processing of OMI collection 3 data DOMINO-v3 was used. For the current version of TROPOMI (v2.8.0/v2.9.1) we use TM5-MP-DOMINO v3.6.5, update 20-10-2024. A reprocessing of both TROPOMI and OMI is currently planned for late 2026/early 2027, and will be based on the same version of TM5-MP-DOMINO.

There are some differences between the OMNO2A Collection 3 algorithm and the Collection 4 algorithm: the Collection 3 algorithm uses a modified Levenberg-Marquardt method [8], [9, Ch. 15] as adapted from the SLATEC mathematical library [10], [11] for the non-linear fitting, whereas the Collection 4 algorithm uses the Optimal Estimation (OE) method [12].

[2] The processing date & time are of the OMNO2A step and are *not* updated in subsequent steps.

## 5.2 The $NO_2$ slant column fit

This section describes the OMNO2A step: the retrieval of the total $NO_2$ slant column density (SCD; $N_s$) from the collection 4 Level-1b radiance and irradiance spectra measured by OMI (see [13]).

### 5.2.1 Spectral fitting

The baseline method to determine $NO_2$ total slant columns is DOAS (see [14], [15]). The DOAS fitting function for OMI follows the current non-linear fitting approach implemented for TROPOMI (see [16], [17], [18], [19]).

The reflectance spectrum observed by the satellite instrument, $R_{\text{meas}}(\lambda)$, is the ratio of the radiance at the top of the atmosphere, $I(\lambda)$, and the extraterrestrial solar irradiance, $E_0(\lambda)$, leaving out the viewing geometry arguments brevity:

$$R_{\text{meas}}(\lambda) = \frac{\pi I(\lambda)}{\mu_0 E_0(\lambda)} \tag{1}$$

where $E_0$ and $I$ are recorded at the same detector row and given on the same wavelength grid (see Sect. 5.2.3), and $\mu_0 = \cos(\theta_0)$ is the cosine of the solar zenith angle. The selection of $E_0$ in the OMI $NO_2$ retrieval is discussed in Sect. 5.2.2.

In space-borne DOAS, $R_{\text{meas}}$ is related to the extinction of light by scattering and absorbing species along the average photon path between sun and satellite instrument. The effective, integrated absorption of $NO_2$ along the average photon path is represented by the total $NO_2$ slant column density ($N_s$). The DOAS spectral fitting is performed for all satellite ground pixels with $\theta_0 < 88°$, so that there is no potential danger from the division by $\mu_0$ in Eq. (1).

The DOAS spectral fitting attempts to find the optimal modelled reflectance spectrum, $R_{\text{mod}}(\lambda)$, by minimising the chi-squared merit function, i.e. the smallest possible differences between the observed and modelled reflectance spectrum:

$$\chi^2 = \sum_{i=1}^{n_\lambda} \left( \frac{R_{\text{meas}}(\lambda_i) - R_{\text{mod}}(\lambda_i)}{\Delta R_{\text{meas}}(\lambda_i)} \right)^2 \tag{2}$$

with $n_\lambda$ the number of wavelengths in the fit window and $\Delta R_{\text{meas}}(\lambda_i)$ the noise on the reflectance, which depends on the radiance and irradiance noise given in the Level-1b product:

$$\Delta R_{\text{meas}}(\lambda_i) = \frac{1}{E_0(\lambda_i)} \sqrt{\left(\Delta I(\lambda_i)\right)^2 + \left(\Delta E_0(\lambda_i)\right)^2 \cdot \left(R_{\text{meas}}(\lambda_i)\right)^2} \tag{3}$$

i.e. on the signal-to-noise ratio (SNR) of the measurements. For numerical reasons a maximum of $2500$ is set to the SNR on the reflectance and $\Delta R_{\text{meas}}$ is adjusted upwards when this SNR limit is exceeded. In the $NO_2$ wavelength range the SNR is typically $500$, but it may be larger over very bright scenes (notably clouds), potentially leading to saturation effects in the spectra, hence limiting the SNR does not pose a significant limit on the lowest SCD error estimates (the lower $\Delta R_{\text{meas}}$ the lower the SCD error is). The magnitude of $\chi^2$ is a measure for how good the fit is. Another measure for the goodness of the fit is the so-called root-mean-square (RMS) error, which is defined as follows:

$$R_{\text{RMS}} = \sqrt{\frac{1}{n_\lambda} \sum_{i=1}^{n_\lambda} \left( R_{\text{meas}}(\lambda_i) - R_{\text{mod}}(\lambda_i) \right)^2} \tag{4}$$

where the difference $R_{\text{meas}}(\lambda) - R_{\text{mod}}(\lambda)$ is usually referred to as the residual of the fit.

Radiance spectral pixels flagged in the Level-1b data as bad or as suffering from saturation or flagged by the outlier removal (see below) are filtered out before doing any further processing step. In the current (January 2026) OMNO2A processor it is not yet possible to manually define one or more wavelength bands that need to be excluded from the DOAS fit, but following the developments in the TROPOMI slant column processor [4] that may be implemented at a later stage. Excluding spectral pixels is done in the processor by setting the error on the reflectance of the spectral pixels that are flagged or to be excluded to $10^4$ times the measurement: $\Delta R_{\text{meas}}(\lambda_i) = 10^4 \times R_{\text{meas}}(\lambda_i)$, as a result of which spectral pixel $i$ does not contribute to the DOAS minimisation of the $\chi^2$ in Eq. (2). In the computation of the RMS error in Eq. (4), these flagged or excluded pixels are skipped.

The baseline model function for OMI collection 4 follows the approach for OMI collection 3 and TROPOMI and reads as follows:

$$R_{\text{mod}}(\lambda) = P(\lambda) \cdot \exp\left[ -\sum_{k=1}^{n_k} \sigma_k(\lambda) \cdot N_{s,k} \right] \cdot \left( 1 + C_{\text{ring}} \frac{I_{\text{ring}}(\lambda)}{E_0(\lambda)} \right) \tag{5}$$

**Table 2**: Main settings of the operational DOAS retrieval of $NO_2$ for OMI and TROPOMI in the operational processing.

| | OMI<br>collection 4 | TROPOMI<br>as of v2.9.1 | OMI<br>QA4ECV v1.1 | OMI<br>collection 3 |
|---|---|---|---|---|
| wavelength range [nm] | $405-465$ | $405-465$ | $405-465$ | $405-465$ |
| omitted range [nm] | — | $428-433$ | — | — |
| secondary trace gases | $O_3$, $H_2O_{vap}$, $O_2$–$O_2$, $H_2O_{liq}$ | $O_3$, $H_2O_{vap}$, $O_2$–$O_2$, $H_2O_{liq}$ | $O_3$, $H_2O_{vap}$, $O_2$–$O_2$, $H_2O_{liq}$ | $O_3$, $H_2O_{vap}$, — |
| pseudo-absorbers | Ring | Ring | Ring | Ring |
| intensity offset correction | no | no | yes | no |
| fitting method | non-linear | non-linear | non-linear | non-linear |
| degree of polynomial | 5 | 5 | 5 | 5 |
| irradiance spectrum | 2005-average | daily | 2005-average | 2005-average |
| slant column processing | NASA / KNMI | PDGS (@ DLR) | NASA / KNMI | NASA / KNMI |
| references | — | [4], [18], [20], [21] | [5], [6], [7] | [16], [17] |

with $\sigma_k(\lambda)$ the absolute cross section and $N_{s,k}$ the slant column amount of molecule $k = 1,\ldots,n_k$ taken into account in the fit ($NO_2$, $O_3$, etc.), $C_{\text{ring}}$ the Ring fitting coefficient and $I_{\text{ring}}(\lambda)$ the synthetic Ring spectrum (generated from an $E_{\text{ref}}(\lambda)$ reference irradiance) and $E_0(\lambda)$ the measured irradiance (cf. Sect. 5.2.2). The Ring spectrum describes the differential spectral signature arising from inelastic Raman scattering of incoming sunlight by $N_2$ and $O_2$ molecules. The last term in Eq. (5) describes both the contribution of elastic scattering to the differential absorption signatures (i.e. the $1$) and the modification of these differential structures by inelastic scattering (the $+C_{\text{ring}} \cdot I_{\text{ring}}(\lambda)/E_0(\lambda)$ term) to the reflectance spectrum. The sources of the reference spectra used are discussed in Sect. 5.2.6.

In the modelled spectrum of Eq. (5) a polynomial of order $n_p$ with coefficients $a_m$:

$$P(\lambda) = \sum_{m=0}^{n_p} a_m \lambda^m \tag{6}$$

is introduced to account for spectrally smooth structures resulting from molecular (single and multiple) scattering and absorption, aerosol scattering and absorption, and surface albedo effects. Because of the polynomial term, only the highly structured differential absorption features contribute to the fit of the slant column densities. In order to prevent the numerical value of the polynomial components in Eq. (6) to become very large or very small (for the $405-465$ nm fit window, for example, usually $n_p = 5$), the wavelengths are scaled to the range $[-1:+1]$ over the fit window in the processor.

In order to remove strong outliers in the DOAS fit residual (caused by, e.g., high-energy particles hitting the CCD detector, variations in the dark current, or bad pixels not correctly flagged in the Level-1b data), a "spike removal" algorithm was implemented in the DOAS processor to detect outliers in the fit residual; after removal of the spectral pixels of such outliers from the measured reflectance, the $NO_2$ DOAS fit is redone to provide the final fit parameters. See Appendix A and [20] for a description and examples of the spike removal.

Table 2 provides an overview of the operational DOAS fit settings used in the OMI collection 4 processor and those used for some current and past UV/Vis backscatter satellite instruments: the fit window, the reference spectra used in the fit (see Sect. 5.2.6) and the degree of the DOAS polynomial. Note that for the processing of GOME(-1) data it was necessary to include a correction for the undersampling of the spectra, i.e. the fact that the spectral sampling is of the same order as the FWHM of the instrument slit function. For the instruments listed in Table 2 this correction is not necessary: their spectral resolution, i.e. the FWHM of the slit function, is 2–3 times as large as their spectral sampling.

### 5.2.2 Selection of the irradiance spectrum

In many trace gas retrievals, including the TROPOMI $NO_2$ retrieval, the irradiance $E_0$ used in the reflectance, defined in Eq. (1), is the irradiance measured once a day by the satellite instrument, selecting the $E_0$ measured closest in time to an orbit of $I$ measurements. OMI also measures $E_0$ once a day, when OMI crosses the

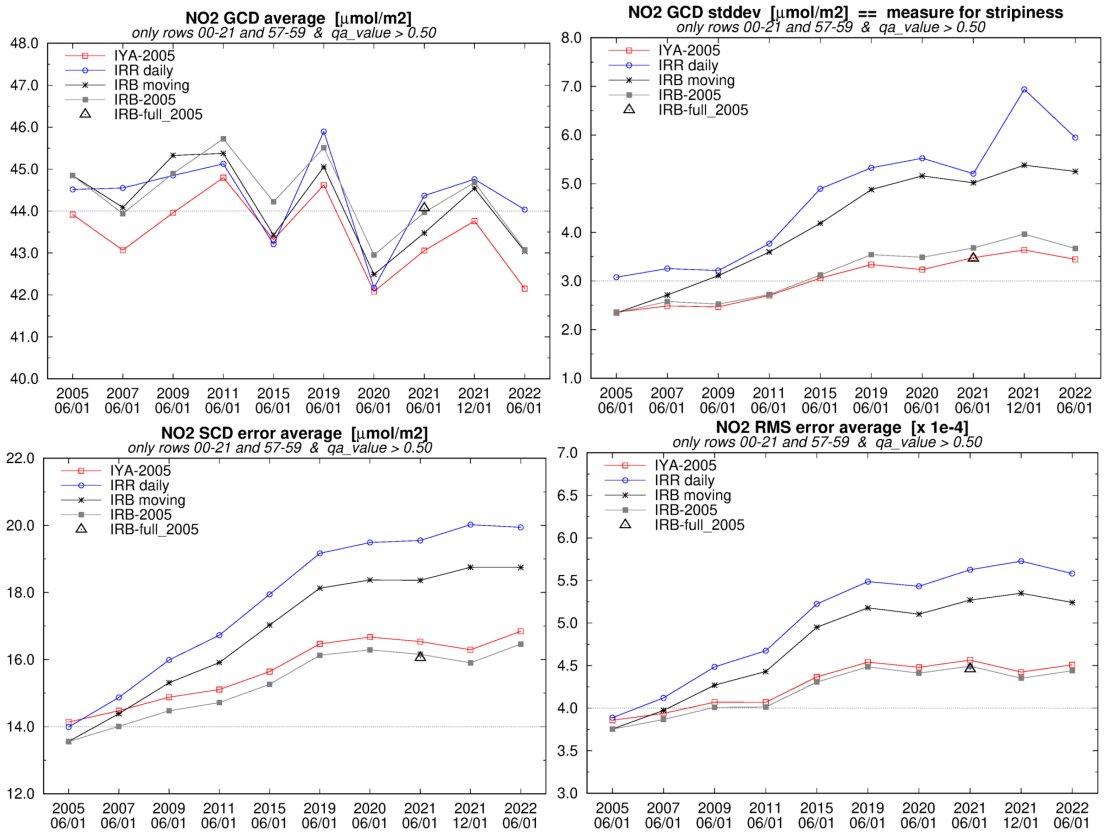


**Figure 3**: Comparison of the main $NO_2$ slant column retrieval results using different irradiances $E_0$ in the reflectance of Eq. (1), as described in the text: the geometric column density (GCD) average (top-left) and standard deviation (top-right), the SCD error (bottom-left) and the RMS error (bottom-right).

terminator along a given orbit. But already at an early stage in the OMI measurement it was clear that this daily irradiance is of insufficient quality, in that it has a low SNR. Instead it was decided, e.g. in the collection 3 processing, to use an irradiance spectrum constructed from a yearly average of daily solar irradiance measurements by OMI during 2005, which has a much lower "irradiance noise" level, $\Delta E_0$, than a individual daily irradiance measurements.

To test this anew for the collection 4 reprocessing, the orbits of 10 days were used (1 June of 2005, 2007, 2009, 2011, 2015, 2019, 2020, 2021 and 2022, as well as 1 Dec. 2021), looking at along-track averages of the geometric column density (GCD):

$$N_{\mathrm{v}}^{\mathrm{geo}} = \frac{N_{\mathrm{s}}}{M^{\mathrm{geo}}} \ , \qquad M^{\mathrm{geo}} = \frac{1}{\cos\theta_0} + \frac{1}{\cos\theta} \tag{7}$$

where $M^{\mathrm{geo}}$ is the geometric air-mass factor (AMF), which accounts for the viewing geometry, and at along-track averages of the SCD error and the RMS error the DOAS fit; the along-track averages where taken over scanlines with latitude range $[-50°:+50°]$ for the nadir pixels. To avoid problems related to the row anomaly (Sect. 5.3.1), only rows $00-21$ & $57-59$ are considered, selecting ground pixels with valid SCD retrieval, i.e. with $\mathtt{qa\_value} > 0.50$ (cf. Sect. 5.5.2) after the SCD retrieval step. Selection of the best irradiance was based on the lowest across-track "stripiness" of the fit result (given by the standard deviation of the across-track average GCD) and lowest across-track average SCD and RMS errors, averaging over all 14 or 15 orbits of each test day.

Three irradiance spectra were investigated for each of the 10 days: the daily measured $E_0$ ("IRR"), a moving average over 100 days prior to the test day radiance mearuments ("IRB") and the 100-day average over the 100days prior to 1 June 2005 used for all test days ("IRB-2005"). For comparison the old 2005 average ("IYA-2005") is also used as a kind of reference, though using this collection 3 irradiance for the collection 4 reprocessing would not be consistent.

Fig. 3 shows the results of this comparison:

- The top-left panel shows that the average GCD – the true value of which is unknown – varies within

about $5\%$ (likely caused by small differences in irradiance and irradiance noise), hence the GCD itself does not lead to a preference of the irradiance choice.

- The top-right panel shows the standard deviation of the average GCD, which can be seen as a measure for the "stripiness" of the GCD data (lacking the across-track stripe amplitude, which will be determined in the processing the OMNO2B / DOMINIO step): the lower this stripiness the more reliable the GCD value is.
- The bottom-left panel shows the average SCD error, i.e. the average estimated DOAS uncertainty of the SCD values.
- The bottom-right panel shows the average RMS error of the fit.

In all cases the average "IRB" (star) is doing better than the daily "IRR" (open cirlce) as expected, while the fixed 100-day average "IRB-2005" (filled squares) is doing best in all cases. Because of this result, it was decided to select a new yearly average over all 2005 irradiances ("IRB-full_2005"), which was tested on the 1 June 2021 data – see the open triangle in Fig. 3: it gives almost the same results at the "IRB-2005" irradiance (filled squares).

The cloud cover data (cf. Sect. 5.3.2) – in particular the cloud fraction, cloud radiance fraction and cloud pressure (not shown here) – do not show differences for the different irradiances (though the cloud pressure appears to be quite sensitive to retrieval details, with per-orbit averages vary $50-60\,$hPa).

For the absorbing aerosol index (AAI), however, the "IRB-full_2005" irradiance is *not* a good idea as it leads to a decrease over time of the average AAI, while the moving average "IRB" irradiance leads to a more or less constant AAI over time, so that for the collection 4 AAI retrieval "IRB" is used [22].

### 5.2.3 Wavelength calibration & common wavelength grid

Both the irradiance and radiance spectra are wavelength calibrated prior to the DOAS fit, using the same wavelength calibration approach, in the $[\lambda_b : \lambda_e]$ wavelength window, with $\lambda_b = 405\,$nm and $\lambda_e = 465\,$nm.

Using the subscripts 'nom' and 'cal' to denote nominal (i.e. from the Level-1b data product) and calibrated wavelengths, respectively, the calibrated irradiance and radiance to be used in Eq. (1) are then given by:

$$\begin{aligned} E_0(\lambda^{E0}_{\mathsf{cal}}) &= E_0(\lambda^{E0}_{\mathsf{nom}} + w^{E0}_s) \\ I(\lambda_{\mathsf{cal}}) &= I(\lambda_{\mathsf{nom}} + w_s + w_q(\lambda_{\mathsf{nom}} - \lambda_0)) \end{aligned} \tag{8}$$

where $w_s$ represents a wavelength shift and $w_q$ a wavelength stretch ($w_q > 0$) or squeeze ($w_q < 0$), with $w_q$ defined w.r.t. the central wavelength of the fit window $\lambda_0$; for the time being the $w_q$ fit parameter will remain turned off. Since in view of numerical stability, the wavelengths are scaled to the range $[-1 : +1]$ over the fit window, computationally $\lambda_0 = 0$. Each wavelength calibration of Eq. (8) comes with its own $\chi^2_w$ as a goodness-of-fit. Note that for the irradiance calibration we only consider a shift.

In order to avoid possible extrapolations, both these steps are performed on a wavelength range that is $1\,$ nm wider than the fit window, i.e. the measured reflectance is formed on the common wavelength grid and then cut to the fit window. where the selection is based on nearest neighbours, providing a more consistent pixel index selection along-track.

The wavelength calibration of Eq. (8) is performed on the irradiance at the start of the processing of a given granule, and per radiance spectrum prior to forming the measured reflectance of Eq. (1). In order to form this reflectance, both (calibrated) spectra $I(\lambda)$ and $E_0(\lambda)$ need to be given on the same wavelength grid. In our approach $E_0(\lambda)$ is converted to the radiance wavelength grid by way of a high-sampling interpolation, taking advantage of the fact that we have additional information from a high-resolution solar reference spectrum $E_{\mathsf{ref}}(\lambda)$. Details of the wavelength calibration and the high-sampling interpolation implemented for OMI are given in Appendices B and C, respectively.

### 5.2.4 Minimising the chi-squared merit function

The slant column densities $N_{\mathsf{s},k}$, the Ring coefficent $C_{\mathsf{ring}}$ and the polynomial coefficients $a_m$ are obtained from a minimisation of the $\chi^2$ of Eq. (2), i.e. the differences between the observed and modelled reflectances. The OMI Collection 3 algorithm uses a modified Levenberg-Marquardt method [8] as adapted from the SLATEC mathematical library [11] for the non-linear fitting, whereas the Collection 4 algorithm uses the Optimal Estimation (OE) method [12]; the latter is also used for the wavelength calibration (Appendix B).

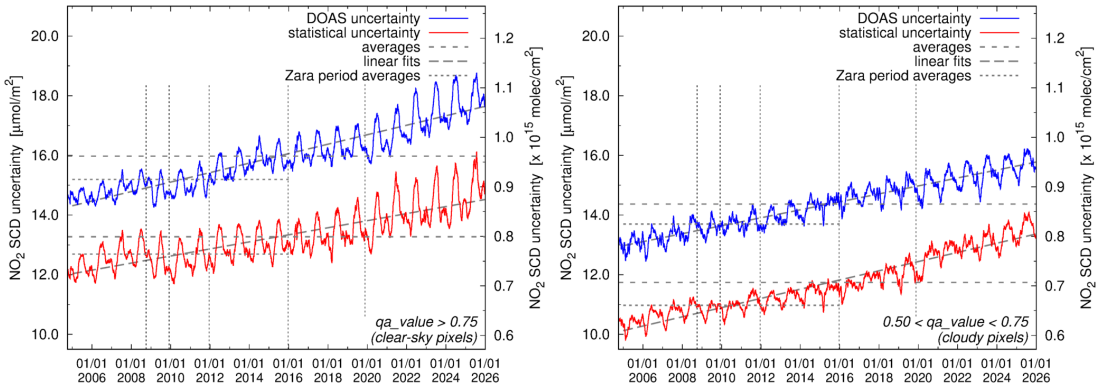


**Figure 4**: DOAS slant column uncertainty estimate (blue) and statistical uncertainty (red) 21-day running means for clear-sky (left) and cloudy (right) pixels with successful retrieval, as function of time. Dashed horizontal lines indicate averages over the full data period, thin-dashed horizontal lines indicate averages over the "Zara period" 2005-2015 [23], and thick-dashed lines indicate a linear fit over calendar years of the full data period. Vertical lines indicate five moments in time when instrument changes took place that affected the data. Figures taken from [24].

For the $\chi^2$-minimisation using the OE solver suitable a-priori values of the fit parameters were selected and the a-priori errors are set very large, so as not to limit the solution of the fit, while for numerical stability reasons a pre-whitening of the data is performed. (Whitening transforms a vector of random variables with a known covariance matrix into a set of new variables whose covariance is the identity matrix, meaning that they are uncorrelated and each have variance $1$; cf. [12], Ch. 2.)

A number of fitting diagnostics is provided by the fitting procedure. Estimated slant column and fitting coefficient uncertainties are obtained from the covariance matrix of the standard errors, which is given as a standard output of the OE procedure. The SCD error estimates are scaled with the square-root of the normalised $\chi^2$, where $\chi^2$ is normalised by $(n_\lambda - D)$, with $n_\lambda$ the number of wavelengths in the fit window and $D$ the degrees of freedom of the fit, which is almost equal to the number of fit parameters. All fitting coefficients are provided in the $NO_2$ output data file as diagnostic data.

### 5.2.5 $NO_2$ slant column uncertainties

The slant column uncertainty $\Delta N_s$ is an estimate which depends on details of the DOAS fit and can therefore not always be compared directly with uncertainties supplied by other DOAS approaches. The spatial variability of the slant columns over a remote Pacific Ocean sector, however, can be used as an independent statistical estimate of the random component of the SCD uncertainty as this is independent of the DOAS approach. Zara et al., 2018 [23] used this approach to compare OMI collection 3 and GOME-2A $NO_2$ and formaldehyde SCDs retrieved to compare the SCD uncertainties following from the different DOAS fits over the years 2005 through 2015 within the QA4ECV project [6].

To monitor the OMI $NO_2$ stability, preliminary collection 4 data has been analysed from the first available data of 1 Oct. 2004 onwards up to end of 2025. Results of this analysis, along with the monitoring of several other quantities, are presented on a set of web pages [24]; Fig. 4 shows an example of these results. The analysis on [24] will be redone once the final collection 4 data is available.

### 5.2.6 Reference spectra

The selection of the reference spectra for the trace gas cross sections in Eq. (5) is driven by whether a species shows substantial absorption in the wavelength range relevant for $NO_2$ retrieval, and exploits the best available sources prior to commissioning phase. Experience with OMI and TROPOMI has shown that $NO_2$, ozone, water vapour, and Rotational Raman Scattering (RRS), i.e. the inelastic part of the Rayleigh scattering (the so-called "Ring effect"), are most relevant in the wavelength interval relevant to $NO_2$. Also taking absorption in liquid water and by the $O_2$–$O_2$ collision complex into account improves the fit [17], hence these are included for OMI collection 4 and TROPOMI.

High-resolution laboratory measured absorption cross sections are convolved with the OMI slit function (or: instrument spectral response function, ISRF) and sampled at a resolution of $0.01$ nm to create the necessary reference spectra in the data files used by the processor. Since the ISRF is (slightly) different for different

detector rows, the convolved reference spectra are determined per detector row. Given the relative smoothness of these convolved cross sections, interpolation to the radiance wavelength grid in Eq. (5) is performed by way of a 4th degree spline interpolation. The final set of reference spectra is:

- trace gas cross sections $\sigma_k(\lambda)$ in Eq. (5):
  - $NO_2$ from [25], [26] at $220$ K
  - $O_3$ from [27] and [28] at $243$ K
  - Water vapour ($H_2O_{vap}$) based on HITRAN 2012 data (see [17] and Sect. 4.1 of [29])
  - $O_2$–$O_2$ from [30] at $293$ K
  - Liquid water ($H_2O_{liq}$) from [31], resampled at $0.01$ nm with a cubic spline interpolation
- an effective Ring spectrum $I_{ring}(\lambda)$ following [32]; see [17] and Sect. 4.2 of [29]
- a high-resolution solar reference spectrum $E_{ref}(\lambda)$ from [33]; see also [34]

The temperature for the $O_3$, $H_2O_{vap}$ and $O_2$–$O_2$ cross section spectra is fixed. Variation of these cross section temperatures has little effect on the fit residual in the retrieval of $NO_2$ slant columns, since the shape of the differential $NO_2$ cross section is in good approximation invariant of temperature. In the case of TROPOMI, the baseline is to use an $NO_2$ cross section that has been measured for $220$ K.

The amplitude of the differential $NO_2$ cross section features, however, has a significant temperature dependency which needs to be accounted for. The resulting $NO_2$ slant column are corrected for deviations from $220$ K in the OMNO2B / TM5-MP-DOMINO processing step.

### 5.2.7 Some remarks regarding an intensity offset correction in DOAS implementations

Several DOAS applications include an intensity offset correction, a constant or linear in wavelength, to improve the retrievals in some spectral ranges. The precise physical origin of such an intensity offset is not known, but it is thought to be related to instrumental issues (e.g. incomplete removal of stray light or dark current in Level-1b spectra) and/or atmospheric issues (e.g. incomplete removal of Ring spectrum structures, vibrational Raman scattering in clear ocean waters); see, for example, [15], [35], [36], [37].

In Eq. (5) such an intensity offset correction would be represented by an additional term on the right hand side:

$$\ldots + \frac{S_{off}}{E_0(\lambda)} \cdot \sum_{m=0}^{n_{off}} c_m \lambda^m \tag{9}$$

with fit parameters $c_m$ and $S_{off}$ a suitable scaling factor which has the unit of the irradiance $E_0$; in most applications $n_{off} = 0$ or $1$ in case an intensity offset is included. (Note that the $NO_2$ slant column retrieval algorithm of OMI collection 3 is not able to handle such an intensity offset correction.)

The possibility of an intensity offset correction has been implemented in the TROPOMI and OMI $NO_2$ slant column processor, but this option is currently turned off as (i) it has not been fully tested yet, (ii) we would first like to understand the physical meaning and implications of such a correction term, and (iii) we need to investigate whether it might be relevant for TROPOMI and/or OMI $NO_2$ retrievals. For a first discussion see [18], based on which an intensity offset correction will not be included in the regular TROPOMI $NO_2$, also because instrumental effects such as straylight and dark current are corrected for in the spectral calibration in the Level 0-to-1b processor ([38]; [39]). For a recent discussion, see also [21].

## 5.3 Other OMI data and external data included by OMNO2A

This section describes the other OMI data and the external data inserted by the OMNO2A processor in the OMI $NO_2$ slant column data product.

### 5.3.1 Row anomaly flags

Since June 2007 OMI has suffered from the so-called "row anomaly" phenomenon, where certain across-track fields of view (rows) are seemingly blocked, resulting in abnormally low radiance reading (cf. [40], [13]).

The Level-1b data product [3] provides an across-track flag for the row anomaly (copied to `xtrack_quality` in the $NO_2$ data product). This flag distinguishes between different reasons for the anomaly [13], but from the $NO_2$ data point of view only "not affected" (flag value zero) or "affected" (flag value larger than zero) is relevant. The row anomaly appears to be affecting different sets of rows south and north of a certain orbit phase ($\varphi_{\text{sat}}$; copied to `satellite_orbit_phase` in the $NO_2$ data product). The collection 4 Level-1b processor makes the separation at $\varphi_{\text{sat}} = 0.580$; in collection 3 the separation was at $\varphi_{\text{sat}} = 0.560$.

The Level-1b row anomaly flagging seems a little too strict at times, too loose at other times, and it is a fixed flagging. To be more flexible a "dynamic row anomaly flag" (written to `xtrack_quality_dynamic` in the $NO_2$ data product) is used, which just has a zero for "not affected" and a one for "affected", while it places the south-north separatation at $\varphi_{\text{sat}} = 0.595$. This dynamic flag is based on a per row along-track average of the absorbing aerosol index (AAI) in the OMAERO product [22]: if the average AAI is above $1.25$ $(1.00)$ for the southern (northen) part then the row is flagged as "affected".

Checking the behaviour of the two flags over time has made us decide to go for a combination of the L1B and the AAI-based flags:

– if both the L1B and AAI flag are set, the row is considered really bad,

– if only the L1B flag is set, the row is possibly affected,

– if the L1B flag is not set, the row is considered to be in good order, no matter what the AAI-based flag is.

Note that this approach ignores the inconsistency in the definitions of the two flags for the 44 or so scanlines with orbit phase $0.580 < \varphi_{\text{sat}} < 0.595$, i.e. between the different north-south separation of the two flags.

Denoting the L1B flag by $f_{\text{L1B}}$ and the AAI flag by $f_{\text{AAI}}$ we suggest to to determine a "row anomaly factor", $f_{\text{QA}}^{\text{anom}}$, with which the `qa_value` (discussed in Sect. 5.5.2) is multiplied, as follows:

$$\begin{array}{l} \underline{\text{if}} \quad f_{\text{L1B}} = 0 \quad \underline{\text{then}} \quad f_{\text{QA}}^{\text{anom}} = 1.00 \\ \underline{\text{else}} \\ \qquad \underline{\text{if}} \quad f_{\text{AAI}} = 0 \quad \underline{\text{then}} \quad f_{\text{QA}}^{\text{anom}} = 0.92 \\ \qquad \underline{\text{else}} \quad f_{\text{QA}}^{\text{anom}} = 0.05 \end{array} \tag{10}$$

where the precise non-zero value of $f_{\text{L1B}}$ is not considered.

The recipe of Eq. (10) was applied to all Pacific Ocean orbits, the result of which is shown in Fig. 5. Those orbits are used for monitoring the $NO_2$ slant column uncertainties mentioned Sect. 5.2.5, results of which are presented on a set of web pages [24].

For pixels with $f_{\text{QA}}^{\text{anom}} < 1.0$ the `row_anomaly_warning` in the `processing_quality_flags` needs to be set; this warning has value 32768 (bit 15). Based solely on $f_{\text{AAI}}$ the current (January 2026) OMNO2A processor

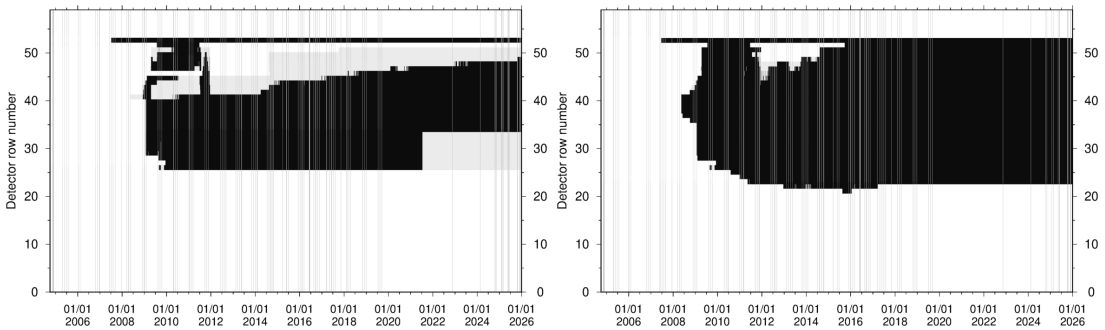


**Figure 5**: Row anomaly flagging in the southern (left) and northern (right) part of the Pacific Ocean orbits following the recipe of Eq. (10): light gray for $f_{\text{QA}}^{\text{anom}}$ = 0.92 and dark gray for $f_{\text{QA}}^{\text{anom}}$ = 0.05; vertical middle gray lines mark missing Pacific ocean orbits, primarily related to switches to zoom mode observations by OMI. Figures taken from [24].

accidentally sets as warning value 16384 (bit 14), which refers to `rebinned_pixel_warning` – this needs to be repaired at some point (cf. Sect. 5.4.1.3). Note that the `row_anomaly_filter` in the `processing_quality_flags`, which has value 99, is not used.

Note that the $f_{\mathrm{QA}}^{\mathrm{anom}}$ values mentioned above are first guesses – they may change once the first OMNO2B / TM5-MP-DOMINO processing tests have been evaluated.

### 5.3.2 Cloud cover and cloud pressure data

Trace gas retrievals from various satellite instruments, including TROPOMI, take the necesarry cloud cover information from near-infrared (NIR) windows, notably the $O_2$ A-band [41] and $O_2$ B-band [42] bands. The OMI spectral range does not cover these NIR bands and most OMI retrievals make use of cloud cover information derived from the $O_2$–$O_2$ absorption feature at 477 nm [43].

The cloud data is derived from the $O_2$–$O_2$ slant column density retrieved in the wavelength window $460 - 490$ nm with a DOAS approach similar to the $NO_2$ retrieval (Sect. 5.2.1), with the model function formulated in Eq. (5) and using cross-sections $\sigma_k(\lambda)$ for $O_2$–$O_2$, $NO_2$ and ozone.

The OMCLDO2 data product, described in the ATBD [44], is read by OMNO2A, which copies the following data fields to the $NO_2$ data product file (cf. Table 3 in Sect. 5.5):

- Cloud albedo ($A_{\mathrm{c}}$), cloud fraction ($f_{\mathrm{eff}}$), and cloud pressure ($p_{\mathrm{c}}$)
- Scene albedo ($A_{\mathrm{sc}}$), and scene pressure ($p_{\mathrm{sc}}$)

The first set provides the cloud pressure that is needed for the calculation of the cloud fraction & cloud radiance fraction in $NO_2$ window (Sect. 5.3.3). Depending on the snow/ice cover (Sect. 5.3.5) in a given ground pixel, the vertical column density step selects between cloudy situations and cloud-free snow/ice situations to calculate an appropriate air-mass factor (AMF).

### 5.3.3 Cloud fraction & cloud radiance fraction in $NO_2$ window

Though the difference in wavelength range of the $NO_2$ fit window and the $O_2$–$O_2$ fit window used for the cloud cover data (Sect. 5.3.2) is small, the same approach as for the TROPOMI $NO_2$ retrieval is used for the OMI collection 4 $NO_2$ retrieval: (i) use the cloud pressure $p_{\mathrm{c}}$ from the cloud product and (ii) retrieve the cloud fraction ($f_{\mathrm{eff,NO_2}}$) and cloud radiance fraction ($w_{\mathrm{NO_2}}$) from the $NO_2$ spectral window itself at $440$ nm. This is done by fitting the continuum reflectance at $440$ nm to a simulated reflectance constructed with the independent pixel approximation and radiative transfer calculations for the clear-sky and cloudy-sky part of the pixel, using the appropriate surface albedo in that spectral window, $A_{\mathrm{s,NOtwo}}$, as forward model parameter. Here $A_{\mathrm{s,NOtwo}}$ is taken from the OMI albedo climatology (cf. Sect. 5.3.4) at $440$ nm, interpolated linearly in time, and using nearest neighbour sampling in latitude and longitude.

The continuum reflectance at $440$ nm could be determined from the observed spectrum, averaged over a small wavelength interval, but that may lead to unexpected values, e.g. in case of spikes in the measurement or missing wavelength pixels. Instead, the modelled reflectance of Eq. (5) is evaluated at $\lambda_{c,\mathrm{NO_2}} = 440$ nm, without taking the absorbing trace gases into account: $R_{\mathrm{TOA}}(\lambda) = P(\lambda) \cdot (1 + C_{\mathrm{ring}})$, where the $C_{\mathrm{ring}}$ term is included because Rayleigh scattering is a combination of elastic Cabannes scattering and inelastic Raman scattering without the spectral structures of the latter. The trace gas absorption is not taken into account here because the reflectance LUT used for the determination of the cloud fraction does not include trace gas absorption either. This reflectance-based approach to determine the cloud fraction in the $NO_2$ window is very similar to FRESCO [41] and explicitly accounts for Rayleigh scattering and involves the calculation of LUTs with the TOA reflectance at $440$ nm as a function of viewing geometry, surface/cloud albedo, and surface/cloud pressure. See Appendix D for details on the cloud fraction retrieval.

In the above calculation the surface albedo is reduced to match the top-of-atmosphere reflectance in case the top of atmosphere reflectance is lower than expected for the climatological surface albedo (see Sect. 5.3.4 and App. D.1), which would lead to negative cloud fractions and unrealistic cloud pressures. Another special case is the treatment of very high cloud fractions. If the scene albedo indicates an elevated scene height and a scene albedo higher than $0.8$, the parameters from the scene retrieval are used: the cloud fraction is set to $1$, the cloud albedo is set to the scene albedo and the cloud pressure is set to the scene pressure. This will prevent odd behaviour for scenes with cloud fraction $f_{\mathrm{eff}} > 1$. Care has been taken to not use this for snow scenes.

### 5.3.4 Surface albedo

The OMI collection 4 $NO_2$ and $O_2$–$O_2$ retrievals use the surface albedo climatology of OMI aggregated to a grid of $0.5° \times 0.5°$; see Kleipool et al. [2008], which describes a climatology made from 3 years of OMI data. Meanwhile the climatology has been improved by using 5 years of data, based on the the same method [46]. For this reason OMNO2A inserts the surface albedo at $477$ nm used for the $O_2$–$O_2$ cloud retrieval ($A_s$) and at $440$ nm for the $NO_2$ retrieval ($A_{s,NO_2}$). The climatological values of the surface albedo are adapted in case the snow/ice flag (cf. Sect. 5.3.5) indicates there may be substantial differences in albedo; see Appendix E for some details on this correction.

Meanwhile, a TROPOMI surface albedo database, based on multiple years of reprocessed and offline level-1b data (see [47], [48]), is available at a resolution of $0.125° \times 0.125°$ and in use for TROPOMI $NO_2$ processing. We are considering to use this database for the final OMI collection-4 $NO_2$ reprocessing.

### 5.3.5 Snow and ice cover

Substantial errors are introduced if the real albedo differs considerably from what is expected from the albedo climatology, for example in the case of the sudden snowfall or ice cover. Correcting the surface albedo from the climatology (which contains a climatological snow cover) using knowledge of actual snow/ice cover will therefore improve the final data product, in terms of the retrieval itself and for flagging such cases. For the $A_{s,NO_2}$ this correction follows the approach included in the OMI cloud data product OMCLDO2 (Sect. 5.3.2) to adapt the surface albedo in the $O_2$–$O_2$ fit window (i.e. at $477$ nm); Appendix E provides some details on this correction of the surface albedo reported in the Level-2 file.

The snow/ice cover data for the OMI processing comes from the Primary Ancillary Data Geo-Colocated to OMI/Aura VIS 1-Orbit L2 Swath 13x24km (OMVANC) data product [49]. The fields in this product include snow cover, sea ice cover, land cover, terrain height, row anomaly flag, and pixel area.

For the TROPOMI processing, the snow/ice information is taken from the ECMWF analysis snow cover data [50], which is derived from synoptic data and from the Interactive Multi-sensor Snow and Ice Mapping System (IMS); [51] show that IMS has better agreement with in situ observations over North America and that NSIDC/NISE [52] misses a significant number of snow-covered pixels. For the OMI collection-4 reprocessing, we could opt for the ERA-5 reanalysis data of ECMWF.

The cloud algorithm (Sect. 5.3.2) provides two sets of data: (i) the effective cloud fraction $f_{\text{eff}}$ and cloud pressure $p_c$ using a cloud albedo $A_c = 0.8$, and (ii) the scene albedo $A_{sc}$ and the scene pressure $p_{sc}$ assuming a cloud fraction $f_{\text{eff}} = 0.0$. With the snow/ice flag the $NO_2$ processing will select which of these two sets is used for the determination of the AMFs and subsequent vertical $NO_2$ columns. When the snow/ice flag indicates that there is more than a 1% snow/ice coverage, the OMNO2B / TM5-MP-DOMINO retrieval will move to scene mode by setting the cloud radiance fraction equal to $1.0$. The $A_{sc}$ and $p_{sc}$ from cloud cover data are then used to determine the effective albedo and pressure of this (ficticious) cloud. Which mode is used can be found via the selection criteria of the `qa_value` definition, listed in Appendix F; see also the $NO_2$ Product User Manual (PUM; [53]).

### 5.3.6 Surface altitude and pressure

The surface altitude (DEM) of a ground pixel is the mean of the sub-pixel altitude derived from the GMTED2010 surface elevation database [3] – it is taken from the Level-1b data product. The surface pressure comes from the GEOS-5 FP-IT 3D Time-Averaged Single-Level Diagnostics Geo-Colocated to OMI/Aura VIS 1-Orbit L2 Swath 13x24km V4 (OMVFPSLV) data product [54]. The fields in this product include boundary layer top pressure, tropopause pressure, surface pressure, surface skin temperature, and vertical wind profiles at 10 m.

The AMFs in the vertical column step of OMNO2B / TM5-MP-DOMINO depend on the surface pressure, $p_s$. For the OMI $NO_2$ retrieval this information is obtained from the TM5-MP model ($1° \times 1°$) driven by ECMWF meteorological data. Because the TM5-MP information is representative for spatially coarse pressures, the TM5-MP results are corrected based on the method described in [55] and [16]. This correction computes a new surface pressure based on the difference between the corresponding spatially coarse terrain height and the actual, pixel-averaged terrain height based on a 3-km resolution digital elevation map [56].

[3] http://topotools.cr.usgs.gov/gmted_viewer/

### 5.3.7 Absorbing Aerosol Index

The OMI absorting aerosol index (AAI) data product, OMAERO [22], provides the $354/388$ nm wavelength pair AAI, which is passed on by OMNO2A to the $NO_2$ data product file as additional flag. The AAI is currently not used in the `qa_value` (App. F), but this may change in an upcoming update.

## 5.4 Updates needed and requested in the OMNO2A processor

The current (January 2026) OMNO2A processor provides data files with the $NO_2$ SCD data (Sect. 5.2) and some auxiliary data (Sect. 5.3). Below follow some requests for corrections and improvements of these files (Sect. 5.4.1) and for improvements in the OMNO2A processor in view of recent developments in the TROPOMI $NO_2$ product (Sect. 5.4.2).

### 5.4.1 Requested corrections of the OMNO2A output files

**5.4.1.1 Global attributes** The OMNO2A netCDF data file has a limited number of global attributes. For tracability and to further aid the data user and inspired by the list of global attributes in a TROPOMI $NO_2$ file, at least the following global attributes are requested:

```
institution = "NASA / KNMI" ;
platform = "EOS-Aura" ;
sensor = "OMI" ;
id = "OMI-Aura_L2A-OMNO2_2005m1003t093252-o006482_v004-2022m1216t204201" ;
identifier_product_doi = "..../...." ;
identifier_product_doi_authority = "http://dx.doi.org/" ;
processing_status = "OFFL-processing slant column product" ;
vcd_processor = "N/A" ;
```

In the OMNO2B / TM5-MP-DOMINO processing step that calculates the VCD, TM5-MP (re)writes the global attributes `processing_status` to "`OFFL-processing nominal product`" and sets `vcd_processor` to hold the correct version information.

**5.4.1.2 Data variable attributes** An attribute of two data variables in the in the OMNO2A files in the `PRODUCT/SUPPORT_DATA/INPUT_DATA` section contains an incorrect reference wavelength value:

```
surface_albedo:radiation_wavelength = 758.
scene_albedo:radiation_wavelength = 758.
```

which need to read:

```
surface_albedo:radiation_wavelength = 477.
scene_albedo:radiation_wavelength = 477.
```

**5.4.1.3 Data variables for dimensions** The $NO_2$ netCDF output file contains a number of dimensions for the data variables:

```
group: PRODUCT
     scanline = 1644 ;
     ground_pixel = 60 ;
     corner = 4 ;
     time = 1 ;
     layer = 34 ;
     vertices = 2 ;

group: PRODUCT/SUPPORT_DATA/DETAILED_RESULTS
     polynomial_exponents = 6 ;
     intensity_offset_polynomial_exponents = 1 ;
```

In the TROPOMI files all the dimensions have associated data variables in the group where the dimensions are defined. In the OMNO2A files that is not the case, with the exception of the `time` dimension, which has a data variable `time` is the data `PRODUCT/SUPPORT_DATA/GEOLOCATIONS`).

Dimensional data variables may be very useful to data users, in particular the along-track `scanline` and across-track `ground_pixel` indices, for example, when making regional data selections. The request therefore is to add data variables for all dimensions.

**5.4.1.4 Row anomaly warning flag** As mentioned in Sect. 5.3.1, in case pixel is flagged for the row anomaly, the OMNO2A processor accidentally sets as warning value 16384 (bit 14), which refers to `rebinned_pixel_warning`, rather then the value 32768 (bit 15) listed for the `row_anomaly_warning`.

### 5.4.2 Improvements in the OMNO2A processor

Since the TROPOMI $NO_2$ processor was ported to OMNO2A collection 4, the TROPOMI $NO_2$ slant column retrieval has undergone some improvents and we would like to include those in OMNO2A as well before making collection 4 slant column data final and proceed with the subsequent vertical column calcualtions in OMNO2B / TM5-MP-DOMINO:

- TROPOMI v2.7.1 included the addition of the Wald-Wolfowitz runs test to evaluate the DOAS fit residual, which has led to two new variables in the data product; see Sect. 6.2.3 in [4] and [21].
- TROPOMI v2.7.1 also included a change in the correction of the surface albedo using the snow/ice flag, described in App. E; see [4].
- TROPOMI v2.9.1 included an adaption of the $NO_2$ DOAS SCD retrieval to deal with an issue in the measured reflectance around $430$ nm; see Sect. 6.2.4 in [4] and the associated publication [21].

In addition there are some possible further improvements to consider:

- The use of the TROPOMI-based surface albedo database instead of the OMI one – see Sect. 5.3.4.
- The use of the ERA-5 ECMWF reanalysis for snow cover data – see Sect. 5.3.5.

## 5.5 The OMNO2A data product

The main product of OMNO2A is the total $NO_2$ slant column ($N_{s,NO_2}$), but the file also contains data variables for the $NO_2$ vertical column data ($N_{v,NO_2}$), which are the main result of the OMNO2B / TM5-MP-DOMINO step. In addation to that the OMI $NO_2$ product file contains data variables for the intermediate steps, such as the results from the DOAS retrieval, output from the data assimilation, cloud information, input database information, flags, uncertainties and the AMF calculation results. The attributes in the file provide full traceability of the data product (including information on processor version, settings, inputs).

The off-line (re)processing of the full OMI $NO_2$ slant and vertical column retrieval algorithm takes place at two locations (cf. Fig. 2):

(1) The first step of the $NO_2$ processing system, the DOAS retrieval of OMNO2A (Sect. 5.2), ingests the Level-1b spectra and is running in the NASA/OMI processing system OSIPS. Also performed in the NASA/OMI system – in separate processing chains – are the OMCLDO2 cloud retrieval (Sect. 5.3.2) and the OMAERO aerosol retrieval (Sect. 5.3.7), needed by several Level-2 data products. The OMNO2A processor uses the slant column and cloud cover data to assemble a $NO_2$ vertical column data product wherein the data fields of the vertical column calculation are left blank.

(2) The OMNO2A data product is then transferred to ECMWF. Once a day the data of all orbits are ingested in the data assimilation / chemistry transport model TM5-MP, to compute the off-line $NO_2$ product using the CTM model profiles computed with the latest analysis ECMWF meteorological fields. This step is referred to as OMNO2B / TM5-MP-DOMINO in Fig. 2 and will be described in a separate ATBD. The full off-line $NO_2$ data product will released for further usage.

Table 3 provides an overview of the data variables in the OMNO2A $NO_2$ data product. The fifth column in that table lists in which step(s) the content of the data variable is set, while the sixth column lists the data group in the netCDF file where the data variable can be found. Data variables not set contain fill values.

### 5.5.1 Unit definitions

In order to comply with the SI unit definitions, the OMI $NO_2$ data product file gives trace gas concentrations in mol/m$^2$, rather than in the commonly used unit molec/cm$^2$. The following multiplication factors enable the user to easily make the conversions, if needed:

- The multiplication factor to convert mol/m$^2$ to molec/cm$^2$ is $6.02214 \times 10^{19}$.
- The multiplication factor to convert mol/m$^2$ to DU is $2241.15$.
- The $O_2$–$O_2$ concentration is given in mol$^2$/m$^5$; the multiplication factor to convert this to the commenly used unit molec$^2$/cm$^5$ is $3.62662 \times 10^{37}$.

### 5.5.2 Ground pixel quality flags

The output for each ground pixel is accompanied by three flags indicating the status of the results of the processing and the retrieval. The main PRODUCT group contains the "quality assurance value" (`qa_value` or $f_{QA}$), which is a continuous variable, ranging from 0 (no output) to 1 (all is well). Warnings that occur during processing or results of the processing can be reasons to decrease the flag value. The `qa_value` is the main flag for data usage:

- `qa_value` $> 0.75$.
  For most users this is the recommended pixel filter. This removes clouds (cloud radiance fraction $> 0.5$), scenes covered by snow/ice, errors and problematic retrievals.
- `qa_value` $> 0.50$.
  This adds the good quality retrievals over clouds and over scenes covered by snow/ice. Errors and problematic retrievals are still filtered out. In particular this may be useful for assimilation and model comparison studies.

The determination of the `qa_value` is described in detail in Appendix F. The `qa_value` indicates whether the footprint is cloud covered or not, and whether there is snow or ice on the surface. It is set to $0$ (zero) if anywhere in the processing an error has occured. Warnings related to the South Atlantic Anomaly, sun glint, or missing

**Table 3**: Overview of the data sets, their units, types and sizes, in the main data output product file, listed alphabetically. All quantities followed by a $*$ in the "symbol" column consist of the value and the associated precision; for the vertical column densities the precisions are listed explicitly to clearly show the different types of precisions.

| *name/data* | *symbol* | *unit* | *description* | *source* [a] | *data group* [b] | *comments* |
|---|---|---|---|---|---|---|
| aerosol absorbing index | — | 1 | L2 AAI $354/388$ nm wavel. pair | X | PSI | added as flag |
| air-mass factor | $M^{\text{trop}}$ | 1 | tropospheric AMF | V | P | — |
| | $M^{\text{trop}}_{\text{clr}}$ | 1 | clear-sky tropospheric AMF | V | PSD | — |
| | $M^{\text{trop}}_{\text{cld}}$ | 1 | cloudy tropospheric AMF | V | PSD | — |
| | $M^{\text{strat}}$ | 1 | stratospheric AMF | V | PSD | — |
| | $M$ | 1 | total AMF | V | P | — |
| averaging kernel | $\mathbf{A}$ | 1 | — | V | P | † |
| chi-squared | $\chi^2$ | 1 | $\chi^2$ of the $NO_2$ DOAS fit | S | PSD | cf. Eq. (2) |
| cloud albedo | $A_{\text{c}}$ | 1 | used in the cloud retrieval | X | PSI | cf. Sect. 5.3.2 |
| cloud fraction | $f_{\text{eff}}$ | 1 | from the cloud retrieval | X | PSI | — |
| cloud fraction $NO_2$ | $f_{\text{eff,NO}_2}$ | 1 | for the $NO_2$ VCD | S | PSD | in $NO_2$ fit window |
| cloud pressure | $p_{\text{c}}$ | Pa | from the cloud retrieval | X | PSI | — |
| cloud radiance fraction | $w_{\text{NO}_2}$ | 1 | for the $NO_2$ VCD | S | PSD | in $NO_2$ fit window |
| cloud selection flag | — | 1 | cloud product selection flag | V | PSD | obsolete variable |
| degrees of freedom | $D$ | 1 | of the slant column fit | S | PSD | — |
| DOAS fit results | $N_{\text{s,NO}_2}$ * | mol/m$^2$ | total $NO_2$ SCD | S | PSD | — |
| | $N^{\text{geo}}_{\text{v,NO}_2}$ | mol/m$^2$ | geometric $NO_2$ column | S | PSD | cf. Eq. (7) |
| | $N_{\text{s,H}_2\text{O}_{\text{liq}}}$ * | m | $H_2O_{liq}$ coeff. in $NO_2$ window | S | PSD | — |
| | $N_{\text{s,H}_2\text{O}_{\text{vap}}}$ * | mol/m$^2$ | $H_2O_{vap}$ SCD in $NO_2$ window | S | PSD | — |
| | $N_{\text{s,O}_2\text{–}\text{O}_2}$ * | mol$^2$/m$^5$ | $O_2$–$O_2$ SCD in $NO_2$ window | S | PSD | — |
| | $N_{\text{s,O}_3}$ * | mol/m$^2$ | $O_3$ SCD in $NO_2$ window | S | PSD | — |
| | $C_{\text{ring}}$ * | 1 | Ring coeff. in $NO_2$ window | S | PSD | — |
| geolocations flag | — | 1 | ground pixel quality flag | I | PSG | — |
| ghost column | $N^{\text{ghost}}_{\text{v}}$ | mol/m$^2$ | $NO_2$ column below the clouds | V | PSD | ‡ |
| ground pixel coordinates | $\delta_{\text{geo}}$ | ° | VIS pixel – latitude | I | PSG | centre, 4 corners |
| | $\vartheta_{\text{geo}}$ | ° | VIS pixel – longitude | I | PSG | centre, 4 corners |
| ground pixel index | — | 1 | across-track pixel index | U | P | cf. Sect. 5.4.1.3 |
| intensity off. coefficients | $c_m$ * | 1 | in the $NO_2$ DOAS fit | S | PSD | cf. Eq. (9) |
| land/water classification | — | 1 | surface type classification | X | PSI | — |
| measurement time | $t$ | s | VIS pixel | I | PSG | — |
| number of wavelengths | $n_\lambda$ | 1 | in the $NO_2$ fit window | S | PSD | # |
| number of iterations | $n_i$ | 1 | from the DOAS fit | S | PSD | — |
| polynomial coefficients | $a_m$ * | 1 | in the $NO_2$ DOAS fit | S | PSD | cf. Eq. (6) § |
| processing error flag | — | 1 | processing quality error flag | ISXV | PSD | cf. Sect. 5.5 |
| processing quality flags | — | 1 | processing quality flags | ISXPV | PSD | cf. Sect. 5.5 |
| qa value | $f_{\text{QA}}$ | 1 | quality assurance value | ISXV | PSD | cf. Sect. 5.5 & F |
| root-mean-square error | $R_{\text{RMS}}$ | 1 | RMS error of the $NO_2$ DOAS fit | S | PSD | cf. Eq. (4) |
| satellite coordinates | $z_{\text{sat}}$ | m | altitude of the satellite | I | PSG | — |
| | $\delta_{\text{sat}}$ | ° | latitude sub satellite point | I | PSG | — |
| | $\vartheta_{\text{sat}}$ | ° | longitude sub satellite point | I | PSG | — |
| | $\varphi_{\text{sat}}$ | 1 | relative offset in orbit | I | PSG | orbit phase |
| scanline index | — | 1 | along-track pixel index | U | P | cf. Sect. 5.4.1.3 |
| scene albedo | $A_{\text{sc}}$ * | 1 | from the cloud retrieval | X | PSI | — |
| scene pressure | $p_{\text{sc}}$ * | Pa | from the cloud retrieval | X | PSI | — |

*Table continues on next page*

non-critical input data lower the `qa_value`. The `qa_value` depends on the solar zenith angle, tropospheric air-mass factor and quality of the DOAS fit, and filters unrealistic albedo values.

The "processing quality error flag" (`processing_quality_error_flag`) contains the first individual event that led to processing failure. The "processing quality flags" (`processing_quality_flags`) contains the individual event that led to processing failure, and/or a precise record of the warnings that occurred during

**Table 3**: — *continued.*

| *name/data* | *symbol* | *unit* | *description* | *source* [a] | *data group* [b] | *comments* |
|---|---|---|---|---|---|---|
| snow-ice flag | — | 1 | snow/ice case flagging | X | PSI | — |
| stripe amplitude | $N^{\text{str}}_{\text{s,NO}_2}$ | mol/m$^2$ | $NO_2$ SCD stripe amplitude | V | PSD | — |
| surface albedo | $A_{\text{s}}$ | 1 | for the cloud retrieval | X | PSI | — |
| surface albedo $NO_2$ | $A_{\text{s,NO}_2}$ | 1 | for cloud fraction $NO_2$ window | X | PSI | cf. Sect. 5.3.2 |
| surface elevation | $z_{\text{s}}$ * | m | VIS pixel | X | PSI | — |
| surface pressure | $p_{\text{s}}$ | Pa | VIS pixel | X | PSI | — |
| TM5 pressure level | $A_l^{\text{TM5}}$ | Pa | — | V | P | ¶ |
| coefficients | $B_l^{\text{TM5}}$ | 1 | — | V | P | ¶ |
| TM5 tropopause layer index | $l_{\text{tp}}^{\text{TM5}}$ | 1 | — | V | P | — |
| vertical column density | $N^{\text{trop}}_{\text{v,NO}_2}$ | mol/m$^2$ | tropospheric $NO_2$ VCD | V | P | — |
| | $\Delta N^{\text{trop}}_{\text{v,NO}_2}$ | mol/m$^2$ | *id.* precision, kernel not applied | V | P | — |
| | $\Delta N^{\text{trop,kernel}}_{\text{v,NO}_2}$ | mol/m$^2$ | *id.* precision, kernel applied | V | P | — |
| | $N^{\text{strat}}_{\text{v,NO}_2}$ | mol/m$^2$ | stratospheric $NO_2$ VCD | V | PSD | — |
| | $\Delta N^{\text{strat}}_{\text{v,NO}_2}$ | mol/m$^2$ | *id.* precision | V | PSD | — |
| | $N_{\text{v,NO}_2}$ | mol/m$^2$ | total $NO_2$ VCD | V | PSD | $\equiv N_{\text{s}}/M$ |
| | $\Delta N_{\text{v,NO}_2}$ | mol/m$^2$ | *id.* precision, kernel not applied | V | PSD | — |
| | $\Delta N^{\text{kernel}}_{\text{v,NO}_2}$ | mol/m$^2$ | *id.* precision, kernel applied | V | PSD | — |
| | $N^{\text{sum}}_{\text{v,NO}_2}$ | mol/m$^2$ | summed $NO_2$ VCD | V | PSD | $\equiv N_{\text{v}}^{\text{trop}} + N_{\text{v}}^{\text{strat}}$ |
| | $\Delta N^{\text{sum}}_{\text{v,NO}_2}$ | mol/m$^2$ | *id.* precision | V | PSD | — |
| viewing geometry data | $\theta_0$ | ° | solar zenith angle | I | PSG | at surface |
| | $\phi_0$ | ° | solar azimuth angle | I | PSG | at surface |
| | $\theta$ | ° | viewing zenith angle | I | PSG | at surface |
| | $\phi$ | ° | viewing azimuth angle | I | PSG | at surface |
| water fraction | — | % | water fraction in percent | X | PSI | — |
| wavelength calibration | $w_s$ * | nm | wavelength shift | S | PSD | cf. Eq. (8) |
| radiance | $w_q$ * | 1 | wavelength stretch | S | PSD | cf. Eq. (8) |
| | $\chi^2_w$ | 1 | $\chi^2$ of the calibration | S | PSD | cf. Eq. (8) |
| wavelength calibration | $w_s^{E0}$ * | nm | wavelength shift | S | PSD | cf. Eq. (8) |
| irradiance | $(\chi_w^{E0})^2$ | 1 | $\chi^2$ of the calibration | S | PSD | cf. Eq. (8) |
| xtrack quality | $f_{\text{L1B}}$ | 1 | row anomaly flag from L1b data | I | PSD | cf. Sect. 5.3.1 |
| xtrack quality dynamic | $f_{\text{AAI}}$ | 1 | dynamic row anomaly flag | S | PSD | cf. Sect. 5.3.1 |

[a] Column 5 lists the source of the data:
I = input from Level-1B data product,
S = from the OMNO2A slant column retrieval (Sect. 5.2),
X = inserted by OMNO2A from external data (Sect. 5.3),
U = to be added in a processor update (Sect. 5.4), and
V = from the vertical column calculation of OMNO2B / TM5-MP-DOMINO.

[b] Column 6 lists the data group of the variable:
P = main PRODUCT group,
PSI = PRODUCT/SUPPORT_Data/INPUT_DATA,
PSG = PRODUCT/SUPPORT_Data/GEOLOCATIONS,
PSD = PRODUCT/SUPPORT_DATA/DETAILED_RESULTS.

† The number of TM5-MP layers is $n_l = 34$.

‡ The $NO_2$ ghost column is the TM5-MP $NO_2$ profile integrated from the surface to the cloud pressure level.

# The actual number of wavelengths $n_\lambda$ used in the fit; cf. Eqs. (2) and (4).

§ The degree of the DOAS polynomial is $n_p = 5$

¶ One set of $n_l + 1$ (see note †) TM5-MP pressure level coefficients per data granule.

processing. The definitions and usage of these flags is harmonised between the Level-2 data products of OMI.

# A Spike removal in the DOAS fit

The OMI $NO_2$ processor contains a "spike removal" in the DOAS fit (Sect. 5.2.1), to remove strong outliers in the fit residual. Below follows a brief description adapted from the TROPOMI $NO_2$ ATBD (see [57], App. F).

Outliers may be caused by, e.g., high-energy particles hitting the CCD detector (so-called transients), variations in the dark current, or bad pixels not correctly flagged in the Level-1b data. After removal of an outlier from the measured reflectance, the $NO_2$ DOAS fit is redone. To avoid ending up in a cycle, the new fit residual is not checked again for outliers.

To detect outliers the so-called box-plot method used by [43] for the OMI $O_2$–$O_2$ cloud algorithm is implemented. This method [58] determines lower and upper values based on the first and third quartiles, $Q_1$ and $Q_3$, i.e. the 25th and 75th percentile of a distribution (the second quartile, $Q_2$, is the median). For data with a Gaussian distribution $Q_1 = -0.67$ and $Q_3 = +0.67$, which means that for normally distributed data, one-half of the data is within 2/3-rd of a standard deviation unit of the mean [59].

If a certain value is larger than $Q_3 + Q_f \cdot Q_{3-1}$ or lower than $Q_1 - Q_f \cdot Q_{3-1}$, with $Q_{3-1} = Q_3 - Q_1$ the inter-quartile range and $Q_f$ a suitable multiplication factor, it is termed an outlier. If $Q_f$ is set too low, valid observations are unjustly removed from the fit and the $NO_2$ SCD error will be underestimated.

For the so-called inner fences, defined by $Q_f = 1.5$ [58] and used by [43], $0.74\%$ of normally distributed data is termed an outlier [59]. This means that on average for each OMI ground pixel 2 or 3 wavelength pixels will be designated as outlier, if the $NO_2$ fit residual on the about $287$ wavelength pixels within the $NO_2$ fit window would be normally distributed. In reality the fit residual is not normally distributed but drops off less steep, hence $Q_f = 1.5$ is clearly too low a factor to use, as this would inevitably lead to the removal of valid observations.

For the so-called outer fences, defined by $Q_f = 3.0$ [58], about $0.002\%$ of normally distributed data is termed an outlier [60], or in terms of the $NO_2$ fit window about 1/100-th of a wavelength pixel (i.e. roughly one wavelength pixel per 100 ground pixels), so that the chance that valid observations are removed is small. Based on this evaluation, the TROPOMI $NO_2$ processor is configured to use $Q_f = 3.0$.

Most ground pixels with outliers have one or two outliers, even those over the South Atlantic Anomaly, where high-energy particle impacts on the CCD detector occur more often. Ground pixels suffering from saturation effects, however, may have a much larger number of outliers. Due to the high SNR of TROPOMI, bright scenes, notably over clouds in the tropics, the CCD may be over-exposed. For OMI, with its lower SNR this is less of a problem, if it occurs at all.

The treatment of saturation and outliers on wavelength pixels in the $NO_2$ SCD retrieval is governed by configuration parameters, listed in Table 4: the settings for $NO_2$-v2.2 are made for use with version-2 Level-1b data in mind; prior $NO_2$ processor versions accepted very few saturation flags.

**Table 4**: Configuration parameters in the $NO_2$ processing related to saturation in the Level-1b spectra and outliers in the $NO_2$ retrieval residual. The first check is disabled in the OMI processing; in the TROPOMI processing the maximum fraction is set to $0.25$.

| configuration parameter | setting |
|---|---|
| The maximum fraction of the radiance spectrum that is allowed to be flagged as saturated before the ground pixel is skipped | N/A |
| The maximum number of outliers that is allowed to be in a spectrum before the ground pixel is skipped | 10 |

# B Wavelength calibration

Wavelength calibration of the OMI radiance and irradiance spectra of the Level-1b input data is performed by the Level-2 processors within the relevant fit window. For the wavelength calibration of the radiances an atmosphere model is needed, especially at the shorter wavelengths where ozone absorption is significant, but also the Ring effect modifies the radiance spectra in ways that have to be taken into account when calibrating the wavelength. In the unlikely case the wavelength calibration of radiance or irradiance fails, the retrieval will be performed using the respective nominal wavelenth grids.

For the calibration of a complete band or a complete detector, the calibration is split up in micro-windows, and a polynomial is drawn through the micro-windows to cover the whole band. When fitting for a specific retrieval window, a single fit covering the retrieval window is more appropriate. The model function that is used for the radiance wavelength calibration is a modified version of a DOAS fit. Sections B.1 and B.2 describe the generic wavelength fit used in most retrieval algorithms for OMI, in section B.3 the actual application to $NO_2$ retrieval is discussed. For the irradiance calibration the same procedure is used, except that atmosphere related effects should be disabled, specifically the Ring effect should *not* be included in this fit. The polynomial order $N$ is set to 1 for the irradiance fit.

## B.1 Description of the problem

The OMI Level-1b radiance spectra have a nominal wavelength scale ($\lambda_{\text{nom}}$), but this wavelength grid is not corrected for inhomogeneous slit illumination. The measurements are also not temperature corrected, but because the instrument itself is temperature stabilized it is expected that this effect can be ignored. The Level-2 processors must correct the nominal wavelength scale of the radiance measurements for inhomogeneous slit illumination due to the presence of clouds in the field of view.

One would like to follow the calibration of the irradiance spectra, for a short wavelength interval. The range $\lambda_{\text{fit}} = [\lambda_-, \lambda_+]$ is the approximate range on which to do the wavelength calibration. To avoid non-linearities this wavelength range is tailored to the specific Level-2 algorithm. For each detector row the nominal wavelength $\lambda_{\text{nom}}$ is adjusted with a wavelength offset (or: shift) $w_s$ and a wavelength stretch $w_q$ to find the calibrated wavelength $\lambda_{\text{cal}}$:

$$\lambda_{\text{cal}} = \lambda_{\text{nom}} + w_s + w_q \left( 2 \frac{\lambda_{\text{nom}} - \lambda_0}{\lambda_+ - \lambda_-} \right) + \cdots \tag{11}$$

with $\lambda_0$ the center of the fit window, $\lambda_-$ the beginning of the fit window and $\lambda_+$ the end of the fit window. In the third term the factor $2$ is used to ensure that the wavelength factor of the stretch lies in the range $[-1:+1]$. The higher order terms in Eq. (11) are ignored, even fitting $w_q$ is optional.

## B.2 Non-linear model function and Jacobian

The model function in the fit is similar to a non-linear DOAS equation. Instead of fitting the reflectance $R$, we fit the radiance $I$ directly, bringing the (model) irradiance $E_{\text{mod}}$ to the other side of the equation. The model function $\mathscr{M}$ is given by:

$$\mathscr{M}(\lambda_{\text{nom}}; a_0, \ldots, a_N, C_{\text{ring}}, w_s, w_q, N_{\text{s},0}, \ldots, N_{\text{s},M}) = P_N(\lambda^*) \cdot \exp\left( \sum_{k=0}^{n_k} -N_{\text{s},k} \sigma_k(\lambda_{\text{cal}}) \right) \cdot \left( E_{\text{mod}}(\lambda_{\text{cal}}) + C_{\text{ring}} I_{\text{ring}}(\lambda_{\text{cal}}) \right) \tag{12}$$

with $\lambda_{\text{cal}}$ the calibrated wavelength as given by the first three terms in Eq. (11),

$$P_N(\lambda^*) = \sum_{k=0}^{N} a_j (\lambda^*)^j \,, \qquad \lambda^* \equiv 2 \frac{\lambda_{\text{nom}} - \lambda_0}{\lambda_+ - \lambda_-} \tag{13}$$

a polynomial of order $N$, $E_{\text{mod}}$ the reference irradiance spectrum, and $I_{\text{ring}}$ the Ring spectrum; both $E_{\text{mod}}$ and $I_{\text{ring}}$ are convolved with the instrument slit function (or: instrument spectral response function; ISRF). The spectra $\sigma_k$ ($k = 0, \ldots, M$) are optional absorption spectra that have a relevant impact on the radiance, for instance the $O_3$ absorption cross section. These additional reference spectra have also been convolved with the ISRF, but note that the DOAS assumption still applies: this merit function is not applicable to line absorbers such as $H_2O_{\text{vap}}$, $CH_4$, CO or $O_2$, and will fail at wavelengths below $\sim 320$ nm because the profile shape of $O_3$

is relevant at those wavelengths. The order of the polynomial is $1 \leq N \leq 5$, depending on the length of the fit window.

The wavelength calibration fit adjusts the parameters $a_0, \ldots, a_N, C_{\text{ring}}, w_s, w_q, N_{\text{s},0}, \ldots, N_{\text{s},M}$ to minimize $\chi^2$:

$$\chi^2 = \frac{1}{m-n} \sum_{i=0}^{m-1} \left( \frac{I_i \; - \; \mathscr{M}(a_0, \ldots, a_N, C_{\text{ring}}, w_s, w_q, N_{\text{s},0}, \ldots, N_{\text{s},M})}{\Delta I_i} \right)^2 \tag{14}$$

with $I_i$ the measured radiance at detector pixel index $i$, $\Delta I_i$ the precision of this radiance, and $m$ the number of spectral points between $\lambda_-$ and $\lambda_+$. The number of degrees of freedom is $m$ minus the number of fit parameters:

$$n = N + 1 + M + 1 + 3 \tag{15}$$

The additional 3 here is when fitting $C_{\text{ring}}$, $w_s$ and $w_q$; if $C_{\text{ring}}$ and/or $w_q$ are not fitted, the number of degrees of freedom increases.

To minimize the number of function calls in the optimisation routine derivatives with respect to the fit parameters as a Jacobian matrix need to be supplied, with $i$ the detector pixel index: The components of the Jacobian are given by Eqs. (16–20) below.

$$\frac{\partial \mathscr{M}_i}{\partial a_j} = (\lambda_i^*)^j \cdot \exp\left( \sum_{k=0}^{M} -N_{\text{s},k} \sigma_k(\lambda_{\text{cal},i}) \right) \cdot \left[ E_{\text{mod}}(\lambda_{\text{cal},i}) + C_{\text{ring}} I_{\text{ring}}(\lambda_{\text{cal},i}) \right] \tag{16}$$

$$\frac{\partial \mathscr{M}_i}{\partial C_{\text{ring}}} = P_N(\lambda_i^*) \cdot \exp\left( \sum_{k=0}^{M} -N_{\text{s},k} \sigma_k(\lambda_{\text{cal},i}) \right) \cdot I_{\text{ring}}(\lambda_{\text{cal},i}) \tag{17}$$

$$\begin{aligned} \frac{\partial \mathscr{M}_i}{\partial w_s} = P_N(\lambda_i^*) \cdot \exp\left( \sum_{k=0}^{M} -N_{\text{s},k} \sigma_k(\lambda_{\text{cal},i}) \right) \times \\ \left\{ \left( - \sum_{k=0}^{M} N_{\text{s},k} \left. \frac{\mathrm{d}\sigma_k}{\mathrm{d}\lambda} \right|_{\lambda=\lambda_{\text{cal},i}} \right) \cdot \left( E_{\text{mod}}(\lambda_{\text{cal},i}) + C_{\text{ring}} I_{\text{ring}}(\lambda_{\text{cal},i}) \right) \right. \\ \left. + \left( \left. \frac{\mathrm{d}E_{\text{mod}}}{\mathrm{d}\lambda} \right|_{\lambda=\lambda_{\text{cal},i}} + C_{\text{ring}} \left. \frac{\mathrm{d}I_{\text{ring}}}{\mathrm{d}\lambda} \right|_{\lambda=\lambda_{\text{cal},i}} \right) \right\} \end{aligned} \tag{18}$$

$$\begin{aligned} \frac{\partial \mathscr{M}_i}{\partial w_q} = P_N(\lambda_i^*) \cdot \exp\left( \sum_{k=0}^{M} -N_{\text{s},k} \sigma_k(\lambda_{\text{cal},i}) \right) \times \\ \left\{ \left( - \sum_{k=0}^{M} N_{\text{s},k} \lambda_i^* \left. \frac{\mathrm{d}\sigma_k}{\mathrm{d}\lambda} \right|_{\lambda=\lambda_{\text{cal},i}} \right) \cdot \left( E_{\text{mod}}(\lambda_{\text{cal},i}) + C_{\text{ring}} I_{\text{ring}}(\lambda_{\text{cal},i}) \right) \right. \\ \left. + \left( \lambda_i^* \left. \frac{\mathrm{d}E_{\text{mod}}}{\mathrm{d}\lambda} \right|_{\lambda=\lambda_{\text{cal},i}} + C_{\text{ring}} \lambda_i^* \left. \frac{\mathrm{d}I_{\text{ring}}}{\mathrm{d}\lambda} \right|_{\lambda=\lambda_{\text{cal},i}} \right) \right\} \end{aligned} \tag{19}$$

$$\frac{\partial \mathscr{M}_i}{\partial N_{\text{s},k}} = -P_N(\lambda_i^*) \cdot \sigma_k(\lambda_{\text{cal},i}) \cdot \exp\left( \sum_{k=0}^{M} -N_{\text{s},k} \sigma_k(\lambda_{\text{cal},i}) \right) \cdot \left( E_{\text{mod}}(\lambda_{\text{cal},i}) + C_{\text{ring}} I_{\text{ring}}(\lambda_{\text{cal},i}) \right) \tag{20}$$

The reference spectra $E_{\text{mod}}(\lambda)$, $I_{\text{ring}}(\lambda)$ and $\sigma_k(\lambda)$ are pre-convolved with the ISRF. During the fitting 4th degree splines are used to represent these spectra. An interesting feature is that a spline of the derivative with respect to the independent variable can be calculated from the parameters of the original spline (given that the derivatives are w.r.t. the wavelength, the resulting spline for these derivaties is 3rd degree).

These equation can be solved with various optimization routines, for instance Levenberg-Marquardt or Gauss-Newton, with or without constraints or regularization methods. After thorough testing the optimal estimation method as implemented in DISAMAR, which is based on Rodgers [12] and uses an unmodified Gauss-Newton to find the state vector for the next iteration, was selected for the OMI wavelength calibration. For this usage, the a-priori error estimates are set very large (see Sect B.2.1), so that these do not limit the solution, and a pre-whitening of the data is performed to improve numerical stability.

### B.2.1 Prior information for the optimal estimation fit

Optimal estimation needs prior information for the regularisation process during the fitting procedure, both a starting value and a covariance value. For input only the diagonal elements of the covariance matrix are specified, on output a full posteriori error covariance matrix is available. The polynomial coefficients are not important, the values and variance were estimated from a large number of retrievals. The Ring coefficient was taken from the same data set. The value for $w_s$ is taken from the spacing of the nominal grid. A $1$-$\sigma$ error of a third of the spacing of the wavelength grid seems reasonable: $\sigma_{\text{prior}}(w_s) = \Delta\lambda/3$. This value will mostly prevent fitting a shift $w_s$ that is larger than half of the grid spacing, which basically means the wavelength is not known at all. The prior value for $w_q$ is $0$ (zero), i.e. no stretch or squeeze. The range depends on the size of the fitting window, a consequence of the use of $\lambda^*$, as defined in Eq. (13). The current value is a deliberate overestimation. The slant column of $O_3$ is typically $0.18$ mol/m$^2$ (about $600$ DU); other trace gases are not included. An overview of the prior information used for OMI is given in Table 5.

**Table 5**: A-priori values and a-priori error for the optimal estimation wavelength fit for OMI. The ozone slant column is expressed in mol/m$^2$; the other quantities are dimensionless.

| Names | $a_0$ | $a_1$ | $a_{2,\ldots,N}$ | $C_{\text{ring}}$ | $w_s$ | $w_q$ | $N_{\text{s},O_3}$ |
|---|---|---|---|---|---|---|---|
| Prior | $1$ | $-0.5$ | $0.01$ | $6\times10^{-2}$ | $0$ | $0$ | $0.25$ |
| Covariance | $(1)^2$ | $(0.5)^2$ | $(0.1)^2$ | $(6\times10^{-2})^2$ | $(\Delta\lambda/3)^2$ | $(0.1)^2$ | $(0.18)^2$ |
| Optional | no | no | yes | yes | no | yes | yes |

## B.3 Application of the wavelength calibration in $NO_2$

For the retrieval of $NO_2$ the $N_{\text{s},O_3}$ is not fitted, as $O_3$ shows little structure and is a weak absorber in band 4, where the slant columns of $NO_2$ (window $405-465$ nm) and $O_2$–$O_2$ (window $460-490$ nm) are fitted. Testing with OMI has shown that there is no significant amount of stretch in the wavelength of the spectra of that instrument in the $405-465$ nm range and hence the $w_q$ fit parameter will remain turned off.

The order of the polynomial in Eq. (13) is set to $2$ and the Ring effect is included in the fit. The a-priori error of $w_s$ is set to $0.07$ nm.

# C High-sampling interpolation

After the wavelength calibration of the radiance spectrum, discussed in Appendix B, the irradiance and radiance observations need to be brought to the same wavelength grid in order to be able to compute the reflectance in Eq. (1). Because of the geometry of the solar observations, these measurements are shifted with respect to the radiance observations due to the Doppler shift caused by the motion of the satellite relative to the sun. Given that the irradiance spectrum is known better than the radiance spectrum, the irradiance spectrum is shifted to the radiance grid and the radiance observations are left without modification:

$$E_0\left(\lambda_{i,\text{earth}}\right) = \frac{E_{\text{high}}\left(\lambda_{i,\text{earth}}\right)}{E_{\text{high}}\left(\lambda_{i,\text{solar}}\right)} E_0\left(\lambda_{i,\text{solar}}\right) \tag{21}$$

with $E_0$ the observed irradiance, $E_{\text{high}}$ a high resolution solar reference spectrum, convolved with the instrument spectral response function, $\lambda_{i,\text{earth}}$ the wavelength of the earth radiance spectrum for pixel $i$, and $\lambda_{i,\text{solar}}$ the wavelength of the solar irradiance spectrum for pixel $i$. The index $i$ is synchronized between the radiance and irradiance observations, such that they refer to the same physical pixel on the detector. On $E_{\text{high}}$ 4th degree spline interpolation is used to find the value at the indicated wavelengths. The input data for the splines have sufficient spectral resolution to allow for this.

Fig. 6 shows the procedure graphically. Panel (d) shows the effect of spline interpolation on the irradiance data to find the values at the earth radiance wavelength grid. Errors are small but systematic. Note that these errors appear directly in the reflectance data. The reflectance in Eq. (1) is then be calculated a at the radiance wavelenth grid: $R_{\text{meas}}(\lambda_{i,\text{earth}}) = \pi I(\lambda_{i,\text{earth}})/\mu_0 E_0(\lambda_{i,\text{earth}})$.

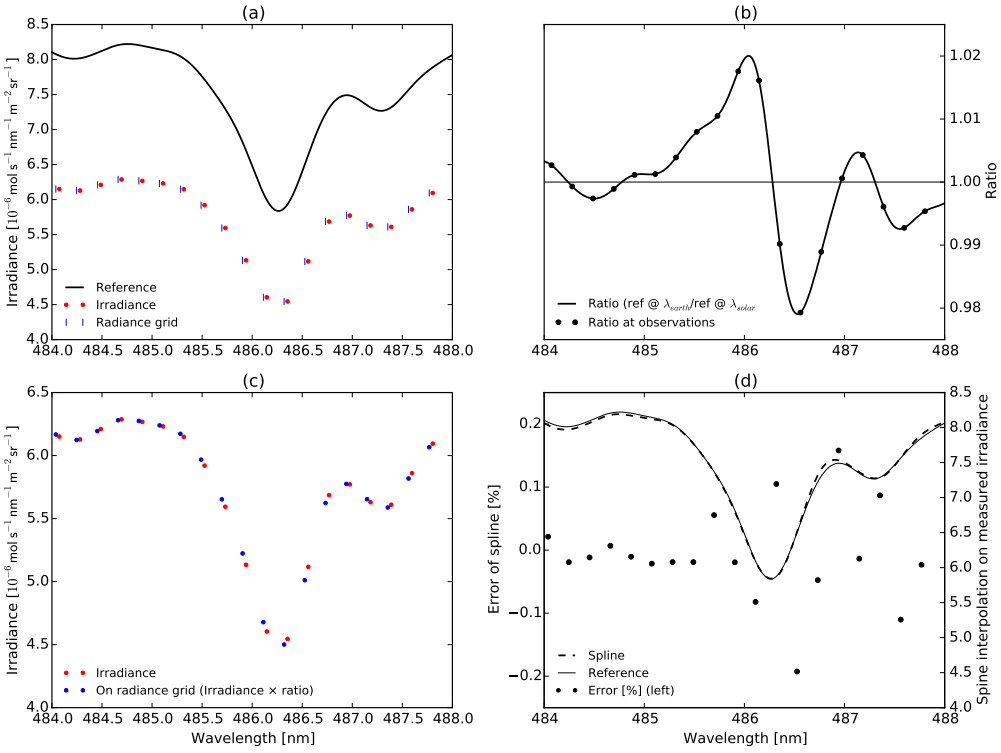


**Figure 6**: High sampling interpolation on part of a solar observation. (**a**) The red dots show the actual observation (taken from GOME-2A). The blue vertical lines indicate the wavelength grid of the radiance observation. The solid line shows a high resolution solar reference spectrum that has been convolved with the instrument spectral response function of the instrument (in this case GOME-2A). (**b**) The ratio $E_{\text{high}}\left(\lambda_{i,\text{earth}}\right)/E_{\text{high}}\left(\lambda_{i,\text{solar}}\right)$. (**c**) In red $E_0\left(\lambda_{i,\text{solar}}\right)$, in blue $E_0\left(\lambda_{i,\text{earth}}\right)$. (**d**) The solid line is the solar reference spectrum. The dash-dotted line is a high resolution irradiance spectrum created by spline interpolation directly on the observed irradiances, brought to the same average level of the window shown here to ease comparisons. The black dots indicate the error in % that are caused by using spline interpolation directly on the irradiance observations. Clear artifacts are caused by this, especially because noise on the observations becomes correlated between nearby points in the spectrum.

# D Effective cloud fraction in the $NO_2$ window

The cloud radiance fraction, $w_{NO_2}$, and the effective cloud fraction, $f_{\text{eff},NO_2}$ (Sect. 5.3.2), in the $NO_2$ fit window, can be computed from a look-up table (LUT) with the top-of-atmosphere (TOA) reflectance at $\lambda_{c,NO_2} = 440$ nm as a function of viewing geometry, surface & cloud albedo, and surface & cloud pressure, based on the continuum reflectance at $440$ nm of the measurement. The continuum reflectance at $440$ nm could be determined from the observed spectrum, averaged over a small wavelength interval, but that may lead to unexpected values, e.g. in case of spikes in the measurement. Instead, we have opted for using the modelled reflectance of Eq. (5) evaluated at $440$ nm. The approach is very similar to FRESCO [61] and explicitly accounts for Rayleigh scattering. The following description is adapted from [62].

The LUT assumes that the measured reflectance at TOA is defined as (cf. Eq. (1)):

$$R_{\text{TOA}}(\lambda) = \frac{\pi I(\lambda)}{\mu_0 E_0(\lambda)} \tag{22}$$

In the independent pixel approximation the cloud faction, $f_{\text{c}}$, for a given wavelength is given by:

$$f_{\text{c}} = \frac{R_{\text{TOA}} - R_{\text{s}}}{R_{\text{c}} - R_{\text{s}}} \tag{23}$$

and the cloud radiance fraction, the fraction of the total radiation that comes from the clouds, is given by:

$$w_{NO_2} = \frac{f_{\text{c}} R_{\text{c}}}{R_{\text{TOA}}} = \frac{f_{\text{c}} R_{\text{c}}}{f_{\text{c}} R_{\text{c}} + (1 - f_{\text{c}}) R_{\text{s}}} \tag{24}$$

where $R_{\text{s}}$ and $R_{\text{c}}$ are the reflectances at surface and cloud, respectively. These are computed from a limited LUT, based on Chandrasekhar ([63], Sect. 72). For bounding surface 'b', i.e. either surface ('s') or cloud ('c'):

$$R_{\text{b}}(\lambda, A_{\text{b}}(\lambda)) = R_0(\lambda) + \frac{A_{\text{b}}(\lambda)\, T(\lambda)}{1 - A_{\text{b}}(\lambda)\, s(\lambda)} \tag{25}$$

where:

| | | |
|---|---|---|
| $R_{\text{b}}(\lambda, A_{\text{b}}(\lambda))$ | = | The reflectance of the combined atmosphere-surface system related to the light coming from the boundary 'b', i.e. either surface ('s') or cloud ('c'). |
| $R_0(\lambda)$ | = | The reflectance of the atmosphere if the surface is perfectly black: $A_{\text{b}} = 0$. |
| $A_{\text{b}}(\lambda)$ | = | The albedo at the bounding surface, either cloud ($A_{\text{c}}$) or surface ($A_{\text{s}}$). |
| $T(\lambda)$ | = | The transmittance of the atmosphere, a measure for the probability that photons travel through the atmosphere, are reflected by a surface with unit albedo, and travel back to the sensor (reflections by the atmosphere back towards the surface are ignored here). |
| $s(\lambda)$ | = | The spherical albedo of the atmosphere for illumination at its lower boundary; $1/[1 - A_{\text{b}}(\lambda) s(\lambda)]$ is the sum of a geometrical series accounting for the reflections between the atmosphere and the surface. |

The transmittance of the atmosphere $T(\lambda)$ is a product of two terms depending on the viewing and solar zenith angles:

$$T(\lambda) = t(\lambda; \mu)\, t(\lambda; \mu_0) \tag{26}$$

where $\mu = \cos(\theta)$ and $\mu_0 = \cos(\theta_0)$ and:

$$t(\lambda; \mu) = \exp(-\frac{\tau(\lambda)}{\mu}) + \int_0^1 2\mu' T_0(\lambda; \mu, \mu') \mathrm{d}\mu' \tag{27}$$

In Eq. (27) we assume a plane parallel atmosphere; for a spherical shell atmosphere the factor $1/\mu$ in $\exp(-\tau/\mu)$ has to be replaced by a different expression.

The TOA reflectance related to the light coming from the boundary 'b', i.e. either surface ('s') or cloud ('c'), is a function of solar and viewing geometries and surface properties: $R_{\text{b}}(\lambda, A_{\text{b}}(\lambda)) = R_{\text{b}}(\lambda; \theta_0, \theta, \phi - \phi_0; p_{\text{b}}, A_{\text{b}}(\lambda))$, where $p_{\text{b}}$ is the pressure at the boundary 'b'. In addition extra dependencies may be needed to account for absorbing species, in particular at shorter wavelengths where absorption by ozone ($O_3$) is significant. A more

**Table 6**: Look-up tables and dimensions for reflectance calculations; no trace gas column entries included.

| | | |
|---|---|---|
| $R_0$ | | Reflectance of the black surface |
| | $\lambda$ | For all wavelengths where a cloud fraction must be computed $[1,\ldots,n]$ |
| | $\mu_0$ | For $\mu_0 = [0.0012141231 : 1.0]$, i.e. $\theta_0 = [89.93° : 0°]$, in 42 steps of $2-5°$ |
| | $\mu$ | For $\mu = [0.0012141231 : 1.0]$, i.e. $\theta = [89.93° : 0°]$, in 42 steps of $2-5°$ |
| | $\phi - \phi_0$ | Dependency stores in three Fourier terms |
| | $p_\mathrm{b}$ | Pressure of the bounding surface (cloud or surface) for $p_\mathrm{b} = [1075$ hPa $: 95$ hPa$]$ in 68 steps * |
| $T$ | | Transmittance of the atmosphere |
| | $\lambda$ | For all wavelengths where a cloud fraction must be computed $[1,\ldots,n]$ |
| | $\mu_0$ | For $\mu_0 = [0.0012141231 : 1.0]$, i.e. $\theta_0 = [89.93° : 0°]$, in 42 steps of $2-5°$ |
| | $\mu$ | For $\mu = [0.0012141231 : 1.0]$, i.e. $\theta = [89.93° : 0°]$, in 42 steps of $2-5°$ |
| | $p_\mathrm{b}$ | Pressure of the bounding surface (cloud or surface) for $p_\mathrm{b} = [1075$ hPa $: 95$ hPa$]$ in 68 steps * |
| $s$ | | Spherical albedo of the atmosphere |
| | $\lambda$ | For all wavelengths where a cloud fraction must be computed $[1,\ldots,n]$ |
| | $p_\mathrm{b}$ | Pressure of the bounding surface (cloud or surface) for $p_\mathrm{b} = [1075$ hPa $: 95$ hPa$]$ in 68 steps * |

*) Through a fixed scale height $p_\mathrm{b}$ is linked to the elevation of the bounding surface: $z_\mathrm{b} = [-55$ m $: 16250$ m$]$.

detailed study is needed to determine if $O_3$ is needed for the cloud fraction, but for $NO_2$ we estimate that ignoring $O_3$ absorption leads to an error of $0.01-0.02$ in the cloud fraction. Raman scattering is ignored here.

The terms used in Eq. (25) have the same or less dependencies: $R_0(\lambda) = R_0(\lambda;\theta_0,\theta,\phi-\phi_0;p_\mathrm{b})$, but crucially not on $A_\mathrm{b}(\lambda)$. Further: $T(\lambda) = T(\lambda;\theta_0,\theta;p_\mathrm{b})$ and $s(\lambda) = s(\lambda;p_\mathrm{b})$. The dependency of $R_\mathrm{b}(\lambda)$ and $R_0(\lambda)$ on $\phi-\phi_0$ can be expressed as a Fourier sum, in case of a Rayleigh atmosphere with three terms. All in all this gives a small set of LUTs for $R_0(\lambda)$, $T(\lambda)$ and $s(\lambda)$; see the overview in Table 6. For use in the $NO_2$ retrieval, the set of LUTs has been computed using DAK at $\lambda_{c,\mathrm{NO_2}} = 440$ nm, the wavelength used for the air-mass factor calculations.

From these LUTs we can calculate the reflectance of the cloudy part of the pixel, $R_\mathrm{c}$, using the cloud pressure, $p_\mathrm{c}$, and cloud albedo, $A_\mathrm{c}$, from the cloud product. And the reflectance of the cloud-free part of the pixel, $R_\mathrm{s}$, using the surface pressure, $p_\mathrm{s}$, from meteorology or a fixed scale height and the surface elevation, $z_\mathrm{s}$, and the surface albedo, $A_\mathrm{s}$, from a climatology. Note that either $p_\mathrm{s}$ or $z_\mathrm{s}$ can be used as entry to the LUT: they are "linked" through the fixed scale height.

## D.1 Adjusting albedo to respect physical limits to the cloud fraction

In order to limit the cloud fraction to the range $[0,1]$, the albedo of the boundary is adjusted to ensure radiative closure [20]. From Eq. (23) it is clear that a negative cloud fraction results when $R_\mathrm{s} > R_\mathrm{TOA}$. Rewriting Eq. (25) to set $R_\mathrm{s} = R_\mathrm{TOA}$ provides an adjusted value for $A_\mathrm{s}$:

$$A_\mathrm{s}(\lambda) = \frac{R_\mathrm{TOA}(\lambda) - R_0(\lambda,p_\mathrm{s})}{T(\lambda,p_\mathrm{s}) + s(\lambda,p_\mathrm{s})\,[R_\mathrm{TOA}(\lambda) - R_0(\lambda,p_\mathrm{s})]} \tag{28}$$

In a similar fashion it is clear from Eq. (23) that a cloud fraction larger than 1 results when $R_\mathrm{TOA} > R_\mathrm{c}$. Rewriting Eq. (25) to set $R_\mathrm{c} = R_\mathrm{TOA}$ provides an adjusted value for $A_\mathrm{c}$:

$$A_\mathrm{c}(\lambda) = \frac{R_\mathrm{TOA}(\lambda) - R_0(\lambda,p_\mathrm{c})}{T(\lambda,p_\mathrm{c}) + s(\lambda,p_\mathrm{c})\,[R_\mathrm{TOA}(\lambda) - R_0(\lambda,p_\mathrm{c})]} \tag{29}$$

Note that in the FRESCO cloud retrieval (Sect. 5.3.2) the surface albedo is adjusted ignoring Rayleigh scattering, which simplifies Eq. (28) to $A_\mathrm{s}(\lambda) = R_\mathrm{TOA}(\lambda)$, and Eq. (29) to $A_\mathrm{c}(\lambda) = R_\mathrm{TOA}(\lambda)$.

# E Surface albedo correction using the snow/ice flag

The retrieval process uses the surface albedo in the $NO_2$ fit window, $A_{s,NO_2}$, as one of the input parameters for the air-mass factor and vertical column calculations (OMNO2B / TM5-MP-DOMINO). This surface albedo is taken from a surface albedo climatology. For OMI that is the $440$ nm data from the 5-year average OMI database [45]; see Sect. 5.3.4. Substantial errors are introduced in the retrieval results if the real surface albedo, $A_s$, differs considerably from what is expected, for example in the case of the sudden snowfall or ice cover. Correcting the surface albedo from the climatology, $A_{clim}$, using knowledge of actual snow/ice cover (Sect. 5.3.5) will therefore improve the final data product, in terms of the retrieval itself and for flagging such cases. For the $A_{s,NO_2}$ this correction follows the approach included in the OMI cloud data product OMCLDO2 [43] to adapt the surface albedo in the $O_2$–$O_2$ fit window (i.e. at $471$ nm).

The basis for the correction are the snow/ice flag values, sampled at the ground pixel centre coordinate, as used in the processing (Sect. 5.3.5). Table 7 provides an overview of these flags, where an asterisk marks snow/ice flag values that lead to an adjustment of the surface albedo in case a certain threshold is exceeded.

**Table 7**: Overview of the snow/ice flag values $f_{NISE}$ used by both the NISE and ECMWF snow/ice data sets, where nrs. 252, 253 and 254 do not occur in the ECMWF data. Flag values marked with an asterisk may lead to adjustment of the climatological surface albedo as described in the text.

| $f_{NISE}$ | meaning | remark |
|---|---|---|
| 000 * | snow-free land | |
| 001-100 * | sea ice concentration (percent) | |
| 101 | permanent ice | |
| 103 * | dry & wet snow | |
| 252 | mixed pixels at coastlines | land-ocean or snow/ice-ocean boundaries |
| 253 | suspect ice value | considered to represent an error |
| 254 | error value | |
| 255 * | ocean | |

The rules for modifying the climatological surface albedo $A_{clim}$ are as follows:

- In case of snow-free land or open ocean (flags 0 and 255) adjust $A_s$ if the difference between $A_{clim}$ and default value $A_{def} = 0.04$ is larger than a given threshold $A_{thrs} = 0.1$, where the albedo is decreased only if $A_{clim} > A_{snow} = 0.6$.

  $$\underline{\text{if}} \quad ( \, (A_{clim} - A_{def}) > A_{thrs} \; \& \; A_{clim} > A_{snow} \, ) \quad \underline{\text{then}} \quad A_s = A_{def} \quad \underline{\text{else}} \quad A_s = A_{clim}$$

- In case of dry or wet snow (flag $f_{NISE} = 103$) adjust $A_s$ if the difference between $A_{clim}$ and $A_{snow} = 0.6$ is larger than a given threshold $A_{thrs} = 0.1$.

  $$\underline{\text{if}} \quad ( \, (A_{clim} - A_{snow}) > A_{thrs} \, ) \quad \underline{\text{then}} \quad A_s = A_{snow} \quad \underline{\text{else}} \quad A_s = A_{clim}$$

- In case of a non-zero sea ice concentration (flags $f_{NISE} = 1 - 100$) adjust $A_s$ if the difference between $A_{clim}$ and a default value $A_{def} = 0.4$ is larger than a given threshold $A_{thrs} = 0.1$.

  $$\underline{\text{if}} \quad ( \, |A_{clim} - A_{def}| > A_{thrs} \, ) \quad \underline{\text{then}} \quad A_s = A_{def} \quad \underline{\text{else}} \quad A_s = A_{clim}$$

  where $A_{def}$ depends on the month of the year $M$ and on which hemisphere $H$ the ice is found:

  $$A_{def} = (1.0 - 0.01 \cdot f_{NISE}) * 0.065 + 0.01 \cdot f_{NISE} \cdot A_{ice}(M, H)$$

  with $A_{ice}(M, H)$ following from the LUT used to determine the OMI surface albedo database [45]:

  $$A_{ice}(M, \text{'north'}) = 0.70,\ 0.73,\ 0.76,\ 0.80,\ 0.84,\ 0.78,\ 0.61,\ 0.61,\ 0.62,\ 0.68,\ 0.67,\ 0.71$$

  $$A_{ice}(M, \text{'south'}) = 0.53,\ 0.50,\ 0.44,\ 0.60,\ 0.61,\ 0.64,\ 0.68,\ 0.76,\ 0.80,\ 0.83,\ 0.78,\ 0.66$$

See also the adjustment of the surface albedo in case the cloud fraction from the cloud retrieval might be negative described in Sect. D.1.

# F Data quality value: the `qa_value` flags

To make the use of the OMI data products easier, a so-called `qa_value` (where 'qa' stands for 'quality assurance') is assigned to each ground pixel, following the same approach as for TROPOMI [4]. The `qa_value` is intended to serve as an easy filter of the observations (dividing the dataset in useful versus not useful observations), depending on how the data is used.

The data files have for each ground pixel the so-called `processing_quality_flags` which provides the user information on processing issues, such as errors that were encountered in the processing, as well as a number of warnings, while the `processing_quality_error_flag` only contains the errors (see the OMI $NO_2$ Product User Manual (PUM) [53]), Some of the warnings have been included in the `qa_value`.

The following differentiation of the `qa_value`, $f_{\mathrm{QA}}$, for usage of the $NO_2$ data product has been made:

| | |
|---|---|
| $0.75 \leq f_{\mathrm{QA}} \leq 1.00$ | The ground pixel is recommended for all applications, including column comparisons, visualisation, trends, monthly/seasonal averages. The data is restricted to cloud-free observations (cloud radiance fraction < 0.5), and snow-ice free observations. |
| $0.50 \leq f_{\mathrm{QA}} < 0.75$ | The ground pixel is recommended for use in data assimilation and comparisons against models or vertical profile observations, given that the averaging kernel is used to specify the sensitivity profile in cloudy situations; this includes good quality retrievals over clouds and snow/ice. |
| $0 < f_{\mathrm{QA}} < 0.50$ | The ground pixel is not recommended for use due to serious retrieval issues. |
| $f_{\mathrm{QA}} = 0$ | A processing error occurred so that the ground pixel cannot be used at all, or the solar zenith angle exceeds the limit set in the data assimilation |

The determination of the `qa_value` is done as follows. Starting from the initial value $f_{\mathrm{QA}} = 1$, $f_{\mathrm{QA}}$ is multiplied by the modification factor $f^i_{\mathrm{QA}}$ of each of the criteria $i$ listed in Table 8 that have been met (i.e. if criterion $i$ is not met then $f^i_{\mathrm{QA}} = 1$).

**Table 8**: Overview of the selection criteria for the `qa_value`, $f_{\mathrm{QA}}$, for the OMNO2A collection 4 processing, maintaining the numbering used for TROPOMI (App. E in [4]). The OMNO2B / TM5-MP-DOMINO processing step recalculates the `qa_value`, in some cases with other factors than those listed below, and includes additional flagging, which are not listed here. In this table $f_{\mathrm{AAI}}$ represents the aerosol index $354/388$ nm pair, which is passed on to the $NO_2$ data product file as added flag (Sect. 5.3.7). Warning flags in `processing_quality_flags` that are not used for $f_{\mathrm{QA}}$ are not listed.

| $i$ | criterion | $f^i_{\mathrm{QA}}$ |
|---|---|---|
| 0 | if fatal error encountered according to `processing_quality_flags` | 0.00 |
| 1 | if `south_atlantic_anomaly_warning` set in `processing_quality_flags` | 1.00 |
| 2 | if `sun_glint_warning` set in `processing_quality_flags` over water | 0.70 |
| 3 | if `pixel_level_input_data_missing_warning` set in `processing_quality_flags` | 0.70 |
| 4 | if `interpolation_warning` set in `processing_quality_flags` | 0.90 |
| 5 | if `solar_eclipse` set in `geolocation_flags` | 0.18 |
| OMI | if `row_anomaly_warning` set in `processing_quality_flags` → see Sect. 5.3.1 | 0.10 |
| 9 | if $\Delta N_{\mathrm{s}} > (\Delta N_{\mathrm{s}})^{\mathrm{max}} = 33.0\times10^{-6}$ mol/m$^2$ ($= 2\times10^{15}$ molec/cm$^2$) | 0.15 |
| 16 | $f_{\mathrm{AAI}} > f^{\mathrm{max}}_{\mathrm{AAI}} = 1.0\times10^{10}$ [for future use] | 0.40 |

*) Note that this criterion means that the system switches to the scene mode if there is 1% or more snow/ice.