Improving algorithms and uncertainty estimates for satellite NO₂ retrievals: Results from the Quality Assurance for Essential Climate Variables (QA4ECV) project

5 K. Folkert Boersma[1,2], Henk J. Eskes[1], Andreas Richter[3], Isabelle De Smedt[4], Alba Lorente[2], Steffen Beirle[5], Jos H. G. M. van Geffen[1], Marina Zara[1], Enno Peters[3], Michel Van Roozendael[4], Thomas Wagner[5], Joannes D. Maasakkers[6], Ronald J. van der A[1], Joanne Nightingale[7], Anne De Rudder[4], Hitoshi Irie[8], Gaia Pinardi[4], Jean-Christopher Lambert[4], and Steven C. Compernolle[4]

10 [1]Royal Netherlands Meteorological Institute, Satellite Observations department, The Netherlands

[2]Wageningen University, Meteorology and Air Quality Group, Wageningen, the Netherlands

[3]Institute of Environmental Physics (IUP-UB), University of Bremen, Bremen, Germany

[4]Belgian Institute for Space Aeronomy (BIRA-IASB), Brussels, Belgium

[5]Max-Planck Institute for Chemistry (MPI-C), Mainz, Germany

15 [6]Harvard University, Cambridge, Massachusetts, United States

[7]National Physics Laboratory (NPL), Teddington, United Kingdom

[8]Center for Environmental Remote Sensing (CEReS), Chiba University, Chiba, Japan

20 *Correspondence to*: K. Folkert Boersma (boersma@knmi.nl)

Abstract. Global observations of tropospheric nitrogen dioxide (NO₂) columns have been shown to be feasible from space, but consistent multi-sensor records do not yet exist, nor are they covered by planned activities on the international level. Harmonised, multi-decadal records of NO₂ columns and their associated uncertainties can provide crucial information how the emissions and concentrations of 25 nitrogen oxides evolve over time. Here we describe the development of a new, community best practice NO₂ retrieval algorithm based on a synthesis of existing approaches. Detailed comparisons of these approaches led us to implement an enhanced spectral fitting method for NO₂, a 1°×1° TM5-MP data

assimilation scheme to estimate the stratospheric background, and improve air mass factor calculations. Guided by the needs expressed by data users, producers, and WMO GCOS guidelines, we incorporated detailed per-pixel uncertainty information in the data product, along with easily traceable information on the relevant quality aspects of the retrieval. We applied the improved QA4ECV NO_2 algorithm on

5   the most actual level-1 data sets to produce a complete 22-year data record that includes GOME (1995-2003), SCIAMACHY (2002-2012), GOME-2(A) (2007 onwards) and OMI (2004 onwards). The QA4ECV NO_2 spectral fitting recommendations and TM5-MP stratospheric column and air mass factor approach are currently also applied to S5P-TROPOMI. The uncertainties in the QA4ECV tropospheric NO_2 columns amount to typically 40% over polluted scenes. First validation results of the QA4ECV

10   OMI NO_2 columns and their uncertainties over Tai'an, China in June 2006 suggests little bias (-2%) and better precision than suggested by uncertainty propagation. We conclude that our improved QA4ECV NO_2 long-term data record is providing valuable information to quantitatively constrain emissions, deposition, and trends in nitrogen oxides on a global scale.

1 Introduction

Nitrogen oxides (NO_x = NO + NO_2) in the atmosphere have far-reaching effects on the Earth system. In the lower troposphere, nitrogen oxides promote the photochemical production of ozone (e.g. Liu et al. [1987]; Grewe et al. [2012]), whereas in the stratosphere, NO_x leads to the catalytic destruction of ozone, and the formation of reservoir species for halogens (e.g. Crutzen et al. [1970]). Nitrogen oxides

20   contribute to aerosol formation, and they are linked to the oxidizing efficiency of the troposphere via ozone which plays an important role in the formation of the hydroxyl radical (OH). NO_2 itself is only a weak greenhouse gas [Solomon et al., 1999], but has considerable relevance for radiative forcing because nitrogen oxides are important precursors of tropospheric ozone, aerosols, and OH. The net effect of nitrogen oxides on climate forcing is modeled to be negative, or 'cooling', with NO_x-driven

25   aerosol screening dominating over tropospheric ozone warming [Shindell et al., 2009]. In 2011, the World Meteorological Organization (WMO) Global Climate Observing System (GCOS) included NO_2 (together with SO_2, HCHO, and CO) in its Implementation Plan for the Global Observing System for

Climate in Support of the UNFCCC [WMO, 2011] "in recognition of the emission-based view on climate forcing of ozone and secondary aerosols, relevant for climate mitigation and important for processes". The formal attribution of NO_2 as precursor to the ozone and aerosols Essential Climate Variables, or ECVs [Bojinski et al., 2014], implies that the scientific community has committed itself to providing reliable, long-term measurement records of NO_2. Apart from its relevance to climate change, atmospheric nitrogen oxides are also important for the health of ecosystems and humans. Deposition of nitrogen to ecosystems may affect the structure and functioning of ecosystems (e.g. Galloway et al. [2003]). Recently, the World Health Organization stated that it is reasonable to infer that NO_2 has direct short-term health effects, such as airway inflammation and reductions in lung function [WHO, 2013], and a literature review of epidemiological studies over a wide geographic area by Hoek et al. [2013] showed that human mortality was significantly associated with long-term exposure to NO_2.

High-quality observations are needed to monitor the concentrations of nitrogen oxides in the atmosphere, both close to the ground, where NO_2 is relevant for deposition and health aspects, as well as aloft, where nitrogen oxides influence atmospheric chemistry and climate. Such measurements are useful for reanalysis studies (e.g. Inness et al. [2013]), contribute to documenting changes in NO_2 concentrations and NO_x emissions (e.g. Zhang et al., [2008]; Vinken et al. [2014]), and to attributing any such changes to their underlying causes (e.g. Verstraeten et al., [2015]; Xu et al., [2013]). This provides policy makers with options for decisions to counter environmental problems (e.g. Witman et al. [2014]). Measurements may also enhance the public's appreciation of the extent and scope of the problem of air pollution. In situ measurements of NO_x concentrations taken on the ground are representative for the quality of the air people breathe close to the measurement station. But such stations are relatively scarce in many countries and cannot provide spatio-temporal continuity on a global scale. Satellite observations on the other hand provide global coverage, thereby offering the unique opportunity to study spatial patterns and temporal variation in NO_2 pollution. For any type of measurement, it holds that they can only be used properly in science or as evidence-basis for policy decisions, if there is unequivocal confidence in the data sets, as well as a proper understanding of their limitations.

The EU Seventh Framework (FP7) project Quality Assurance for Essential Climate Variables (QA4ECV [2018], www.qa4ecv.eu) was designed to demonstrate how reliable climate data sets can be generated, along with detailed and traceable information on the quality of such data. Specifically, for NO_2, the goals of this project are: (1) to generate a multi-decadal (1995-2017) satellite data record of tropospheric and stratospheric NO_2 column densities based on calibrated satellite data and state-of-the-art retrievals, and (2) to provide fully traceable uncertainty metrics for this record, ready for ingestion in models or in other interpretation efforts. Obtaining global, long-term, and stable satellite observations with validated accuracy and precision is not straightforward. The GOME (1995-2003; Burrows et al. [1999]), SCIAMACHY (2002-2012; Bovensmann et al. [1999]), OMI (from 2004 onwards; Levelt et al. [2006]), and GOME-2A (from 2007 onwards; Munro et al. [2006]) instruments have been providing global observations of NO_2 over the last 22 years, but there are important differences in overpass time, instrumental artefacts (e.g. calibration and design differences), and signal-to-noise levels that need to be taken into account. To be used properly, the information content of the NO_2 products needs to be validated over a variety of regions, and users need guidance provided by well-established quality information to help them judge the fitness-for-purpose of the NO_2 products.

In this work, we demonstrate our approach to improve a retrieval algorithm and apply it to generate a multi-decadal record of NO_2 columns with a consortium of European retrieval groups. We follow the guidelines for the generation of ECV datasets from WMO [2010]. Our efforts are inspired by the QA4ECV project goals described above, but also by recent studies showing that there is still room for substantial improvement in all sub-steps of the retrieval (e.g. Richter et al. [2011]; Lin et al. [2014]; van Geffen et al. [2015]; Krotkov et al. [2016]), by the outcome of validation studies showing that various state-of-science retrievals have biases on the order of tens of percents (e.g. Jin et al. [2016]; Drosoglu et al. [2017]; Kollonige et al. [2018]), and the considerable structural uncertainty in retrieved tropospheric NO_2 columns emerging when different retrieval methodologies are applied to the exact same satellite observations (e.g. van Noije et al. [2006]; Lorente et al. [2017]). The efforts from five European retrieval groups within the QA4ECV consortium allow us to perform a detailed comparison of current approaches to various retrieval sub-steps. These comparisons have proven to be helpful in reducing and

better quantifying the uncertainty of the NO_2 retrieval. The improved quality of the QA4ECV NO_2 record itself, and the improved knowledge on the uncertainties, should make the QA4ECV satellite data record better fit for the purpose of trend analysis, data assimilation, and inverse modelling studies.

The manuscript is organised as follows: in section 2 we discuss how NO_2 data user requirements, the expertise from NO_2 data providers, and the quality requirements defined by GCOS are providing direction for this study. In section 3 we assess the quality of the best currently available level 1 data sets for NO_2 retrieval from GOME, SCIAMACHY, OMI, and GOME-2(A), and discuss how this guides the selection of spectral fitting approaches. Section 4 focuses on the algorithm design and the traceability of the retrieval approach and external data used. In section 5, we give an overview of the main lessons learned in the inter-comparisons of retrieval sub-steps. Section 6 summarizes the uncertainty information provided in the QA4ECV data product, and how these uncertainty estimates compare to the inter-comparison results from section 5. We conclude with a first validation of our new QA4ECV OMI NO_2 tropospheric columns and their uncertainties against independent MAX-DOAS measurements collected during a one-month campaign over Tai'an, China.

2 User needs and expert recommendations

2.1 User survey

At the start of the QA4ECV project, we have identified the requirements of data users in terms of uncertainty information and usability of the data product. This included a survey of 22 NO_2 data users, and interviews with 3 NO_2 'champion users', who provided more detailed written answers to questions. The questionnaire was aimed at establishing what users need in terms of quality flags, traceability information, and product uncertainty description. The main outcome of the survey for NO_2 is summarized in the Supplement. Briefly, users need detailed quality flags, specific information on random and systematic contributions to the uncertainties, traceability information on the product, and on validation of the product and algorithm. The full survey also includes results for the HCHO and CO data products and can be found in QA4ECV Deliverable 1.1 [Nightingale et al., 2015].

2.2 Producer requirements

We also carried out a survey of data producer requirements for quality assurance in satellite data records, and discussed retrieval priorities and quality assurance (QA) needs with retrieval experts from different groups within the consortium (BIRA-IASB, IUP Bremen, KNMI, Max Planck Institute for Chemistry, and Wageningen University in alphabetical order). Producers of data products (other than those involved in QA4ECV) that we interviewed recognized the need for processing chain information to be more transparent and more easily accessible for data users.

There was also a strong intrinsic motivation from NO_2 data producers to improve the retrieval algorithms and generate a long-term NO_2 data set from available satellite reflectance measurements. The NO_2 retrieval groups in the QA4ECV consortium discussed priorities for retrieval improvement, based on their collective experience with the retrieval, validation, and use of existing individual NO_2 data products for different sensors. The central idea was to arrive at a QA4ECV 'consortium algorithm' based on best practices derived from lessons learned from inter-comparisons between approaches for all relevant retrieval sub-steps and extend the steps initiated within the ESA S5P verification project [DLR, 2015].

2.3 QA4ECV consortium activities

Retrieval of tropospheric NO_2 columns is based on a 3-step approach. First, a set of absorption cross-sections, including NO_2, is fitted to the measured top-of-atmosphere reflectance spectrum, which provides the so-called slant column densities (SCDs, N_s). Then (step 2), the stratospheric contribution to the SCD ($N_{s,strat}$) is estimated and subtracted from the SCD. In the third step, the tropospheric air mass factor (or AMF, M_{trop}) is calculated based on knowledge of the satellite viewing conditions and assumptions on the state of the atmosphere in order to convert the residual tropospheric SCD into a tropospheric vertical column density VCD ($N_{v,trop}$). The retrieval equation is:

$$N_{v,trop} = \frac{N_s - N_{s,strat}}{M_{trop}} \tag{1}$$

The following activities leading to retrieval improvement were identified and conducted during the QA4ECV project:

(1) An inter-comparison of spectral fitting approaches between institutes. NO_2 SCDs were computed by all groups for the same orbits of level-1 data and results were compared. This resulted in a quantification of the level of agreement on the slant columns and a better understanding of the factors responsible for remaining differences. This is a relevant exercise in view of the substantial revisions of spectral fitting approaches over the last years (e.g. Richter et al. [2011]; van Geffen et al. [2015]; Marchenko et al. [2015]; Anand et al. [2015]), and resulted in the definition of the QA4ECV 'best practices' spectral fitting algorithm.

(2) An evaluation of the algorithm SCD uncertainties against an independent statistical uncertainty estimates [Zara et al., 2018].

(3) A comparison of stratospheric NO_2 fields and associated tropospheric residues from different approaches: consistency and plausibility checks, and quantification of differences. Recent improvements in the KNMI data assimilation approach [Maasakkers, 2013], and the newly developed STREAM scheme [Beirle et al., 2016], provided more insight in the stratospheric correction, and in the associated uncertainties.

(4) A comparison of altitude-dependent, or 'box' air mass factors (AMFs) for simplified scenarios. This comparison established the degree of consistency between radiative transfer models, pointed out discrepancies, and provided hints for possible improvements. The resulting spread between the (box) AMFs can be interpreted as the structural uncertainty[1] when using different radiative transfer models, vertical layering and interpolation schemes [Lorente et al., 2017].

(5) Comparisons of tropospheric AMFs calculated by different groups with an increasing number of differences in algorithm choices: from identical settings (wherein only model, vertical layering, and interpolation differ between groups), via preferred settings (every group using their own preferred information on clouds, albedo, NO_2 profile, etc.), to a wider 'round robin' comparison

---

[1]Structural uncertainty can be identified with the metrology concept "uncertainty of measurement method" (see the Guide to the expression of uncertainty in measurement [GUM, 2008], section F.2.5): *uncertainty associated with the method of*

wherein also groups outside of Europe participated. This last comparison was 'unguided', i.e. groups could freely decide how to calculate their AMFs, deciding for themselves whether to include aerosol corrections, using look-up tables, correcting for residual clouds etc. The spread between the round robin AMFs is indicative of the structural uncertainty in the AMF calculation [Lorente et al., 2017].

It is impossible at the algorithm development stage to have a full understanding which settings and approaches lead to the best results. This led the consortium to consider it beneficial to include more than one 'best practices' approach for the stratospheric correction and AMF calculation sub-steps. Specifically, apart from the proposed 'default' stratospheric correction method, also stratospheric NO_2 column estimates from the independent STREAM method has been included in the QA4ECV NO_2 data product. For the tropospheric AMF calculation, it was decided to provide both the standard tropospheric AMF (linear combination of a partly cloudy, partly clear-sky AMF) but also to include the clear-sky AMF in the data product. This allows data producers to directly test different 'retrieval options' (correcting for residual clouds vs. cloud 'clearing') at the validation stage, and provides users with the possibility to test the robustness of the data product beyond the quoted retrieval uncertainty alone.

2.4 GCOS requirements and GCOS guidelines for dataset generation

The Global Climate Observing System (GCOS) published a set of requirements that tropospheric NO_2 columns should fulfill. The requirements from GCOS report 154 [WMO, 2011] are listed in Table 1 below. The recently published requirements from GCOS report 200 [GCOS, 2016], are not considered here yet.

Table 1. GCOS requirements for satellite retrievals of tropospheric NO_2 columns [WMO, 2011].

	Horizontal Resolution	Vertical Resolution	Temporal Resolution	Uncertainty	Stability/decade[2]

measurement, as there can be other methods, some of them as yet unknown or in some way impractical, that would give systematically different results of apparently equal validity.

NO$_2$ tropospheric column	5-10km	N/A	4h	max(20%; 0.03 DU[1])	2%

[1]An uncertainty of 0.03 DU (Dobson Units) corresponds to 0.8×10^{15} molec.cm^{-2}. The 0.03 DU holds for tropospheric NO$_2$ columns up to 4.0×10^{15} molec.cm^{-2}, for larger column values the relative uncertainty of 20% holds. Note that we replaced the heading 'accuracy' of [WMO, 2011] by 'uncertainty' to be compliant with ISO standard on metrology [VIM]. Indeed, [WMO, 2011] states that "the [accuracy]

5   requirements are indicative of acceptable overall levels for the uncertainties of product values."

[2] According to GCOS, the user requirement for stability is a requirement on the extent to which the uncertainty of a measurement remains constant over a long period (GCOS-200, 2016).

The GCOS requirements, especially those on resolution, can be discussed for their adequacy. These are

10   'target requirements', which should be advanced towards when generating a long-term record of tropospheric NO$_2$ column measurements. The resolution requirements listed above cannot be met by the satellite sensors capable of measuring NO$_2$ that have been operational over the last 20 years, because of limitations in their instrument design, with the exception of the recently launched S5P-TROPOMI sensor, which does meet the requirement. Indeed, the GCOS report states that "products at lower spatial

15   and temporal resolution" than 5-10 km (that is the NO$_2$ products currently available from GOME, SCIAMACHY, OMI, and GOME-2) "…would be sufficient to provide an independent instrument data record of long-term precursor trends to assist in the attribution of changes in ozone and aerosol".

The target requirements for uncertainty and stability are possibly within reach, judging from validation

20   studies, and these have motivated the QA4ECV consortium to find ways to reduce the retrieval uncertainties, and to better estimate the systematic error component of the retrieval uncertainty.

GCOS has also established guidelines for the generation of climate datasets [GCOS, 2010]. Those guidelines serve as a checklist, against which ECV producers can evaluate their production and

25   documentation process [Nightingale et al., 2018]. Section 3 of the Supplement provides a point-by-point

overview of how these guidelines have been taken into account for the generation of the QA4ECV NO_2 data product. A comprehensive comparison with respect to these and other GCOS requirements [GCOS, 2016; WMO, 2010, 2011] is available in QA4ECV deliverable D6.1 [Compernolle, 2018].

3 Quality of level 1 data

5   In the early stages of the QA4ECV project design, it was decided to use GOME, SCIAMACHY, OMI, and GOME-2 (on MetOp-A) to generate a data record for tropospheric and stratospheric NO_2 vertical columns spanning the period 1995-2017. Table 2 lists the relevant instrument specifics for these instruments. For all instruments, the most recent and corrected level-1 datasets are used.

10   **Table 2.** Satellite instruments and level-1 data contributing to the QA4ECV NO_2 ECV data product.

Instrument	Local overpass time	Spatial resolution	Calibrated level-1 datasets	Spectral resolution /sampling	Main level-1 issue
GOME (1995-2003)[1]	10:30 hrs	320×40 km^2	version 5[2]	0.40 nm, 0.20 nm	Spectral structures in Solar irradiance caused by diffuser plate
SCIAMACHY (2002-2012)	10:00 hrs	60×30 km^2	version 8	0.44 nm, 0.24 nm	
OMI (2004-)	13:40 hrs	24×13 km^2 (at nadir)	collection003	0.63 nm, 0.21 nm	Row anomaly (blockage), stripes
GOME-2(A) (2007-)	09:30 hrs	80×40 km^2; 40×40	EUMETSAT/ R/5_12	0.50 nm, 0.20 nm	Throughput loss resulting in more noise

---

[2] The level 1 data are not exactly the same as in Coldewey-Egbers et al. [2018], which is v5.1. The main difference between v5 and v5.1 is the consistency of orbits, and not the radiances themselves [Dehn, personal communication].

		km^2 after 15 July 2013			

[1]GOME lv1 data is in principle available up until September 2011, but for a limited area of the globe only. In June 2003 the on-board tape recorder failed resulting in reduced coverage of GOME-observations, since data could only be downlinked in 'real-time' during overpasses above ground receiving stations.

Prior to algorithm testing, we assessed the quality of the relevant level-1 data. Here we briefly discuss our findings, and discuss how the quality of the level-1 data may affect the retrieval of NO$_2$ SCDs and their uncertainties.

10 **GOME**

GOME level-1 data with global coverage is available from July 1995 to June 2003. ESA has produced a GOME level-1 data set for the mission called version 5.1 [GOME Products and Algorithms, 2018] that is sufficiently well characterized and complete. An important concern with GOME level-1 data is that the solar irradiance signal is detected after reflection from a diffuser plate, whereas the radiance signal

15 is not. The reflection on the diffuser plate created large and seasonally varying artificial spectral structures in the solar irradiance [Richter and Wagner, 2001]. This makes it very difficult for GOME to use solar irradiance spectra as reference in the DOAS spectral fitting. To avoid the issue, Earthshine radiances over remote regions can be used as reference spectra. The implication is that only differential NO$_2$ SCDs are retrieved. To provide the total NO$_2$ SCDs necessary for an ECV dataset of stratospheric

20 and tropospheric NO$_2$ columns, a background correction, typically estimated from an external source, is required.

Detector degradation is another relevant issue for NO$_2$ and cloud retrievals in the visible channel. This degradation in the level-1 data has been estimated to amount to approximately 15% between 1995 and

25 2003 [Slijkhuis et al., 2015], and is anticipated to result in modest increases in the GOME NO$_2$ SCD

uncertainty. The quality of the GOME level-1 data has also been affected by other instrument-related issues, but these may be of less relevance to the quality of the NO_2 spectral fits. Different GOME scan angles (East, Nadir, West) are affected differently in terms of throughput degradation and dichroic mirror degradation, possibly resulting in systematic differences in NO_2 SCDs and cloud products for the different scan angles (or stripes).

SCIAMACHY

SCIAMACHY lv1 data is available from August 2002 until April 2012. SCIAMACHY lv1 version 7.04 data has been made available by ESA in 2016. One particular feature of the SCIAMACHY level 1 data is that co-adding of spectra was performed on-board of SCIAMACHY, prior to downlinking the data from the satellite to receiving stations. The cluster 424-527 nm was read out more frequently (than other spectral bands) in order to minimize the co-addition of spectra and thereby optimizing the spatial resolution for NO_2 to 60×30 km^2 (30 x 30 km^2 in some latitude bands). A consequence of this is that only spectral data from the 424-527 nm cluster are available for DOAS NO_2 spectral fitting. Similar as for GOME, SCIAMACHY solar irradiances suffer from spectral structures from the diffuser plate. A second diffuser was therefore included in the instrument, mounted on the backside of the azimuthal scan mirror. Using solar irradiances from this azimuthal scan mirror strongly reduces the apparent seasonality in NO_2 introduced by the diffuser, although some structures still remain [Richter et al., 2011].

Over its lifetime, the SCIAMACHY instrument suffered from degradation of its optical components. This degradation is the result of a complex mixture of aging of the front optics through UV radiation and photochemical reactions, detector contamination by water vapour deposition and changes in the thermal equilibrium of the platform. As a result, the throughput of SCIAMACHY decreased over the years, in particular in the UV. In addition, small changes in spectral sensitivity over time, for example

from etaloning[3], cancel when using daily irradiance spectra for DOAS spectral fits, but this prevents the use of a single solar irradiance for the full time series. As degradation of the scan mirror leads to scan angle dependent degradation, and scan angle dependent biases, or stripes, can therefore develop over time in the NO$_2$ SCDs.

OMI

The OMI instrument produces stable (to ~2% over the mission time, in the row anomaly-free areas) lv1 radiances over the period 2004-2017 for rows not affected by the row anomaly. The OMI level-1 data are from the Collection 3 data. Processing of this Collection 3 data started in February 2010 with the

10   version 1.1.3 of the Ground Data Processing System software [Dobber et al., 2008], and has produced a complete level-1 dataset for the entire OMI mission. The main issue of the OMI level-1 data is the so-called row anomaly (RA). From June 2007 onwards, several rows of the CCD detector (each corresponding to a specific part of the OMI nadir field-of-view), received less light from the Earth, and some other rows appear to receive Sunlight scattered off a peeling piece of spacecraft insulation. A

15   plausible reason for these effects is a partial obscuration of the entrance port by insulating layer material that may have come loose on the outside of the instrument. For rows affected by the RA, successful spectral fits can still be achieved for NO$_2$, but the cloud retrievals suffer from large errors that cannot be overcome, thus the affected pixels have to be removed from further analysis. Figure 1 shows the rows flagged in the Collection 3 level 1 data over time. By 2017, 38% of the available data was affected by

20   the row anomaly. All rows affected are flagged with a specific 'row anomaly flag' in the QA4ECV OMI NO$_2$ data product, addressing the user needs expressed in section 2.1.

Spurious across-track variability, or stripes, are apparent in current OMI NO$_2$ data products. The stripes appear as discrete jumps in NO$_2$ SCDs from one viewing angle to the other. The origin of the stripes is

25   probably related to small differences in spectral calibration and detector sensitivity from one viewing

---

[3]Etaloning refers to unintended multiple reflection of radiation between optical elements leading to constructive and destructive interference for some wavelengths.

angle to the other. There is currently no solution via the level 1 data, but application of a destriping correction (e.g. Boersma et al. [2011) reduces the systematic stripes to within acceptable limits. The magnitude of the NO_2 destriping corrections has increased from 0.3×10^{15} to 0.5×10^{15} molec. cm^{-2} between 2004 and 2016, related to the use of an annual mean (2005) irradiance spectrum as reference in the DOAS spectral fits.

Optical throughput changes in OMI's (visible) irradiance channel is on the order of 1-1.5% over the mission period (for rows not affected by the RA). For a signal-to-noise ratio of approximately 500, a deterioration of 1.5% leads to only marginal increases in NO_2 fitting uncertainties [Zara et al., 2018]. Spectral stability, important for the accuracy of DOAS retrievals, has also been very good in the visible channel at 0.002 nm. Such wavelength shifts, if unaccounted for, cause NO_2 SCD errors of less than 1%. For more details, please see section 2.2 of QA4ECV Deliverable 4.2 [Müller et al., 2016] and Schenkeveld et al. [2017].

GOME-2(A)

GOME-2 on EUMETSAT's MetOp-A satellite is an improved version of the GOME instrument [Munro et al., 2016]. Level-1 data, version 6.0, is available from January 2007 onwards. A key concern is the accuracy of the long-term record of GOME-2(A) level-1 data. Like GOME and SCIAMACHY, GOME-2 suffered from degradation of its optical components during its lifetime. The optical parts of GOME-2(A) are thought to be increasingly contaminated by outgassing coating material that was meant to protect the detector electronics [Hassinen et al., 2016]. This contamination resulted in a progressive wavelength-dependent loss of the instrument throughput. The discontinuity appearing in September 2009 reflects the so-called 2nd throughput test, during which the temperature of the GOME-2 instrument was changed in a controlled way to observe whether or not there was a recovery in performance at any point during the heating. Although the test did not recover the degradation already suffered, it did succeed in stabilizing the throughput from September 2009 onwards. The main impact of this degradation is an increase of the noise due to throughput loss. As a result uncertainties from random

error on the NO_2 slant columns are expected to increase with time, especially between January 2007 and September 2009. Compared to GOME and SCIAMACHY, degradation of GOME-2(A) started immediately after launch and proceeded faster, but was stabilized after the throughput test.
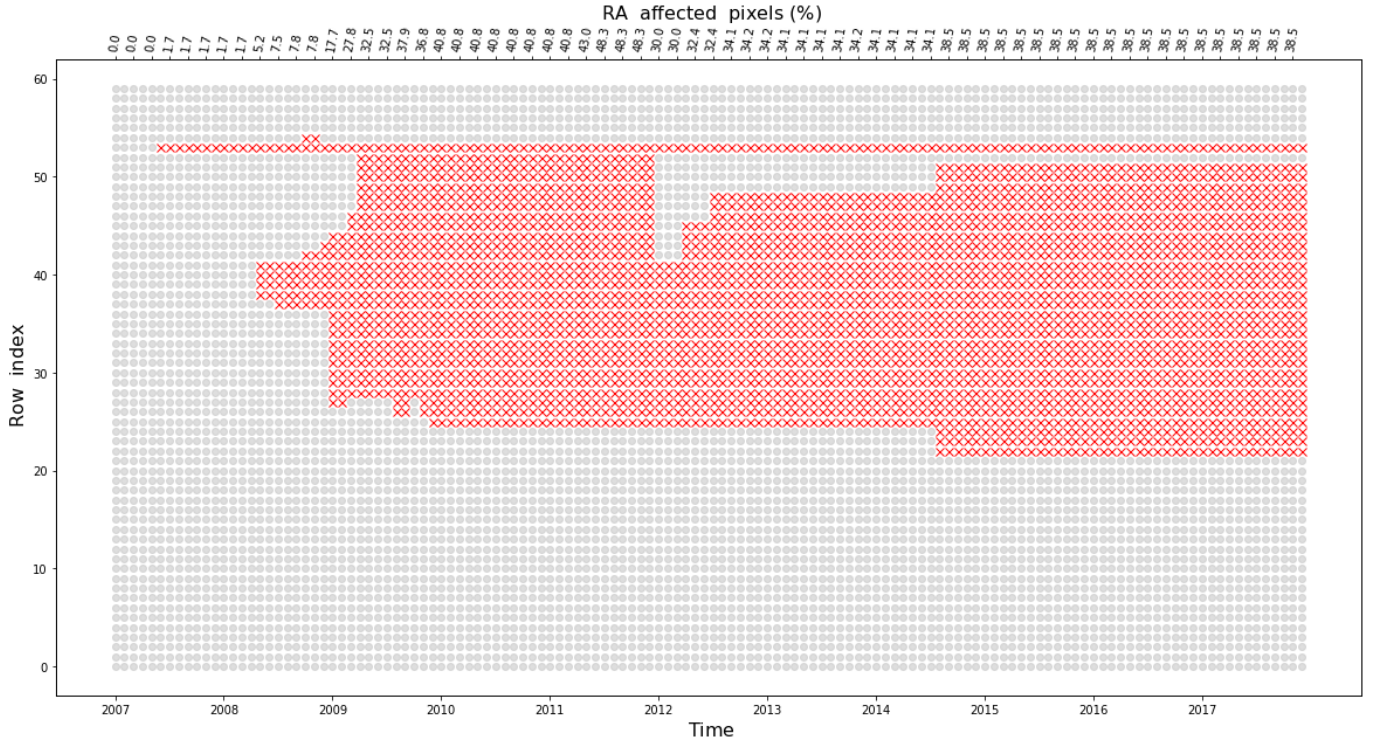
Figure 1. OMI row anomaly in the UV-2/VIS channel as a function of time throughout the OMI mission (2007-2017). Prior to 2007, there was no row anomaly. Affected rows (red crosses) are suffering from a partial blockage of light entering the instrument, so that the absolute radiance levels become compromised. The upper x-axis indicates the percentage of OMI pixels (defined as the 100% × ratio of the number of pixels affected by the row anomaly to the total number of pixels) being affected at a particular month.

In-flight analysis of the GOME-2(A) instrument slit function using a non-linear fitting of Gaussian lineshapes to the Kurucz solar atlas has revealed significant time variations of the GOME-2 slit function in channel 3 (e.g. Dikty et al., [2011]). Specifically, the nominal width of the slit function (0.50 nm) has decreased over time, probably due to thermal fluctuations of the GOME-2(A) optical bench associated with seasonal and long-term changes in the solar irradiance [Munro et al., 2016]. In QA4ECV, this issue is addressed by including the GOME-2(A) slit function as a fit parameter in the DOAS spectral fitting procedure. However, it is unlikely that this fully resolves the issue, so that further increases in NO_2 SCD uncertainties over time should be anticipated [Zara et al., 2018]. Compared to GOME and SCIAMACHY, GOME-2(A) solar irradiances suffer much less from spectral structures caused by the diffuser plate, but some small effects remain [Richter et al, 2011]. One minor issue is the sensitivity for polarization structures in the level-1 spectra. In principle, this is corrected for in the level 0-to-1 algorithm [Munro et al., 2016], but some residual small spectral features remain that may interact with atmospheric absorbers in the DOAS fitting.

The instrument specifics, intrinsic quality, and degradation of the 4 instruments' level-1 data have guided us in selecting the basic settings for spectral fitting of QA4ECV NO_2 SCDs. We used these guiding principles:

- Select as much as possible just one or in any case overlapping fitting windows for different instruments. NO_2 SCDs are known to be sensitive to the selection of fitting window, as shown in van Geffen et al. [2015], and in the S5P-verification report [S5P/TROPOMI Verification Report, 2015].
- Select a wide fitting window including more NO_2 absorption features for an instrument with a relatively low signal-to-noise, i.e. OMI. This is known to reduce the random component of the uncertainty in the NO_2 SCDs (e.g. Bucsela et al. [2006]; Boersma et al. [2007]).
- Select the most practical reference spectrum for the DOAS spectral fitting. Ideally these spectra are daily solar irradiances, as in the case of OMI, but if these are compromised in any way, they may be replaced by an average irradiance spectrum, or by daily Earthshine spectra, as is done for GOME. For the latter, a correction for the amount of NO_2 absorption signature in the

Earthshine reference spectrum is still required.

4 Algorithm design and traceability

An important ambition of the QA4ECV project is to provide full traceability on retrieval algorithms. Usually, a condensed flow diagram for the retrieval algorithm is included in an Algorithm Theoretical

5 Baseline Document (ATBD). The drawback is that ATBDs are often not easily accessible, and that it is not immediately clear which ancillary information has been used in particular algorithm sub-steps. We therefore generated an algorithm 'traceability chain', a web-hosted interactive flow diagram that shows how the QA4ECV NO_2 algorithm is put together, which external pieces of information are embedded in the retrieval process, and where details on those pieces of information can be found. The traceability

10 chain has different layers (Figure 2). The main entry for users is the overall algorithm flow chart. Users can click on algorithm process elements, which takes them a level deeper into the algorithm. Figure 2 shows how to interact with the NO_2 traceability chain at multiple levels. The chain is provided as a clickable option on the QA4ECV website along with the options 'Data Access' and 'User Forum'. Providing these options at the same 'entrance level' allows users to obtain good understanding of how

15 the algorithm works and where ancillary data is coming from. The 'Traceability Chain' button, leads to the full chain (1st layer). Next, as an example, when clicking the 'DOAS + wavelength calibration' step will lead to details on that sub-process (2nd layer). The absorption cross sections used in the DOAS step, are available under 'Laboratory Absorption Cross Sections', which contains the references to the cross-section data and papers describing them (3rd layer). The references themselves are linked to the digital

20 object identifiers (DOI's) and take users directly to the relevant paper.
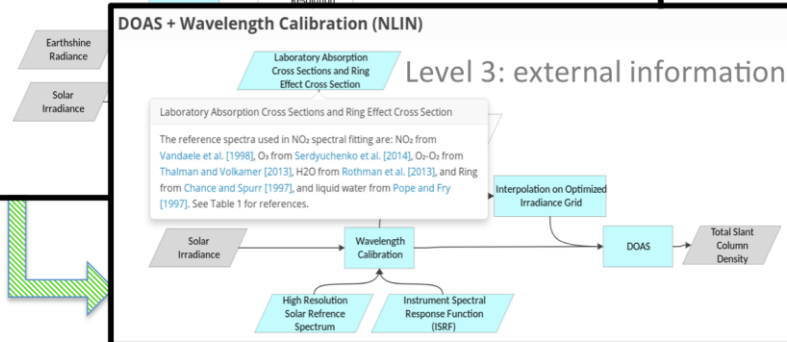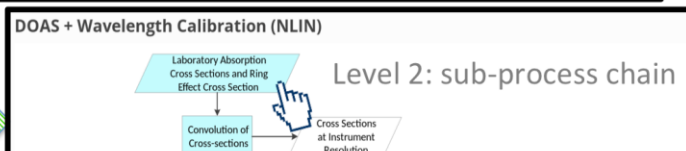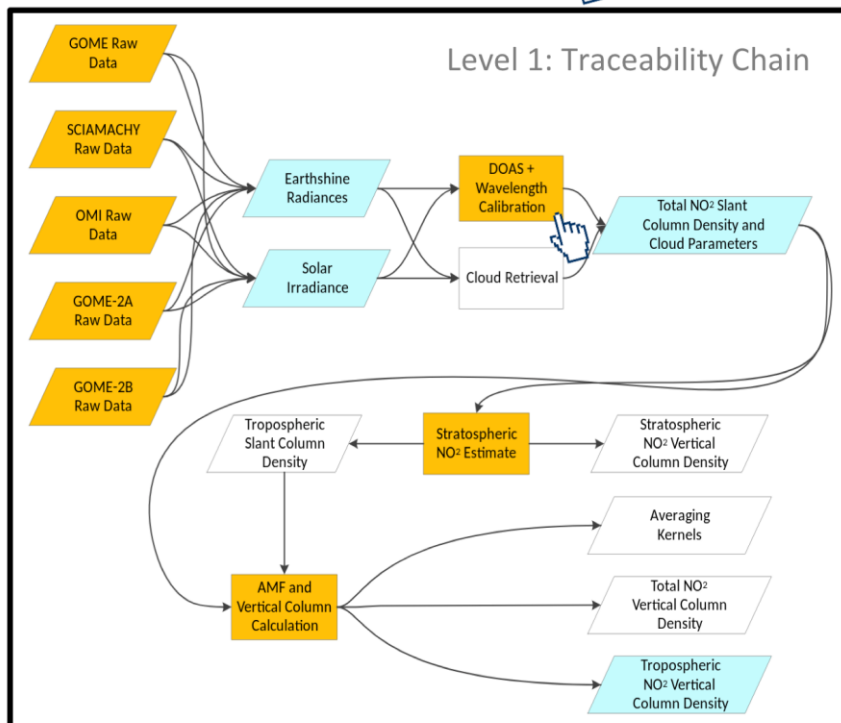
Figure 2. Traceability Chain for the QA4ECV NO$_2$ retrieval algorithm. The orange blocks (rectangles) are the building blocks of the retrieval, and in the main chain these are clickable to see more details in deeper layers. The light blue blocks are also clickable and will provide more information on that process in a pop-up window. The parallelograms provide information on algorithm choices and input data sets. The interactive traceability chain is available at: http://www.qa4ecv.eu/ecv/no2-pre.

5 Inter-comparison of retrieval sub-steps and algorithm selection

Differences between NO$_2$ retrievals from different retrieval groups can be traced back to different settings and to different a priori parameters used in the individual retrievals. We made a systematic step-by-step analysis of all components of the NO$_2$ retrieval, by documenting and comparing approaches from the consortium institutes, and analysing their contribution to differences and their benefits. These tests, evaluations, and innovations have guided the development of the QA4ECV consortium 'best practice' algorithm for generating a multi-decadal record for NO$_2$, and helped to characterize the uncertainties of each retrieval sub-step.

5.1 Evaluation of spectral fitting approaches

NO$_2$ spectral fitting approaches by BIRA-IASB, IUP Bremen, KNMI, and MPI-C were compared in two rounds, with emphasis on OMI and GOME-2, for

    (1) common (as much as possible identical) settings for the same level-1 data,

    (2) preferred retrieval settings defined by each group..

The inter-comparison comprised 4 full days: winter and summer early and late in the mission in order to investigate the agreement of retrieval codes with respect to seasonal and instrumental changes. Table 3 shows the details of the spectral fitting retrieval code from the four participating institutes. The retrieval algorithms are based on the same principles, but have been implemented differently, and use different software packages. The KNMI code applies a wavelength shift prior to the DOAS fit, and does not include an intensity offset in the intensity-fitting model. The common settings are listed in the caption of Table 3.

Table 3. Overview of OMI SCD retrieval codes from the QA4ECV consortium's institutes. The common settings used for round 1 were: 405-465 nm fitting window, polynomial degree 4, inclusion of O_3, NO_2, O_2-O_2, H_2O, Ring cross-sections, use of mean solar irradiance as reference spectrum. The cross-sections have been convolved with the OMI slit function for each row separately.

Institute	Retrieval	Code	Method	Wavelength calibration	Intensity offset	Spike removal	Reference
BIRA-IASB	QDOAS	C	Optical depth[4], non-linear least squares regression (Levenberg-Marquard)	Via Fraunhofer atlas, and shift and squeeze	Yes	Yes	Fayt and Van Roozendael [2001]
IUP Bremen	NLIN	PASCAL/ DELPHI	Optical depth, non-linear least squares regression (Levenberg-Marquard)	Via Fraunhofer atlas, and shift and squeeze	Yes	Yes	Richter [1997]

---

[4]An 'optical depth' fitting model is of the form: $\ln\left(\frac{I(\lambda)}{I_0(\lambda)}\right) = -\sum_i \sigma_i(\lambda)N_{s,i} + \sum_j a_j\lambda^j$ with $I(\lambda)$ the radiance, and $I_0(\lambda)$ the irradiance spectrum, $\sigma_i(\lambda)$ the absorption cross section spectrum of trace gas i, $N_{s,i}$ the fitting coefficient, or slant column density of trace gas i, and a_j the coefficients of a low order polynomial.

KNMI	OMNO2A v2	C	Intensity fit[5], non-linear least squares regression (Levenberg-Marquard)	Via Fraunhofer atlas, and shift	No	Yes	van Geffen et al. [2015]
MPI-C	MPI-C	MATLAB	Optical depth, non-linear least squares regression, shift and squeeze accounted for by pseudo-absorbers	Via Fraunhofer atlas, and shift and squeeze (non-linear)	Yes	Yes	Beirle et al. [2013]

The common settings inter-comparison of OMI NO$_2$ SCDs for all orbits on 2 February 2005, 16 August 2005, 4 February 2013, and 4 August 2013 showed very good agreement between the different algorithms. The correlation between SCDs from each pair of retrieval codes is always >99.8% for all

5   OMI orbits within the 4 selected days. The correlation is slightly less (but still > 99%) between the KNMI code and the other 3 codes, suggesting that algorithms agree in capturing the full dynamical range of NO$_2$ SCDs. The remaining differences appear over background regions and can be attributed to using a non-linear intensity fitting model instead of a linear optical density fit (resulting in NO$_2$ SCD differences over the oceans up to 1×10^{15} molec.cm^{-2}, see Figure 3), and to including or excluding an

10 intensity offset term in the set of fit parameters (differences up to 1×10^{15} molec.cm^{-2}, reducing contrast

[5]An 'intensity' fitting model is of the form: $I(\lambda) = I_0(\lambda)e^{-\sum_i \sigma_i(\lambda)N_{s,i} + \sum_j a_j \lambda^j}$

between bright and dark scenes). Retrieval on optical densities has the advantage of being a linear fit and has traditionally been used in DOAS applications. Fitting intensities has the advantage of a more transparent treatment of the Ring effect and has been applied in operational OMI data retrieval [van Geffen et al., 2015]. While none of the two approaches is better by definition, the results differ in particular in combination with the offset correction applied. Based on these outcomes, it was recommended to use optical density fitting and include the intensity offset in the QA4ECV fitting model, even though the exact physical meaning of this term is not entirely clear. Including the intensity-offset term appears to account for spectral signatures originating from vibrational Raman scattering in open water [Oldeman, 2018], associated incomplete Ring corrections, and prevents O_3 misfits over water and over land. Excluding the intensity offset term results in larger NO_2 SCD uncertainties, and in (spurious) spatial patterns in the O_3 SCDs that resemble the spatial patterns in TOA reflectance. In QA4ECV, the spectral fitting is approximated by Eq. (2) in Zara et al. [2018] (QDOAS), and a variation thereof for NLIN. For more details, see QA4ECV Deliverable 4.2, section 2.3 [Müller et al., 2016].
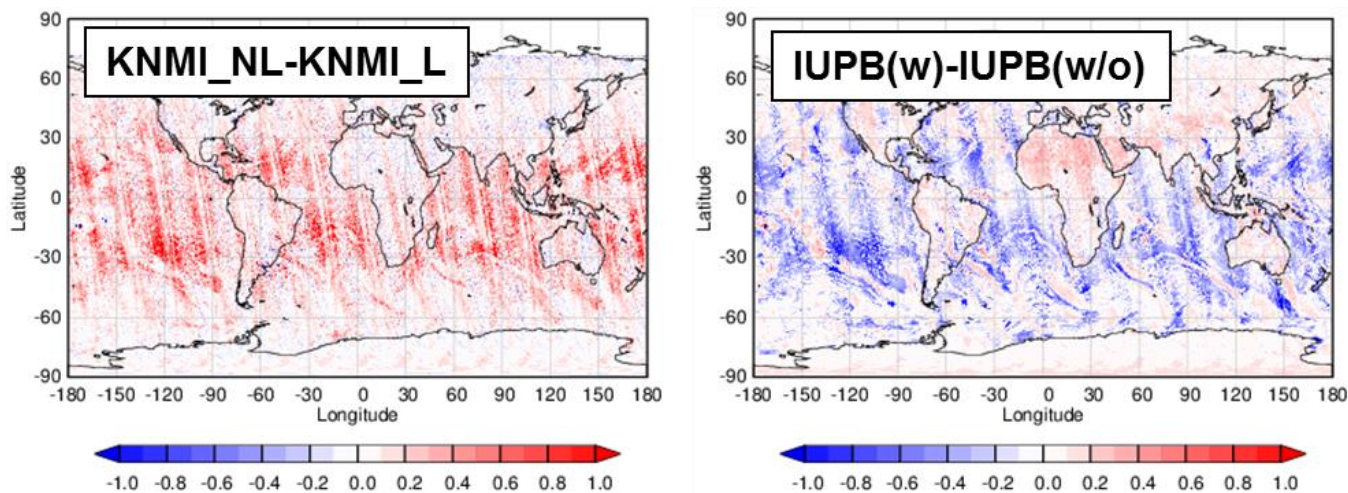
Figure 3. OMI NO_2 slant column differences between KNMI intensity (KNMI_NL) and optical density fit (KNMI_L) and IUPB fit including (IUPB(w)) and excluding the intensity offset (IUPB(w/o))(right). Data from 2 February 2005. The units of the colour bar in both panels are 10^{15} molec. cm^{-2}.

In round 2, each institute applied preferred settings to retrieve OMI NO_2 SCDs for the same set of days. The KNMI settings are identical as in round 1 (Table 3). Relative to the common settings, IUP Bremen used the 425-497 nm fitting window, and included a signature for sand absorption (see Richter et al. [2011]) in the fitting model, BIRA-IASB applied a 425-460 nm window, and included both sand and CHO-CHO signatures, and MPI applied a 431-460 nm window, and excluded liquid water absorption from the fit. The inter-comparison of preferred settings SCDs again showed very good agreement between the algorithms. The correlation between the different pairs is >98%, and the average differences between the different sets are $<1\times10^{15}$ molecules cm^{-2}. The largest offset ($+0.9\times10^{15}$ molecules cm^{-2}) appears between KNMI and IUP Bremen (Figure 4(a)). The higher KNMI SCDs are explained by the intensity fit used by KNMI (left panel Figure 3) and by the relatively large difference in centre wavelengths of the fitting window between these algorithms (435 nm for KNMI, 461 nm for IUP Bremen). Between 405-435 nm, the O_3 optical thickness is smaller, and photon paths through the stratosphere are slightly longer than in the 435-500 nm spectral region, located in the flanks of the Chappuis band. DAK simulations indeed show 1.5% higher air mass factors at 405 nm than at 500 nm (Figure 4(b)). For the majority of SCDs retrieved over unpolluted regions, the use of an intensity fit, together with the 'bluer' fitting window explains the differences between KNMI and IUP Bremen retrievals. It was not possible to point out a clear 'winner' among the different fitting approaches, but including an intensity offset, and liquid water absorption in the fit model, reduced fitting residuals and improved NO_2 and O_3 fit results. NO_2 SCDs are most sensitive to the fitting approach, i.e. intensity fit or optical density fit.
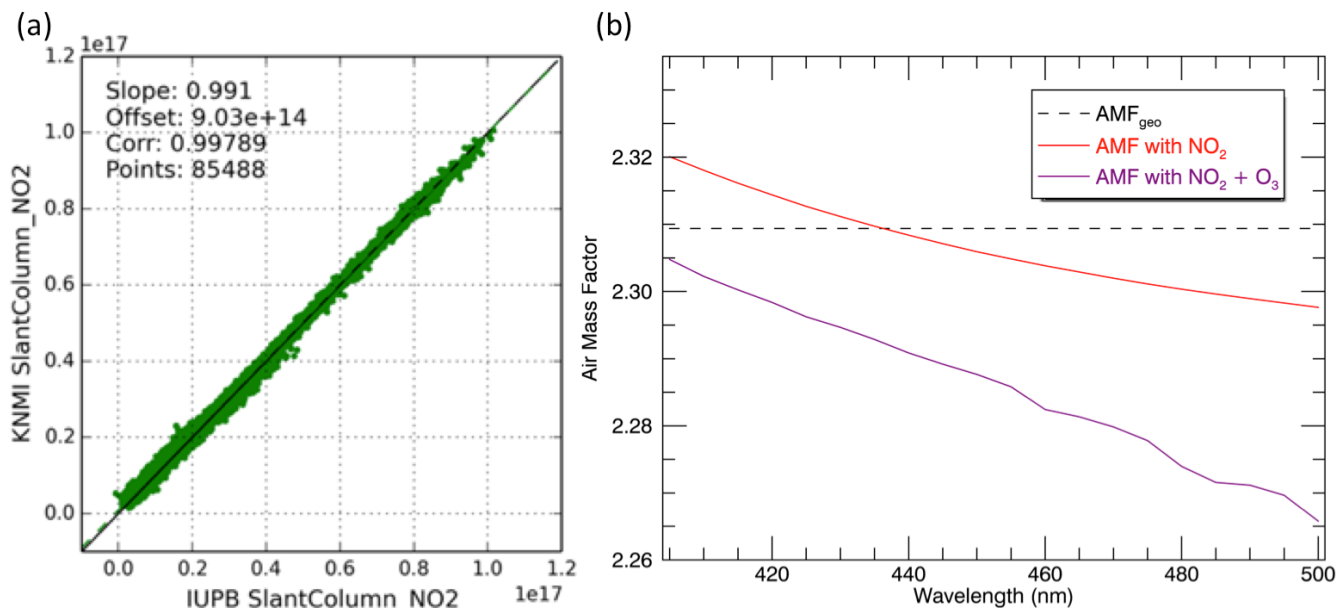
Figure 4. (a) Correlation plots of IUP Bremen (425-497 nm fitting window) and KNMI (405-465 nm) NO$_2$ slant columns retrieved using preferred fit settings for OMI orbit OMIL2_2005m0202t0339, including only pixels with SZA < 88° and intensity < 1.1x10^{14}. (b) Wavelength dependency of the total air mass factor for a scenario with SZA=VZA=30° (geometrical AMF = 2.31), as calculated with DAK
5 for a mid-latitude standard atmospheric profile with a total NO$_2$ column of 5.9×10^{15} molecules cm^{-2} (mostly situated in the stratosphere) (red curve), and for the same mid-latitude standard atmospheric profile but now with absorption by both NO$_2$ and O$_3$ (total column of 322 DU, purple curve).

The comparisons of the fitting approaches led to a number of clear recommendations for spectral fitting
10 of NO$_2$ for the QA4ECV record. A complete list can be found in QA4ECV Deliverable 4.2 [Müller et al., 2016]. We highlight the most important ones here:

- Include an intensity-offset correction.
- Given the sensitivity to selecting intensity fit or optical density fit (systematic bias up to 1×10^{15} molecules cm^{-2}), it is recommended to use one and the same fit model for all sensors.
15 - For the 405-465 nm fitting window, the absorption spectrum of liquid water should be included (not necessary for the smaller windows)

Together with the recommendations driven by level-1 data quality considerations shown in Table 3, this has led to the definition of spectral fitting of NO_2 and data processing from GOME, SCIAMACHY, OMI, and GOME-2 as summarized in Table 4. Here, the rationale was to maintain as much as possible the same fit approach and fit settings for GOME-2, SCIAMACHY, and GOME as for OMI (for details see Table 2 in Muller et al. [2018].

For the morning sensors GOME-2, SCIAMACHY, and GOME, tests were done to evaluate the consistency between results of the spectral fitting approaches, since some settings such as the fitting window had to be different in order to avoid SCIAMACHY and GOME feature for wavelengths <425 nm interfering with the spectral fit. The results of these tests are reported in the QA4ECV Deliverable document 4.5 [Muller et al., 2018], and we summarize them here. Monthly mean normalized NO_2 SCDs from GOME-2(A) and SCIAMACHY agreed very well in space and time despite the differences between the instruments in terms of coverage, pixel size, and fitting window (Figure 17 and 18 in Muller et al. [2018]). Because GOME suffers from a diffuser plate artifact emerging in the irradiance files, we used daily radiance spectra obtained over the Pacific Ocean as substitute for the irradiance spectrum. The region over the Pacific is largely free of tropospheric NO_2. We then determined an offset correction as the difference between the normalized SCD values from SCIAMACHY and GOME between August 2002 and June 2003, when both instruments were operational over the reference sector. It amounts to 1.48×10^{15} molec.cm^{-2}. A subsequent matching of the corrected GOME stratospheric columns to the SCIAMACHY stratospheric columns over the reference region, showed that the robustness of the correction is excellent, with only small deviations ($\pm 10^{14}$ molec.cm^{-2}) between GOME and SCIAMACHY for the period of overlap (Figure 20 in Muller et al. [2018]).

Table 4. Recommended settings for QA4ECV NO_2 spectral fitting for the retrieval of NO_2 slant columns from GOME, SCIAMACHY, OMI, and GOME-2(A) to generate a multi-decadal data record for the period 1995-2017.

	OMI	**GOME-2(A)**	**SCIAMACHY**	**GOME**
DOAS processor	QDOAS version	NLIN 7.25 (2007-	NLIN 7.35-7.37	NLIN 7.55

	Globalcalib6	2011) QDOAS (2012-2016)		
Fitting window	405-465 nm	405-465 nm	425-465 nm	425-465 nm
Fitting method	Optical density	Optical density	Optical density	Optical density
Selection reference spectrum	Annual mean (2005) solar reference	Daily solar reference	Daily solar reference (from azimuthal scan mirror diffuser)	Daily Pacific radiance[6] plus offset of 1.476×10^{15} molec. cm^{-2})
Polynomial	4th order	4th order	4th order	4th order
Fitting parameters	O_3, NO_2, O_2-O_2, H_2O, Ring, liquid water	O_3, NO_2, O_2-O_2, H_2O, Ring, liquid water	O_3, NO_2, O_2-O_2, H_2O, Ring, liquid water	O_3, NO_2, O_2-O_2, H_2O, Ring, liquid water
Undersampling correction[7]	No	No	No	Yes
Eta correction	No	No	No	Yes

5.2 Evaluation of stratosphere-troposphere separation

We compared stratospheric correction approaches by IUP Bremen, KNMI, and MPI-C to establish best practices for this algorithm step. The stratospheric correction approach from IUP Bremen is based on
5   scaling model-simulated (B3dCTM model) stratospheric vertical columns to match satellite

---

[6] Averaged for the area enclosed by 160-260°E, 10°S-10°N. The offset has been determined from a comparison with coincident SCIAMACHY SCDs (2002-2003)

[7] GOME is the only sensor that requires an undersampling and eta (polarisation) correction because the spectral sampling is too coarse for the FWHM of the GOME slit function [Chance et al., 2005] and the polarization structure needs to be accounted for. For the other sensors this is not needed.

observations over the remote Pacific [Hilboll et al., 2013]. In the KNMI-approach, NO_2 SCDs are assimilated in the TM4 model, so that model simulations of stratospheric NO_2 columns agree well with the retrieved slant columns over regions away from strong tropospheric pollution [Dirksen et al., 2011]. MPI-C uses a modified reference sector approach called STREAM [Beirle et al., 2016]. This approach estimates the stratospheric vertical columns from retrievals over regions where tropospheric NO_2 is assumed to be negligible, and over regions with high clouds, where the tropospheric column is shielded. The derived stratospheric field is then smoothed and interpolated globally, based on the assumption that the spatial pattern of stratospheric NO_2 does not feature strong gradients.

The inter-comparison of stratospheric correction approaches focused on 2 individual days (1 January 2005 and 19 July 2005) and 2 monthly means (January and July 2005). This comparison should be regarded as a 'preferred settings' round, where SCD inputs were identical, but the stratospheric AMFs, and methods to estimate the stratospheric NO_2 columns varied between the groups. We evaluated the "success" of the stratospheric corrections via checks on the smoothness of stratospheric patterns, and on the plausibility of the tropospheric residues (defined as N_v-$N_{v,strat}$) over remote regions where values are expected to be low and not strongly negative. The comparisons (Section 2.4 of QA4ECV Deliverable 4.2 [Müller et al., 2016]) indicated that the different schemes showed similar stratospheric NO_2 columns and tropospheric residues, and each of the approaches would be appropriate for use in the QA4ECV NO_2 algorithm. The quantitative differences between the stratospheric NO_2 columns were generally smaller than 0.5×10^{15} molecules cm^{-2}, a number that can be regarded as an upper limit for the 'structural uncertainty' in the stratospheric estimate, but the patterns also revealed that IUP Bremen and KNMI stratospheric NO_2 columns were biased high at high solar zenith angles in the winter hemisphere. In Lorente et al. [2017], we attributed this bias to the SCIATRAN and DAK radiative transfer models not fully accounting for the sphericity of the atmosphere in describing photon transport after backscattering. The McArtim model does account for the sphericity of the atmosphere for both incoming and backscattered light, resulting in lower stratospheric AMFs especially for extreme solar zenith angles.

The KNMI data assimilation was selected as the default approach for estimating the stratospheric NO_2 column in the QA4ECV algorithm. This ensures consistent knowledge of the state of the atmosphere (NO_2 and temperature profiles, stratospheric dynamics) derived from the same model that predicts the a priori tropospheric NO_2 profile shape required by the tropospheric AMF calculation. We decided to update the model framework for assimilation from TM4 to TM5-MP [Williams et al., 2017]. Moreover, the data assimilation approach has incorporated a correction for sphericity via McArtim, as described in Lorente et al. [2017]. Retrieval results point out that the stratospheric AMFs, together with improvements in the data assimilation scheme, lead to much fewer negative tropospheric columns for retrievals at extreme viewing geometries, also at mid-latitudes. As a second option, the consortium selected MPI-C STREAM as a complementary algorithm for stratospheric NO_2 estimates to be included in the QA4ECV data product. STREAM is based on the measurements alone without involving models. This allows QA4ECV data users to switch approaches, which may be beneficial under certain circumstances. Especially in situations with strong stratospheric NO_2 gradients, such as near the polar vortex, assimilation is the preferred approach. It has been shown that the data assimilation captures the strong spatial gradients occurring near the vortex [Dirksen et al., 2011], whereas the STREAM method by design results in zonally smooth structures in those regions. STREAM could be a useful alternative to data assimilation for studies into weak NO_x sources, such as emissions from soil, ships, and small, isolated anthropogenic sources. The strength of STREAM is that it based on measurements and does not rely on models. Data assimilation is potentially somewhat vulnerable to misinterpret tropospheric contributions as stratospheric NO_2, so that STREAM could be used in areas away from strong stratospheric gradients (where the zonally smooth structure of the stratospheric field is of little consequence). Furthermore, the differences between the two methods are useful as a measure of structural uncertainty in the stratospheric correction, beyond the typical uncertainties of 0.2×10^{15} molecules cm^{-2} derived from the Observation − Forecast statistics of the assimilation scheme [Dirksen et al., 2011]. Regions of enhanced structural uncertainty are relevant especially over areas with small tropospheric NO_2 enhancements, such as from outflow of continental pollution over oceans, shipping lanes, and over areas with soil NO_x emissions.

As an example, Figure 5 shows OMI stratospheric NO$_2$ estimates from both the data assimilation and STREAM approach for the QA4ECV v1.1 product on 2 February 2005. The upper panel illustrates that the latitudinal gradients in NO$_2$ between the data assimilation and STREAM agree reasonably well. It is evident that the data assimilation approach captures more variability along a zonal band, resulting on this day in lower stratospheric NO$_2$ over North America and Europe, and higher over northeastern Asia than in the STREAM method. The differences are up to 1×10^{15} molec. cm^{-2}, such that they have a substantial impact on the tropospheric column retrievals.
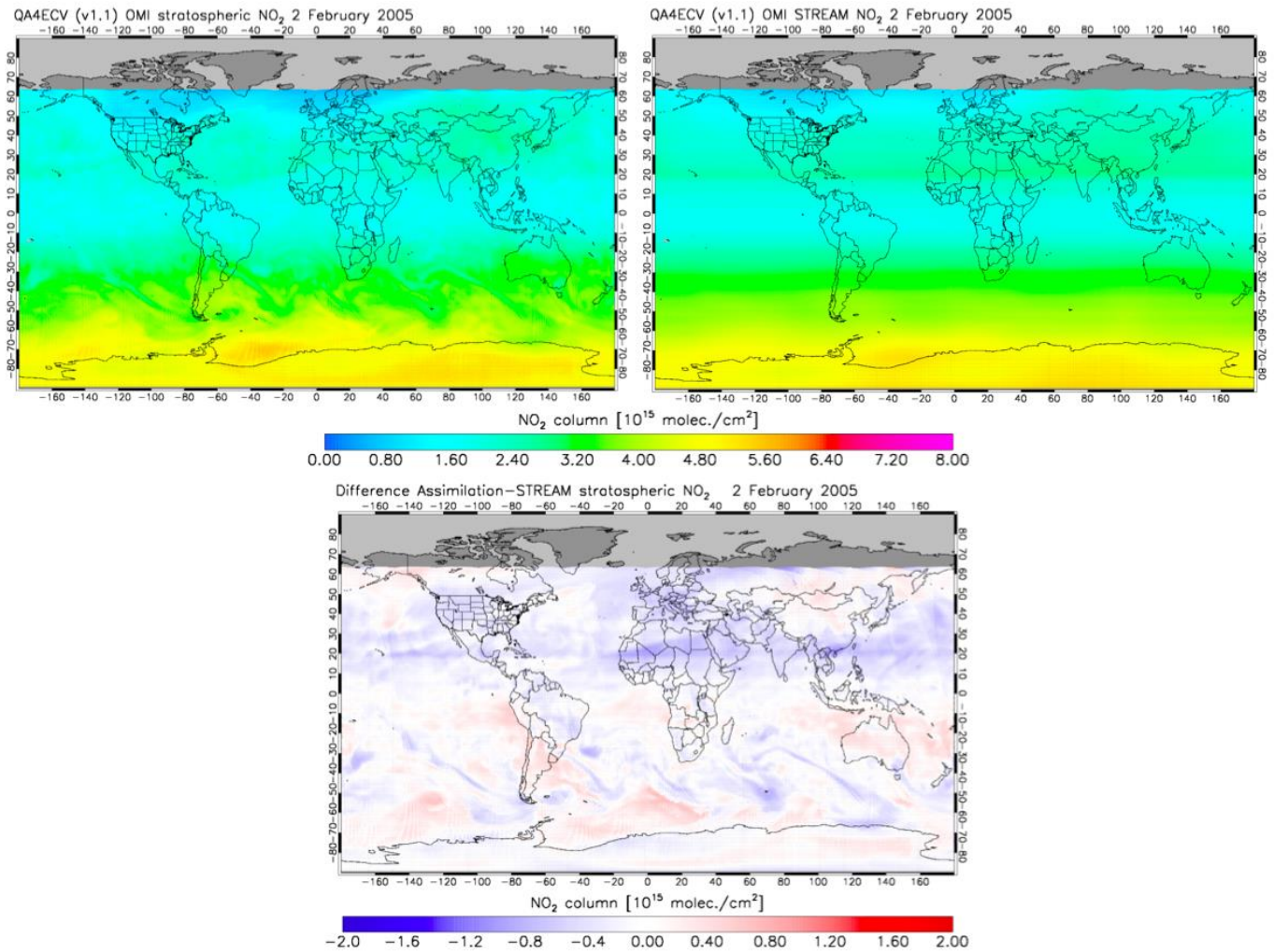
Figure 5. Stratospheric NO$_2$ columns from OMI on 2 February 2005, estimated with the data

assimilation method (upper left), and with the STREAM method (upper right). The lower panel shows the differences between the stratospheric NO_2 estimates.

5.3 Evaluation of air mass factor calculations

We performed a comparison of approaches to calculate AMFs for NO_2 and mapped the uncertainties associated with these approaches. Much of this comparison has been reported in Lorente et al. [2017] and in Section 2.5 of QA4ECV Deliverable 4.2 [Müller et al., 2016], so we give only a brief summary here. First, we compared radiative transfer models from the consortium (LIDORT, SCIATRAN, DAK, McArtim) for their top-of-atmosphere reflectances and their capacity to compute vertically resolved, or 'box' AMFs. The agreement between reflectances from the 4 models' at 440 nm (and also at 340 nm) was excellent. Mean relative differences between models were generally small (<1%), with the exception of high solar zenith angles (>80°), where systematic differences with the McArtim model amount to up to 10%. McArtim is the only model that simulates radiative transfer in full sphericity for the direct and the diffuse light [Deutschmann et al., 2011]. Other differences such as different layering schemes, polarization description, refractive index, and Rayleigh scattering cross section spectrum, only lead to small differences (<1%) between the models.

To establish the QA4ECV NO_2 algorithm settings, we selected the appropriate wavelength for calculating the NO_2 box AMFs. We investigated the wavelength dependency of the NO_2 AMFs for retrieval scenarios with substantial tropospheric pollution ($N_{v,trop} = 16 \times 10^{15}$ molec.cm^{-2}), and considered that the AMF calculated at a single wavelength should be representative for the fit window average AMF. Tropospheric NO_2 AMFs were calculated between 400-500 nm with 1 nm steps. Figure 6 shows a distinct increase of AMF with wavelength. This increase reflects the increasing transparency of the lower troposphere towards the 'green' part of the spectrum where Rayleigh scattering is weakening. In general, tropospheric AMFs increase by 0.2-0.3% per nm redshift. The purple, blue, and light blue lines show the AMF averaged over all spectral points in 3 relevant fitting windows used within QA4ECV and by individual groups.
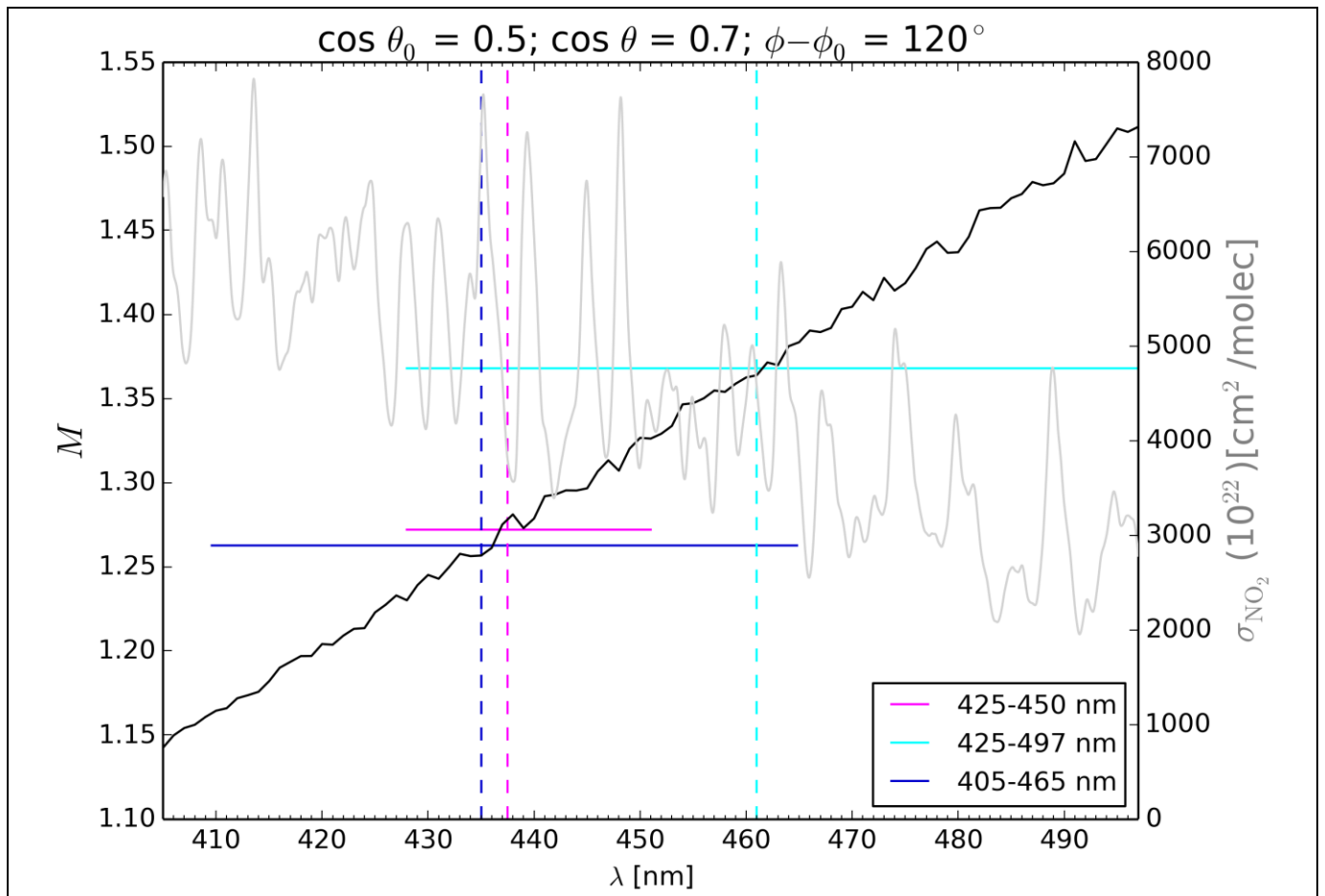
Figure 6. NO$_2$ tropospheric air mass factor (black)[8] as a function of wavelength computed with DAK for a polluted boundary layer for a specific viewing geometry (θ=60°, θ_o=45.6°). Horizontal lines show averaged multi-wavelength AMF for different fitting windows (purple, 425-450 nm, blue 405-465 nm and light blue 425-497 nm). The grey line shows NO$_2$ absorption cross-section from Vandaele et al. (1998) at 220 K. A mid-latitude standard atmosphere was used including O$_3$. The AMF was computed for a polluted boundary layer with 16×10^{15} molec/cm^2, without aerosols, a boundary layer height of 1 km and surface albedo 0.05.

---

[8]The non-smooth behaviour of the black line is because the spectral resolution of the AMF is not sufficient to resolve the NO$_2$ cross section used in the calculation. If a constant cross section value is used in the RTM for calculating TOA reflectance the increasing AMF with wavelength would be spectrally smooth.

We saw in Figure 4 that total NO_2 AMFs weakly decrease with wavelength (-0.01%/nm redshift). Figure 6 shows that tropospheric NO_2 AMFs increase with wavelength (+0.2-0.3%/nm). This difference can be understood from Rayleigh scattering, occurring mostly in the lowest kilometres of the troposphere. The bulk scattering increasingly screens NO_2 in the boundary layer towards the UV, so that tropospheric AMFs are smallest for shorter wavelengths. For the fitting windows considered for QA4ECV NO_2 retrievals (425-465 nm and 405-465 nm), we recommend calculating the NO_2 AMF at 437.5 nm for all sensors. The blue and purple lines in Fig. 8 indicate that 437.5 nm is a representative wavelength to calculate the NO_2 AMF. 437.5 nm is reasonably near to the centre wavelength of both windows (435 and 445 nm respectively) and the 437.5 nm AMF is within 2% of the window-average AMF for both windows. Uncertainties related to the exact choice of AMF wavelength calculation are much smaller than other AMF uncertainties, such as clouds, albedo, trace gas and aerosol profiles, as discussed below and in Lorente et al. [2017].

We compared altitude resolved AMFs and tropospheric AMFs calculated with the 4 different radiative transfer models. We found that the agreement is very good (within 3% and 6% respectively), if identical ancillary data (surface albedo, terrain height, cloud parameters and trace gas profile) and cloud and aerosol corrections are being used. This shows that the choice of RTM for calculation of the tropospheric AMF introduces a modest uncertainty of no more than 6%, which is intrinsic to the calculation method, and cannot be avoided.

To assess the full impact of preferred settings and methods for AMF calculations, we organized a round robin comparison. Six groups joined this round robin, each using their preferred setting to calculate the tropospheric AMFs. Besides the QA4ECV-partners KNMI/WUR, BIRA-IASB, and IUP Bremen, also NASA GSFC, Leicester University, and Peking University participated. The 6 groups used widely different calculation methods (RTMs, temperature, cloud and aerosol corrections) and preferred ancillary data on albedo, terrain height, NO_2 profile, etc. [Müller et al., 2016]. The ensemble mean AMF served as a reference against which to compare the AMFs by the individual groups. The round robin

exercise focused on China because it provides challenging retrieval conditions and Peking University only calculates AMFs over that region. The overall spatial pattern of AMF values was well-reproduced by all groups. AMFs generally agree to within 10% over unpolluted areas, but show differences of up to 40% with respect to the ensemble mean over polluted regions in eastern China and Korea. These differences can be traced back to differences in the preferred surface albedo, clouds, and a priori NO_2 profiles used in the AMF calculation. It is not possible to identify the single most important forward model dependency for the AMF calculation. The analysis in QA4ECV Deliverable 4.2 [Müller et al., 2016] and in Lorente et al. [2017] suggests that accurate knowledge on surface albedo, clouds, and a priori NO_2 profiles are of similar importance, and their interplay, in combination with the choices for cloud and aerosol correction methods is driving the structural uncertainty in the NO_2 AMFs.

Based on the results from the comparisons discussed above, the following recommendations for calculating QA4ECV NO_2 AMFs were made:

- calculate the NO_2 AMFs at 437.5 nm for all instruments
- apply the independent pixel approximation for cloud correction, but also include clear-sky AMFs in the product
- use cloud information (cloud fraction, cloud pressure) from FRESCO+ for GOME, SCIAMACHY, GOME-2A [Wang et al., 2008] and OMCLDO2 for OMI [Veefkind et al., 2016]. These have been derived using the same physical principles as in the AMF calculation.
- apply implicit aerosol correction (via the cloud correction). This correction is effective in most retrieval scenarios with moderate aerosol pollution. When accurate, observation-based aerosol information becomes available from e.g. ECMWF CAMS or NASA GMAO, explicit aerosol corrections will be considered.
- use surface albedo climatologies (as close as possible to the 437.5 nm AMF wavelength) consistent with the ones used in the cloud retrievals. For GOME, SCIAMACHY, and GOME-2A, this is the albedo climatology from Tilstra et al. [2017], and for OMI the updated 5-year climatology [Kleipool et al., 2008].
- use the DEM_3KM pixel-average terrain height.

- use spatially interpolated (to pixel centre) NO_2 profiles simulated by TM5-MP at $1°×1°$. TM5-MP is the model used for the data assimilation of NO_2 SCDs to estimate the stratospheric contribution (section 5.2).

6 QA4ECV NO₂ uncertainty estimates

6.1 Theoretical algorithm uncertainty

The QA4ECV NO_2 product contains an algorithm uncertainty estimate associated with each individual pixel's tropospheric NO_2 column. This estimate is calculated theoretically via uncertainty propagation based on the principal retrieval equation (Eq. (1)):

$$\sigma = \sqrt{\left(\frac{\sigma_{N_S}}{M_{tr}}\right)^2 + \left(\frac{\sigma_{N_{s,strat}}}{M_{tr}}\right)^2 + \left(\frac{(N_S - N_{S,strat})\sigma_{M_{tr}}}{M_{tr}^2}\right)^2} \tag{2}$$

The uncertainty propagation accounts for spectral fitting uncertainties (σ_{N_S}), and contributions from uncertainties in a priori and ancillary data required for calculating the stratospheric NO_2 background ($\sigma_{N_{s,strat}}$) and the AMF ($\sigma_{M_{tr}}$). The uncertainty in the tropospheric AMF, or AMF covariance is written as:

$$\sigma_{M_{tr}}^2 = \left(\frac{\partial M}{\partial A_s}\sigma_{A_s}\right)^2 + \left(\frac{\partial M}{\partial f_{cl}}\sigma_{f_{cl}}\right)^2 + \left(\frac{\partial M}{\partial p_{cl}}\sigma_{p_{cl}}\right)^2 + (0.1 M_{tr})^2 + 2\left(\frac{\partial M}{\partial A_s}\frac{\partial M}{\partial f_{cl}}\langle\epsilon_{f_{cl}}\epsilon_{A_s}\rangle\right) \tag{3}$$

where $\frac{\partial M}{\partial A_s}$ represents the local sensitivity of the the air mass factor to surface albedo A_s, and σ_{A_s} the best estimate of the uncertainty in the surface albedo, and so on. The fourth term on the right hand side represents the contribution from uncertainty in the a priori profile shapes, and is tentatively approximated as 10% of the tropospheric AMF. This term is absent when using the averaging kernel in satellite data applications [Eskes and Boersma, 2003], which removes the dependence on the a-priori

profile. The last term represents the contribution from error correlation between cloud fractions and surface albedo $\langle \epsilon_{f_{cl}} \epsilon_{A_s} \rangle$; surface albedo influences AMF directly, and indirectly because cloud fractions are sensitive to surface reflectance (see Eq. (20) and Eq. (A2) in Boersma et al. [2004] and Lorente et al. [2018] for more detail). As $\langle \epsilon_{f_{cl}} \epsilon_{A_s} \rangle$ and $\frac{\partial M}{\partial f_{cl}}$ are negative, and $\frac{\partial M}{\partial A_s}$ is positive, this last term gives a positive contribution to $\sigma^2_{M_{tr}}$.

The uncertainty σ should be interpreted as the best guess of the retrieval uncertainty for one specific measurement. This uncertainty contains random and systematic error components, and the different systematic error components (due to errors in profile shape, surface albedo, etc.) each have their own spatial and temporal scale. Therefore, when averaging over multiple pixels (spatially) or over time, part of the error will cancel out or be smoothed, but (an unknown) part of the systematic error will remain even after averaging, see Boersma et al. [2016].

We recommend using Eq. (4) below to estimate the uncertainty σ_o for spatially or temporally averaged data. This method takes the area-weighted (statistical) retrieval uncertainty σ, and then accounts for a partial correlation in the errors between pixels as in Eskes et al. [2003]:

$$\sigma_0 = \sigma \sqrt{\frac{1-c}{n} + c} \tag{4}$$

with c the error correlation between the n retrievals. In Boersma et al. [2016], $c=0.15$ is proposed based on the consideration that errors in surface albedo, clouds, a priori NO_2 profile, and aerosols (or lack of description thereof) are typically correlated at the spatiotemporal scales of moderate resolution (global) models, i.e. down to $0.5° \times 0.5°$ and over one month (for example the surface albedo is from a monthly climatology). Eq. (4) with $c=0.15$ implies that the spatially or temporally averaged uncertainty cannot reduce to below 39% of the level of typical single-pixel uncertainties (σ), even when many observations are available.

6.2 Algorithm uncertainties and quality flags

Table 5 gives an overview of the most important uncertainties and the quality flags of QA4ECV NO_2 provided in the data product. Note that the uncertainty estimates and quality flags provide clearly different information to the user. The uncertainty characterizes the dispersion of the NO_2 column, given the value of the measured column, and our best understanding of the retrieval process. Quality flags indicate whether the retrieved value and the uncertainty estimate have been obtained under conditions where they are expected to be valid.

Table 5. Overview of the main uncertainty estimates and quality flags included in the QA4ECV NO_2 ECV precursor product. The third column indicates if the estimate is unique for that pixel, derived from a global estimate, or a blend of individual and global estimates.

Name	Meaning	Per pixel or global	Symbol
Tropospheric NO_2 column uncertainty	Algorithm uncertainty estimate of the tropospheric NO_2 column	Per pixel	σ
Tropospheric NO_2 column uncertainty when averaging kernel is applied	Algorithm uncertainty estimate, as above, but contribution from profile uncertainty removed	Per pixel	σ_{AK}
Stratospheric NO_2 column uncertainty	Global estimate of uncertainty in the stratospheric VCD	Global	$\sigma_{N_{strat}}$
Uncertainty of the sum of the tropospheric and stratospheric vertical NO_2 columns	Algorithm uncertainty estimate of the total NO_2 column	Per pixel	
NO_2 SCD uncertainty	Uncertainty estimated from the DOAS spectral fitting of NO_2	Per pixel	σ_{N_S}
Slant column related	First term on the right hand side of	Per pixel	$\left(\frac{\sigma_{N_S}}{M_{tr}}\right)$

uncertainty of the NO_2 tropospheric vertical column	Eq. (2)		
Stratospheric column related uncertainty of the NO_2 tropospheric vertical column	Second term on the right hand side of Eq. (2)	Mix of pixel and global	$\left(\dfrac{\sigma_{N_{s,strat}}}{M_{tr}}\right)$
Total tropospheric AMF related uncertainty of the tropospheric NO_2 vertical column	Third term on the right hand side of Eq. (2).	Per pixel	$\left(\dfrac{(N_S - N_{S,strat})\sigma_{M_{tr}}}{M_{tr}^2}\right)$
Surface albedo related uncertainty of the tropospheric vertical NO_2 column	Contribution to the uncertainty of uncertainties in the surface albedo in the tropospheric AMF[9]	Per pixel	$\left(\dfrac{\partial M}{\partial A_s}\sigma_{A_s}\right)\dfrac{N_v}{M_{tr}}$
Cloud fraction related uncertainty of the tropospheric vertical NO_2 column	Contribution to the uncertainty of uncertainties in the cloud fraction in the tropospheric AMF	Per pixel	$\left(\dfrac{\partial M}{\partial f_{cl}}\sigma_{f_{cl}}\right)\dfrac{N_v}{M_{tr}}$
Cloud pressure related uncertainty of the tropospheric vertical NO_2 column	Contribution to the uncertainty of uncertainties in the cloud pressure in the tropospheric AMF	Per pixel	$\left(\dfrac{\partial M}{\partial p_{cl}}\sigma_{p_{cl}}\right)\dfrac{N_v}{M_{tr}}$
TM5 profile related uncertainty of the tropospheric vertical NO_2 column	Global estimate of the contribution to the uncertainty of uncertainties in the TM5 NO_2 profile in the tropospheric AMF	Global	$0.1N_v$

[9] The term between brackets indicates the AMF uncertainty caused by uncertainty in the forward model parameter, here surface albedo. These terms are the same as in Eq. (3) in this work and Eq. (12) in Boersma et al. [2004]. To arrive at the contribution of the forward model parameter uncertainties to the NO_2 column uncertainty, we ratio the AMF uncertainty contribution by the AMF itself, and multiply with the actual tropospheric NO_2 column.

processing error flag	Flag indicating whether the processing was successful (0) or failed (-1)	Per pixel	
processing quality flags	Flags indicating conditions that affect the quality of the retrieval	Per pixel	

6.3 Evaluating the sub-step uncertainty estimates

An innovative aspect of the QA4ECV project is the evaluation of the uncertainty estimates of retrieval sub-steps against independent estimates of the same metric and structural uncertainties.

6.3.1 Evaluation of NO$_2$ SCD uncertainties

We compared the DOAS uncertainty estimates (σ_{N_S}) from the spectral fitting algorithm against independent estimates obtained from the spatial variability of an ensemble of DOAS SCDs over areas with little geophysical variability using a statistical approach [Boersma et al., 2007]. Our SCD

10   uncertainty evaluation is described in detail in QA4ECV Deliverable 5.5 [Boersma et al., 2017[a]] and in Zara et al. [2018] for OMI and GOME-2A, and we summarize results here. For both instruments, we found that the improved QA4ECV OMI NO$_2$ retrieval shows smaller uncertainties than other OMI algorithms and good agreement between the DOAS and statistical SCD uncertainties. This suggests that the recommendations made in section 5.1 and in QA4ECV D4.2 [Müller et al., 2016] have improved the

15   spectral fitting of NO$_2$ such, that the typical mission-average SCD uncertainties for both instruments amount to 0.7-0.8×10^{15} (was ~1.0×10^{15}) molec. cm^{-2}. For OMI, this uncertainty is dominated by random contributions from propagation of measurement noise, but we also noticed a 30% systematic contribution from stripe effects. For OMI, the trend in SCD uncertainties was small (<2%/yr) in line with the known radiometric stability of the instrument [Schenkeveld et al., 2017], but for GOME-2A,

20   the NO$_2$ SCD uncertainties increased by 8%/yr until September 2009, and after heating the instrument by <3%/yr over 2009-2015. The structural (systematic) uncertainty, estimated from the differences

between NO$_2$ SCDs calculated with different but equally plausible fitting methods (with or without intensity offset correction, see section 5.1) is larger than but of similar magnitude as the theoretical and statistical estimates. Table 6 gives an overview of the various estimates of uncertainty for the NO$_2$ SCDs.

6.3.2 Evaluation of uncertainties in the stratospheric correction

The uncertainty of the stratospheric NO$_2$ vertical column in QA4ECV NO$_2$ product is based on a global statistical analysis of results from the data assimilation procedure, and documented as 0.2×10^{15} molecules cm^{-2} [Dirksen et al., 2011]. The assimilation predicts stratospheric NO$_2$ columns from an

10 observation-constrained (analyzed) start field and TM5-modeled transport and chemistry. The average discrepancies between the 24-hour forecast and actual satellite-observed NO$_2$ slant column fields over pristine areas are regarded as a measure for the uncertainty in the stratospheric NO$_2$ field. In QA4ECV Deliverable 5.5 [Boersma et al., 2017[a]], we verified that the observation minus forecast (O-F) assimilation statistics over the Pacific are indeed consistent with an uncertainty estimate of 0.2×10^{15}

15 molecules cm^{-2} for the stratospheric column.

To further evaluate the estimate of the stratospheric column uncertainty, we compare the QA4ECV data assimilation and STREAM OMI stratospheric NO$_2$ column estimates for 2 February 2005. There are considerable methodological differences between the data assimilation and STREAM techniques. Yet

20 the data assimilation and STREAM stratospheric NO$_2$ distributions agree to reasonable extent, with data assimilation stratospheric columns generally smaller and their spatial features sharper than in STREAM. The TM5-MP assimilation approach distinguishes stratospheric NO$_2$ from free-tropospheric background contributions, while STREAM does not do this. This may be a main reason for the structurally lower values in the assimilation. This is illustrated in Figure 7, which shows the meridional variability in the

25 stratospheric NO$_2$ column from data assimilation and from STREAM along 40°N on 2 February 2005. Between 75°-125°W over the United States and between 0-40° E (Europe), the data assimilation stratospheric columns values are $0.2\text{-}0.5 \times 10^{15}$ molecules cm^{-2} lower than the STREAM values. Over

eastern Asia (100-140° E), data assimilation and STREAM agree to within $\pm 0.3 \times 10^{15}$ molecules cm^{-2}. These differences reflect the structural uncertainty in stratospheric (vertical) NO_2 columns, arising when different retrieval methodologies are applied to the same satellite observations, and both uncertainty estimates are included in Table 6.
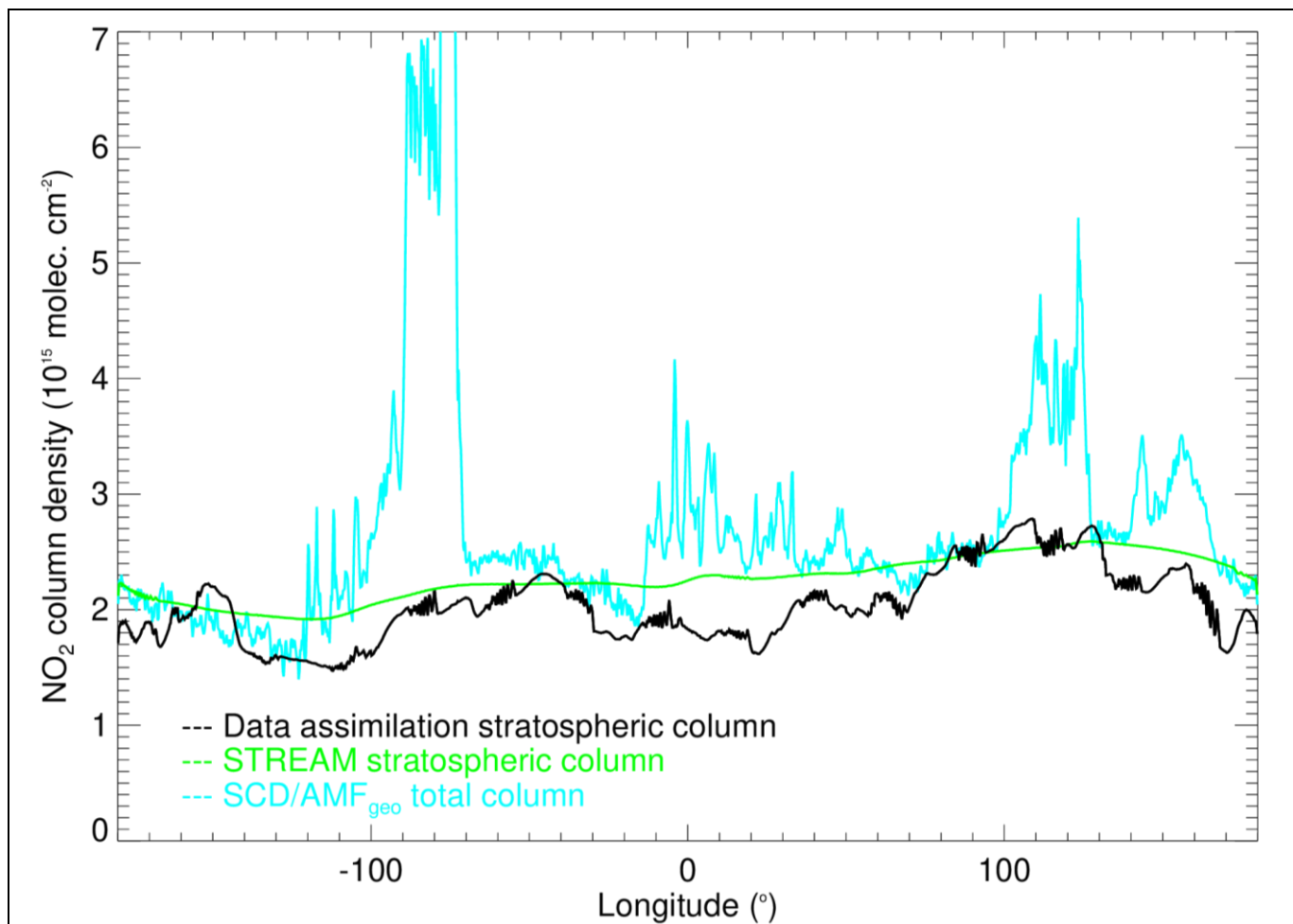
Figure 7. Meridional average QA4ECV OMI NO_2 column averaged over 39-41°N on 2 February 2005. No cloud radiance, albedo, or AMF filtering has been applied. Data assimilation and STREAM stratospheric columns are indicated in the black and green lines, the total slant columns divided by the geometric AMF is light blue. Both data assimilation and STREAM stratospheric column estimates are included in the QA4ECV NO_2 product.

6.3.3 Evaluation and breakdown of uncertainties in the air mass factors

The uncertainty in the tropospheric AMF is calculated via the uncertainty propagation from Eq. (3). The contribution of each parameter to the overall AMF uncertainty depends on the specific observation conditions for each pixel. The air mass factor sensitivities (e.g. $\frac{\partial M}{\partial A_s}$) describe the sensitivity of the AMF to changes in the local parameter value, evaluated around the specific value for the parameter at the pixel. The uncertainties in the cloud parameters ($\sigma_{f_{cl}}, \sigma_{p_{cl}}$), surface albedo ($\sigma_{A_s}$), and the a priori profile shape have been estimated from the literature or derived from comparisons with independent data. For QA4ECV OMI NO_2, we use an uncertainty in the surface albedo of 0.015, based on various comparisons of albedo databases (e.g. Boersma et al., [2011]), uncertainties of 0.025 and 50 hPa in the OMI O_2-O_2 cloud fraction and cloud pressure estimates, respectively, based on recent improvements in the cloud algorithm [Veefkind et al., 2016], and a 10% contribution from NO_2 profile uncertainty. The latter is based on comparing AMFs calculated with simulated a priori profiles to AMFs calculated with measured NO_2 profiles from aircraft and lidar (e.g. Hains et al. [2010] and references therein).

Apart from the overall AMF uncertainty estimate, the QA4ECV NO_2 ECV precursor data product also provides the individual contributions by the cloud parameters, surface albedo, and a priori profile shapes. Figure 8 presents the relative monthly average tropospheric AMF uncertainties and their individual contributions from surface albedo, cloud and profile uncertainties (not shown because they have been set at the 10% level) for OMI throughout 2005 over Europe, the United States, China and Johannesburg, South Africa, regions polluted with NO_2. The largest contribution to AMF uncertainty is from surface albedo-cloud cross-term contribution (10-20%), with surface albedo contributions contributing substantially ($\pm$10%). In winter the uncertainty in cloud pressure is a substantial contributor in Europe and China. The strong surface albedo-cloud fraction cross-term ($2\left(\frac{\partial M}{\partial A_s}\frac{\partial M}{\partial f_{cl}}\langle\epsilon_{f_{cl}}\epsilon_{A_s}\rangle\right)$) can be understood from the strong sensitivity of the cloud fraction to the surface albedo, especially when cloud fractions are small (see Appendix in Boersma et al. [2004]). The overall tropospheric AMF uncertainties are estimated to be 20-25%, comparable to earlier estimates for GOME tropospheric NO_2

presented in Boersma et al. [2004].

Figure 8. Average (single-pixel) QA4ECV OMI tropospheric AMF uncertainty (black line) estimated for Europe (40-55°N, 10°W-15°E), United States (35-45°N, 100°W-75°W), China (35-45°N, 110-140°E), and Johannesburg (24-28°S, 26-30°E) in 2005. The coloured lines indicate the contribution to AMF uncertainty from the various inputs to the AMF calculation indicated in the legend. The contribution from a priori profile uncertainty is assumed to be constant at 10% of the AMF uncertainty (not plotted).

We quantified the structural uncertainty in tropospheric AMFs by comparing an ensemble of different

5 AMF calculation methods and parameter assumptions over eastern China, a region with high amounts

and a complex mixture of aerosols, clouds, and NO_2 pollution [Lorente et al., 2017]. Retrieval groups used their preference for ancillary data and preferred cloud and aerosol corrections. The outcome of the comparisons suggested systematic AMF differences of up to 15% in Summer and 40% in Winter between the groups. We consider these structural uncertainty estimates to be conservative, as they have been calculated for the particularly challenging retrieval regime of eastern China in 2005. Including the structural uncertainties in the overall budget, as done for the QA4ECV HCHO ECV precursor product [De Smedt et al., 2017], would bring tropospheric AMF uncertainties to ±30% in Summer and ±50% in Winter.

Table 6. Comparison of uncertainty estimates for the main QA4ECV OMI NO_2 retrieval steps. The SCD and stratospheric SCD uncertainties are representative for all possible retrieval scenarios. AMF uncertainties are representative for situations with high NO_2.

	Algorithm uncertainty (molec. cm^{-2})	Independent uncertainty estimate (molec. cm^{-2})	Structural uncertainty (molec. cm^{-2})
SCD (σ_{N_S})	0.8×10^{15} [a]	0.7×10^{15} [a]	1.0×10^{15} [b]
Stratospheric SCD ($\sigma_{N_{s,strat}}$)	$0.2 \times 10^{15} \cdot M_{strat}$ [c]		$(0.2\text{-}0.5) \times 10^{15} \cdot M_{strat}$ [d]
Tropospheric AMF ($\sigma_{M_{tr}}$)	20% (Summer) 25% (Winter)		15% (Summer)[e] 40% (Winter)[e]

[a]Zara et al. [2018]

[b]Section 5.1 of this work

[c]Dirksen et al. [2011] and analysis of data assimilation observation minus forecast differences QA4ECV Deliverable 5.5 [Boersma et al., 2017[a]].

[d]Figure 7 of this work.

[e]Lorente et al. [2017]

6.4 Overall uncertainties in tropospheric NO_2 columns

6.4.1 Uncertainties in single-pixel tropospheric NO_2 columns

Here we present estimates of typical algorithm, single-pixel uncertainties for the QA4ECV NO_2 columns in four regions: Europe, United States, and China as showcases for typical polluted regions, and the Pacific Ocean as an example of a remote region, with low, background levels. These uncertainty estimates should be interpreted as representative for typical, single-pixel uncertainties encountered by users interpreting the data. We see from Figure 9 that over the polluted regions in wintertime, the single-pixel retrieval uncertainty is dominated by the uncertainty in the tropospheric AMF. In Summer, contributions from uncertainties in the SCD are largest, but with comparable contributions from uncertainties in the stratospheric correction and the tropospheric AMF. On average a single pixel is 35%-45% uncertain in the polluted regions. Over the background region (Pacific Ocean), we see that the tropospheric NO_2 column uncertainty exceeds 100%, and is dominated year-round by the uncertainties in SCD and stratospheric column estimate.
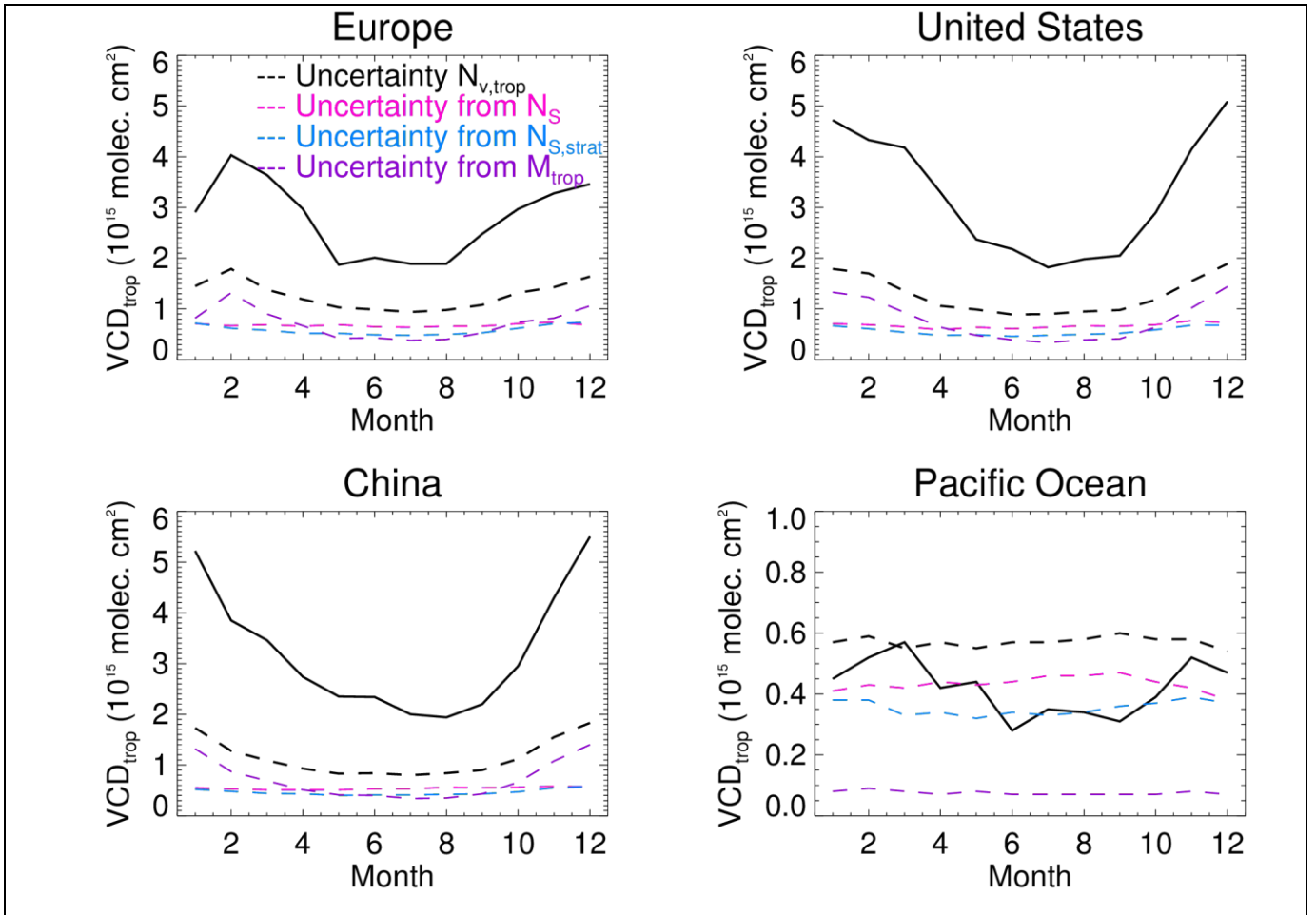
Figure 9. Average (single-pixel equivalent) QA4ECV OMI tropospheric NO_2 columns (solid line) and associated total uncertainties (dashed black line) for Europe (40-50°N, 10°W-15°E), United States (35-45°N, 100-75°W), eastern China (30-45°N, 110-140°E), and the Pacific Ocean (35-45°N, 160-140°W) in 2005. The dashed coloured lined indicate the contributions to the tropospheric NO_2 column uncertainty from SCD (pink, $\left(\frac{\sigma_{N_S}}{M_{tr}}\right)$), stratospheric correction (light blue, $\left(\frac{\sigma_{N_{S,strat}}}{M_{tr}}\right)$), and the tropospheric AMF (purple, $\left(\frac{(N_S - N_{S,strat})\sigma_{M_{tr}}}{M_{tr}^2}\right)$).

6.4.2 Uncertainties in averaged tropospheric NO₂ columns

When averaging tropospheric columns over space, uncertainties may be considerably reduced. For example, over regions such as the Pacific Ocean, where the uncertainty is dominated by random SCD

error, the tropospheric column uncertainty will be much reduced when averaging over a month or over a larger region. Over polluted regions, dominated by uncertainties in the tropospheric AMF, averaging will also reduce the tropospheric column uncertainties, but an unknown systematic component will remain. For both retrieval situations, we adopt Eq. (4), to account for possible systematic errors arising

5   from imperfect knowledge of surface albedo, a priori NO_2 profile, clouds, and correlations between these.

In model-column comparisons and in trend analysis studies, it is often important to have knowledge of temporally averaged uncertainties. Because the temporal variability in tropospheric NO_2 columns is

10   typically strong (because of the diurnal cycle, day-to-day variability, weekly cycles, etc.), this implies considerable variability in day-to-day uncertainties. To obtain the uncertainty in a monthly mean tropospheric NO_2 column over a certain region, we recommend taking whichever is largest: (a) the temporally averaged values for σ_o (Eq. (4)), or (b) the standard deviation of the mean (standard error) of the daily tropospheric NO_2 columns. If there is substantial temporal variability (from changes in

15   photochemistry, transport events), the standard error will be a good representation of the uncertainty in the monthly mean tropospheric NO_2 column. Figure 10 shows a comparison of monthly averaged uncertainties σ_o and the local standard deviation of the mean NO_2 columns for four small regions ($0.25°$ $\times$ $0.25°$). The figure confirms that the averaged uncertainties provide an optimistic estimate of the uncertainty, at $\pm 10\%$, in the monthly mean NO_2 columns. For the polluted regions, the standard

20   deviation of the mean is 15-30%, exceeding the average uncertainties. This illustrates that calculating the uncertainty in a monthly mean over a small region such as a city, is more driven by sampling limitations, than by the intrinsic uncertainty of the retrieval.
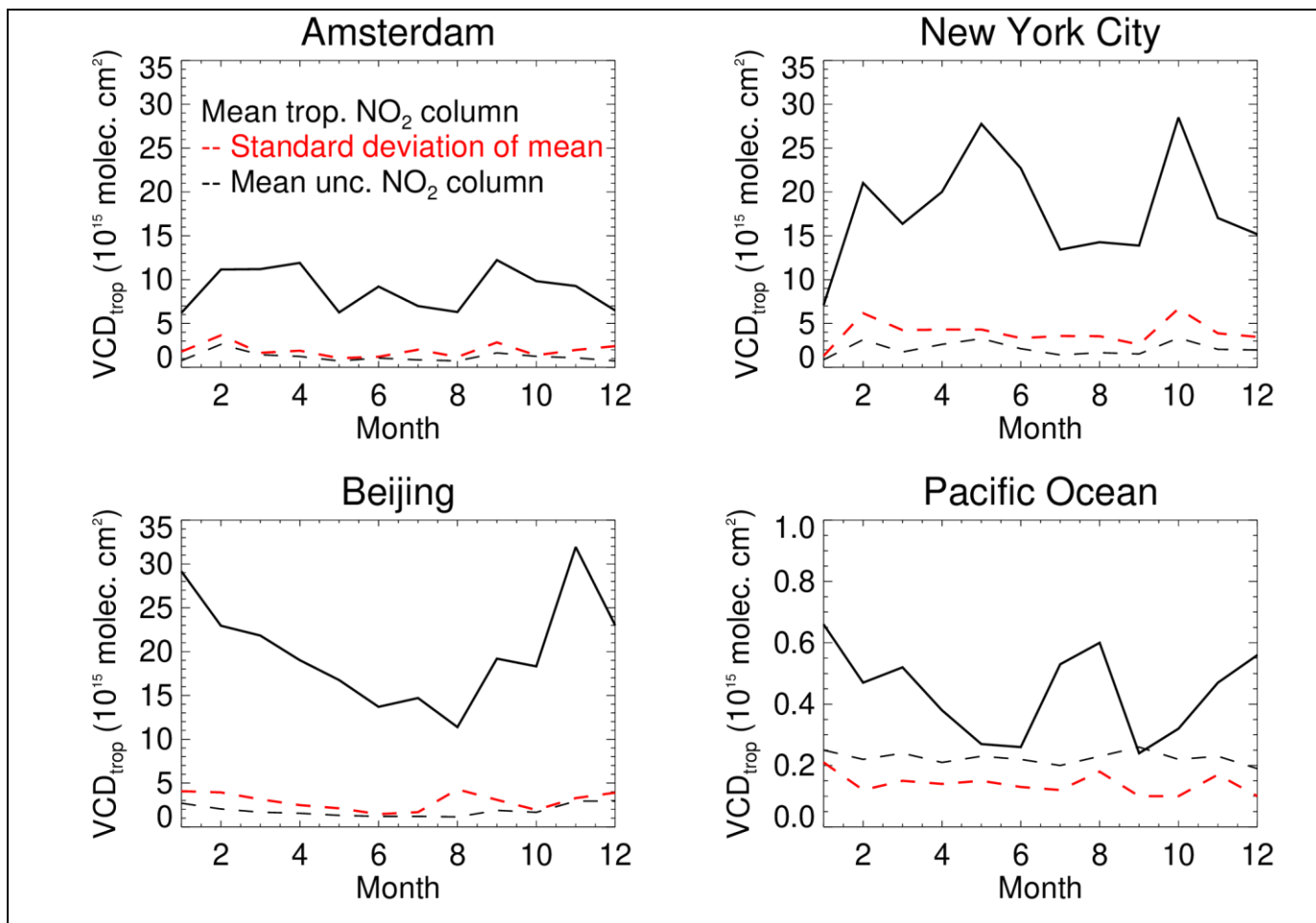
Figure 10. Monthly mean single grid-cell QA4ECV OMI tropospheric NO_2 columns (solid line), standard deviation of the mean (standard error, dashed red line), and super-observation uncertainty (σ_0, dashed black line) in 2005 over Amsterdam (52.375°N, 4.875°E), New York City (40.875°N, 73.875°W), Beijing (39.875°N, 116.375°E), and the Pacific (39.875°N, 149.875°W). Grid cell size 0.25° × 0.25°. Only pixels with cloud radiance fraction < 0.5 were included in the calculation.

7 Validation of QA4ECV NO_2 columns and uncertainties

As an example of the validation efforts taken within QA4ECV, we here compare the QA4ECV OMI tropospheric NO_2 with independent MAX-DOAS column measurements in the polluted city of Tai'an, China. We compare OMI pixels measured within 20 km and 30 minutes of a MAX-DOAS

measurement in Tai'an. We validate both the QA4ECV v1.1 and the well-established DOMINO v2 product for reference.

The MAX-DOAS measurements were conducted by Irie et al. [2008] in the Chinese city of Tai'an in May-June 2006 when pollution levels were substantial. The instrumentation and retrieval technique have been described extensively in Irie et al. [2008; 2012]. The slant column retrievals have been tested in a semi-blind inter-comparison exercise in Cabauw, The Netherlands, indicating agreement to within 10% with other groups [Roscoe et al., 2010]. Uncertainties in the MAX-DOAS NO_2 columns are driven by noise, air mass factor and temperature uncertainties amounting to approximately 15% uncertainty. The representative horizontal 'footprint' of the MAX-DOAS measurement is on the order of 10 km. It was suggested by Irie et al. [2012] that the spatial distribution of NO_2 tropospheric columns around Tai'an during their observation period was rather homogeneous compared to other sites used for their validation comparisons. More quantitative characterization on this aspect will be discussed below.

We compare OMI NO_2 tropospheric columns measured with a pixel center within 20 km of the location of the MAX-DOAS instrument in Tai'an (for some days more than one pixel can be matched up with a MAX-DOAS measurement). This coincidence criterion limits spatial representativeness mismatches between MAX-DOAS and OMI, and is consistent with the spatial dimensions of the MAX-DOAS (±10 km) and OMI (20-30 km) footprints (we excluded pixels from the outer 4 OMI rows). We furthermore require that the OMI columns were measured within 30 minutes of the coinciding MAX-DOAS measurement, have a pixel footprint area < 700 km^2, and that the satellite retrieval was done under mostly clear-sky conditions (cloud radiance fraction <0.5), which is in line with recommendations on the appropriate use of QA4ECV data as documented in the Product Specification Document [QA4ECV Deliverable D4.6]. Earlier studies (e.g. Pinardi et al. [2017]; Drosoglou et al. [2017]) found largest discrepancies between MAX-DOAS and satellite NO_2 columns over strongly polluted regions. Such discrepancies are at least partly due to spatial inhomogeneity in the NO_2 field around the station location. To quantify the spatial representativeness of the Tai'an MAX-DOAS site for the OMI pixels included in the comparison, we calculated the campaign-mean spatial tropospheric NO_2 column

distribution (Figure 11). We then use the ratio of this campaign-mean column at Tai'an to the campaign-mean column at the location of the individual OMI pixel to project individual OMI NO_2 columns ($N_{V,p}$, i.e. what is usually validated) within our criteria to values more representative for the location of the Tai'an ($N_{V,T}$):

$$N_{V,T} = \left(\frac{\overline{N_{V,T}}}{\overline{N_{V,p}}}\right) \cdot N_{V,p} \qquad (5)$$

For example, for a pixel observed directly southwest of Tai'an, where pollution levels are somewhat higher than directly over Tai'an, the scaling factor will be smaller than 1. For the coincidence criterion

10 of 20 km used here, the scaling factors stay close to 1 and modifications do not exceed 1×10^{15} molec.cm^{-2} (<20% of the Tai'an column, see Figure S2).

Figure 11. Campaign mean (30 May – 30 June 2006) QA4ECV tropospheric NO_2 column distribution over eastern China for clear-sky situations (cloud radiance fraction<0.5). The black circle indicates the location of Tai'an, where Chiba University operated the MAX-DOAS instrument. One cell corresponds to $0.1° \times 0.1°$. On average there are 15 satellite pixels per cell used to calculate the campaign mean.

We match each OMI pixel fulfilling the spatio-temporal coincidence criteria with the corrected MAX-DOAS NO_2 columns. Discarding pixels with effective cloud pressures >875 hPa (often indicative of aerosol haze), we find 31 QA4ECV OMI pixels matching 13 independent MAX-DOAS measurements collected over 7 different days. Figure 12(a) shows a scatter plot of QA4ECV vs. MAX-DOAS tropospheric NO_2 columns for Tai'an. We find a bias (mean difference) of -0.15×10^{15} molec.cm^{-2} (-2%) and the root mean squared deviation is 1.08×10^{15} molec.cm^{-2} (16%). Not applying the scaling factors

from Eq. (5) leads to a bias of -0.47×10^{15} molec.cm^{-2} (-7%) and a root mean squared deviation of 1.19×10^{15} molec.cm^{-2} (18%). Using a reduced major axis regression analysis, we find a relationship between QA4ECV (y) and MAX-DOAS NO$_2$ columns (x) as $y = -0.86 \times 10^{15}$ molec.cm^{-2} + 1.10x (R^2=0.26, n=31). Including also pixels with high effective cloud pressures (>875 hPa) leads to a bias of

5 -0.48×10^{15} molec.cm^{-2} (-6.6%, n=37) and a root mean squared deviation of 1.35×10^{15} molec.cm^{-2} (20%).

Figure 12(b) shows the scatter plot of DOMINO v2 vs. MAX-DOAS NO$_2$ columns for Tai'an. There are now 45 DOMINO v2 pixels matching with 17 independent MAX-DOAS measurements. This higher number of matches can be explained from the previous version of the OMI O$_2$-O$_2$ cloud product

10 [Acarreta et al., 2004], used in the DOMINO v2 retrieval, containing effective cloud pressures that are too low compared to independent information [Boersma et al., 2011; Veefkind et al., 2016], so that more OMI pixels pass the selection criteria. The bias for DOMINO v2 is $+0.85 \times 10^{15}$ molec.cm^{-2} (+11%, n=45), with a root mean squared deviation of 2.66×10^{15} molec.cm^{-2} (35%).
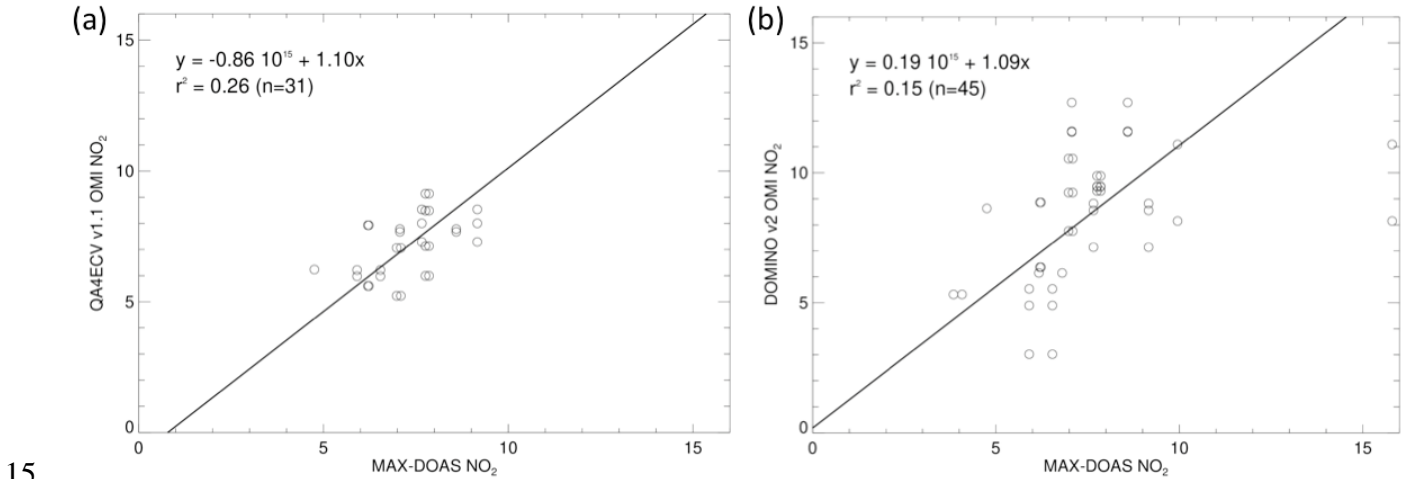
Figure 12. (a) Scatterplot of QA4ECV OMI vs. MAX-DOAS tropospheric NO$_2$ columns for Tai'an (China) in May-June 2006. The solid line shows the result of a reduced major axis regression to the data. Only pixels measured a cloud radiance fraction < 0.5 and an effective cloud pressure <875 hPa, within 20 km and 30 minutes of a MAX-DOAS measurement have been selected. (b) Same as (a), but

20 now for DOMINO v2 vs. MAX-DOAS tropospheric NO$_2$.

The differences between OMI and MAX-DOAS NO_2 columns provide an opportunity to evaluate the uncertainties of the satellite retrievals. This relies on good knowledge of the MAX-DOAS uncertainties and relatively small uncertainties associated with the representativeness of MAX-DOAS for the coincident OMI columns. Assuming that the retrieval errors between OMI and QA4ECV are independent and follow a normal distribution, we expect that the distribution of the differences between OMI and MAX-DOAS takes on a Gaussian form characterized by width $\sigma = (\sigma_O^2 + \sigma_{MD}^2 + \sigma_R^2)^{1/2}$, with σ_O the uncertainty reported (in the data files) for QA4ECV OMI NO_2 columns, σ_{MD} the uncertainty reported for MAX-DOAS NO_2 columns, and σ_R the uncertainty from spatiotemporal mismatches between the satellite and ground-based measurement (Table 7). The mean reported uncertainties (σ_O and σ_{MD}) are regarded as random errors here (see discussion in Section 2.3 in Boersma et al. [2004], and below). In Figure 13 we compare the distribution of differences predicted from the above Gaussian function based on the uncertainties reported in the OMI and MAX-DOAS data files and a 10% representativeness difference error (estimated from deviations from the Tai'an value shown in Figure S2), to the actually observed differences (individual pairs of OMI and MAX-DOAS NO_2 column values). We see from Figure 13 that the differences between OMI and MAX-DOAS NO_2 columns are more narrowly distributed than expected from algorithm uncertainties and theory, although the sample size is small (*n*=31 for QA4ECV, *n*=45 for DOMINO v2). This holds for QA4ECV differences, which are 39% smaller than expected, but also for the DOMINO v2 differences, 16% smaller than expected over Tai'an. The tighter distribution of the observed differences implies that: (1) the uncertainties in OMI and MAX-DOAS retrievals possess some degree of correlation (for instance in situations when OMI is biased high, also MAX-DOAS may be biased high, limiting the magnitude of the differences), (2) OMI and/or MAX-DOAS algorithm uncertainty estimates are too conservative, (3) OMI or MAX-DOAS uncertainties contain an unknown persistent error component, so that σ_O or σ_{MD} have been overestimated, or (4) a combination of the above. The MAX-DOAS NO_2 retrieval technique suffers from some similar error contributions (a priori NO_2 profile shape, aerosols) but it is also different by design from the satellite retrieval (no albedo or stratospheric correction dependence, ground-based perspective), so we should expect some but not full error correlation. If there would be a substantial

systematic and persistent error component to σ_O or σ_{MD} (and we have no indication for this nor do we know about its magnitude or sign), we would have needed to reduce our estimates for σ_O and σ_{MD} in Table 7 and expect a distribution of the differences that is more narrowly Gaussian and peaking at a typical systematic difference (or bias). Figure 13 shows little bias for QA4ECV. We therefore conclude

5   that the OMI (and MAX-DOAS) retrieval uncertainties estimates could be too conservative, although our findings are based on a small sample. In the case of QA4ECV, a reduction of both the OMI and MAX-DOAS uncertainties by 35%, would be in much better agreement with the observed differences at the Tai'an station.

10   **Table 7.** Expected and observed differences between OMI and MAX-DOAS NO$_2$ columns observed over Tai'an in June 2006 for the QA4ECV (n=31) and DOMINO v2 (n=45) ensemble. Summary of uncertainties for the all (31 or 45) matching pixels. σ_O (reported OMI uncertainty) and σ_M (reported MAX-DOAS uncertainty) are the mean of 31 or 45 individual values, and σ_R is considered to be a 10% contribution from mismatches.

	Expected differences (QA4ECV)	Observed differences (QA4ECV)	Expected differences (DOMINO v2)	Observed differences (DOMINO v2)
σ	2.11×10^{15} molec. cm^{-2}	1.29×10^{15} molec. cm^{-2}	2.57×10^{15} molec. cm^{-2}	2.16×10^{15} molec. cm^{-2}
σ_O	1.71×10^{15} molec. cm^{-2}		2.21×10^{15} molec. cm^{-2}	
σ_M	1.08×10^{15} molec. cm^{-2}		1.27×10^{15} molec. cm^{-2}	
σ_R	0.60×10^{15} molec. cm^{-2}		0.60×10^{15} molec. cm^{-2}	

This first validation is based on a limited time range and one site. A more comprehensive validation work, based on several MAXDOAS sites and several years of data is in preparation [Compernolle 2018].
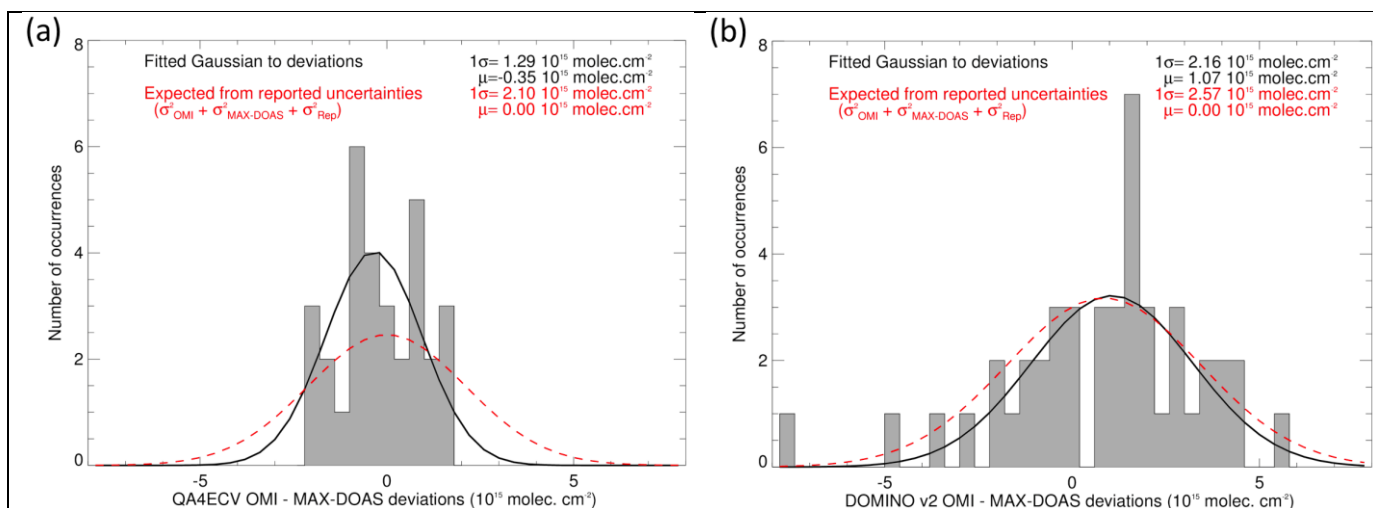
Figure 13. (a) Histogram of differences of QA4ECV OMI vs. MAX-DOAS tropospheric NO_2 columns for Tai'an (China) in May-June 2006. The black line shows a Gaussian fit to the observed differences, and the red dashed lines shows the Gaussian expected from the uncertainties reported in the QA4ECV and MAX-DOAS data products. (b) Same as (a), but now for differences between DOMINO v2 vs. MAX-DOAS tropospheric NO_2.

8 Data availability and DOI

The QA4ECV NO_2 Essential Climate Variable precursor product contains vertical NO_2 columns for the period 1995-2017. The dataset contains: (1) the tropospheric vertical column density, (2) the stratospheric vertical column density, and (3) the total vertical column density. The NO_2 ECV precursor data provides geophysical information for each and every ground pixel observed by GOME, SCIAMACHY, OMI, and GOME-2(A). The QA4ECV NO_2 data product is available online via www.qa4ecv.eu, under 'ECV data'. The data product has been processed with the coherent algorithm described in this work.

For GOME, data is available from 1 July 1995 until 30 June 2003 (8 years). For SCIAMACHY, data is available from 1 August 2002 until 30 April 2012 (9 years and 9 months). For GOME-2(A), data is available from 1 January 2007 until 31 December 2017, and for OMI from 1 October 2004 until 31

December 2017, so that the total length of the data set exceeds 22 years at the time of writing. For each of the data sets, digital object identifiers have been registered [Boersma et al., 2017[b,c,d,e]]. Detailed information on how to use the data can be found in the Product Specification Document for NO_2 ECV Precursor product [Boersma et al., 2017[f]].

9 Summary

We have developed an improved algorithm and uncertainty assessment for tropospheric NO_2 satellite retrievals from UV/VIS satellite sensors. Our effort has resulted in the generatation of a 1995-2017 climate data record of tropospheric NO_2 columns with fully traceable uncertainty metrics that can be

10   readily used for model evaluation, for estimating NO_x emissions and nitrogen deposition. In designing our new algorithm, we followed advice from the user and producer community and from WMO GCOS best practices on generating climate data records. Specifically, we extended the information content on flags and uncertainties in the data files, and present a so-called traceability chain along with the data files. This traceability chain is an easily accessible web-hosted interactive flow diagram that shows the

15   components of the QA4ECV NO_2 algorithm, and how external information is embedded in the retrieval process, providing details on where those pieces of information can be found.

The QA4ECV project involved detailed comparisons of different approaches between groups for the DOAS slant column retrievals and the estimate of the stratospheric sub-column and air mass factors.

20   Using the latest and best available level-1 data for GOME, SCIAMACHY, GOME-2A, and OMI from the relevant space agencies, the comparisons led us to improve the spectral fitting of NO_2 by accounting for liquid water absorption, and an intensity-offset correction. This improved the quality of the NO_2 fit over clear-sky ocean scenes by up to 30% [Zara et al., 2018], but did not substantially affect the NO_2 fits over polluted scenes. We compared three alternate methods for estimating the stratospheric NO_2

25   background: data assimilation [Dirksen et al., 2011], and the model-based [Hilboll et al., 2013] and modified reference sector (STREAM, Beirle et al., [2016]) approaches. Data assimilation was considered to be the most viable option for the QA4ECV algorithm because it provides a coherent framework for stratospheric corrections as well as air mass factor (AMF) calculations. We based the

data assimilation on the TM5-MP chemistry transport model with $1° \times 1°$ horizontal resolution, a major step forward compared to earlier assimilation schemes based on TM4 ($3° \times 2°$), and include corrections for sphericity effects on atmospheric radiative transfer, as described in Lorente et al. [2017]. Our new stratospheric correction leads to fewer negative tropospheric NO_2 columns for retrievals at extreme viewing geometries. We then tested various models and approaches to calculate tropospheric AMFs under challenging retrieval scenarios. AMFs calculated with different radiative transfer models agree well, as long as assumptions and ancillary data inputs are consistent. With groups using their own preferred settings, we find differences (or structural uncertainty) in AMFs up to 40% with respect to the ensemble mean, stressing the importance of adequate traceability. Many of the lessons learned for QA4ECV algorithm development, are currently being applied to NO_2 retrievals from S5P-TROPOMI.

The QA4ECV NO_2 product contains an algorithm uncertainty estimate associated with each individual observation. We obtain this estimate via uncertainty propagation calculations, accounting for pixel-specific sensitivities to state parameters (Jacobians) such as surface reflectance, clouds, and the NO_2 vertical profile. The uncertainties are highest in the cold season, when AMFs are particularly uncertain, and typically amount to 40% over polluted area. For averaged QA4ECV NO_2 data, associated uncertainties may be reduced, but part of the uncertainty due to systematic error will remain. Our work provides recommendations on how to estimate the uncertainty for spatially or temporally averaged data, taking into account a partial correlation in the errors between pixels. We evaluated the algorithm uncertainties against independent assessments of structural uncertainties for each retrieval step and find that the structural uncertainties are of similar magnitude or exceed the algorithm uncertainties for all retrieval sub-steps. Finally, we used MAX-DOAS NO_2 column measurements obtained over the polluted Tai'an (China) region in June 2006 to validate the OMI QA4ECV NO_2 columns and their uncertainties. Accounting for spatial differences between the pixel and the location of Tai'an, we found good agreement between the QA4ECV and MAX-DOAS NO_2 columns (bias = -2%, RMS differences 16%, $n=31$), much better than the agreement between DOMINO v2 and MAX-DOAS (bias = +11%, RMS 35%, $n=45$). The small differences between coinciding QA4ECV and MAX-DOAS NO_2 columns

suggest that our QA4ECV algorithm uncertainties are likely on the conservative side, at least over Tai'an.

Acknowledgments

5   This research has been supported by the EU FP7 Project Quality Assurance for Essential Climate Variables (QA4ECV), grant no. 607405. We acknowledge the free use of the GOME-2 data provided by EUMETSAT. We thank all the persons who have kindly and carefully responded to QA4ECV user survey. We appreciate the efforts by Michael Barkley (Leicester University), Lok Lamsal (NASA GSFC), Jin-Tai Lin and Mengyao Liu (Peking University) who kindly contributed to the NO_2 AMF
10   intercomparison.

References

Anand, J. S., Monks, P. S., and Leigh, R. J.: An improved retrieval of tropospheric NO_2 from space over polluted regions using an Earth radiance reference, Atmos. Meas. Tech., 8, 1519-1535,
15   https://doi.org/10.5194/amt-8-1519-2015, 2015.

Beirle, S., Sihler, H., and Wagner, T.: Linearisation of the effects of spectral shift and stretch in DOAS analysis, Atmos. Meas. Tech., 6(3), 661-675, doi:10.5194/amt-6-661-2013, 2013.

20   Beirle, S., Hörmann, C., Jöckel, P., Liu, S., Penning de Vries, M., Pozzer, A., Sihler, H., Valks, P., and Wagner, T.: The STRatospheric Estimation Algorithm from Mainz (STREAM): estimating stratospheric NO_2 from nadir-viewing satellites by weighted convolution, Atmos. Meas. Tech., 9, 2753-2779, https://doi.org/10.5194/amt-9-2753-2016, 2016.

25   Boersma, K. F., Eskes, H. J., and Brinksma, E. J.: Error analysis for tropospheric NO_2 retrieval from space, J. Geophys. Res., 109, D04311, doi:10.1029/2003JD003962, 2004.

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., Gleason, J. F., and Bucsela, E. J.: Near-real time retrieval of tropospheric NO_2 from OMI, Atmos. Chem. Phys., 7, 2103–2118, doi:10.5194/acp-7-2103-2007, 2007.

5   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.: An improved tropospheric NO_2 column retrieval algorithm for the Ozone Monitoring Instrument, Atmos. Meas. Tech., 4, 1905-1928, https://doi.org/10.5194/amt-4-1905-2011, 2011.

10   Boersma, K. F., Vinken, G. C. M., and Eskes, H. J.: Representativeness errors in comparing chemistry transport and chemistry climate models with satellite UV–Vis tropospheric column retrievals, Geosci. Model Dev., 9, 875-898, https://doi.org/10.5194/gmd-9-875-2016, 2016.

Boersma, K. F., De Smedt, I., George, M., Compernolle, S., Eskes, H. J., Zara, M., van Geffen, J., 15 Lorente, A., Richter, A., Peters, E., Hilboll, A., Yu. H., Van Roozendael, M., Beirle, S., Dörner, S., Wagner, T., Nightingale, J., Lambert, J.-C., Coheur, P.-F., and Clerbaux, C.: Report on the assessment and characterization of uncertainties in the retrieval algorithms for Atmosphere ECV records, 71 pp., http://www.qa4ecv.eu/sites/default/files/D5.5_v1.0.compressed.pdf, last access: 12 April 2018, 2017[a].

20   Boersma, K. F., Eskes, H., Richter, A., De Smedt, I., Lorente, A., Beirle, S., Van Geffen, J., Peters, E., Van Roozendael, M. and Wagner, T.: QA4ECV NO_2 tropospheric and stratospheric vertical column data from GOME (Version 1.1) [Data set]. Royal Netherlands Meteorological Institute (KNMI), http://doi.org/10.21944/qa4ecv-no2-gome-v1.1, 2017[b].

25   Boersma, K. F., Eskes, H., Richter, A., De Smedt, I., Lorente, A., Beirle, S., Van Geffen, J., Peters, E., Van Roozendael, M. and Wagner, T.: QA4ECV NO_2 tropospheric and stratospheric vertical column data from SCIAMACHY (Version 1.1) [Data set]. Royal Netherlands Meteorological Institute (KNMI), http://doi.org/10.21944/qa4ecv-no2-scia-v1.1, 2017[c].

Boersma, K. F., Eskes, H., Richter, A., De Smedt, I., Lorente, A., Beirle, S., Van Geffen, J., Peters, E., Van Roozendael, M. and Wagner, T.: QA4ECV NO$_2$ tropospheric and stratospheric vertical column data from GOME-2A (Version 1.1) [Data set], Royal Netherlands Meteorological Institute (KNMI),

5 http://doi.org/10.21944/qa4ecv-no2-gome2a-v1.1, 2017[d].

Boersma, K. F., Eskes, H., Richter, A., De Smedt, I., Lorente, A., Beirle, S., Van Geffen, J., Peters, E., Van Roozendael, M. and Wagner, T.: QA4ECV NO$_2$ tropospheric and stratospheric vertical column data from OMI (Version 1.1) [Data set], Royal Netherlands Meteorological Institute (KNMI).

10 http://doi.org/10.21944/qa4ecv-no2-omi-v1.1, 2017[e].

Boersma, K. F., van Geffen, J., Eskes, H., van der A, R. J., De Smedt, I., Van Roozendael, M., Yu, H., Richter, A., Peters, E., Beirle, S., Wagner, T., Lorente, A., Scanlon, T., Compernolle, S., and Lambert, J.-C.: Product Specification Document for the QA4ECV NO$_2$ Precursor Product (Version 1.1), available

15 at: http://temis.nl/qa4ecv/no2col/QA4ECV_NO2_PSD_v1.1.compressed.pdf (last access: 25 April 2018), 2017[f].

Bojinski, S., Verstraete, M., Peterson, T. C., Richter, C., Simmons, A., and Zemp, M.: The concept of Essential Climate Variables in support of climate research, applications, and policy, Bull. Am. Meteor.

20 Soc., 1431-1443, doi:10.1175/BAMS-D-13-00047.1, 2014.

Bovensmann, H., Burrows, J. P., Buchwitz, M., Frerick, J., Noel, S., Rozanov, V. V., Chance, K. V., and Goede, A. P. H.: SCIA- MACHY: Mission objectives and measurement modes, J. Atmos. Sci., 56, 127–150, 1999.

Bucsela, E. J., Celarier, E. A., Wenig, M. O., Gleason, J. F., Veefkind, J. P., Boersma, K. F., and Brinksma, E. J.: Algorithm for NO$_2$ vertical column retrieval from the Ozone Monitoring Instrument, IEEE trans. on Geosci. Rem. Sens., 44(5), doi:10.1109/TGRS.2005.863715, 2006.

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.: The Global Ozone Monitoring Experiment (GOME): Mission Concept and First Scientific Results, J. Atmos. Sci., 5   56, 151–175, 1999.

Chance, K., Kurosu, T. P. and Sioris, C. E.: Undersampling correction for array detector-based satellite spectrometers., Appl. Opt., 44(7), 1296–1304, doi:10.1364/AO.44.001296, 2005.

10 Compernolle, S., Lambert, J.-C., Boersma, K. F., Schulz, J., Müller, J.-P., Coheur, P.-F., De Smedt, I., Van Roozendael, M., Blessing, S., George, M., and Gobron, N.: Report on the compliance of ECV records with GCOS and user requirements (QA4ECV Deliverable 6.1), 69 pp., http://www.qa4ecv.eu/sites/default/files/QA4ECV_D6p1_FINAL%2BSMM.pdf, last access: 12 April 2018, 2018.

Crutzen, P. J.: The influence of nitrogen oxides on the atmospheric ozone content, Q.J.R. Meteorol. Soc., 96, 320–325, doi:10.1002/qj.49709640815, 1970.

Deutschmann, T., Beirle, S., Friess, U., Grzegorski, M., Kern, C., Kritten, L., Platt, U., Prados-Román, 20 C., Puķite, J., Wagner, T., Werner, B., Pfeilsticker, K.,: The Monte Carlo atmospheric radiative transfer model McArtim: Introduction and validation of Jacobians and 3D features, Journal of Quantitative Spectroscopy and Radiative Transfer, 112(6), 1119-1137, https://doi.org/10.1016/j.jqsrt.2010.12.009, 2011.

25 Dikty, S., Richter, A., Weber, M., Noel, S., Bovensmann, H., Wittrock, F., and Burrows, J. P.: GOME-2 on MetOp-A Support for Analysis of GOME-2 In-Orbit Degradation and Impacts on Level 2 Data Products – Final Report, Tech. rep., Inst. of Environ. Phys., Bremen, Germany, available at: www.eumetsat.int, document ITT 09/10000262 (last access: 25 September 2012), 2011.

Dirksen, R. J., Boersma, K. F., Eskes, H. J., Ionov, D. V., Bucsela, E. J., Levelt, P. F., and Kelder, H. M.: Evaluation of stratospheric NO_2 retrieved from the ozone monitoring instrument: intercomparison, diurnal cycle, and trending, J. Geophys. Res., 116, D08305, doi:10.1029/2010JD014943, 2011.

DLR, Max-Planck Institute, IUP, RAL Space, Royal Netherlands Meteorological Institute: Sentinel-5P TROPOMI Science Verification Report, Issue 2.1, S5P-IUP-L2-ScVR-RP, Technical Document, 314 pp, https://earth.esa.int/web/sentinel/user-guides/sentinel-5p-tropomi/document-library/-/asset_publisher/w9Mnd6VPjXlc/content/sentinel-5p-tropomi-science-verification-report, last access on 10 22 August 2017, 2015.

Dobber, M. R., Kleipool, Q., Dirksen, R., Levelt, P. F., Jaross, G., Taylor, S., Kelly, T., Flynn, L., Leppelmeier, G., and Rozemeijer, N.: Validation of ozone monitoring instrument level 1b data products, J. Geophys. Res., 113, D15S06, doi:10.1029/2007JD008665, 2008.

Drosoglou, T., Bais, A. F., Zyrichidou, I., Kouremeti, N., Poupkou, A., Liora, N., Giannaros, C., Koukouli, M. E., Balis, D., and Melas, D.: Comparisons of ground-based tropospheric NO_2 MAX-DOAS measurements to satellite observations with the aid of an air quality model over the Thessaloniki area, Greece, Atmos. Chem. Phys., 17, 5829-5849, https://doi.org/10.5194/acp-17-5829-2017, 2017.

Eskes, H. J., van Velthoven, P. F. J., Valks, P. J. M., and Kelder, H. M.: Assimilation of GOME total-ozone satellite observations in a three‐dimensional tracer‐transport model, Q.J.R. Meteorol. Soc., 129, 1663-1681, doi:10.1256/qj.02.14, 2003.

25 Eskes, H. J. and Boersma, K. F.: Averaging kernels for DOAS total-column satellite retrievals, Atmos. Chem. Phys., 3, 1285-1291, https://doi.org/10.5194/acp-3-1285-2003, 2003.

Fayt, C., and Van Roozendael, M.: WinDOAS 2.1 software user manual. Uccle, Belgium, BIRA-IASB, 2001.

Galloway, J. N., Aber, J. D., Erisman, J. W., Seitzinger, S. P., Howarth, R. W., Cowling, E. B., Cosby, B. J.: The Nitrogen Cascade, BioScience, 53(4), 341–356, https://doi.org/10.1641/0006-3568(2003)053[0341:TNC]2.0.CO;2, 2003.

GCOS: Implementation plan for the Global Observing System for climate in support of the UNFCC (2010 Update), https://library.wmo.int/opac/doc_num.php?explnum_id=3851, last access: 12 April 2018, GCOS-138, 180 pp., 2010.

GCOS-200: The Global Observing System for Climate: Implementation Needs, GCOS 2016 Implementation Plan, GCOS No. 200, November 2016. Available at https://library.wmo.int/opac/doc_num.php?explnum_id=3417.

GOME Products and Algorithms, https://earth.esa.int/web/sppa/mission-performance/esa-missions/ers-2/gome/products-and-algorithms/products-information, 2018.

Grewe, V., Dahlmann, K., Matthes, S., and Steinbrecht, W.: Attributing ozone to NO_x emissions: Implications for climate mitigation measures, Atmos. Environm., 59, 102-107, doi:10.1016/j.atmosenv.2012.05.002, 2012.

GUM: Joint Committee for Guides in Metrology (JCGM/WG 1) 100:2008, Evaluation of measurement data – Guide to the expression of uncertainty in a measurement (GUM), ISO/IEC Guide 98-3:2008, http://www.bipm.org/utils/common/documents/jcgm/JCGM_100_2008_E.pdf

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., Wittrock, F., Schoenhardt, A., Wagner, T., Ibrahim, O. W., Van Roozendael, M., Pinardi, G., Gleason, J. F., Veefkind, J. P., and Levelt, P.: Testing and improving OMI DOMINO tropospheric NO_2 using observations from the DANDELIONS and INTEX-B validation campaigns, J. Geophys. Res., 115, D05301, doi:10.1029/2009JD012399, 2010.

Hassinen, S., Balis, D., Bauer, H., Begoin, M., Delcloo, A., Eleftheratos, K., Gimeno Garcia, S., Granville, J., Grossi, M., Hao, N., Hedelt, P., Hendrick, F., Hess, M., Heue, K.-P., Hovila, J., Jønch-Sørensen, H., Kalakoski, N., Kauppi, A., Kiemle, S., Kins, L., Koukouli, M. E., Kujanpää, J., Lambert, J.-C., Lang, R., Lerot, C., Loyola, D., Pedergnana, M., Pinardi, G., Romahn, F., van Roozendael, M., Lutz, R., De Smedt, I., Stammes, P., Steinbrecht, W., Tamminen, J., Theys, N., Tilstra, L. G., Tuinder, O. N. E., Valks, P., Zerefos, C., Zimmer, W., and Zyrichidou, I.: Overview of the O3M SAF GOME-2 operational atmospheric composition and UV radiation data products and data availability, Atmos. Meas. Tech., 9, 383-407, https://doi.org/10.5194/amt-9-383-2016, 2016.

Hilboll, A., Richter, A., Rozanov, A., Hodnebrog, Ø., Heckel, A., Solberg, S., Stordal, F., and Burrows, J. P.: 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, https://doi.org/10.5194/amt-6-565-2013, 2013.

Hoek, G., Krishnan, R. M., Beelen, R., Peters, A., Ostro, B., Brunekreef, B., and Kaufman, J. D.: Long-term air pollution exposure and cardio- respiratory mortality: a review, Environmental Health,12(43), https://doi.org/10.1186/1476-069X-12-43, 2013.

Irie, H., Kanaya, Y., Akimoto, H., Tanimoto, H., Wang, Z., Gleason, J. F., and Bucsela, E. J.: Validation of OMI tropospheric NO_2 column data using MAX-DOAS measurements deep inside the North China Plain in June 2006: Mount Tai Experiment 2006, Atmos. Chem. Phys., 8, 6577-6586, https://doi.org/10.5194/acp-8-6577-2008, 2008.

Irie, H., Boersma, K. F., Kanaya, Y., Takashima, H., Pan, X., and Wang, Z. F.: Quantitative bias estimates for tropospheric NO_2 columns retrieved from SCIAMACHY, OMI, and GOME-2 using a common standard for East Asia, Atmos. Meas. Tech., 5, 2403-2411, https://doi.org/10.5194/amt-5-2403-2012, 2012.

Inness, A., Baier, F., Benedetti, A., Bouarar, I., Chabrillat, S., Clark, H., Clerbaux, C., Coheur, P., Engelen, R. J., Errera, Q., Flemming, J., George, M., Granier, C., Hadji-Lazaro, J., Huijnen, V., Hurtmans, D., Jones, L., Kaiser, J. W., Kapsomenakis, J., Lefever, K., Leitão, J., Razinger, M., Richter, A., Schultz, M. G., Simmons, A. J., Suttie, M., Stein, O., Thépaut, J.-N., Thouret, V., Vrekoussis, M., Zerefos, C., and the MACC team: The MACC reanalysis: an 8 yr data set of atmospheric composition, Atmos. Chem. Phys., 13, 4073-4109, https://doi.org/10.5194/acp-13-4073-2013, 2013.

Jin, J., Ma, J., Lin, W., Zhao, H., Shaiganfar, R., Beirle, S., and Wagner, T.: MAX-DOAS measurements and satellite validation of tropospheric NO2 and SO2 vertical column densities at a rural site of North China, Atmospheric Environment, 133, 12-25, https://doi.org/10.1016/j.atmosenv.2016.03.031, 2016.

Kleipool, Q. L., Dobber, M. R., de Haan, J. F., and Levelt, P. F.: Earth surface reflectance climatology from 3 years of OMI data, J. Geophys. Res.: Atmospheres, 113, https://doi.org/10.1029/2008JD010290, d18308, 2008.

Krotkov, N. A., McLinden, C. A., Li, C., Lamsal, L. N., Celarier, E. A., Marchenko, S. V., Swartz, W. H., Bucsela, E. J., Joiner, J., Duncan, B. N., Boersma, K. F., Veefkind, J. P., Levelt, P. F., Fioletov, V. E., Dickerson, R. R., He, H., Lu, Z., and Streets, D. G.: Aura OMI observations of regional SO_2 and NO_2 pollution changes from 2005 to 2015, Atmos. Chem. Phys., 16, 4605-4629, https://doi.org/10.5194/acp-16-4605-2016, 2016.

Kollonige, D. E., Thompson, A. M., Josipovic, M., Tzortziou, M., Beukes, J. P., Burger, R., Martins, D. K. van Zyl, P. G., Vakkari, V., and Laakso, L.: OMI satellite and ground-based Pandora observations and their application to surface NO_2 estimations at terrestrial and marine sites, *Journal of Geophysical Research: Atmospheres*, 123, 1441–1459, https://doi.org/10.1002/2017JD026518, 2018.

Levelt, P. F., Van den Oord, G. H. J., Dobber, M. R., Malkki, A., Visser, H., de Vries, J., Stammes, P., Lundell, J. O. V., and Saari, H.: The Ozone Monitoring Instrument, IEEE T. Geosci. Remote, 44, 1093–1101, 2006.

10   Liu, S. C., Trainer, M., Fehsenfeld, F. C., Parrish, D. D., Williams, E. J., Fahey, D. W., Hübler, G., and Murphy, P. C.: Ozone production in the rural troposphere and the implications for regional and global ozone distributions, J. Geophys. Res., 92(D4), 4191–4207, doi:10.1029/JD092iD04p04191, 1987.

Lorente, A., Folkert Boersma, K., Yu, H., Dörner, S., Hilboll, A., Richter, A., Liu, M., Lamsal, L. N., 15   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.: Structural uncertainty in air mass factor calculation for NO_2 and HCHO satellite retrievals, Atmos. Meas. Tech., 10, 759-782, https://doi.org/10.5194/amt-10-759-2017, 2017.

20   Lorente, A., Boersma, K. F., Stammes, P., Tilstra, L. G., Richter, A., Yu, H., Kharbouche, S., and Muller, J.-P.: The importance of surface reflectance anisotropy for cloud and NO_2 retrievals from GOME-2 and OMI, Atmos. Meas. Tech., 11, 4509-4529, https://doi.org/10.5194/amt-11-4509-2018, 2018.

25   Maasakkers, J. D.: Vital improvements to the retrieval of tropospheric NO_2 columns from the Ozone Monitoring instrument, M.Sc. thesis, Eindhoven University of Technology, The Netherlands, 67 pp., 2013.

Marchenko, S., Krotkov, N. A., Lamsal, L. N., Celarier, E. A., Swartz, W. H., and Bucsela, E. J.: Revising the slant column density retrieval of nitrogen dioxide observed by the Ozone Monitoring Instrument. *J. Geophys. Res. Atmos.*, 120, 5670–5692. doi:10.1002/2014JD022913, 2015.

5 Miyazaki, K., Eskes, H. J., and Sudo, K.: Global NO_x emission estimates derived from an assimilation of OMI tropospheric NO_2columns, Atmos. Chem. Phys., 12, 2263-2288, https://doi.org/10.5194/acp-12-2263-2012, 2012.

Müller, J.-P., Kharbouche, S., Gobron, N., Scanlon, T., Govaerts, Y., Danne, O., Schultz, J., Lattanzio,
10 A., Peters, E., De Smedt, I., Beirle, S., Lorente, A., Coheur, P. F., George, M., Wagner, T., Hilboll, A., Richter, A., Van Roozendael, M., and Boersma, K. F.: Recommendations (scientific) on best practices for retrievals for Land and Atmosphere ECVs (QA4ECV Deliverable 4.2 version 1.0), 186 pp., http://www.qa4ecv.eu/sites/default/files/D4.2.pdf, last access: 12 April 2018, 2016.

15 Munro, R., Eisinger, M., Anderson, C., Callies, J., Corpaccioli, E., Lang, R., Lefebvre, A., Livschitz, Y., and Albiñ̃ana, A. P.: GOME-2 on MetOp, Proc. of The 2006 EUMETSAT Meteoro- logical Satellite Conference, Helsinki, Finland, 12–16 June 2006, EUMETSAT P.48, 2006.

Munro, R., Lang, R., Klaes, D., Poli, G., Retscher, C., Lindstrot, R., Huckle, R., Lacan, A., Grzegorski,
20 M., Holdak, A., Kokhanovsky, A., Livschitz, J., and Eisinger, M.: The GOME-2 instrument on the Metop series of satellites: instrument design, calibration, and level 1 data processing − an overview, Atmos. Meas. Tech., 9, 1279-1301, https://doi.org/10.5194/amt-9-1279-2016, 2016.

Nightingale, J., De Rudder, A., Boersma, F., Scanlon, T., Farquhar, C., Muller, J.-P., and Fox, N.:
25 Results from the QA4ECV user requirements survey on quality assurance in satellite data products (QA4ECV Deliverable 1.1 version 2.0), 16 pp., http://www.qa4ecv.eu/sites/default/files/QA4ECV_D.1.1_survey_report_V2.0_20150922.pdf, last access: 10 April 2018, 2015.

Nightingale J., Boersma, K. F., Muller, J.-P., Compernolle, S., Lambert, J.-C., Blessing, S., Giering, R., Gobron, N., De Smedt, I., Coheur, P., George, M., Schulz, J., and Wood, A.: Quality Assurance Framework Development Based on Six New ECV Data Products to Enhance User Confidence for Climate Applications, *Remote Sensing,* 10(8), 1254; doi: 10.3390/rs10081254, 2018.

Oldeman, A.: Effect of including an intensity offset in the DOAS NO_2 retrieval of TROPOMI, Internship report, R-1944-SE, Eindhoven University of Technology/KNMI, Eindhoven, May 2018, https://kfolkertboersma.files.wordpress.com/2018/06/report_oldeman_22052018.pdf, last access: 30 October 2018, 2018.

Pinardi, G., Van Roozendael, M., Hendrick, F., et al.: Satellite nadir NO2 validation based on direct-sun and MAXDOAS network observations, 8th International DOAS Workshop, Yokohama, Japan, 4-6 September, 2017, O2-01, 2017.

Richter, A.: Absorptionsspektroskopische Messungen stratosphärischer Spurengase über Bremen, 53°N, PhD-Thesis, University of Bremen, June 1997 (in German), 1997.

Richter, A. and Wagner, T.: Diffuser plate spectral structures and their influence on GOME slant columns, Tech. Rep., Inst. Env. Phys., Univ. Bremen and Inst. Env. Phys., Univ. Heidelberg, 2001.

Richter, A., Begoin, M., Hilboll, A. and Burrows, J. P.: An improved NO_2 retrieval for the GOME-2 satellite instrument, Atmos. Meas. Tech., 4(6), 1147–1159, doi:10.5194/amt-4-1147-2011, 2011.

Roscoe, H. K., Van Roozendael, M., Fayt, C., du Piesanie, A., Abuhassan, N., Adams, C., Akrami, M., Cede, A., Chong, J., Clémer, K., Friess, U., Gil Ojeda, M., Goutail, F., Graves, R., Griesfeller, A., Grossmann, K., Hemerijckx, G., Hendrick, F., Herman, J., Hermans, C., Irie, H., Johnston, P. V., Kanaya, Y., Kreher, K., Leigh, R., Merlaud, A., Mount, G. H., Navarro, M., Oetjen, H., Pazmino, A.,

5   Perez-Camacho, M., Peters, E., Pinardi, G., Puentedura, O., Richter, A., Schönhardt, A., Shaiganfar, R., Spinei, E., Strong, K., Takashima, H., Vlemmix, T., Vrekoussis, M., Wagner, T., Wittrock, F., Yela, M., Yilmaz, S., Boersma, F., Hains, J., Kroon, M., Piters, A., and Kim, Y. J.: Intercomparison of slant column measurements of NO_2 and O_4 by MAX-DOAS and zenith-sky UV and visible spectrometers, Atmos. Meas. Tech., 3, 1629-1646, https://doi.org/10.5194/amt-3-1629-2010, 2010.

S5P/TROPOMI Science Verification Report, S5P-IUP-L2-ScVR-RP, Richter A. and the Verification Teams: European Space Agency, 314 pp., available at: https://earth.esa.int/web/sentinel/user-guides/sentinel-5p-tropomi/document-library/-/asset_publisher/w9Mnd6VPjXlc/content/sentinel-5p-tropomi-science-verification-report (last access: 12 April 2018), 2015.

Schenkeveld, V. M. E., Jaross, G., Marchenko, S., Haffner, D., Kleipool, Q. L., Rozemeijer, N. C., Veefkind, J. P., and Levelt, P. F.: In-flight performance of the Ozone Monitoring Instrument, Atmos. Meas. Tech., 10, 1957-1986, https://doi.org/10.5194/amt-10-1957-2017, 2017.

20   Shindell, D. T., Faluvegi, G., Koch, D. M., Schmidt, G. A., Unger, N., and Bauer, S. E.: Improved Attribution of Climate Forcing to Emissions, Science, 325, 716-718, doi:10.1126/science.1174760, 2009.

Slijkhuis, S., Aberle, B., Coldewey-Egbers, M., Loyola, D., Dehn, A., Fehr, T.: GOME/ERS-2: New
25   homogeneous Level 1B data from an old instrument, Proceeding at the ESA ATMOS conference, 8 – 12 June, University of Crete Heraklion, Greece, 2015.

Solomon, S., Portmann, R. W., Sanders, R. W., Daniel, J. S., Madsen, W., Bartram, B., and Dutton, E. G.: On the role of nitrogen dioxide in the absorption of solar radiation, J. Geophys. Res., 104(D10), 12047–12058, doi:10.1029/1999JD900035, 1999.

5   Tilstra, L. G., Tuinder, O. N. E., Wang, P., and Stammes, P.: Surface reflectivity climatologies from UV to NIR determined from Earth observations by GOME-2 and SCIAMACHY, J. Geophys. Res.: Atmospheres, 122, 4084–4111, https://doi.org/10.1002/2016JD025940, 2016JD025940, 2017.

van Geffen, J. H. G. M., Boersma, K. F., Van Roozendael, M., Hendrick, F., Mahieu, E., De Smedt, I.,
10   Sneep, M., and Veefkind, J. P.: Improved spectral fitting of nitrogen dioxide from OMI in the 405–465 nm window, Atmos. Meas. Tech., 8, 1685-1699, https://doi.org/10.5194/amt-8-1685-2015, 2015.

Veefkind, J. P., de Haan, J. F., Sneep, M., and Levelt, P. F.: Improvements to the OMI O_2-O_2 operational cloud algorithm and comparisons with ground-based radar–lidar observations, Atmos.
15   Meas. Tech., 9, 6035–6049, https://doi.org/10.5194/amt-9-6035-2016, 2016.

Verstraeten, W. W., Neu, J. L., Williams, J. E., Bowman, K. W., Worden, J. R., and Boersma, K. F.: Rapid increases in tropospheric ozone production and export from China, Nature Geoscience, 8(9), 690-695, http://dx.doi.org/10.1038/ngeo2493, 2015.

Vinken, G. C. M., Boersma, K. F., Maasakkers, J. D., Adon, M., and Martin, R. V.: Worldwide biogenic soil NO_x emissions inferred from OMI NO_2 observations, Atmos. Chem. Phys., 14, 10363-10381, doi:10.5194/acp-14-10363-2014, 2014.

25   Wang, P., Stammes, P., van der A, R., Pinardi, G., and van Roozendael, M.: FRESCO+: an improved O_2 A-band cloud retrieval algorithm for tropospheric trace gas retrievals, Atmos. Chem. Phys., 8, 6565-6576, https://doi.org/10.5194/acp-8-6565-2008, 2008.

Williams, J. E., Boersma, K. F., Le Sager, P., and Verstraeten, W. W.: The high-resolution version of TM5-MP for optimized satellite retrievals: description and validation, Geosci. Model Dev., 10, 721-750, https://doi.org/10.5194/gmd-10-721-2017, 2017.

5   Witman, S., Holloway, T., and Reddy, P.: Integrating Satellite Data into Air Quality Management: Experience from Colorado, Environmental Manager (EM) Magazine, February 2014 Issue, 2014.

World Health Organization: Review of evidence on health aspects of air pollution REVIHAAP Project, Tech.   rep.,   Copenhagen,   Denmark,   302   pp.,   available   at:
10   http://www.euro.who.int/__data/assets/pdf_file/0004/193108/REVIHAAP-Final-technical-report-final-version.pdf?ua=1 (last access: 1 April 2018), 2013.

World Meteorological Organization: Guideline for the Generation of Datasets and Products Meeting GCOS   Requirements,   Geneva,   Switzerland,   12   pp.,   available   at
15   https://library.wmo.int/opac/doc_num.php?explnum_id=3854 (last access: 1 April 2018), 2010.

World Meteorological Organization: Systematic Observation Requirements for Satellite-Based Data Products   for   Climate   –   2011   Update,   Geneva,   Switzerland,   128   pp.,   available   at: https://library.wmo.int/opac/doc_num.php?explnum_id=3710 (last access: 1 April 2018), 2011.

Xu, X., Wang, J., Henze, D. K., Qu, W., and Kopacz, M.: Constraints on aerosol sources using GEOS-Chem adjoint and MODIS radiances, and evaluation with multisensor (OMI, MISR) data, J. Geophys. Res. Atmos., 118, 6396–6413, doi:10.1002/jgrd.50515, 2013.

25   Zara, M., Boersma, K. F., De Smedt, I., 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.: Improved slant column density retrieval of nitrogen dioxide and formaldehyde for OMI and GOME-2A from

QA4ECV: intercomparison, uncertainty characterisation, and trends, Atmos. Meas. Tech., 11, 4033-4058, https://doi.org/10.5194/amt-11-4033-2018, 2018.

Zhang, L., Jacob, D. J., Boersma, K. F., Jaffe, D. A., Olson, J. R., Bowman, K. W., Worden, J. R.,
5   Thompson, A. M., Avery, M. A., Cohen, R. C., Dibb, J. E., Flock, F. M., Fuelberg, H. E., Huey, L. G., McMillan, W. W., Singh, H. B., and Weinheimer, A. J.: Transpacific transport of ozone pollution and the effect of recent Asian emission increases on air quality in North America: an integrated analysis using satellite, aircraft, ozonesonde, and surface observations, Atmos. Chem. Phys., 8, 6117-6136, https://doi.org/10.5194/acp-8-6117-2008, 2008.