

Reply to Reviewer Comments on “Improving algorithms and uncertainty estimates for satellite NO2 retrievals: Results from the Quality Assurance for Essential Climate Variables (QA4ECV) project”

Folkert Boersma, 12 November 2018

5 We thank the reviewer for the positive and useful comments, and for careful reading of the paper. We have addressed the questions as follows, with our response in blue. Below please also find the revised manuscript with track changes.

**Reviewer Comment 1**

10 The manuscript summarizes the long-term effort in building a consistent, multi- decade record of NO2 columns based on the practically all available data sets acquired from space. The article offers a plenitude of technical details aiming to improve the traceability of the described products as well as provides valuable recommendations that should help improving quality of the NO2 retrievals. The article deserves a prompt publication.

15 [Thank you for this comment.](#)

I would like the authors to consider the following corrections/amendments:

=====

The main text:

20 I think that Section 2, especially Sect. 2.1, should be substantially shortened, since, by the author’s remark, the survey’s outcome was published elsewhere. I am not sure Fig. 1 adds any valuable content to the main objectives of the article. I would consider its removal. Section 2.1 could be shortened to 1-2 paragraphs by mentioning only the user’s suggestions that have been implemented in the current version of the products). I would leave out all ‘things-to-be-done’ or ‘things-to-be-considered’.

25

The same applies to Sect. 2.2 - I would concentrate exclusively on the already implemented items.

If the authors perceive the full review of the user's comments as very important, then I suggest moving it to Appendix.

5 This is a good point. We feel it is important to document in an easily accessible manner the outcomes of the surveys, and this is now done in the Supplementary material. We shortened the text and removed original Figure 1 accordingly.

Also, consider putting the text from p.8, l.25 through p.10, l.21 into a new Sect. 2.3, thus ascribing the current Sect. 2.3 to 2.4. This particular text, [p.8,l.25 - p.10,l.21], does not belong in Sect. 2.2.

10 Agreed. We now introduce a new section 2.3 titled 'QA4ECV consortium activities'.

The Footnote #2 to Table 1 does not help understand the meaning of the quoted 2%. Please either rephrase or remove.

15 We have rephrased this as follows: "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)."

20 Table 2: is GOME v5 L1 used in the article different from the most recent L1 version described by Coldewey-Egbers et al. (AMTD, 2018)? Adding a footnote may help to remove the ambiguity. Also, Shah et al. (2018, AMT, 11, 2345) describes SCIAMACHY V8 while the authors use v7. This should be commented on just as well.

25 Thanks for the opportunity to clarify this. In our GOME data product, we worked with what was available to us at the time of project finalization, which was v5 level 1 data. This is 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], which is not a

strong concern.

For SCIA, we indeed use V8 level 1 data, so our original manuscript wrongfully stated that it was v7.04-w. This has now been corrected.

5

p.15, l.5. OMI shows different degradation rates in radiances and irradiances. Moreover, various instrument-performance metrics show far superior stability than the quoted 2%. The wording should be changed to something like: "The OMI instrument produces stable (to  $\pm 2\%$  over the mission time, in the row anomaly-free areas) L1B radiances".

10

Done.

p.15, l.11. '... directly from the Sun...'. Please re-phrase, since 'directly' does not apply to the sunlight presumably scattered by a peeling piece of insulation.

15

Done.

Fig. 2. The percentages shown at the upper axis do not correspond to the RA-marked areas in the plot. Either correct or clarify.

20

The percentages in the upper x-axis do not correspond to  $(\#badrow/\#allrows) \times 100\%$  but to the "percentage of affected pixels" meaning:  $(\#badpixels/\#totalpixels) \times 100\%$  from all rows. For some rows, the row anomaly does not affect the entire orbit. This is now clarified in the caption of the figure concerned.

25

p.16, l.1. Replace 'detector degradation' by 'optical throughput changes in the irradiance channel'. The CCD detector does not change at the quoted rates.

Agreed. We updated the sentence.

p.18, l.9. In addition to the mentioned time-dependent slit-function changes, the GOME-2A slit function varies with the scan angle and along the orbit, though at a much lower level compared to the dominant long-term changes. Are these cross-track and along-orbit changes accounted for? Please comment on.

5 In the GOME-2 NO2 fits, no slit function fit was performed so in-orbit changes are not accounted for. In the visible, the change of the GOME-2(A) slit function is not important (unlike in the UV-range). If any cross-track changes in the slant columns would still persist, these are likely dampened by the stripe-correction, which was applied on GOME-2A data in the same fashion as for OMI.

10 Figure 3. In the present format some details are inevitably lost due to the limited resolution of a printout. The same applies to the screen viewing, no matter the zoom. I may suggest: 1. Substantially enlarging the upper section. 2. Substantially overlapping two lower sections (these are practically the same) and expanding both of them at the same time.

15 Thanks for this good suggestion. We followed up on it.

p.23, l.4. Do you include or reject the RA-affected retrievals in your stats? Please clarify.

20 We have applied XTrack flagging using the mask provided in the lv1 files, which means that the RA-affected retrievals have not been included in the stats.

Fig. 4. Define the units of the color bars under both plots.

Done.

25

p.23, l.13. The intensity offset proves to be very important, so this particular subject deserves more detailed discussion, either in the main text or in a separate Appendix. In particular, the authors should provide a mathematical description and discuss the particulars of the 'best-practice' implementation of

the offset. Indeed, the cited Müller et al. [2016] provides valuable information, but it is not possible to conclude which form of the intensity offset is considered as the best-practice by the authors. I suggest, besides providing mathematical description in the text, putting some specifics of the applied approach in Table 3. Currently the Table carries 'yes' or 'no' entries in the respective column. Does this mean that all 'yes' entries ascribe to the same intensity offset mathematical form and implementation?

We now include a mathematical description of the intensity offset, which has been applied for all sensors. The selected form of the intensity offset was driven by the optical density nature of the fit.

Table 4. Describe the 'undersampling' and 'Eta' corrections. These weren't mentioned in the text.

Good point. We now elaborate on these corrections in the margin of Table 4.

p.31, Footnote 6 : replace '...behavior of the red line...' by '...behavior of the black line...'

15

Done.

p.32, 1.1. The 1st sentence mentions AMF\_strat in Fig.5, while Fig.5 caption says AMF\_total. Which one is true?

20

Thank you for pointing this out. We corrected the first sentence on page 32.

Table 5. If all listed values are derived on a pixel basis, then I would put a note in the caption and remove 'per pixel' from the fields. If not, then each field must carry either 'global' or 'per pixel' notation. Or, better yet, the table could be segregated into the 'pixel' and 'global' parts.

25

Thank you for this valuable suggestion. We followed up on this suggestion and introduced a category indicating whether the estimate is pixel-specific, 'global', or a mix of these categories.

Fig. 9. All the colors but blue and light-blue are described in the caption. Either add these two descriptions or remove all and mention that additional info is provided in the figure legends.

5 Done.

Fig.13. The caption quotes  $P < 850$  hPa while the text (p. 50) says  $P < 875$  hPa. Which one applies?

875 hPa applies. Corrected.

10

p.52, 1.7. The common Gauss-process error-combination formula gives  $\sigma = \sqrt{\sum(\sigma_i^2)}$ . The text mentions  $\sigma = 1 / \sqrt{\sum(\sigma_i^2)}$ . Please explain the difference.

This was corrected and now reads  $\sigma = \sqrt{\sum(\sigma_i^2)}$ .

15

p.50 quotes -2% bias and 16% RMS, while Summary provides -2% and 13%, respectively. And, the same applies to DOMINO results: [+11% and 35%] in p.51 vs. [+11% and 28%] in Summary. Please clarify.

20 Corrected. The numbers in the text apply, and those in the summary has been corrected.

The list of references: a quick spot-check shows the cited in the text, but missing in the list - Boersma et al. [2004], Irie et al. [2009], Krotkov et al. [2016]. Thus, I suspect such omissions to be more numerable. Please double-check the list of references.

25

Done.

Also, please add letters b through f to the Boersma et al. [2017] works, to match the citations in the

main text.

Done.

5 =====

Additional corrections:

p.23, l.15 - "... results in larger NO2..."

Done.

10

p.24, l.10 - "The inter-comparison of preferred-setting SCDs..."

Done.

15 p.29, l.4 - remove the 2nd 'agree reasonably well'

Done.

## Reviewer Comment 2

The paper “Improving algorithms and uncertainty estimates for satellite NO<sub>2</sub> retrievals: Results from the Quality Assurance for Essential Climate Variables (QA4ECV) project” by Boersma et al. describes the improvement of NO<sub>2</sub> retrieval algorithms for GOME, SCIAMACHY, OMI and GOME-2, and the production of a multi-decadal record of global NO<sub>2</sub> columns. This manuscript provides an overview of the NO<sub>2</sub> retrieval results from the QA4ECV project, which have been partially reported in earlier papers. Furthermore, first comparisons have been made between the OMI tropospheric NO<sub>2</sub> data and ground-based MAX-DOAS measurements at one site in China. The topic of the manuscript is within the scope of AMT and it is of interest to the scientific community. It can be recommended for publication, if the authors make an effort to address the comments listed below, and improve the manuscript accordingly.

[Thank you for this comment.](#)

15 Section 3

Table 2 The spatial resolution of GOME-2A was changed to 40x40 km<sup>2</sup> on 15 July 2013.

[We have now included this in Table 2.](#)

20 P18 The diffuser plate also causes spectral structures in the GOME-2 solar irradiances, but the effect is much smaller than for GOME/ERS-2.

[We have now update the text in Section 3.](#)

25 Section 5.1

The manuscript discusses a consistent NO<sub>2</sub> retrieval for GOME, SCIAMACHY, OMI and GOME-2, and Table 4 lists the specific DOAS settings for the 4 satellite instruments. However, in the text only the analyses of the fitting results for OMI are discussed in detail. Considering the specific instrument

characteristics affecting the DOAS retrieval (as described in Sect. 3), a short discussion on the NO<sub>2</sub> fitting results for the other instruments should be included, and the consistency between the results of the 4 instruments discussed.

- 5 Good point. Although the focus within QA4ECV was mostly on OMI, we also evaluated the consistency in the NO<sub>2</sub> SCDs from the other instruments. We have now added a discussion of our tests on the consistency of the NO<sub>2</sub> fitting results from GOME-2, SCIAMACHY, and GOME.

P23-24 The differences in the NO<sub>2</sub> slant column shown in Fig. 4 are significant. As described in the  
10 text, the differences are mainly a result of the intensity vs optical density fit and the intensity offset (while the stratospheric AMF effect is very small). These two key DOAS subjects should be discussed in more detail in this section. In particular, I suggest adding a discussion/recommendation about the intensity vs optical density fit method.

- 15 We discuss in more detail now the use of the optical density fits and intensity offset used in the QA4ECV fitting process. For a detailed discussion of the tests done, we refer to QA4ECV D4.2, and a work in preparation on these test by the University of Bremen.

Fig 5b The illustrated scenario ( $SZA=VZA=0$ ) is not really a typical OMI viewing geometry. Please use  
20 a realistic mid-latitude OMI measurement scenario.

We have done this, and now show the scenario for  $SZA = VZA = 30^\circ$ .

Section 5.2

- 25 P27 Here it is mentioned that NO<sub>2</sub> SCDs are assimilated in TM4, while at the end of Sect. 5.3 TM5 is mentioned.

In the testing phase, we used the DOMINO v2 code, where assimilation is done in TM4 [Boersma et al.,

2011; Dirksen et al., 2011]. In the retrieval development phase, we updated the model framework to TM5. The text now reflects this.

P28 Please provide examples of “particular applications” of the two different STS approaches.

5

The purpose of either STS approach is to correct for the stratospheric contribution to the observed SCD, in order to retrieve a high-quality tropospheric product. Both corrections perform similarly over regions with strong anthropogenic NO<sub>2</sub> pollution.

- 10 Data assimilation can be applied for global retrievals. The strength of data assimilation is that it provides a correction that is based on broad and consistent knowledge on the state of the atmosphere (NO<sub>2</sub> and temperature profile, stratospheric dynamics). Especially in situations with strong stratospheric NO<sub>2</sub> 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
- 15 [Dirksen et al., 2011], whereas the STREAM method by design results in zonally smooth structures in those regions.

- STREAM can be applied especially for studies into small sources, such as emissions from soil, ships, and small, isolated anthropogenic sources. The strength of STREAM is that it does not rely on
- 20 chemistry transport models. Data assimilation is potentially somewhat vulnerable to misinterpret tropospheric contributions as stratospheric NO<sub>2</sub>. This qualifies STREAM as a useful alternative for data assimilation for studies in areas away from strong stratospheric gradients (where the zonally smooth structure of the stratospheric field is of little consequence). The text has been updated accordingly.

- 25 Section 6.2

Tabel 5 Are the “Symbols” for the three “Contribution to the uncertainty of . . .” entries correct?

They are indeed correct, but could be clarified a bit more. The first term, between brackets, indicates the

AMF uncertainty caused by uncertainty in the forward model parameters. These terms are the same as Eq. (12) in Boersma et al. [2004]. Since we're interested here in the contribution of these uncertainties to the NO<sub>2</sub> column, we ratio the AMF uncertainty by the AMF itself, and multiply with the actual tropospheric NO<sub>2</sub> column. We now include a clarification in a footnote to Table 5.

5

#### Section 7

As also mentioned in the text, the ground-based validation discussed in the paper is for one site only and has a very limited time range. Although it is understandable that in this manuscript only a first validation is presented and a dedicated validation paper is in preparation, it is problematic to draw  
10 conclusions about OMI retrieval uncertainties based on regression/differences analysis of such a small data sample. If the validation and uncertainty analysis are illustrated for one site then a longer data period should be used to increase the number of measurements and to be able to account for seasonal variations. Why has the site Tai'an been selected if only campaign data for a short period is available? (also considering the fact that the authors have access to longer MAX DOAS data records for several  
15 other sites).

A more comprehensive validation work, with focus on other sites and over longer periods, is in preparation by our Belgian colleagues (as was mentioned in the manuscript). We chose to do a first evaluation of the OMI QA4ECV NO<sub>2</sub> product and their uncertainties with the Tai'an data available to  
20 us, because these data have been used before to validate the DOMINO v2 product with good agreement [Irie et al., 2012]. This anchors our evaluation of the new QA4ECV product to previous, well-established efforts. Moreover, the scope in section 7 is not just on validation but as much on bringing forward ideas on evaluating satellite data, such as correcting for spatial gradients, and using the deviations to check on the reported uncertainties.

25

# Improving algorithms and uncertainty estimates for satellite NO<sub>2</sub> retrievals: Results from the Quality Assurance for Essential Climate Variables (QA4ECV) project

5

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

10

<sup>1</sup>Royal Netherlands Meteorological Institute, Satellite Observations department, The Netherlands

<sup>2</sup>Wageningen University, Meteorology and Air Quality Group, Wageningen, the Netherlands

<sup>3</sup>Institute of Environmental Physics (IUP-UB), University of Bremen, Bremen, Germany

<sup>4</sup>Belgian Institute for Space Aeronomy (BIRA-IASB), Brussels, Belgium

15 <sup>5</sup>Max-Planck Institute for Chemistry (MPI-C), Mainz, Germany

<sup>6</sup>Harvard University, Cambridge, Massachusetts, United States

<sup>7</sup>National Physics Laboratory (NPL), Teddington, United Kingdom

<sup>8</sup>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 (NO<sub>2</sub>) 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 NO<sub>2</sub> columns and their associated uncertainties can provide crucial information how the emissions and concentrations of nitrogen oxides evolve over time. Here we describe the development of a new, community best practice NO<sub>2</sub> retrieval algorithm based on a synthesis of existing approaches. Detailed comparisons of these approaches led us to implement an enhanced spectral fitting method for NO<sub>2</sub>, a 1°×1° TM5-MP data

25

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<sub>2</sub> algorithm on 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<sub>2</sub> 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<sub>2</sub> columns amount to typically 40% over polluted scenes. First validation results of the QA4ECV OMI NO<sub>2</sub> 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<sub>2</sub> long-term data record is providing valuable information to quantitatively constrain emissions, deposition, and trends in nitrogen oxides on a global scale.

## 15 **1 Introduction**

Nitrogen oxides (NO<sub>x</sub> = NO + NO<sub>2</sub>) 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<sub>x</sub> leads to the catalytic destruction of ozone, and the formation of reservoir species for halogens (e.g. Crutzen et al. [1970]). Nitrogen oxides 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<sub>2</sub> 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<sub>x</sub>-driven 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<sub>2</sub> (together with SO<sub>2</sub>, 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<sub>2</sub> 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<sub>2</sub>. 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<sub>2</sub> 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<sub>2</sub>.

High-quality observations are needed to monitor the concentrations of nitrogen oxides in the atmosphere, both close to the ground, where NO<sub>2</sub> 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<sub>2</sub> concentrations and NO<sub>x</sub> 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<sub>x</sub> 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<sub>2</sub> 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](http://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<sub>2</sub>, the goals of this project are: (1) to generate a multi-decadal (1995-2017) satellite data record of tropospheric and stratospheric NO<sub>2</sub> 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<sub>2</sub> 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<sub>2</sub> 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<sub>2</sub> products.

In this work, we demonstrate our approach to improve a retrieval algorithm and apply it to generate a multi-decadal record of NO<sub>2</sub> 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<sub>2</sub> 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<sub>2</sub> retrieval. The improved quality of the QA4ECV NO<sub>2</sub> 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<sub>2</sub> data user requirements, the expertise from NO<sub>2</sub> 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<sub>2</sub> 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<sub>2</sub> 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<sub>2</sub> data users, and interviews with 3 NO<sub>2</sub> ‘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<sub>2</sub> 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].

Folkert Boersma 2-11-2018 18:07

**Deleted:** Here we briefly summarize

Folkert Boersma 2-11-2018 18:07

**Deleted:** :

Folkert Boersma 2-11-2018 18:07

**Deleted:** .

## 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<sub>2</sub> data producers to improve the retrieval algorithms and generate a long-term NO<sub>2</sub> data set from available satellite reflectance measurements. The NO<sub>2</sub> 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<sub>2</sub> 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<sub>2</sub> columns is based on a 3-step approach. First, a set of absorption cross-sections, including NO<sub>2</sub>, 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:

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

Folkert Boersma 2-11-2018 18:09

Deleted: -

... [1]

Folkert Boersma 3-11-2018 18:23

Deleted: -

... [2]

Folkert Boersma 2-11-2018 18:19

**Deleted:** *They also admit that traceable input (ancillary data) files, read-in software, sensitivity analysis documentation, and publications are not always provided along with the data. Data producers stated that direct communication with their data users is important, mostly on issues including read-in software, product format, flagging and filtering procedures, and the uncertainty budget. In general, data producers thought that the necessary traceability and quality information is in principle available, but cannot always be easily found. These recommendations helped shape QA4ECV QA System further. Within this system, data producers have the possibility to provide all these pieces of information, and users can obtain a quick overview of the maturity and completeness of the data product. Data producers were generally positive about benchmarking their satellite data product against other scientific standards, such as from cross-calibrated global validation networks. They noticed that quality information for independent reference data is often not available. For more detailed outcomes of the ECV data producer survey, please see QA4ECV Deliverable 1.1 [Nightingale et al., 2015].*

Folkert Boersma 28-10-2018 21:07

Formatted: Font:Italic

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

- 5 (1) An inter-comparison of spectral fitting approaches between institutes. NO<sub>2</sub> 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.
- 10 (2) An evaluation of the algorithm SCD uncertainties against an independent statistical uncertainty estimates [Zara et al., 2018].
- 15 (3) A comparison of stratospheric NO<sub>2</sub> 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.
- 20 (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<sup>1</sup> when using different radiative transfer models, vertical layering and interpolation schemes [Lorente et al., 2017].
- 25 (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<sub>2</sub> profile, etc.), to a wider ‘round robin’ comparison

---

<sup>1</sup>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<sub>2</sub> column estimates from the independent STREAM method has been included in the QA4ECV NO<sub>2</sub> 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<sub>2</sub> 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<sub>2</sub> columns [WMO, 2011].

	<b>Horizontal Resolution</b>	<b>Vertical Resolution</b>	<b>Temporal Resolution</b>	<b>Uncertainty</b>	<b>Stability/decade<sup>2</sup></b>
--	------------------------------	----------------------------	----------------------------	--------------------	-------------------------------------

*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 <sub>2</sub> tropospheric column	5-10km	N/A	4h	max(20%; 0.03 DU <sup>1</sup> )	2%
---	--------	-----	----	------------------------------------	----

<sup>1</sup>An uncertainty of 0.03 DU (Dobson Units) corresponds to  $0.8 \times 10^{15}$  molec.cm<sup>-2</sup>. The 0.03 DU holds for tropospheric NO<sub>2</sub> columns up to  $4.0 \times 10^{15}$  molec.cm<sup>-2</sup>, 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] requirements are indicative of acceptable overall levels for the uncertainties of product values.”

<sup>2</sup>[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 ‘target requirements’, which should be advanced towards when generating a long-term record of tropospheric NO<sub>2</sub> column measurements. The resolution requirements listed above cannot be met by the satellite sensors capable of measuring NO<sub>2</sub> 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 and temporal resolution” than 5-10 km (that is the NO<sub>2</sub> 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 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 documentation process [Nightingale et al., 2018]. Section 3 of the Supplement provides a point-by-point

Folkert Boersma 28-10-2018 21:19

**Deleted:** is in general understood to represent the extent to which the error of a product remains constant over a long period, typically a decade or more [WMO, 2011].

Folkert Boersma 3-11-2018 18:33

**Deleted:** 1

overview of how these guidelines have been taken into account for the generation of the QA4ECV NO<sub>2</sub> 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 [Compernelle, 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<sub>2</sub> 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<sub>2</sub> ECV data product.

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

Folkert Boersma 28-10-2018 21:57  
Deleted: 7.04-w

Folkert Boersma 30-10-2018 16:41  
Deleted: 8

<sup>2</sup> 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 <sup>2</sup> <u>after</u> <u>15 July</u> <u>2013</u>			
--	--	---	--	--	--

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.

5

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<sub>2</sub> 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 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<sub>2</sub> SCDs are retrieved. To provide the total NO<sub>2</sub> SCDs necessary for an ECV dataset of stratospheric and tropospheric NO<sub>2</sub> columns, a background correction, typically estimated from an external source, is required.

Detector degradation is another relevant issue for NO<sub>2</sub> 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 2003 [Slijkhuis et al., 2015], and is anticipated to result in modest increases in the GOME NO<sub>2</sub> 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<sub>2</sub> 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<sub>2</sub> 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<sub>2</sub> to 60 × 30 km<sup>2</sup> (30 x 30 km<sup>2</sup> in some latitude bands). A consequence of this is that only spectral data from the 424-527 nm cluster are available for DOAS NO<sub>2</sub> spectral fitting. Similar 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<sub>2</sub> 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<sup>3</sup>, 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<sub>2</sub> SCDs.

5

## 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

10 are from the Collection 3 data. Processing of this Collection 3 data started in February 2010 with the 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<sub>2</sub>, 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<sub>v</sub> 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<sub>2</sub> data product, addressing the user needs expressed in section 2.1.

Spurious across-track variability, or stripes, are apparent in current OMI NO<sub>2</sub> data products. The stripes appear as discrete jumps in NO<sub>2</sub> 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

---

<sup>3</sup>Etaloning refers to unintended multiple reflection of radiation between optical elements leading to constructive and destructive interference for some wavelengths.

Folkert Boersma 28-10-2018 21:59

**Deleted:** The OMI instrument performance and level-1 data product have proven stable (to within 2%) since the mission started in October 2004.

Folkert Boersma 28-10-2018 22:03

**Deleted:** directly from the Sun

Folkert Boersma 3-11-2018 18:36

**Deleted:** 2

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<sub>2</sub> destriping corrections has increased from  $0.3 \times 10^{15}$  to  $0.5 \times 10^{15}$  molec. cm<sup>-2</sup> 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<sub>2</sub> 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<sub>2</sub> 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].

#### 15 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 2<sup>nd</sup> 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 succeeded 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

Folkert Boersma 29-10-2018 09:39

**Deleted:** In OMI's visible

Folkert Boersma 29-10-2018 09:38

**Deleted:** detector degradation

error on the NO<sub>2</sub> 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.



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<sub>2</sub> 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<sub>2</sub> SCDs. We used these guiding principles:

- Select as much as possible just one or in any case overlapping fitting windows for different instruments. NO<sub>2</sub> SCDs are known to be sensitive to the selection of fitting window, as shown in van Geffen et al. [2015], and in the SSP-verification report [SSP/TROPOMI Verification Report, 2015].
- Select a wide fitting window including more NO<sub>2</sub> 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<sub>2</sub> 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<sub>2</sub> absorption signature in the

Earthshine reference spectrum is still required.

#### 4 Algorithm design and traceability

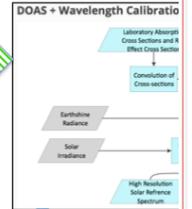
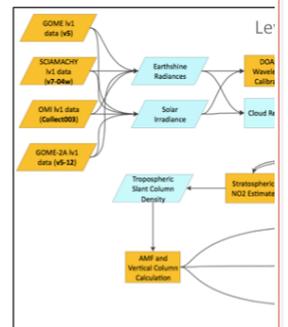
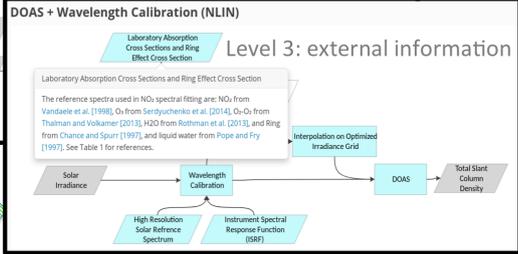
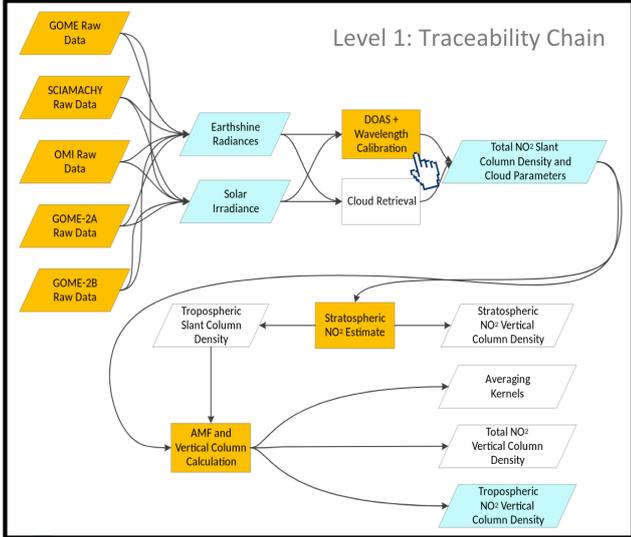
An important ambition of the QA4ECV project is to provide full traceability on retrieval algorithms. Usually, a condensed flow diagram for the retrieval algorithm is included in an Algorithm Theoretical  
5 Baseline Document (ATBD). The drawback is that ATBDs are often not easily accessible, and that 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<sub>2</sub> 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<sub>2</sub> 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 (1<sup>st</sup> layer). Next, as an example, when clicking the ‘DOAS + wavelength calibration’ step will lead to details on that sub-process (2<sup>nd</sup> 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 (3<sup>rd</sup> layer). The references themselves are linked to the digital  
20 object identifiers (DOI’s) and take users directly to the relevant paper.

Folkert Boersma 3-11-2018 18:36

Deleted: 3

Folkert Boersma 3-11-2018 18:36

Deleted: 3



Deleted:

**Figure 2**, Traceability Chain for the QA4ECV NO<sub>2</sub> 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.

5 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<sub>2</sub> 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<sub>2</sub> 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<sub>2</sub>, and helped to characterize the uncertainties of each retrieval sub-step.

### 5.1 Evaluation of spectral fitting approaches

15 NO<sub>2</sub> 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.

Folkert Boersma 3-11-2018 18:36

Deleted: Figure 3

Folkert Boersma 8-11-2018 20:31

Deleted: :

Folkert Boersma 8-11-2018 20:32

Deleted: comparison of NO<sub>2</sub> SCDs retrieved by different groups for the same level-1 data with common

Folkert Boersma 8-11-2018 20:32

Deleted: as (1), but now with each group using their own, preferred retrieval settings

**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<sub>3</sub>, NO<sub>2</sub>, O<sub>2</sub>-O<sub>2</sub>, H<sub>2</sub>O, 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.

5

Institute	Retrieval	Code	Method	Wavelength calibration	Intensity offset	Spike removal	Reference
BIRA-IASB	QDOAS	C	Optical depth <sup>4</sup> , non-linear least squares regression (Levenberg-Marquard)	Via Fraunhofer atlas, and shift and squeeze	Yes	Yes	Fayt and Van Roozendaal [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]
KNMI	OMNO2A	C	Intensity fit <sup>5</sup>	Via	No	Yes	van Geffen

<sup>4</sup>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.

<sup>5</sup>An ‘intensity’ fitting model is of the form:  $I(\lambda) = I_0(\lambda)e^{-\sum_i \sigma_i(\lambda)N_{s,i} + \sum_j a_j \lambda^j}$

	v2		non-linear least squares regression (Levenberg-Marquard)	Fraunhofer atlas, and shift			et al. [2015]
MPI-C	MPI-C	MATLAB	Optical depth, non-linear least squares regression, <u>shift and squeeze accounted for by pseudo-absorbers,</u>	Via Fraunhofer atlas, and shift and squeeze <u>(non-linear)</u>	Yes	Yes	Beirle et al. [2013]

Folkert Boersma 8-11-2018 20:32

Deleted: (Levenberg-Marquard)

The common settings inter-comparison of OMI NO<sub>2</sub> 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

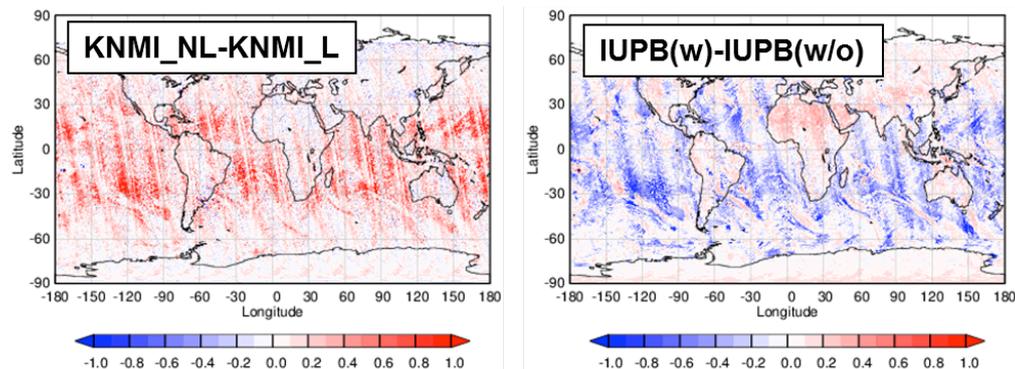
5 OMI orbits within the 4 selected days. The correlation is slightly less (but still > 99%) between the KNMI code and the other 3 codes, suggesting that algorithms agree in capturing the full dynamical range of NO<sub>2</sub> SCDs. The remaining differences appear over background regions and can be attributed to using a non-linear intensity fitting model instead of a linear optical density fit (resulting in NO<sub>2</sub> SCD

10 differences over the oceans up to  $1 \times 10^{15}$  molec.cm<sup>-2</sup>, see Figure 3), and to including or excluding an intensity offset term in the set of fit parameters (differences up to  $1 \times 10^{15}$  molec.cm<sup>-2</sup>, reducing contrast 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

Folkert Boersma 3-11-2018 18:36

Deleted: 4

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<sub>3</sub> misfits over water and over land. Excluding the intensity offset term results in larger NO<sub>2</sub> SCD uncertainties, and in (spurious) spatial patterns in the O<sub>3</sub> 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].

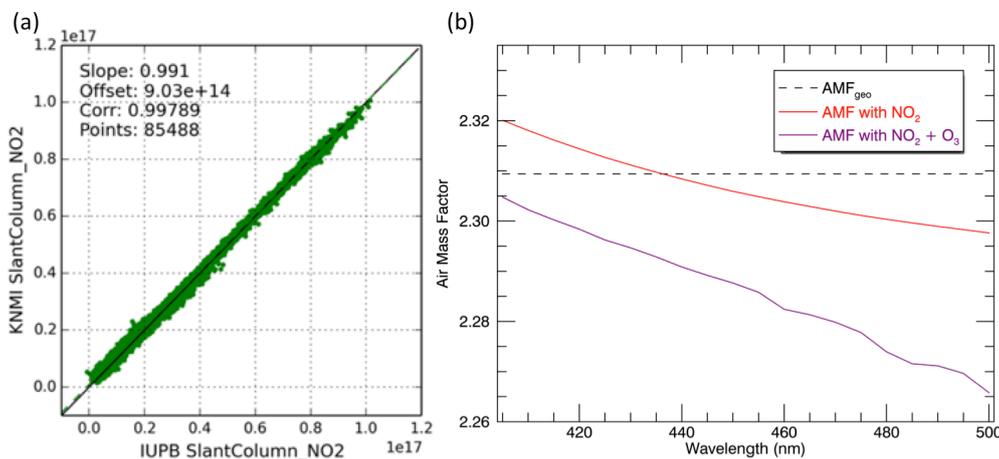


**Figure 3.** OMI NO<sub>2</sub> slant column differences between KNMI intensity (KNMI\_NL) and optical density fit (KNMI\_L) and IUPB fit including (IUPB(w)) and excluding the intensity offset (IUPB(w/o))(right). Data from 2 February 2005. The units of the colour bar in both panels are 10<sup>15</sup> molec. cm<sup>-2</sup>.

In round 2, each institute applied preferred settings to retrieve OMI NO<sub>2</sub> 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

Folkert Boersma 3-11-2018 18:36  
Deleted: 4

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  $<1 \times 10^{15}$  molecules  $\text{cm}^{-2}$ . The largest offset ( $+0.9 \times 10^{15}$  molecules  $\text{cm}^{-2}$ ) 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  $\text{O}_3$  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  $\text{NO}_2$  and  $\text{O}_3$  fit results.  $\text{NO}_2$  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)  $\text{NO}_2$  slant columns retrieved using preferred fit settings for OMI orbit OMIL2\_2005m0202t0339,

Folkert Boersma 3-11-2018 18:37

Deleted: 5

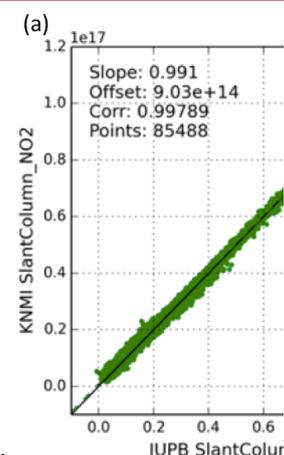
Folkert Boersma 3-11-2018 18:37

Deleted: 4

Folkert Boersma 3-11-2018 18:37

Deleted: 5

Folkert Boersma 5-11-2018 17:25



Deleted:

Folkert Boersma 3-11-2018 18:37

Deleted: 5

including only pixels with  $SZA < 88^\circ$  and intensity  $< 1.1 \times 10^{14}$ . (b) Wavelength dependency of the total air mass factor for a scenario with  $SZA=VZA=30^\circ$  (geometrical AMF = 2.31), as calculated with DAK for a mid-latitude standard atmospheric profile with a total  $NO_2$  column of  $5.9 \times 10^{15}$  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 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  $1 \times 10^{15}$  molecules  $cm^{-2}$ ), it is recommended to use one and the same fit model for all sensors.
- 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

Folkert Boersma 30-10-2018 17:04

Deleted: following

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<sub>2</sub>. 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  $1.48 \times 10^{15}$  molec.cm<sup>-2</sup>. 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<sup>-2</sup>) between GOME and SCIAMACHY for the period of overlap (Figure 20 in Muller et al. [2018]).

**Table 4.** Recommended settings for QA4ECV NO<sub>2</sub> spectral fitting for the retrieval of NO<sub>2</sub> 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 Globalcalib6	NLIN 7.25 (2007-2011) QDOAS (2012-2016)	NLIN 7.35-7.37	NLIN 7.55
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 <sup>6</sup> plus offset of $1.476 \times 10^{15}$ molec.cm <sup>-2</sup> )

<sup>6</sup> 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)

Folkert Boersma 31-10-2018 10:03

Deleted:

Folkert Boersma 8-11-2018 20:36

Deleted: The Pacific Ocean reference sector has been defined as

Polynomial	4th order	4th order	4th order	4th order
Fitting parameters	O <sub>3</sub> , NO <sub>2</sub> , O <sub>2</sub> -O <sub>2</sub> , H <sub>2</sub> O, Ring, liquid water	O <sub>3</sub> , NO <sub>2</sub> , O <sub>2</sub> -O <sub>2</sub> , H <sub>2</sub> O, Ring, liquid water	O <sub>3</sub> , NO <sub>2</sub> , O <sub>2</sub> -O <sub>2</sub> , H <sub>2</sub> O, Ring, liquid water	O <sub>3</sub> , NO <sub>2</sub> , O <sub>2</sub> -O <sub>2</sub> , H <sub>2</sub> O, Ring, liquid water
Undersampling correction <sup>7</sup>	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 scaling model-simulated (B3dCTM model) stratospheric vertical columns to match satellite observations over the remote Pacific [Hilboll et al., 2013]. In the KNMI-approach, NO<sub>2</sub> SCDs are assimilated in the TM4 model, so that model simulations of stratospheric NO<sub>2</sub> 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<sub>2</sub> 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<sub>2</sub> does not feature strong gradients.
- 15 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,

<sup>7</sup> 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.

and methods to estimate the stratospheric NO<sub>2</sub> 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 5 4.2 [Müller et al., 2016]) indicated that the different schemes showed similar stratospheric NO<sub>2</sub> columns and tropospheric residues, and each of the approaches would be appropriate for use in the QA4ECV NO<sub>2</sub> algorithm. The quantitative differences between the stratospheric NO<sub>2</sub> columns were generally smaller than  $0.5 \times 10^{15}$  molecules cm<sup>-2</sup>, 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 10 stratospheric NO<sub>2</sub> 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.

15 The KNMI data assimilation was selected as the default approach for estimating the stratospheric NO<sub>2</sub> column in the QA4ECV algorithm. This ensures consistent knowledge of the state of the atmosphere (NO<sub>2</sub> and temperature profiles, stratospheric dynamics) derived from the same model that predicts the a priori tropospheric NO<sub>2</sub> 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, 20 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<sub>2</sub> estimates to be included 25 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<sub>2</sub> gradients, such as near the polar

Folkert Boersma 8-11-2018 20:38

**Deleted:** model used by MPI-C to compute stratospheric AMFs (

Folkert Boersma 8-11-2018 20:38

**Deleted:** )

Folkert Boersma 31-10-2018 13:13

**Deleted:** .i.e.

Folkert Boersma 2-11-2018 10:17

**Deleted:** s

Folkert Boersma 2-11-2018 10:19

**Deleted:** for particular applications

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<sub>x</sub> 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<sub>2</sub>, 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<sup>15</sup> molecules cm<sup>-2</sup> 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<sub>2</sub> enhancements, such as from outflow of continental pollution over oceans, shipping lanes, and over areas with soil NO<sub>x</sub> emissions.

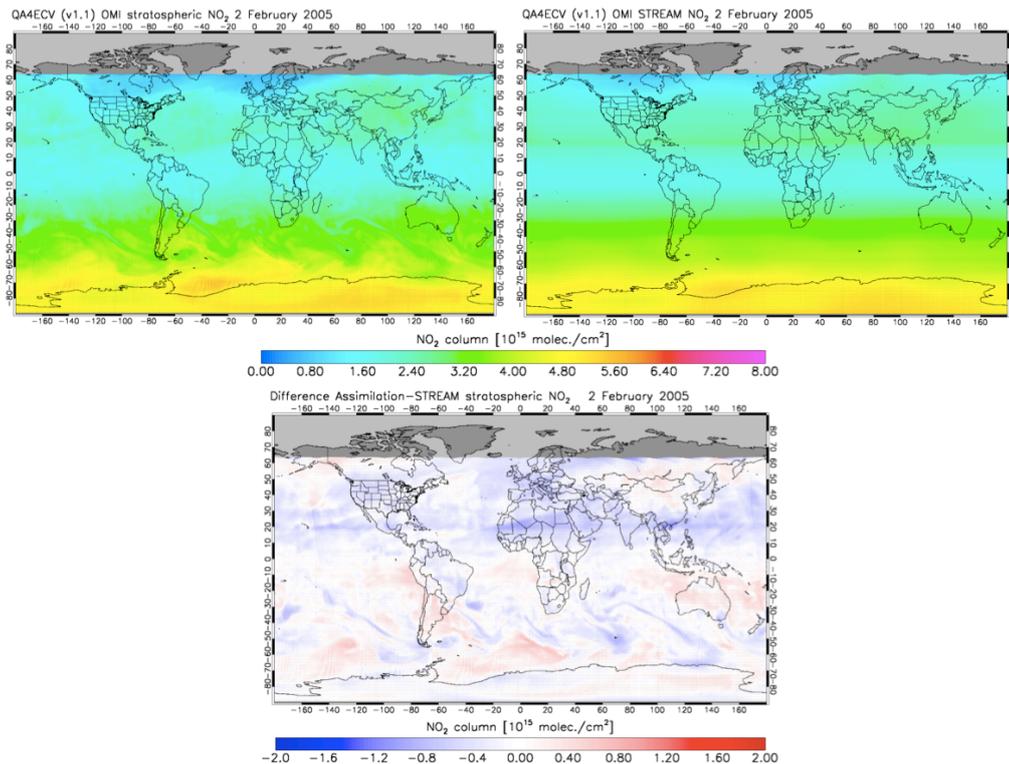
As an example, Figure 5 shows OMI stratospheric NO<sub>2</sub> estimates from both the data assimilation and STREAM approach for the QA4ECV v1.1 product on 2 February 2005. The upper panel illustrates that the latitudinal gradients in NO<sub>2</sub> between the data assimilation and STREAM agree reasonably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

Folkert Boersma 3-11-2018 18:37

Deleted: 6

Folkert Boersma 30-10-2018 16:35

Deleted: agree reasoably well



**Figure 5.** Stratospheric NO<sub>2</sub> 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<sub>2</sub> estimates.

Folkert Boersma 3-11-2018 18:37

Deleted: 6

### 5 5.3 Evaluation of air mass factor calculations

We performed a comparison of approaches to calculate AMFs for NO<sub>2</sub> 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,

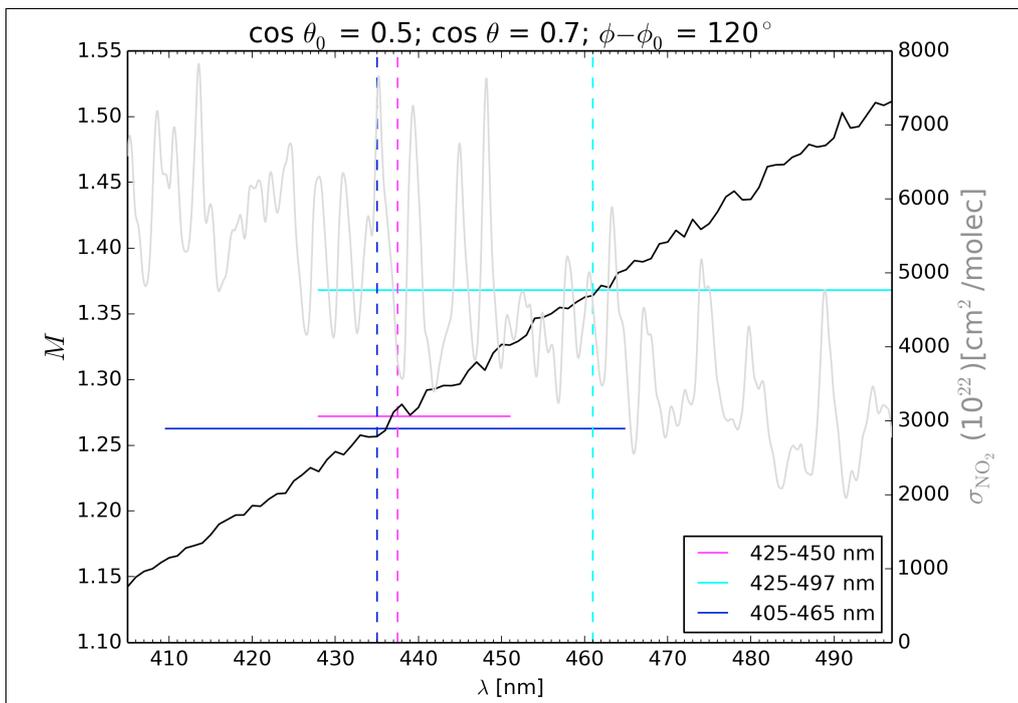
10 McArtim) for their top-of-atmosphere reflectances and their capacity to compute vertically resolved, or

‘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<sub>2</sub> algorithm settings, we selected the appropriate wavelength for calculating the NO<sub>2</sub> box AMFs. We investigated the wavelength dependency of the NO<sub>2</sub> AMFs for retrieval scenarios with substantial tropospheric pollution ( $N_{v,trop} = 16 \times 10^{15}$  molec.cm<sup>-2</sup>), and considered that the AMF calculated at a single wavelength should be representative for the fit window average AMF. Tropospheric NO<sub>2</sub> 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.

Folkert Boersma 3-11-2018 18:37

Deleted: 7



**Figure 6.** NO<sub>2</sub> tropospheric air mass factor (black)<sup>8</sup> as a function of wavelength computed with DAK for a polluted boundary layer for a specific viewing geometry ( $\theta=60^\circ$ ,  $\theta_0=45.6^\circ$ ). 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<sub>2</sub> absorption cross-section from Vandaele et al. (1998) at 220 K. A mid-latitude standard atmosphere was used including O<sub>3</sub>. The AMF was computed for a polluted boundary layer with  $16 \times 10^{15}$  molec/cm<sup>2</sup>, without aerosols, a boundary layer height of 1 km and surface albedo 0.05.

<sup>8</sup>The non-smooth behaviour of the black line is because the spectral resolution of the AMF is not sufficient to resolve the NO<sub>2</sub> 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.

Folkert Boersma 3-11-2018 18:37

Deleted: 7

Folkert Boersma 30-10-2018 10:13

Deleted: red

We saw in Figure 4 that total NO<sub>2</sub> AMFs weakly decrease with wavelength (-0.01%/nm redshift). Figure 6 shows that tropospheric NO<sub>2</sub> 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<sub>2</sub> in the boundary layer towards the UV, so that tropospheric AMFs are smallest for shorter wavelengths. For the fitting windows considered for QA4ECV NO<sub>2</sub> retrievals (425-465 nm and 405-465 nm), we recommend calculating the NO<sub>2</sub> 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<sub>2</sub> 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<sub>2</sub> 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

Folkert Boersma 3-11-2018 18:37

Deleted: 5

Folkert Boersma 30-10-2018 10:16

Deleted: stratospheric

Folkert Boersma 3-11-2018 18:38

Deleted: 7

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<sub>2</sub> 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<sub>2</sub> 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<sub>2</sub> AMFs.

Based on the results from the comparisons discussed above, the following recommendations for calculating QA4ECV NO<sub>2</sub> AMFs were made:

- calculate the NO<sub>2</sub> 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<sub>2</sub> profiles simulated by TM5-MP at 1°×1°. TM5-MP is the model used for the data assimilation of NO<sub>2</sub> SCDs to estimate the stratospheric contribution (section 5.2).

## 5 6 QA4ECV NO<sub>2</sub> uncertainty estimates

### 6.1 Theoretical algorithm uncertainty

The QA4ECV NO<sub>2</sub> product contains an algorithm uncertainty estimate associated with each individual pixel's [tropospheric NO<sub>2</sub> column](#). This estimate is calculated theoretically via uncertainty propagation based on the principal retrieval equation (Eq. (1)):

10

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

The uncertainty propagation accounts for spectral fitting uncertainties ( $\sigma_{N_S}$ ), and contributions from uncertainties in a priori and ancillary data required for calculating the stratospheric NO<sub>2</sub> background  
 15 ( $\sigma_{N_S, strat}$ ) and the AMF ( $\sigma_{M_{tr}}$ ). The uncertainty in the tropospheric AMF, or AMF covariance is written as:

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

20 where  $\frac{\partial M}{\partial A_s}$  represents the local sensitivity of the the air mass factor to surface albedo  $A_s$ , and  $\sigma_{A_s}$  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 ( $\epsilon_{f_{cl}} \epsilon_{A_s}$ ); 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

5 positive contribution to  $\sigma_{M_{tr}}^2$ .

The uncertainty  $\sigma$  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  
 10 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  $\sigma_0$  for spatially or temporally averaged  
 15 data. This method takes the area-weighted (statistical) retrieval uncertainty  $\sigma$ , and then accounts for a partial correlation in the errors between pixels as in Eskes et al. [2003]:

$$\sigma_0 = \sigma \sqrt{\frac{1-c}{n} + c} \quad (4)$$

20 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  $\text{NO}_2$  profile, and aerosols (or lack of description thereof) are typically correlated at the spatiotemporal scales of moderate resolution (global models, i.e. down to  $0.5^\circ \times 0.5^\circ$  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  
 25 reduce to below 39% of the level of typical single-pixel uncertainties ( $\sigma$ ), 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<sub>2</sub> 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<sub>2</sub> 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<sub>2</sub> 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 <sub>2</sub> column uncertainty	Algorithm uncertainty estimate of the tropospheric NO <sub>2</sub> column	Per pixel	$\sigma$
Tropospheric NO <sub>2</sub> column uncertainty when averaging kernel is applied	Algorithm uncertainty estimate, as above, but contribution from profile uncertainty removed	Per pixel	$\sigma_{AK}$
Stratospheric NO <sub>2</sub> 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 <sub>2</sub> columns	Algorithm uncertainty estimate of the total NO <sub>2</sub> column	Per pixel	
NO <sub>2</sub> SCD uncertainty	Uncertainty estimated from the DOAS spectral fitting of NO <sub>2</sub>	Per pixel	$\sigma_{N_S}$
Slant column related	First term on the right hand side of	Per pixel	$\left(\frac{\sigma_{N_S}}{M_{tr}}\right)$

Folkert Boersma 30-10-2018 10:23

Deleted: Per pixel

Folkert Boersma 30-10-2018 10:23

Deleted: a

Folkert Boersma 30-10-2018 10:23

Deleted: Per pixel

Folkert Boersma 30-10-2018 10:23

Deleted: a

Folkert Boersma 30-10-2018 10:23

Deleted: . Same

Folkert Boersma 30-10-2018 10:24

Deleted: Per pixel

Folkert Boersma 30-10-2018 10:24

Deleted: a

uncertainty of the NO <sub>2</sub> tropospheric vertical column	Eq. (2)		
Stratospheric column related uncertainty of the NO <sub>2</sub> tropospheric vertical column	Second term on the right hand side of Eq. (2)	<u>Mix of pixel and global</u>	$\left(\frac{\sigma_{N_{s, strat}}}{M_{tr}}\right)$
Total tropospheric AMF related uncertainty of the tropospheric NO <sub>2</sub> vertical column	Third term on the right hand side of Eq. (2).	<u>Per pixel</u>	$\left(\frac{(N_s - N_{s, strat})\sigma_{M_{tr}}}{M_{tr}^2}\right)$
Surface albedo related uncertainty of the tropospheric vertical NO <sub>2</sub> column	Contribution to the uncertainty of uncertainties in the surface albedo in the tropospheric AMF <sup>9</sup>	<u>Per pixel</u>	$\left(\frac{\partial M}{\partial A_s} \sigma_{A_s}\right) \frac{N_v}{M_{tr}}$
Cloud fraction related uncertainty of the tropospheric vertical NO <sub>2</sub> column	Contribution to the uncertainty of uncertainties in the cloud fraction in the tropospheric AMF	<u>Per pixel</u>	$\left(\frac{\partial M}{\partial f_{cl}} \sigma_{f_{cl}}\right) \frac{N_v}{M_{tr}}$
Cloud pressure related uncertainty of the tropospheric vertical NO <sub>2</sub> column	Contribution to the uncertainty of uncertainties in the cloud pressure in the tropospheric AMF	<u>Per pixel</u>	$\left(\frac{\partial M}{\partial p_{cl}} \sigma_{p_{cl}}\right) \frac{N_v}{M_{tr}}$
TM5 profile related uncertainty of the tropospheric vertical NO <sub>2</sub> column	Global estimate of the contribution to the uncertainty of uncertainties in the TM5 NO <sub>2</sub> profile in the tropospheric AMF	<u>Global</u>	0.1N <sub>v</sub>

<sup>9</sup> 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<sub>2</sub> column uncertainty, we ratio the AMF uncertainty contribution by the AMF itself, and multiply with the actual tropospheric NO<sub>2</sub> column.

processing error flag	Flag indicating whether the processing was successful (0) or failed (-1)	<a href="#">Per pixel</a>	
processing quality flags	Flags indicating conditions that affect the quality of the retrieval	<a href="#">Per pixel</a>	

### 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.

5

#### 6.3.1 Evaluation of NO<sub>2</sub> SCD uncertainties

We compared the DOAS uncertainty estimates ( $\sigma_{N_S}$ ) 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 uncertainty evaluation is described in detail in QA4ECV Deliverable 5.5 [Boersma et al., 2017<sup>a</sup>] 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<sub>2</sub> 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 spectral fitting of NO<sub>2</sub> such, that the typical mission-average SCD uncertainties for both instruments amount to 0.7-0.8×10<sup>15</sup> (was ~1.0×10<sup>15</sup>) molec. cm<sup>-2</sup>. 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, the NO<sub>2</sub> 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<sub>2</sub> 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<sub>2</sub> SCDs.

5

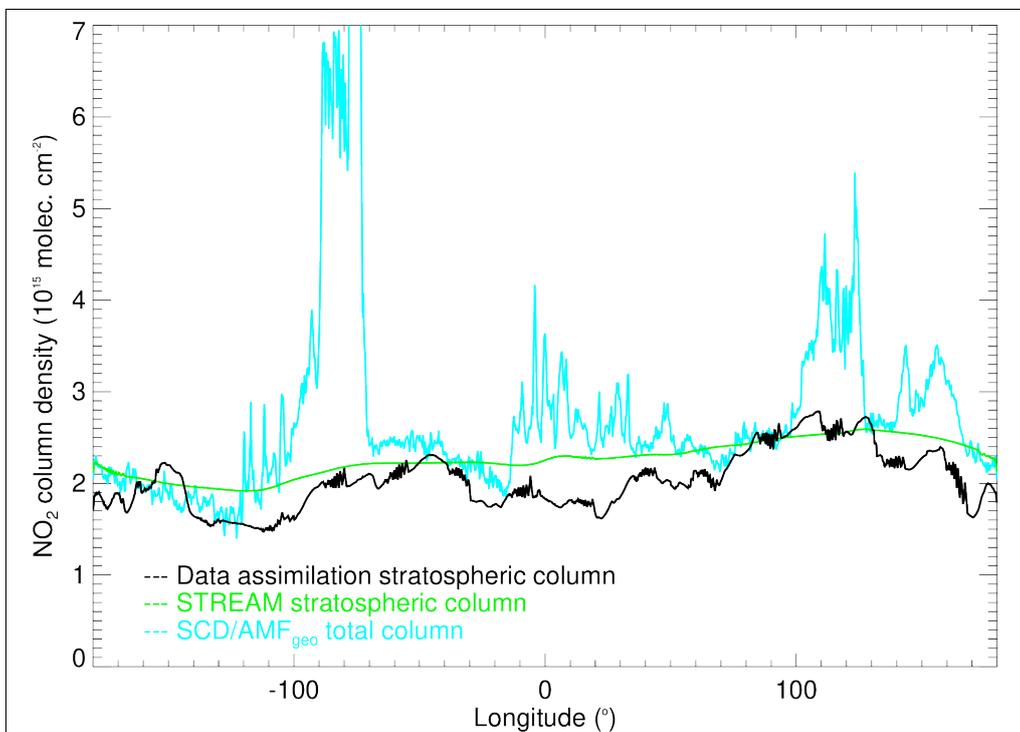
### 6.3.2 Evaluation of uncertainties in the stratospheric correction

The uncertainty of the stratospheric NO<sub>2</sub> vertical column in QA4ECV NO<sub>2</sub> product is based on a global statistical analysis of results from the data assimilation procedure, and documented as  $0.2 \times 10^{15}$  molecules cm<sup>-2</sup> [Dirksen et al., 2011]. The assimilation predicts stratospheric NO<sub>2</sub> columns from an observation-constrained (analyzed) start field and TM5-modeled transport and chemistry. The average discrepancies between the 24-hour forecast and actual satellite-observed NO<sub>2</sub> slant column fields over pristine areas are regarded as a measure for the uncertainty in the stratospheric NO<sub>2</sub> field. In QA4ECV Deliverable 5.5 [Boersma et al., 2017<sup>a</sup>], we verified that the observation minus forecast (O-F) assimilation statistics over the Pacific are indeed consistent with an uncertainty estimate of  $0.2 \times 10^{15}$  molecules cm<sup>-2</sup> 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<sub>2</sub> column estimates for 2 February 2005. There are considerable methodological differences between the data assimilation and STREAM techniques. Yet the data assimilation and STREAM stratospheric NO<sub>2</sub> 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<sub>2</sub> 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 stratospheric NO<sub>2</sub> 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  $0.2-0.5 \times 10^{15}$  molecules cm<sup>-2</sup> 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  $\text{cm}^{-2}$ . These differences reflect the structural uncertainty in stratospheric (vertical)  $\text{NO}_2$  columns, arising when different retrieval methodologies are applied to the same satellite observations, and both uncertainty estimates are included in Table 6.

5

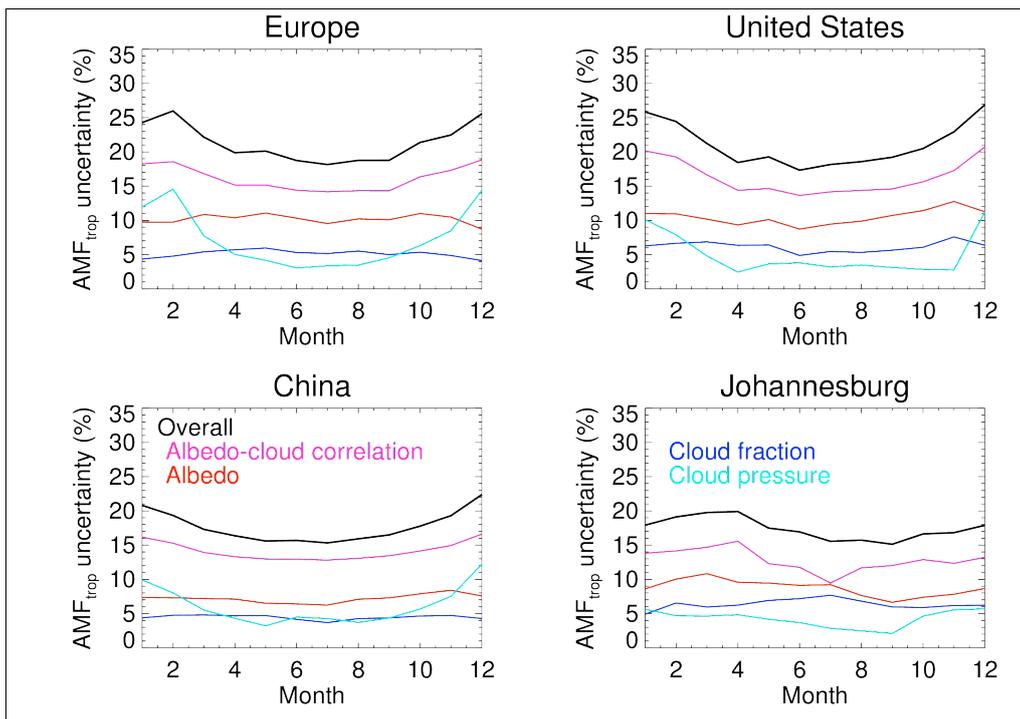


**Figure 7.** Meridional average QA4ECV OMI  $\text{NO}_2$  column averaged over 39-41°N on 2 February 2005. No cloud radiance, albedo, or AMF filtering has been applied. Data assimilation and STREAM stratospheric columns are indicated in the black and green lines, the total slant columns divided by the geometric AMF is light blue. Both data assimilation and STREAM stratospheric column estimates are included in the QA4ECV  $\text{NO}_2$  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<sub>2</sub>, 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<sub>2</sub>-O<sub>2</sub> cloud fraction and cloud pressure estimates, respectively, based on recent improvements in the cloud algorithm [Veeffkind et al., 2016], and a 10% contribution from NO<sub>2</sub> profile uncertainty. The latter is based on comparing AMFs calculated with simulated a priori profiles to AMFs calculated with measured NO<sub>2</sub> profiles from aircraft and lidar (e.g. Hains et al. [2010] and references therein).

Apart from the overall AMF uncertainty estimate, the QA4ECV NO<sub>2</sub> 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<sub>2</sub>. 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 ( $2 \left( \frac{\partial M}{\partial A_s} \frac{\partial M}{\partial f_{cl}} (\epsilon_{f_{cl}} \epsilon_{A_s}) \right)$ ) 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<sub>2</sub>



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

Folkert Boersma 30-10-2018 10:28

Deleted: pink

Folkert Boersma 30-10-2018 10:28

Deleted: s

Folkert Boersma 30-10-2018 10:29

Deleted: surface albedo-cloud fraction error correlations, the red line indicates the AMF uncertainties due to surface albedo uncertainty.

and a complex mixture of aerosols, clouds, and NO<sub>2</sub> 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  $\pm 30\%$  in Summer and  $\pm 50\%$  in Winter.

10 **Table 6.** Comparison of uncertainty estimates for the main QA4ECV OMI NO<sub>2</sub> retrieval steps. The SCD and stratospheric SCD uncertainties are representative for all possible retrieval scenarios. AMF uncertainties are representative for situations with high NO<sub>2</sub>.

	Algorithm uncertainty (molec. cm <sup>-2</sup> )	Independent uncertainty estimate (molec. cm <sup>-2</sup> )	Structural uncertainty (molec. cm <sup>-2</sup> )
SCD ( $\sigma_{N_s}$ )	$0.8 \times 10^{15}$ <sup>(a)</sup>	$0.7 \times 10^{15}$ <sup>(a)</sup>	$1.0 \times 10^{15}$ <sup>(b)</sup>
Stratospheric SCD ( $\sigma_{N_s, strat}$ )	$0.2 \times 10^{15} \cdot M_{strat}$ <sup>(c)</sup>		$(0.2-0.5) \times 10^{15} \cdot M_{strat}$ <sup>(d)</sup>
Tropospheric AMF ( $\sigma_{M_{tr}}$ )	20% (Summer) 25% (Winter)		15% (Summer) <sup>(e)</sup> 40% (Winter) <sup>(e)</sup>

<sup>(a)</sup>Zara et al. [2018]

<sup>(b)</sup>Section 5.1 of this work

15 <sup>(c)</sup>Dirksen et al. [2011] and analysis of data assimilation observation minus forecast differences  
| QA4ECV Deliverable 5.5 [Boersma et al., 2017<sup>d</sup>].

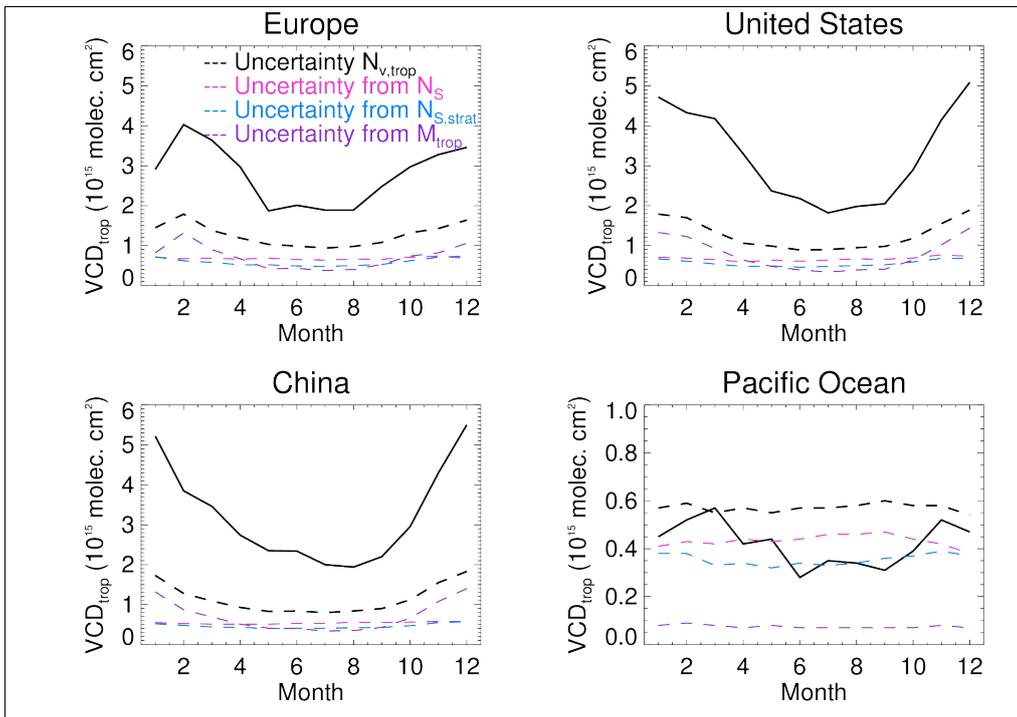
<sup>(d)</sup>Figure 7 of this work.

<sup>(e)</sup>Lorente et al. [2017]

## 6.4 Overall uncertainties in tropospheric NO<sub>2</sub> columns

### 6.4.1 Uncertainties in single-pixel tropospheric NO<sub>2</sub> columns

Here we present estimates of typical algorithm, single-pixel uncertainties for the QA4ECV NO<sub>2</sub> 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<sub>2</sub> column uncertainty exceeds 100%, and is dominated year-round by the uncertainties in SCD and stratospheric column estimate.



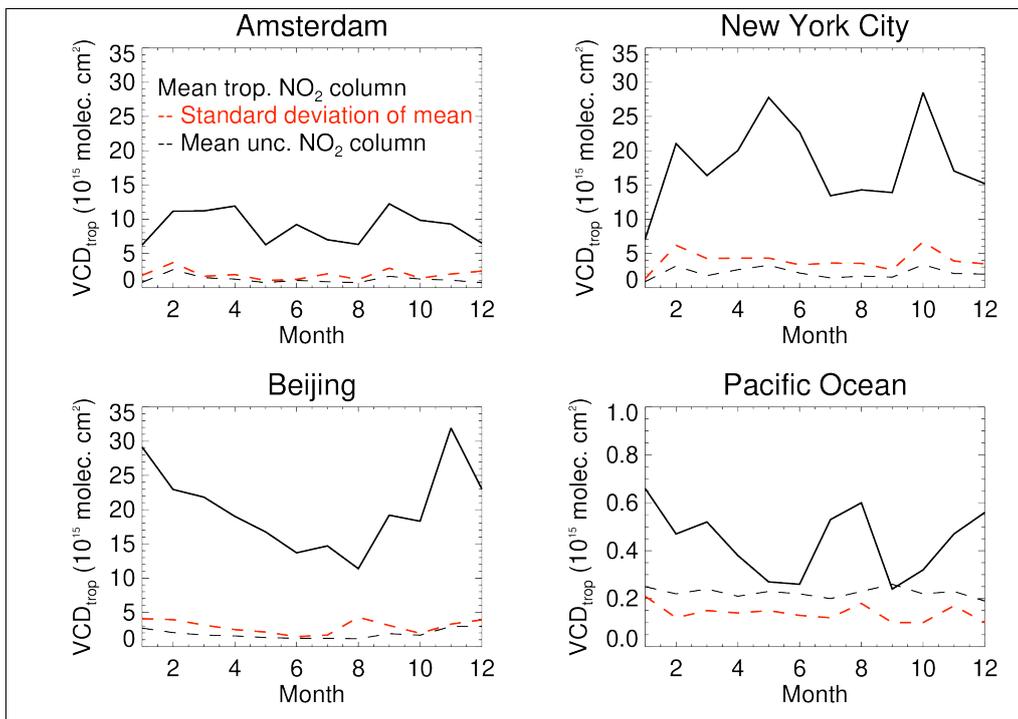
**Figure 9.** Average (single-pixel equivalent) QA4ECV OMI tropospheric NO<sub>2</sub> columns (solid line) and associated total uncertainties (dashed black line) for Europe (40-50°N, 10°W-15°E), United States (35-45°N, 100-75°W), eastern China (30-45°N, 110-140°E), and the Pacific Ocean (35-45°N, 160-140°W) in 2005. The dashed coloured lines indicate the contributions to the tropospheric NO<sub>2</sub> column uncertainty from SCD (pink,  $(\frac{\sigma_{N_s}}{M_{tr}})$ ), stratospheric correction (light blue,  $(\frac{\sigma_{N_{s, strat}}}{M_{tr}})$ ), and the tropospheric AMF (purple,  $(\frac{(N_s - N_{s, strat})\sigma_{M_{tr}}}{M_{tr}^2})$ ).

#### 6.4.2 Uncertainties in averaged tropospheric NO<sub>2</sub> 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 from imperfect knowledge of surface albedo, a priori NO<sub>2</sub> 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<sub>2</sub> columns is 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<sub>2</sub> column over a certain region, we recommend taking whichever is largest: (a) the temporally averaged values for  $\sigma_o$  (Eq. (4)), or (b) the standard deviation of the mean (standard error) of the daily tropospheric NO<sub>2</sub> columns. If there is substantial temporal variability (from changes in photochemistry, transport events), the standard error will be a good representation of the uncertainty in the monthly mean tropospheric NO<sub>2</sub> column. Figure 10 shows a comparison of monthly averaged uncertainties  $\sigma_o$  and the local standard deviation of the mean NO<sub>2</sub> columns for four small regions (0.25° × 0.25°). The figure confirms that the averaged uncertainties provide an optimistic estimate of the uncertainty, at ±10%, in the monthly mean NO<sub>2</sub> columns. For the polluted regions, the standard 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 ( $\sigma_0$ , dashed black line) in 2005 over Amsterdam ( $52.375^\circ N$ ,  $4.875^\circ E$ ), New York City ( $40.875^\circ N$ ,  $73.875^\circ W$ ), Beijing ( $39.875^\circ N$ ,  $116.375^\circ E$ ), and the Pacific ( $39.875^\circ N$ ,  $149.875^\circ W$ ). Grid cell size  $0.25^\circ \times 0.25^\circ$ . 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  
5 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<sub>2</sub> columns are driven  
by noise, air mass factor and temperature uncertainties amounting to approximately 15% uncertainty.  
10 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<sub>2</sub> 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.

15 We compare OMI NO<sub>2</sub> 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  
20 require that the OMI columns were measured within 30 minutes of the coinciding MAX-DOAS  
measurement, have a pixel footprint area < 700 km<sup>2</sup>, 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  
25 discrepancies between MAX-DOAS and satellite NO<sub>2</sub> columns over strongly polluted regions. Such  
discrepancies are at least partly due to spatial inhomogeneity in the NO<sub>2</sub> 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<sub>2</sub> 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<sub>2</sub> columns ( $N_{V,p}$ , i.e. what is usually validated) within our criteria to values more representative for the location of the Tai'an ( $N_{V,T}$ ):

5

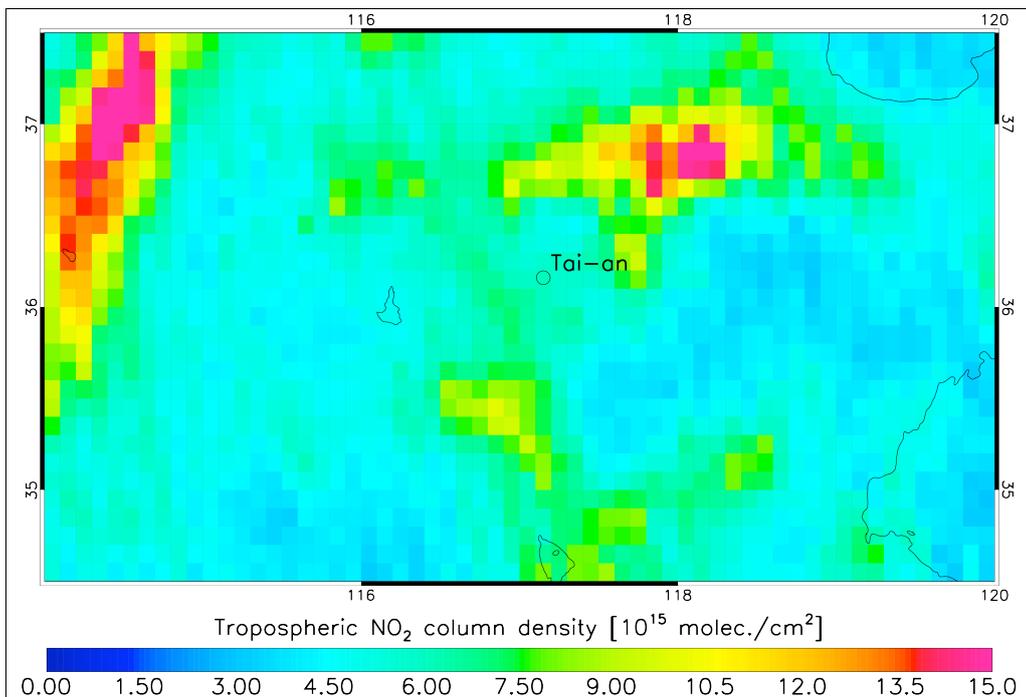
$$N_{V,T} = \left( \frac{\overline{N_{V,T}}}{N_{V,p}} \right) \cdot N_{V,p} \quad (5)$$

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  $1 \times 10^{15}$  molec.cm<sup>-2</sup> (<20% of the Tai'an column, see Figure S2).

Folkert Boersma 3-11-2018 18:33

Deleted: 1

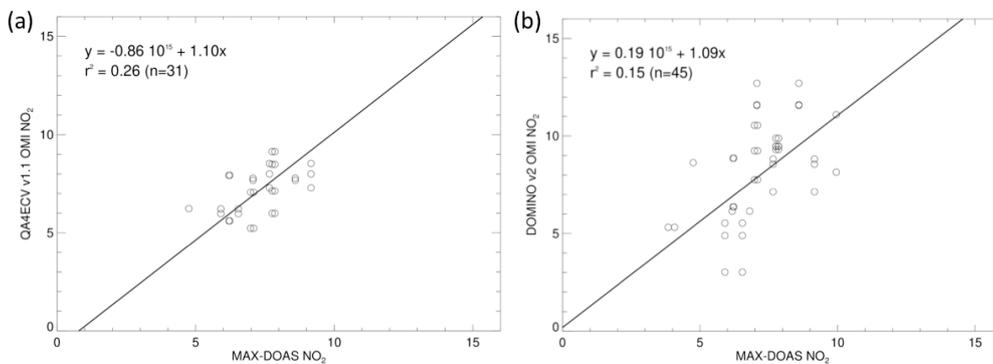


**Figure 11.** Campaign mean (30 May – 30 June 2006) QA4ECV tropospheric NO<sub>2</sub> column distribution over eastern China for clear-sky situations (cloud radiance fraction < 0.5). The black circle indicates the location of Tai'an, where Chiba University operated the MAX-DOAS instrument. One cell corresponds to 0.1° × 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<sub>2</sub> 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<sub>2</sub> columns for Tai'an. We find a bias (mean difference) of  $-0.15 \times 10^{15}$  molec.cm<sup>-2</sup> (-2%) and the root mean squared deviation is  $1.08 \times 10^{15}$  molec.cm<sup>-2</sup> (16%). Not applying the scaling factors

from Eq. (5) leads to a bias of  $-0.47 \times 10^{15}$  molec.cm<sup>-2</sup> (-7%) and a root mean squared deviation of  $1.19 \times 10^{15}$  molec.cm<sup>-2</sup> (18%). Using a reduced major axis regression analysis, we find a relationship between QA4ECV ( $y$ ) and MAX-DOAS NO<sub>2</sub> columns ( $x$ ) as  $y = -0.86 \times 10^{15}$  molec.cm<sup>-2</sup> + 1.10x ( $R^2=0.26$ ,  $n=31$ ). Including also pixels with high effective cloud pressures (>875 hPa) leads to a bias of  $-0.48 \times 10^{15}$  molec.cm<sup>-2</sup> (-6.6%,  $n=37$ ) and a root mean squared deviation of  $1.35 \times 10^{15}$  molec.cm<sup>-2</sup> (20%).

Figure 12(b) shows the scatter plot of DOMINO v2 vs. MAX-DOAS NO<sub>2</sub> 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<sub>2</sub>-O<sub>2</sub> cloud product [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; Veeffkind 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<sup>-2</sup> (+11%,  $n=45$ ), with a root mean squared deviation of  $2.66 \times 10^{15}$  molec.cm<sup>-2</sup> (35%).



**Figure 12.** (a) Scatterplot of QA4ECV OMI vs. MAX-DOAS tropospheric NO<sub>2</sub> 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 now for DOMINO v2 vs. MAX-DOAS tropospheric NO<sub>2</sub>.

The differences between OMI and MAX-DOAS NO<sub>2</sub> 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  $\sigma_O$  the uncertainty reported (in the data files) for QA4ECV OMI NO<sub>2</sub> columns,  $\sigma_{MD}$  the uncertainty reported for MAX-DOAS NO<sub>2</sub> columns, and  $\sigma_R$  the uncertainty from spatiotemporal mismatches between the satellite and ground-based measurement (Table 7). The mean reported uncertainties ( $\sigma_O$  and  $\sigma_{MD}$ ) 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<sub>2</sub> column values). We see from Figure 13 that the differences between OMI and MAX-DOAS NO<sub>2</sub> 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  $\sigma_O$  or  $\sigma_{MD}$  have been overestimated, or (4) a combination of the above. The MAX-DOAS NO<sub>2</sub> retrieval technique suffers from some similar error contributions (a priori NO<sub>2</sub> 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  $\sigma_O$  or  $\sigma_{MD}$  (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  $\sigma_O$  and  $\sigma_{MD}$  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 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<sub>2</sub> 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.  $\sigma_O$  (reported OMI uncertainty) and  $\sigma_M$  (reported MAX-DOAS uncertainty) are the mean of 31 or 45 individual values, and  $\sigma_R$  is considered to be a 10% contribution from mismatches.

	Expected differences (QA4ECV)	Observed differences (QA4ECV)	Expected differences (DOMINO v2)	Observed differences (DOMINO v2)
$\sigma$	$2.11 \times 10^{15}$ molec. cm <sup>-2</sup>	$1.29 \times 10^{15}$ molec. cm <sup>-2</sup>	$2.57 \times 10^{15}$ molec. cm <sup>-2</sup>	$2.16 \times 10^{15}$ molec. cm <sup>-2</sup>
$\sigma_O$	$1.71 \times 10^{15}$ molec. cm <sup>-2</sup>		$2.21 \times 10^{15}$ molec. cm <sup>-2</sup>	
$\sigma_M$	$1.08 \times 10^{15}$ molec. cm <sup>-2</sup>		$1.27 \times 10^{15}$ molec. cm <sup>-2</sup>	
$\sigma_R$	$0.60 \times 10^{15}$ molec. cm <sup>-2</sup>		$0.60 \times 10^{15}$ molec. cm <sup>-2</sup>	

15

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 [Compernelle 2018].

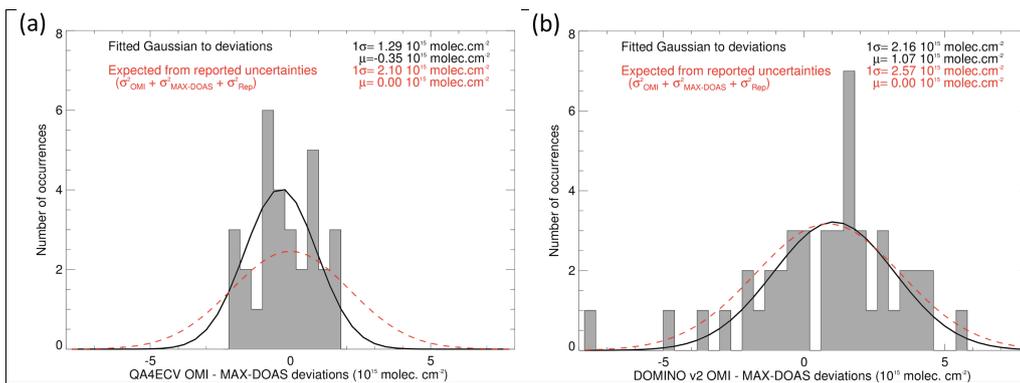
20

Folkert Boersma 8-11-2018 20:56

Deleted: it is likely that

Folkert Boersma 8-11-2018 20:57

Deleted: are



**Figure 13.** (a) Histogram of differences of QA4ECV OMI vs. MAX-DOAS tropospheric NO<sub>2</sub> 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<sub>2</sub>.

## 8 Data availability and DOI

The QA4ECV NO<sub>2</sub> Essential Climate Variable precursor product contains vertical NO<sub>2</sub> 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<sub>2</sub> ECV precursor data provides geophysical information for each and every ground pixel observed by GOME, SCIAMACHY, OMI, and GOME-2(A). The QA4ECV NO<sub>2</sub> data product is available online via [www.qa4ecv.eu](http://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<sup>b,c,d,e</sup>]. Detailed information on how to use the data can be found in the Product Specification Document for NO<sub>2</sub> ECV Precursor product [Boersma et al., 2017<sup>f</sup>].

5

## 9 Summary

We have developed an improved algorithm and uncertainty assessment for tropospheric NO<sub>2</sub> satellite retrievals from UV/VIS satellite sensors. Our effort has resulted in the generation of a 1995-2017 climate data record of tropospheric NO<sub>2</sub> columns with fully traceable uncertainty metrics that can be readily used for model evaluation, for estimating NO<sub>x</sub> 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 components of the QA4ECV NO<sub>2</sub> 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. 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<sub>2</sub> by accounting for liquid water absorption, and an intensity-offset correction. This improved the quality of the NO<sub>2</sub> fit over clear-sky ocean scenes by up to 30% [Zara et al., 2018], but did not substantially affect the NO<sub>2</sub> fits over polluted scenes. We compared three alternate methods for estimating the stratospheric NO<sub>2</sub> 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  $1^\circ \times 1^\circ$  horizontal resolution, a major step forward compared to earlier assimilation schemes based on TM4 ( $3^\circ \times 2^\circ$ ), 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<sub>2</sub> 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<sub>2</sub> retrievals from S5P-TROPOMI.

The QA4ECV NO<sub>2</sub> 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<sub>2</sub> 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<sub>2</sub> 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<sub>2</sub> column measurements obtained over the polluted Tai'an (China) region in June 2006 to validate the OMI QA4ECV NO<sub>2</sub> 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<sub>2</sub> 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<sub>2</sub> columns

Folkert Boersma 30-10-2018 10:43

Deleted: 3

Folkert Boersma 30-10-2018 10:43

Deleted: 2

Folkert Boersma 30-10-2018 10:43

Deleted: 8

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<sub>2</sub> AMF  
10 intercomparison.

### References

- Anand, J. S., Monks, P. S., and Leigh, R. J.: An improved retrieval of tropospheric NO<sub>2</sub> 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<sub>2</sub> 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<sub>2</sub> retrieval from space, *J. Geophys. Res.*, 109, D04311, doi:10.1029/2003JD003962, 2004.

- Boersma, K. F., Eskes, H. J., Veeffkind, J. P., Brinkma, 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<sub>2</sub> 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., Veeffkind, 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<sub>2</sub> 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., Compernelle, 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](http://www.qa4ecv.eu/sites/default/files/D5.5_v1.0.compressed.pdf), last access: 12 April 2018, 2017<sup>a</sup>.
- 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<sub>2</sub> 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<sup>b</sup>.
- 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<sub>2</sub> 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<sup>c</sup>.

Boersma, K. F., Eskes, H., Richter, A., De Smedt, I., Lorente, A., Beirle, S., Van Geffen, J., Peters, E., Van Roozendaal, M. and Wagner, T.: QA4ECV NO<sub>2</sub> 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<sup>d</sup>.

Boersma, K. F., Eskes, H., Richter, A., De Smedt, I., Lorente, A., Beirle, S., Van Geffen, J., Peters, E., Van Roozendaal, M. and Wagner, T.: QA4ECV NO<sub>2</sub> 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<sup>e</sup>.

Boersma, K. F., van Geffen, J., Eskes, H., van der A, R. J., De Smedt, I., Van Roozendaal, M., Yu, H., Richter, A., Peters, E., Beirle, S., Wagner, T., Lorente, A., Scanlon, T., Compernelle, S., and Lambert, J.-C.: Product Specification Document for the QA4ECV NO<sub>2</sub> Precursor Product (Version 1.1), available 15 at: [http://temis.nl/qa4ecv/no2col/QA4ECV\\_NO2\\_PSD\\_v1.1.compressed.pdf](http://temis.nl/qa4ecv/no2col/QA4ECV_NO2_PSD_v1.1.compressed.pdf) (last access: 25 April 2018), 2017<sup>f</sup>.

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. 25

Bucsele, E. J., Celarier, E. A., Wenig, M. O., Gleason, J. F., Veefkind, J. P., Boersma, K. F., and Brinksma, E. J.: Algorithm for NO<sub>2</sub> 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.*, 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 Compernelle, S., Lambert, J.-C., Boersma, K. F., Schulz, J., Müller, J.-P., Coheur, P.-F., De Smedt, I., Van Roozendaal, 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](http://www.qa4ecv.eu/sites/default/files/QA4ECV_D6p1_FINAL%2BSMM.pdf), last access: 12 April 2018, 2018.
- 15 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](http://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<sub>2</sub> retrieved from the ozone monitoring instrument: intercomparison, diurnal cycle, and trending, *J. Geophys. Res.*, 116, D08305, doi:10.1029/2010JD014943, 2011.
- 5 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](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.
- 15 Drosoglou, T., Bais, A. F., Zyrichidou, I., Kouremeti, N., Poupkou, A., Liora, N., Giannaros, C., Koukoulis, M. E., Balis, D., and Melas, D.: Comparisons of ground-based tropospheric NO<sub>2</sub> 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.
- 20 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](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](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](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<sub>x</sub> emissions: Implications for climate mitigation measures, *Atmos. Environ.*, 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](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<sub>2</sub> 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<sub>2</sub> 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<sub>2</sub> 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<sub>2</sub> 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 NO<sub>2</sub> and SO<sub>2</sub> 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., Bucsele, E. J., Joiner, J., Duncan, B. N., Boersma, K. F., Veeffkind, 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<sub>2</sub> and NO<sub>2</sub> 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<sub>2</sub> estimations at terrestrial and marine sites, *Journal of Geophysical Research: Atmospheres*, 123, 1441–1459, <https://doi.org/10.1002/2017JD026518>, 2018.
- 5
- 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 Roozendaal, 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<sub>2</sub> 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<sub>2</sub> retrievals from GOME-2 and OMI, *Atmos. Meas. Tech.*, 11, 4509-4529, <https://doi.org/10.5194/amt-11-4509-2018>, 2018.
- 25
- Maasackers, J. D.: Vital improvements to the retrieval of tropospheric NO<sub>2</sub> 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<sub>x</sub> emission estimates derived from an assimilation of OMI tropospheric NO<sub>2</sub> columns, *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 Albin~ana, A. P.: GOME-2 on MetOp, Proc. of The 2006 EUMETSAT Meteorological 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](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., Compernelle, 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](https://doi.org/10.3390/rs10081254), 2018.

10 [Oldeman, A.: Effect of including an intensity offset in the DOAS NO<sub>2</sub> retrieval of TROPOMI, Internship report, R-1944-SE, Eindhoven University of Technology/KNMI, Eindhoven, May 2018, \[https://kfolkertboersma.files.wordpress.com/2018/06/report\\\_oldeman\\\_22052018.pdf\]\(https://kfolkertboersma.files.wordpress.com/2018/06/report\_oldeman\_22052018.pdf\), last access: 30 October 2018, 2018.](https://kfolkertboersma.files.wordpress.com/2018/06/report_oldeman_22052018.pdf)

Pinardi, G., Van Roozendael, M., Hendrick, F., et al.: Satellite nadir NO<sub>2</sub> validation based on direct-sun and MAXDOAS network observations, 8<sup>th</sup> International DOAS Workshop, Yokohama, Japan, 4-6 September, 2017, O2-01, 2017.

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

20 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<sub>2</sub> retrieval for the GOME-2 satellite instrument, *Atmos. Meas. Tech.*, 4(6), 1147–1159, [doi:10.5194/amt-4-1147-2011](https://doi.org/10.5194/amt-4-1147-2011), 2011.

- Roscoe, H. K., Van Roozendaal, 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<sub>2</sub> and O<sub>4</sub> 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.
- 10
- 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](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.
- 15
- 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.
- 25
- Slijkhuis, S., Aberle, B., Coldewey-Egbers, M., Loyola, D., Dehn, A., Fehr, T.: GOME/ERS-2: New 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 Roozendaal, 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<sub>2</sub>-O<sub>2</sub> 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.
- 20 Vinken, G. C. M., Boersma, K. F., Maasakkers, J. D., Adon, M., and Martin, R. V.: Worldwide biogenic soil NO<sub>x</sub> emissions inferred from OMI NO<sub>2</sub> 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 Roozendaal, M.: FRESCO+: an improved O<sub>2</sub> 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](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](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:  
20 [https://library.wmo.int/opac/doc\\_num.php?explnum\\_id=3710](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 Roozendaal, 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.

5 Zhang, L., Jacob, D. J., Boersma, K. F., Jaffe, D. A., Olson, J. R., Bowman, K. W., Worden, J. R., 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.