Royal Netherlands
Meteorological Institute
Ministry of Infrastructure
and Water Management

TROPOMI ATBD of the total and tropospheric NO_2 data products

document number	:	S5P-KNMI-L2-0005-RP
authors	:	J.H.G.M. van Geffen, H.J. Eskes, K.F. Boersma, J.D. Maasakkers and J.P. Veefkind
CI identification	:	CI-7430-ATBD
issue	:	1.4.0
date	:	2019-02-06
status	:	released

Document approval record

	digital signature
Prepared:	Jos van Geffen Digitally signed by Jos van Geffen Date: 2019.01.31 17:25:20 +01'00'
Co-author:	Henk Eskes Digitaal ondertekend door Henk Eskes Datum: 2019.02.01 14:29:01 +01'00'
Checked:	S. Beirle, A. Hilboll, A. Richter, B. Sanders
Approved PM:	Arnoud Apituley Digitally signed by Arnoud Apituley Date: 2019.02.04 09:37:55 +01'00'
Approved PI:	J. Pepijn Veefkind Digitally signed by J. Pepijn Veefkind Date: 2019.02.05 12:28:26 +01'00'

Document change record

issue	date	item	comments
0.0.1	2012-08-17	All	Initial draft version
0.0.2	2012-09-12	All	Major reordering, adding text and references throughout
0.0.3	2012-09-26	6.1–6.2 7.1–7.2	Several small corrections and additions
0.1.0	2012-09-26	—	First official release
0.2.0	2012-11-15	5.1–5.4 6.1–6.2	Large number of updates throughout the text after internal reviewing by the TROPOMI Level-2 Working Group
0.2.1	2013-04-10	5, 6, 7	Document number corrected, reorganisation of Sect. 6; removal of the discussion on an alternative retrieval approach; various minor corrections, updates and additions
0.2.2	2013-06-03	5, 6, 7 8, 9, 10	Major updates and further reorganisation of Sect. 5–7 first versions of Sect. 8–10
0.3.0	2013-06-04	—	Release for Level-2 Working Group review
0.3.1	2013-06-19	5.1–5.3 6.2–6.6 7.3–7.5 8.4, 9.2	Correction and additions resulting from v0.3.0 internal Level-2 Working Group review, and other minor corrections and additions
0.5.0	2013-06-20	—	Release for external review
0.5.1	2013-09-05	5.1–5.2 6.1–6.6 7.2, 7.5	First round of updates taking into account comments and suggestions of the reviewers of version 0.5.0
0.5.2	2013-11-21	5.5 6.2–6.7 7.1, 7.2 8.2, 8.3	Further updates in view of the reviewers comments, and providing more details regarding the fit procedure and the processing chain
0.9.0	2013-11-27	—	Release to ESA and external reviewers
0.9.1	2014-03-17	4.	Instrument overview section now in separate document; section numbering of this document unchanged to maintain change record
0.9.2	2014-04-07	6.2	Some small typographic corrections and updates
0.10.0	2014-04-15	—	Release to ESA
0.10.1	2014-09-19	6.1–6.6 7.1–7.4	Minor updates and corrections in text and tables; old Sects. 6.5 and 6.7 combined; Sects. 6.6 updated; references updated. Descriptions updated and tables of input and output data added. Notation of variables improved or clarified in view of the IODD [RD1].
0.10.2	2014-09-25	6.6	Table added with overview user applications and data the users need
0.11.0	2014-10-02	—	Release to ESA
0.11.1	2015-08-27	6.4.1 6.4.4.1 6.6, 7.4	Update of AMF look-up table entries Update text regarding using cloud fraction from NO_2 spectral window Minor updates in product data set tables, incl. dataset units
0.13.0	2015-09-14	—	Limited release to S5P Validation Team
0.13.1	2015-12-08	6.2, 7.5 7.1	Minor textual corrections Correction of units of input datasets
0.14.0	2015-12-11	—	Release to ESA
0.14.1	2016-01-21	6.4.2 6.4.6	Temperature correction equation updated Description of de-striping implementation clarified

Document change record – *continued*

issue	date	item	comments
1.0.0	2016-02-05	—	Public release
1.0.1	2017-01-31	All	Corrections and additions in response to internal review comments
1.0.2	2017-06-13	A – D	Appendices added
1.0.3	2017-07-13	All	Finalising the corrections and additions
1.1.0	2017-08-16	—	Updated public release for commissioning phase (E1)
1.1.1	2018-02-02	6.2	Formulation of the Ring term to match operational implementation
		6.2	Improved description of the wavelength calibration
		6.2.3	New section with text from main section to improve readability
		6.4.6	Expanded description of a possible de-striping algorithm
		7.4	Updated detailed product overview table to match output product
1.1.2	2018-05-31	6.2	Updated to match the operational implementation
		C, D	Updated to match the operational implementation
		E	Appendix on the `qa_value` definition added
1.1.3	2018-06-08	5 – 9	Many textual corrections and improvements to match the operational implementation and to incorporate results based on the TROPOMI measurements from the commissioning phase
1.2.0	2018-06-11	—	Release for operational phase (E2) — processor version 1.0.0
1.2.1	2018-10-15	—	Version with main text changes v1.1.0 to v1.2.0 marked for reviewers
1.2.2	2018-11-08	—	Figures 1, 2, 4, 6, 18 redone using TROPOMI data
		5.4	Text and Table updated
		5.6	Text updated and Table 2 added
		6.3.1	Text on updates in TM5-MP and Figure 7 added
		6.4.6	Text on de-striping expanded and Figure 12 added
		6.6	Improved treatment of snow/ice cases described and Figure 15 added
		7.4	Tables 11 and 12 updated to include stripe amplitude
		8.2	Text updated and superfluous figure removed
		E	Updated `qa_value` for treatment of snow/ice cases (cf. Sect. 6.6)
		—	Further textual updates
1.3.0	2018-11-08	—	Release — processor version 1.2.0
1.3.1	2019-01-31	6.4.4.1	Text on treatment of cloud data improved; Figure 10 added
		6.6	Product overview Table 5 and product usage Table 6 updated
		7.4	Detailed product overview Table 11 updated
		E	Updated `qa_value` for $M^{\text{trop}}/M^{\text{geo}}$ threshold
		—	Further minor textual updates and corrections
1.4.0	2019-02-06	—	Release — processor version 1.3.0

Contents

List of Tables

List of Figures

1 Introduction to the document

1.1 Identification

This document, identified as S5P-KNMI-L2-0005-RP, is the Algorithm Theoretical Basis Document (ATBD) for the TROPOMI total and tropospheric NO_2 data products. It is part of a series of ATBDs describing the TROPOMI Level-2 data products. The latest public release version of the ATBD is available via [ER1].

This version of the ATBD describes NO_2 processor version 1.3.0, operational as of February / March 2019.

1.2 Purpose and objective

The purpose of this document is to describe the theoretical basis and the implementation of the NO_2 Level-2 algorithm for TROPOMI. The document is maintained during the development phase and the lifetime of the data products. Updates and new versions will be issued in case of changes of the algorithms.

1.3 Document overview

Sections 2 and 3 list the applicable and reference documents and the terms and abbriviations specific for this document; references to peer-reviewed papers and other scientific publications are listed in Appendix F. Section 4 provides a reference to a general description of the TROPOMI instrument, which is common to all ATBDs of the TROPOMI Level-2 data products. Section 5 provides an introduction to the NO_2 data products, their heritage, the set-up of their retrieval, the requirements of the products, and their availability. Section 6 gives an overview of the TROPOMI NO_2 data processing system and important aspects of the various steps in the processing. Section 7 lists some aspects regarding the feasibility of the NO_2 data products, such as the computational effort and the auxiliary information needed for the processing. Section 8 deals with an error analysis of the NO_2 data product. Section 9 gives a brief overview of validation issues and possibilities, such as campaigns and satellite intercomparions. Section 10 formulates some conclusion regarding the NO_2 data products.

1.4 Acknowledgements

The authors would like to thank the following people for useful discussions, information, reviews of earlier versions of this document and other contributions: Andreas Hilboll, Andreas Richter, Bram Sanders, Deborah Stein – Zweers, Dominique Brunner, Huan Yu, Isabelle De Smedt, Jason Williams Johan de Haan, Maarten Sneep, Marina Zara, Mark ter Linden, Michel Van Roozendael, Piet Stammes, Pieter Valks, Ronald van der A, Steffen Beirle, Thomas Wagner.

2 Applicable and reference documents

2.1 Applicable documents

[AD1] Level 2 Processor Development – Statement of Work.
source: ESA/ESTEC; **ref:** S5P-SW-ESA-GS-053; **issue:** 1.1; **date:** 2012-05-21.

[AD2] GMES Sentinel-5 Precursor – S5p System Requirement Document (SRD).
source: ESA/ESTEC; **ref:** S5p-RS-ESA-SY-0002; **issue:** 4.1; **date:** 2011-04-xx.

2.2 Standard documents

There are no standard documents

2.3 Reference documents

[RD1] Sentinel 5 precursor/TROPOMI KNMI and SRON level 2 Input Output Data Definition.
source: KNMI; **ref:** S5P-KNMI-L2-0009-SD; **issue:** 4.0.0; **date:** 2015-11-02.

[RD2] Terms, definitions and abbreviations for TROPOMI L01b data processor.
source: KNMI; **ref:** S5P-KNMI-L01B-0004-LI; **issue:** 3.0.0; **date:** 2013-11-08.

[RD3] Terms and symbols in the TROPOMI Algorithm Team.
source: KNMI; **ref:** S5P-KNMI-L2-0049-MA; **issue:** 2.0.0; **date:** 2016-05-17.

[RD4] TROPOMI Instrument and Performance Overview.
source: KNMI; **ref:** S5P-KNMI-L2-0010-RP; **issue:** 0.10.0; **date:** 2014-03-15.

[RD5] GMES Sentinels 4 and 5 Mission Requirements Document.
source: ESA/ESTEC; **ref:** EOP-SMA/1507/JL-dr; **issue:** 3; **date:** 2011-09-21.

[RD6] QA4ECV - Quality Assurance for Essential Climate Variables.
source: KNMI; **ref:** EU-project 607405, SPA.2013.1.1-03; **date:** November 2012.

[RD7] Science Requirements Document for TROPOMI. Volume I: Mission and Science Objectives and Observational Requirements.
source: KNMI, SRON; **ref:** RS-TROPOMI-KNMI-017; **issue:** 2.0.0; **date:** 2008-10-30.

[RD8] CAPACITY: Operational Atmospheric Chemistry Monitoring Missions – Final report and technical notes of the ESA study.
source: KNMI; **ref:** CAPACITY; **date:** Oct. 2005.

[RD9] CAMELOT: Observation Techniques and Mission Concepts for Atmospheric Chemistry – Final report of the ESA study.
source: KNMI; **ref:** RP-CAM-KNMI-050; **date:** Nov. 2009.

[RD10] TRAQ: Performance Analysis and Requirements Consolidation – Final report of the ESA study.
source: KNMI; **ref:** RP-ONTRAQ-KNMI-051; **date:** Jan. 2010.

[RD11] Sentinel-5P Calibration and Validation Plan for the Operational Phase.
source: ESA; **ref:** ESA-EOPG-CSCOP-PL-0073; **issue:** 1; **date:** 2017-11-6.

[RD12] S5P Mission Performance Centre Nitrogen Dioxide [L2__NO2___] Readme.
source: ESA; **ref:** S5P-MPC-KNMI-PRF-NO2; **issue:** 1.0.1; **date:** 2018-10-31.

[RD13] NO2 PGE Detailed Processing Model.
source: Space Sytems Finland; **ref:** TN-NO2-0200-SSF-001; **issue:** 1.2; **date:** 2010-04-21.

[RD14] Algorithm theoretical basis document for the TROPOMI L01b data processor.
source: KNMI; **ref:** S5P-KNMI-L01B-0009-SD; **issue:** 8.0.0; **date:** 2017-06-01.

[RD15] 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.

[RD16] 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.

[RD17] An improved temperature correction for OMI NO_2 slant column densities from the 405-465 nm fitting window.
source: KNMI; **ref:** TN-OMIE-KNMI-982; **issue:** 1.0; **date:** 2017-01-24.

[RD18] Cloud retrieval algorithm for GOME-2: FRESCO+.
source: EUMETSAT/KNMI; **ref:** EUM/CO/09/4600000655/RM; **issue:** 1.3; **date:** 2010-10-18.

[RD19] Sentinel-5 L2 Prototype Processor – Algorithm Theoretical Baseline Document for Cloud data product.
source: KNMI; **ref:** KNMI-ESA-S5L2PP-ATBD-005; **issue:** 3.0; **date:** 2018-12-15.

[RD20] S5P/TROPOMI ATBD Cloud Products.
source: DLR; **ref:** S5P-DLR-L2-ATBD-400I; **issue:** 1.5.0; **date:** 2018-04-30.

[RD21] Dutch OMI NO_2 (DOMINO) data product v2.0 – see URL `http://www.temis.nl/airpollution/no2.html`.
source: KNMI; **ref:** OMI_NO2_HE5_2.0_2011; **date:** 18 August 2011.

[RD22] Preparing elevation data for Sentinel 5 precursor.
source: KNMI; **ref:** S5P-KNMI-L2-0121-TN; **issue:** 2.0.0; **date:** 2015-09-11.

[RD23] Wavelength calibration in the Sentinel-5 precursor Level 2 data processors.
source: KNMI; **ref:** S5P-KNMI-L2-0126-TN; **issue:** 1.0.0; **date:** 2015-09-11.

[RD24] Determine the effective cloud fraction for a specific wavelength.
source: KNMI; **ref:** S5P-KNMI-L2-0115-TN; **issue:** 1.0.0; **date:** 2013-10-15.

2.4 Electronic references

[ER1] TROPOMI level-2 product ATBD documents. URL `http://www.tropomi.eu/documents/atbd/`.

[ER2] TEMIS website: NO2 data product page. URL `http://www.temis.nl/airpollution/no2.html`.

[ER3] QA4ECV website. URL `http://www.qa4ecv.eu/`.

[ER4] QA4ECV NO2 ECV precursor data. URL `http://www.qa4ecv.eu/ecv/no2-pre`.

[ER5] Copernicus Open Access S5P Data Hub. URL `https://s5phub.copernicus.eu`.

[ER6] TROPOMI website. URL `http://www.tropomi.eu/`.

[ER7] S5P/TROPOMI ISRF. URL `http://www.tropomi.eu/data-products/isrf-dataset`.

[ER8] Vandaele et al. NO2 cross sections. URL `http://spectrolab.aeronomie.be/no2.htm`.

[ER9] TM5 website. URL `http://www.projects.science.uu.nl/tm5/`.

[ER10] Q. L. Kleipool, M. R. Dobber, J. F. De Haan et al.; OMI Surface Reflectance Climatology (2010). URL `http://disc.sci.gsfc.nasa.gov/Aura/data-holdings/OMI/omler_v003.shtml`.

[ER11] A. Nolin, R.L. Armstrong and J. Maslanik; Near Real-Time SSM/I EASE–Grid Daily Global Ice Concentration and Snow Extent. Boulder, CO, USA: National Snow and Ice Data Center. Digital media (2005). Updated daily; URL `http://nsidc.org/data/docs/daac/nise1_nise.gd.html`.

[ER12] EUMETSAT Ocean & Sea Ice Satellite Application Facility. Updated daily; URL `http://osisaf.met.no/`.

[ER13] TROPOMI level-2 product PUM documents. URL `http://www.tropomi.eu/documents/pum/`.

[ER14] J.J. Danielson and D.B. Gesch; Global Multi-resolution Terrain Elevation Data 2010 (GMTED2010) (2011). URL `http://topotools.cr.usgs.gov/gmted_viewer/`.

[ER15] L. G. Tilstra, O. N. E. Tuinder, P. Wang et al.; Surface reflectivity climatologies from UV to NIR determined from Earth observations by GOME-2 and SCIAMACHY (2017). URL http://temis.nl/surface/gome2_ler.html.

[ER16] TROPOMI Mission Performance Centre. URL http://www.tropomi.eu/data-products/mission-performance-centre.

[ER17] S5P Mission Performance Centre VDAF website. URL http://mpc-vdaf.tropomi.eu/.

[ER18] S5P First Public Release Validation Workshop June 25-26, 2018. URL https://nikal.eventsair.com/QuickEventWebsitePortal/sentinel-5p-first-product-release-workshop/sentinel-5p.

3 Terms, definitions and abbreviated terms

Terms, definitions and abbreviated terms that are used in development program for the TROPOMI L0-1b data processor are described in [RD2]. Terms, definitions and abbreviated terms that are used in development program for the TROPOMI L2 data processors are described in [RD3]. Terms, definitions and abbreviated terms that are specific for this document can be found below.

3.1 Terms and definitions

The most important symbols related to the data product described in this document – some of which are not in [RD3] – are the following; see also the data product overview list in Table 11.

M	total air-mass factor
M_{cld}	cloudy air-mass factor
M_{clr}	clear-sky air-mass factor
M^{trop}	tropospheric air-mass factor
M^{strat}	stratospheric air-mass factor
N_{s}	total slant column density
$N_{\mathrm{s}}^{\mathrm{trop}}$	tropospheric slant column density
$N_{\mathrm{s}}^{\mathrm{strat}}$	stratospheric slant column density
N_{v}	total vertical column density
$N_{\mathrm{v}}^{\mathrm{trop}}$	tropospheric vertical column density
$N_{\mathrm{v}}^{\mathrm{strat}}$	stratospheric vertical column density
$N_{\mathrm{v}}^{\mathrm{sum}}$	sum of tropospheric and stratospheric vertical column density

3.2 Acronyms and abbreviations

AAI	Absorbing Aerosol Index
ACE	Atmospheric Chemistry Experiment
AMF	Air-mass factor
CTM	Chemistry Transport Model
DAK	Doubling-Adding KNMI
DOAS	Differential Optical Absorption Spectroscopy
DOMINO	Dutch OMI NO_2 data products of KNMI for OMI
ECMWF	European Centre for Medium-Range Weather Forecast
ENVISAT	Environmental Satellite
EOS-Aura	Earth Observing System (Chemistry & Climate Mission)
ERBS	Earth Radiation Budget Satellite
ERS	European Remote Sensing satellite
FRESCO	Fast Retrieval Scheme for Clouds from the Oxygen A band
GOME	Global Ozone Monitoring Experiment
HALOE	Halogen Occultation Experiment
IDAF-L2	Instrument Data Analysis Facility, Level 2 (at KNMI)
IPA	Independent pixel approximation
ISRF	Instrument Spectral Response Function (*aka* slit funtion)
LER	ambertian equivalent reflectivity
LUT	Look-up table
MACC	Monitoring Atmospheric Composition and Climate
MAX-DOAS	Multi-axis DOAS
MERIS	Medium Resolution Imaging Spectrometer

MetOp	Meteorological Operational Satellite
MPC	S5P Mission Performance Centre
NISE	Near-real-time Ice and Snow Extent
NRT	near-real time (i.e. processing within 3 hours of measurement)
OMI	Ozone Monitoring Instrument
OMNO2A	OMI NO2 slant column data product (at NASA)
OSIRIS	Optical Spectrograph and Infrared Imager System
OSISAF	Ocean & Sea Ice Satellite Application Facility
PANDORA	not an acronym; direct Sun UV-visible spectrometer
PDGS	Sentinel-5Precursor Payload Data Ground Segment (at DLR)
POAM	Polar Ozone and Aerosol Measurements
QA4ECV	European "Quality Assurance for Essential Climate Variables" project
S5P	Sentinel-5 Precursor (satellite carrying TROPOMI)
SAGE	Stratospheric Gas and Aerosol Experiment
SAOZ	Systeme d'Analyse par Observations Zenithales instrument
SCIAMACHY	Scanning Imaging Absorption Spectrometer for Atmospheric Cartography
SME	Solar Mesosohere Explorer
SNR	Signal-to-Noise Ratio
SPOT	Système Pour l'Observation la Terre
STREAM	STRatospheric Estimation Algorithm from Mainz
TM4, TM5	Data assimilation / chemistry transport model (version 4 or 5)
TM4NO2A	NO_2 data products of KNMI for GOME, SCIAMACHY and GOME-2
TROPOMI	Tropospheric Monitoring Instrument
UARS	Upper Atmosphere Research Satellite
VDAF	Validation Facility of the MPC

4 TROPOMI instrument description

A description of the TROPOMI instrument and performance, referred to from all ATBDs, can be found in [RD4]. See also the overview paper of Veefkind et al. [2012].

5 Introduction to the TROPOMI NO_2 data products

5.1 Nitrogen dioxide in troposphere and stratosphere

Nitrogen dioxide (NO_2) and nitrogen oxide (NO) – together usually referred to as nitrogen oxides (NO_x = NO + NO_2) – are important trace gases in the Earth's atmosphere, present in both the troposphere and the stratosphere. They enter the atmosphere as a result of anthropogenic activities (notably fossil fuel combustion and biomass burning) and natural processes (such as microbiological processes in soils, wildfires and lightning). Approximately 95% of the NO_x emissions is in the form of NO. During daytime, i.e. in the presence of sunlight, a photochemical cycle involving ozone (O_3) converts NO into NO_2 (and vice versa) on a timescale of minutes, so that NO_2 is a robust measure for concentrations of nitrogen oxides (Solomon [1999], Jacob [1999]).

In the troposphere NO_2 plays a key role in air quality issues, as it directly affects human health [World Health Organisation, 2003]. In addition nitrogen oxides are essential precursors for the formation of ozone in the troposphere (e.g. Sillman et al. [1990]) and they influence concentrations of OH and thereby (shorten) the lifetime of methane (CH_4) (e.g. Fuglestvedt et al. [1999]). Although NO_2 is a minor greenhouse gas in itself, the indirect effects of NO_2 on global climate change are probably larger, with a presumed net cooling effect mostly driven by a growth in aerosol concentrations through nitrate formation from nitrogen oxides and enhanced levels of oxidants (e.g. Shindell et al. [2009]). Deposition of nitrogen is of great importance for eutrification [Dentener et al., 2006], the response of the ecosystem to the addition of substances such as nitrates and phosphates – negative environmental effects include the depletion of oxygen in the water, which induces reductions in fish and other animal populations.

For typical levels of OH the lifetime of NO_x in the lower troposphere is less than a day. For Riyadh, for example, Beirle et al. [2011] find a lifetime of about 4.0 ± 0.4 hours, while at higher latitudes (e.g. Moscow) the lifetime can be considerably longer, up to 8 hour in winter, because of a slower photochemistry in that season. For Switzerland Schaub et al. [2007] report lifetimes of 3.6 ± 0.8 hours in summer and $13.1 \pm (3.8)$ hours in winter. With lifetimes in the troposphere of only a few hours, the NO_2 will remain relatively close to its source, making the NO_x sources well detectable from space. As an example, Fig. 1 shows distinct hotspots of NO_2 pollution over the highly industrialised and urbanised regions of London, Rotterdam and the Ruhr area in the monthly average tropospheric NO_2 for April 2018 over Europe derived from TROPOMI data.

In the stratosphere NO_2 is involved in some photochemical reactions with ozone and thus affects the ozone layer (Crutzen et al. [1970]; Seinfeld and Pandis [2006]). The origin of NO_2 in the stratosphere is mainly from oxidation of N_2O in the middle stratosphere, which leads to NO_x, which in turn acts as a catalyst for ozone

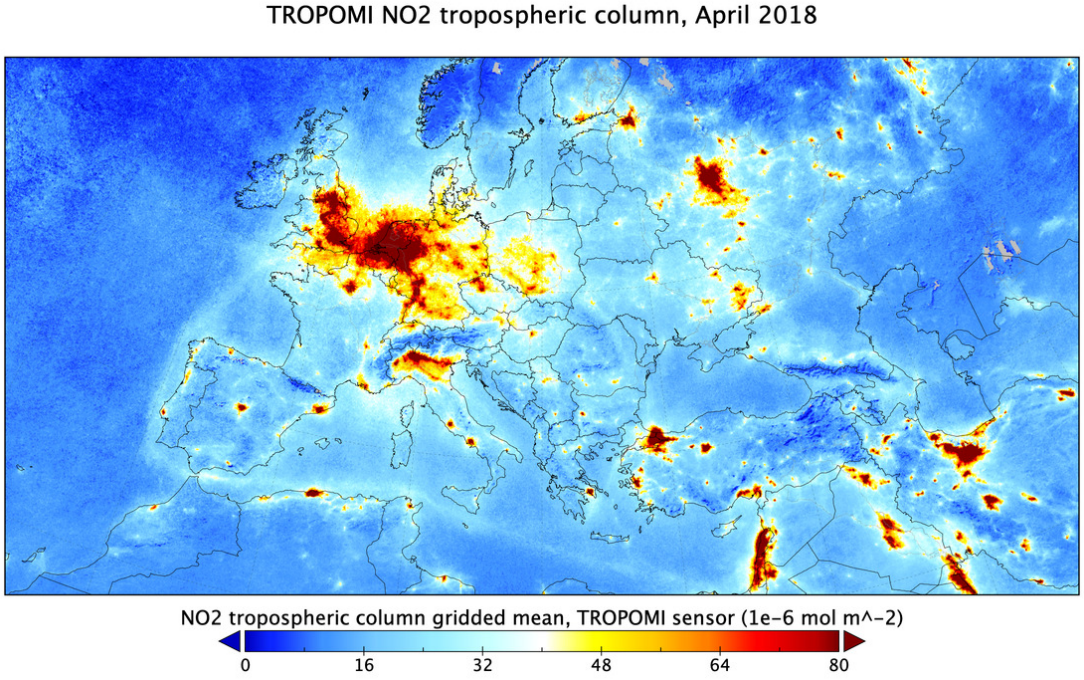

Figure 1: Monthly average distribution of tropospheric NO_2 columns for April 2018 over Europe based on TROPOMI data, derived with processor version 1.2.0.

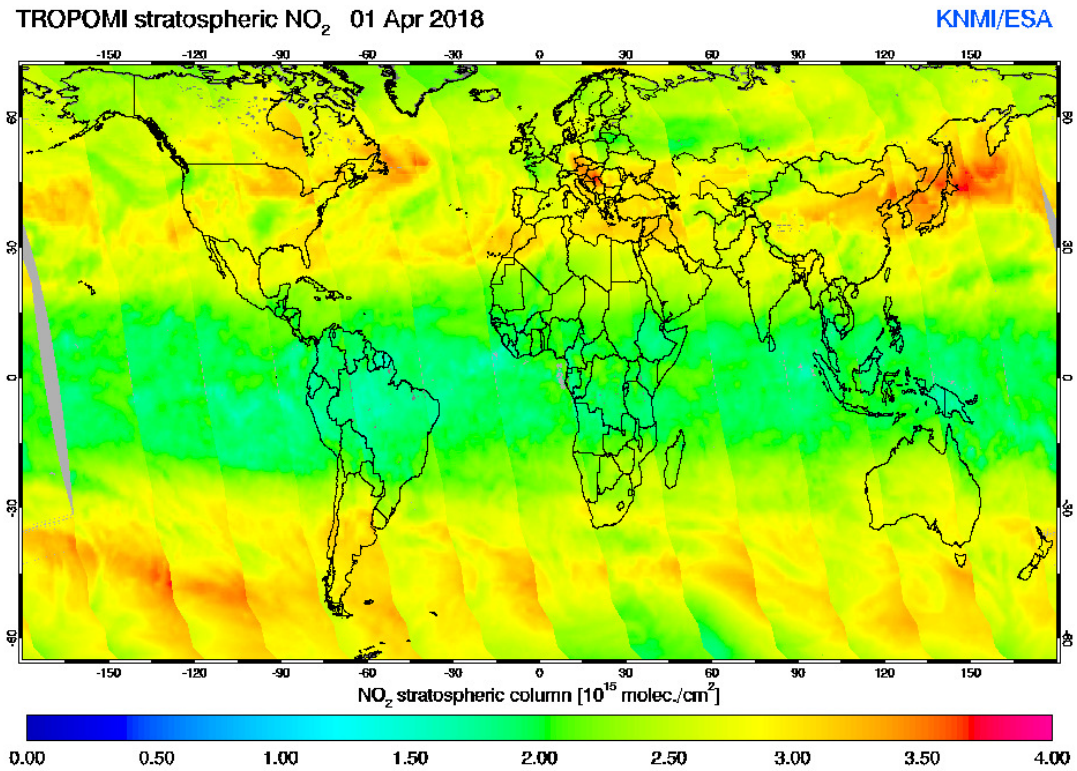

Figure 2: Distribution of stratospheric NO_2 on 1 April 2018 along the individual TROPOMI orbits, derived with processor version 1.2.0. The image shows that atmospheric dynamics creates variability in the stratospheric columns, mainly at mid-latitudes. Furthermore we can see the effect of the increase of NO_2 in the stratosphere during daytime leading to jumps from one orbit to the next. Note that the colour scale range is different from range in Fig. 1.

destruction (Crutzen et al. [1970]; Hendrick et al. [2012]). But NO_x can also suppress ozone depletion by converting reactive chlorine and hydrogen compounds into unreactive reservoir species (such as $ClONO_2$ and HNO_3; Murphy et al. [1993]).

Fig. 2 shows, as an example, the stratospheric NO_2 distribution derived from TROPOMI measurements on 1 April 2018. In a study into the record ozone loss, triggered by enhanced NO_x levels, in the exceptionally strong Arctic polar vortex in Spring 2011, Adams et al. [2013] showed the usefulness of such data when investigating the anomalous dynamics and chemistry in the stratosphere. With its higher spatial resolution and signal-to-noise ratio, TROPOMI will clearly be well-suited to help understand the stratospheric NO_2 content and its implications for the ozone distribution.

From observed trends in N_2O emissions one would expect a trend in stratospheric NO_2 with potential implications for persistent ozone depletion well into the 21st century [Ravishankara et al., 2009]. There have been some reports of such trends in stratospheric NO_2, for instance from New Zealand [Liley et al., 2000] and northern Russia [Gruzdev and Elokhov, 2009]. On the other hand, Hendrick et al. [2012] report that changes in the NO_x partitioning in favour of NO may well conceal the effect of trends in N_2O. TROPOMI will continue the important record of stratospheric NO_2 observations that started with GOME in 1995, and improve the detectability of trends.

Over unpolluted regions most NO_2 is located in the stratosphere (typically more than 90%). For polluted regions 50–90% of the NO_2 is located in the troposphere, depending on the degree of pollution. Over polluted regions, most of the tropospheric NO_2 is found in the planetary boundary layer, as has been shown among others in campaigns using measurements made from aeroplanes, such as INTEX (e.g. Hains et al. [2010]). In areas with strong convection, enhanced NO_2 concentrations are observed at higher altitudes due to production of NO_x by lightning (e.g. Ott et al. [2010]).

The important role of NO_2 in both troposphere and stratosphere implies that it is not only important to know the total column density of NO_2, but rather the tropospheric NO_2 and stratospheric NO_2 concentrations separately. A proper separation between the two is therefore important, in particular for areas with low pollution,

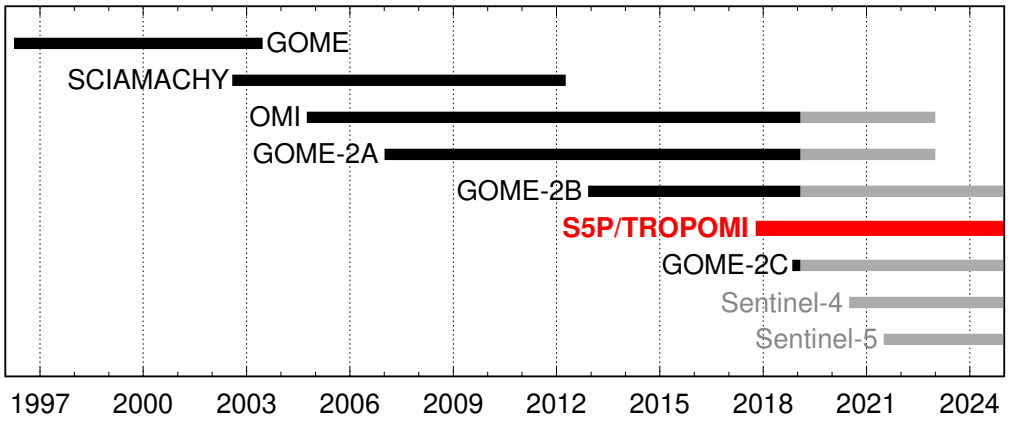

Figure 3: Overview of the European UV/Vis polar orbiting and geostationary backscatter satellite instruments capable of retrieving tropospheric and stratospheric NO_2 column data since the launch of GOME aboard ERS-2.

where the stratospheric concentration forms a significant part of the total column.

5.2 NO_2 satellite retrieval heritage

Tropospheric concentrations of NO_2 are monitored all over the world by a variety of remote sensing instruments – ground-based, in-situ (balloon, aircraft) or satellite-based – each with its own specific advantages, and to some extent still under development.

Stratospheric NO_2 has been measured by a number of satellite instruments since the 1980s, such as the spectrometer aboard SME (1981-1989; Mount et al. [1984]), SAGE-II/III (ERBS/Meteor-3M, 1984-2005; Chu and McCornick [1986]), HALOE (UARS, 1991-2005; Gordley et al. [1996]), POAM (SPOT-3, 1993-1996; Randall et al. [1998]), SCIAMACHY (ENVISAT, 2002–2012; Bovensmann et al. [1999], Sierk et al. [2006]), OSIRIS (Odin, 2001–present; Llewellyn et al. [2004], Adams et al. [2016]), and ACE (SCISAT-1, 2003–present; Bernath et al. [2005]).

Over the past 22 years tropospheric NO_2 has been measured from UV/Vis backscatter satellite instruments such as GOME (ERS-2, 1995–2011; Burrows et al. [1999]), SCIAMACHY (ENVISAT, 2002–2012; Bovensmann et al. [1999]), OMI (EOS-Aura, 2004–present; Levelt et al. [2006]), the GOME-2 instruments [Munro et al., 2006] aboard MetOp-A (2007–present), MetOp-B (2012–present) and MetOp-C (2019–present), and the OMPS instrument on the Suomi NPP platform [Yang et al., 2014]. TROPOMI (see [RD4]; Veefkind et al. [2012]) will extend the records of these observations, and in turn will be followed up by several forthcoming instruments including Sentinel 5 and the geostationary platforms GEMS [Bak et al., 2013], TEMPO [Zoogman et al., 2017] and Sentinel 4 [Ingmann et al., 2012], [RD5]. Fig. 3 shows the timelines of the NO_2 data records of these instruments. Note that TROPOMI, OMI, the GOME-2 instruments and Sentinel-5 provide (near-)global coverage in one day, and that Sentinel-4 is a geostationary instrument.

For the UV/Vis backscatter instruments that observe NO_2 down into the troposphere, KNMI has operated – in close collaboration with BIRA-IASB, NASA and DLR – a real-time data processing system, the results of which are freely available via the TEMIS website [ER2]. The data has been used for a variety of studies in areas like validation (see e.g. Boersma et al. [2009], Hains et al. [2010], Lamsal et al. [2010]), trends (see e.g. Van der A et al. [2008], Stavrakou et al. [2008], Dirksen et al. [2011], Castellanos and Boersma [2012], DeRuyter et al. [2012]), and NO_x emission and lifetime estimates (see e.g. Lin et al. [2010], Beirle et al. [2011], Mijling and Van der A [2012], Wang et al. [2012]).

The DOMINO approach for OMI (or similar approach called TM4NO2A for GOME, SCIAMACHY and GOME-2) is based on a DOAS retrieval, a pre-calculated air-mass factor (AMF) look-up table and a data assimilation / chemistry transport model for the separation of the stratospheric and tropospheric contributions to the NO_2 column (see Sect. 6 for details). The differences between the processing systems for the different instruments are small and related to instrument issues, such as available spectral coverage and wavelength calibration, other absorbing trace gases fitted along, and details of the cloud cover data retrieval.

The European Quality Assurance for Essential Climate Variables (QA4ECV) project ([RD6], [ER3], Boersma et al. [2018]) has led to a homogeneous reprocessing dataset of NO_2 for the sensors GOME, SCIAMACHY,

OMI and GOME-2A. This project has investigated and improved all the individual steps/modules in the NO_2 retrieval. The new NO_2 datasets are available via the QA4ECV project website at [ER4]. This new release replaces the DOMINO-v2 OMI NO_2 dataset and TM4NO2A datasets for the other sensors.

The TROPOMI NO_2 processor includes many of the developments from the QA4ECV project, including improvements in the TM5-MP-domino chemistry modelling-retrieval-assimilation approach, DOAS optimisations and air-mass factor lookup table. On top of that, several further improvements have been implemented, notably in the TM5-MP-domino system and the output data file.

5.3 Separating stratospheric and tropospheric NO_2 with a data assimilation system

The NO_2 processing system starts with a DOAS retrieval step that determines the NO_2 slant column density, which represents the total amount of NO_2 along the line of sight, i.e. from Sun via Earth's atmosphere to the satellite. To determine the tropospheric NO_2 slant column density, the stratospheric NO_2 slant column density is subtracted from the total slant column provided by a DOAS retrieval performed on a spectrum of backscattered light measured by a satellite instrument, after which the two slant sub-columns are converted to the tropospheric and stratospheric vertical NO_2 column, respectively.

Several approaches to estimate the stratospheric NO_2 amount have been introduced in the past. The DOMINO / TM4NO2A approach uses information from a chemistry transport model by way of data assimilation to simulate the instantaneous stratospheric NO_2 distribution and to force consistency between the stratospheric NO_2 column and the satellite measurement [Boersma et al., 2004]. Other methods applied elsewhere include the following.

a) The wave analysis method uses subsets of satellite measurements over unpolluted areas to remove known areas of pollution, i.e. areas with potentially large amounts of tropospheric NO_2, from a 24-hour composite of the satellite measured NO_2 and expands the remainder with a planetary wave analysis across the whole stratosphere, followed where necessary by a second step to mask pollution events (e.g. Bucsela et al. [2006]). This approach has been used between 2004 and 2012 for the OMI NO_2 Standard Product (SP) of NASA/KNMI.

b) The reference sector method method uses a north-to-south region over the Pacific Ocean that is assumed to be free of tropospheric NO_2, as there are no (surface) sources of NO_2, so that all NO_2 measured is assumed to be in the stratosphere (e.g. Richter and Burrows [2002], Martin et al. [2002]). This stratospheric NO_2 is then assumed to be valid in latitudinal bands for all longitudes. In some implementaions this method is extended with a spatial filtering to include other relatively clean areas across the world (e.g. Bucsela et al. [2006], Valks et al. [2011]).

c) Image processing techniques assume that the stratospheric NO_2 shows only smooth and low-amplitude latitudinal and longitudinal variations (e.g. Leue et al. [2001], Wenig et al. [2003]). This approach will probably miss the finer details in the stratospheric NO_2 distribution (as is the case for methods *a* and *b* above). The next version of NASA's OMI NO_2 SP will use a similar approach [Bucsela et al., 2013].

d) Independent stratospheric NO_2 data, such as collocated limb measurements (e.g. Beirle et al. [2010], Hilboll et al. [2013b]) or data taken from a chemistry transport model (e.g. Hilboll et al. [2013a]), can be subtracted from the total (slant) column measurements to find the tropospheric NO_2 concentrations. Unfortunately limb collocated stratospheric measurements are not available for satellite retrievals from the GOME(-2), OMI, and TROPOMI sensors. Nevertheless this approach is potentially very useful for comparison and validation studies. Possible cross-calibration problems between the stratospheric and the total measurements would complicate the approach.

e) STRatospheric Estimation Algorithm from Mainz (STREAM), [Beirle et al., 2016]. The STREAM approach is based on the total column measurements over clean, remote regions as well as over clouded scenes where the tropospheric column is effectively shielded. STREAM is a flexible and robust interpolation algorithm and does not require input from chemical transport models. It was developed as a verification algorithm for the upcoming satellite instrument TROPOMI, as a complement to the operational stratospheric correction based on data assimilation. STREAM was successfully applied to the UV/vis satellite instruments GOME 1/2, SCIAMACHY, and OMI. It overcomes some of the artifacts of previous algorithms, as it is capable of reproducing some of the gradients of stratospheric NO_2, e.g., related to the polar vortex, and reduces interpolation errors over continents.

These ways of treating the stratospheric NO_2 field may not be accurate enough to capture the variability of the stratospheric NO_2 in latitudinal and longitudinal direction, as well as in time. At the same time it is not certain whether these methods do actually separate stratospheric NO_2: some of the NO_2 interpreted as

Table 1: NO_2 data product requirements for the TROPOMI NO_2 data products, where accuraries are split in the systematic and random components. The numbers are taken from [RD11] (see also [RD12]).

NO_2 data product	Vertical resolution	Bias	Random
Stratospheric NO_2	Stratospheric column	$< 10\%$	0.5×10^{15} molec/cm^2
Tropospheric NO_2	Tropospheric column	$25-50\%$	0.7×10^{15} molec/cm^2

"stratospheric" may be in the (higher) troposphere.

Also the assimilation approach suffers from these uncertainties, but in a different way since actual meteorological fields are used to model the dynamical and chemical variability of NO_x in the stratosphere. The assimilation analyses the retrieved total slant column with a strong forcing to the observations over clean regions (regions with small tropospheric column amounts). The data assimilation ensures that the model simulations of the stratospheric NO_2 column agrees closely with the satellite measurements. The modelled stratospheric NO_2 (slant column) amount is subtracted from the full column observation to derive the tropospheric column.

The use of a data assimilation system to provide stratospheric NO_2 concentrations has been shown to provide realistic results, as indicated by validation studies. For example, Hendrick et al. [2012] found very good agreement between satellite retrievals using data assimilation to estimate the stratospheric NO_2 column (GOME, SCIAMACHY and GOME-2) and ground-based measurements at the station of Jungfraujoch.

The advantages of the use of stratospheric transport-chemistry modelling in combination with data assimilation are:

- Data assimilation provides a realistic error estimate of the stratospheric NO_2 column [Dirksen et al., 2011].
- The height of the tropopause, obtained from the meteorological data, provides an accurate point of separation of the stratospheric from the tropospheric NO_2 column.
- The result of the data assimilation is a comprehensive understanding of 3-D NO_2 distributions that covers the whole world, taking into account the spatial and temporal variability of the NO_2 profiles.

5.4 NO_2 data product requirements

S5P/TROPOMI mission requirements have been discussed in several documents, including the GMES Sentinels-4, -5 and -5Precursor Mission Requirements Document [RD5] and the Science Requirements Document for TROPOMI [RD7] provide the requirements for the TROPOMI mission, aboard the Sentinel-5Precursor (S5P) mission. These requirements are based on the findings of the CAPACITY [RD8], CAMELOT [RD9] and TRAQ [RD10] studies. For the TROPOMI NO_2 column data products the set of requirements which are used as baseline in the routine validation work are the NO_2 data product requirement listed in Table 1; these are given in the "Sentinel-5P Calibration and Validation Plan for the Operational Phase" document [RD11] and also given in the NO_2 Product ReadMe File [RD12].

The uncertainties stated in Table 1 include retrieval errors as well as instrument errors. Over polluted areas retrieval errors will dominate the uncertainties; these relate to the presence of clouds and aerosols and to the surface albedo. Over rural areas, with low NO_2 concentrations, errors in tropospheric NO_2 are mostly driven by random noise related to the instrument's Signal-to-Noise Ratio (SNR), to estimates of the stratospheric NO_2 column, and to uncertainties in the NO_2 profile.

5.5 NO_2 retrieval for TROPOMI

The TROPOMI retrieval of total and tropospheric NO_2 is based on the DOMINO system (see Sect. 6.1), thus extending the long-term record of NO_2 data, produced using a reliable, well-established and well-described processing system (see Boersma et al. [2004], Boersma et al. [2007] and Boersma et al. [2011]). In particular, the inclusion of many of the retrieval developments of the QA4ECV project ([RD6], [ER3]) in the TROPOMI NO_2 retrieval will ensure a good continuity from the QA4ECV OMI and GOME-2 NO_2 records to TROPOMI. For the OMI NO_2 retrieval a number of improvements are related to spectral fitting [Van Geffen et al., 2015] and to the chemistry modelling, stratosphere-troposphere separation and the air-mass factor [Maasakkers et al., 2013].

The TROPOMI NO_2 processing is performed in two locations (Sect. 6.5). The first step of the processing, the DOAS retrieval, takes place at the official Level-2 processing site at DLR, the Sentinel-5Precursor Payload Data Ground Segment (PDGS). The data assimilation system providing the information necessary to split

Table 2: Overview of periods of operation of the operational NO_2 processor versions in the near-real time (NRT) and the off-line (OFFL) data streams, starting from the

Processor version	Data stream	In operation from		In operation until	
		orbit	date	orbit	date
01.00.00	NRT	03745	2018-07-04	03946	2018-07-18
	OFFL	03661	2018-06-28	03847	2018-07-11
01.01.00	NRT	03947	2018-07-18	05333	2018-10-24
	OFFL	03848	2018-07-11	05235	2018-10-17
01.02.00	NRT	05336	2018-10-24	[*current version*]	
	OFFL	05236	2018-10-17	[*current version*]	

the total slant column into its stratospheric and tropospheric contributions and providing NO_2 profile data, is running in the Instrument Data Analysis Facility (IDAF) at KNMI. The final step, the conversion of the slant column into the tropospheric and stratospheric NO_2 columns, will take place (i) in the NRT processing mode at the PDGS (at DLR), and (ii) in off-line or reprocessing mode in the IDAF at KNMI after which the results are pushed to the PDGS.

In order to comply with the SI unit definitions, the TROPOMI NO_2 data product file (described further in Sect. 6.6) provides the trace gas columns in mol/m^2, rather then in the commonly used unit molec/cm^2. For convenience sake, most of the text and figures of this document will remain in the latter unit; only the tables listing the input (Sect. 7.1) and output (Sect. 7.4) dataset use the SI based units.

5.6 NO_2 data product availability and access

The NO_2 processing has started directly after "first light", providing data for initial checks and validations. The data is planned to be made available in June 2018, about nine months after launch. TROPOMI NO_2 data is processed in near-real time (NRT), and the data is available within 3 hours after measurement; this data stream uses a forecast of the TM5-MP data (see next sections). A few days later, the data is processed in off-line mode (OFFL), using TM5-MP analysis data. Table 2 provides an overview of the operation TROPOMI NO_2 data processor versions. All data (near-real time, offline and reprocessed) is freely accessible via the Copernicus Open Access Data Hub [ER5]. See the TROPOMI website [ER6] for more information on data dissemination.

6 Algorithm description

6.1 Overview of the NO_2 retrieval algorithm

The TROPOMI NO_2 processing system is based on the DOMINO and QA4ECV processing systems, with improvements related to specific TROPOMI aspects and new scientific insights. The basis for the processing at KNMI is a retrieval-assimilation-modelling system which uses the 3-dimensional global TM5 chemistry transport model as an essential element. The retrieval consists of a three-step procedure, performed on each measured Level-1b spectrum:

1. the retrieval of a total NO_2 slant column density (N_s) from the Level-1b radiance and irradiance spectra measured by TROPOMI using a DOAS (Differential Optical Absorption Spectroscopy) method,
2. the separation of the N_s into a stratospheric (N_s^{strat}) and a tropospheric (N_s^{trop}) part on the basis of information coming from a data assimilation system, and
3. the conversion of the tropospheric slant column density (N_s^{trop}) into a tropospheric vertical column density (N_v^{trop}) and of the stratospheric slant column density (N_s^{strat}) into a stratospheric vertical column density (N_v^{strat}), by applying an appropriate AMF based on a look-up table of altitude-dependent AMFs and actual, daily information on the vertical distribution of NO_2 from the TM5-MP model on a $1^\circ \times 1^\circ$ grid. The altitude-dependent AMF depends on the satellite geometry, terrain height, cloud fraction and height and surface albedo.

These steps are described in detail in the sections below.

6.2 Spectral fitting

The baseline method to determine NO_2 total slant columns is DOAS (see Platt [1994], Platt and Stutz [2008]). The DOAS fitting function for TROPOMI follows the current non-linear fitting approach for OMI (Boersma et al. [2011], Van Geffen et al. [2015], [RD13]).

The reflectance spectrum $R_{meas}(\lambda)$ observed by the satellite instrument is the ratio of the radiance at the top of the atmosphere, $I(\lambda)$, and the extraterrestrial solar irradiance, $E_0(\lambda)$ (where I also depends on the viewing geometry, but those arguments are left out for brevity):

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

where E_0 and I are recorded at the same detector row and given on the same wavelength grid (see below), and $\mu_0 = \cos(\theta_0)$ is the cosine of the solar zenith angle. The E_0 is measured once a day, when TROPOMI crosses the terminator along a given orbit, and this daily irradiance spectrum is used for the radiance measurements of subsequent orbits until a new E_0 is available. Since in Eq. (1) the factor π/μ_0 becomes very large at high solar zenith angle, the algorithm internally actually uses the sun-normalised radiance $I(\lambda)/E_0(\lambda)$; note that this ratio is sometimes also called reflectance.

In space-borne DOAS, R_{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 $\theta_0 < 88^\circ$, so that there is no potential danger from the division by μ_0 in Eq. (1). The FRESCO+ cloud data product (Sect. 6.4.4.1) uses this θ_0 cut-off as well.

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

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

with N_λ the number of wavelengths in the fit window and $\Delta R_{meas}(\lambda_i)$ the precision of the measurements, which depends on the precision of the radiance and irradiance measurements as given in the Level-1b product:

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

The magnitude of χ^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:

$$R_{\mathrm{RMS}} = \sqrt{\frac{1}{N_\lambda}\sum_{i=1}^{N_\lambda}\Big(R_{\mathrm{meas}}(\lambda_i) - R_{\mathrm{mod}}(\lambda_i)\Big)^2} \tag{4}$$

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

The baseline fitting function for TROPOMI follows the approach for OMI and reads as follows:

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

with $\sigma_k(\lambda)$ the absolute cross section and $N_{\mathrm{s},k}$ the slant column amount of molecule $k = 1,\ldots,N_k$ taken into account in the fit (NO_2, O_3, etc.), C_{ring} the Ring fitting coefficient and $I_{\mathrm{ring}}(\lambda)/E_0(\lambda)$ the sun-normalised synthetic Ring spectrum. 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 $+C_{\mathrm{ring}}\cdot\sigma_{\mathrm{ring}}(\lambda)$ term) to the reflectance spectrum.

In the modelled spectrum of Eq. (5) a polynomial of order N_p with coefficients a_m is defined:

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

This polynomial 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 of the polynomial are scaled to the range $[-1:+1]$ over the fit window.

Fig. 4 shows an example of a reflectance spectrum observed by TROPOMI on 4 July 2018 during orbit 03747, along with the modelled spectrum obtained from the DOAS fit using Eq. (5), with cross-sections for NO_2, ozone (O_3) and water vapour (H_2O_{vap}), a Ring spectrum and a 5th order polynomial as fitting parameters. The (almost cloud-free) ground pixels lies over the industrial area of Rotterdam (scanline 2012, row 323, $\theta_0 = 27.82^\circ$, $\theta = 31.47^\circ$). Fit results: $N_{\mathrm{s,NO_2}} = 4.54\times10^{-4}$ mol/m^2, $N_{\mathrm{v,NO_2}} = 2.66\times10^{-4}$ mol/m^2, $N^{\mathrm{trop}}_{\mathrm{v,NO_2}} = 2.52\times10^{04}$ mol/m^2, RMS $= 1.59\times10^{-4}$ sr^{-1}. The residual (shown in the bottom panel of Fig. 4) is of the order of 10^{-4}, corresponding to an unexplained differential optial depth of that magnitude.

In principle we can expect that the TROPOMI Level-1b solar irradiance spectra are well wavelength calibrated [RD14]. Nevertheless, to be on the safe side, we perform a wavelength calibration of the irradiance specifically for the NO_2 fit window. The earth radiance spectra, however, only have been assigned wavelengths, and thus certainly need to be calibrated before they are usable. Both types of spectra are calibrated prior to the DOAS fit, and using the same wavelength calibration approach. Using the subscripts 'nom' and 'cal' to denote nominal (i.e. from the Level-1b data product) and calibrated wavelengths, respectively, the calibrated (ir)radiance to be used in Eq. (1) is then given by:

$$\begin{aligned} E_0(\lambda^{E0}_{\mathrm{cal}}) &= E_0(\lambda^{E0}_{\mathrm{nom}} + w^{E0}_s) \\ I(\lambda_{\mathrm{cal}}) &= I(\lambda_{\mathrm{nom}} + w_s + w_q(\lambda_{\mathrm{nom}} - \lambda_0)) \end{aligned} \tag{7}$$

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 λ_0. Since in view of numerical stability, the wavelengths are scaled to the range $[-1:+1]$ over the fit window (similar to the wavelengths of the DOAS polynomial in Eq. (6), computationally $\lambda_0 = 0$. Note that for the irradiance calibration we only conside a shift. Each wavelength calibration of Eq. (7) comes with its own χ^2_w as a goodness-of-fit. Once TROPOMI Level-1b spectra are available we will investigate whether including w_q is necessary; initially we fix $w_q = 0$.

The wavelength calibration of Eq. (7) 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 case the $E_0(\lambda)$ is converted to the radiance wavelength grid by way of a high-sampling interpolation, taking

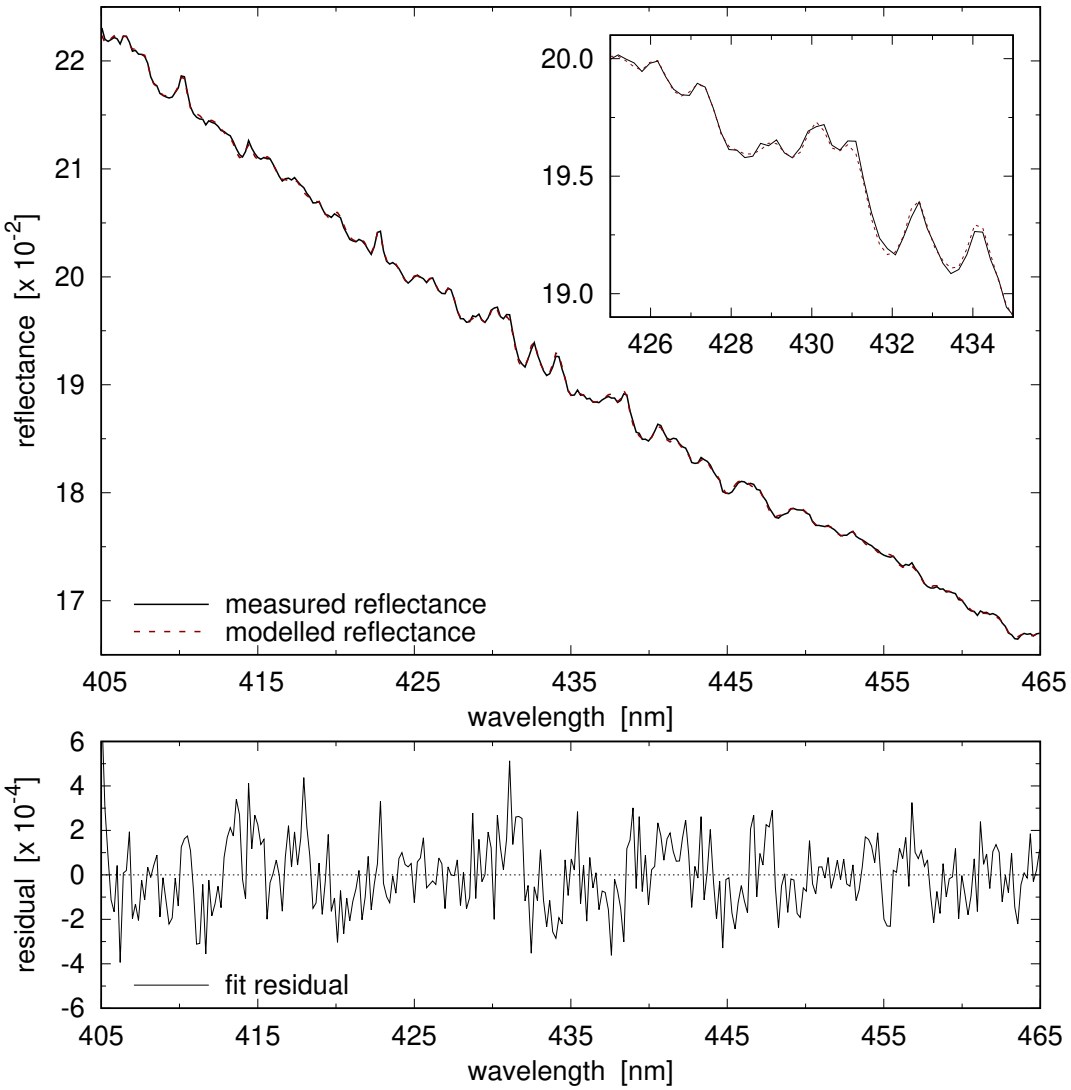

Figure 4: The *top panel* shows an example of a reflectance spectrum (black solid line) obtained by TROPOMI on 4 July 2018 during orbit 03747, the spectrum modelled in the DOAS fit procedure (dashed red line); the inset shows an enlargement of a 10 nm wide part of the fit window. The *bottom panel* shows the residual of the DOAS fit, i.e. the measured minus the modelled reflectance spectrum; note that the vertical scale is a factor of 100 smaller than the scale in the top panel.

advantage of the fact that we have additional information from a high-resolution solar reference spectrum $E_{\text{ref}}(\lambda)$. Details of the wavelength calibration and the high-sampling interpolation implemented for TROPOMI are given in Appendices A and B, respectively.

Slant column densities $N_{\text{s},k}$, the Ring coefficent C_{ring}, and the polynomial coefficients a_m are obtained from a minimisation of the χ^2 of Eq. (2), i.e. the differences between the observed and modelled reflectances. In the initial TROPOMI NO_2 DOAS, we implented a version of the OMI NO_2 DOAS processor, called OMNO2A, which uses a non-linear least squares fitting based on routines available in the SLATEC mathematical library [Vandevender and Haskell, 1982]. During the commissioning phase, however, we discovered that this implementation suffered from some issues (i.e. the χ^2 and/or the slant column error estimates were scaled incorrectly) that could not be solved due to inflexibility of the OMNO2A code. To solve this issue, we chose to use the optimal estimation (OE) routine already available in the processor, since it was implemented for the wavelength calibration; see Appendix A. For the χ^2 minimisation 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.

A number of fitting diagnostics is provided by the fitting procedure. Estimated slant column and fitting

Table 3: Main settings of the operational DOAS retrieval of NO_2 for TROPOMI, and for the current and previous satellite instruments in the operational processing of KNMI, which converts the NO_2 slant column data products into tropospheric and stratospheric vertical column data. For OMI the settings used for the QA4ECV v1.1 processing ([RD6], [ER3]) are given; these are an extention of the settings used for the DOMINO v2 processing (see Sect. 6.2.2 for a brief discussion).

	TROPOMI	OMI (QA4ECV v1.1)	GOME-2 (TM4NO2A v2.3)	SCIAMACHY (TM4NO2A v2.3)
wavelength range [nm]	$405-465$	$405-465$	$425-450$	$426.5-451.5$
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	O_3, H_2O_{vap}, O_2–O_2
pseudo-absorbers	Ring	Ring	Ring	Ring
fitting method	non-linear	non-linear	linear	linear
degree of polynomial	5	5	3	2
polarisation correction	no	no	no	yes
slant column processing	PDGS (@ DLR)	NASA / KNMI	DLR / BIRA-IASB	BIRA-IASB
references	—	[Boersma et al., 2011] [Van Geffen et al., 2015]	[Valks et al., 2011] [Liu et al., 2018]	[Van Roozendael et al., 2006]

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 normalised χ^2, where χ^2 is normalised by $(N_\lambda - D)$, with N_λ 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.

Table 3 provides an overview of the operational DOAS fit settings planned for TROPOMI 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. 6.2.1) 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 3 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. For TROPOMI, for example, the spectral sampling is about 0.2 nm and the FWHM is about 0.55 nm [RD4].

6.2.1 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 will exploit the best available sources. Experience with OMI 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. Van Geffen et al. [2015] (cf. Sect. 6.2.2) show that including also absorption in liquid water and by the O_2–O_2 collision complex improves the fit, hence these will be included for TROPOMI.

High-resolution laboratory measured absorption cross sections are convolved with the TROPOMI slit function (or: instrument spectral response function, ISRF; available via [ER7]), and sampled at a resolution of 0.01 nm to create the necessary cross sections. 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 spline interpolation. The final set of reference spectra (see also [RD15] and Fig. 5) is:

- trace gas cross sections $\sigma_k(\lambda)$ in Eq. (5):
  - NO_2 from Vandaele et al. [1998] at 220 K; see [ER8]
  - O_3 from Gorshelev et al. [2014] and Serdyuchenko et al. [2014] at 243 K
  - Water vapour (H_2O_{vap}) based on HITRAN 2012 data (see Van Geffen et al. [2015] and Sect. 4.1 of [RD15])
  - O_2–O_2 from Thalman and Volkamer [2013] at 293 K

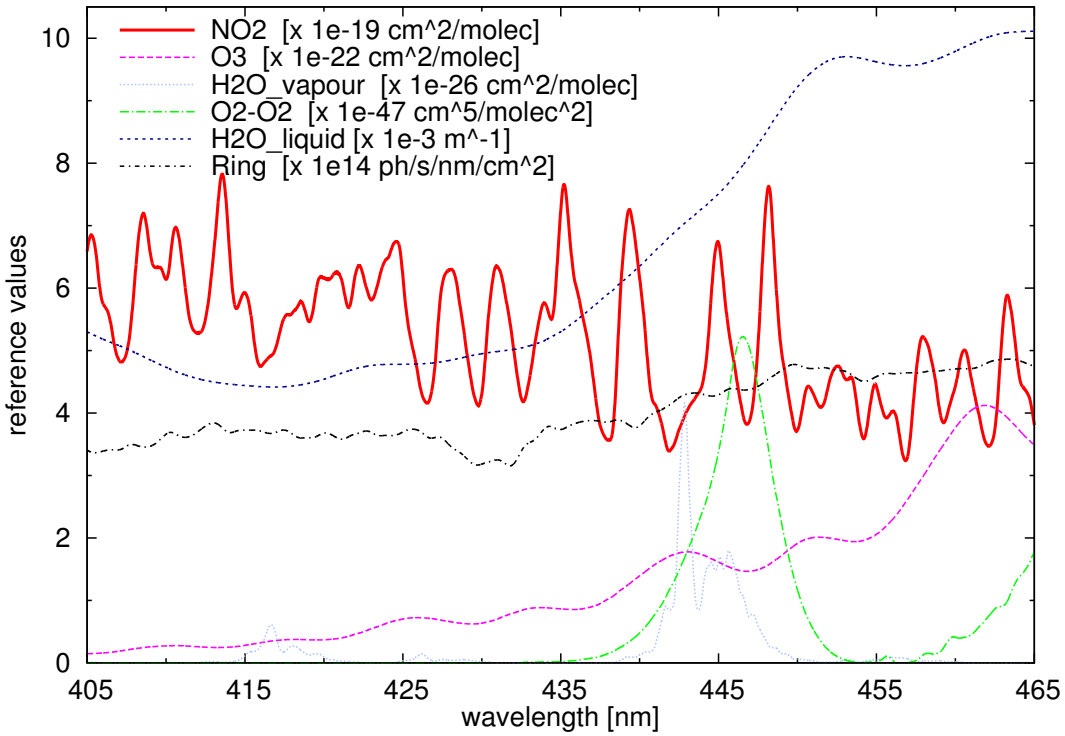

Figure 5: Absorption cross sections $\sigma_k(\lambda)$ for NO_2, O_3, water vapour, O_2–O_2 and liquid water, as well as the Ring spectrum $I_{\text{ring}}(\lambda)$, the pseudo-absorber which accounts for the Ring effect, in Eq. (5) for the $405-465$ nm wavelength range used in the OMI data processor. The reference spectra have been multiplied by the factors given in the plot legend to make the spectral signatures visible in one plot.

  - Liquid water (H_2O_{liq}) from Pope and Frey [1997], resampled at 0.01 nm with a cubic spline interpolation
- an effective Ring spectrum $I_{\text{ring}}(\lambda)$ from Chance and Spurr [1997] (see Van Geffen et al. [2015] and Sect. 4.2 of [RD15])

The inclusion of absorption by soil (as discussed by e.g. Richter et al. [2011]; Merlaud et al. [2012]) is not considered for TROPOMI, as its potential absorption signal lies well above 465 nm, the upper limit of the fit window considered for the retrieval. Also currently not being considered for inclusion in the fit is the vibrational Raman scattering in clear ocean waters (e.g. Vasilkov et al. [2002], Vountas et al. [2003]), as its potential effect on the fit is currently poorly understood; cf. Sect. 6.2.3.

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.

Note that the amplitide of the differential cross section features has a significant temperature dependence which is important to account for. The resulting NO_2 slant column are corrected for deviations from 220 K at later retrieval steps, as described in Sect. 6.4.2.

6.2.2 DOAS fit details for OMI and TROPOMI

Comparisons of OMI NO_2 data from the DOMINO v2 processing system to independent data from other instruments have shown that OMI slant NO_2 columns are higher than columns derived from GOME-2 and SCIAMACHY (as first stated by N. Krotkov at the OMI Science Meeting in Sept. 2012), as well as columns derived from groundbased measurements. Due to the separation between stratospheric and tropospheric NO_2, which proceeds in the same way for the three satellite instruments, the high bias in the NO_2 slant columns is propagated to the stratospheric column [Belmonte Rivas et al., 2014].

Van Geffen et al. [2015] show that improving the OMI wavelength calibration of the Level-1b spectra in the OMNO2A data processing of the NO_2 slant columns used by DOMINO v2 reduces both the total NO_2 slant

column values and the RMS of the DOAS fit. Van Geffen et al. [2015] further show that including both O_2–O_2 and H_2O_{liq} (discussed by e.g. Richter et al. [2011], Lerot et al. [2010]) in the fit improves the OMI NO_2 fit results and ensures that fit coefficients for O_3 and O_2–O_2 have realistic values. Criteria for establishing what are the best settings for the fit can be summarised as follows: (a) a low error on the NO_2 slant column, (b) a low RMS error value, (c) inclusion of secondary trace gases that clearly improve the fit, e.g. by removing specific features in the fit residual, (d) physically realistic values for the slant column values of these secondary trace gases.

The improvements described by Van Geffen et al. [2015] for OMNO2A have been used for the processing of OMI NO_2 data within the QA4ECV-project ([RD6], [ER3]). These improvements are incorporated in the TROPOMI NO_2 slant column processing.

The OMNO2A radiance wavelength calibration determines only a spectral shift. For the TROPOMI NO_2 processing a spectral shift and stretch/squeeze are included in the NO_2 slant column fit, as described by Eq. (7), though initially only a spectral shift will be fitted; the stretch/squeeze option may be turned on at a later stage, if that appears to be necessary.

6.2.3 Some notes regarding other DOAS implementations

Many implementations of DOAS deploy a linearised version of Eq. (5), including the Ring effect as a pseudo-absorber, giving the equation in terms of optical depth rather than in terms of reflectances:

$$ln[R_{\text{mod}}(\lambda)] = P^*(\lambda) - \sum_{k=1}^{N_k} \sigma_k(\lambda) \cdot N_{\text{s},k} - \sigma_{\text{ring}}(\lambda) \cdot C^*_{\text{ring}} \tag{8}$$

where the Ring coefficient C^*_{ring} and the polynomial $P^*(\lambda)$ are essentially different from C_{ring} and $P(\lambda)$ in Eq. (5). In this approach the Ring cross section $\sigma_{\text{ring}}(\lambda)$ is constructed from the Ring radiance spectrum $I_{\text{ring}}(\lambda)$ divided by a reference solar spectrum minus a low-order polynomial (so that $\sigma_{\text{ring}}(\lambda)$ varies around zero). This linearisation allows then for the use of a linear least squares fitting routine, which is computationally faster than a non-linear solver needed when using Eq. (5).

We feel, however, that the Ring effect is physically described better by the non-linear approach of Eq. (5) and we will therefore use that in the NO_2 data processing for TROPOMI. Apart from dropping the physical description of the Ring effect, a disadvantage of the linearised approach is that error propagation is no longer straightforward, because taking the logarithm of the observed spectra implies that the error no longer has a Gaussian distribution.

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 know, but it is thought to be related to instrumental issues (e.g. incomplete removal of straylight or dark current in Level-1b spectra) and/or atmospheric issued (e.g. incomplete removal of Ring spectrum structures, vibrational Raman scattering in clear ocean waters); see, for example, Platt and Stutz [2008], [Richter et al., 2011], [RD16], [Lampel et al., 2015].

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

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

with fit parameters c_m and S_{off} a suitable scaling factor; in most applications $N_{\text{off}} = 0$ or 1 if an intensity offset is included. The NO_2 slant column retrieval algorithm of OMI (OMNO2A) is not able to handle such an intensity offset correction. The possibility of an intensity offset correction has been implemented in the TROPOMI NO_2 slant column processor, but this option is initially turned off as (i) it has not been 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 NO_2 retrievals. Note that TROPOMI is expected to show a very low level of dark current.

6.3 Separation of stratospheric and tropospheric NO_2

The baseline method for the TROPOMI NO_2 algorithm to separate the stratospheric and tropospheric contribution to the NO_2 total slant columns is by data assimilation of slant columns in the TM5-MP CTM [Williams et al., 2017]. KNMI has considerable experience with this method, and in the absence of collocated independent (e.g. limb) information on stratospheric NO_2, we consider this to be the most viable method to distinguish

stratospheric from tropospheric NO_2. This approach explicitly accounts for chemistry and dynamics in the stratosphere.

The central idea of the data assimilation is to regularly update a CTM simulation of the three-dimensional, coupled troposphere-stratosphere NO_2 distribution with available measurement data in such a way that the CTM simulation of the stratospheric NO_2 column achieves close agreement with the TROPOMI slant columns over areas known to have little or no tropospheric NO_2. The assimilation effectively relies on slant columns observed over regions where the model predicts the NO_2 column to be dominated by stratospheric NO_2 (e.g. over the remote oceans). For those regions and times, the modeled slant column, i.e. the inner product of the observation operator **H** and the simulated vertical distribution $\vec{x}$, is effectively forced to the observed state. For regions and times where the model predicts large tropospheric contributions, the slant column is not as good a proxy for stratospheric NO_2, and the analysis adjustment is only very small. Because total reactive nitrogen (NO_y) is a well-conserved quantity in the stratosphere, with relatively small source and sink contributions, the information from the observations can be stored in the model over long time periods. The stratosperic wind will transport the stratospheric analysis results from the oceans and remote regionas to the polluted areas.

The assimilation scheme is based on the Kalman filter technique, with a prescribed parameterisation of the horizontal correlations between forecast errors. The assimilation time step in the model is 30 minutes. A full orbit of TROPOMI observations is analysed simultaneously by the Kalman filter. This is done in order to avoid discontinuities in the analysis that may occur at the location where the orbit is divided. The mid-time of the orbit is used to determine the model time step of the analysis. The analysed profile field $\vec{x}_a$ includes NO_2 in both troposphere and stratosphere, and is calculated from the forecast $\vec{x}_f$ and the 2-D field of so-called superobservations $\vec{y}$ by:

$$\vec{x}_a = \vec{x}_f + \mathbf{P}\mathbf{H}^T(\mathbf{H}\mathbf{P}\mathbf{H}^T + \mathbf{R})^{-1}(\vec{y} - \vec{y}_f) \tag{10}$$

with matrix **H** the observation operator, **P** the forecast error covariance matrix, and **R** the combined observation and representativeness error covariance (Eskes et al. [2003]; Dirksen et al. [2011]). The term $\mathbf{P}\mathrm{H}^T(\mathbf{H}\mathrm{P}\mathrm{H}^T + \mathbf{R})^{-1}$ determines the weight given to the observations depending on the uncertainty of the observation versus the uncertainty of the model forecast. The departure $(\vec{y} - \vec{y}_f)$ is the difference between observed and forecasted model column (observation minus forecast). The ratio between analysis and forcast following from the Kalman equation is also applied to species that are chemically closely related to NO_2 in the stratosphere, i.e. NO, NO_3, N_2O_5 and HNO_4 [Dirksen et al., 2011]. A simplified modelling of the observation error is introduced [Dirksen et al., 2011], with a fixed small error attached to the stratospheric part of the slant column (where the AMF is well known) and a large error attached to the tropospheric contribution to the slant column (where the AMF is uncertain). These uncertainties (0.2 and 6.0×10^{15} molec/cm^2 for the stratosphere/troposphere respectively) are fixed and have been optimised with sensitivity runs in such a way that the impact of major source regions on the analysis is minimal, and the forcing over clean regions is strong. These coefficients have recently been fine-tuned for OMI, and the same coefficients are used for TROPOMI.

The observation operator **H** is a combination of a horizontal interpolation and the application of the averaging kernel [Eskes and Boersma, 2003], a N_l-element vector that contains the sensitivity of TROPOMI to NO_2 in each model layer. The kernel itself is proportional to the height-dependent air-mass factor and also contains a correction factor for the dependence of the NO_2 cross section on temperature, using temperature profiles from the ECMWF model. The scalar product of the observation operator vector and the model NO_2 profile at the location of the individual TROPOMI observations yields the slant column that would be observed by TROPOMI. The average of all TROPOMI observations (and model equivalents) with center coordinates inside a $1^\circ \times 1^\circ$ (longitude $\times$ latitude) model grid cell is treated as a single measurement, a so-called superobservation [Boersma et al., 2016]. To further speed up the analysis step by about a factor of 8, we assimilate only half of the superobservations, those grid cells with $i + j =$ even, where i and j are the longitude and latitude indices of the TM5-MP grid (the "checkerboard" approach). $\vec{y}_f$ is the model forecast of the superobservations, given by $\mathbf{H}\vec{x}_f$. In order to reduce the computational effort, the Kalman filter is applied for these superobservations, rather than for all individual observations. An additional advantage of this approach is that because of averaging the TROPOMI observations into superobservations, much of the noise in the TROPOMI observations cancels out, allowing a tight observational constraint on the model state.

We use the TM5-MP CTM (Williams et al. [2017]; see also Huijnen et al. [2010a]; Huijnen et al. [2010b]; [ER9]) for the assimilation of TROPOMI NO_2 slant columns. This is a major improvement over the DOMINO v2 data assimilation systems operated at KNMI for GOME, SCIAMACHY, OMI, and GOME-2, which use an older version of the TM CTM (TM4; e.g. Dentener et al. [2003]). The main advantage of the transition to TM5-MP is the better spatial resolution ($1^\circ \times 1^\circ$), updated information on (NO_x) emissions, and an improved description of relevant physical (photolysis rate constants) and chemical (reaction rate constants) processes in that model

Williams et al. [2017]. The assimilation system operates at a resolution of $1^\circ \times 1^\circ$ (longitude $\times$ latitude), with N_l sigma pressure layers up to 0.1 hPa in the vertical direction. TM5-MP uses forecast and analysed 3-hourly meteorological fields from the European Centre for Medium Range Weather Forecast (ECMWF) operational model. These fields include global distributions of wind, temperature, surface pressure, humidity, (liquid and ice) water content, precipitation and surface parameters.

Once the TROPOMI slant columns have been assimilated, the integral from the layer above the tropopause to the upper TM5-MP layer provides the stratospheric slant column that can be isolated from the total slant column, giving the tropospheric slant column (cf. Sect. 6.4):

$$N_{\mathrm{s}}^{\mathrm{trop}} = N_{\mathrm{s}} - N_{\mathrm{s}}^{\mathrm{strat}} \tag{11}$$

For the tropopause definition the WMO-1985 temperature gradient criterion is followed, but other definitions would not lead to significantly different results (e.g. Bucsela et al. [2013]). NO_x has a C-shape profile and the air around the tropopause has only a small contribution to the total column. Compared to the QA4ECV processing of OMI a new routine has been introduced for TROPOMI which reduces the fine-scale jumps in tropopause level.

The data assimilation system provides the following information, necessary for the subsequent processing in the calculation of the AMF (see Sect. 6.4) needed for the conversion of the tropospheric slant column to the tropospheric vertical column and the final NO_2 data product (see Sect. 6.6):

- the stratospheric slant and vertical columns: $N_{\mathrm{s}}^{\mathrm{strat}}$ and $N_{\mathrm{v}}^{\mathrm{strat}}$
- an estimate of the error on the stratospheric vertical column: $\Delta N_{\mathrm{v}}^{\mathrm{strat}}$
- the NO_2 profile: $n_{l,\mathrm{NO_2}}$, with l the index for the layer number $1, 2, \ldots, N_l$ – this is represented by $\vec{x}_f$ in Eq. (10)
- the temperature profile at the layers: T_l^{TM5}, for $l = 1, 2, \ldots, N_l$
- the pressure level coefficients: A_l^{TM5}, B_l^{TM5}, for $l = 0, 1, \ldots, N_l$
- the index of the pressure level of the tropopause: $l_{\mathrm{tp}}^{\mathrm{TM5}}$
- the surface elevation and pressure: $z_{\mathrm{s}}^{\mathrm{TM5}}$ and $p_{\mathrm{s}}^{\mathrm{TM5}}$, at the $1^\circ \times 1^\circ$ model resolution

Note that the model divides the atmosphere in N_l layers. The pressure level coefficients determine the pressure at the $N_l + 1$ levels separating the layers: $p_l = A_l^{\mathrm{TM5}} + B_l^{\mathrm{TM5}} \cdot p_{\mathrm{s}}$, for $l = 0, 1, \ldots, N_l$, with p_{s} the surface pressure for the given TROPOMI ground pixel. The pressure for the layer l, for which the concentration (volume mixing ratio) $n_{l,\mathrm{NO_2}}$ and the temperature T_l^{TM5} are given, is then midway between the level pressures p_{l-1}^{TM5} and p_l^{TM5}. The layer with index $l_{\mathrm{tp}}^{\mathrm{TM5}}$ contains the tropopause. Note that $n_{l,\mathrm{NO_2}}$ is represented by $\vec{x}_f$ in Eq. (10).

6.3.1 Stratospheric chemistry in the TM5-MP model

TM5-MP is primarily a tropospheric chemistry model [Williams et al., 2017]. NO_x-O_x-HO_y chemical processes are implemented according to the Carbon Bond 05 (CB05) chemistry scheme, which includes non-methane hydrocarbons to account for loss by reactions with OH [Williams et al., 2017]. Because the chemistry version of TM5-MP does not simulate N_2O, the actual source of NO_x to the stratosphere, NO_x is derived from simulated HNO_3 concentrations, which follow climatological HNO_3:O_3 ratios observed by ODIN [Maasakkers et al., 2013] and the multi-sensor reanalysis of stratospheric O_3 columns [Van der A et al., 2015] with climatological ozone profile shapes. In this way the model partly compensates for the biases that occur due to the missing N_2O source globally, and the missing reactions involving halogens which are important in the polar vortex. Processes included in the TM5-MP tracer evolution are advection, convection, diffusion, photolysis and deposition. Rapid changes in stratospheric NO_2 due to e.g. sudden stratospheric warmings or changes in the vortex edge location are largely accounted for through the use of the ECMWF analysis. Solar proton events are not included in the model, but the related biases are largely removed by the assimilation. NO_x emissions are based on the RETRO-REAS emission inventories for 2006. For more details, the reader is referred to Dirksen et al. [2011].

For QA4ECV and for TROPOMI the original implementation of the stratospheric climatologies has been improved by a better interpolation to the TM5-MP vertical levels and by adding an extra nudging to NOx observations from HALOE in the upper stratosphere which is not well constrained by HNO_3 [Grooß and Russell, 2005].

The data assimilation provides a regular update of the TM5-MP simulation, with a time step of 30 minutes, of the NO_2 distribution in the atmosphere on the basis of available observations: if NO_2 slant columns are available with a measurement time within 15 minutes of the model time, the model field is updated, i.e. the forecast TM5-MP state is adjusted towards the observations. The stratospheric error estimate is based on "observation minus forecast" statistics (over relatively unpolluted areas) in the assimilation. Our experience

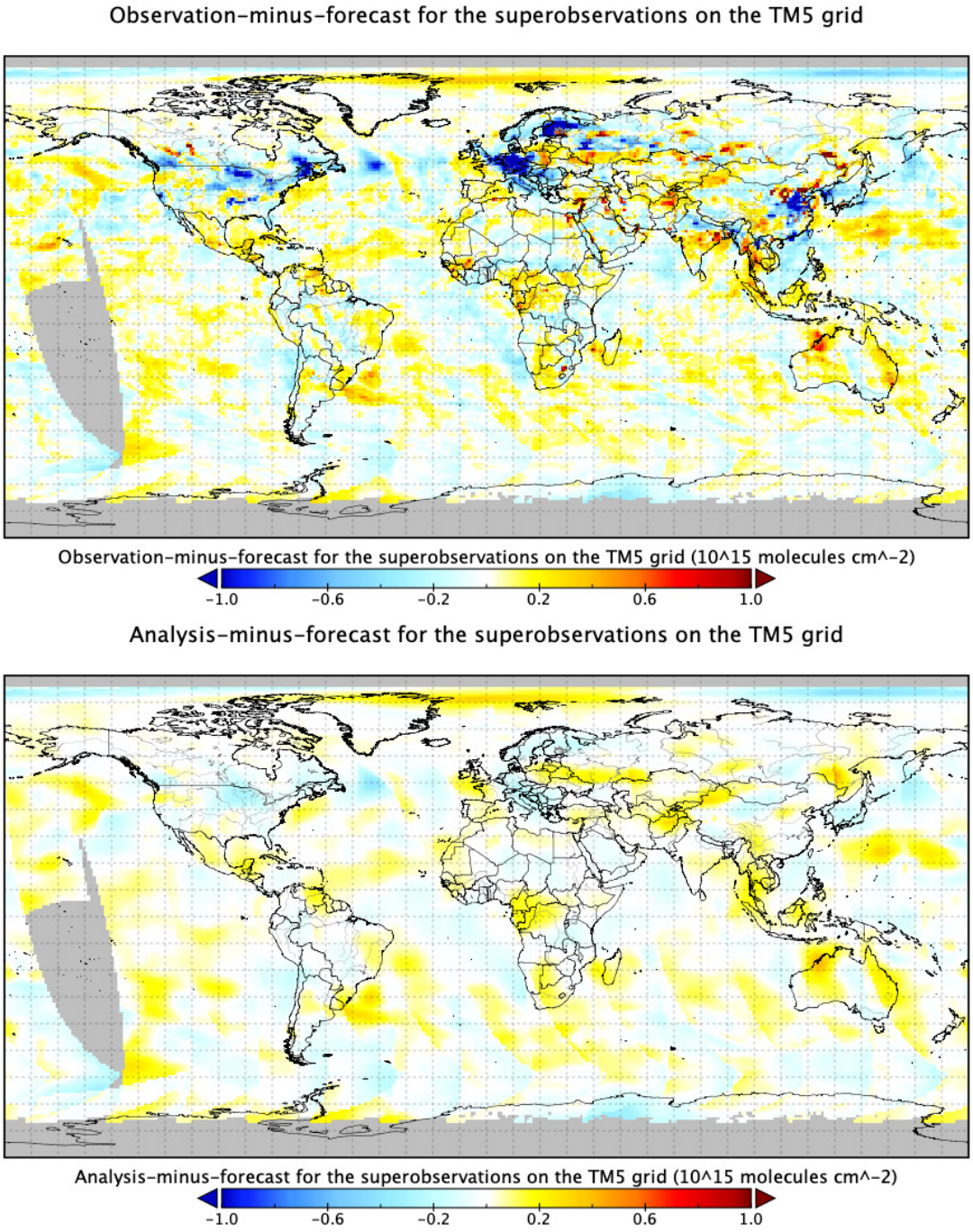

Figure 6: Observation-minus-forecast (OmF, *top panel*) and analysis-minus-forecast (AmF, *bottom panel*) differences in NO_2 slant columns divided by the geometric AMF, for 1 April 2018. The observations are averaged to "superobservations" on the $1^\circ \times 1^\circ$ grid of the TM5-MP model. The model forecast is simulating the observations using the kernels and air-mass factors. The OmF demonstrates clear differences (dark-blue and bright-red spots) between the model forcast and TROPOMI concerning the fine-scale distribution of tropospheric pollution. The AmF plot shows that the assimilation hardly changes the tropospheric distribution, but efficiently updates the stratospheric fields over the more unpolluted regions like the oceans.

with NO_2 data assimilation using GOME, SCIAMACHY, OMI, and GOME-2 in TM has shown that the model chemistry responds smoothly to the updates forced by the satellite measurements.

Fig. 6 provides an example of the "observation minus forecast" (O–F) and the model forcing ("analysis minus forcast", A–F) for TROPOMI data of 1 April 2018. The difference between the two panels of Fig. 6 illustrates the effect of the assimilation: considerable O–F differences, resulting mostly from (anthropogenic)

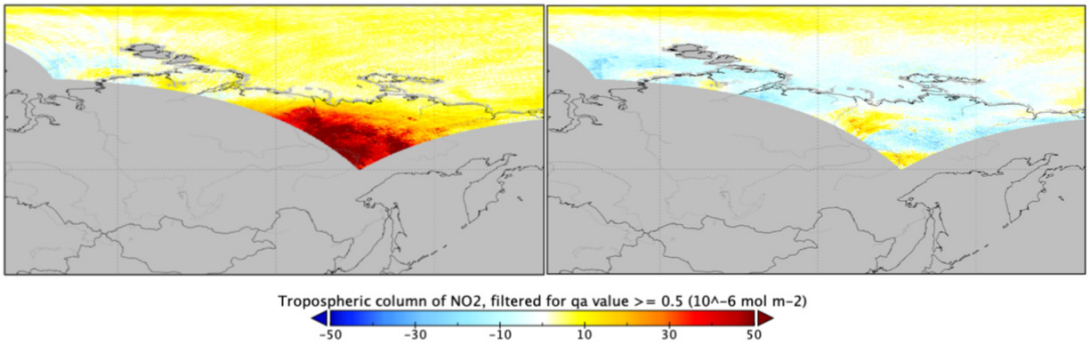

Figure 7: The NO_2 tropospheric column retrievals for the descending part of TROPOMI orbit 3623, 25 June 2018, 19h UTC, over Siberia. Version 1.0.2 is shown on the left, and version 1.2.0 on the right. Prominent unrealistic positive biases are observed in v1.0.2 and v1.1.0 for the highest solar zenith angles on the left side of the orbit, while v1.2.0 has much more realistic values close to zero with a tendency towards a weak negative bias.

tropospheric NO_2 sources, have only a minor influence on the analysis. On the other hand, synoptic-scale structures in O–F persist in the A–F differences. That the A–F differences are much smaller (generally less than $\pm 0.15 \times 10^{15}$ molec/cm^2) than the O–F differences (up to $\pm 1.0 \times 10^{15}$ molec/cm^2) demonstrates that most tropospheric contributions are effectively discounted by the assimilation procedure.

For the second release of the TROPOMI NO_2 product, version 1.2.0 of 24 October 2018, several improvements have been included. The TM5-MP model was upgraded to the latest version, including some bug fixes. In the TM5-MP model the photolysis for SZA $> 85^\circ$ was improved, impacting in particular the stratospheric NO2 columns at high latitudes. Furthermore the assimilation of NO_2 observations is now restricted to the ascending part of the orbit, which is especially important during the spring-summer months (June-July). These changes have improved the retrieval for high SZA and in the polar regions (see Fig. 7).

6.4 Air-mass factor and vertical column calculations

The TROPOMI NO_2 algorithm uses as default pre-calculated air-mass factor look-up tables to convert the tropospheric and stratospheric slant columns into meaningful vertical columns Lorente et al. [2017]. The AMF, denoted by the symbol M, is the ratio of the slant column density of the absorbing trace gas along the (slant) optical path from sun to satellite, and the vertical column density above the point at the surface area the satellite is viewing. The total vertical column density then follows from the retrieved total slant column density:

$$N_\mathrm{v} = N_\mathrm{s}/M \tag{12}$$

The AMF depends on the vertical profile of the trace gas and can be written as (Palmer et al. [2001]; Eskes and Boersma [2003]):

$$M = \frac{\sum_l m_l\, n_l\, c_l}{\sum_l n_l}\,, \qquad m_l \equiv \delta N_\mathrm{s}/\delta n_l \tag{13}$$

with m_l the altitude-dependent AMFs or box AMF (see Sect. 6.4.1) that describe the vertically resolved sensitivity to NO_2, n_l the column density, and c_l the temperature correction term discussed below (see Sect. 6.4.2) for layer $l = 1, 2, \ldots, N_l$ [Boersma et al., 2004]. The altitude-dependent AMFs depend on retrieval (forward model) parameters, including the satellite viewing geometry, as well as surface albedo and surface pressure, cloud fraction, and cloud pressure.

The TM5-MP model provides an estimate for the stratospheric vertical profile. Summation over the layers above the tropopause level ($l > l_\mathrm{tp}^\mathrm{TM5}$) to top-of-atmosphere ($l = N_l$) provides the stratospheric AMF, from which the stratospheric slant column can then be calculated:

$$N_\mathrm{s}^\mathrm{strat} = N_\mathrm{v}^\mathrm{strat} * M^\mathrm{strat} = \sum_{l=l_\mathrm{tp}^\mathrm{TM5}+1}^{N_l} m_l\, n_l\, c_l \tag{14}$$

Subtracting this from the total slant column and using the tropospheric AMF, determined by adding up the layers from the surface ($l = 1$) up to and including the tropopause level ($l = l_\mathrm{tp}^\mathrm{TM5}$) in Eq. (13), then gives the

tropospheric vertical column:

$$N_{\mathrm{s}}^{\mathrm{trop}} = N_{\mathrm{s}} - N_{\mathrm{s}}^{\mathrm{strat}} \qquad \Rightarrow \qquad N_{\mathrm{v}}^{\mathrm{trop}} = N_{\mathrm{s}}^{\mathrm{trop}} / M^{\mathrm{trop}} \tag{15}$$

Note that the total vertial column N_{v} in Eq. (12) is *not* the same as sum of the partial vertical columns:

$$N_{\mathrm{v}}^{\mathrm{sum}} \equiv N_{\mathrm{v}}^{\mathrm{trop}} + N_{\mathrm{v}}^{\mathrm{strat}} \neq N_{\mathrm{v}} \tag{16}$$

Our best physical estimate of the NO_2 vertical column at any given place is the sum $N_{\mathrm{v}}^{\mathrm{sum}}$. Users who, for example, wish to assimilate NO_2 total columns should, however, use the total column N_{v} for this. The total column N_{v} depends strongly on the modelled ratio of the stratospheric and tropospheric sub-columns, a dependency which is partly removed in the summed product. For data assimilation use is made of the averaging kernels, and in this way the resulting analyses are not dependent on the a-priori (including the ratio of the model tropospheric and stratospheric column).

In the absence of atmospheric scattering, a so-called geometric AMF, denoted by M^{geo}, can be defined by way of a simple function of the solar zenith angle θ_0 and of the viewing zenith angle θ:

$$M^{\mathrm{geo}} = \frac{1}{\cos\theta_0} + \frac{1}{\cos\theta} \tag{17}$$

This quantity is used in the criteria for the `qa_value` (see Appendix E) but not written to the output data product.

6.4.1 Altitude dependent AMFs

The altitude-dependent AMFs, or vertical sensitivities, have been calculated with a radiative transfer model by adding a finite amount of NO_2 to the model atmosphere layer l for an atmosphere that is otherwise devoide of NO_2, and subsequently ratioing the excess NO_2 slant column (simulated with a radiative transfer model) to the vertical column added to that layer ($m_l = \delta N_{\mathrm{s}} / \delta n_l$) [Lorente et al., 2017]. The model atmosphere does not include aerosols and describes the Earth's surface as a Lambertian reflector.

As radiative transfer model we will use the Doubling-Adding KNMI (DAK) radiative transfer model (De Haan et al. [1987]; Stammes et al. [2001]), version 3.2, which has the possibility to include a pseudo-sphericity correction. The radiative transfer calculations will take the sphericity of the atmosphere into account, with Rayleigh scattering (including multiple scattering effects) and polarisation correction included (see Boersma et al. [2011] and references therein). The DAK model atmosphere consists of a Lambertian surface albedo, and an adjustable number of atmospheric layers. Atmospheric data are from the standard AFGL midlatitude summer profile. We calculate the AMF at 437.5 nm, near the middle of the spectral fitting window, for the corresponding TROPOMI NO_2 slant column retrievals; this was an ideal compromise between the small ($425-450$ nm) and wide ($405-465$ nm) fit windows, as demonstrated in the QA4ECV-project ([RD6], [ER3], see document [RD16]).

The altitude-dependent AMFs are stored in a look-up table (LUT) as a function of solar zenith angle (θ_0), viewing zenith angle (θ), relative azimuth angle (ϕ_{rel}), Lambertian surface albedo (A_{s}), surface pressure (p_{s}), and (midlevel) atmospheric pressure (p_l). This 6-dimensional LUT is to be extended with more reference points compared to earlier versions in order to respect the anticipated increase in variability of TROPOMI retrieval parameters (coarser OMI pixels have less variability in spatially smeared surface albedo and surface pressure values than anticipated for TROPOMI) and to minimise interpolation errors when looking up the appropriate altitude-dependent AMF. Pixel-specific altitude-dependent AMFs are obtained by using the best estimates for forward model parameters and a 6-D linear interpolation scheme.

Table 4 gives an overview of the reference points for the quantities that make up the 6 dimensions. The dimensions for the LUT are chosen to balance sufficiently accurate 6-dimensional linear interpolation with computational efficiency and resource economy. For out-of-bounds values (there is a slight chance that this occurs for surface pressure or atmospheric pressure) we use the point nearest to the LUT reference point (e.g. when surface pressure would be 1051 hPa, we will use the 1048 hPa reference point). In the current OMI NO_2 data product only ground pixels with $\theta_0 < 80^\circ$ ($\cos(\theta_0) = 0.174$) are used in the conversion to vertical columns. For TROPOMI and future OMI NO_2 data products the slant to vertical column conversion will not be limited in terms of θ_0; in practice this means the range will be the same as for the FRESCO+ cloud retrieval: $\theta_0 < 88^\circ$ (i.e. $\cos(\theta_0) = 0.035$), hence the lower limit of $\cos(\theta_0)$ of 0.03 in Table 4.

The TROPOMI `qa_value` (see Appendix E) will indicate that observations with $\theta_0 > 81.2$ should not be used. Experience has shown that (observation-forecast) differences increase rapidly above this point.

Table 4: Quantities and their reference points in the AMF look-up table to be used in the TROPOMI NO_2 data processing to convert the tropospheric slant column into the tropospheric vertical column. The lower limit of $\cos(\theta)$ in the list is related to the maximum value of θ for TROPOMI, which is 72° (as for OMI).

Quantity	Number of reference points	Values at reference points
Solar zenith angle $\cos(\theta_0)$	17	1.00, 0.95, 0.90, 0.80, 0.70, 0.60, 0.50, 0.45, 0.40, 0.35, 0.30, 0.25, 0.20, 0.15, 0.10, 0.05, 0.03
Viewing zenith angle $\cos(\theta)$	11	1.00, 0.95, 0.90, 0.80, 0.70, 0.60, 0.50, 0.45, 0.40, 0.35, 0.30
Relative azimuth angle $180^\circ - \lvert\phi - \phi_0\rvert$	10	0°, 20°, 40°, 60°, 80°, 100°, 120°, 140°, 160°, 180°
Surface albedo A_s	26	0.00, 0.01, 0.02, 0.03, 0.04, 0.05, 0.06, 0.07, 0.08, 0.09, 0.10, 0.12, 0.14, 0.16, 0.18, 0.20, 0.25, 0.30, 0.35, 0.40, 0.50, 0.60, 0.70, 0.80, 0.90, 1.00
Surface pressure p_s [hPa]	14	1048, 1036, 1024, 1013, 978, 923, 840, 754, 667, 554, 455, 372, 281, 130
Atmospheric pressure p_l [hPa]	174	1054.995, 1042.82, 1030.78, 1018.89, 1007.13, 995.51, 984.0309, 972.67, 961.45, 950.35, 939.39, 928.55, 917.84, 907.24, 896.71, 886.24, 875.88, 865.65, 855.54, 845.54, 835.67, 825.90, 816.26, 806.72, 797.12, 787.47, 777.93, 768.51, 759.21, 750.01, 740.93, 731.96, 723.09, 714.33, 705.65, 697.04, 688.54, 680.14, 671.85, 663.65, 655.56, 647.56, 639.66, 631.86, 624.07, 616.30, 608.62, 601.03, 593.54, 586.15, 578.85, 571.63, 564.51, 557.48, 550.44, 543.39, 536.43, 529.56, 522.77, 516.08, 509.47, 502.9492, 496.50, 490.14, 483.75, 477.32, 470.97, 464.71, 458.53, 452.44, 446.42, 440.49, 434.63, 428.86, 423.12, 417.42, 411.80, 406.26, 400.79, 395.39, 390.07, 384.82, 379.64, 374.52, 369.43, 364.37, 359.37, 354.44, 349.57, 344.78, 340.05, 335.38, 330.78, 326.24, 321.70, 317.15, 312.66, 308.24, 303.89, 299.59, 295.35, 291.18, 287.06, 283.00, 261.31, 225.35, 193.41, 165.49, 141.03, 120.12, 102.68, 87.82, 75.12, 64.30, 55.08, 47.20, 40.535, 34.79, 29.86, 25.70, 22.14, 19.08, 16.46, 14.20, 12.30, 10.69, 9.29, 8.06, 6.70, 6.11, 5.37, 4.70, 4.10, 3.57, 3.12, 2.74, 2.41, 2.12, 1.87, 1.65, 1.46, 1.29, 1.141, 1.01, 0.89, 0.79, 0.69, 0.61, 0.54, 0.48, 0.42, 0.37, 0.33, 0.29, 0.23, 0.18, 0.13, 0.10, 0.07, 0.05, 0.04, 0.030, 0.020, 0.014, 0.0099, 0.0066, 0.004471, 0.002997, 0.002005, 0.001352, 0.0009193, 0.0006300, 0.0004387, 0.000307

6.4.2 Temperature correction

For the TROPOMI NO_2 retrieval, a temperature correction is applied in the air-mass factor step (see Eq. (13)). The NO_2 cross-sections used in the DOAS retrieval, taken from Vandaele et al. [1998] [ER8], are valid for NO_2 at a temperature of 220 K. The temperature at which the NO_2 cross-section is evaluated does significantly influence the fit: amplitudes of the differential NO_2 absorption features decrease with increasing temperature, while the overall shape of the differential cross-section is in good approximation independent of temperature.

To account for the temperature sensitivity, a correction factor has been determined for the difference between the effective temperature of the NO_2 (which is derived from the ECMWF temperature profile and the modelled profiles in the data assimilation system) and the temperature of the cross-section, where the

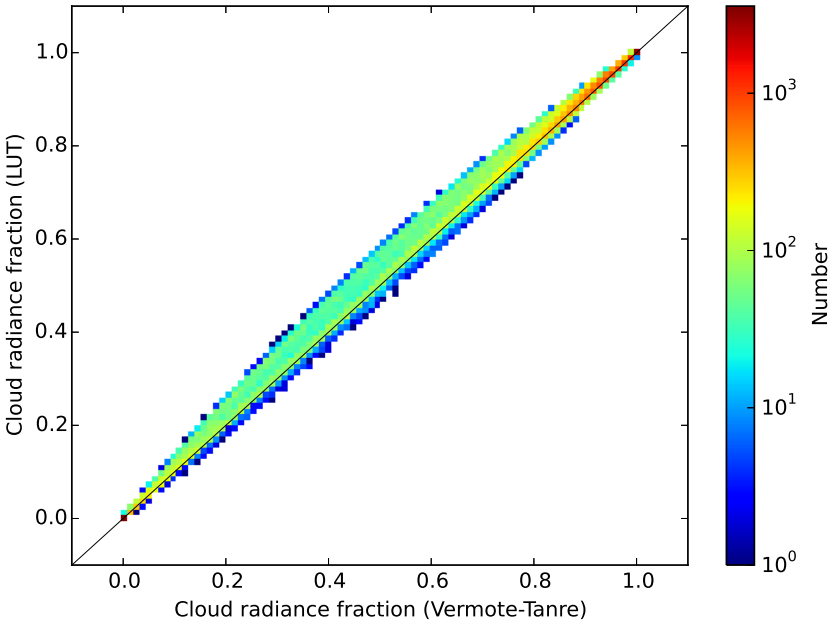

Figure 8: Comparison between the calculation of the cloud radiance fraction for the NO_2 vertical column calculation with the previously used analytical approach of Vermote and Tanré [1992] (x-axis) and the new approach using LUTs (y-axis) using data from one OMI orbit, excluding snow/ice ground pixels. The average difference between the two value is 0.013 ± 0.019.

temperature dependence is assumed to be linear. For layer l of the NO_2 profile the correction factor c_l is:

$$c_l = 1 - 0.00316(T_l - T_\sigma) + 3.39 \times 10^{-6}(T_l - T_\sigma)^2 \tag{18}$$

with T_l and T_σ the temperature of the profile layer and cross-section, respectively. The function in Eq. (18) is an update [RD17] w.r.t. the correction used for the OMI NO_2 data in DOMINO v2 (Boersma et al. [2002], Boersma et al. [2004], Bucsela et al. [2013]). Note that the temperature sensitivity given in the above equation is determined for the default wavelength window $405 - 465$ nm used for the fit; depending on the fit window and on TROPOMI's spectral resolution details, the function may need to be adapted.

6.4.3 Cloud correction

The AMF formulation accounts for cloud-contaminated pixels. Following Martin et al. [2002] and Boersma et al. [2002], the independent pixel approximation (IPA) is used to express the AMF as a linear combination of a cloudy AMF (M_{cld}) and a clear-sky AMF (M_{clr}), both for the total column and the tropospheric column:

$$M = w M_{\text{cld}} + (1-w) M_{\text{clr}} \ , \qquad M^{\text{trop}} = w M_{\text{cld}}^{\text{trop}} + (1-w) M_{\text{clr}}^{\text{trop}} \tag{19}$$

with w the radiance weighted cloud fraction, which depends on the effective cloud fraction (f_{eff}):

$$w = \frac{f_{\text{eff}} I_{\text{cl}}}{R} = \frac{f_{\text{eff}} I_{\text{cl}}}{f_{\text{eff}} I_{\text{cl}} + (1 - f_{\text{eff}}) I_{\text{cr}}} \tag{20}$$

where I_{cl} is the radiance from the cloudy part of the pixel, I_{cr} the radiance from the clear part of the pixel, and R the total scene radiance. Both I_{cl} and I_{cr} depend on the viewing geometry, the assumed (cloud) albedo, the surface pressure and the cloud pressure, following from the FRESCO-S algorithm (Sect. 6.4.4.1). In the DOMINO v2 and TM4NO2A processing of data from OMI, GOME-2 and their predecessors, these radiances were calculated following the analytical approach of Vermote and Tanré [1992], using f_{eff} from the cloud retrieval process for the same instrument. For TROPOMI, the cloud (radiance) fraction is determined from the radiance in the NO_2 fit window using LUTs, as detailed below (Sect. 6.4.4.1); Fig. 8 shows a comparison between these two approaches.

6.4.4 Retrieval parameters

6.4.4.1 Cloud cover data

The FRESCO+ algorithm (Wang et al. [2008]; [RD18], available via [ER1]) retrieves cloud information from the O_2 A-band around 758 nm: the cloud fraction and the cloud pressure, for all satellite ground pixels with solar zenith angle $\theta_0 < 88^\circ$. The surface albedo database that will be used by the FRESCO+ algorithm is based on GOME-2 observations (Tilstra et al. [2017]) at 758 and 772 nm; see Sect. 6.4.4.2.

Due to the high spectral resolution of TROPOMI compared to GOME-2, the FRESCO+ algorithm needed to be re-written and the corresponding lookup tables have been generated once more. The result, called FRESCO-S, is used for the TROPOMI NO_2 product and will be used for other Sentinels as well. See the Sentinel-5 FRESCO-S ATBD, [RD19], for details.

FRESCO+ / FRESCO-S does not provide the geometric cloud fraction but rather a radiometric equivalent cloud fraction: an effective cloud fraction, f_{eff}, that results in the same top-of-atmosphere radiance as the real cloud, based on an optically thick Lambertian cloud with a fixed albedo of $A_{\text{c}} = 0.8$ (which may be adapted in case of very bright scenes) at the cloud pressure level, p_{c}. This approach has proven to be useful for trace gas retrieval, and the errors introduced to the trace gas retrievals are usually small (and minimal for a fixed cloud albedo of $A_{\text{c}} = 0.8$; see Wang et al. [2008], who evaluated this for ozone and NO_2) when compared to scattering cloud models (Koelemeijer et al. [2001]; Stammes et al. [2008]).

Because of the large difference in wavelength between the O_2 A-band and the NO_2 retrieval window, the cloud fraction retrieved by FRESCO+ in the O_2 A-band may not be exactly representative for the cloud fraction in the NO_2 window, although Van Diedenhoven et al. [2007] found that cloud parameters retrieved from UV and O_2 A-band measurements showed good consistency for cloud fractions > 0.2; for mostly clear skies, FRESCO+ provides somewhat higher cloud fractions than UV-based retrievals.

In addition, and more importantly, a misalignment between ground pixel field-of-view of the VIS and NIR bands [RD4], containing the NO_2 retrieval window and the O_2 A-band, respectively, exists for the TROPOMI measurements.

For these reasons, the baseline option for the TROPOMI NO_2 retrieval is to (i) use the cloud pressure p_{c} from FRESCO-S and (ii) retrieve the cloud fraction and cloud radiance fraction 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, $A_{\text{s,NOtwo}}$, as forward model parameter. Here $A_{\text{s,NOtwo}}$ is taken from the OMI albedo climatology (cf. Sect. 6.4.4.2) 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, we have opted for using the modelled reflectance of Eq. (5) evaluated at 440 nm.

This approach determining the cloud fraction in the NO_2 window is very similar to FRESCO-S 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 C for details.

Fig 9 shows a comparison of the effective cloud fractions from the O_2 A-band and in the NO_2 fit window for small surface albedo in the O_2 A-band ($A_{\text{s}} < 0.1$) and for intermedidate surface albedo levels ($0.1 < A_{\text{s}} < 0.6$). The latter albedos are typical for vegetation scenes, for which the surface albedo is strongly wavelength dependent, which clearly leads to different cloud fractions. Given that the cloud fractions depends strongly on the underlying surface albedo and that for the two retrievals different surface albedo climatologies are used (GOME-2 for the O_2 A-band and OMI for the NO_2 window), even for low surface albedo ($A_{\text{s}} < 0.1$) small differences in the cloud fraction can be expected.

With processor version v1.3.0 some improvements were made in the FRESCO-S algorithm and in the way the FRESCO-S results are treated in the NO_2 algorithm. Previously in FRESCO-S the surface albedo from the database was used without modifications other than snow or ice at the surface (see Sect. 6.4.4.2 and Appendix D). In the update the surface albedo is reduced to the top of atmosphere reflectance if the top of atmosphere reflectance is lower than the surface albedo. Before the update these cases would lead to negative cloud fractions and unrealistic cloud pressures. Another change in FRESCO-S 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 $f_{\text{eff}} > 1$. Care has been taken to not use this for snow scenes. To be on the safe side, the NO_2 algorithm will discard cases where the cloud fraction in the NO_2 window ($f_{\text{eff,NO}_2}$) is negative. Fig. 10 gives an

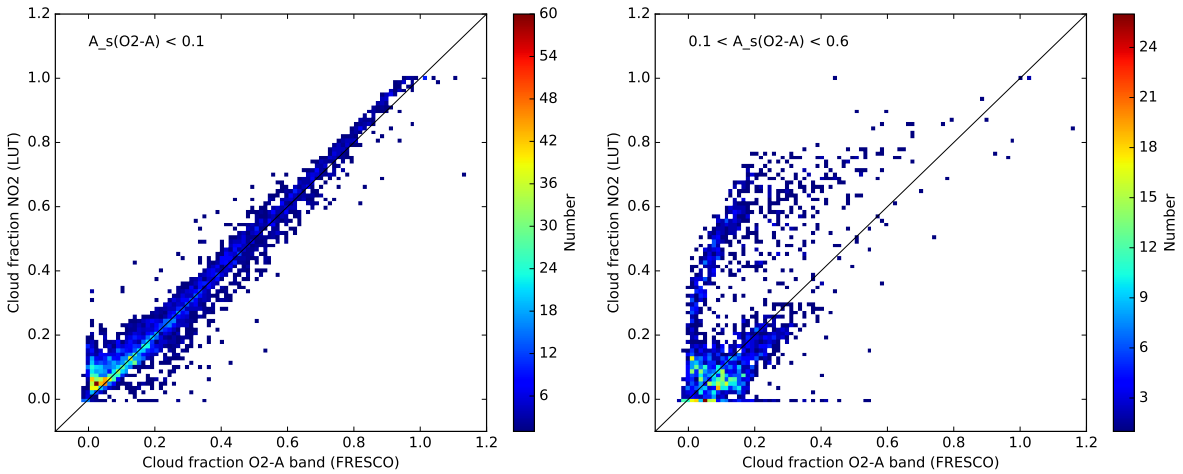

Figure 9: Comparison between the cloud fraction retrieved in the O_2 A-band by the FRESCO+ algorithm (f_{eff}; x-axis) and in the NO_2 fit window using the new LUT approach ($f_{\text{eff,NO}_2}$; y-axis) from one GOME-2A orbit, excluding snow/ice ground pixels. The comparison is shown for ground pixels with small surface albedo in the O_2 A-band ($A_s < 0.1$; *left panel*), with an average difference between the two of 0.013 ± 0.001, and intermediate surface albedo levels ($0.1 < A_s < 0.6$; *right panel*), typical for vegetation scenes.

example of the results of these improvements on the cloud pressure (top panels) and on the tropospheric NO_2 column (lower panels).

Apart from the support product FRESCO-S, TROPOMI cloud parameters will be provided by an algorithm currently under development at DLR ([RD20], available via [ER1]). Once the validity and reliability of this cloud

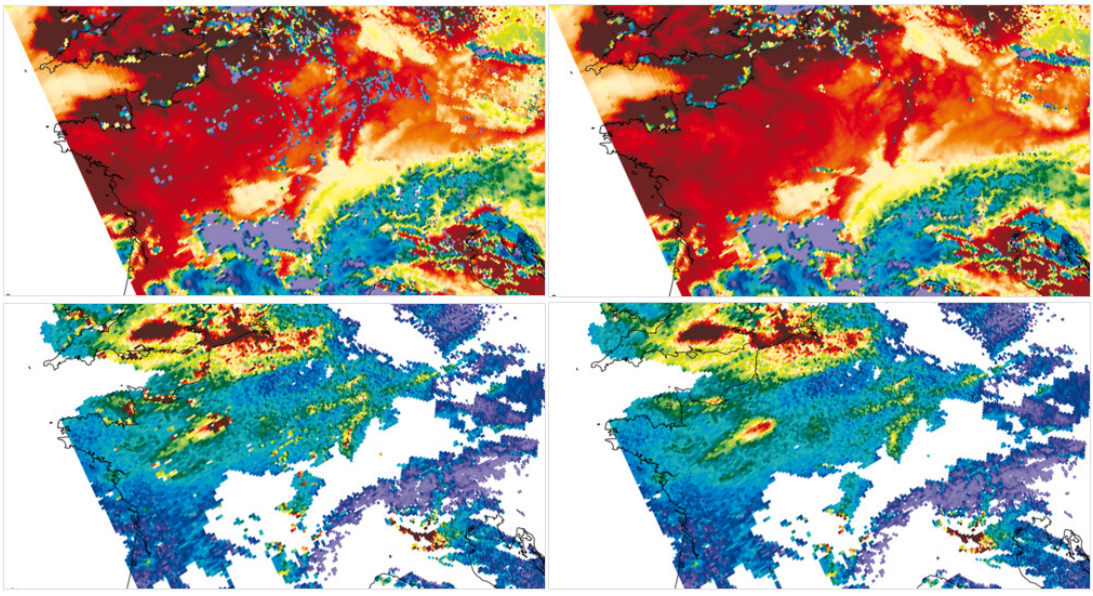

Figure 10: The retrieval of tropospheric NO_2 over France on 24 February 2018, for processor version 1.2.0 or older (left) and version 1.3.0 (right). The old FRESCO-S implementation was producing unrealistic low cloud top pressures for negative cloud fractions, the blue spots in the top-left panel. These have been removed in the new implementation (top-right). In the (relatively rare) case that the cloud fraction from the NO_2 spectral window is positive, but FRESCO-S (v1.2.0) retrieves a negative cloud fraction with unrealistic cloud pressures, this was resulting in high, noisy NO_2 spots around Paris (left panel, v1.2.0). The new treatment in FRESCO-S results in more realistic cloud pressures, the spots disappear, and we observe a well-defined pollution plume from Paris transported by the wind from the north-east (right panel, v1.3.0). Note that the NO_2 data has been filtered for clouds (white area in the lower panels).

data product is established, its cloud parameters will be tested in the NO_2 processor and the results will be compared against the results found with FRESCO-S cloud data.

6.4.4.2 Surface albedo

The baseline surface albedo climatology for TROPOMI NO_2 retrievals is the OMI database, aggregated to a grid of $0.5^\circ \times 0.5^\circ$; 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 [ER10]. This 5 years based climatology (version 3) has been used for the DOMINO v2 and QA4ECV OMI NO_2 retrievals, and is also used for the TROPOMI NO_2 retrievals. For the surface albedo in the NO_2 window, $A_{\mathrm{s,NO_2}}$, the 440 nm data is used. The climatological value of the surface albedo are adapted in case the snow/ice flag (cf. Sect 6.4.4.3) indicates there may be substantial differences in albedo; see Appendix D for some details on this correction.

The OMI albedo climatology is considered to be the best currently available source of information for the surface albedo, because of its spectral coverage in the NO_2 fit region, its relatively high spatial resolution, and the seamless transition between land and sea. An additional advantage is that the OMI climatology has been derived from observations taken at similar local times and under similar viewing conditions as the TROPOMI observations will be taken.

The Kleipool surface albedo climatology is based on OMI data, which does not cover the near-infrared wavelengths in use by the FRESCO-S algorithm to derive cloud properties (Sect. 6.4.4.1). Instead, the surface albedo database that is used by the FRESCO-S algorithm is based on GOME-2 observations [Tilstra et al., 2017] at 758 and 772 nm, given at $0.25^\circ \times 0.25^\circ$ resolution. The relatively coarse spatial resolution of GOME-2 measurements underlying the climatology and the fact that the overpass time of the OMI measurements used for the Kleipool climatology are quite similar to the overpass times of TROPOMI (i.e. the measurements are taken under similar viewing geometries), while the overpass time of GOME-2 is several hours earlier, are in favour of our choice for the Kleipool surface albedo climatology for the NO_2 retrieval, and for our choice to determine the cloud fraction in the NO_2 window.

Accounting for the anisotropic properties of surface reflectance (so-called BRDF effects) is currently not implemented in the TROPOMI NO_2 retrieval algorithm. Accounting for BRDF in OMI NO_2 retrievals has a generally small effect (<5%) with substantial effects only occurring at extreme viewing angles at high solar zenith angles [Zhou et al., 2010]. The specific choice of albedo data set is considered to be more important than accounting for surface BRDF effects. Nevertheless, both BRDF and improved horizontal resolution for the albedo dataset is considered a high priority for coming upgrades. In the near future we will consider actual developments in generating improved surface albedo data, for instance from the ADAM ([RD15], Sect. 6.1) and QA4ECV ([RD6], [ER3]) projects, and the recent work by Vasilkov et al. [2017].

6.4.4.3 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 coverage) 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 $A_{\mathrm{s,NO_2}}$ this correction follows the approach included in the OMI cloud data product OMCLDO2 [Veefkind et al., 2016] to adapt the surface albedo in the O_2–O_2 fit window (i.e. at 471 nm); Appendix D provides some details on this correction of the surface albedo reported in the Level-2 file. The baseline for snow/ice cover information are the daily updated NISE [ER11] data sets; alternatively the snow/ice data provided by the ECMWF could be used (see also [RD1]). An alternative for the snow/ice cover might come from OSISAF [ER12], but implementation is not straightforward as this are currently separate products for land and for sea.

The FRESCO algorithm (Sect. 6.4.4.1) provides two sets of data (Wang et al. [2008]; [RD18]): (i) the effective cloud fraction f_{eff} and cloud pressure p_{c} using a cloud albedo $A_{\mathrm{c}} = 0.8$ (this cloud albedo may adapted by FRESCO+ over bright scenes), and (ii) the scene albedo A_{sc} and the scene pressure p_{sc} assuming a cloud fraction $f_{\mathrm{eff}} = 0.0$. With the snow/ice flag – which is incorporated in the TROPOMI support FRESCO-S cloud product and transferred to the NO_2 data product – 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 NISE snow/ice flag indicates that there is more than a 1% snow/ice coverage, the retrieval will move to scene mode by setting the cloud radiance fraction Eq. (20) equal to 1.0. The A_{sc} and p_{sc} from FRESCO-S 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 E; see also the NO_2 Product User Manual

(PUM; available via [ER13]).

6.4.4.4 Surface pressure

The (altitude-dependent) AMFs in Eq. (13) depend on the surface pressure, p_s. For the TROPOMI NO_2 retrieval this information will be obtained from the TM5-MP model ($1^\circ \times 1^\circ$) driven by ECMWF meteorological data. Because the TM5-MP information is representative for spatially coarse pressures, the TM5-MP results will be corrected based on the method described in Zhou et al. [2009] and Boersma et al. [2011]. 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 [Maasakkers et al., 2013].

6.4.4.5 A priori vertical NO_2 profiles

A CTM is considered to be the best source of information for a priori NO_2 vertical profiles. The baseline for the TROPOMI NO_2 retrieval algorithm is to use TM5-MP Williams et al. [2017] vertical NO_2 profiles simulated at a $1^\circ \times 1^\circ$ (longitude $\times$ latitude) spatial resolution for $N_l = 34$ layers covering troposphere and stratosphere. In future updates this layer choice may be further optimised. The a priori profiles are calculated at the centre of the TROPOMI ground pixel via weighted linear interpolation based on the four nearest neighbour TM5-MP cell centres. Using TM5-MP instead of TM4 constitutes a significant improvement in itself: TM5 v3 is a benchmarked recent model version (Huijnen et al. [2010a]; Huijnen et al. [2010b]; [ER9]; Williams et al. [2017]), with more up-to-date NO_x emissions (from the MACCity inventory), chemistry, and ongoing improvements of ship, soil and lightning NO_x emission descriptions.

Using TM5-MP with a global $1^\circ \times 1^\circ$ resolution is an important improvement over previous global satellite NO_2 retrievals that used vertical profiles computed at spatial resolutions of $2^\circ \times 2.5^\circ$ or $3^\circ \times 2^\circ$ (e.g. Lamsal et al. [2010], Boersma et al. [2011]). Obviously, there are spatial gradients in NO_2 concentrations over scales smaller than a degree, but a resolution of $1^\circ \times 1^\circ$ should capture the most relevant gradients much better than a resolution of $3^\circ \times 2^\circ$. Using higher resolution models in combination with the TROPOMI averaging kernels will in effect further improve the spatial resolution in the a priori NO_2 fields for advanced users interested in regionally focused investigations (e.g. Huijnen et al. [2010b]).

The effect of the improved spatial resolution is illustrated by Fig. 11, which shows the difference between averaged tropospheric NO_2 columns from the OMI sensor from 20–30 October 2004 retrieved with TM5 at $3^\circ \times 2^\circ$ and at $1^\circ \times 1^\circ$. The retrieval with the higher resolution profile shapes leads to more pronounced contrasts between the sources of pollution and background (ventilated) pollution. To better capture the sources of air pollution is an important target of the TROPOMI mission.

6.4.5 Averaging kernels

For each ground pixel, the TROPOMI data product provides the averaging kernel. The averaging kernel for DOAS retrievals is defined as the altitude-dependent AMF ratioed (decoupled from the NO_2 vertical distribution) by the total air-mass factor [Eskes and Boersma, 2003]. Furthermore, the kernels include the height-dependent correction factors that correct for the temperature difference between the cross section used in the DOAS fit and the actual temperature in a given layer. The tropospheric averaging kernel is obtained by scaling the kernel by M/M^{trop} (see [RD21]) and setting all elements of the kernel to zero above the tropopause layer, i.e. for $l > l_{\text{tp}}^{\text{TM5}}$. Using the averaging kernel is important for data users who wish to minimise the discrepancies between the assumptions in the TROPOMI retrieval and their application of interest, for example for validation, data assimilation, or comparison to a model (e.g. Silver et al. [2013]; Boersma et al. [2016]). In particular, comparisons that make use of the kernel are no longer depending on the a-priori TM5-MP profile shape [Eskes and Boersma, 2003].

The averaging kernel should be used in validation exercises, model evaluations, and assimilation or inversion attempts with TROPOMI NO_2 columns whenever possible. The recipe for using the averaging kernel **A** for the purpose of obtaining a model estimate of the tropospheric NO_2 column ($N_{\text{v}}^{\text{trop}}$) that can be compared to TROPOMI is as follows:

$$N_{\text{v}}^{\text{trop}} = \mathbf{A}\vec{x}_m = \sum_{l=1}^{N_l} A_l\, S_l\, x_{m,l} \tag{21}$$

where S_l are the components at the l-th vertical layer of an operator that executes a mass-conserving vertical interpolation, followed by a conversion to sub-columns (molec/cm^2) in case the model vertical distribution $x_{m,l}$

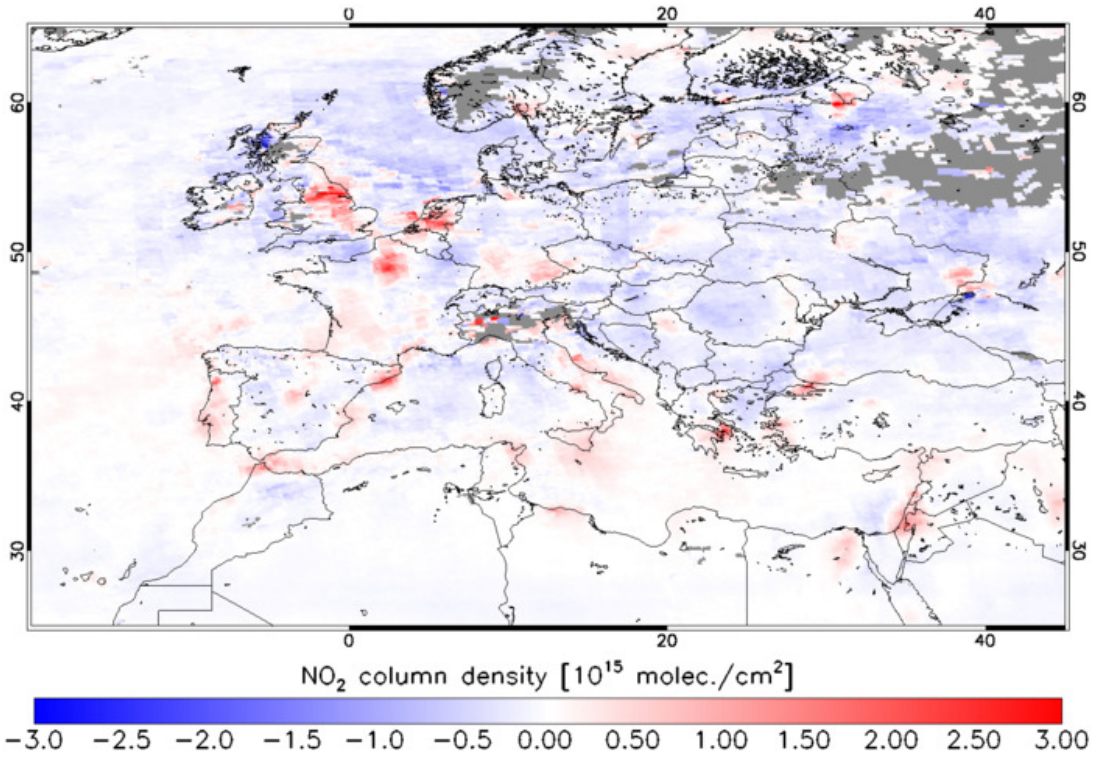

Figure 11: Tropospheric NO_2 from OMI retrieved with TM5-MP at a resolution of $1^\circ \times 1^\circ$ minus retrieved with TM5-MP at a resolution of $3^\circ \times 2^\circ$ for 20–30 October 2004 over Europe.

is not given in those units.

6.4.6 De-striping the NO_2 data product

The OMI measurements show across-track biases (stripes) in NO_2 resulting from viewing zenith angle dependent calibration errors in the OMI backscatter reflectances. For the DOMINO v2 NO_2 data product, Boersma et al. [2011] developed an empirical post-hoc de-striping correction based on the daily mean across-track dependency of the NO_2 slant columns. This correction is applied in the final step of the NO_2 processing, i.e. after the conversion to vertical columns. A new destriping approach was developed within the QA4ECV project, which is now fully integrated in the retrieval approach, avoiding the extra post-processing step.

Given that TROPOMI is measuring with a CCD detector similar to the one used by OMI, calibration related across-track biases will also be present in the TROPOMI NO_2 data. For this reason an option has been included in the Level-2 processor that allows for a de-striping correction on the NO_2 slant column data, similar to the approach implemented for QA4ECV.

Because the striping amplitude was found to be much smaller than for OMI, the stripe correction was switched off in the first release of the TROPOMI NO_2 product, v1.0.2 and v1.1.0, of July 2018. In the first update to v1.2.0 (24 October 2018) it was however decided to turn on the destriping to remove small but systematic across-track features and further improve the product quality in this way.

The de-striping is performed on orbits over the Pacific Ocean (longitude between $150^\circ\text{W} - 180^\circ\text{W}$), in order to avoid interference by tropospheric pollution hotspots. Observations are averaged over a 30° latitude range in the tropical belt. A slant column stripe amplitude is determined for each viewing angle. The stripe amplitude factors ($N^{\text{corr}}_{\text{s,NO}_2}$) are defined as the difference between the measured total slant columns ($N_{\text{s,NO}_2}$) and total slant columns derived from the model profiles ($N^{\text{mod}}_{\text{s,NO}_2}$), using the averaging kernels (**A**) and air-mass factor (M) from the retrieval.

In order to retain only features which are slowly varying over time, and in order to reduce the sensitivity to features observed during a single overpass, the stripe correction factors are averaged over a time period of 7 days, or about 7 Pacific orbits.

The slant column stripe amplitudes, one for each viewing angle, are stored in three places:

- For off-line (OFFL) processing, in a separate daily data file which contains the stripe amplitudes

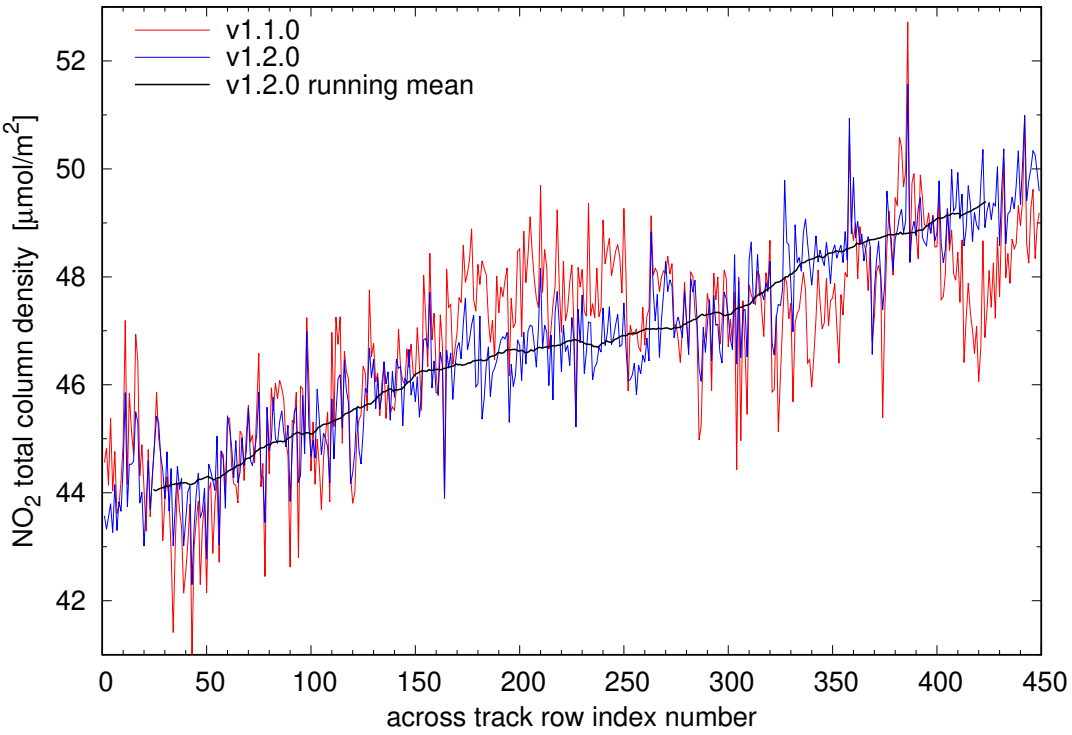

Figure 12: Comparison of the total column (i.e. stratosphere plus troposphere), averaged over the tropical Pacific Ocean on 15 July 2018 (orbit 03747) as a function of row index. The red curve is the v1.1.0 results without destriping, the blue curve is the v1.2.0 result with the destriping (the stripe correction is averaged over a week), and the black shows a 50-row running mean of the blue curve. See the text for a discussion

determined on the previous day. This file is read during every restart of the TM5-MP/DOMINO system in order to initialise the stripe correction. These files are written during the TM5-MP/DOMINO run after every update of the stripe amplitudes, when processing a Pacific orbit.

- The stripe amplitudes are written to the NO_2 Level-2 datafiles.
- The stripe amplitudes are written to the TM5-MP output files, which are used by the near-real time (NRT) NO_2 Level-2 processor.

Fig. 12 shows an example of the impact of the de-striping for TROPOMI of the total NO_2 column (i.e. stratosphere plus troposphere), averaged over the tropical Pacific Ocean on 15 July 2018 (orbit 03747) as a function of the viewing zenith angle, or row index. The red curve is the v1.1.0 results without destriping, the blue curve is the v1.2.0 result with the destriping (the stripe correction is averaged over a week), and the black shows a 50-row running mean of the blue curve. The red curve shows single-row spikes, as well as correlated structures, such as the high values around row 200 and the low values around rows 40, 320 or 420. These correlated features are currently under investigation together with the Level 0-to-1b team. The plot shows that the stripe filtering removes the major part of both the high and low frequency variability. Note that the amplitude of the structures in the red curve is small, generally within 5% of the column over the clean Pacific Ocean. Also note that we expect an increase of the total column in the stratosphere from left to right, as indicated by the black curve, due to the diurnal cycle of stratospheric NO_x chemistry.

6.5 Processing chain elements

6.5.1 Off-line (re)processing

The off-line (re)processing of the TROPOMI NO_2 retrieval algorithm, schematically displayed in Fig. 13, will take place at two locations (for more details, see [RD1]):

(1) The first step of the NO_2 processing system, illustrated in the top left part of Fig. 13, the DOAS retrieval, ingests the Level-1b spectra and is running in the PDGS TROPOMI processing system at DLR. Also

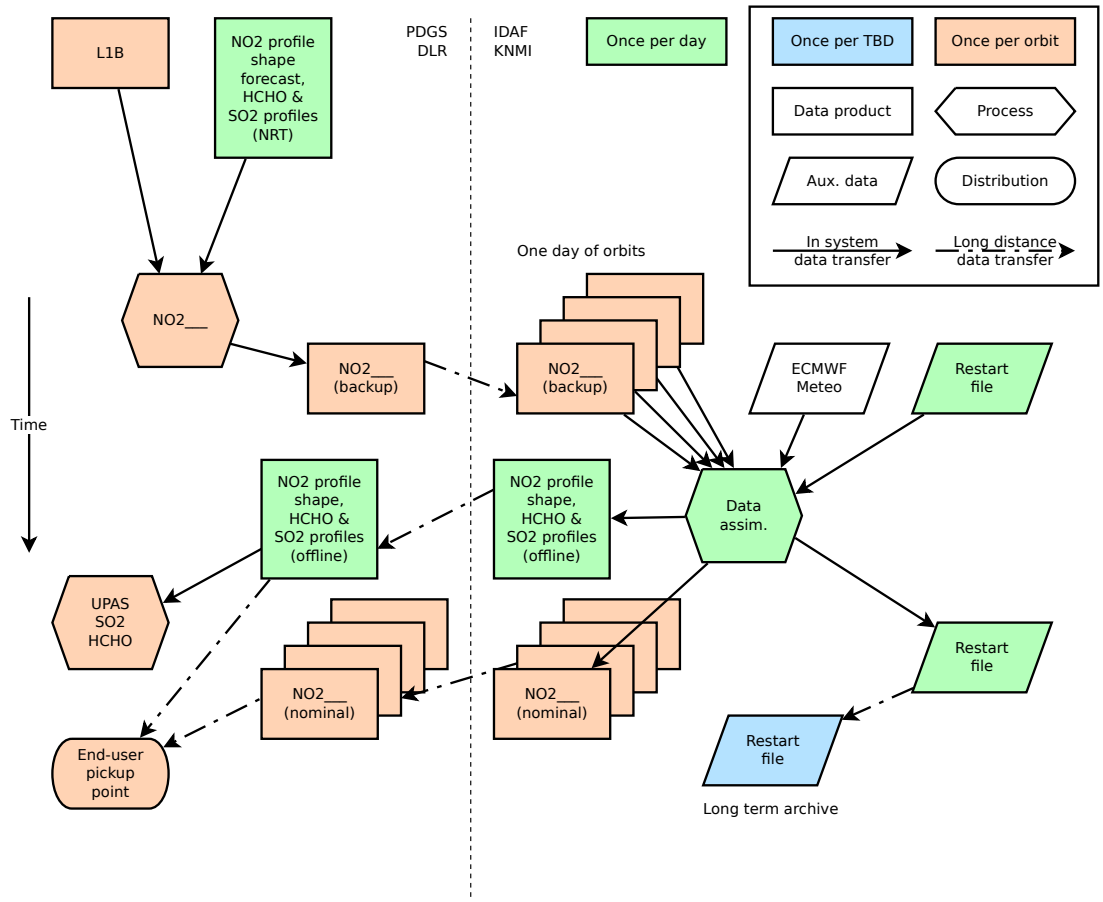

Figure 13: Schematic representation of the TROPOMI processing of tropospheric NO_2 data from a Level-1b spectrum received in the PDGS in the off-line mode. The dotted line marks the division of the processing locations: the Payload Data Ground Segment (PDGS) at DLR on the left and the Instrument Data Analysis Facility (IDAF-L2) at KNMI on the right. (Source of the figure: [RD1].)

performed in the PDGS, in a separate processing chain (not shown), is the retrieval of cloud information, needed by several Level-2 data products. The processor uses the slant column and cloud cover data to assemble a "backup" NO_2 vertical column product, based on the TM5-MP NO_2 vertical profile forecast produced for the NRT processing at the observation date (cf. Sect. 6.5.2).

(2) The NO_2 backup data product is then transferred to the IDAF at KNMI, where once a day the data of all orbits is ingested in the data assimilation / chemistry transport model TM5-MP, as illustrated in the right part of Fig. 13, to compute the off-line NO_2 product using the CTM model profiles computed with the latest ECMWF meteorological fields.

(3) The off-line or nominal NO_2 data product is then transferred back to the PDGS (bottom left part of Fig. 13), where it is made available for the users.

The motivation for this set-up is to take full advantage of the available processing elements at DLR and KNMI, and at the same time keep the number of data transfers limited. DLR will operate in the PDGS a suite of processors geared to handling large amounts of TROPOMI spectra, including the processing of NO_2 column data from TROPOMI spectra. The IDAF at KNMI hosts a complete data assimilation system based on the TM5-MP model, and has considerable experience in both the off-line and on-line retrieval of NO_2 from the GOME, SCIAMACHY, OMI, and GOME-2 instruments. The essential inputs for the processing of TROPOMI NO_2 data are (1) the Level-1b spectra measured by TROPOMI at the PDGS, and (2) the ECMWF meteo data at the IDAF.

As illustrated by Fig. 13, the data assimilation system not only provides vertical profiles for the processing of NO_2 data, but also for other TROPOMI data products: formaldehyde (HCHO) and sulphur dioxide (SO_2). Unlike NO_2, HCHO and SO_2 are not assimilated in the TM5-MP model: their profiles are output of the TM5-MP

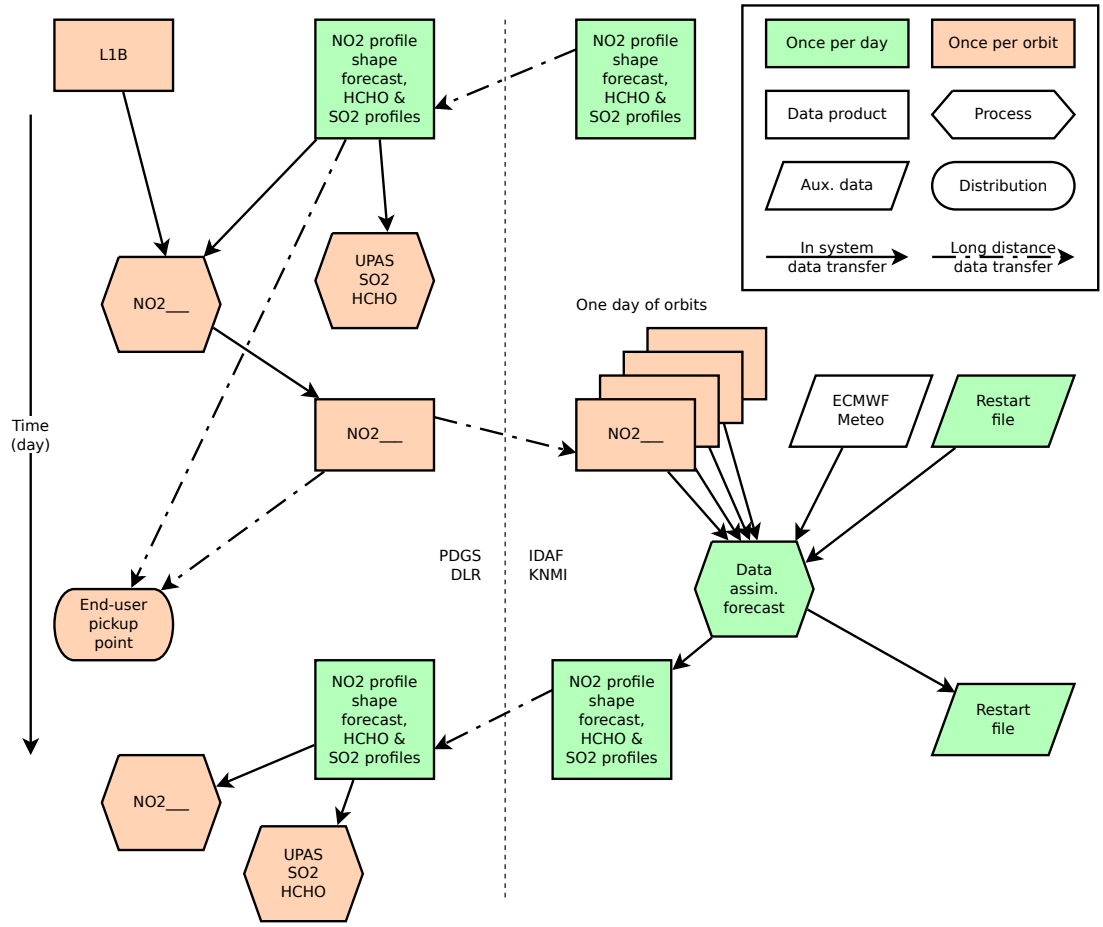

Figure 14: Schematic representation of the TROPOMI processing of tropospheric NO_2 data from a Level-1b spectrum received in the PDGS in the NRT mode. The dotted line marks the division of the processing locations: the Payload Data Ground Segment (PDGS) at DLR on the left and the Instrument Data Analysis Facility (IDAF-L2) at KNMI on the right. (Source of the figure: [RD1].)

model, based on the chemistry involving these species.

6.5.2 Near-real time processing

The NRT processing of TROPOMI NO_2 is based on the same principles as the off-line processing, described in Sect. 6.5.1. The main difference between the NRT processing, depicted in Fig. 14, and the off-line processing (Fig. 13) is the timing of the data assimilation step and the use of ECMWF meteorological forecasts rather than analysed ECMWF meteorological fields. For the NRT processing of TROPOMI data, the TM5-MP model is run once per day in the IDAF at KNMI, and ingests the NO_2 slant columns from the orbits that have been observed thus far. Based on the assimilated "state" of day i, the TM5-MP model provides a forecast of the NO_2 vertical distribution for days $i+1$ until $i+4$. This information is then transferred to the PDGS, as illustrated in Fig. 14, for the NO_2 NRT data product.

This procedure ensures that as soon as new TROPOMI measurements are available in NRT, all necessary information from the TM5-MP model is ready to be processed in the PDGS to provide an NO_2 vertical column data product, without the need for a (time consuming) model run first. With NO_2 profile data available in a 5-day forecast, an interruption of the data stream from the IDAF system is not an immediate problem for the NRT processing system. In case the interruption lasts longer than 5 days, the PDGS processing system will use the latest available NO_2 profile as a fall-back to be able to continue providing NO_2 data in NRT. As fall-back the latest available NO_2 profile is used, rather than NO_2 profile data from a climatology, because a switch to climatology data would constitute an evident discontinuity in the NO_2 data.

In the NRT processing, the TM5-MP data assimilation run is started just after midnight, as soon as the ECMWF meteo data has arrived. In that run, the system incorporates all the NO_2 slant column data that has

Table 5: Overview of data sets for each ground pixel in the final NO_2 data product assembled for dissemination via the TROPOMI website, the Sentinel-5P Core Service. Where relevant, the precision of a data set is provided as well. Data sets marked with $*$ are not part of the official Level-2 data product, but will be provided in a separate support data file. A more detailed overview can be found in Tables 11 and 12.

origin of data set	*for each ground pixel*	*symbols*
Level-1b spectrum	measurement time	t
	ground pixel centre and corner coordinates	ϑ_{geo}, δ_{geo}
	viewing geometry data	θ_0, θ, ϕ_0, ϕ
Databases	surface albedo in the NO_2 window	$A_{\text{s,NO}_2}$
	surface albedo used for the cloud retrieval	A_{s}
	surface elevation and pressure	z_{s}, p_{s}
Cloud retrieval	cloud fraction and cloud pressure FRESCO	f_{eff}, p_{c}
	scene pressure and scene albedo FRESCO	p_{sc}, A_{sc}
	cloud fraction in the NO_2 window	$f_{\text{eff,NO}_2}$
	cloud radiance fraction in the NO_2 window	w_{NO_2}
DOAS retrieval	NO_2 slant column	$N_{\text{s,NO}_2}$
	slant columns of secondary trace gases	$N_{\text{s,O}_3}$, $N_{\text{s,H}_2\text{O}_{\text{vap}}}$, $N_{\text{s,}\ldots}$
	Ring effect coefficient	C_{ring}
	polynomial coefficients	$a_m \ \ [m = 0, 1, ..., N_p]$
	intensity offset coefficients	$c_m \ \ [m = 0, 1, ..., N_{\text{off}}]$
	number of spectral points	N_λ
	degrees of freedom of the fit	D
	RMS error and χ^2 of the fit	R_{RMS}, χ^2
	wavelength calibration coefficients	w_s^{E0}, w_s, w_q
Data assimilation & AMF calculation	NO_2 tropospheric vertical column	$N_{\text{v}}^{\text{trop}}$
	NO_2 stratospheric vertical column	$N_{\text{v}}^{\text{strat}}$
	NO_2 total vertical columns	$N_{\text{v}} \equiv N_{\text{s}}/M$
		$N_{\text{v}}^{\text{sum}} \equiv N_{\text{v}}^{\text{trop}} + N_{\text{v}}^{\text{strat}}$
	NO_2 slant column stripe amplitude	$N_{\text{s,NO}_2}^{\text{corr}}$
	tropospheric AMF	M^{trop}, $M_{\text{clr}}^{\text{trop}}$, $M_{\text{cld}}^{\text{trop}}$
	stratospheric and total AMF	M^{strat}, M
	averaging kernel	$\mathbf{A}$
	TM5 tropopause layer index	$l_{\text{tp}}^{\text{TM5}}$
	TM5 pressure level coefficients	A_l^{TM5}, B_l^{TM5}
	$*$ NO_2 profile for stratosphere and troposphere	n_{l,NO_2}
	$*$ TM5 temperature profile	T_l^{TM5}
	$*$ TM5 surface elevation and pressure	$z_{\text{s}}^{\text{TM5}}$, $p_{\text{s}}^{\text{TM5}}$
Flags	quality assurance value (`qa_value`)	f_{QA}
	processing quality flags	—
	absorbing aerosol index	—
	snow/ice flag and land/water classification	—

been processed since the previous data assimilation run (from 24 hours before). Since the (ECMWF) forecast is provided up to 5 days ahead, the NRT processing is capable of providing tropospheric NO_2 data, even after a period of missing data. Previous analysis has shown [Boersma et al., 2007] that the forecast is accurate enough to provide reliable NO_2 tropospheric columns for a few days ahead. The differences between the off-line and NRT NO_2 product are found to be small.

Table 6: Overview of different user applications of NO_2 data and the data sets from the TROPOMI NO_2 data product the users will need. In addition all users may need pixel related data, such as measurement time, geolocation, viewing geometry, etc., as well as the processing and data quality flags.

	user application	*data sets needed*
# 1	Tropospheric chemistry / air quality model evaluation and data assimilation Validation with tropospheric NO_2 profile measurements (aircraft, balloon, MAX-DOAS)	$N_\mathrm{v}^\mathrm{trop}$, $\Delta N_\mathrm{v}^\mathrm{trop,kernel}$ M^trop, M, $\mathbf{A}$ † A_l^TM5, B_l^TM5, $l_\mathrm{tp}^\mathrm{TM5}$, p_s
# 2	Tropospheric column comparisons, e.g. with other NO_2 column retrievals	$N_\mathrm{v}^\mathrm{trop}$, $\Delta N_\mathrm{v}^\mathrm{trop}$
# 3	Stratospheric chemistry model evaluation and data assimilation Validation with stratospheric NO_2 profile measurements (limb/occultation satellite observations)	$N_\mathrm{v}^\mathrm{strat}$, $\Delta N_\mathrm{v}^\mathrm{strat}$ M^strat, M, $\mathbf{A}$ ‡ A_l^TM5, B_l^TM5, $l_\mathrm{tp}^\mathrm{TM5}$, p_s
# 4	Stratospheric column comparisons, e.g. with ground-based remote sensors	$N_\mathrm{v}^\mathrm{strat}$, $\Delta N_\mathrm{v}^\mathrm{strat}$
# 5	Whole atmosphere (troposphere + stratosphere) data assimilation systems	N_v, $\Delta N_\mathrm{v}^\mathrm{kernel}$ § $\mathbf{A}$ A_l^TM5, B_l^TM5, $l_\mathrm{tp}^\mathrm{TM5}$, p_s
# 6	Whole atmosphere (troposphere + stratosphere) comparisons with ground-based remote sensing (e.g. Pandora)	$N_\mathrm{v}^\mathrm{sum}$, $\Delta N_\mathrm{v}^\mathrm{sum}$ §
# 7	Visualisation of the NO_2 product	$N_\mathrm{v}^\mathrm{trop}$, $N_\mathrm{v}^\mathrm{strat}$, $N_\mathrm{v}^\mathrm{sum}$ §

† The tropospheric kernel $\mathbf{A}^\mathrm{trop}$ is derived from the total kernel $\mathbf{A}$ and the air-mass factors M and M^trop.
‡ The stratospheric kernel $\mathbf{A}^\mathrm{strat}$ is derived from the total kernel $\mathbf{A}$ and the air-mass factors M and M^strat.
§ Note that the total NO_2 vertical column $N_\mathrm{v} \equiv N_\mathrm{s}/M$ is *not* the same as the sum $N_\mathrm{v}^\mathrm{sum} \equiv N_\mathrm{v}^\mathrm{trop} + N_\mathrm{v}^\mathrm{strat}$

6.6 The NO_2 data product

The NO_2 vertical column data product contains the data sets listed in Table 5. The main product is the tropospheric NO_2 column, but the file also contains all intermediate steps such as the results from the DOAS NO_2 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).

Table 6 provides a list of six main classes of possible TROPOMI NO_2 data usage, and lists the data sets that these users will need for their applications. For notes on applying the averaging kernel, see Sect. 6.4.5. More information on the content and usage of the data product can be found in the NO_2 Product User Manual (PUM; available via [ER13]).

In order to comply with the SI unit definitions, the TROPOMI 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 – also provided as attributes to the data sets – enabling the user to easily make the conversions, if needed:

- The multiplication factor to convert mol/m^2 to molec/cm^2 is 6.02214×10^{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 3.62662×10^{37}.

The output for each ground pixel is accompanied by two flags indicating the status of the results of the processing and the retrieval. The "quality assurance value" (`qa_value` of f_QA) 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

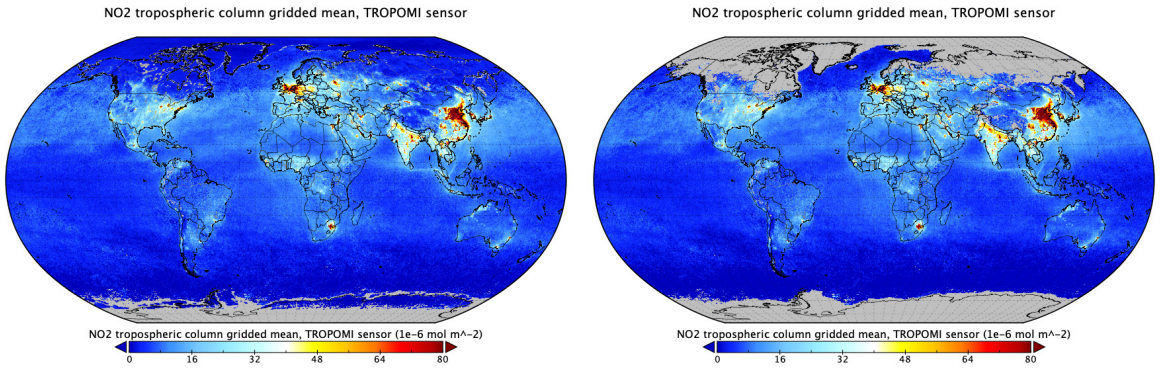

Figure 15: In TROPOMI NO_2 product version 1.2 (operational since 24 October 2018) the filtering over snow and ice is relaxed, greatly improving the coverage. *Left panel:* monthly mean NO_2 concentration for April 2018, v1.2; *right panel:* same but for v1.0.

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 E. 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 if anywhere in the processing an error occured, as indicated by the `processing_quality_flags`. Warnings related to the South Atlantic Anomaly, sun glint, or missing 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.

In the upgrade to TROPOMI NO_2 product version 1.2 (operational since 24 October 2018) the rather strict filtering of snow/ice covered scenes was abandoned and retrievals receive a `qa_value` > 0.75 when the scene pressure from the cloud retrieval is close to the surface pressure. This significantly enhances the coverage over high latitudes, as shown in Figire 15.

The "processing quality flags" (`processing_quality_flags`) contains the individual event that led to processing failure, or a precise record of the warnings that occurred during processing. The definitions and usage of these two flags is harmonised between the Level-2 data products of TROPOMI and is documented in the NO_2 Product User Manual (PUM; available via [ER13]).

The NO_2 data product provides the Absorbing Aerosol Index (AAI; the ATBD of the AAI is available via [ER1]) and a snow/ice flag (see Sect. 6.4.4.3) as additional information for the NO_2 data users, both in the off-line and the NRT processing mode. The AAI is not yet used in the flags discussed above, but this may be added in an upcoming update.

The data product will consist of two files: one with the main retrieval results and a separate TM5-MP model data file with vertical information on atmospheric NO_2, SO_2 and HCHO profile and temperature at the $1^\circ \times 1^\circ$ grid of TM5-MP on a half-hourly basis. The additional model datafile is large and will probably not be used by most NO_2 data users. For some advanced users the model profiles have shown to be useful.

7 Feasibility

7.1 Required input

The processing of TROPOMI NO_2 data poses different demands for different retrieval steps, as described in Sect. 6.5. The dynamic and static input data needed in the PDGS for the off-line and NRT processing of the NO_2 data product are listed below and summarised in Tables 7 and 8, respectively. Table 7 also mentions what the fall-back is in the processing of a given ground pixel in case the dynamical data is not available. The FRESCO/KNMI cloud product is a TROPOMI Level-2 support product, provided by KNMI software running in the PDGS. The S5P/DLR cloud product is optional; the actual use of this product will be investigated. For the snow/ice cover data, NISE [ER11] and ECMWF assimilated data are requested, at least one is required, with daily updates near the polar region, less frequent updates closer to the equator. (See also the general TROPOMI documents [RD15] and [RD1].)

7.1.1 Inputs at the PDGS for spectral fitting and air-mass factor calculation

In the PDGS at DLR, the following input is required, making a distinction between: (a) static (constant) input data and dynamic input data, which changes every orbit, and (b) data needed for the spectral fitting, and information needed in the subsequent processing step. After the DOAS NO_2 retrieval, the PDGS assembles the NO_2 vertical column data product, using information from the data assimilation system from the IDAF at KNMI for further processing, as llustrated in Fig. 13.

Spectral fitting input data for the DOAS fit

- Dynamic input:
  - Level-1b Earthshine and Solar spectra
- Static input:
  - Reference spectra (convolved with the TROPOMI slit function; [ER7]) for NO_2, O_3, H_2O_{vap}, O_2–O_2, H_2O_{liq}, Ring effect

Note that since the TROPOMI slit function differs for each of the viewing directions, i.e. for each of the detector rows, there is one set of reference spectra for each viewing direction.

NO_2 data product input data

- Dynamic input:
  - NO_2 slant column density & errors from the DOAS fit
  - NO_2 profile from the TM5-MP data assimilation system
  - Temperature and pressure profile, orography and tropopause level from TM5-MP / ECMWF
  - Geolocation data (incl. pixel corner coordinates)
  - Viewing geometry
  - Effective cloud fraction and cloud pressure
  - Scene albedo and scene pressure
  - Snow and ice cover data
- Static input:
  - Pixel-average representative (interpolated) surface albedo at 440 nm (representative for the NO_2 window)
  - Pixel-average representative (interpolated) terrain height from a digital elevation map, including a land/water classification
  - Altitude-dependent AMF look-up table
  - Cloud fraction and cloud radiance fraction look-up table

7.1.2 Inputs at the IDAF for the data assimilation

In the IDAF at KNMI, the NO_2 slant column data received from the PDGS is used in the data assimilation system to determine the NO_2 profile needed for the conversion of the NO_2 slant columns from the DOAS fit into the stratospheric and tropospheric NO_2 columns at the PDGS. For this step the following input is required, where making a distinction between the input needed for the data assimilation system.

Table 7: Overview of the dynamic input data needed for both the off-line and the NRT NO_2 data processing in the PDGS. The table does not list the meteorological input needed by the data assimilation system in the IDAF. See Sect. 7.1 for further remarks.

name/data	*symbol*	*unit*	*source*	*pre-process needs*	*backup if not available*	*comments*
S5P Level-1b Earth radiance VIS band	$I(\lambda)$	mol/s/m^2/nm/sr	S5P Level-1b product	per pixel	no retrieval	—
S5P Level-1b Solar irradiance VIS band	$E_0(\lambda)$	mol/s/m^2/nm	S5P Level-1b product	per pixel	use previous	—
NO_2 profile	$n_{l,\mathrm{NO_2}}$	mixing ratio	TM5-MP model	per pixel	latest available †	NRT
					N/A	off-line
FRESCO/KNMI cloud product	f_{eff}	1	S5P Level-2 support product	—	no VCD product	—
	p_{c}	Pa				
	A_{c}	1				
	A_{sc}	1				
	p_{sc}	Pa				
S5P/DLR cloud product	f_{eff}	1	S5P Level-2 cloud product	—	no VCD product	optional
	p_{c}	Pa				
snow/ice cover flag	—	—	NISE [ER11]	per pixel	previous day ‡	—
					ECMWF	—
aerosol absorbing index	AAI	1	S5P Level-2 AAI 354/388 nm pair	—	set AAI fill value	NRT
					set AAI fill value	off-line

[†] Latest available forecast NO_2 profile for that day.
[‡] If NISE value for the day is not available, the value previous day is used; if that value is unavailable, the snow/ice flag from ECMWF is used, with fall-back to a climatological value.

Table 8: Overview of the static input data needed for both the off-line and the NRT NO_2 data processing in the PDGS. The reference spectra convolved with the TROPOMI slit function (see column 5) are given each of the detector rows. The table does not list the input needed by the data assimilation system in the IDAF. See Sect. 7.1 for further remarks.

name/data	*symbol*	*unit*	*source*	*pre-process needs*	*comments*
absorption cross sections					
NO_2	$\sigma_{\mathrm{NO_2}}(\lambda)$	m^2/mol	Vandaele et al. [1998]	convolution	—
O_3	$\sigma_{\mathrm{O_3}}(\lambda)$	m^2/mol	Gorshelev et al. [2014] & Serdyuchenko et al. [2014]	convolution	—
O_2–O_2	$\sigma_{\mathrm{O_2-O_2}}(\lambda)$	m^5/mol^2	Thalman and Volkamer [2013]	convolution	—
H_2O_{vap}	$\sigma_{\mathrm{H_2O_{vap}}}(\lambda)$	m^2/mol	HITRAN 2012 data	convolution	†
H_2O_{liq}	$\sigma_{\mathrm{H_2O_{liq}}}(\lambda)$	1/m	Pope and Frey [1997]	convolution	—
Ring reference spectrum	$I_{\mathrm{ring}}(\lambda)$	mol/s/m^2/nm/sr	Chance and Spurr [1997]	convolution	†
retrieval input settings	—	—	KNMI	—	—
air-mass factor lookup table	—	—	KNMI	—	—
cloud fraction lookup table	—	—	KNMI	—	‡
digital elevation map	z_{s}	m	GMTED2010 [ER14]	per pixel	§
surface albedo database	$A_{\mathrm{s,NO_2}}$	1	Kleipool et al. [2008]; [ER10]	per pixel	#
			Tilstra et al. [2017]; [ER15]	per pixel	

† Created e.g. as in Van Geffen et al. [2015]; see also [RD15].
‡ For the cloud fraction retrieval in the NO_2 fit window and for the cloud radiance fraction.
# Climatological value may be adjusted based on the dynamical snow/ice flag; cf. Sect. 6.4.4.2 & App. D.
§ The elevation data is constructed as an average over a 10-km circular region (with standard deviation and extrema), which is representative for the TROPOMI pixel, sampled per 3 km [RD22].

Data assimilation input data

- Dynamic input:
  - ECMWF meteorological fields (pressure, temperature, wind, ...)
  - The "NO_2 data product input data" listed above
  - TM5-MP start field from the previous day

Table 9: Approximate computational effort for the off-line TROPOMI NO_2 processing. Any delays introduced by the different processing steps having to wait for data to be available are not included. In May 2018 the retrieval loop inside TM5-MP was made MPI-parallel, which led to a significant speedup of the code (from 5h to about 1h per day).

	Time needed for processing one TROPOMI orbit	*Time needed for processing one day of TROPOMI data*
Spectral fitting + AMF Data transfer DLR $\rightarrow$ KNMI AMF/assimilation/modelling (TM5-MP) Data transfer KNMI $\rightarrow$ DLR	10 min (9 cores) $<$ 1 min 4 mins (20 cores) $<$ 1 min	2.5 hours (9 cores) $<$ 15 min 1 hour (20 cores) $<$ 15 min
Total processing time	15 min	4 hours

Table 10: Estimate of the computational effort for the near-real time TROPOMI NO_2 processing. Any delays introduced by the different processing steps having to wait for data to be available are not included.

	Time needed for processing one TROPOMI orbit in NRT
Spectral fitting & AMF Data transfer DLR $\rightarrow$ KNMI Data assimilation with TM5-MP Data transfer KNMI $\rightarrow$ DLR	8 min (16 cores) $<$ 1 min N/A $<$ 15 min (once a day)
Total processing time	25 min

  - (In case de-striping is turned on:) Destriping coefficients from the previous day
- Static input:
  - TM5-MP static input data: NO_x (and other) emission inventories, climatologies, ...

7.2 Computational effort

Table 9 contains an overview of the processing time needed for the NO_2 product; for one pixel the spectral fitting and AMF calculation takes about 0.006 seconds (excluding overhead). The code was developed at KNMI, mainly in C++, and was transferred to and tested at DLR.

Compared to OMI, TROPOMI has about 10 times more observations, which implies a factor 10 extra computing time. Especially the TM5-MP processing of TROPOMI was taking quite some CPU time. The original implementation took about 5h to process one day of TROPOMI. In April-May 2018 the AMF retrieval loop inside TM5-MP was parallelised over observations. Together with other optimisations this has resulted in a speed-up of a factor 5, or 1h processing on 20 processors for one day of TROPOMI retrievals.

The assimilation of 15 million observations per day does not lead to a slow down of the analysis step compared to OMI. The TROPOMI pixels are binned to so-called superobservations at $1^\circ \times 1^\circ$ (Sect. 6.3). The number of TROPOMI superobservations is comparable to the number for OMI (before the row anomaly occurred). The number of superobservations to be assimilated is thinned out by a factor 2 (checkerboard approach) to further reduce the computational burden.

7.3 Near-real time timeliness

For the NRT Level-2 data to be available within the required 3 hours after measurement, it is required that the processing of Level-2 data does not take more than about 30 minutes per orbit. Table 10 shows that the actual processing time is about 8 minutes (using 16 cores), which is well within the NRT constraints.

The data assimilation run is done at KNMI once a day (just after midnight) to provide a forecast of the NO_2 profile for the coming 5 days, based on assimilation of TROPOMI slant columns observed over the previous day. These forecast runs have recently been accelerated and takes $\approx$3h for one forecast. Note that the NRT chain does not need to wait for this, as mentioned in Sect. 6.5.2.

Table 11: Overview of the data sets, their units, types and sizes, in the main data output product file, listed alphabetically; cf. Table 5. All quantities followed by a $*$ in the "symbol" column consist of the value and the associated precision (for these the number of data per pixel is doubled in the 6th column); for the vertical column densities the precisions are listed explicitly to clearly show the different types of precisions. In the last column 'PV' denotes the processor version when this variable was introduced. The data sets in the support data file are listed in Table 12.

name/data	*symbol*	*unit*	*description*	*type*	*data per pixel*	*comments*
aerosol absorbing index	—	1	L2 AAI $354/388$ nm wavel. pair	float	1	added as flag
air-mass factor	M^{trop}	1	tropospheric AMF	float	1	—
	$M^{\text{trop}}_{\text{clr}}$	1	clear-sky tropospheric AMF	float	1	since PV 1.3.0
	$M^{\text{trop}}_{\text{cld}}$	1	cloudy tropospheric AMF	float	1	since PV 1.3.0
	M^{strat}	1	stratospheric AMF	float	1	—
	M	1	total AMF	float	1	—
averaging kernel	$\mathbf{A}$	1	—	float	N_l	†
chi-squared	χ^2	1	χ^2 of the NO_2 DOAS fit	float	1	cf. Eq. (2)
cloud albedo	A_{c} *	1	used in the cloud retrieval	float	1 ×2	—
cloud pressure	p_{c} *	Pa	from the cloud retrieval	float	1 ×2	—
cloud fraction	f_{eff} *	1	from the cloud retrieval	float	1 ×2	—
cloud fraction NO_2	$f_{\text{eff,NO}_2}$	1	for the NO_2 VCD	float	1	in NO_2 fit window
cloud radiance fraction	w_{NO_2}	1	for the NO_2 VCD	float	1	in NO_2 fit window
degrees of freedom	D	1	of the slant column fit	float	1	—
DOAS fit results	$N_{\text{s,NO}_2}$ *	mol/m^2	total NO_2 SCD	float	1 ×2	—
	$N_{\text{s,H}_2\text{O}_{\text{liq}}}$ *	m	H_2O_{liq} coeff. in NO_2 window	float	1 ×2	—
	$N_{\text{s,H}_2\text{O}_{\text{vap}}}$ *	mol/m^2	H_2O_{vap} SCD in NO_2 window	float	1 ×2	—
	$N_{\text{s,O}_2\text{–O}_2}$ *	mol^2/m^5	O_2–O_2 SCD in NO_2 window	float	1 ×2	—
	$N_{\text{s,O}_3}$ *	mol/m^2	O_3 SCD in NO_2 window	float	1 ×2	—
	C_{ring} *	1	Ring coeff. in NO_2 window	float	1 ×2	—
ghost column	$N_{\text{v}}^{\text{ghost}}$	mol/m^2	NO_2 column below the clouds	float	1	‡
ground pixel coordinates	δ_{geo}	°	VIS pixel – latitude	float	5	centre, 4 corners
	ϑ_{geo}	°	VIS pixel – longitude	float	5	centre, 4 corners
ground pixel index	—	1	across-track pixel index	int	1	—
intensity off. coefficients	c_m *	1	in the NO_2 DOAS fit	float	$\left(N_{\text{off}}+1\right)$ ×2	cf. Eq. (9)
measurement time	t	s	VIS pixel	float	2	—
number of wavelengths	N_λ	1	in the NO_2 fit window	int	1	#
number of iterations	N_i	1	from the DOAS fit	int	1	—
polynomial coefficients	a_m *	1	in the NO_2 DOAS fit	float	(N_p+1) ×2	cf. Eq. (6) §
processing quality flags	—	1	—	int	1	cf. Sect. 6.6
qa value	f_{QA}	1	quality assurance value	float	1	cf. Sect. 6.6 & E
root-mean-square error	R_{RMS}	1	RMS error of the NO_2 DOAS fit	float	1	cf. Eq. (4)
satellite coordinates	z_{sat}	m	altitude of the satellite	float	1	—
	δ_{sat}	°	latitude sub satellite point	float	1	—
	ϑ_{sat}	°	longitude sub satellite point	float	1	—
	φ_{sat}	1	relative offset in orbit	float	1	—
scanline index	—	1	along-track pixel index	int	1	—
scene albedo	A_{sc} *	1	from the cloud retrieval	float	1 ×2	—
scene pressure	p_{sc} *	Pa	from the cloud retrieval	float	1 ×2	—
snow-ice flag	—	1	snow/ice case flagging	int	1	—
stripe amplitude	$N_{\text{s,NO}_2}^{\text{corr}}$	mol/m^2	NO_2 SCD stripe amplitude	float	0	cf. Sect. 6.4.6 ⊗
surface albedo	A_{s}	1	for the cloud retrieval	float	1	—
surface albedo NO_2	$A_{\text{s,NO}_2}$	1	for cloud fraction NO_2 window	float	1	—
surface classification	—	1	land/water classification	int	1	—
surface elevation	z_{s} *	m	VIS pixel	float	1 ×2	—
surface pressure	p_{s}	Pa	VIS pixel	float	1	—
TM5 pressure level coefficients	A_l^{TM5}	Pa	—	float	0	¶
	B_l^{TM5}	1	—	float	0	¶
TM5 tropopause layer index	$l_{\text{tp}}^{\text{TM5}}$	1	—	int	1	—

Table continues on next page

Table 11: — *continued.*

name/data	*symbol*	*unit*	*description*	*type*	*data per pixel*	*comments*
vertical column density	$N_{\mathrm{v,NO_2}}^{\mathrm{trop}}$	mol/m^2	tropospheric NO_2 VCD	float	1	—
	$\Delta N_{\mathrm{v,NO_2}}^{\mathrm{trop}}$	mol/m^2	*id.* precision, kernel not applied	float	1	cf. Sect. 8.4
	$\Delta N_{\mathrm{v,NO_2}}^{\mathrm{trop,kernel}}$	mol/m^2	*id.* precision, kernel applied	float	1	cf. Sect. 8.4
	$N_{\mathrm{v,NO_2}}^{\mathrm{strat}}$	mol/m^2	stratospheric NO_2 VCD	float	1	—
	$\Delta N_{\mathrm{v,NO_2}}^{\mathrm{strat}}$	mol/m^2	*id.* precision	float	1	—
	$N_{\mathrm{v,NO_2}}$	mol/m^2	total NO_2 VCD	float	1	$\equiv N_{\mathrm{s}}/M$
	$\Delta N_{\mathrm{v,NO_2}}$	mol/m^2	*id.* precision, kernel not applied	float	1	cf. Sect. 8.4
	$\Delta N_{\mathrm{v,NO_2}}^{\mathrm{kernel}}$	mol/m^2	*id.* precision, kernel applied	float	1	cf. Sect. 8.4
	$N_{\mathrm{v,NO_2}}^{\mathrm{sum}}$	mol/m^2	summed NO_2 VCD	float	1	$\equiv N_{\mathrm{v}}^{\mathrm{trop}} + N_{\mathrm{v}}^{\mathrm{strat}}$
	$\Delta N_{\mathrm{v,NO_2}}^{\mathrm{sum}}$	mol/m^2	*id.* precision	float	1	—
viewing geometry data	θ_0	$^\circ$	solar zenith angle	float	1	at surface
	ϕ_0	$^\circ$	solar azimuth angle	float	1	at surface
	θ	$^\circ$	viewing zenith angle	float	1	at surface
	ϕ	$^\circ$	viewing azimuth angle	float	1	at surface
wavelength calibration	w_s *	nm	wavelength shift	float	1 ×2	cf. Eq. (7)
radiance	w_q *	1	wavelength stretch	float	1 ×2	cf. Eq. (7)
	χ_w^2	1	χ^2 of the calibration	float	1	cf. Eq. (7)
wavelength calibration	w_s^{E0} *	nm	wavelength shift	float	0 ×2	cf. Eq. (7) ⊗
irradiance	$(\chi_w^{E0})^2$	1	χ^2 of the calibration	float	0	cf. Eq. (7) ⊗

† The number of TM5-MP layers is $N_l = 34$.
‡ The NO_2 ghost column is the NO_2 profile from TM5-MP integrated from the surface to the cloud pressure level.
# The actual number of wavelengths N_λ used in the fit (cf. Eq. (2)), i.e. after removel of, for example, bad pixels within the fit window.
§ 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.
⊗ One set per detector row.

Table 12: Overview of the data set units, types and sizes in the support output product file; this file is also used to store the profiles of HCHO and SO_2, delivered along with the NO_2 profile by the TM5-MP model. The data is provided on the TM5-MP grid resolution of $1^\circ \times 1^\circ$ on a half-hourly basis, rather than on TROPOMI pixel basis. The data sets in the main data file are listed in Table 11.

name/data	*symbol*	*unit*	*description*	*type*	*data per grid cell*	*comments*
HCHO profile	$n_{l,\mathrm{HCHO}}$	1	volume mixing ratio	float	N_l	†
NO_2 profile	$n_{l,\mathrm{NO_2}}$	1	volume mixing ratio	float	N_l	†
SO_2 profile	$n_{l,\mathrm{SO_2}}$	1	volume mixing ratio	float	N_l	†
TM5 temperature profile	T_l^{TM5}	K	—	float	N_l	†
TM5 pressure level coefficients	A_l^{TM5}	Pa	—	float	0	¶
	B_l^{TM5}	1	—	float	0	¶
TM5 surface elevation	$z_{\mathrm{s}}^{\mathrm{TM5}}$	m	—	float	1	‡
TM5 surface pressure	$p_{\mathrm{s}}^{\mathrm{TM5}}$	Pa	—	float	1	—
TM5 tropopause layer index	$l_{\mathrm{tp}}^{\mathrm{TM5}}$	1	—	int	1	—
stripe amplitude	$N_{\mathrm{s,NO_2}}^{\mathrm{corr}}$	mol/m^2	NO_2 SCD stripe amplitude	float	0	cf. Sect. 6.4.6 ⊗
date & time	—	1	year, month, day, hour, min, sec	int	0	§
time	d	days	no. of days since 1 Jan. 1950	float	0	§

† The number of TM5 layers is $N_l = 34$.
¶ One set of $N_l + 1$ (see note †) TM5-MP pressure level coefficients per data file.
‡ This data set is provided via a separate static TROPOMI digital elevation map file.
§ One set per data file.
⊗ One set per detector row.

7.4 NO_2 product description and size

The TROPOMI NO_2 data output product consists of the retrieved tropospheric and stratospheric NO_2 columns, along with error estimates and the (total) averaging kernel. A general overview of the data product contents is given in Sect. 6.6 and Table 5. Table 11 provides a more detailed overview of the data sets, their unit, type, etc. in the main output data product.

Given the number of data per ground pixel listed in Table 11, the Level-2 NO_2 output file is about 340 MB for one TROPOMI orbit (about 1.4 million observations).

The averaging kernel describes how the retrieved NO_2 columns relate to the true NO_2 profile [Eskes and Boersma, 2003]. The averaging kernel should be used in validation exercises, model evaluations, and assimilation or inverse modelling attempts with TROPOMI NO_2 data (cf. Sect 6.4.5). The output product will also contain the necessary information (surface pressure and TM5-MP sigma coordinates) to construct the pressure grid to which the averaging kernel values correspond.

For advanced users, a separate support file will be made available that contains the temperature and NO_2 (SO_2, HCHO) vertical profile. This data is given at the TM5-MP grid resolution of $1^\circ \times 1^\circ$, rather than on TROPOMI pixel basis, on a half-hourly basis for one day per file (i.e. 48 time steps); each file is about 1.6GB). The temperature and NO_2 profiles are not included in the standard Level-2 product, because most users will not need these and because vertical profiles will drastically increase the size of the TROPOMI Level-2 retrieval files. Table 12 provides an overview of the data sets in the support output data product.

8 Error analysis

The TROPOMI NO_2 retrieval algorithm generates stratospheric and tropospheric vertical column densities for all pixels. Since assumptions differ considerably for stratospheric and tropospheric retrievals, the error budget for each case will be treated separately below.

The overall error for the retrieved tropospheric columns is determined through propagation of the three main error sources: (a) measurement noise and spectral fitting affecting the slant columns, (b) errors related to the separation of stratospheric and tropospheric NO_2, and (c) systematic errors due to uncertainties in model parameters such as clouds, surface albedo, and a priori profile shape, affecting the tropospheric air-mass factor. For the stratospheric NO_2 column, the errors are driven by slant column errors, errors in the estimate of the stratospheric contribution to the slant column, and stratospheric AMF (observation operator) errors.

For NO_2, the overall error budget thus consists of several different error source terms. Errors in the slant columns will be driven in part by instrumental noise (random errors), and in part by necessary choices on the physical model and reference spectra used (systematic errors). Errors in the AMF will be mostly systematic (e.g. assumptions on albedo) but will also have random contributions (e.g. from observed cloud parameters, or sampling / interpolation errors). It is thus not possible to make a clear distinction between these error types in the total error reported in the TROPOMI NO_2 data product. This implies that by averaging TROPOMI pixels over time or over a larger area, the random part of the overall error can be largely eliminated, but systematic effects may still persist in averaged retrievals.

Experience with errors in OMI NO_2 over polluted regions, largely stemming from theoretical error analysis and practical validation studies, indicates that overall errors on the order of 25% for individual tropospheric NO_2 column retrievals may be expected. Validation studies show that the systematic part of this error is on the order of 10-15% (e.g. Hains et al. [2010]; Irie et al. [2012]; Ma et al. [2013]). For stratospheric NO_2 columns, the errors are considerably smaller and depend mostly on the absolute accuracy of the slant columns, and on the separation of the stratospheric and tropospheric contributions. The stratospheric NO_2 column error is expected to have errors on the order of 5-10% (e.g. Hendrick et al. [2012]) or $0.15-0.2\times10^{15}$ molec/cm^2.

8.1 Slant column errors

Instrument noise is the main source of errors in the spectral fitting of TROPOMI Level-1b spectra. The radiometric signal-to-noise ratios (SNR) of TROPOMI in the $400-500$ nm range turns out to be $1400-1500$ for a individual Level-1b spectra [RD4]. Experience with OMI spectral fitting in the $405-465$ nm spectral domain showed that the uncertainty in OMI NO_2 slant column densities of about 0.75×10^{15} molec/cm^2 in 2005 (when the SNR of OMI was $900-1000$) to about 0.90×10^{15} molec/cm^2 in 2015. [Boersma et al., 2007], [Zara et al., 2018]. Other, potentially systematic, errors include inaccuracies in the NO_2 cross-section spectrum (Vandaele et al. [1998]; [ER8]), in other reference spectra, notably in the Ring spectrum, and in the temperature dependence of the NO_2 cross section, but these have been shown to be of little concern for the slant column errors [Boersma et al., 2002].

Fig. 16 shows as function of the SNR an estimate of the uncertainty of the retrieved slant column density determined by a DOAS fit in the wavelength window $405-465$ nm with polynomial degree 5. Spectra were simulated with a radiative transfer code using an atmosphere with two NO_2 profiles, taken from the CAMELOT study [RD9], with the same profile shape in the stratosphere:

(*a*) European background profile, simulated with a total vertical column $N_v = 2.5\times10^{15}$ molec/cm^2

(*b*) European polluted profile, simulated with a total vertical column $N_v = 7.5\times10^{15}$ molec/cm^2

The simulations are performed with surface albedo $A_s = 0.05$, no clouds, solar zenith angle $\theta_0 = 50^\circ$, and looking down in nadir. The legend of Fig. 16 gives the total slant column N_s in 10^{15} molec/cm^2. The retrieved N_s varies very little with the SNR: about 3×10^{12} molec/cm^2 between SNR$=700$ and 1100. For profile (*a*) the retrieved N_s is within 5% of the inital N_s and for profile (*b*) it is within 3%. Given this a good accuracy of the DOAS fits can be expected, with uncertainties in the range of $10-15\%$ for background NO_2 cases and $5-10\%$ for polluted cases.

Fig. 17 shows a direct comparison of the NO_2 slant column error of OMI and TROPOMI as reported by the DOAS fits on actual observations from both instruments. The figure demonstrates that TROPOMI has an SCD error of about $8-10$ μmol/m^2, or $0.5-0.6\times10^{15}$ molec/cm^2. The OMI noise level in 2005 is about 40% higher than what is observed with TROPOMI. This is in agreement with the theoretical dependence on SNR described above.

During the TROPOMI commissioning phase it became clear that over bright scenes, e.g. high thick clouds in the tropics, the measurements at the visible wavelengths may become saturated. One very sensitive parameter to detect saturation is the NO_2 slant column uncertainty, and therefore this is used as one of the quality criteria

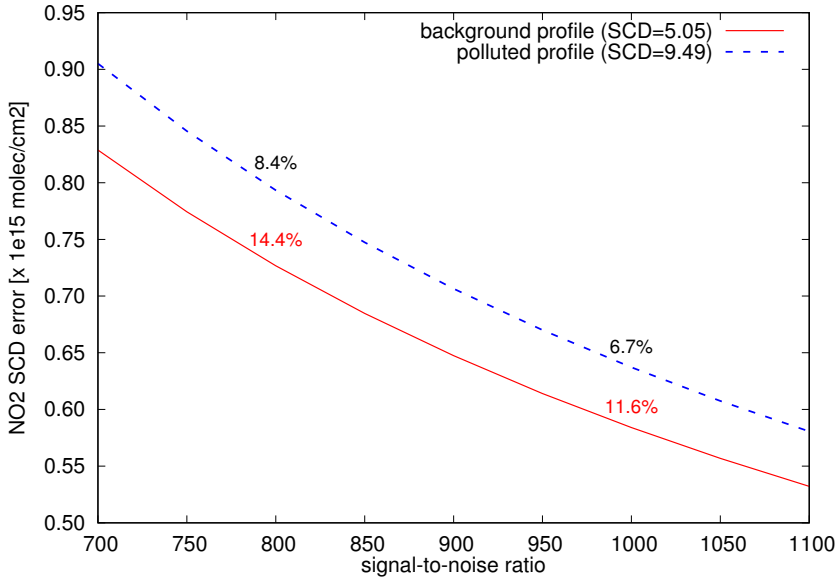

Figure 16: DOAS retrieval slant column uncertainty estimate [in 10^{15} molec/cm^2] as function of the SNR for two NO_2 profiles. The plot legend gives the retrieved slant column in 10^{15} molec/cm^2. At SNR equal 800 and 1000 the relative slant column uncertainty is indicated. For further details see the text.

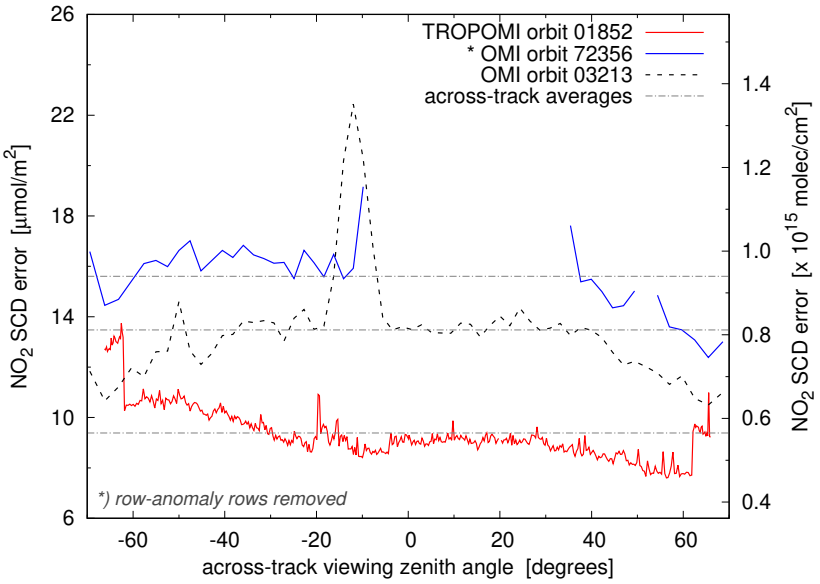

Figure 17: NO_2 slant column error estimates for the Pacific Ocean orbit on 20 Feb. 2018 of TROPOMI are compared to those for the almost overlapping OMI orbit (with rows affected by the row anomaly removed) and to those of a similar OMI orbit from 20 Feb. 2005. Data of scanlines with nadir latitude in the range $[-20°:+20°]$ is averaged along-track; OMI data is processed within the QA4ECV project [RD6], [ER3].

(`qa_value`, see Appendix E): when the slant column uncertainty exceeds 33 μmol/m^2, the pixel is flagged as bad quality.

8.2 Errors in the stratospheric (slant) columns

Data assimilation of TROPOMI NO_2 slant columns in TM5-MP provides the estimate of the stratospheric contribution to the NO_2 slant columns. The accuracy of these estimates is largely determined by the accuracy of the slant columns, as the TM5-MP stratospheric NO_2 distributions are scaled to become consistent with the retrieved slant columns. Random error estimates are derived from the assimilation approach: a considerable

advantage of the assimilation scheme is that it provides a statistical estimate of the uncertainties in the stratospheric (slant) columns through the standard deviation of the differences between the TM5-MP model analysis and forecast stratospheric NO_2 ("A–F"). Generally, the uncertainty for the stratospheric NO_2 columns is of the order of $0.1-0.2\times10^{15}$ molec/cm^2, similar to OMI [Dirksen et al., 2011]. This similarity with OMI is partly the result of using superobservations, which reduces the random contribution to the errors in the stratospheric slant column estimates. Fig. 6, bottom panel, shows the average A–F difference for 1 April 2018 in the data assimilation system based on TM5-MP. The A–F differences are on average 0.15×10^{15} molec/cm^2, and O–F over unpolluted scenes are about 0.2×10^{15} molec/cm^2. The latter is used as estimate of the uncertainties of the stratospheric NO_2 columns.

Forward (radiative transfer) model calculations are important for, but contribute little to errors in the assimilation procedure. The observation operator **H** (see Eq. (10)) is proportional to the averaging kernel [Eskes and Boersma, 2003], the vector that contains the vertical sensitivity of TROPOMI to NO_2 in each layer. The scalar product of the observation operator vector and the TM5-MP NO_2 profile at the location of the individual TROPOMI observations yields the slant column that would be observed by TROPOMI given the modeled profile. Stratospheric radiative transfer calculations around 435 nm are relatively straightforward compared to those for the troposphere, where multiple scattering occurs, and the effects of clouds and aerosols interact with the vertical distribution of NO_2. The main forward model parameter influencing errors in the stratospheric estimate is the a priori stratospheric NO_2 profile shape (and associated temperature correction), but sensitivity tests suggest that uncertainties in the exact shape of this profile are of little influence to the overall error of the stratospheric NO_2 column.

One potential source of error is the sphericity correction in the radiative transfer model. These errors are negligible for most viewing geometries, but need to be considered for far off-nadir viewing angles and high solar zenith angles. Lorente et al. [2017] investigated the differences between stratospheric NO_2 AMFs calculated with a model simulating radiative transfer for an atmosphere spherical for incoming, single-scattered, and multiple-scattered light (McArtim), and a model with an atmosphere that is spherical for incoming light, but plane-parallel for scattered sunlight. When solar and viewing zenith angles are both large, the DAK model overestimates the stratospheric AMFs by 5-10%. For TROPOMI, we therefore use an AMF LUT that is based on DAK radiative transfer simulations, but whose values for extreme viewing geometries have been made consistent with the McArtim simulations. This is the same AMF LUT that is being used in the QA4ECV retrievals of NO_2 from OMI and GOME-2A ([RD6], [ER3]).

8.3 Errors in the tropospheric air-mass factors

The tropospheric AMF is calculated with a forward model (here version 3.2 of the DAK radiative transfer model) and depends on the a priori assumed profile shape and forward model parameters (cloud fraction, cloud pressure, surface albedo, surface pressure and aerosol properties). The AMF also depends on the solar zenith, viewing zenith and relative azimuth angles, but the measurement geometry is known with high accuracy and therefore does not contribute significantly to the AMF errors. The forward model itself is assumed to represent the physics of the measurement accurately, so that forward model errors can be characterised in terms of model parameters only.

The most important AMF errors are cloud fraction, surface albedo, and a priori profile shape. Cloud parameters are obtained from TROPOMI observations, and these will have random as well as systematic components. Surface albedo and NO_2 profile shape are obtained from a priori assumptions (i.e. a pre-calculated climatology and CTM simulations, respectively), and much depends on the accuracy of these assumptions that will be different for different retrieval situations (e.g. season, surface type etc.). Because the retrieved cloud fraction depend on similar (if not the same) surface albedo assumptions as the NO_2 air-mass factors, errors will be dampened to some extent [Boersma et al., 2004].

In Table 13 the most probable uncertainties of the forward model parameters to provide a cautious error prediction for TROPOMI NO_2 AMFs are listed. For this the theoretical error propagation framework used in Boersma et al. [2004] is followed. This approach takes into account the sensitivity of the AMF to uncertainties around the actual value of a particular forward model parameter (e.g. the AMF is much more sensitive to albedo errors for dark surfaces than for brighter surfaces).

Aerosol-related errors are intimately coupled to cloud parameter errors. The O_2 A-band cloud algorithm currently does not correct for the presence of aerosols, so that an effective cloud fraction and cloud pressure will be retrieved. It is a matter of ongoing research whether or not the disentanglement of aerosol and cloud effects will improve the quality of the AMFs (Leitão et al. [2010]; Boersma et al. [2011]; Lin et al. [2014]).

The results in Table 13 provide a general estimate of overall retrieval uncertainties that may be expected for TROPOMI NO_2 data under polluted conditions. In these conditions, AMF uncertainties contribute most to

Table 13: Estimate of the contributions to the error in the AMF due to individual error sources ('BL' stands for Boundary Layer.) The estimated AMF errors are considered to be representative of 'typical' retrieval scenarios over regions of interest, i.e. with substantial NO_2 pollution for mostly clear-sky situations, and non-extreme boundary conditions for surface albedo and pressure. Note that the uncertainties can be substantially lager for specific condition, e.g. for very small albedo and large SZA.

Error type	*Estimated error*	*Corresponding AMF error*
Cloud fraction	± 0.02	$\pm 10\%$
Cloud pressure	± 50 hPa	$\pm[0-10]\%$
Surface albedo	± 0.015	$\pm 10\%$
Surface pressure	± 20 hPa	$\pm[0-5]\%$
A priori NO_2 profile shape	BL height & mixing schemes	$\pm 10\%$
A priori NO_x emissions	$\pm[0-25]\%$	$\pm[0-10]\%$
Aerosol-related errors		$\pm[0-10]\%$
Overall error		$\pm[15-25]\%$

Table 14: Relative tropospheric NO_2 vertical column per pixel uncertainty due to the tropospheric AMF uncertainty only. Estimates based on QA4ECV OMI NO_2 data for selected regions for the year 2005, taken from Boersma et al. [2018].

	Average	*box size ranges*	
Region	*AMF uncertainty*	*longitude*	*latitude*
China	17 - 22 %	$110:140$	$35:45$
USA	17 - 27 %	$-100:-75$	$35:45$
Europe	18 - 26 %	$-10:15$	$40:55$
Johannesburg	15 - 20 %	$26:30$	$-28:-24$

the retrieval uncertainties. But error analysis for individual retrievals show considerable variability on these estimates [Boersma et al., 2004]. For instance, regions with a low surface albedo are very sensitive to albedo uncertainties, and this can be reflected in AMF errors of more than 50%. For TROPOMI NO_2 a full error propagation that takes these sensitivities into account will be provided, and as well as a unique error estimate for every pixel.

Table 13 shows the settings used for OMI in the QA4ECV retrieval reprocessing. For TROPOMI we have increased the a-priori tropospheric profile shape error to 20%. A motivation for this is the increased resolution of TROPOMI which leads to an increased variability in profile shapes. Also the uncertainties in free tropospheric NO_2 lead to additional errors, and the OMI estimate may be too optimistic.

8.4 Total errors in the tropospheric NO_2 columns

The overall error in the TROPOMI tropospheric NO_2 columns is driven by error propagation of the error terms discussed before, i.e. (1) slant column errors, (2) errors associated with the separation of the stratospheric and tropospheric contributions to the slant column, and (3) tropospheric air-mass factor errors.

The overall error variance for each pixel is written as in Boersma et al. [2004]:

$$\left\langle \varepsilon^2 \right\rangle = \left(\frac{\sigma(N_\mathrm{s})}{M^\mathrm{trop}} \right)^2 + \left(\frac{\sigma(N_\mathrm{s}^\mathrm{strat})}{M^\mathrm{trop}} \right)^2 + \left(\frac{(N_\mathrm{s} - N_\mathrm{s}^\mathrm{strat}) \cdot \sigma(M^\mathrm{trop})}{(M^\mathrm{trop})^2} \right)^2 \tag{22}$$

with $\sigma(N_\mathrm{s})$ the slant column error, $\sigma(N_\mathrm{s}^\mathrm{strat})$ the stratospheric slant column error and $\sigma(M^\mathrm{trop})$ the estimated error in the tropospheric air-mass factor ($\pm 25\%$). The total error depends on details in the retrieval and therefore differs from one pixel to the next. For small tropospheric excess slant columns, the overall retrieval uncertainty is dominated by the random errors in spectral fitting, whereas for large tropospheric slant columns, the retrieval uncertainty is dominated by air-mass factor uncertainties (the last term in Eq. (22)).

Fig. 18 shows the absolute and relative error in the tropospheric NO_2 column retrieved for clear-sky scenes from TROPOMI data on 17 April 2018. We see that over the oceans and the remote continental regions, the

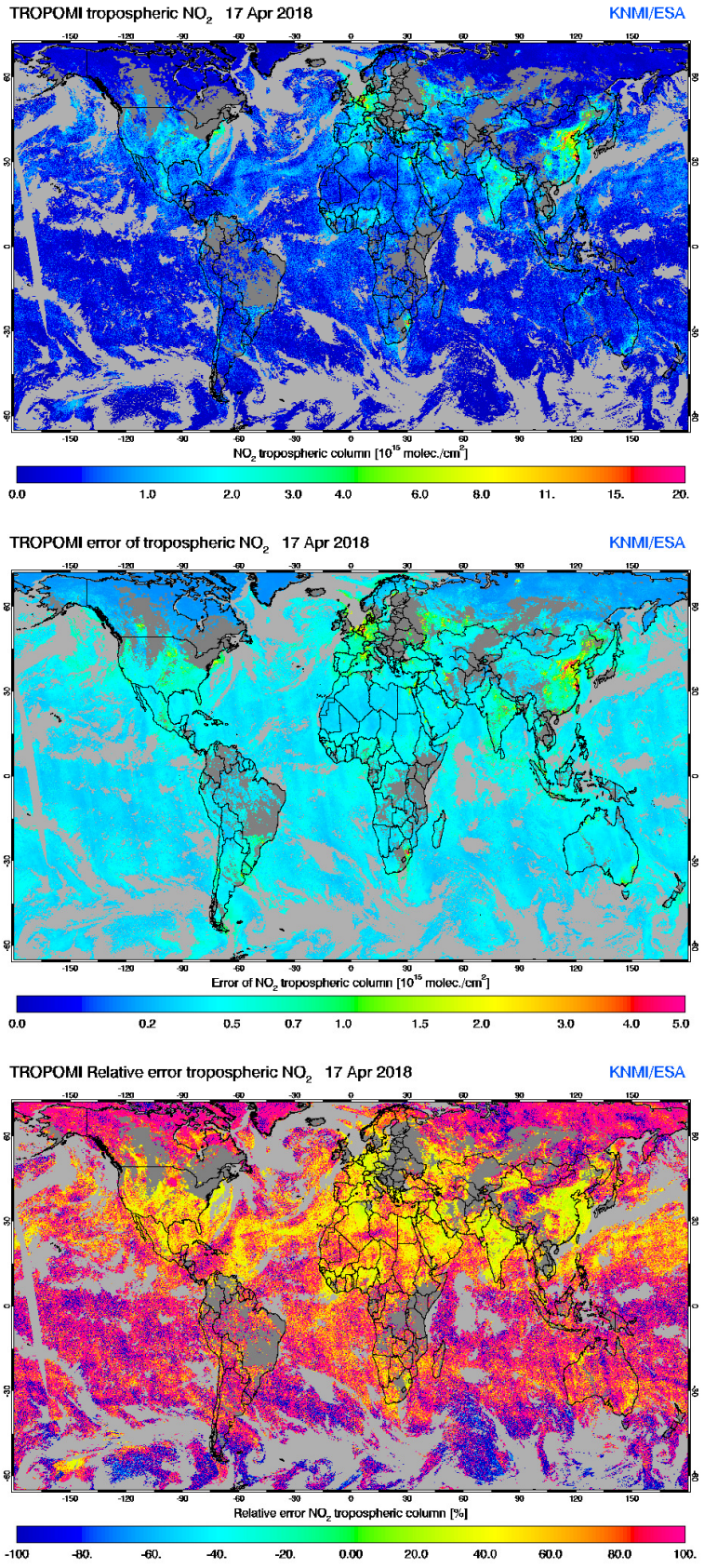

Figure 18: TROPOMI tropospheric NO_2 vertical column values (*top panel*; in 10^{15} molec/cm^2), the corresponding absolute error estimate (*middle panel*; in 10^{15} molec/cm^2; note that the scale range is reduced by a factor 4), and the relative error (*bottom panel*; in %) for 17 April 2018. Large relative errors are seen mostly over areas with small NO_2 column values: oceans and remote continental regions. These errors reflect uncertainties in the slant and stratospheric column. Over the very polluted hotspots typical errors are in the $25-40$% range, reflecting uncertainties in the air-mass factor.

overall tropospheric retrieval uncertainty is dominated by errors in the spectral fitting and the stratospheric column estimate and is typically more than 100% (indicated by purple colours in the bottom panel of Fig. 18). For larger columns over continental areas, the relative uncertainty in the retrieved column reduces to $15-50$%, and is dominated by the uncertainty in the tropospheric air-mass factor. Retrieval results are generally best for regions with strong NO_2 sources and/or high surface albedos.

Based on the instrumental performance for TROPOMI, and our experience with OMI tropospheric NO_2 retrievals (see Fig. 18 and Table 14), the overall error budget for individual TROPOMI tropospheric NO_2 retrievals can tentatively be approximated as $\varepsilon = 0.5\times10^{15}$ molec/cm^2 + [0.2 to 0.50] $\cdot N_v^{trop}$. This is a more complete and realistic error statement than the requirements from [RD5] ($\varepsilon = 1.3\times10^{15}$ molec/cm^2 + $0.1\cdot N_v^{trop}$ for a horizontal resolution of $5-20$ km; cf. Table 1).

The error components can be split in two classes: input parameter plus DOAS related uncertainties (cloud, albedo, aerosol, stratosphere, slant column) and a-priori related uncertainties (profile shape). In Rodgers optimal estimation formalism the latter may be called the smoothing error. It depends on the use of the data which uncertainty should be used. When the NO_2 vertical columns are used without knowledge of the NO_2 profiles, then the uncertainty, $\Delta N_{v,NO_2}^{trop,kernel}$, is the sum of input parameter, DOAS and smoothing. When profile information is available (e.g. when comparisons with models are performed) and the kernels are used, the uncertainty, $\Delta N_{v,NO_2}^{trop}$, is the sum of input parameter and DOAS only, without the smoothing error contribution. Both uncertainty estimates for the tropospheric vertical column will be made available in the product: one for applications with the kernel, one for applications without.

The individual components of the total uncertainty of the tropospheric column are available in the code and provided in the NO_2 data files of the QA4ECV project ([RD6], [ER3]). In the current TROPOMI NO_2 processor (mid-2017) only the total error is made available in the data product. In the next upgrade we may consider to add the tropospheric column error components due to the slant column uncertainties, stratospheric estimate and the air-mass factor, and contributions of this AMF uncertainty due to cloud fraction, cloud pressure, albedo and profile shape uncertainties.

9 Validation

9.1 Validation requirements

The most important validation need for TROPOMI NO_2 is for tropospheric NO_2 under polluted and clean conditions. Under polluted NO_2 conditions, column and profile information in the lower troposphere is essential for column density validation. The NO_2 data from OMI has been validated in several studies over the past years, based on satellite inter-comparisons (e.g. GOME-2, SCIAMACHY) as well as comparisons against ground-based (e.g. MAX-DOAS, Lidar) and in-situ (e.g. aircraft, balloon) measurements. And these validation activities have covered both tropospheric and stratospheric NO_2 data.

Despite these validation activities large uncertainties remain. These uncertainties are partly related to the quality of the independent NO_2 data used for the validation. And they are partly related to the issue of representativity of the often point-size ground-based and in-situ measurements w.r.t. the finite-sized satellite ground pixels.

The routine validation of TROPOMI is organised through the S5P Mission Performance Centre (MPC), see [ER16]. From November 2018, the MPC generates routine validation results for the TROPOMI Level-2 products in the form of up-to-date validation results and consolidated validation reports. These are available through the MPC VDAF website [ER17]. Further validation is performed by the S5P Validation Team (S5PVT) members. Preliminary results were reported at the S5P First Public Release Validation Workshop (ESA/ESRIN, June 25-26, 2018; individual contributions to the workshop are available in [ER18]). The objective of the MPC is to provide an operational service-based solution for:

- The routine quality control of the operational S5P Level 1 and Level 2 data products;
- The long-term monitoring of the TROPOMI instrument sensor performance and ageing as well as the S5P operational products;
- The in-flight calibration and characterisation of the TROPOMI instrument on-board Sentinel-5 Precursor;
- The validation of the Level-2 data products using external and independent datasets;
- The maintenance and evolution of the calibration, validation and processing algorithms.

The MPC activity includes comparisons with MAX-DOAS and PANDORA observations, SAOZ observations, and satellite observations (comparisons with OMI in particular).On top of the MPC activities, the TROPOMI data will be evaluated against any campaign data that will become available in the coming years, organised by Europe and partners outside Europe. These validation activities make a distinction between stratospheric and tropospheric NO_2.

There is also a need for correlative surface albedo data to investigate the accuracy of the OMI-based surface albedo climatology. Retrieval of TROPOMI NO_2 depends on cloud information, thus cloud properties (from the FRESCO+ retrieval and cloud fraction retrieval in the NO_2 window) must be validated carefully with correlative measurements.

Below plans for activities related to validation of the TROPOMI NO_2 data product are listed. Post-launch activities include comparisons of geophysical data comparisons between TROPOMI and correlative NO_2 data from a variety of sources.

9.2 Algorithm testing and verification

This important activity provides confidence in the retrieval algorithms, including forward and inverse models, based on simulations, and comparisons between different techniques and software programs. Much of TROPOMI's verification phase will address this issue in a thorough way. This activity also includes reviews and updates of the TROPOMI NO_2 ATBD.

In the (pre-launch) verification phase, the TROPOMI retrieval algorithm has been tested on data from existing satellite instruments, especially OMI and GOME-2. Fig. 19 shows a direct comparison of the NO_2 retrieval – in the form of the "geometric" vertical columns, defined here as N_s/M^{geo} – of OMI and TROPOMI as reported by the DOAS fits on actual observations from both instruments. The figure demonstrates that TROPOMI and OMI has lead to roughly the same absolute retrieved NO_2 slant columns, but that TROPOMI has a much smaller across-track variation ("stripiness").

An extensive comparison of the TROPOMI NO_2 with the TROPOMI verification algorithm will provide useful validation and verification of the retrieval algorithm proposed here. Similar activities have already taken place for GOME-2A and OMI in the framework of the FP7 QA4ECV project ([RD6], [ER3]). These efforts have demonstrated that the OMNO2A v2 (KNMI; VanGeffen et al. [2015]) and QA4ECV (Bremen, BIRA; [RD16])

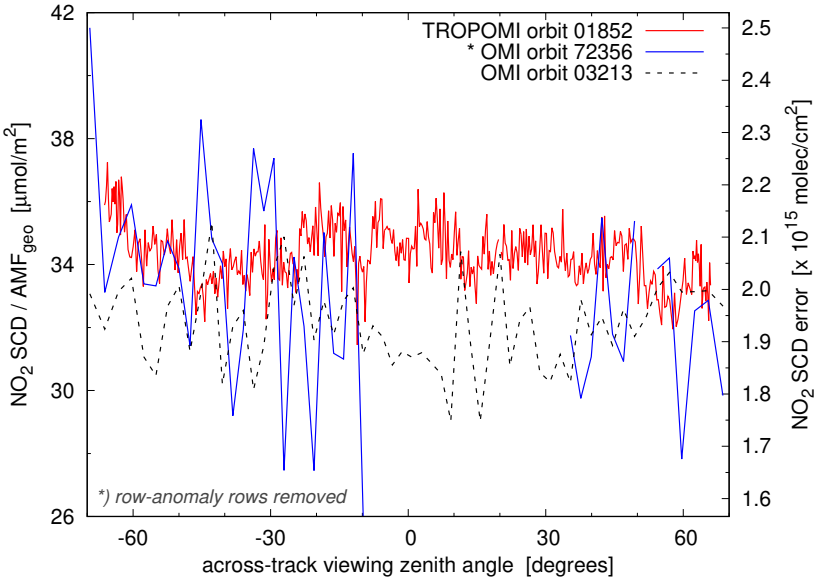

Figure 19: NO_2 "geometric" vertical columns, defined here as N_s/M^{geo}, for the Pacific Ocean orbit on 20 Feb. 2018 of TROPOMI are compared to those for the almost overlapping OMI orbit (with rows affected by the row anomaly removed) and to those of a similar OMI orbit from 20 Feb. 2005. Same data as in Fig. 17.

spectral fitting approaches lead to consistent results, but with better precision for the QA4ECV spectral fitting algorithm including so-called intensity offset correction terms [Zara et al., 2018]. In the early stages of the TROPOMI mission we have tested and compared several DOAS software packages on the TROPOMI Level-1b spectra: the TROPOMI DOAS software based on an optimal estimation solver, QDOAS from BIRA-IASB, and the OMNO2A v2 algorithm. In general good agreement was found between these algorithms. More extended comparisons will be conducted against the DOAS retrievals of IUP-Bremen University.

During the QA4ECV project, the representation of stratospheric NO_y in the model has been improved by nudging ODIN HNO3:O3 ratios, leading to more realistic NO_2 concentrations in the free-running mode. These improvements are applied to TROPOMI as well. The profile shape in the stratosphere is crucial for a good quality stratospheric column estimate. This profile shape should be monitored on a regular basis, and should be compared with (satellite limb) observations and climatology. Increasing the number of stratospheric and tropospheric layers in the TM5-MP model is something that will be tested, and the modelled diurnal cycle of stratospheric NO_2 will be investigated.

The QA4ECV-project led to the generation of a new NO_2 AMF LUT for QA4ECV NO_2 retrievals for OMI, GOME-2A, etc. This LUT was generated with DAK, and now includes a simple sphericity correction based on detailed comparisons between DAK and McArtim as described in Lorente et al. [2017]. This LUT is also used in the TROPOMI NO_2 algorithms.

9.3 Stratospheric NO_2 validation

For stratospheric NO_2 columns, correlative (column and profile) measurements are needed in regions that are representative for a complete zonal band, and hence need to be relatively unpolluted. The currently operational NDACC network covers this need in principle, although there are concerns about the accuracy of the standard SAOZ and FTIR measurements techniques for some stations (e.g. Dirksen et al. [2011]). Comparisons against the SAOZ network are reported extensively in the S5P MPC validation reports. Nevertheless, the measurements of stratospheric NO_2 concentrations taken from high-altitude ground stations such as the Jungfraujoch station in Switzerland, are particularly valuable for validation (e.g. Hendrick et al. [2012]). Other useful sources of stratospheric NO_2 data are satellite instruments that measure in limb view, SCIAMACHY (Beirle et al. [2010], Hilboll et al. [2013b]), HIRDLS and MLS [Belmonte et al., 2014], OSIRIS [Adams et al., 2016] These measurements can provide vertical profiles of NO_2 in the stratosphere, but there are difficulties in using them for direct validation as they are often only sparsely validated themselves.

In our view the main priority of the validation efforts should lie on a better characterisation of the vertical profile of stratospheric NO_2, as these profiles are an essential input to the data assimilation system in use

for the separation between tropospheric and stratospheric NO_2 columns from the TROPOMI measurements. To this end the validation measurements should be well characterised, in terms of the quality of the data and of estimates of the error on the data. Improving the knowlegde of spatial and seasonal variations in the stratospheric NO_2 profiles is also important.

Stratospheric NO_2 measurements near the Arctic vortex in late winter and early spring would be useful to better test the capability of the data assimilation scheme (and other stratosphere-troposphere separation schemes) in capturing the influence of stratospheric air masses low in NO_x on stratospheric NO_2 at lower latitudes. Such excursions are known to occur and may lead to systematic errors in the separation scheme (e.g. Dirksen et al. [2011]; Bucsela et al. [2013]). Independent measurements may provide important information on how to improve these issues.

9.4 Tropospheric NO_2 validation

For validation of tropospheric NO_2 data, correlative (column and profile) measurements are needed in the highly populated polluted regions at mid-latitudes, and also in regions with natural sources of nitrogen oxides, e.g. from biomass burning, microbial soil activity and lightning. NDACC-instruments unfortunately do not meet this need, as they are often located in relatively remote and clear areas.

Information on tropospheric NO_2 concentrations – with the NO_2 in the planetary boundary layer and/or in the free troposphere – comes from in-situ instruments (at the ground, in masts, or on low-flying balloons) and from remote-sensing instruments at the ground, on balloons or aircraft. The emerging suite of MAX-DOAS and PANDORA instruments are particularly valuable for validation of TROPOMI measurements, but homogenisation and cross-calibration of these measurements should remain a priority for successful validation.

An important issue when comparing independent NO_2 measurements with data derived from satellite-based instruments is the question of representativity. We recommend a careful investigation into the spatial representativity of any independent NO_2 measurement, in order to facilitate a meaningful comparison with the 7×3.5 km^2 TROPOMI pixel. Aircraft remote sensing observations with mapping instruments is a very valuable source of validation data for TROPOMI, e.g. Nowlan et al. [2016], Tack et al. [2018].

The local overpass time of OMI and TROPOMI are nearly identical. This implies that a direct comparison of OMI and TROPOMI for overlapping orbits is a key element of the TROPOMI validation. A first example of this is shown in Fig. 20. The figure demonstrates a good initial quantitative agreement between OMI and TROPOMI measurements, but also between the OMI-QA4ECV and TROPOMI retrieval software.

Important for the validation as well as for the data assimilation system in use for the separation between tropospheric and stratospheric NO_2 columns from the TROPOMI measurements is a good understanding of the vertical profile of the tropospheric NO_2. The best source of information on vertical profiles of NO_2 is still from incidental aircraft campaigns. Alternatively, experimental NO_2 profiles from (tethered) balloon sondes, and measurement towers, will provide valuable information on the vertical distribution of NO_2.

Since tropospheric retrievals depend on the concept of the air-mass factor, which has to rely on a priori information, it is important to also validate the inputs and assumptions that go into the air-mass factor calculation. This mostly concerns cloud parameters – cloud fraction and cloud pressure – that should be well characterised. Another critical issue, about which very little is known as yet, is the effect of the presence of aerosols on the NO_2 retrieval. Collocated information on the aerosol profile – e.g. coming from the TROPOMI Aerosol Layer Height data product – could be useful for this.

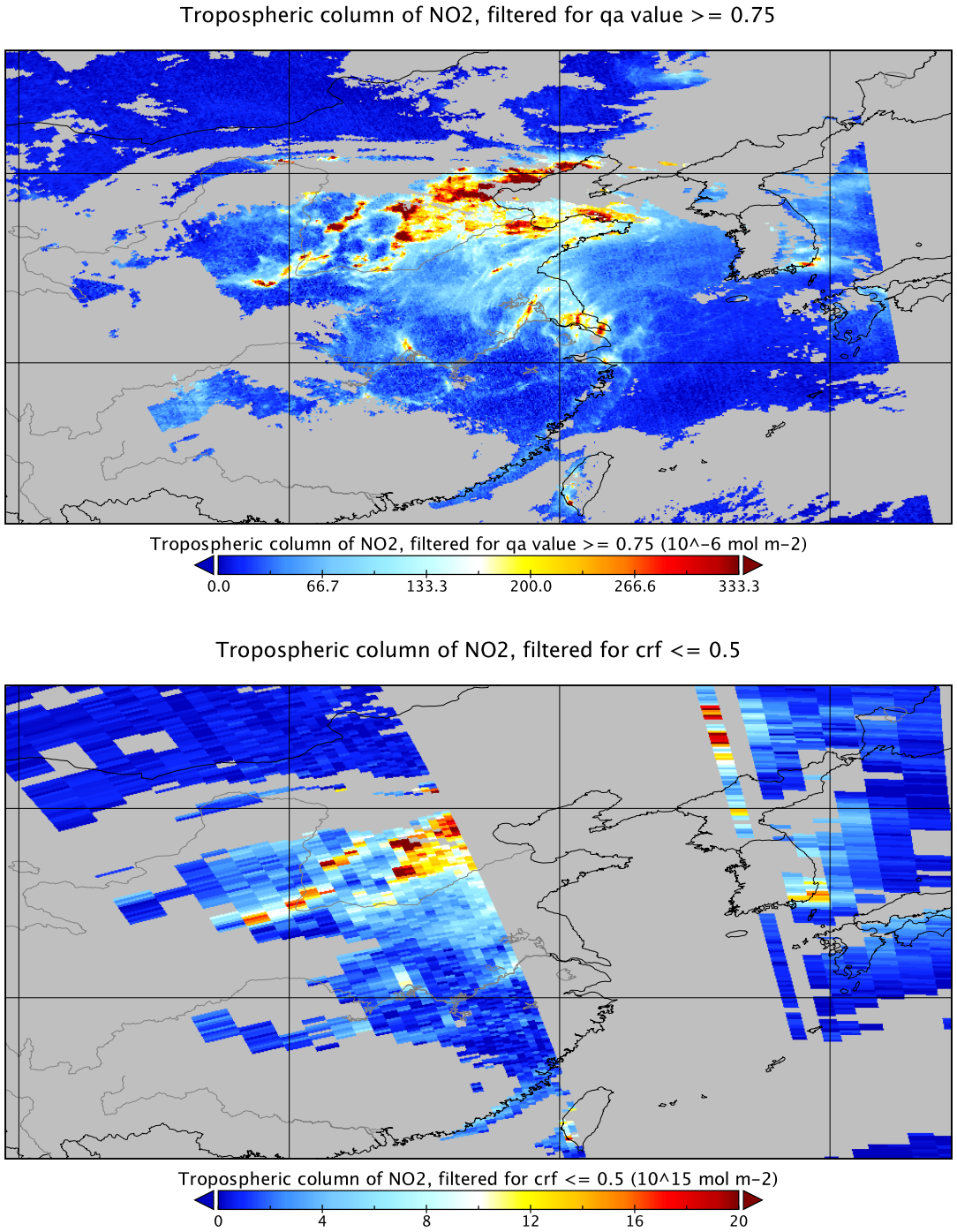

Figure 20: Tropospheric NO_2 vertical column values retrieved from TROPOMI observations (*top panel*) on 23 February 2018 (unit μ mol/m^2), compared with the corresponding OMI NO_2 tropospheric column observations for the same day (*lower panel*) (unit 10^{15} molec/cm^2). The scales have been chosen to allow a quantitative comparison. Note that the OMI data has been stripe corrected, while no stripe correction was applied to TROPOMI.

10 Conclusion

We have presented the baseline approach for the retrieval of the operational tropospheric and stratospheric NO_2 column products from the TROPOMI sensor. The NO_2 data are deliverd both as an off-line product for the NO_2 data record and as a near-real time product, with the NO_2 data delivered within 3 hours after observation. The TROPOMI NO_2 data products pose an improvement over previous NO_2 data sets, particularly in their unprecedented spatial resolution (7×3.5 km^2), but also in the separation of the stratospheric and tropospheric contributions of the retrieved slant columns, and in the calculation of the air-mass factors used to convert slant to total columns.

The backbone of the retrieval system is the TM5-MP chemistry transport model, that is operated at a global resolution of $1^\circ \times 1^\circ$. The assimilation of NO_2 slant columns in TM5-MP ensures that the modelled stratospheric state becomes consistent with the TROPOMI slant columns over regions with small tropospheric NO_2 amounts. The information from the data assimilation system is used to separate the slant column into its stratospheric and tropospheric components and to provide the a priori NO_2 vertical profile required by the air-mass factor calculation.

For each TROPOMI pixel an air-mass factor (AMF) is calculated, using altitude-dependent AMFs from a look-up table calculated with the DAK radiative transfer model, in combination with the vertical distribution of NO_2 provided by the TM5-MP chemistry transport model (in assimilation mode) at a spatial resolution of $1^\circ \times 1^\circ$. The AMF calculation uses local surface albedos from the OMI surface reflectance climatology that is based on 5 years of OMI measurements. It accounts for cloud scattering using information on effective cloud fraction and cloud pressure retrieved for every TROPOMI pixel from the reflectance at the AMF wavelength and from the FRESCO-S retrieval algorithm, respectively.

Several additional algorithm improvements w.r.t. the OMI / DOMINO v2 processing have been implemented, such as the inclusion of additional reference spectra in the DOAS spectral fit to improve the accuracy of the retrieved NO_2 slant columns, major updates to the data assimililation / chemistry transport model used to determine the vertical column densities, and a more careful quality filtering of the measurements, reflected in the `qa_value`. Residuals resulting from tests with the TROPOMI prototype fitting algorithm on OMI spectra suggest a need to include absorption by liquid water, in any case over cloud-free ocean scenes without substantial oceanic chlorophyll. Revisiting the OMI spectra also re-emphasised the importance of an appropriate spectral calibration that is representative for the complete fitting window. Using TM5-MP at a spatial resolution of $1^\circ \times 1^\circ$ (instead of a lower spatial resolution) has been shown to provide more accurate estimates of the NO_2 profiles. The conversion of the slant to vertical columns has been improved by using an air-mass factor look-up table with more nodes, in order to reduces interpolation errors.

The TROPOMI NO_2 processing chain enables us to provide a realistic error budget. The retrieval error is dominated by the spectral fitting error over oceans and regions with low tropospheric NO_2 amounts. Over the polluted regions, air-mass factor errors contribute substantially to the overall error, which can be generally approximated as 0.5×10^{15} molec/cm^2 $+ 25\%$ for an individual pixel.

Besides a complete error analysis, the TROPOMI data product will also provide the averaging kernel, which describes the sensitivity of TROPOMI to NO_2 in each model layer, for every pixel. The averaging kernel is especially relevant for data users who wish to minimise the discrepancies between the assumptions in the TROPOMI retrieval and their own application of interest, e.g. for data assimilation, validation, or comparison studies.

TROPOMI's high spatial resolution, with 7×3.5 km^2 ground pixels at nadir, will enable monitoring NO_2 columns with an unprecedented accuracy, both in the troposphere and the stratosphere. From these measurements we will learn more about the distribution of NO_2, its sources and sinks, its transport through the atmosphere, its role in stratospheric and tropospheric chemistry, as well as in climate issues, notably through the important role that nitrogen oxides play in the formation of secondary pollutants ozone and aerosol. The early-afternoon NO_2 data record, which started with OMI, will be extended by TROPOMI, alongside the mid-morning measurements of the GOME-2 instruments, thus providing essential information on the diurnal cycle of NO_2. Over the past 20-odd years various UV/Vis backscatter instruments have been used to monitor NO_2 on a global scale. The operational TROPOMI NO_2 data processing is consistent with the NO_2 retrieval record generated at KNMI, and will continue and improve that record.

A Wavelength calibration

The S5P/TROPOMI radiance spectra in the Level-1b input data are not wavelength calibrated, because most Level-2 processors will perform a wavelength calibration on the fitting window specific to the algorithm. If the calibration had been done in the Level 0-to-1b processor, it would have been done for the whole fitting window, and this may or may not have met the science requirements for the wavelength calibration for the Level-2 trace gas retrievals. 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.

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 A.1 and A.2 describe the generic wavelength fit used in most retrieval algorithms for S5P/TROPOMI, in section A.3 the actual application to NO_2 retrieval is discussed.

Note that during phase E1 the Level-1b irradiance data will not be calibrated, so this has to be done by the NO_2 processor. For this the procedure described in below for the radiance data 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.

A.1 Description of the problem

The S5P/TROPOMI Level-1b radiance spectra have a nominal wavelength scale (λ_{nom}), but this wavelength grid is not corrected for inhomogeneous slit illumination [RD14, section 28]. 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 λ_{nom} is adjusted with a wavelength offset (or: shift) w_s and a wavelength stretch w_q to find the calibrated wavelength λ_{cal}:

$$\lambda_{\text{cal}} = \lambda_{\text{nom}} + w_s + w_q \left(2\,\frac{\lambda_{\text{nom}} - \lambda_0}{\lambda_+ - \lambda_-} \right) + \cdots \tag{23}$$

with λ_0 the center of the fit window, λ_- the beginning of the fit window and λ_+ 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. (23) are ignored, even fitting w_q is optional.

A.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_{mod} to the other side of the equation. The model function $\mathscr{M}$ is given by:

$$\mathscr{M}(\lambda_{\text{nom}}; a_0, \ldots, a_N, C_{\text{ring}}, w_s, w_q, N_{\text{s},0}, \ldots, N_{\text{s},M}) = \\ P_N(\lambda^*) \cdot \exp\left(\sum_{k=0}^{M} -N_{\text{s},k}\,\sigma_k(\lambda_{\text{cal}}) \right) \cdot \left(E_{\text{mod}}(\lambda_{\text{cal}}) + C_{\text{ring}} I_{\text{ring}}(\lambda_{\text{cal}}) \right) \tag{24}$$

with λ_{cal} the calibrated wavelength as given by the first three terms in Eq. (23),

$$P_N(\lambda^*) = \sum_{k=0}^{N} a_j (\lambda^*)^j\,, \qquad \lambda^* \equiv 2\,\frac{\lambda_{\text{nom}} - \lambda_0}{\lambda_+ - \lambda_-} \tag{25}$$

a polynomial of order N, E_{mod} the reference irradiance spectrum, and I_{ring} the Ring spectrum; both E_{mod} and I_{ring} are convolved with the instrument slit function (or: instrument spectral response function; ISRF; available via [ER7]). The spectra σ_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_{vap}, CH_4, CO or O_2, and will fail at wavelengths below ~ 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 χ^2:

$$\chi^2 = \frac{1}{m-n} \sum_{i=0}^{m-1} \left(\frac{I_i - \mathscr{M}(a_0, \ldots, a_N, C_{\text{ring}}, w_s, w_q, N_{\text{s},0}, \ldots, N_{\text{s},M})}{\Delta I_i} \right)^2 \tag{26}$$

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

$$n = N + 1 + M + 1 + 3 \tag{27}$$

The additional 3 here is when fitting C_{ring}, w_s and w_q; if C_{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. (28–32) below.

$$\frac{\partial \mathscr{M}_i}{\partial a_j} = (\lambda_i^*)^j \cdot \exp\left(\sum_{k=0}^{M} -N_{\text{s},k} \sigma_k(\lambda_{\text{cal},i}) \right) \cdot \left[E_{\text{mod}}(\lambda_{\text{cal},i}) + C_{\text{ring}} I_{\text{ring}}(\lambda_{\text{cal},i}) \right] \tag{28}$$

$$\frac{\partial \mathscr{M}_i}{\partial C_{\text{ring}}} = P_N(\lambda_i^*) \cdot \exp\left(\sum_{k=0}^{M} -N_{\text{s},k} \sigma_k(\lambda_{\text{cal},i}) \right) \cdot I_{\text{ring}}(\lambda_{\text{cal},i}) \tag{29}$$

$$\begin{aligned} \frac{\partial \mathscr{M}_i}{\partial w_s} = P_N(\lambda_i^*) \cdot \exp\left(\sum_{k=0}^{M} -N_{\text{s},k} \sigma_k(\lambda_{\text{cal},i}) \right) \times \\ \left\{ \left(- \sum_{k=0}^{M} N_{\text{s},k} \left. \frac{\mathrm{d}\sigma_k}{\mathrm{d}\lambda} \right|_{\lambda=\lambda_{\text{cal},i}} \right) \cdot \left(E_{\text{mod}}(\lambda_{\text{cal},i}) + C_{\text{ring}} I_{\text{ring}}(\lambda_{\text{cal},i}) \right) \right. \\ \left. + \left(\left. \frac{\mathrm{d}E_{\text{mod}}}{\mathrm{d}\lambda} \right|_{\lambda=\lambda_{\text{cal},i}} + C_{\text{ring}} \left. \frac{\mathrm{d}I_{\text{ring}}}{\mathrm{d}\lambda} \right|_{\lambda=\lambda_{\text{cal},i}} \right) \right\} \end{aligned} \tag{30}$$

$$\begin{aligned} \frac{\partial \mathscr{M}_i}{\partial w_q} = P_N(\lambda_i^*) \cdot \exp\left(\sum_{k=0}^{M} -N_{\text{s},k} \sigma_k(\lambda_{\text{cal},i}) \right) \times \\ \left\{ \left(- \sum_{k=0}^{M} N_{\text{s},k} \lambda_i^* \left. \frac{\mathrm{d}\sigma_k}{\mathrm{d}\lambda} \right|_{\lambda=\lambda_{\text{cal},i}} \right) \cdot \left(E_{\text{mod}}(\lambda_{\text{cal},i}) + C_{\text{ring}} I_{\text{ring}}(\lambda_{\text{cal},i}) \right) \right. \\ \left. + \left(\lambda_i^* \left. \frac{\mathrm{d}E_{\text{mod}}}{\mathrm{d}\lambda} \right|_{\lambda=\lambda_{\text{cal},i}} + C_{\text{ring}} \lambda_i^* \left. \frac{\mathrm{d}I_{\text{ring}}}{\mathrm{d}\lambda} \right|_{\lambda=\lambda_{\text{cal},i}} \right) \right\} \end{aligned} \tag{31}$$

$$\frac{\partial \mathscr{M}_i}{\partial N_{\text{s},k}} = -P_N(\lambda_i^*) \cdot \sigma_k(\lambda_{\text{cal},i}) \cdot \exp\left(\sum_{k=0}^{M} -N_{\text{s},k} \sigma_k(\lambda_{\text{cal},i}) \right) \cdot \left(E_{\text{mod}}(\lambda_{\text{cal},i}) + C_{\text{ring}} I_{\text{ring}}(\lambda_{\text{cal},i}) \right) \tag{32}$$

The reference spectra $E_{\text{mod}}(\lambda)$, $I_{\text{ring}}(\lambda)$ and $\sigma_k(\lambda)$ are pre-convolved with the ISRF. During the fitting 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.

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 uses an unmodified Gauss-Newton to find the state vector for the next iteration, was selected for the S5P/TROPOMI wavelength calibration. For this usage, the a-priori error estimates are set very large (see Sect A.2.1), so that these do not limit the solution, and a pre-whitening of the data is performed to improve numerical stability.

A.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-σ 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 λ^*, as defined in Eq. (25). 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 S5P/TROPOMI is given in Table 15.

Table 15: A priori values and a priori error for the optimal estimation wavelength fit for S5P/TROPOMI. 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_{ring}	w_s	w_q	N_{s,O_3}
Prior	1	-0.5	0.01	6×10^{-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

A.3 Application of the wavelength calibration in NO_2

For the retrieval of NO_2 the N_{s,O_3} is not fitted, as O_3 shows little structure in the NO_2 and it is a weak absorber in the NO_2 fit window ($405-465$ nm).

Testing with OMI [RD23] has shown that there is no significant amount of stretch in the wavelength of the spectra of that instrument. Given the similarities of OMI and S5P/TROPOMI, the initial baseline is to not fit w_q. After launch the need to include w_q in the fit can be investigated; turning it on is a simple configuration change. The order of the polynomial N 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.

B High-sampling interpolation

After the wavelength calibration of the radiance spectrum, discussed in Appendix A, 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:

$$E_0\left(\lambda_{i,\text{earth}}\right) = \frac{E_{\text{high}}\left(\lambda_{i,\text{earth}}\right)}{E_{\text{high}}\left(\lambda_{i,\text{solar}}\right)}\, E_0\left(\lambda_{i,\text{solar}}\right) \tag{33}$$

with E_0 the observed irradiance, E_{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_{high} 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. 21 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: $R_{\text{meas}}(\lambda_{i,\text{earth}}) = \pi I(\lambda_{i,\text{earth}})/\mu_0 E_0(\lambda_{i,\text{earth}})$.

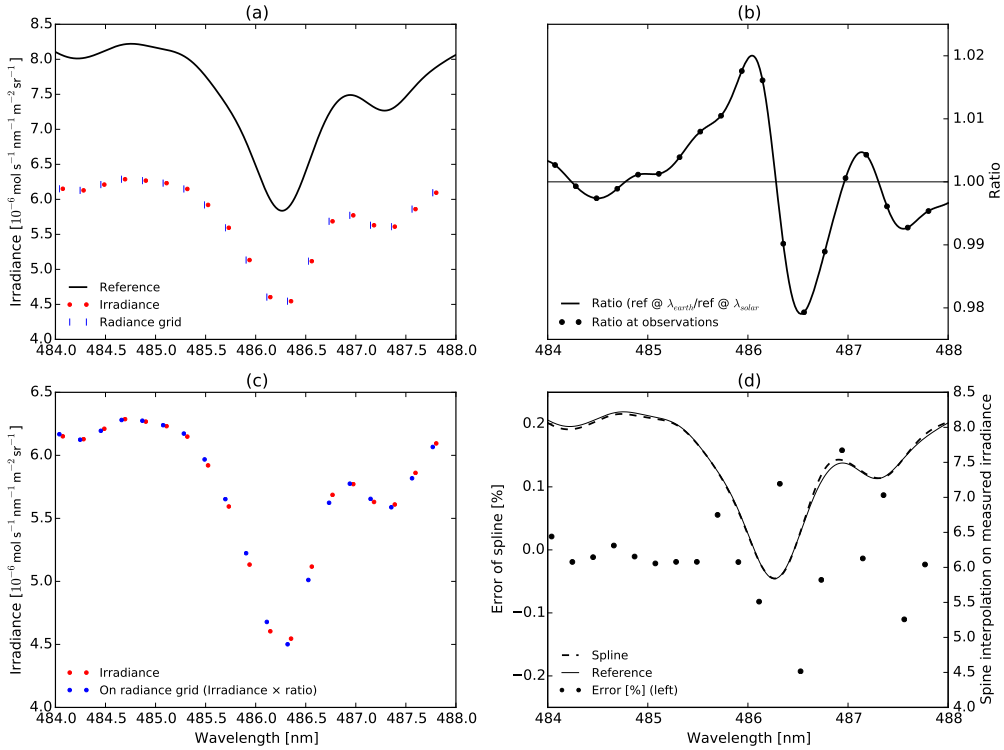

Figure 21: High sampling interpolation on part of a solar observation. (**a**) The red dots show the actual observation (taken from GOME-2A). The blue vertical lines indicate the wavelength grid of the radiance observation. The solid line shows a high resolution solar reference spectrum that has been 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.

C Effective cloud fraction in the NO_2 window

The cloud radiance fraction, w_{NO_2} (Sect 6.4.3), and the effective cloud fraction, f_{eff,NO_2} (Sect. 6.4.4.1), 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+ [RD19] and explicitly accounts for Rayleigh scattering. The following description is adapted from [RD24].

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

$$R_{\text{TOA}}(\lambda) = \frac{\pi I(\lambda)}{\mu_0 E_0(\lambda)} \tag{34}$$

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

$$f_{\text{c}} = \frac{R_{\text{TOA}} - R_{\text{s}}}{R_{\text{c}} - R_{\text{s}}} \tag{35}$$

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

$$w_{NO_2} = \frac{f_{\text{c}} R_{\text{c}}}{R_{\text{TOA}}} = \frac{f_{\text{c}} R_{\text{c}}}{f_{\text{c}} R_{\text{c}} + (1 - f_{\text{c}}) R_{\text{s}}} \tag{36}$$

where R_{s} and R_{c} are the reflectances at surface and cloud, respectively. These are computed from a limited LUT, based on Chandrasekhar (Chandrasekhar et al. [1950], Sect. 72). For bounding surface 'b', i.e. either surface ('s') or cloud ('c'):

$$R_{\text{b}}(\lambda, A_{\text{b}}(\lambda)) = R_0(\lambda) + \frac{A_{\text{b}}(\lambda)\, T(\lambda)}{1 - A_{\text{b}}(\lambda)\, s(\lambda)} \tag{37}$$

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_{c}) or surface (A_{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:

$$T(\lambda) = t(\lambda; \mu)\, t(\lambda; \mu_0) \tag{38}$$

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

$$t(\lambda; \mu) = \exp(-\frac{\tau(\lambda)}{\mu}) + \int_0^1 2\mu' T_0(\lambda; \mu, \mu') \mathrm{d}\mu' \tag{39}$$

In Eq. (39) 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_{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

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. (37) 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 below. 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_c, using the cloud pressure, p_c, and cloud albedo, A_c, from the cloud product. And the reflectance of the cloud-free part of the pixel, R_s, using the surface pressure, p_s, from meteorology or a fixed scale height and the surface elevation, z_s, and the surface albedo, A_s, from a climatology. Note that either p_s or z_s can be used as entry to the LUT: they are "linked" through the fixed scale height.

Set of LUTs for reflectance calculations: entries and dimensions (no trace gas column entries included).

R_0		Reflectance of the black surface
	λ	For all wavelengths where a cloud fraction must be computed $[1,\ldots,n]$
	μ_0	For $\mu_0 = [0.0012141231 : 1.0]$, i.e. $\theta_0 = [89.93^\circ : 0^\circ]$, in 42 steps of $2-5^\circ$
	μ	For $\mu = [0.0012141231 : 1.0]$, i.e. $\theta = [89.93^\circ : 0^\circ]$, in 42 steps of $2-5^\circ$
	$\phi-\phi_0$	Dependency stores in three Fourier terms
	p_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
	λ	For all wavelengths where a cloud fraction must be computed $[1,\ldots,n]$
	μ_0	For $\mu_0 = [0.0012141231 : 1.0]$, i.e. $\theta_0 = [89.93^\circ : 0^\circ]$, in 42 steps of $2-5^\circ$
	μ	For $\mu = [0.0012141231 : 1.0]$, i.e. $\theta = [89.93^\circ : 0^\circ]$, in 42 steps of $2-5^\circ$
	p_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
	λ	For all wavelengths where a cloud fraction must be computed $[1,\ldots,n]$
	p_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_b is linked to the elevation of the bounding surface: $z_\mathrm{b} = [-55$ m $: 16250$ m$]$.

D Surface albedo correction using NISE 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 (Sect. 6.4). This surface albedo is taken from a surface albedo climatology. For TROPOMI that is the 440 nm data from the 5-year average OMI database [Kleipool et al., 2008]; see Sect. 6.4.4.2. 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. 6.4.4.3) 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 [Veefkind et al., 2016] 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 NISE snow/ice flag database [ER11], which is sampled at the ground pixel centre coordinate. Table 16 provides an overview of these flags, where an asterisk marks NISE flags that lead to an adjustment of the surface albedo in case a certain threshold is exceeded.

Table 16: Overview of the NISE snow/ice flags f_{NISE}. 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 (NISE 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 (NISE 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 (NISE flags $f_{NISE} = 1 - 100$) adjust A_s f 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 [Kleipool et al., 2008]:

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

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

E Data quality value: the `qa_value` flags

To make the use of the TROPOMI data product files easier, a so-called `qa_value` (where 'qa' stands for 'quality assurance') is assigned to each ground pixel. 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. Some of these warnings have been included in the `qa_value`. The meaning of the `processing_quality_flags` values is detailed in Appendix A of the NO_2 Product User Manual (PUM; available via [ER13]).

The following differentiation of the `qa_value`, f_{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_{QA} is multiplied by the modification factor f^i_{QA} of each of the criteria i listed in Table 17 that have been met (i.e. if criterion i is not met then $f^i_{\mathrm{QA}} = 1$).

Table 17: Overview of the selection criteria for the `qa_value`, f_{QA}. Some quantities have a minimum or maximum value; these values are configuration parameters in the processing. In this table f_{NISE} stands for the NISE snow/ice flags listed in Table 16. And f_{AAI} represents the aerosol index $354/388$ nm pair, which is passed on to the NO_2 data product file as added flag.

i	criterion	f^i_{QA}
1	fatal error encountered according to `processing_quality_flags`	0.00
2	`south_atlantic_anomaly_warning` set in `processing_quality_flags`	0.95
3	`sun_glint_warning` set in `processing_quality_flags`	0.93
4	`pixel_level_input_data_missing_warning` set in `processing_quality_flags`	0.90
5	if $\theta_0 > \theta_0^{\mathrm{max},1} = 81.2°$	0.30
6	if $\theta_0 > \theta_0^{\mathrm{max},2} = 84.5°$	0.00
7	if $M^{\mathrm{trop}}/M^{\mathrm{geo}} < M^{\mathrm{trop}}_{\mathrm{min}} = 0.1$	0.45
8	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
	if $f_{\mathrm{NISE}} < f^{\mathrm{max}}_{\mathrm{NISE}} = 1$ * or $f_{\mathrm{NISE}} = 252$ or $f_{\mathrm{NISE}} = 255$ then [no snow or ice]	
9	$A_{\mathrm{s,NO_2}} > A^{\mathrm{max}}_{\mathrm{s}} = 0.3$	0.20
10	$w_{\mathrm{NO_2}} > w^{\mathrm{max}}_{\mathrm{NO_2}} = 0.5$	0.74
	else-if ($f_{\mathrm{NISE}} \neq 253$ and $f_{\mathrm{NISE}} \neq 254$) [snow/ice case]	
11	if ($f_{\mathrm{NISE}} > 80$ and $f_{\mathrm{NISE}} < 104$ and $p_{\mathrm{sc}} > 0.98 \cdot p_{\mathrm{s}}$) [cloud-free snow/ice]	0.88
12	else [cloudy snow/ice]	0.73
13	if $p_{\mathrm{sc}} < p^{\mathrm{min}}_{\mathrm{sc}} = 3.0\times10^4$ Pa	0.25
14	else [snow/ice error]	0.00
15	$f_{\mathrm{AAI}} > f^{\mathrm{max}}_{\mathrm{AAI}} = 1.0\times10^{10}$ [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.

F References

[Adams et al., 2016] Adams, C., Normand, E. N., McLinden, C. A., Bourassa, A. E., Lloyd, N. D., Degenstein, D. A., Krotkov, N. A., Rivas, M. B., Boersma, K. F., and Eskes, H. (2016). Limb-nadir matching using non-coincident no2 observations: proof of concept and the omi-minus-osiris prototype product. *Atmos. Meas. Tech.*, 9:4103–4122.

[Adams et al., 2013] Adams, C., Strong, K., Zhao, X., Bourassa, A. E., Daffer, W. H., Degenstein, D., Drummond, J. R., Farahani, E. E., Fraser, A., Lloyd, N. D., Manney, G. L., McLinden, C. A., Rex, M., Roth, C., Strahan, S. E., Walker, K. A., and Wohltmann, I. (2013). The spring 2011 final stratospheric warming above Eureka: anomalous dynamics and chemistry. *Atmos. Chem. Phys.*, 13:611–624.

[Bak et al., 2013] Bak, J., Kim, J. H., Liu, X., Chance, K., and Kim, J. (2013). Evaluation of ozone profile and tropospheric ozone retrievals from GEMS and OMI spectra. *Atmos. Meas. Tech.*, 6:239–249.

[Beirle et al., 2011] Beirle, S., Boersma, K. F., Platt, U., Lawrence, M. G., and Wagner, T. (2011). Megacity emissions and lifetimes of nitrogen oxides probed from space. *Science*, 333:1737–1739.

[Beirle et al., 2016] Beirle, S., Hörmann, C., P., J., Liu, S., Penning de Vries, M., Pozzer, A., Sihler, H., Valks, P., and Wagner, T. (2016). The STRatospheric Estimation Algorithm from Mainz (STREAM): estimating stratospheric NO2 from nadir-viewing satellites by weighted convolution. *Atmos. Meas. Tech.*, 9:2753–2779.

[Beirle et al., 2010] Beirle, S., Kühl, S., Puķīte, J., and Wagner, T. (2010). Retrieval of tropospheric column densities of NO_2 from combined SCIAMACHY nadir/limb measurements. *Atmos. Meas. Tech.*, 3:283–299.

[Belmonte-Rivas et al., 2014] Belmonte-Rivas, M., Veefkind, P., Boersma, F., Levelt, P., Eskes, H., and Gille, J. (2014). Intercomparison of daytime stratospheric NO_2 satellite retrievals and model simulations. *Atmos. Meas. Tech.*, 7:2203–2225.

[Bernath et al., 2005] Bernath, P. F., McElroy, C. T., Abrams, M. C., Boone, C. D., and M. Butler, e. a. (2005). Atmospheric Chemistry Experiment (ACE): Mission overview. *Geophys. Res. Lett.*, 32(L15S01):5 pp.

[Boersma et al., 2002] Boersma, K. F., Bucsela, E., Brinksma, E., and Gleason, J. F. (2002). NO_2. In *OMI Algorithm Theoretical Basis Document – Vol. 4: OMI Trace Gas Algorithms, ATBD-OMI-02 Vers. 2.0*, pages 13–28. NASA Goddard Space Flight Cent., Greenbelt, Md.

[Boersma et al., 2004] Boersma, K. F., Eskes, H. J., and Brinksma, E. J. (2004). Error analysis for tropospheric NO_2 retrieval from space. *J. Geophys. Res.*, 109(D04311):20 pp.

[Boersma et al., 2011] Boersma, K. F., Eskes, H. J., Dirksen, R. J., Van der A, R. J., Veefkind, J. P., Stammes, P., Huijnen, V., Kleipool, Q. L., Sneep, M., Claas, J., Leitão, J., Richter, A., Zhou, Y., and Brunner, D. (2011). An improved retrieval of tropospheric NO_2 columns from the Ozone Monitoring Instrument. *Atmos. Meas. Tech.*, 4:1905–1928.

[Boersma et al., 2018] Boersma, K. F., Eskes, H. J., Richter, A., Smedt, I. D., Lorente, A., Beirle, S., van Geffen, J. H. G. M., Zara, M., Peters, E., Roozendael, M. V., Wagner, T., Maasakkers, J. D., van der A, R. J., Nightingale, J., Rudder, A. D., Irie, H., Pinardi, G., Lambert, J.-C., and Compernolle4, S. (2018). 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. Discuss.*, in review.

[Boersma et al., 2007] Boersma, K. F., Eskes, H. J., Veefkind, J. P., Brinksma, E. J., Van der A, R. J., Sneep, M., Van den Oord, G. H. J., Levelt, P. F., Stammes, P., F., G. J., and Bucsela, E. J. (2007). Near-real time retrieval of tropospheric NO_2 from OMI. *Atmos. Chem. Phys.*, 7:2013–2128.

[Boersma et al., 2009] Boersma, K. F., Jacob, D. J., Trainic, M., Rudich, Y., DeSmedt, I., Dirksen, R., and Eskes, H. J. (2009). Validation of urban NO_2 concentrations and their diurnal and seasonal variations observed from the SCIAMACHY and OMI sensors using in situ surface measurements in israeli cities. *Atmos. Chem. Phys.*, 9:3867–3879.

[Boersma et al., 2016] Boersma, K. F., Vinken, G. C. M., and Eskes, H. J. (2016). Representativeness errors in comparing chemistry transport and chemistry climate models with satellite UV-Vis tropospheric column retrievals. *Geosci. Model Dev.*, 9:875–898.

[Bovensmann et al., 1999] Bovensmann, H., Burrows, J. P., Buchwitz, M., Frerick, J., Noel, S., Rozanov, V. V., Chance, K. V., and Goede, A. P. H. (1999). SCIAMACHY: mission objectives and measurement modes. *J. Atmos. Sci.*, 56:127–150.

[Bucsela et al., 2006] Bucsela, E. J., Celarier, E. A., Wenig, M. O., Gleason, J. F., Veefkind, J. P., Boersma, K. F., and Brinksma, E. J. (2006). Algorithm for NO_2 vertical column retrieval from the ozone monitoring instrument. *IEEE Trans. Geosci. Rem. Sens.*, 44:1245–1258.

[Bucsela et al., 2013] Bucsela, E. J., Krotkov, N. A., Celarier, E. A., Lamsal, L. N., Swartz, W. H., Bhartia, P. K., Boersma, K. F., Veefkind, J. P., Gleason, J. F., and Pickering, K. E. (2013). A new stratospheric and tropospheric NO_2 retrieval algorithm for nadir-viewing satellite instruments: applications to OMI. *Atmos. Meas. Tech.*, 6:2607–2626.

[Burrows et al., 1999] Burrows, J. P., Weber, M., Buchwitz, M., Rozanov, V., Ladstätter-Weißenmayer, A., Richter, A., Debeek, R., Hoogen, R., Bramstedt, K., Eichmann, K.-U., Eisinger, M., and Perner, D. (1999). The Global Ozone Monitoring Experiment (GOME): Mission concept and first results. *J. Atmos. Sci.*, 56:151–175.

[Castellanos and Boersma, 2012] Castellanos, P. and Boersma, K. F. (2012). Reductions in nitrogen oxides over Europe driven by environmental policy and economic recession. *Scientific Reports*, 2:7 pp.

[Chance and Spurr, 1997] Chance, K. V. and Spurr, R. J. D. (1997). Ring effect studies: Rayleigh scattering, including molecular parameters for rotational Raman scattering and the Fraunhofer spectrum. *Appl. Opt.*, 36(21):5224–5230.

[Chandrasekhar, 1960] Chandrasekhar, S. (1960). *Radiative Transfer*. Dover Publications, New York.

[Chu and McCormick, 1986] Chu, W. P. and McCormick, M. P. (1986). Sage observations of strastospheric nitrogen dioxide. *J. Geophys. Res.*, 91(D5):5465–5476.

[Crutzen, 1970] Crutzen, P. J. (1970). The influence of nitrogen oxides on the atmospheric ozone content. *Quart. J. R. Meteorol. Soc.*, 96:320–325.

[De Haan et al., 1987] De Haan, J. F., Bosma, P. B., and Hovenier, J. W. (1987). The adding method for multiple scattering in a non-homogeneous Rayleigh atmosphere. *Astron. & Astroph.*, 183:371–391.

[De Ruyter de Wildt et al., 2012] De Ruyter de Wildt, M., Eskes, H., and Boersma, K. F. (2012). The global economic cycle and satellite-derived NO_2 trends over shipping lanes. *Geophys. Res. Lett.*, 39(L01802):6 pp.

[Dentener et al., 2006] Dentener, F., Drevet, J., Lamarque, J. F., Bey, I., Eickhout, B., Fiore, A. M., Hauglustaine, D., Horowitz, L. W., Krol, M., Kulshrestha, U. C., Lawrence, M., Galy-Lacaux, C., Rast, S., Shindell, D., Stevenson, D., Noije, T. V., Atherton, C., Bell, N., Bergman, D., Butler, T., Cofala, J., Collins, B., Doherty, R., Ellingsen, K., Galloway, J., Gauss, M., Montanaro, V., Müller, J. F., Pitari, G., Rodriguez, J., Sanderson, M., Solmon, F., Strahan, S., Schultz, M., Sudo, K., Szopa, S., and Wild, O. (2006). Nitrogen and sulfur deposition on regional and global scales: A multimodel evaluation. *Global Biogeochem. Cycles*, 20(GB4003):21 pp.

[Dentener et al., 2003] Dentener, F., Van Weele, M., Krol, M., Houweling, S., and van Velthoven, P. (2003). Trends and inter-annual variability of methane emissions derived from 1979-1993 global CTM simulations. *Atmos. Chem. Phys.*, 3:73–88.

[Dirksen et al., 2011] Dirksen, R. J., Boersma, K. F., Eskes, H. J., Ionov, D. V., Bucsela, E. J., Levelt, P. F., and Kelder, H. M. (2011). Evaluation of stratospheric NO_2 retrieved from the Ozone Monitoring Instrument: Intercomparison, diurnal cycle, and trending. *J. Geophys. Res.*, 116(D08305):22 pp.

[Eskes and Boersma, 2003] Eskes, H. J. and Boersma, K. F. (2003). Averaging kernels for DOAS total-column satellite retrievals. *Atmos. Chem. Phys.*, 3:1285–1291.

[Eskes et al., 2003] Eskes, H. J., Van Velthoven, P. F. J., Valks, P., and Kelder, H. M. (2003). Assimilation of GOME total ozone satellite observations in a three-dimensional tracer transport model. *Q. J. R. Meteorol. Soc.*, 129(590):1663–1681.

[Fuglestvedt et al., 1999] Fuglestvedt, J. S., Berntsen, T., Isaksen, I. S. A., Mao, H., Liang, X.-Z., and Wang, W.-C. (1999). Climatic forcing of nitrogen oxides through changes in tropospheric ozone and methane. *Atmos. Environ.*, 33(3):961–977.

[Gordley et al., 1996] Gordley, L. L., III, J. M. R., Mickley, L. J., Frederick, J. E., Park, J. H., Stone, K. A., Beaver, G. M., McInerney, J. M., Deaver, L. E., Toon, G. C., Murcray, F. J., Blatherwick, R. D., Gunson, M. R., Abbatt, J. P. D., III, R. L. M., Mount, G. H., Sen, B., and Blavier, J.-F. (1996). Validation of nitric oxide and nitrogen dioxide measurements made by the Halogen Occultation Experiment for UARS platform. *J. Geophys. Res.*, 101(D6):10241–10266.

[Gorshelev et al., 2014] Gorshelev, V., Serdyuchenko, A., Weber, M., and Burrows, J. P. (2014). High spectral resolution ozone absorption cross-sections: Part I. Measurements, data analysis and comparison around 293K. *Atmos. Meas. Tech.*, 7:609–624.

[Grooß and Russel, 2005] Grooß, J.-U. and Russel, J. M. (2005). Technical note: A stratospheric climatology for O3, H2O, CH4, NOx, HCl and HF derived from HALOE measurements. *Atmos. Chem. Phys*, 5:2797–2807.

[Gruzdev and Elokhov, 2009] Gruzdev, A. N. and Elokhov, A. S. (2009). Validating NO_2 measurements in the vertical atmospheric column with the OMI instrument aboard the EOS Aura satellite against ground-based measurements at the Zvenigorod Scientific Station. *Izv. Atmos. Oceanic Phys.*, 45(4):444–455.

[Hains et al., 2010] Hains, J. C., Boersma, K. F., Kroon, M., Dirksen, R. J., Cohen, R. C., Perring, A. E., Bucsela, E., Volten, H., Swart, D. P. J., Richter, A. andWittrock, F., Schoenhardt, A., Wagner, T., Ibrahim, O. W., Roozendael, V., M., Pinardi, G., Gleason, J. F., Veefkind, J. P., and Levelt, P. (2010). Testing and improving OMI DOMINO tropospheric NO_2 using observations from the DANDELIONS and INTEX-B validation campaigns. *J. Geophys. Res.*, 115(D05301):20 pp.

[Hendrick et al., 2012] Hendrick, F., Mahieu, E., Bodeker, G. E., Boersma, K. F., Chipperfield, M. P., De Mazière, M., De Smedt, I., Demoulin, P., Fayt, C., Hermans, C., Kreher, K., Lejeune, B., Pinardi, G., Servais, C., Stübi, R., Van der A, R., Vernier, J.-P., and Van Roozendael, M. (2012). Analysis of stratospheric NO_2 trends above Jungfraujoch using ground-based UV-visible, FTIR, and satellite nadir observations. *Atmos. Chem. Phys.*, 12:8851–8864.

[Hilboll et al., 2013a] Hilboll, A., Richter, A., and Burrows, J. P. (2013a). Long-term changes of tropospheric NO_2 over megacities derived from multiple satellite instruments. *Atmos. Chem. Phys.*, 13:4145–4169.

[Hilboll et al., 2013b] Hilboll, A., Richter, A., Rozanov, A., Hodnebrog, O., Heckel, A., Solberg, S., Stordal, F., and Burrows, J. P. (2013b). Improvements to the retrieval of tropospheric NO_2 from satellite – stratospheric correction using SCIAMACHY limb/nadir matching and comparison to Oslo CTM2 simulations. *Atmos. Meas. Tech.*, 6:565–584.

[Huijnen et al., 2010a] Huijnen, V., Eskes, H. J., Poupkou, A., Elbern, H., Boersma, K. F., Foret, G., Sofiev, M., Valdebenito, A., Flemming, J., Stein, O., Gross, A., Robertson, L., D'Isidoro, M., Kioutsioukis, I., Friese, E., Amstrup, B., Bergstrom, R., Strunk, A., Vira, J., Zyryanov, D., Maurizi, A., Melas, D., Peuch, V.-H., and Zerefos, C. (2010a). Comparison of OMI NO_2 tropospheric columns with an ensemble of global and European regional air quality models. *Atmos. Chem. Phys.*, 10:3273–3296.

[Huijnen et al., 2010b] Huijnen, V., Williams, J., Van Weele, M., Van Noije, T., Krol, M., Dentener, F., Segers, A., Houweling, S., Peters, W., De Laat, J., Boersma, F., Bergamaschi, P., Van Velthoven, P., Le Sager, P., Eskes, H., Alkemade, F., Scheele, R., Nédélec, P., and Pätz, H.-W. (2010b). The global chemistry transport model tm5: description and evaluation of the tropospheric chemistry version 3.0. *Geosc. Model Dev.*, 3(2):445–473.

[Ingmann et al., 2012] Ingmann, P., Veihelmann, B., Langen, J., Lamarre, D., Stark, H., and Courrèges-Lacoste, G. B. (2012). Requirements for the GMES Atmosphere Service and ESA's implementation concept: Sentinels-4/-5 and -5p. *Rem. Sens. Environment*, 120:58–69.

[Irie et al., 2012] Irie, H., Boersma, K. F., Kanaya, Y., Takashima, H., Pan, X., and Wang, Z. F. (2012). First quantitative bias estimates for tropospheric NO_2 columns retrieved from SCIAMACHY, OMI, and GOME-2 using a common standard. *Atmos. Meas. Tech.*, 5:2403–2411.

[Jacob, 1999] Jacob, D. J. (1999). *Introduction to Atmospheric Chemistry*. Princeton University Press.

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

[Koelemeijer et al., 2001] Koelemeijer, R. B. A., Stammes, P., W., H. J., and De Haan, J. F. (2001). A fast method for retrieval of cloud parameters using oxygen A band measurements from the Global Ozone Monitoring Experiment. *J. Geophys. Res.*, 106(D4):3475–3490.

[Lampel et al., 2015] Lampel, J., Frieß, U., and U., P. (2015). The impact of vibrational raman scattering of air on doas measurements of atmospheric trace gases. *Atmos. Meas. Tech.*, 8:3767–3787.

[Lamsal et al., 2010] Lamsal, L. N., Martin, R. V., Van Donkelaar, A., Celarier, E. A., Bucsela, E. J., Boersma, K. F., Dirksen, R., Luo, C., and Wang, Y. (2010). Indirect validation of tropospheric nitrogen dioxide retrieved from the OMI satellite instrument: Insight into the seasonal variation of nitrogen oxides at northern midlatitudes. *J. Geophys. Res.*, 115(D05302):15 pp.

[Leitão et al., 2010] Leitão, J., Richter, A., Vrekoussis, M., Kokhanovsky, A., Zhang, Q., Beekmann, M., and Burrows, J. P. (2010). On the improvement of NO_2 satellite retrievals – aerosol impact on the airmass factors. *Atmos. Meas. Tech.*, 3:475–493.

[Lerot et al., 2010] Lerot, C., Stavrakou, T., De Smedt, I., Müller, J.-F., and Van Roozendael, M. (2010). Glyoxal vertical columns from GOME-2 backscattered light measurements and comparisons with a global model. *Atmos. Chem. Phys.*, 10:12059–12072.

[Leue et al., 2001] Leue, C., Wenig, M., Wagner, T., Klimm, O., Platt, U., and Jähne, B. (2001). Quantitative analysis of NO_x emissions from Global Ozone Monitoring Experiment satellite image sequences. *J. Geophys. Res.*, 106(D6):5493–5505.

[Levelt et al., 2006] Levelt, P. F., van den Oord, G. H. J., Dobber, M. R., Mälkki, A., Visser, H., de Vries, J., Stammes, P., Lundell, J. O. V., and Saari, H. (2006). The Ozone Monitoring Instrument. *IEEE Trans. Geosci. Rem. Sens.*, 44:1093–1101.

[Liley et al., 2000] Liley, J. B., Johnston, P. V., McKenzie, R. L., Thomas, A. J., and Boyd, I. S. (2000). Stratospheric NO_2 variations from a long time series at Lauder, New Zealand. *J. Geophys. Res.*, 105(D9):11,633–11,640.

[Lin et al., 2014] Lin, J., Martin, R. V., Boersma, K. F., Sneep, M., Stammes, P., Spurr, R., Wang, P., Roozendael, M. V., Clémer, K., and Irie, H. (2014). Retrieving tropospheric nitrogen dioxide over China from the Ozone Monitoring Instrument: Effects of aerosols, surface reflectance anisotropy and vertical profile of nitrogen dioxide. *Atmos. Chem. Phys.*, 14:1441–1461.

[Lin et al., 2010] Lin, J.-T., McElroy, M. B., and Boersma, K. F. (2010). Constraint of anthropogenic NO_x emissions in china from different sectors: a new methodology using multiple satellite retrievals. *Atmos. Chem. Phys.*, 10:63–78.

[Liu et al., 2018] Liu, S., Valks, P., Pinardi, G., De Smedt, I., Yu, H., Beirle, S., and Richter, A. (2018). An Improved Total and Tropospheric NO2 Column Retrieval for GOME-2. *Atmos. Meas. Tech. Discuss.*, in review.

[Llewellyn et al., 2004] Llewellyn, E. J., Lloyd, N. D., Degenstein, D. A., Gattinger, R. L., and S. V. Petelina, e. a. (2004). The OSIRIS instrument on the Odin spacecraft. *Canadian Journal of Physics*, 82(6):411–422.

[Lorente et al., 2017] Lorente, A., Folkert Boersma, K., Yu, H., Dörner, S., Hilboll, A., Richter, A., Liu, M., Lamsal, L. N., Barkley, M., De Smedt, I., Van Roozendael, M., Wang, Y., Wagner, T., Beirle, S., Lin, J.-T., Krotkov, N., Stammes, P., Wang, P., Eskes, H. J., and Krol, M. (2017). Structural uncertainty in air mass factor calculation for NO2 and HCHO satellite retrievals. *Atmos. Meas. Tech.*, 10:759–782.

[Ma et al., 2013] Ma, J. Z., Beirle, S., Jin, J. L., Shaiganfar, R., Yan, P., and Wagner, T. (2013). Tropospheric NO_2 vertical column densities over Beijing: results of the first three years of ground-based MAX-DOAS measurements (2008-2011) and satellite validation. *Atmos. Chem. Phys.*, 13:1547–1567.

[Maasakkers et al., 2013] Maasakkers, J. D., Boersma, K. F., Williams, J. E., Van Geffen, J., Vinken, G. C. M., Sneep, M., Hendrick, F., Van Roozendael, M., and Veefkind, J. P. (2013). Vital improvements to the retrieval of tropospheric NO_2 columns from the Ozone Monitoring Instrument. *Geophys. Res. Abstracts*, 15(EGU2013-714):1. EGU General Assembly 2013.

[Martin et al., 2002] Martin, R. V., Chance, K., Jacob, D. J., Kurosu, T. P., Spurr, R. J. D., Bucsela, E., Gleason, J. F., Palmer, P. I., Bey, I., Fiore, A. M., Li, Q., Yantosca, R. M., and Koelemeijer, R. B. A. (2002). An improved retrieval of tropospheric nitrogen dioxide from gome. *J. Geophys. Res.*, 107(D20):4437–4457.

[Merlaud et al., 2012] Merlaud, A., Van Roozendael, M., Van Gent, J., Fayt, C., Maes, J., Toledo-Fuentes, X., Ronveaux, O., and De Maziére, M. (2012). DOAS measurements of NO_2 from an ultralight aircraft during the Earth Challenge expedition. *Atmos. Meas. Tech.*, 5:2057–2068.

[Mijling and Van der A, 2012] Mijling, B. and Van der A, R. J. (2012). Using daily satellite observations to estimate emissions of short-lived air pollutants on a mesoscopic scale. *J. Geophys. Res.*, 117(D17302):20 pp.

[Mount et al., 1984] Mount, G. H., Rusch, D. W., Noxon, J. F., Zawodny, J. M., and Barth, C. A. (1984). Measurements of stratospheric NO_2 from the Solar Mesosphere Explorer Satellite: 1. An overview of the results. *J. Geophys. Res.*, 89(D1):1327–1340.

[Munro et al., 2006] Munro, R., Eisinger, M., Anderson, C., Callies, J., Corpaccioli, E., Lang, R., Lefebvre, A., Livschitz, Y., and Albinana, A. P. (2006). GOME-2 on MetOp. In *Proceedings of the Atmospheric Science Conference 2006*, SP 628. ESA, ESA, Paris.

[Murphy et al., 1993] Murphy, D. M., Fahey, D. W., Proffitt, M. H., Liu, S. C., Chan, K. R., Eubank, C. S., Kawa, S. R., and Kelly, K. K. (1993). Reactive nitrogen and its correlation with ozone in the lower stratosphere and upper troposphere. *J. Geophys. Res.*, 98(D5):8751–8773.

[Nowlan et al., 2016] Nowlan, C. R., Liu, X., Leitch, J. W., Chance, K., González Abad, G., Liu, C., Zoogman, P., Cole, J., Delker, T., Good, W., Murcray, F., Ruppert, L., Soo, D., Follette-Cook, M. B., Janz, S. J., Kowalewski, M. G., Loughner, C. P., Pickering, K. E., Herman, J. R., Beaver, M. R., Long, R. W., Szykman, J. J., Judd, L. M., Kelley, P., Luke, W. T., Ren, X., and Al-Saadi, J. A. (2016). Nitrogen dioxide observations from the Geostationary Trace gas and Aerosol Sensor Optimization (GeoTASO) airborne instrument: Retrieval algorithm and measurements during DISCOVER-AQ Texas 2013. *Atmos. Meas. Tech.*, 9:2647–2668.

[Ott et al., 2010] Ott, L. E., Pickering, K. E., Stenchikov, G. L., Allen, D. J., DeCaria, A. J., Ridley, B., Lin, R.-F., Lang, S., and Tao, W.-K. (2010). Production of lightning NO and its vertical distribution calculated from three-dimensional cloud-scale chemical transport model simulations. *J. Geophys. Res.*, 115(D04301):19 pp.

[Palmer et al., 2001] Palmer, P. I., Jacob, D., Chance, K., Martin, R. V., Spurr, R. J. D., Kurosu, T. P., Bey, I., Yantosca, R., Fiore, A., and Li, Q. (2001). Air-mass factor formulation for spectroscopic measurements from satellite: applications to fomaldehyde retrievals from the Global Ozone Monitoring Experiment. *J. Geophys. Res.*, 106:14539–14550.

[Platt, 1994] Platt, U. (1994). Differential Optical Absorption Spectroscopy (DOAS). *Air monitoring by spectroscopic techniques, Chem. Anal.*, 127:27–76. edited by M.W. Sigrist.

[Platt and Stutz, 2008] Platt, U. and Stutz, Z. (2008). *Differential Optical Absorption Spectroscopy, Principles and Applications*. Springer, Heidelberg, Germany.

[Pope and Fry, 1997] Pope, R. M. and Fry, E. S. (1997). Absorption spectrum (380-700 nm) of pure water. II. Integrating cavity measurements. *Appl. Opt.*, 36(33):8710–8723.

[Randall et al., 1998] Randall, C. E., Rusch, D. W., Bevilacqua, R. M., Hoppel, K. W., and Lumpe, J. D. (1998). Polar Ozone and Aerosol Measurement (POAM) II stratospheric NO_2 1993-1996. *J. Geophys. Res.*, 103(D21):28361–28371.

[Ravishankara et al., 2009] Ravishankara, A. R., Daniel, J. S., and Portmann, R. W. (2009). Nitrous Oxide (N2O): The Dominant Ozone-Depleting Substance Emitted in the 21st Century. *Science*, 326(5949):123–125.

[Richter et al., 2011] Richter, A., Begoin, M., Hilboll, A., and Burrows, J. P. (2011). An improved NO_2 retrieval for the GOME-2 satellite instrument. *Atmos. Meas. Tech.*, 4:1147–1159.

[Richter and Burrows, 2002] Richter, A. and Burrows, J. P. (2002). Tropospheric NO_2 from GOME measurements. *Adv. Space Res.*, 29(11):1673–1683.

[Schaub et al., 2007] Schaub, D., Brunner, D., Boersma, K. F., Keller, J., Folini, D., Buchmann, B., Berresheim, H., and Staehelin, J. (2007). SCIAMACHY tropospheric NO_2 over Switzerland: estimates of NO_x lifetimes and impact of the complex Alpine topography on the retrieval. *Atmos. Chem. Phys.*, 7:5971–5987.

[Seinfeld and Pandis, 2006] Seinfeld, J. H. and Pandis, S. N. (2006). *Atmospheric Chemistry and Physics - From Air Pollution to Climate Change (2nd Edition)*. John Wiley & Sons.

[Serdyuchenko et al., 2014] Serdyuchenko, A., Gorshelev, V., Weber, M., Chehade, W., and Burrows, J. P. (2014). High spectral resolution ozone absorption cross-sections: Part II. Temperature dependence. *Atmos. Meas. Tech.*, 7:625–636.

[Shindell et al., 2009] Shindell, D. T., Faluvegi, G., Koch, D. M., Schmidt, G. A., Unger, N., and Bauer, S. E. (2009). Improved attribution of climate forcing to emissions. *Science*, 326(5953):716–718.

[Sierk et al., 2006] Sierk, B., Richter, A., Rozanov, A., Von Savigny, C., Schmoltner, A. M., Buchwitz, M., Bovensmann, H., and Burrows, J. P. (2006). Retrieval and monitoring of atmospheric trace gas concentrations in nadir and limb geometry using the space-borne SCIAMACHY instrument. *Environmental Monitoring and Assessment*, 120:65–73.

[Sillman et al., 1990] Sillman, S., Logan, J. A., and Wofsy, S. C. (1990). The sensitivity of ozone to nitrogen oxides and hydrocarbons in regional ozone episodes. *J. Geophys. Res.*, 95(D2):1837–1851.

[Silver et al., 2013] Silver, J. D., Brandt, J., Hvidberg, M., Frydendall, J., and Christensen, J. H. (2013). Assimilation of OMI NO_2 retrievals into the limited-area chemistry-transport model DEHM (V2009.0) with a 3-D OI algorithm. *Geosc. Model Dev.*, 6:1–16.

[Solomon, 1999] Solomon, S. (1999). Stratospheric ozone depletion: A review of concepts and history. *Rev. Geophys.*, 37(3):275–316.

[Stammes, 2001] Stammes, P. (2001). Spectral radiance modeling in the UV-visible range. In Smith, W. and Timofeyev, Y., editors, *IRS 2000: Current Problems in Atmospheric Radiation*, pages 385–388. A. Deepak, Hampton, Va.

[Stammes et al., 2008] Stammes, P., Sneep, M., De Haan, J. F., Veefkind, J. P., Wang, P., and Levelt, P. F. (2008). Effective cloud fractions from the Ozone Monitoring Instrument: Theoretical framework and validation. *J. Geophys. Res.*, 113(D16S38):12 pp.

[Stavrakou et al., 2008] Stavrakou, T., Müller, J.-F., Boersma, K. F., De Smedt, I., and Van der A, R. J. (2008). Assessing the distribution and growth rates of NO_2 emission sources by inverting a 10-year record of NO_2 satellite columns. *Geophys. Res. Lett.*, 35(L10801):5 pp.

[Tack et al., 2018] Tack, F., Merlaud, A., Meier, A. C., Vlemmix, T., Ruhtz, T., Iordache, M.-D., Ge, X., van der Wal, L., Schuettemeyer, D., Ardelean, M., Calcan, A., Schönhardt, A., Meuleman, K., Richter, A., and Van Roozendael, M. (2018). Intercomparison of four airborne imaging DOAS systems for tropospheric NO_2 mapping – The AROMAPEX campaign . *Atmos. Meas. Tech. Discuss.*, in review.

[Thalman and Volkamer, 2013] Thalman, R. and Volkamer, R. (2013). Temperature dependant absorption cross-sections of O2-O2 collision pairs between 340 and 630 nm at atmospherically relevant pressure. *Phys. Chem. Chem. Phys.*, 15:15371–15381.

[Tilstra et al., 2017] Tilstra, L. G., Tuinder, O. N. E., Wang, P., and Stammes, P. (2017). Surface reflectivity climatologies from UV to NIR determined from Earth observations by GOME-2 and SCIAMACHY. *J. Geophys. Res.*, 122:4084–4111.

[Valks et al., 2011] Valks, P., Pinardi, G., Richter, A., Lambert, J.-C., Hao, N., Loyola, D., Van Roozendael, M., and Emmadi, S. (2011). Operational total and tropospheric NO_2 column retrieval for GOME-2. *Atmos. Meas. Tech.*, 4:1491–1514.

[Van der A et al., 2015] Van der A, R. J., Allaart, M. A. F., and Eskes, H. J. (2015). Extended and refined multi sensor reanalysis of total ozone for the period 1970-2012. *Atmos. Meas. Tech.*, 8:3021–3035.

[Van der A et al., 2008] Van der A, R. J., Eskes, H. J., Boersma, K. F., Van Noije, T. P. C., Van Roozendael, M., De Smedt, I., Peters, D. H. M. U., Kuenen, J. J. P., and Meijer, E. W. (2008). Trends, seasonal variability and dominant NO_x source derived from a ten year record of NO_2 measurd from space. *J. Geophys. Res.*, 113(D04302):12 pp.

[Van Diedenhoven et al., 2007] Van Diedenhoven, B., Hasekamp, O. P., and Landgraf, J. (2007). Retrieval of cloud parameters from satellite-based reflectance measurements in the ultraviolet and the oxygen A-band. *J. Geophys. Res.*, 112(D15208):15 pp.

[Van Geffen et al., 2015] Van Geffen, J. H. G. M., Boersma, K. F., Van Roozendael, M., Hendrick, F., Mahieu, E., De Smedt, I., M., S., and Veefkind, J. P. (2015). Improved spectral fitting of nitrogen dioxide from OMI in the 405 – 465 nm window. *Atmos. Meas. Tech.*, 8:1685–1699.

[Van Roozendael et al., 2006] Van Roozendael, M., Loyola, D., Spurr, R., Balis, D., Lambert, J.-C., Livschitz, Y., Valks, P., Ruppert, T., Kenter, P., Fayt, C., and Zehner, C. (2006). Ten years of GOME/ERS-2 total ozone data – The new GOME data processor (GDP) version 4: 1. Algorithm description. *J. Geophys. Res.*, 111(D14311):21 pp.

[Vandaele et al., 1998] Vandaele, A. C., Hermans, C., Simon, P. C., Carleer, M., Colin, R., Fally, S., Mérienne, M. F., Jenouvrier, A., and Coquart, B. (1998). 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*, 59:171–184.

[Vandevender and Haskell, 1982] Vandevender, W. H. and Haskell, K. H. (1982). The SLATEC mathematical subroutine library. *ACM SIGNUM Newsletter*, 17.3:16–21.

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

[Vasilkov et al., 2002] Vasilkov, A. P., Joiner, J., Gleason, J., and Bhartia, P. K. (2002). Ocean Raman scattering in satellite backscatter UV measurements. *Geophys. Res. L*, 29(17):1837–1840.

[Veefkind et al., 2012] Veefkind, J. P., Aben, I., McMullan, K., Förster, H., De Vries, J., Otter, G., Claas, J., Eskes, H. J., De Haan, J. F., Kleipool, Q., Van Weele, M., Hasekamp, O., Hoogeveen, R., Landgraf, J., Snel, R., Tol, P., Ingmann, P., Voors, R., Kruizinga, B., Vink, R., Visser, H., and Levelt, P. F. (2012). TROPOMI on the ESA Sentinel-5 Precursor: A GMES mission for global observations of the atmospheric composition for climate, air quality and ozone layer applications. *Rem. Sens. Environment*, 120:70–83.

[Veefkind et al., 2016] Veefkind, J. P., De Haan, J. F., Sneep, M., and Levelt, P. (2016). Improvements of the OMI O2-O2 operational cloud algorithm and comparisons with ground-based radar-lidar observations. *Atmos. Meas. Techn.*, 9:6035–6049.

[Vermote and Tanré, 1992] Vermote, E. and Tanré, D. (1992). Analytic expressions for radiative properties of planar Rayleigh scattering media, including polarization contributions. *J. Quant. Spectrosc. & Radiat. Transfer*, 47:305–314.

[Vountas et al., 2003] Vountas, M., Richter, A., Wittrock, F., and Burrows, J. P. (2003). Inelastic scattering in ocean water and its impact on trace gas retrievals from satellite data. *Atmos. Chem. Phys.*, 3:1365–1375.

[Wang et al., 2008] Wang, P., Stammes, P., Van der A, R., Pinardi, G., and Van Roozendael, M. (2008). FRESCO+: an improved O_2 A-band cloud retrieval algorithm for tropospheric trace gas retrievals. *Atmos. Chem. Phys.*, 8:6565–6576.

[Wang et al., 2012] Wang, S. W., Zhang, Q., Streets, D. G., He, K., Martin, R. V., Lamsal, L. N., Chen, D., Lei, Y., and Lu, Z. (2012). Growth in NO_x emissions from power plants in China: bottom-up estimates and satellite observations. *Atmos. Chem. Phys.*, 12:4429–4447.

[Wenig et al., 2003] Wenig, M., Spichtinger, N., Stohl, A., Held, G., Beirle, S., Wagner, T., Jähne, B., and Platt, U. (2003). Intercontinental transport of nitrogen oxide pollution plumes. *Atmos. Chem. Phys.*, 3:387–393.

[Williams et al., 2017] Williams, J. E., Boersma, K. F., Le Sager, P., and Verstraeten, W. W. (2017). The high-resolution version of TM5-MP for optimized satellite retrievals: description and validation. *Geosci. Model Dev.*, 10:721–750.

[World Health Organisation, 2003] World Health Organisation (2003). *Health Aspects of Air Pollution with Particulate Matter, Ozone and Nitrogen Dioxide*. World Health Organisation, Bonn.

[Yang et al., 2014] Yang, K., Carn, S. A., Ge, C., Wang, J., and Dickerson, R. R. (2014). Advancing measurements of tropospheric NO2 from space: New algorithm and first global results from OMPS. *Geophys. Res. Lett.*, 41:4777–4786.

[Zara et al., 2018] Zara, M., Boersma, K. F., De Smedt, E., Richter, A., Peters, E., Van Geffen, J. H. G. M., Beirle, S., Wagner, T., Van Roozendael, M., Marchenko, S., Lamsal, L. N., and Eskes, H. J. (2018). 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.*, 11:4033–4058.

[Zhou et al., 2009] Zhou, Y., Brunner, D., Boersma, K. F., Dirksen, R., and Wang, P. (2009). An improved tropospheric NO_2 retrieval for omi observations in the vicinity of mountainous terrain. *Atmos. Meas. Tech.*, 2:401–416.

[Zhou et al., 2010] Zhou, Y., Brunner, D., Spurr, R. J. D., Boersma, K. F., Sneep, M., Popp, C., and Buchmann, B. (2010). Accounting for surface reflectance anisotropy in satellite retrievals of tropospheric NO_2. *Atmos. Meas. Tech.*, 3:1185–1203.

[Zoogman et al., 2017] Zoogman, P., Liu, X., Suleiman, R. M., Pennington, W. F., and Flittner, D. E., e. a. (2017). Tropospheric emissions: Monitoring of pollution (TEMPO). *J. Quant. Spectrosc. Radiat. Transf.*, 186:17–39.