



Validation of Aura-OMI QA4ECV NO₂ climate data records with ground-based DOAS networks: the role of measurement and comparison uncertainties

Steven Compernelle¹, Tijn Verhoelst¹, Gaia Pinardi¹, José Granville¹, Daan Hubert¹, Arno Keppens¹, Sander Niemeijer², Bruno Rino², Alkis Bais³, Steffen Beirle⁴, Folkert Boersma^{5,8}, John P. Burrows⁶, Isabelle De Smedt¹, Henk Eskes⁵, Florence Goutail⁷, François Hendrick¹, Alba Lorente⁸, Andrea Pazmino⁷, Ankie PETERS⁵, Enno Peters⁶, Jean-Pierre Pommereau⁷, Julia Remmers⁴, Andreas Richter⁶, Jos van Geffen⁵, Michel Van Roozendael¹, Thomas Wagner⁴, and Jean-Christopher Lambert¹

¹Royal Belgian Institute for Space Aeronomy (BIRA-IASB), Uccle, Belgium

²Stamps Corporation, Delft, the Netherlands

³Aristotle University of Thessaloniki, Laboratory of Atmospheric Physics (AUTH), Thessaloniki, Greece

⁴Max Planck Institute for Chemistry (MPIC), Mainz, Germany

⁵Royal Netherlands Meteorological Institute (KNMI), De Bilt, the Netherlands

⁶Institute of Environmental Physics, University of Bremen (IUP-B), Bremen, Germany

⁷Laboratoire Atmosphères, Milieux, Observations Spatiales, CNRS, Guyancourt, France

⁸Wageningen University, Meteorology and Air Quality Group, Wageningen, the Netherlands

Correspondence: Steven Compernelle (steven.compernelle@aeronomie.be)

Received: 27 September 2019 – Discussion started: 2 January 2020

Revised: 30 April 2020 – Accepted: 24 May 2020 – Published: 10 July 2020

Abstract. The QA4ECV (Quality Assurance for Essential Climate Variables) version 1.1 stratospheric and tropospheric NO₂ vertical column density (VCD) climate data records (CDRs) from the OMI (Ozone Monitoring Instrument) satellite sensor are validated using NDACC (Network for the Detection of Atmospheric Composition Change) zenith-scattered light differential optical absorption spectroscopy (ZSL-DOAS) and multi-axis DOAS (MAX-DOAS) data as a reference. The QA4ECV OMI stratospheric VCDs have a small bias of ~ 0.2 Pmolec. cm⁻² (5%–10%) and a dispersion of 0.2 to 1 Pmolec. cm⁻² with respect to the ZSL-DOAS measurements. QA4ECV tropospheric VCD observations from OMI are restricted to near-cloud-free scenes, leading to a negative sampling bias (with respect to the unrestricted scene ensemble) of a few peta molecules per square centimetre (Pmolec. cm⁻²) up to -10 Pmolec. cm⁻² (-40%) in one extreme high-pollution case. The QA4ECV OMI tropospheric VCD has a negative bias with respect to the MAX-DOAS data (-1 to -4 Pmolec. cm⁻²), which is a feature also found for the OMI OMNO2 standard data

product. The tropospheric VCD discrepancies between satellite measurements and ground-based data greatly exceed the combined measurement uncertainties. Depending on the site, part of the discrepancy can be attributed to a combination of comparison errors (notably horizontal smoothing difference error), measurement/retrieval errors related to clouds and aerosols, and the difference in vertical smoothing and a priori profile assumptions.

1 Introduction

Nitrogen oxides (NO_x = NO₂ + NO) play a significant role in the atmosphere, as they catalyse tropospheric ozone formation via a suite of chemical reactions, impact the oxidizing capacity of the atmosphere and, thus, influence the atmospheric burdens of major pollutants like methane and carbon monoxide (Seinfeld and Pandis, 1997). In addition, they are responsible for secondary aerosol formation (Sillman et al., 1990). Fossil fuel combustion is the dominant source of the

global NO_x emission budget (~ 50 %), followed by natural emissions from soils, lightning and open vegetation fires (Delmas et al., 1997). High ozone, aerosol and NO_x have adverse effects on human health (Hoek et al., 2013; World Health Organization, 2013), and the recommended limits from the European Union (EU) and the World Health Organization (WHO) are often exceeded, especially in densely populated and industrialized regions (European Environment Agency, 2018). Therefore, emissions of NO_x have been the main target of abatement strategies worldwide (e.g. the Protocol of Gothenburg, 1999). The effects of NO_x emissions on climate are complex and are currently not fully understood. On the one hand, emissions of NO_x result in an increase in ozone and, thus, a net warming (as ozone is a greenhouse gas); on the other hand, they lead to a decrease in methane abundances at longer timescales and, therefore, to a cooling effect (Myhre et al., 2013). Due to their indirect impact on radiative forcing and potential affect on climate (Shindell et al., 2009), NO_x has been identified as an “Essential Climate Variable” (ECV) precursor by the Global Climate Observing System (GCOS; GCOS, 2016). NO_x is also present in the stratosphere (Noxon, 1979), where it contributes to the catalytic destruction of ozone (Crutzen, 1970).

Observations from satellite nadir-viewing sensors are essential for mapping the global multiyear picture of the NO_x distribution and trend. However, the quality of these data sets needs to be carefully assessed using ground-based measurements at different sites (see e.g. Petritoli et al., 2004; Pinardi et al., 2014; Heue et al., 2005; Brinksmas et al., 2008; Celarier et al., 2008, for validations on GOME, Global Ozone Monitoring Experiment; GOME-2, Global Ozone Monitoring Experiment-2; SCIAMACHY, Scanning Imaging Absorption Spectrometer for Atmospheric Cartography; and OMI, Ozone Monitoring Instrument data). A limitation often encountered is that uncertainties in satellite and/or ground-based data are not adequately characterized, and the ground-based data sets are generally not harmonized across networks.

The EU Seventh Framework Programme (FP7) QA4ECV (Quality Assurance for Essential Climate Variables) project (<http://www.qa4ecv.eu>, last access: 20 April 2020) demonstrated how reliable and traceable quality information can be provided for satellite and ground-based measurements of climate and air quality parameters. Here, we highlight three of its achievements:

1. The development of a quality assurance framework for climate data records (CDRs; Nightingale et al., 2018), covering aspects such as product traceability, uncertainty description, validation and documentation, following international standards (QA4EO, 2019; Joint Committee for Guides in Metrology, 2008, 2012). Among its components are a generic validation protocol (Compornolle et al., 2018, building upon Keppens et al., 2015), a compilation of recommended terminology for

CDR quality assessment (Compornolle and Lambert, 2017; Compornolle et al., 2018) and a validation server (Compornolle et al., 2016; Rino et al., 2017); the latter is a prototype for the operational validation servers for S5P-MPC (Sentinel-5P Mission Performance Center) and CAMS (Copernicus Atmosphere Monitoring Service).

2. The establishment of multi-decadal CDRs for six ECVs following the guidelines of the quality assurance framework; among them are the QA4ECV NO₂ (Lorente et al., 2017; Zara et al., 2018; Boersma et al., 2018) and the HCHO (De Smedt et al., 2018) version 1.1 satellite products, which are available for several sensors.
3. The development of an NO₂ and HCHO long-term ground-based data set for 10 MAX-DOAS instruments, harmonized with respect to measurement protocol and data format and with an extensive uncertainty characterization (Hendrick et al., 2016; Richter et al., 2016).

A general across-community issue in the geophysical validation of satellite data sets with respect to ground-based reference measurements is the additional uncertainty that appears when comparing data sets characterized by different temporal/spatial/vertical sampling and smoothing properties (Loew et al., 2017). This is especially critical for short-lived tropospheric gases (Richter et al., 2013b). This issue was the focus of the EU Horizon 2020 GAIA-CLIM (Gap Analysis for Integrated Atmospheric ECV CLimate Monitoring; Verhoelst et al., 2015; Verhoelst and Lambert, 2016) project.

In this work, we report a comprehensive validation of the QA4ECV NO₂ version 1.1 data product on the OMI sensor using the ground-based measurements acquired by DOAS (differential optical absorption spectroscopy) UV–Vis instrument networks developed in the context of the Network for the Detection of Atmospheric Composition Change (NDACC) as a reference. Zenith-scattered light DOAS (ZSL-DOAS) data obtained routinely as part of NDACC monitoring activities are used to validate the stratospheric vertical column density (VCD), while multi-axis DOAS (MAX-DOAS) data, either from NDACC or further harmonized within the QA4ECV project, are used to validate the tropospheric VCD. We focus on how well the *ex ante*¹ uncertainties and comparison errors explain the observed discrepancies, making use of the framework and methodology developed within the QA4ECV and GAIA-CLIM projects.

This paper is structured as follows. In Sect. 2, the satellite and reference data sets are described. Section 3.1 provides details about the validation methodology. In Sect. 3.2, we outline how the quality screening of QA4ECV OMI NO₂,

¹An *ex ante* quantity does not rely on a statistical comparison with external data (von Clarmann, 2006). This is to be contrasted with *ex post* quantities like the mean difference of satellite data vs. reference data.

notably the exclusion of cloudy scenes, leads to underestimated early afternoon tropospheric NO₂ VCDs. Section 3.3 presents the comparison of the QA4ECV OMI stratospheric NO₂ VCD with ZSL-DOAS. In Sect. 3.4, the satellite tropospheric VCD is compared with measurements from 10 MAX-DOAS instruments. The differences are analysed in relation to the uncertainties and the comparison errors. Potential causes of the discrepancies (e.g. horizontal smoothing difference error, low-lying clouds or aerosols, and profile shape uncertainty) and attempts to resolve the discrepancies are then discussed. Finally, the conclusions are formulated in Sect. 4.

2 Description of the data sets

2.1 Satellite data

2.1.1 QA4ECV OMI NO₂

The QA4ECV NO₂ OMI version 1.1 data product is retrieved from Level 1 UV–Vis spectral measurements (OMI-Aura_L1-OML1BRVG radiance files) from the Dutch–Finnish UV–Vis nadir-viewing OMI (Ozone Monitoring Instrument) spectrometer on NASA’s Earth Observing System Aura (EOS-Aura) polar satellite. The nominal footprint of the OMI ground pixels is 24 × 13 km² (across × along track) at nadir to 165 × 13 km² at the edges of the 2600 km swath, and the ascending node local time is 13:42 LT. For more details on the instrument, see Levelt et al. (2006). The data product provides a Level 2 (L2) tropospheric, stratospheric and total NO₂ VCD.

The QA4ECV algorithm includes the following steps: (i) retrieving the total slant column density (SCD) N_s using differential optical absorption spectroscopy (DOAS), (ii) estimating the stratospheric SCD $N_{s, \text{strat}}$ from data assimilation using the TM5 (Tracer Model, version 5) chemistry transport model (CTM), (iii) obtaining the tropospheric contribution by subtraction and (iv) calculating the tropospheric air mass factors (AMFs) M_{trop} by converting the SCD to a VCD $N_{v, \text{trop}}$ (see Table 1). The retrieval equation is as follows:

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

More information can be found in the “Product Specification Document for the QA4ECV NO₂ ECV precursor product” (Boersma et al., 2017b) and in Zara et al. (2018) and Boersma et al. (2018). A preliminary evaluation of the data indicated that QA4ECV NO₂ values are 5%–20% lower than the earlier version of the OMI NO₂ data product, DOMINO v2, over polluted regions, and that they agree slightly better with MAX-DOAS NO₂ VCD measurements in Tai’an (China) and De Bilt (the Netherlands) than the DOMINO v2 VCDs (Lorente et al., 2017; Lorente Delgado, 2019).

The data product files contain a comprehensive amount of metadata. For each pixel, the satellite data product pro-

vides a total ex ante uncertainty on the retrieved tropospheric VCD as well as a breakdown of the uncertainty u_{SAT} into an ex ante uncertainty budget with the following uncertainty source components: uncertainty in total SCD u_{SAT, N_s} ; stratospheric SCD $u_{\text{SAT}, N_{s, \text{strat}}}$; and tropospheric AMF $u_{\text{SAT}, M_{\text{trop}}}$, which contains contributions from uncertainties in surface albedo u_{SAT, A_s} , cloud fraction (CF) $u_{\text{SAT}, f_{\text{cl}}}$, cloud pressure $u_{\text{SAT}, p_{\text{cl}}}$ and a priori profile shape u_{SAT, S_a} ; and an albedo-CF cross-term, with $c_{A_s, f_{\text{cl}}}$ representing the error correlation coefficient between both properties (Boersma et al., 2018, Sect. 6).

$$\begin{aligned} u_{\text{SAT}}^2 &= u_{\text{SAT}, N_s}^2 + u_{\text{SAT}, N_{s, \text{strat}}}^2 + u_{\text{SAT}, M_{\text{trop}}}^2 \\ u_{\text{SAT}, M_{\text{trop}}}^2 &= u_{\text{SAT}, A_s}^2 + u_{\text{SAT}, f_{\text{cl}}}^2 + u_{\text{SAT}, p_{\text{cl}}}^2 + u_{\text{SAT}, S_a}^2 \\ &\quad + 2c_{A_s, f_{\text{cl}}} u_{\text{SAT}, A_s} u_{\text{SAT}, f_{\text{cl}}} \end{aligned} \quad (2)$$

Furthermore, the satellite data files provide several relevant instrument parameters, influence quantities (e.g. cloud fraction, surface albedo and terrain height), intermediate quantities (e.g. SCD, AMF and stratospheric SCD) and the column averaging kernel \mathbf{a}_{SAT} , which relates the retrieved VCD to the true profile. The a priori NO₂ profiles (simulated with TM5) are not stored in the data files. If users have to adapt a (measured or modelled) profile \mathbf{x}_h at a high vertical resolution to the vertical sensitivity of the satellite, they can apply Eq. (11) from Eskes and Boersma (2003):

$$\mathbf{a}_{\text{SAT}} \cdot \mathbf{x}_h = \mathbf{x}_{h, \text{sm}}, \quad (3)$$

where the a priori profile $\mathbf{x}_{\text{SAT}, a}$ is not explicit. The dependence of the retrieval on $\mathbf{x}_{\text{SAT}, a}$ is already implicit via the averaging kernel \mathbf{a}_{SAT} .

However, the reference data in the current work are column retrievals or profile retrievals with a limited vertical resolution and are based on an a priori profile that is different from the satellite retrieval. Before smoothing, satellite and reference retrievals should be adjusted such that they use the same a priori profile (Rodgers and Connor, 2003); therefore, knowledge of the satellite a priori profile is relevant. These can be derived from the TM5-MP data files (Huijnen et al., 2010; Williams et al., 2017), which are available upon request (see Boersma et al., 2017b, for contact details), by spatially interpolating the profiles to the location of the satellite ground pixel.

In this work, we considered data from 2004 up to and including 2016 for the tropospheric VCD and up to and including 2017 for the stratospheric VCD.

2.1.2 OMI STREAM stratospheric NO₂

The Stratospheric Estimation Algorithm From Mainz (STREAM; Beirle et al., 2016) was included as an alternative stratospheric estimation scheme in the QA4ECV NO₂ data files. In STREAM, the estimate of stratospheric columns is based on satellite observations with a negligible tropospheric

contribution, i.e. generally over regions with low tropospheric NO₂ levels, and for satellite pixels with high clouds, where the tropospheric column is shielded. The stratospheric field is then smoothed and interpolated globally, assuming that the spatial pattern of stratospheric NO₂ does not feature strong gradients.

2.1.3 NASA OMNO2 data product

Although not the main focus of this work, we do include NASA's OMI NO₂ data – OMNO2 version 3.1 (Bucsela et al., 2016; Krotkov et al., 2017) – as a benchmark comparison of an alternative retrieval product with QA4ECV MAX-DOAS. Like QA4ECV OMI NO₂, OMNO2 is also based on the DOAS approach, although nearly all retrieval steps are different between the QA4ECV and NASA OMI NO₂ algorithms (Table 1). A detailed comparison of the QA4ECV and NASA fitting approaches showed small differences between NO₂ SCDs (Zara et al., 2018); thus, differences between the spectral fitting approaches only explain a small part of the differences in the tropospheric VCDs. The stratospheric correction approach differs between the two algorithms. Although the QA4ECV and NASA stratospheric SCDs have not been compared directly, previous evaluations suggest that differences between the approaches typically lead to small but spatially widespread differences of up to $0.5\text{--}1.0 \times 10^{15}$ molec.cm⁻² in tropospheric VCDs. This leaves differences between the tropospheric AMF calculations (and especially the prior information used in their calculations) as the most likely explanation for the lower NASA values compared with QA4ECV NO₂ VCDs (e.g. Goldberg et al., 2017).

2.2 Ground-based data

2.2.1 Zenith-scattered light DOAS

The ZSL-DOAS data are part of the Network for the Detection of Atmospheric Composition Change (NDACC; De Mazière et al., 2018, see also <http://www.ndaccdemo.org/>, last access: 22 April 2020), which is a major contributor to the WMO's Global Atmospheric Watch. A significant part of the multi-decadal ZSL-DOAS data is provided by the Système d'Analyse par Observation Zénithale (SAOZ; see Pomereau and Goutail, 1988) subnetwork from the IPSL Atmospheres Laboratory (LATMOS), using SAOZ instrumentation in automated data acquisition mode and with fast data delivery.

Zenith-sky measurements are performed during twilight at sunrise and sunset. Due to this measurement geometry with a long optical path in the stratosphere, the measured column is about 14 times more sensitive to stratospheric NO₂ than to tropospheric NO₂ (Solomon et al., 1987). Moreover, it allows for usable measurements to also be made during cloudy conditions. Processing followed the NDACC standard

operation procedure (http://ndacc-uvvis-wg.aeronomie.be/tools/NDACC_UVVIS-WG_NO2settings_v4.pdf, last access: 22 April 2020), as implemented, for instance, in the LATMOS_v3 SAOZ processing. From slant column inter-comparisons, Vandaele et al. (2005) deduce an uncertainty of about 4 %–7 %, but this excludes the uncertainty on the AMF required to convert the slant to vertical columns. Ionov et al. (2008) estimate a total uncertainty on the vertical columns of 21 %, but this is probably an overestimation for the most recent processing, as Bognar et al. (2019) now suggest a 13 % total uncertainty. A visualization of the geographical distribution of the instruments is provided in Fig. 1. More details about the particular co-location scheme, considering the large horizontal smoothing of these measurements and the photochemical adjustment required to convert twilight measurements to satellite overpass times, are provided in Sect. 3.1.

2.2.2 Multi-axis DOAS

The tropospheric NO₂ VCD data used as a reference are a long-term record of MAX-DOAS (multi-axis DOAS) measurements from 10 instruments, reprocessed by different teams for the QA4ECV project (see Table 2). MAX-DOAS instruments measure scattered sunlight under different viewing elevations from the horizon to the zenith (Platt and Stutz, 2008). The observed light travels a long path (the length is dependent on the elevation angle) in the lower troposphere, while the stratospheric contribution is removed by a reference zenith measurement. Two different MAX-DOAS data processing methods were used for the current validation study, QA4ECV MAX-DOAS and bePRO (Belgian Profiling) MAX-DOAS (Clémer et al., 2010), with the latter being part of NDACC.

Thanks to an extensive harmonization effort within the QA4ECV project, reference QA4ECV MAX-DOAS data sets were produced by the different teams for all 10 instruments. These data sets are available at http://uv-vis.aeronomie.be/groundbased/QA4ECV_MAXDOAS/index.php (last access: 22 April 2020). This effort was based on a four-step approach (see http://uv-vis.aeronomie.be/groundbased/QA4ECV_MAXDOAS/QA4ECV_MAXDOAS_readme_website.pdf, last access: 22 April 2020; Hendrick et al., 2016; Richter et al., 2016; Peters et al., 2017), including (i) the establishment of recommendations for DOAS analysis settings from an intercomparison of NO₂ slant column densities retrieved from common spectra, (ii) the development of NO₂ AMF look-up tables (LUTs) to harmonize the conversion of SCDs into VCDs, (iii) the establishment of a first harmonized error budget and (iv) the generation of MAX-DOAS data files in the Generic Earth Observation Metadata Standard (GEOMS) as a common format. It is worth noting that as only SCDs measured at a relatively high elevation angle (typically 30°) are used to minimize the impact of aerosols

Table 1. The OMI satellite data products considered in this work.

Data product	Spectral fitting	Stratospheric correction	Tropospheric AMF
OMI QA4ECV v1.1	Zara et al. (2018)	Data assimilation in TM5-MP (Boersma et al., 2018)	Surface albedo from Kleipool et al. (2008) 5-year climatology at $0.5^\circ \times 0.5^\circ$; clouds from OMI O ₂ –O ₂ algorithm (OMCLDO2 data product, Veefkind et al., 2016); a priori NO ₂ profiles from daily TM5-MP at $1^\circ \times 1^\circ$
OMI STREAM ^a		Weighted (observations with a negligible tropospheric contribution – clean regions and cloudy pixels) convolution (Beirle et al., 2016)	
OMNO2 v3.1	Marchenko et al. (2015)	Three-step (interpolation, filtering and smoothing) stratospheric field reconstruction to fill in the tropospheric contaminated scenes (Bucsela et al., 2013)	Surface albedo from Kleipool et al. (2008) 5-year climatology at $0.5^\circ \times 0.5^\circ$; clouds from OMI O ₂ –O ₂ algorithm (OMCLDO2 data product), a priori profiles from monthly Global Modelling Initiative data at $1^\circ \times 1.25^\circ$ (Strahan et al., 2013)

^a OMI STREAM stratospheric VCD is contained in the OMI QA4ECV v1.1 data files.

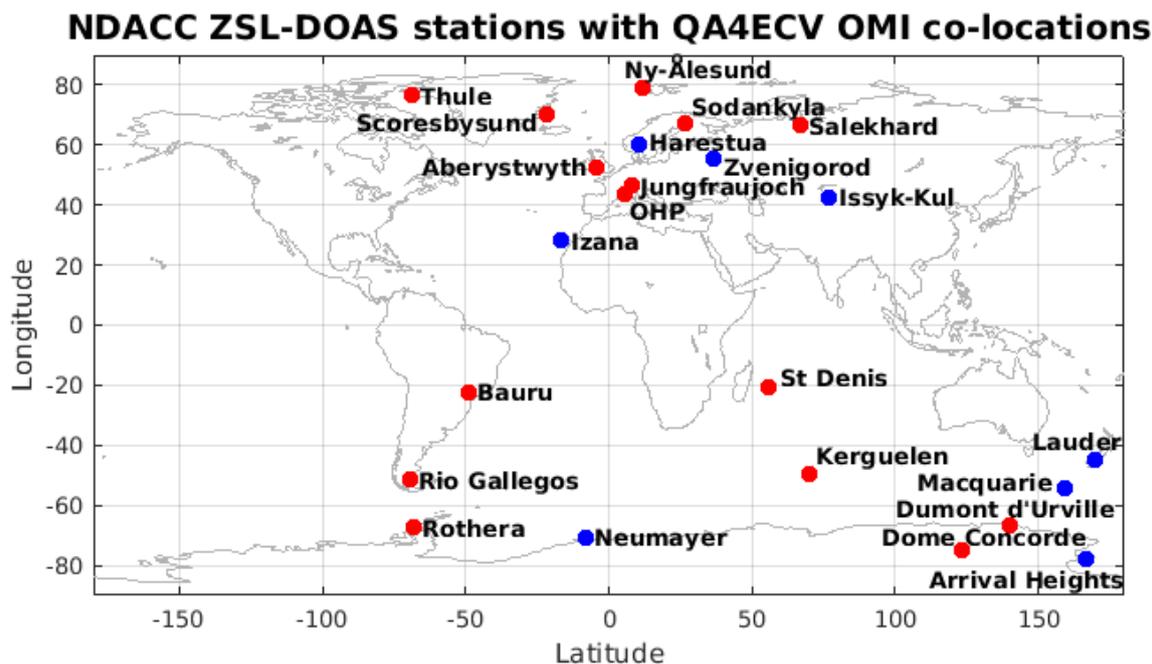


Figure 1. Global distribution of the ZSL-DOAS instruments used in this study. Red markers indicate SAOZ instruments, and blue markers indicate other NDACC ZSL-DOAS instruments.

and a priori profile shape on the retrieval in this QA4ECV approach, the horizontal location of the centre of the effectively probed air mass is close to the instrument location (typically within 1 km). The NO₂ AMF LUTs are produced using the bePRO/LIDORT radiative transfer suite (Clémer et al., 2010; Spurr, 2008). This tool uses the following input (among others): a set of NO₂ vertical profile shapes,

vertical averaging kernel LUTs, geometry parameters (e.g. solar angles and viewing angles) and aerosol optical density (AOD) vertical profile shapes. Column averaging kernel LUTs were calculated based on the Eskes and Boersma (2003) approach, using the bePRO/LIDORT radiative transfer model initialized with similar parameter values to those used for the calculation of the AMF LUTs. Interpolated

AMFs as well as the corresponding vertical profile shapes and column averaging kernels are generated by the tool. More detail is provided in Hendrick et al. (2016).

The second processing method, bePRO MAX-DOAS (Cl mer et al., 2010; Hendrick et al., 2014; Vlemmix et al., 2015), is available for three BIRA-IASB instruments (at Bujumbura, Uccle and Xianghe). This approach, which is based on the optimal estimation method (OEM; see Rodgers, 2000), provides profile measurements, albeit with a limited degree of freedom for signal in the vertical dimension, which is typically ~ 2 (Bujumbura and Uccle) or ~ 3 (Xianghe). The horizontal extension of the air masses probed by profile retrieval MAX-DOAS is about 5–15 km from the instrument in the viewing direction (Richter et al., 2013a). The extension depends on the atmospheric visibility (lower extension for lower visibility) and the altitude of the NO₂ layer (lower extension with decreasing profile height). This is in line with the typical distances estimated by studies such as Irie et al. (2011, their Fig. 17). The horizontally projected area of the air mass probed by the MAX-DOAS is estimated to be of the order of 0.01 to 0.2 km² for QA4ECV MAX-DOAS and ~ 1 km² for bePRO MAX-DOAS, assuming a 1° field of view and a simple geometrical approximation.

There is a clear distinction between the QA4ECV MAX-DOAS and bePRO retrieval algorithms. In the QA4ECV MAX-DOAS algorithm, the VCD is obtained by dividing a differential SCD by a differential AMF at a single elevation angle (see Sect. 1.3 of Hendrick et al., 2016). In the bePRO approach (Cl mer et al., 2010; Hendrick et al., 2014; Vlemmix et al., 2015), a VCD is obtained by integrating a vertical NO₂ profile retrieved by an optimal estimation method using measurements at several elevation angles.

MAX-DOAS instruments probe the lower troposphere, with the highest sensitivity (described by the column averaging kernel) close to the surface, typically in the lowest 1.5 km of the atmosphere. Nevertheless, the vertical grid extends to ~ 10 km for QA4ECV MAX-DOAS and ~ 3 km for bePRO MAX-DOAS.

The MAX-DOAS sites span a wide range of NO₂ levels, from relatively low at Observatoire de Haute-Provence (OHP) and Bujumbura, with a mean tropospheric MAX-DOAS VCD around the OMI overpass time of ~ 3 Pmolec. cm⁻², to strongly polluted at Xianghe, with a mean MAX-DOAS value of ~ 24 Pmolec. cm⁻² (see Fig. 3c, black boxplots), whereas the other sites are moderately polluted (mean value of between 5.6 and 11 Pmolec. cm⁻²).

The MAX-DOAS tropospheric VCD is provided with an ex ante uncertainty in the GEOMS data files. Unfortunately the uncertainty estimation approach employed is not harmonized among all data providers. Therefore, we set the total uncertainty at 22.2 % of the retrieved VCD for QA4ECV MAX-DOAS instead, following the QA4ECV deliverable D3.9 recommendation (Richter et al., 2016). Using sensitivity tests, aerosol effects (20 %) and the NO₂ a priori profile shape (8 %) were identified as the main contributors to the

MAX-DOAS uncertainty, whereas the uncorrelated instrument noise was only 2 %. However, we did not follow D3.9 (Richter et al., 2016) in its recommended division of the uncertainty into the random error and systematic error contributions² and consider only a total uncertainty. Regarding bePRO MAX-DOAS, we consider a 12 % total uncertainty for Uccle and Xianghe (following Hendrick et al., 2014), and a 21 % total uncertainty for Bujumbura (following Gielen et al., 2017). We finally note that an absolute scale uncertainty estimate might be more appropriate for clean sites.

We note that the bePRO profile retrieval algorithm has recently been compared to several other retrieval algorithms (Frie  et al., 2019; Tirpitz et al., 2020). In future validation work, the consideration of other retrieval algorithms that performed well in the intercomparison exercises of Frie  et al. (2019) and Tirpitz et al. (2020) would be of high interest.

As the accuracy of satellite or ground-based remote sensing can be affected by the presence of aerosol, tracking aerosol optical depth (AOD) is useful. The bePRO MAX-DOAS provides AOD measurements at the same temporal sampling resolution as the NO₂ measurements. The QA4ECV MAX-DOAS provides an AOD climatology (Hendrick et al., 2016) based on AERONET (Aerosol Robotic Network) data (Giles et al., 2019); however, we found that the precision of this climatological data set was inadequate for the current work, especially for urban sites. Instead, we considered AOD directly from AERONET (Giles et al., 2019; <http://aeronet.gsfc.nasa.gov>, last access: 22 September 2019), whose measurements are based on Cimel Electronique Sun–sky radiometers. Level 2.0 AOD at a wavelength of 440 nm was chosen, which is within the QA4ECV MAX-DOAS retrieval window of 425–490 nm. Note that the AERONET data are already cloud filtered.

A limitation when investigating AOD dependencies in satellite MAX-DOAS comparisons using AERONET AOD with QA4ECV MAX-DOAS tropospheric NO₂ VCD data (compared with using bePRO AOD with bePRO NO₂ data) is that it implies subsetting: for part of the QA4ECV MAX-DOAS NO₂ data, no co-located AERONET AOD data are available. Moreover, as opposed to the bePRO

²In D3.9, the systematic error uncertainty is set at 3 %, arising from absorption cross-section-related systematic error uncertainty on the SCD, whereas the random error uncertainty is set at 22 %, arising from uncertainty on the AMF. However, the assumption that an error in a priori profile shape, for example, would translate to a random error on the retrieved column is not evident in our opinion. In a later analysis (Hendrick et al., 2018), a comparison of QA4ECV MAX-DOAS with more advanced MAX-DOAS profiling methods was performed. This highlighted systematic differences between -12 % and $+7$ %, which are considerably larger than the systematic error uncertainty of 3 % recommended by the D3.9. This suggests that a larger part of the total uncertainty is due to systematic error. Therefore, in this work, we only consider a total uncertainty of 22.2 %, which is derived from the sum of the recommended systematic and random components in quadrature.

Table 2. Overview of contributing sources for the QA4ECV MAX-DOAS reference data set.

Station	Location	Start and end time	Class	Contributor ^a
Bremen (DE)	53.10° N, 8.85° E	Feb 2005–Dec 2016	Urban	IUP-UB
De Bilt (NL) ^c	52.10° N, 5.18° E	Mar 2011–Nov 2017	Suburban	KNMI
Cabauw (NL) ^{c,d}	51.97° N, 4.93° E			
Uccle (BE) ^{b,d}	50.80° N, 4.36° E	Apr 2011–Jun 2015	Urban	BIRA-IASB
Mainz (DE) ^d	49.99° N, 8.23° E	Jun 2013–Dec 2015	Urban	MPG
Observatoire de Haute-Provence (FR) ^d	43.94° N, 5.71° E	Feb 2005–Dec 2016	Rural/background	BIRA-IASB
Thessaloniki (GR) ^d	40.63° N, 22.96° E	Jan 2011–May 2017	Urban	AUTH
Xianghe (CHN) ^{b,d}	39.75° N, 116.96° E	Apr 2010–Jan 2017	Suburban	BIRA-IASB
Athens (GR) ^d	38.05° N, 23.86° E	Sep 2012–Oct 2016	Urban	IUP-UB
Nairobi (KEN)	1.23° S, 36.82° E	Jan 2004–Nov 2014	Rural/urban	IUP-UB
Bujumbura (BU) ^{b,d}	3.38° S, 29.38° E	Jan 2014–Dec 2016	Suburban	BIRA-IASB

^a Contributing teams: the Aristotle University of Thessaloniki (AUTH), the Royal Belgian Institute of Space Aeronomy (BIRA-IASB), the Institute of Environmental Physics at the University of Bremen (IUP-UB), the Max Planck Institute (MPG) and the Royal Netherlands Meteorological Institute (KNMI). ^b For this sensor, bePRO MAX-DOAS data (providing profile data) are also available. ^c The same instrument was operated at two different locations, De Bilt and Cabauw, which are approximately 30 km apart. ^d An AERONET instrument, measuring aerosol optical depth, is located at this site or in close proximity.

NO₂/bePRO AOD combination, co-located QA4ECV MAX-DOAS NO₂/AERONET AOD data pairs have a temporal co-location mismatch and (where instruments are at different locations) a spatial co-location mismatch. A test was performed (results not shown) using the bePRO NO₂/AERONET AOD combination, and it was generally found that the results are less clear than for the bePRO NO₂/bePRO AOD combination.

3 Validation

3.1 Validation methodology

The generic validation protocol is similar to that outlined by Keppens et al. (2015), and it is tailored within the QA4ECV project for the ECVs NO₂, HCHO and CO (Compernelle et al., 2018). Terms and definitions applicable to the quality assurance of ECV data products have been agreed upon within QA4ECV (Compernelle et al., 2018); the full set can be found in Compernelle and Lambert (2017). The discussion and analysis of comparison error follows the terminology and framework detailed within the GAIA-CLIM project (Verhoelst et al., 2015; Verhoelst and Lambert, 2016).

In the following sections, we detail the baseline validation methodology.

3.1.1 Screening criteria

Filters are applied to the satellite data product following the recommendations in the QA4ECV NO₂ product specification document (PSD; Boersma et al., 2017b) as well as to minimize comparison error with MAX-DOAS.

Following the QA4ECV NO₂ PSD (Boersma et al., 2017b), satellite data are kept for tropospheric NO₂ validation if the following conditions are met:

- (1) no raised error flag;
- (2) a satellite solar zenith angle (SZA) less than 80°;
- (3) the so-called “snow–ice flag” indicates “snow-free land”, “ice-free ocean” or a sea ice coverage below 10 %;
- (4) the ratio of the tropospheric AMF over the geometric AMF, $\frac{M_{\text{trop}}}{M_{\text{geo}}}$, must be higher than 0.2 in order to avoid scenes with a very low tropospheric AMF (which typically occur when the TM5 model predicts a large amount of NO₂ close to the surface in combination with aerosols or clouds, effectively screening this NO₂ from detection);
- (5) an effective cloud fraction (CF) less than 0.2. This last filter is comparable to the PSD recommendation of a cloud radiance fraction (CRF) below 0.5, and it was chosen because the effective cloud fraction is a more general property than the CRF. Note that the satellite-retrieved cloud fraction and cloud height are effective properties that are sensitive to both aerosol and cloud (Boersma et al., 2004). It should be mentioned that cloudy pixel retrievals – although subject to larger errors than clear-sky pixels – can still be used (e.g. in data assimilation), provided that the averaging kernel is taken into account (Schaub et al., 2006).
- (6) This condition is not mentioned in the PSD, but it is applied by Boersma et al. (2018) and is a filter to limit the impact of aerosol haze and low clouds. In Boersma et al. (2018), this was accomplished by excluding ground pixels with a high retrieved cloud pressure, i.e. $p_c > 850$ hPa. Unfortunately, this filter can remove a substantial portion of the data; therefore, a less strict filter was required in the current work. Low cloud can lead to a

high uncertainty in the retrieved tropospheric NO₂ value when it is uncertain if the cloud is located above the trace gas (a high screening effect and, therefore, a low AMF) or is at similar height (a partial screening effect and partial surface albedo effect and, therefore, a higher AMF). This is registered in the uncertainty component due to the cloud pressure u_{SAT,p_c} available within the data product. Data analysis reveals that a relatively small number of ground pixels are responsible for an important contribution to the root mean square (RMS) of the ex ante satellite uncertainty for several sites (Xi-anqhe, Uccle, De Bilt, Bremen and Athens) via the cloud pressure component u_{SAT,p_c} . Most of these high-uncertainty ground pixels have a low retrieved effective cloud pressure (Fig. S1 in the Supplement), which is indicative of aerosol haze or low-lying cloud. The aforementioned cloud pressure filter used by Boersma et al. (2018) would effectively remove these suspicious ground pixels but many other pixels with a low u_{SAT,p_c} as well. Therefore, we chose to apply filter (6) instead, which is a one-sided sigma-clipping on u_{SAT,p_c} : data where $u_{SAT,p_c,i} > \text{mean}(u_{SAT,p_c,i}) + 3 \times \text{SD}(u_{SAT,p_c,i})$ are removed. This sigma-clipping removes a smaller percentage of the data, while still achieving its goal of limiting u_{SAT,p_c} and u_{SAT} . After this filtering step, u_{SAT,p_c} is only a minor contributor to the OMI uncertainty budget.

- (7) Finally, satellite ground pixels with a footprint greater than 950 km², corresponding to the five outermost rows at each swath edge of the OMI orbit, are removed in order to limit the horizontal smoothing difference error with the MAX-DOAS data. Filter (7) is not a filter on satellite data quality but rather a limit on the scope of the validation.

Regarding stratospheric NO₂ validation, only filters (1)–(3) are applied. Hence, both cloudy and non-cloudy scenes are used.

Regarding the OMNO2 data product, we followed the recommendation of Bucsele et al. (2016) by only including ground pixels for which the least significant bit of the VcdQualityFlags variable is zero (indicating good data). Furthermore, the effective cloud fraction must be less than 0.2 and the pixel area must be less than 950 km².

No screening was applied to the ground-based reference data sets. In particular, filtering is not applied on the MAX-DOAS cloud flag as a baseline, as it is not available for all data sets. It should be noted that clouds can impact the quality of MAX-DOAS retrievals (see e.g. radiative transfer simulations of Ma et al., 2013, and Jin et al., 2016).

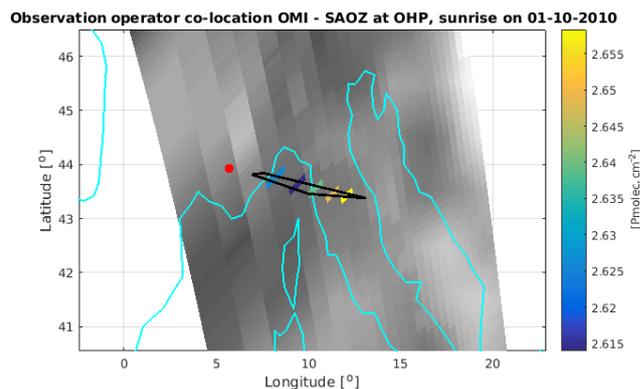


Figure 2. Illustration of a single co-location between OMI and a sunrise ZSL-DOAS measurement using the dedicated observation operator. The red dot marks the location of the ground instrument, the cyan lines indicate the coastlines of this part of the Mediterranean, the greyscale background contains the full orbit data and the coloured pixels are those that have their centre within the observation operator (black polygon), i.e. those that are averaged to obtain a satellite measurement comparable with that of the ZSL-DOAS instrument.

3.1.2 Co-location criteria and processing

Stratospheric column

The air mass to which a ZSL-DOAS measurement is sensitive spans over many hundreds of kilometres towards the rising or setting Sun (e.g. Solomon et al., 1987). The co-location scheme employed here takes this into account by averaging all OMI ground pixels of a temporally co-located orbit (with a maximum allowed time difference of 12 h) that have their centre within the ZSL-DOAS observation operator. This observation operator is a 2-D polygon that results from the parametrization of the actual extent of the air mass to which the ZSL-DOAS measurement is sensitive. Its horizontal dimensions were derived using the UVSPEC/DISORT ray tracing code (Mayer and Kylling, 2005), mapping the 90 % interpercentile range of the stratospheric vertical column to a projection on the ground, and then parameterizing it as a function of the solar zenith and azimuth angles during the twilight measurement, where the SZA during a nominal single measurement sequence is assumed to range from 87 to 91° (at the location of the station). Note that the station location is not part of the area of actual measurement sensitivity. The average OMI stratospheric column over this observation operator can then be compared to the column measured by the ZSL-DOAS instrument. An illustration of a single such co-location is presented in Fig. 2. Note that the above-mentioned SZA range may not be covered entirely at polar sites. For more details, we refer the reader to Lambert et al. (1996) and Verhoelst et al. (2015).

To account for effects of the photochemical diurnal cycle of stratospheric NO₂, the ZSL-DOAS measurements, ob-

tained twice daily at twilight at each station, are adjusted to the OMI overpass time using a model-based factor. The latter is extracted from LUTs that are calculated with the PSCBOX 1-D stacked-box photochemical model (Errera and Fonteyn, 2001; Hendrick et al., 2004) initiated by daily atmospheric composition and meteorological fields from the SLIMCAT chemistry transport model (Chipperfield, 1999). The amplitude of the adjustment depends strongly on the effective SZA assigned to the ZSL-DOAS measurements; it is taken here to be 89.5°. The uncertainty related to this adjustment is of the order of 10 % or 1 to 2 10¹⁴ molec. cm⁻².

Tropospheric column

Regarding the tropospheric column validation, satellite data are kept if the satellite ground pixel covers the MAX-DOAS instrument location and if a MAX-DOAS measurement is within a 1 h interval centred on the satellite measurement time. The average of all MAX-DOAS measurements within this 1 h interval is taken. The typical number of MAX-DOAS measurements taken within this time interval was between two and four for most sites. This procedure was applied to both QA4ECV OMI NO₂ and the OMNO2 comparisons.

3.2 Impact of quality screening

Quality screening is a necessary step before a satellite data product can be used, but it can be a limit to the data product's scope. Figure 3a presents the remaining fractions of the satellite overpass data at the MAX-DOAS sites at each of the seven successive filter steps described in Sect. 3.1.1. Note that the Cabauw and De Bilt sites are not included, as the results are very close to that of Uccle.

With respect to the data filtering conditions mentioned earlier, the error flag (1) removes ~ 10 %–30 % of the data; filters on the SZA and the snow–ice flag (2 and 3) have a relatively small impact; the filter on the AMF ratio (4) has a large impact on the Bremen, Mainz, Cabauw, De Bilt, Uccle and Xianghe sites (35 %–40 % of data removed); and the filter on CF (5) has an important screening impact at all sites (see Fig. 3), removing up to 60 % of the data at the Bujumbura site. As an alternative to the CF filter, we also tested the CRF < 0.5 filter; for most sites the CRF and CF filters have a near-identical impact, although for Bujumbura and Nairobi the CRF filter is more restrictive (results not shown). In combination, the quality filters recommended by the PSD (filters 1–5) remove between 56 % (Athens) and 90 % (Bremen) of the data.

Filter (6), the filter on the uncertainty component due to cloud pressure u_{SAT, p_c} , removes 5 % of the data at most at the Xianghe site, whereas the alternative filter on cloud pressure would have removed 15 % of the data (Fig. S1). The filter on ground pixel size (7) removes an additional 3 %–16 % of the data.

The screening can have a strong seasonal effect; for example, the winter months are strongly underrepresented for the western European urban sites (Fig. 3b). Figure 3c presents box plots of co-located MAX-DOAS tropospheric NO₂ measurements for each MAX-DOAS site before (black) and after (blue) screening. Both the mean and median values decrease due to the filtering step. We conclude that the quality screening tends to reject scenes with a high tropospheric NO₂ VCD, i.e. the restriction to quality-screened scenes leads to a negative sampling bias with respect to the ensemble of all scenes. On an absolute scale, the screening effect is strongest at the Xianghe site, leading to a decrease in the yearly mean tropospheric NO₂ from 24 to 15 Pmolec. cm⁻² (40 % decrease). At Nairobi, Thessaloniki, Bremen, De Bilt and Cabauw, the tropospheric VCD is reduced by several peta molecules per square centimetre (Pmolec. cm⁻²). The cloud filter is a main contributor to this sampling bias. This is in accordance with the results of Ma et al. (2013), who found that higher tropospheric NO₂ was measured by MAX-DOAS in Beijing under cloudy conditions compared with clear-sky conditions. Indeed, cloudy conditions lead to less photochemical loss of tropospheric NO₂, as explained by the model results (Boersma et al., 2016). In comparisons of OMI tropospheric NO₂ with independent data, care should be taken to ensure that the independent data are also sampled for clear-sky conditions (Boersma et al., 2016). A systematic influence of clouds on the MAX-DOAS retrievals might contribute to the observed sampling bias effect.

It can be argued that the AMF ratio filter (filter 4) is too restrictive. In Sect. S2 in the Supplement results are presented for the less restrictive $\frac{\text{AMF}_{\text{trop}}}{\text{AMF}_{\text{geo}}} \geq 0.05$. The remaining data fraction is slightly increased at the Bremen, Mainz, Uccle, De Bilt and Cabauw sites (from ~ 8 % to ~ 10 %), and the winter months are better represented (see Fig. S2). The negative sampling bias at De Bilt and Bremen is reduced, whereas it is removed at Mainz. As will be shown in Sect. 3.4.6, this adapted filtering generally has no negative impact on the satellite vs. MAX-DOAS comparisons.

3.3 Comparison of OMI stratospheric NO₂ with ZSL-DOAS

Figure 4 contains time series of stratospheric NO₂ columns, from both satellite (QA4ECV product) and ground-based instruments, at two illustrative ground sites: Kerguelen in the southern Indian Ocean, which is representative of very clean background conditions, and the Observatoire de Haute-Provence in France, which is affected by significant tropospheric pollution in local winter that often exceeds the wintertime stratospheric column. The graphs show the well-known seasonal cycle in stratospheric NO₂, which is captured similarly by satellite measurements and the ZSL-DOAS instrument. It is already evident from perusal of the results at OHP that the stratospheric comparison is hardly affected by the peaks in tropospheric pollution, e.g. in winter

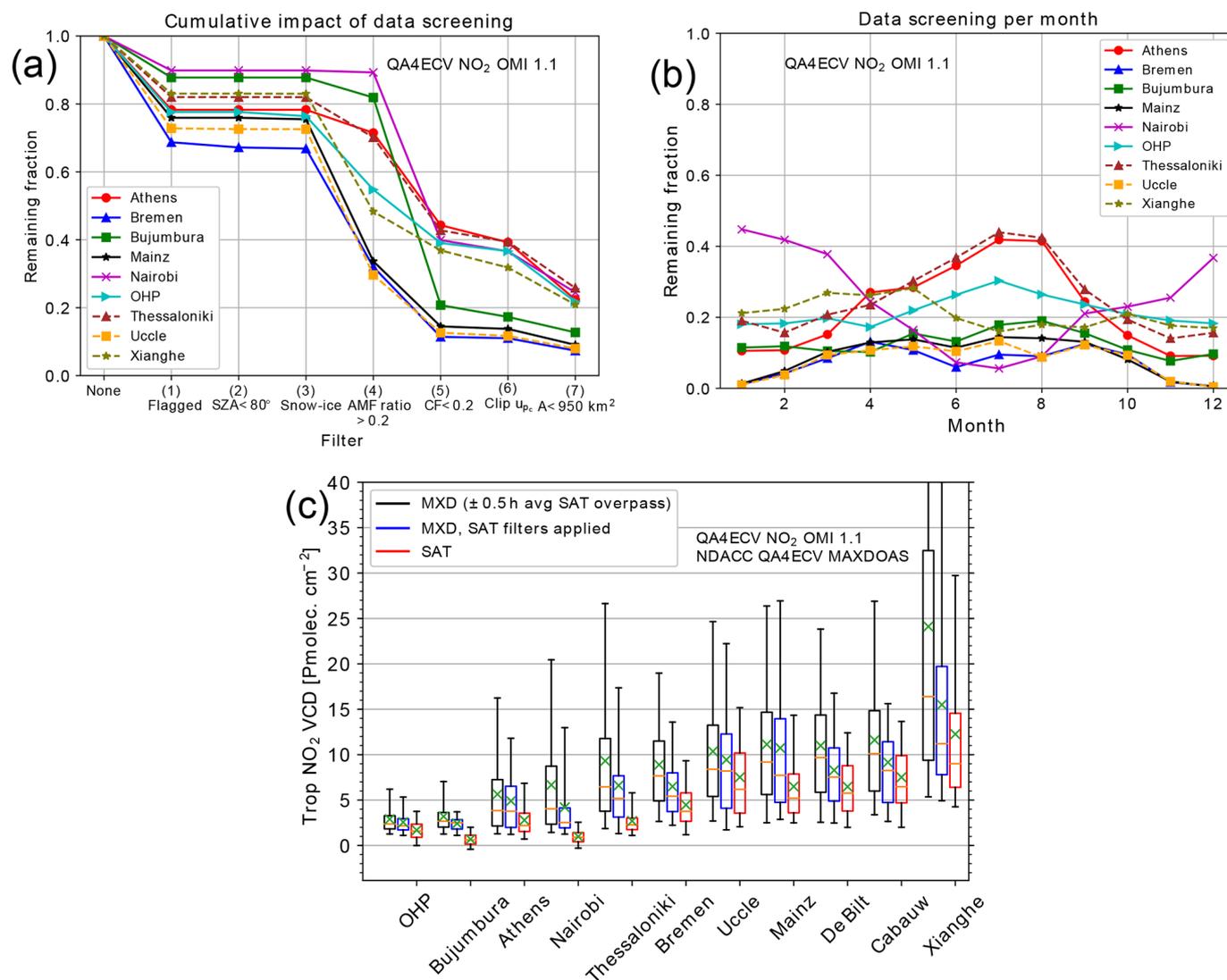


Figure 3. Starting from satellite data with a ground pixel covering the MAX-DOAS site, panel (a) shows the remaining data fraction after the application of each of the seven filter criteria. The criteria are explained in Sect. 3.1.1. The Cabauw and De Bilt sites are not included here, as the fractions are very close to that of Uccle. Panel (b) shows the remaining fraction per month after the application of all filters. Panel (c) shows boxplots of QA4ECV MAX-DOAS data (“MXD”) co-located with QA4ECV OMI for each site, before the application of the filters (black), after the application of the filters (blue) and QA4ECV OMI co-located with MAX-DOAS after the application of the filters (“SAT”, red). The sites are sorted according to the median MAX-DOAS value before filtering. The box edges represent the 1st and 3th quartiles, the orange line represents the median, the green cross represents the mean, and the whiskers represent the 5th and 95th percentiles.

2005–2006, indicating a good separation between the troposphere and stratosphere in the QA4ECV OMI retrievals.

To better reveal differences in the representation of the seasonal cycle, Fig. 5 presents the full time series at these two stations as a function of the day of the year (DoY), with a 1-month moving mean applied. While the seasonal cycle is generally well represented, with accurate levels in local summer, the QA4ECV OMI stratospheric NO₂ column does appear to be a little lower than the ground-based value in local winter at these two sites. However, this is not a network-wide feature; this is illustrated in Fig. 6, which shows the median

difference for each day of the year for every station, ordered by latitude, where the median is taken over the entire 14-year time series.

From this figure, it is clear that the agreement is poorer at high latitudes, owing to more difficult measurement conditions (such as a high SZA) and at times a highly variable atmosphere (e.g. vortex dynamics), which amplify errors due to imperfect co-location. At more moderate latitudes, some seasonal features can be observed, but their sign varies from station to station, e.g. for Lauder and Kerguelen. A potential source of seasonal errors lies in the use of NO₂ cross

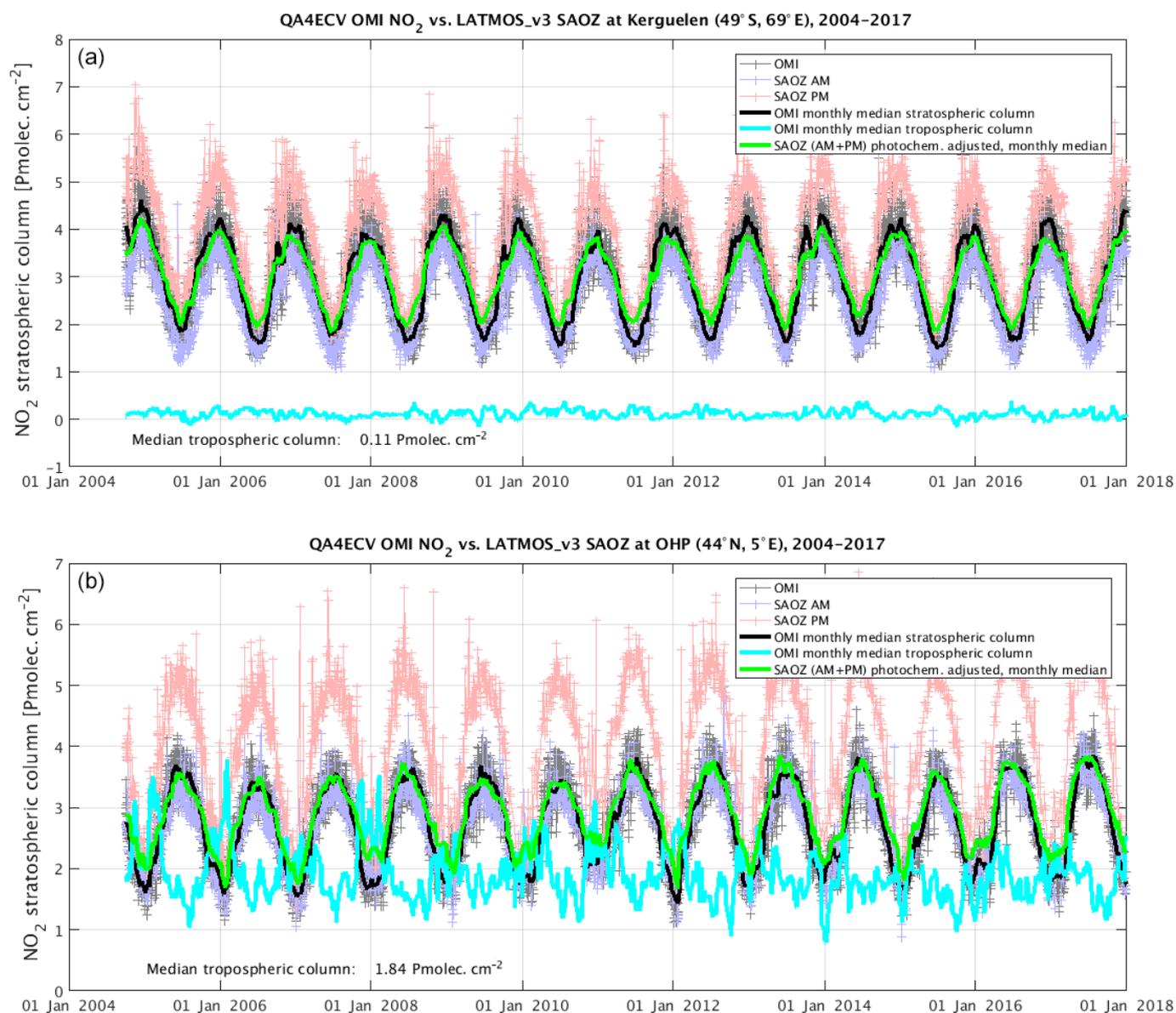


Figure 4. (a) Time series of OMI and SAOZ stratospheric NO₂ above the Kerguelen NDACC station in the Indian Ocean, which is typical of clean background conditions. Panel (b) is similar to (a) but for the Observatoire de Haute-Provence (OHP), which shows more significant tropospheric columns in winter due to anthropogenic pollution.

sections at a fixed temperature. The QA4ECV NO₂ retrieval includes a second-order a posteriori temperature correction to adjust for the difference in the absorption cross section between the assumed 220 K and the true effective temperature (Zara et al., 2017). However, the ZSL-DOAS data were not temperature corrected, and Hendrick et al. (2012) estimate the impact to range between a 2.4 % overestimation in local winter and a 3.6 % underestimation in local summer for ZSL-DOAS measurements at Jungfraujoch. In other words, the amplitude of the seasonal cycle should be about 6 % larger than that currently reported by the ZSL-DOAS at mid-latitudes for an assumed effective stratospheric tem-

perature of 220 K. Therefore, this effect could explain part of the discrepancy between satellite and ground-based seasonal cycles at sites such as Kerguelen, but it requires confirmation with a proper ZSL-DOAS temperature correction. Developmental work on this is ongoing (François Hendrick, personal communication, 2019), but it is beyond the scope of the current paper. The excellent agreement between sunrise and sunset ZSL-DOAS measurements after mapping them to the OMI overpass time at Kerguelen suggests that the photochemical adjustment works well, but it does not exclude the presence of biases that are common to sunrise and sunset measurements. At OHP, the wintertime agreement between

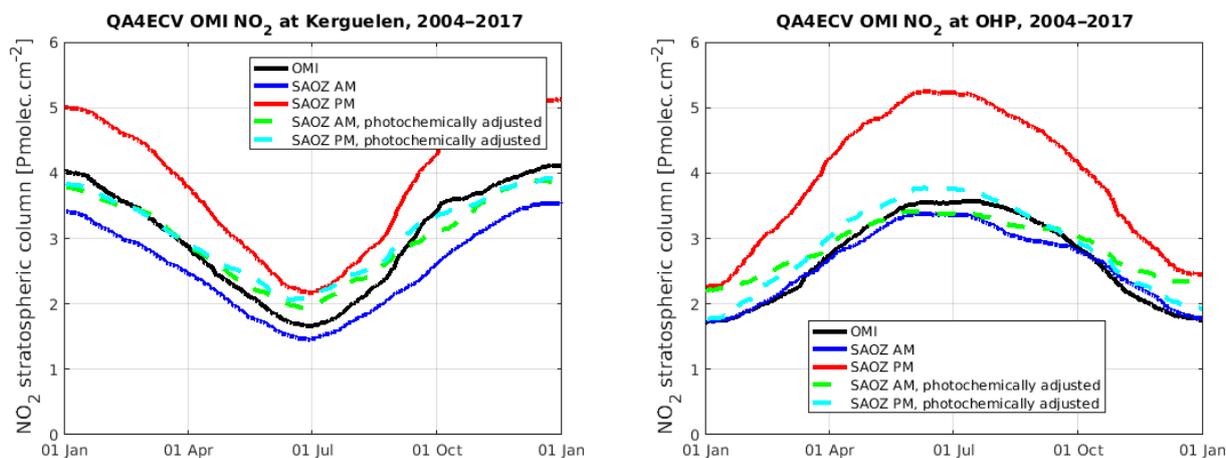


Figure 5. Climatological, i.e. all years mapped to a single year and with a 1-month smoothing function applied, comparison between QA4ECV OMI stratospheric NO₂ and the ZSL-DAOS instruments at Kerguelen and the Observatoire de Haute-Provence (OHP), revealing overall good agreement.

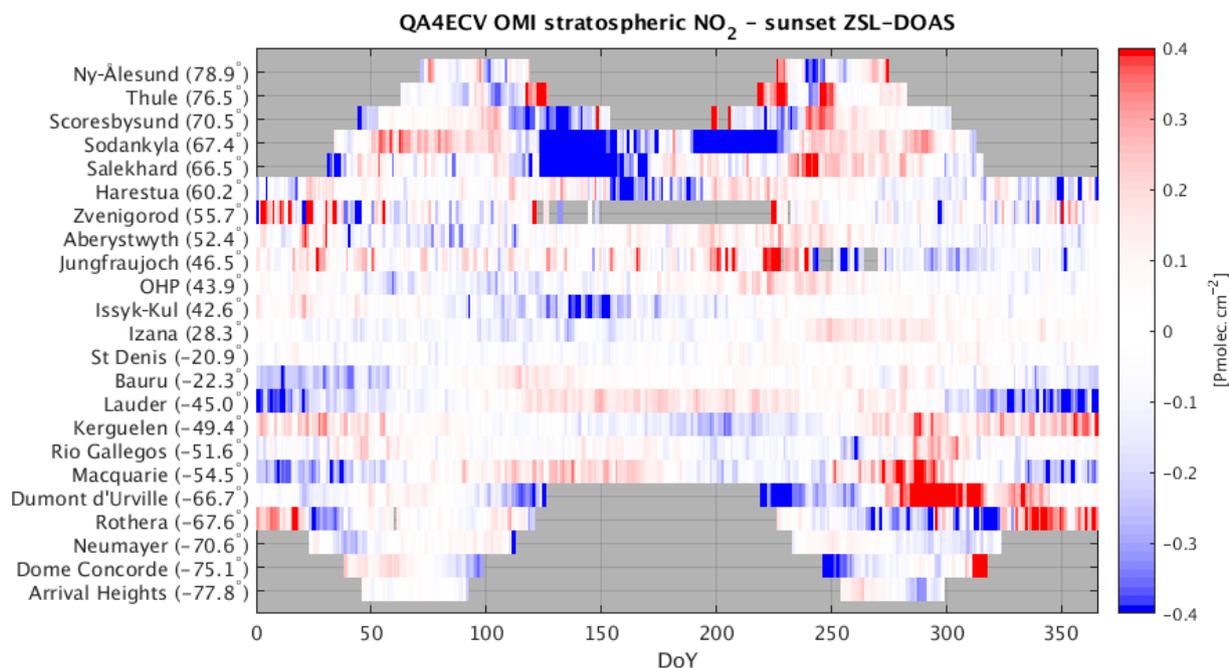


Figure 6. Median difference for each station (ordered by latitude) and for each day of the year, taken over the entire 14-year record, between QA4ECV OMI stratospheric NO₂ and the co-located, photochemically adjusted, sunset ZSL-DOAS measurements.

sunrise and sunset after photochemical adjustment is not as good. Contamination by tropospheric pollution is expected to be similar for both sunrise and sunset measurements, as it contributes to the air mass below the scattering altitude, i.e. the column above the station, as opposed to the large and offset area of sensitivity in the stratosphere. Differences between sunrise and sunset contamination could still be caused by a diurnal cycle in the tropospheric column, but an analysis of that diurnal cycle (e.g. from MAX-DOAS data) is beyond the scope of this work.

Figure 7 presents the network-wide results in terms of the bias and comparison spread for each station as a function of latitude. On average, QA4ECV OMI stratospheric NO₂ seems to have a minor negative bias ($-0.2 \text{ Pmolec. cm}^{-2}$) with respect to the ground-based network. In view of the station-to-station scatter of the order of $0.3 \text{ Pmolec. cm}^{-2}$ and the uncertainties on the ground-based data, this is hardly significant and is roughly in line with validation results for other OMI stratospheric NO₂ data sets (e.g. Celarier et al., 2008; Dirksen et al., 2011). Interestingly, the STREAM

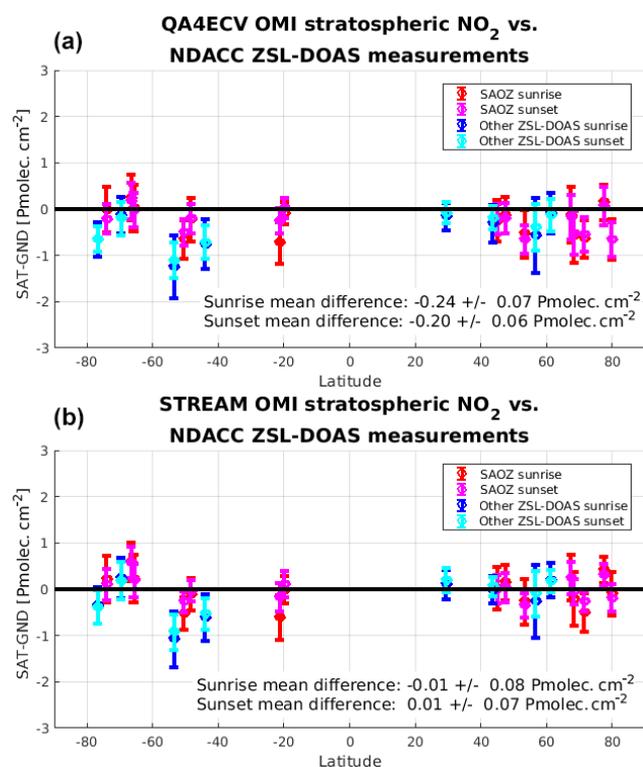


Figure 7. Meridian dependence of the mean (the circular markers) and standard deviation ($\pm 1\sigma$ error bars) of the individual differences between QA4ECV (a) and STREAM (b) OMI stratospheric NO₂ column data and ZSL-DOAS reference data represented at individual stations from the Antarctic to the Arctic. The values in the legend correspond to the mean and standard error of all mean (for each station) differences.

stratospheric NO₂ product, also included in the data files but based on a very different approach (Beirle et al., 2016), does not present this negative bias (see Fig. 7b). This deserves further exploration but is outside the scope of the current paper. The comparison spread at a single station varies from 0.2 to 0.5 Pmolec. cm⁻², corresponding to about 10 % of the stratospheric column. Raw comparisons at Zvenigorod, Russia, yielded a higher comparison spread (1.2 Pmolec. cm⁻²) due to very large pollution events in the Moscow area affecting the ZSL-DOAS measurements; however, for Fig. 7 these were excluded by filtering out co-located pairs with an OMI tropospheric column larger than 3 Pmolec. cm⁻².

3.4 Comparison of OMI tropospheric NO₂ with MAX-DOAS

A key issue in the geophysical validation of satellite data sets with respect to suborbital reference measurements is the additional uncertainty that appears when comparing different perceptions of the inhomogeneous and variable atmosphere, i.e. when comparing data sets characterized by different sampling and smoothing properties, both in space and time,

which is a main topic of the European GAIA-CLIM project (Verhoelst et al., 2015; Verhoelst and Lambert, 2016). Potential comparison error sources for satellite vs. MAX-DOAS are discussed in Sect. 3.4.1–3.4.5, following the framework and terminology of Verhoelst et al. (2015) and Verhoelst and Lambert (2016). The impact of the horizontal smoothing difference error on the bias is presented in a qualitative way in Figs. 8 and S5–S8.

The results of the comparison of QA4ECV OMI with MAX-DOAS are provided in Sect. 3.4.6. The overall bias and dispersion are provided in boxplots of the differences for each site (Fig. 9); comparisons of the NASA OMI data product OMNO2 with MAX-DOAS are also shown. The seasonality of the bias for each site is shown in Figs. 10 and 11. Figure 12 presents the overall discrepancy between QA4ECV OMI and MAX-DOAS as given by the mean-squared deviation (MSD), which is split into bias, seasonally cyclic and residual components. This figure also presents the consistency of the RMSD with the combined ex ante uncertainty. The impact of adapting screening criteria on bias and dispersion is shown in Figs. S9–S13. A priori profile harmonization and vertical smoothing is presented in Fig. 13 for the bePRO sites at Uccle and Xianghe. The discussion of these figures is given point by point in Sect. 3.4.6. Table 3 gives an overview of the error source attributions.

3.4.1 Sources of comparison errors: overview

Part of the discrepancies between the OMI and the MAX-DOAS data sets are due to comparison errors. Starting from the general comparison equation (Verhoelst et al., 2015; Verhoelst and Lambert, 2016), the difference between satellite and reference measurements can be approximated in this specific case as

$$N_{v,\text{trop},\text{SAT}} - N_{v,\text{trop},\text{REF}} = e_{\text{total}} \\ = -e_{\text{REF}} + e_{\text{SAT}} + e_{\text{Sr}} + e_{\Delta r} + e_{\Delta t} + e_{\Delta z}, \quad (4)$$

where $N_{v,\text{trop},\text{SAT}}$ and $N_{v,\text{trop},\text{REF}}$ are the tropospheric VCD values measured by satellite and reference ground-based sensors respectively, e_{SAT} and e_{REF} are the errors in both measurements, e_{Sr} is the horizontal smoothing difference error (as the horizontal projection of the probed air mass of satellite and ground-based measurements is different), and $e_{\Delta r}$, $e_{\Delta t}$ and $e_{\Delta z}$ are the horizontal, temporal and vertical sampling difference error respectively (as satellite and ground-based measurement are not taken at exactly the same space and time).

3.4.2 Temporal sampling difference error

The temporal sampling difference error and the MAX-DOAS uncorrelated random error are already mitigated by averaging the MAX-DOAS measurements within a 1.0 h interval. We found that using larger time intervals can lead to an increase in the bias, which is likely due to the photochemical evolu-

tion and transport of the NO₂ molecule, but at this small time window the temporal sampling difference error has a random character³. The residual uncertainty can be estimated by taking the uncertainty of the mean of the MAX-DOAS values within each time interval. Subtracting this component in quadrature from the RMSD, the $N_{v,trop,SAT} - N_{v,trop,REF}$ discrepancies at the different sites would be reduced by less than 0.1 Pmolec. cm⁻² for the OHP, Bujumbura, Athens and Nairobi sites, and by 0.1 to 0.5 Pmolec. cm⁻² at most for the other sites. Therefore the temporal sampling difference error and the MAX-DOAS uncorrelated random error can be considered to be insignificant contributions to the $N_{v,trop,SAT} - N_{v,trop,REF}$ discrepancies, and they are not discussed further here. In agreement with this, Wang et al. (2017) found that the impact of the temporal sampling difference error on satellite vs. MAX-DOAS tropospheric NO₂ VCD comparisons was negligible.

3.4.3 Horizontal sampling difference error

Tropospheric NO₂ has a large spatial variability, especially at polluted sites; therefore, random and systematic features in the true NO₂ field at the scale of the distance between the MAX-DOAS location and the co-located satellite ground pixel (typically a few kilometres to a few tens of kilometres, ~ 10–14 km on average) can be expected. However, one must realize that (i) there is no directional preference in the co-locations, meaning that directional features are averaged out in the comparison, and (ii) the satellite measurements are strongly spatially smoothed.

To estimate the impact of the horizontal sampling difference error, we compare two sets of QA4ECV OMI NO₂ tropospheric VCDs. Regarding the first set ($N_{v,trop,SAT1}$), it is required that its ground pixel covers the MAX-DOAS site and its pixel centre is within 5 km of the site. The second set ($N_{v,trop,SAT2}$) has its ground pixel second-nearest to the site, within the same overpass. SAT₁ pixels are within 3–4 km of the site on average, SAT₂ pixels are within 11–12 km of the site, and the distance between SAT₁ and SAT₂ pixels is typically 13.6 km, i.e. comparable to the mean distances encountered in the OMI vs. MAX-DOAS comparisons. Note that the discrepancy between $N_{v,trop,SAT1}$ and $N_{v,trop,SAT2}$ is due to both the horizontal sampling difference error and the random noise error.

Details on the analysis are given in Sect. S3 in the Supplement. The main conclusions are as follows:

- The bias caused by the horizontal sampling difference error reaches ~ -0.6 Pmolec. cm⁻² at most (at Athens, Bremen and Mainz) and is therefore only a very minor contributor to the observed bias between OMI and MAX-DOAS (discussed later in Sect. 3.4.6).

³This is checked by comparing MAX-DOAS measurements before and after the satellite overpass time for the different overpasses.

- The dispersion of $N_{v,trop,SAT2} - N_{v,trop,SAT1}$ can, in principle, be due to variation in the total slant column, in the AMF or in the stratospheric slant column (see Eq. 1). It is shown in the Supplement that it is largely due to variation in the slant column. It follows that uncorrelated random noise error mainly originates from the slant column, not from AMF or stratospheric column (as these do not vary much between neighbouring pixels). This then justifies the use of the ex ante uncertainty component due to SCD uncertainty, u_{SAT,N_s} , as an estimate of the total random error uncertainty. Note that u_{SAT,N_s} was scaled such that it only accounts for the random error of the slant column (Zara et al., 2018), not for systematic error.
- At the Bujumbura and Nairobi sites, $u_{SAT1,N_s}^2 + u_{SAT2,N_s}^2$ exceeds the variance of the difference, indicating that u_{SAT,N_s} is sometimes overestimated.
- The standard deviation caused by the horizontal sampling difference error (obtained by subtracting the dispersion due to random noise in quadrature) is minor compared with the discrepancies encountered in the OMI vs. MAX-DOAS comparisons.

3.4.4 Vertical sampling difference error

Two sources of vertical sampling difference error can be identified. First, the surface altitude of the ground-based MAX-DOAS sensor and the mean surface altitude of the OMI ground pixel are not exactly the same. To estimate a correction, we applied a VMR-conserving vertical shift of the satellite a priori profile, described by Zhou et al. (2009). The ground levels are shifted by 0.03 km on average (Cabauw and De Bilt) to 0.4 km (Athens and Bujumbura). This hardly changed $N_{v,trop,SAT} - N_{v,trop,REF}$ (bias changes of 0.3 Pmolec. cm⁻² or less). This VMR-conserving approach probably underestimates the discrepancy at the Athens and Bujumbura sites which have a complicated orography. The MAX-DOAS instrument at Athens is located on one of the hills surrounding the city at an altitude of 527 m, while the mean surface altitude of the co-located satellite pixels is ~ 200 m. Therefore, the MAX-DOAS measurement misses the lowest part of the tropospheric column, and correcting for this would increase the already negative bias. The MAX-DOAS instrument at Bujumbura is at an altitude of 860 m at the edge of the city, which is located in a valley surrounded by 2000–3000 m high mountains (Gielen et al., 2017); this causes the mean surface altitude of the co-located satellite pixels (~ 1.2 km) to be higher than the MAX-DOAS instrument.

A second source of vertical sampling difference error is the fact that the MAX-DOAS only measures the lower tropospheric NO₂ VCD, whereas the satellite measures the full tropospheric VCD. This is, in principle, a source of positive bias in $N_{v,trop,SAT} - N_{v,trop,REF}$ and, therefore, cannot ex-

plain the observed negative bias in the comparison. A proper quantification of this bias source depends critically on the assumed vertical profile shape and is outside the scope of the current work.

3.4.5 Horizontal smoothing difference error

Ideally, subpixel variation in the tropospheric VCD would be estimated using a high-resolution model with grid cell area comparable to the MAX-DOAS horizontally projected area of the probed air mass. Instead, we employ two semi-quantitative approaches here to estimate the bias from horizontal smoothing difference error.

In the first approach, the horizontal smoothing effect is estimated from the QA4ECV OMI NO₂ data. The “superpixel” OMI tropospheric VCDs are constructed by averaging OMI VCDs of individual pixels of a relatively small size ($\leq 500 \text{ km}^2$) within a 20 km radius centred on the MAX-DOAS site. For each overpass, a superpixel VCD is compared with the individual ground pixel VCD covering the MAX-DOAS site. Using this procedure, a superpixel consists of three individual ground pixels on average. The mean difference per season, from 2004 to 2016, is presented in Fig. 8. The second approach is similar, but uses S5P TROPOMI NO₂ data, from May 2018 to May 2019, RPRO (reprocessed) + OFFL (offline) data with processor version 01.02.00–01.02.02, and the superpixel tropospheric VCD is constructed by averaging VCDs of individual pixels that are within a latitude, longitude box of $\Delta\text{lat} = 0.14^\circ$, $\Delta\text{long} = 0.7^\circ$ centred on the MAX-DOAS site for each overpass. TROPOMI has a similar overpass time to OMI (early afternoon) and a considerably finer resolution ($3.5 \times 7 \text{ km}^2$ at nadir). The area of this superpixel corresponds to $\sim 700\text{--}900 \text{ km}^2$, i.e. about the size of a bigger OMI pixel, and typically contains 20 TROPOMI ground pixels.

The OMI-based approach has as the advantage that the time range is appropriate, but it is limited by the large ground pixel size. Regarding the finer resolution TROPOMI data, one should keep in mind that its ground pixel size is still large compared with the horizontally projected area of the probed air mass of the MAX-DOAS⁴; hence, the contribution of the horizontal smoothing difference error to the bias and comparison might still be underestimated. Another limitation is that the TROPOMI time range considered does not overlap with the time range considered for OMI. Both approaches suggest a negative bias contribution due to the horizontal smoothing difference error at the Mainz and Thessaloniki sites and no such bias contribution at OHP, whereas

⁴The horizontal distance of the QA4ECV MAX-DOAS measurements is small compared with a TROPOMI pixel in both the viewing and the perpendicular direction. Regarding bePRO MAX-DOAS, although it has a small field of view, its probed distance in the viewing direction ($\sim 10 \text{ km}$) is of similar or slightly larger magnitude compared with the cross section of a TROPOMI ground pixel.

the results are mixed for other sites (the bias varies over the seasons, and/or different results are found between the OMI- and TROPOMI-based calculations). Differences between the OMI- and TROPOMI-based calculations are likely caused by (i) the much larger central pixel of OMI compared with TROPOMI, which leads to a lower sensitivity to fine-scale variations in Fig. 8a, and (ii) the evolution in parameters such as NO₂ concentration patterns, which are captured differently by the different temporal ranges used in Fig. 8a and b. A case in point is the positive mean differences in JFM (January–February–March) and OND (October–November–December) captured in the TROPOMI-based calculation but not in the OMI-based calculation. Both MAX-DOAS sensors are not located in urban centres, although pollution centres are located nearby. Therefore, the positive mean differences in JFM and OND captured by TROPOMI are likely due to NO₂ fields in the periphery of the TROPOMI superpixel. This is in agreement with the very recent work by Pinardi et al. (2020) on the horizontal smoothing effect. The estimated “horizontal dilution factors” in Fig. S3 of Pinardi et al. (2020) are positive for Cabauw and Xianghe, indicating that NO₂ is higher on the periphery than at the MAX-DOAS location.

A tropospheric NO₂ monthly field with subpixel variability is derived from the QA4ECV OMI NO₂ data using a variant of the temporal averaging approach of Wenig et al. (2008)⁵ (as shown in Figs. S5–S8) for months with a minimal (left column) and maximal (right column) OMI vs. MAX-DOAS bias (as derived from Figs. 10–11). Fields are constructed for each month by averaging over the 2004–2016 period. The resulting field is horizontally smoothed; the variability is an underestimate of the true horizontal NO₂ variability. Subpixel enhanced tropospheric NO₂ approximately centred on the MAX-DOAS site can be identified in high-bias months at Nairobi, Thessaloniki and Mainz, whereas this is clearly not the case for the OHP, Bujumbura, Uccle, De Bilt/Cabauw and Xianghe sites. In Athens the pollution peak centre is some 10 km from the sensor, and for Bremen no clear peak is identified.

The contribution of the horizontal smoothing difference error to the (OMI–MAX-DOAS) bias at Mainz is consistent with the results of Drosoglou et al. (2017), who achieved a significant bias reduction by adjusting the OMI data with factors derived from air quality simulations at a high spatial resolution of 2 km.

Similar maps have been constructed in studies such as Ma et al. (2013) and Chen et al. (2009) to estimate the impact of the horizontal smoothing effect on satellite vs. DOAS comparisons.

⁵Here, the arithmetic average of covering satellite ground pixels is taken for each 0.02×0.02 grid cell rather than using a weighted average, as done by Wenig et al. (2008). Only ground pixels with an area less than 950 km^2 are considered.

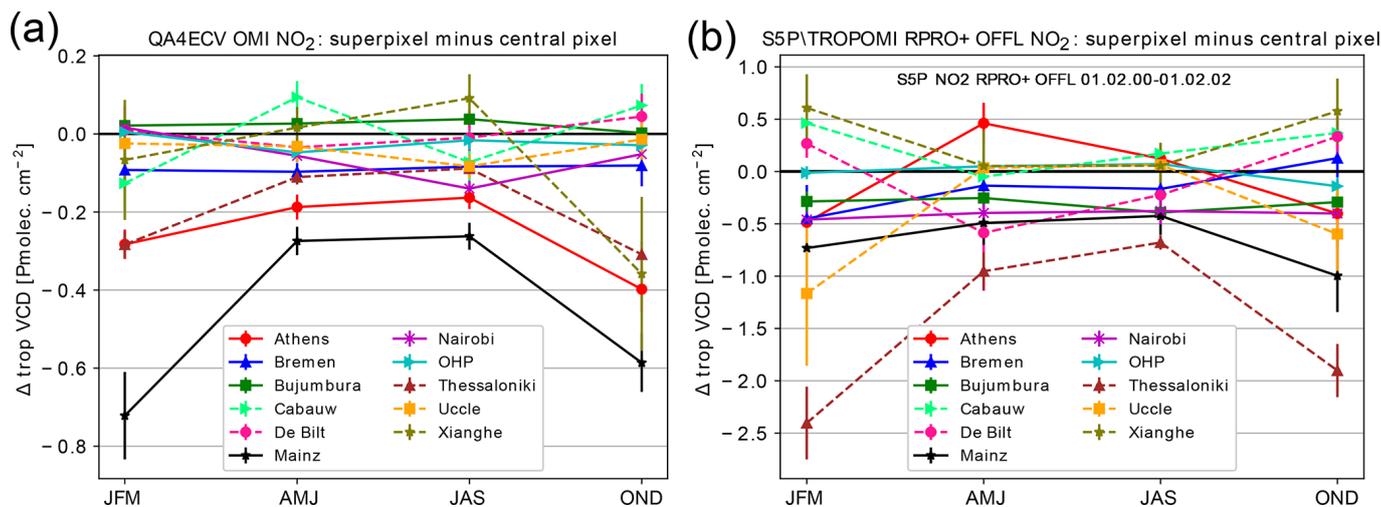


Figure 8. (a) Mean difference for each season between the QA4ECV OMI superpixel (ground pixels averaged within 20 km of the central site) and the central OMI ground pixel using data from 2004 to 2016. (b) Similar to (a) but using TROPOMI NO₂ data from April 2018 to May 2019, and the superpixel is defined within a latitude, longitude box of $\Delta\text{lat} = 0.14^\circ$, $\Delta\text{long} = 0.7^\circ$ centred on the MAX-DOAS site.

3.4.6 Comparison results

Bias and dispersion

Figure 9 (black boxplots) presents boxplots of the difference of QA4ECV OMI with co-located QA4ECV MAX-DOAS for each MAX-DOAS site. At all sites, the bias of QA4ECV OMI with respect to QA4ECV MAX-DOAS is negative. On an absolute scale, it is the lowest at the less polluted OHP and Bujumbura sites (mean difference of -0.9 and -1.7 Pmolec. cm⁻² respectively), and highest at the Thessaloniki and Mainz sites (mean difference of ~ -4 Pmolec. cm⁻²). On a relative scale, the bias is lowest (median relative difference between -15% and -20%) at the Uccle, Cabauw, De Bilt and Xianghe sites and highest (median relative difference of $\sim -70\%$) at Bujumbura and Nairobi. The difference dispersion, expressed as the interquartile range (IQR), is lowest at the Bujumbura, OHP and Nairobi sites (~ 1 – 2 Pmolec. cm⁻²) and largest at the Mainz and Xianghe sites (~ 5 – 6 Pmolec. cm⁻²).

As discussed in Sect. 3.4.2–3.4.5, among the different comparison error components, only the horizontal smoothing difference error is expected to induce an important negative bias, and this is only true for some sites (e.g. Thessaloniki and Mainz), whereas for other sites (e.g. OHP and Xianghe) this is not expected. This means that the bias is (at least in some cases) due to error in the satellite and/or MAX-DOAS measurement, not due to comparison error.

We present in the same figure boxplots of the tropospheric NO₂ VCD difference between OMNO2 data and QA4ECV MAX-DOAS measurements (blue boxplots). The bias of OMNO2 vs. QA4ECV MAX-DOAS is comparable to that of QA4ECV OMI NO₂ vs. QA4ECV MAX-DOAS, although slightly more negative at all sites except Cabauw. If one only

considers the subset of the OMNO2 pixels where QA4ECV OMI has an accepted pixel, the OMNO2 bias becomes closer to that of QA4ECV OMI for most sites. Although bePRO MAX-DOAS has a better correction for aerosols and vertical profile effects compared to QA4ECV MAX-DOAS in principle, the bias of QA4ECV OMI with respect to bePRO MAX-DOAS (Fig. 9, green boxes, only for the Bujumbura, Uccle and Xianghe sites) is comparable to that of QA4ECV OMI vs. QA4ECV MAX-DOAS.

In most cases, we conclude that mutual differences between the tropospheric NO₂ VCD of the two OMI satellite data products and between both MAX-DOAS processing methods are small compared with the differences between the satellite OMI data products and the MAX-DOAS measurements. The main exception is at OHP, where the median difference and relative difference of OMNO2 vs. QA4ECV MAX-DOAS (-1.4 Pmolec. cm⁻², -60%) is considerably more negative than that of QA4ECV OMI vs. QA4ECV MAX-DOAS (-0.8 Pmolec. cm⁻², -30%). The observation of higher MAX-DOAS tropospheric VCD compared with satellite measurements is a common finding in the literature (e.g. Ma et al., 2013; Kanaya et al., 2014; Chan et al., 2015; Jin et al., 2016; Drosoglou et al., 2017, 2018). Therefore, the negative bias is not specific to a particular satellite or MAX-DOAS data product.

Seasonal cycle of the bias

Figures 10 and 11 present a seasonal plot (i.e. all data mapped to 1 year) of QA4ECV OMI tropospheric NO₂ VCD, of QA4ECV MAX-DOAS, and of the difference for each site. Also indicated are the rolling monthly mean and median as well as outliers identified by iterative 4σ clipping.

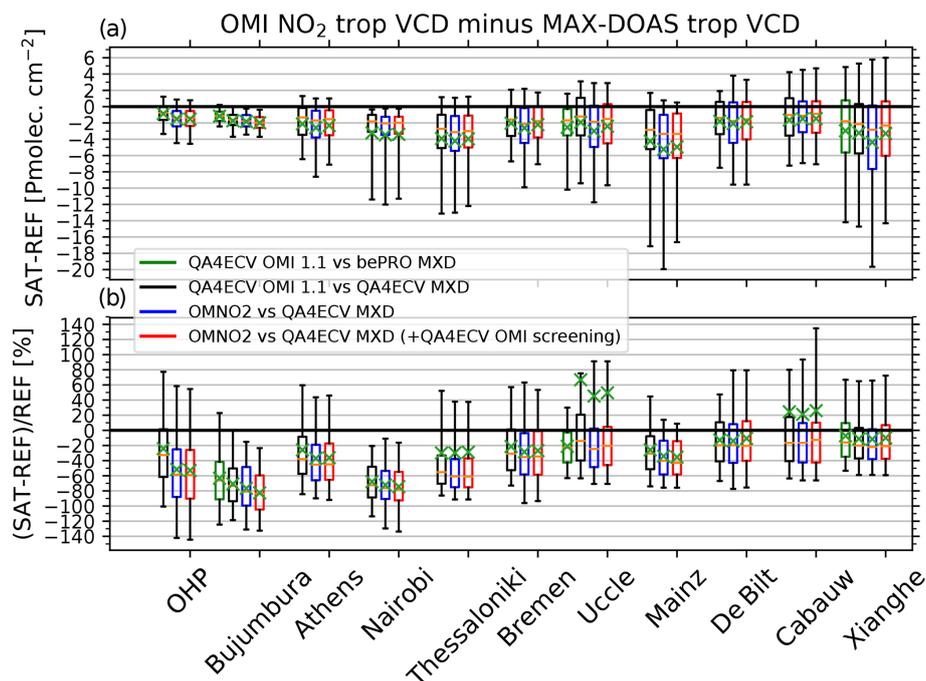


Figure 9. Boxplots for each site showing QA4ECV OMI NO₂ vs. QA4ECV MAX-DOAS (black boxes), QA4ECV OMI NO₂ vs. bePRO MAX-DOAS (green boxes, only for three sites), OMNO2 vs. QA4ECV MAX-DOAS (blue boxes) and OMNO2 vs. QA4ECV MAX-DOAS, but for the subset of OMNO2 pixels where QA4ECV OMI has an accepted pixel only (red boxes). Panel (a) displays boxplots of SAT-REF tropospheric VCD differences, and panel (b) displays boxplots of (SAT-REF)/REF. The same boxplot conventions are used as in Fig. 3. Outlying mean relative differences (green crosses) can occur when low REF values are present.

A seasonal cycle in the bias, with a larger underestimation in seasons with high NO₂, is a recurring feature (Fig. 10). This is the case at the more polluted sites such as Xianghe, Mainz and Thessaloniki in winter months and is in agreement with several literature results (Ma et al., 2013; Kanaya et al., 2014; Jin et al., 2016). Note, however, that we also find a seasonal cycle in the bias at the relatively clean OHP site. A very strong seasonal cycle in bias (tenfold increase) is present at Nairobi, where the MAX-DOAS sensor measures a strongly elevated NO₂, peaking in July and August, which is either not detected or hardly detected by the satellite. This is likely a (spatially) local phenomenon, which would be consistent with the locally enhanced NO₂ in Fig. S5. This site is characterized by local traffic. The enhanced NO₂ concentrations in July and August (as measured by MAX-DOAS) are possibly related to meteorology. This season is characterized by low precipitation, low wind speeds (see <https://weather-and-climate.com/average-monthly-Rainfall-Temperature-Sunshine,Nairobi,Kenya>, last access: 22 September 2019) and high cloud cover (as indicated by the QA4ECV OMI cloud fraction measurements) that limits NO₂ photolysis; therefore, a build-up of locally emitted NO₂ is a possible explanation. The fact that OMI hardly measures this elevated NO₂ could be due to the local character of the emissions.

Overall discrepancy and consistency with ex ante uncertainty

The discrepancy, as measured by the root-mean-squared difference (RMSD) between satellite measurements and MAX-DOAS data, exceeds the combined ex ante uncertainty for all sites⁶ (see Fig. 12, for the squared quantities). Clearly, comparison error contributes significantly to the RMSD, and/or there are underestimated/unrecognized errors in the satellite or reference data.

The mean-squared difference in Fig. 12 is split into three additive components: (i) the squared mean difference (bias component), (ii) the variance of the rolling monthly mean difference (seasonal cycle component) and (iii) the variance of the residual difference. The first two components can be attributed to systematic error, whereas the third component can be attributed to random error and any uncharacterized systematic error. The leading component can be different for

⁶Although the root-mean-squared error (RMSE) and uncertainty are not exactly equivalent, they should be roughly comparable if all error sources are well characterized. If all error is purely random, the RMSE equals the standard deviation of errors, of which the uncertainty is an ex ante estimate. If the error is fully systematic and constant, the RMSE equals the absolute value of the bias, which is expected to be smaller than twice the uncertainty with a 95 % probability.

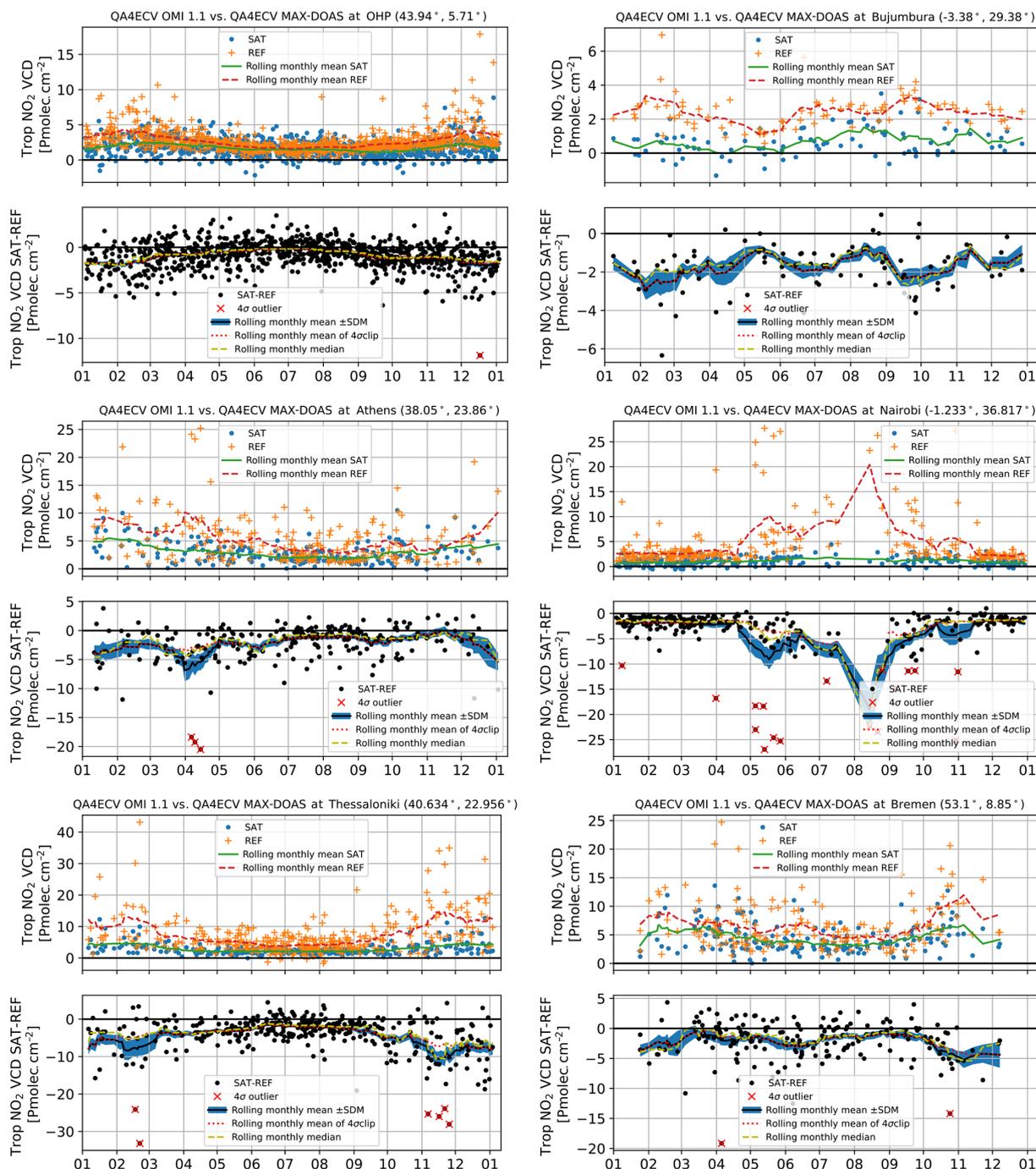


Figure 10. Seasonal cycle plots for the OHP, Bujumbura, Athens, Nairobi, Thessaloniki and Bremen sites. The top panel for each site shows the tropospheric VCD of QA4ECV OMI NO₂ and QA4ECV MAX-DOAS as well as the rolling monthly mean and median of both. The bottom panel for each site shows the differences between QA4ECV OMI NO₂ and QA4ECV MAX-DOAS, the outliers indicated by 4σ clipping, and rolling monthly mean and median of the difference.

each site (e.g. the bias component at Bujumbura, the seasonal component at Nairobi, and the residual at Mainz and Xianghe).

The satellite and reference data products do not provide the information to split the squared uncertainty according

to the random or systematic nature of the error source. Instead, the squared uncertainty in Fig. 12 is separated into additive components according to origin: (i) uncertainty in the MAX-DOAS measurement u_{GB} , (ii) uncertainty in the satellite measurement due to error in the SCD (expected to

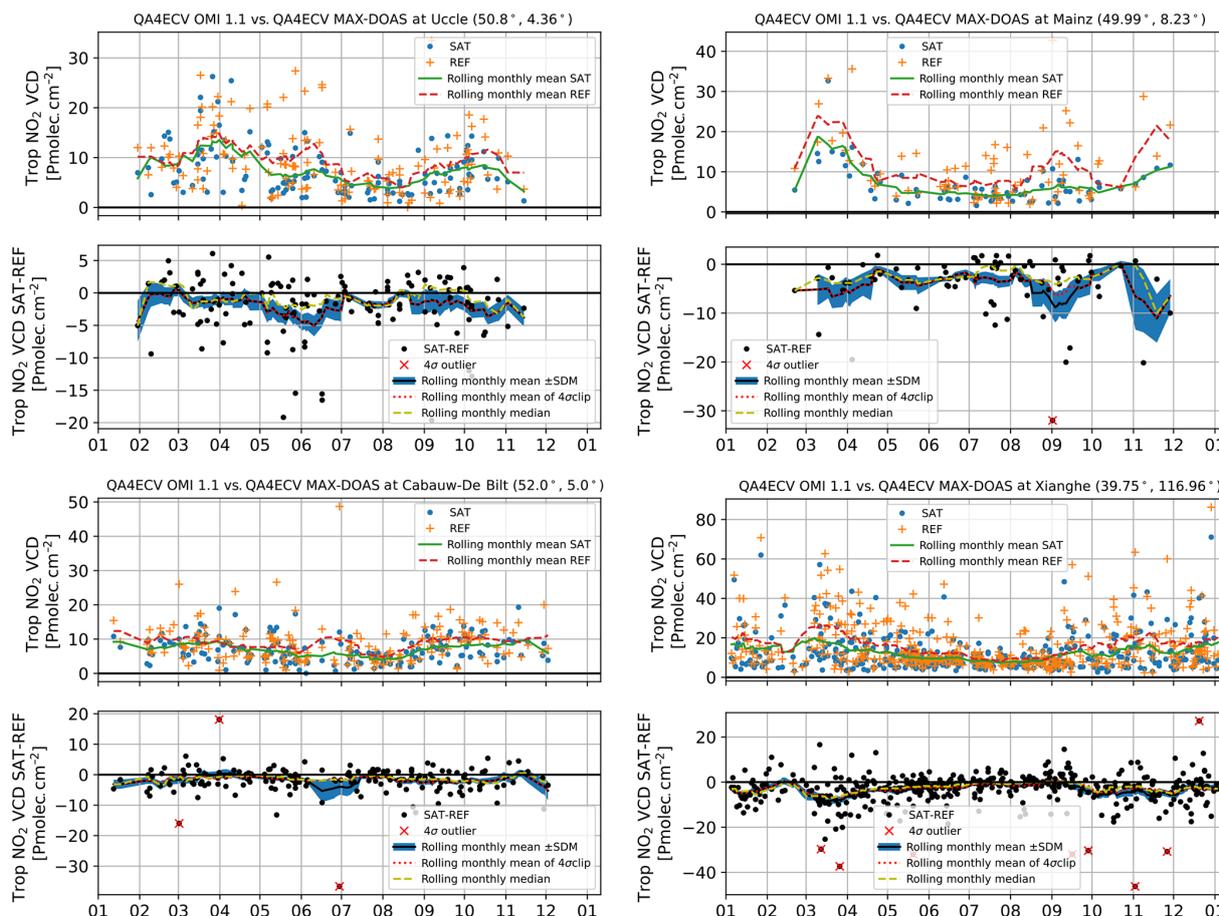


Figure 11. As in Fig. 10 but for the Uccle, Mainz, Cabauw/De Bilt and Xianghe sites.

be mainly random in nature) u_{SAT, N_s} , (iii) stratospheric SCD $u_{SAT, N_{s, strat}}$ and (iv) uncertainty in satellite measurement due to error in tropospheric AMF $u_{SAT, M_{trop}}$. For the sites with the lowest NO₂ levels (OHP and Bujumbura), uncertainty in the SCD is the main contributor, whereas the MAX-DOAS uncertainty becomes the leading component for the other sites.

By analysing and intercomparing the tropospheric AMF calculation methods between different retrieval groups, Lorente et al. (2017) concluded that the uncertainty due to differences in retrieval methodology (i.e. methodological uncertainty – termed structural uncertainty by Lorente et al., 2017) is 32 % under unpolluted conditions and 42 % under polluted conditions, which is mostly due to the different choices regarding ancillary data surface albedo, a priori profile and cloud parameters by different groups. In Fig. 12, this AMF component of the methodological uncertainty, $u_{SAT, meth, M_{trop}}$, is presented as an alternative to the ex ante $u_{SAT, M_{trop}}$ obtained by uncertainty propagation, where we classified OHP and Bujumbura as unpolluted sites and the others as polluted. In all cases, the methodological uncertainty exceeds the ex ante uncertainty $u_{SAT, M_{trop}}$. At four sites, the discrepancy between OMI and MAX-DOAS can

be explained for the most part (Uccle and Cabauw/De Bilt) or even completely (Xianghe) using this methodological uncertainty, but this is not the case at the other sites.

Modifying screening criteria

Applying a more strict screening protocol can, at least in principle, mitigate discrepancies in bias and dispersion, at the expense of data loss. In the case at hand, results are mixed for the different sites (see Figs. S9–S13); stricter criteria do not resolve bias or dispersion for all sites. For the Uccle, Mainz, Cabauw and Xianghe sites strong reductions in bias and/or dispersion (~ 0.5 – 2 Pmolec. cm⁻²) can be achieved by filtering more strictly on the effective cloud properties cloud fraction, cloud pressure, the uncertainty component due to cloud pressure $u_{SAT, p_{cl}}$, the MAX-DOAS cloud flag (removing scenes with thick or broken clouds) or the AOD. This suggests that part of the discrepancy is caused by clouds and/or aerosol. More minor reductions in the bias and/or dispersion are achieved for the Bujumbura, Nairobi, Athens, Bremen and De Bilt sites.

Screening more strictly on the ground pixel area leads to improvements in the bias for Mainz and Thessaloniki, con-

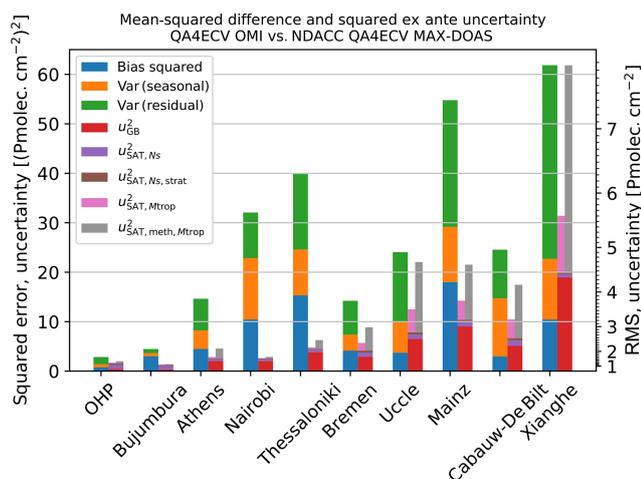


Figure 12. Two stacked bar plots are provided for each site. The left bar shows the mean-squared difference of QA4ECV OMI NO₂ vs. QA4ECV MAX-DOAS, which is split into three components: (i) the square of the mean difference, (ii) the variance of the rolling monthly mean difference and (iii) the variance of the residual difference. The right bar shows the combined ex ante uncertainty of QA4ECV OMI NO₂–QA4ECV MAX-DOAS, which is split into four components: (i) the MAX-DOAS squared uncertainty, (ii) the QA4ECV OMI squared uncertainty contribution from the total SCD and (iii) from the stratospheric SCD, and (iv) the QA4ECV OMI squared uncertainty contribution from the tropospheric AMF. Also shown is the AMF component of the methodological uncertainty, derived by intercomparing the retrieval methodologies by Lorente et al. (2017), which is referred to as structural uncertainty in this work. The right y axis provides a square-root scaling of the corresponding RMS.

firming (see Sect. 3.4.5) that the horizontal smoothing difference error is a component of the bias. Improvements in dispersion are found for Mainz, Thessaloniki, Uccle and Xianghe.

Using a stricter filter on effective cloud properties, the RMSD can be made consistent with the ex ante uncertainty for the Uccle and Cabauw-De Bilt sites (results not shown). For Mainz, this can be achieved if ground pixels larger than 400 km² are also excluded (keeping only 25 % of the data). Finally, we note that the RMSD and uncertainty are consistent in the months from May to (and including) August (when NO₂ values are low) at the OHP site without the need for stricter filtering.

For most sites, additional screening (within reasonable limits) cannot lower the RMSD enough that it matches the uncertainty. Some uncertainty components in either OMI or MAX-DOAS data are likely underestimated or not included.

While we found that stricter screening using the uncertainty component due to cloud pressure, $u_{\text{SAT}, p_{\text{cl}}}$, often leads to better results, the threshold values obtained are quite low. This indicates that $u_{\text{SAT}, p_{\text{cl}}}$ is underestimated in the satellite data product. As expected, relaxing the cloud fraction filter

beyond the baseline can lead to an increase in bias and/or dispersion (see e.g. Bujumbura, Nairobi and Uccle in Figs. S9–S13), motivating the $\text{CF} \leq 0.2$ (or almost equivalently $\text{CRF} \leq 0.5$) recommendation. On the other hand, relaxing the AMF ratio filter beyond the baseline has no large impact on the comparison, whereas further restricting it sometimes has a negative impact (e.g. an increase in the bias and/or dispersion at Uccle, Xianghe and Cabauw). Therefore, the current recommended baseline lower bound ($\frac{\text{AMF}_{\text{trop}}}{\text{AMF}_{\text{geo}}} \geq 0.2$) can be replaced by a lower value (e.g. 0.1 or 0.05).

Vertical smoothing

The nonuniform vertical sensitivity of the satellite measurement, combined with an approximate a priori profile shape, is a source of error in the satellite measurement. The bePRO MAX-DOAS provides not only column but also profile shape information (albeit with a limited vertical resolution) and, therefore, allows for this error source to be assessed separately. Figure 13 shows the impact of directly applying Eq. (3) on the bePRO MAX-DOAS profile (after vertical alignment using the method from Zhou et al., 2009) on the mean-squared deviation (MSD) as well as its bias, seasonal cycle and residual components for the Uccle and Xianghe sites. While direct smoothing of the MAX-DOAS profile improves the MSD for Uccle, it increases it for Xianghe because the seasonal cycle component increases. The increase in seasonal variance is caused by the interplay of the seasonal variation in the MAX-DOAS vertical profile and of the satellite vertical averaging kernel. Specifically, for the Xianghe case, it is found that averaging kernels have higher values close to the surface in wintertime, while MAX-DOAS NO₂ profiles can also be peaked at the surface. This combination causes increased MAX-DOAS columns upon vertical smoothing. This is also seen in cases such as the comparison of the GOME-2 AC SAF GDP 4.8 NO₂ product with MAX-DOAS at Xianghe (see Fig. 7.14 and 7.15 of Hovila et al., 2018, and Figs. S3 and S5 of Liu et al., 2019). While the a priori harmonization seems to mitigate this effect, it does not resolve it. Thus, whether the situation could be improved by improved MAX-DOAS a priori profiles and/or improved satellite averaging kernels should be a focus of future research.

However, one should consider that the retrieved bePRO profiles have a low vertical resolution and depend on their own a priori profile shape. As is well known (Eq. 10 of Rodgers and Connor, 2003, see also the general profile harmonization overview of Keppens et al., 2019), a priori profiles of satellite and reference data should be harmonized before comparison and smoothing. Here, we aligned the surface levels of the profiles following Zhou et al. (2009) and changed the a priori shape profile of the bePRO data to that of the satellite while keeping the bePRO a priori VCD size (which is actually obtained from measurement, see Hendrick et al., 2014) intact. More detail on the operations applied is

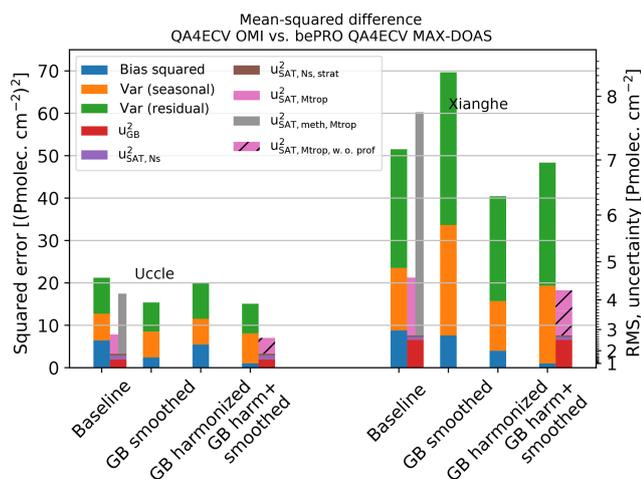


Figure 13. Mean-squared deviation of QA4ECV OMI vs. bePRO MAX-DOAS tropospheric VCD at Uccle and Xianghe split into the squared mean difference (blue), variance of the rolling monthly mean difference (orange) and variance of the residual difference (green) components. For each site, from left to right, (i) baseline comparison, (ii) MAX-DOAS profile smoothed by the OMI averaging kernel, (iii) MAX-DOAS a priori replaced with that of the satellite and (iv) a priori harmonization followed by smoothing are shown. Details of the operations are provided in Sect. S6. At the baseline (i), the squared ex ante uncertainty (divided into components) is also provided. The same squared ex ante uncertainty minus the satellite profile shape uncertainty contribution is provided in (iv).

provided in Sect. S6. The harmonization operation reduces all components of the MSD (bias, seasonal cycle and residual component) for the Xianghe site. When smoothing is also applied after the a priori harmonization, the bias component (blue bar in Fig. 13) is almost completely removed, but the other two components increase. Therefore, application of the averaging kernel does not necessarily lead to an improvement in all aspects of a comparison; this should be the focus of further research. As the bias component is almost completely removed after harmonization and smoothing at Uccle and Xianghe, one can conclude that – at least at these two sites – the bias is largely due to errors in the a priori profile shape. Therefore, using better quality a priori profiles in both satellite and MAX-DOAS data (e.g. from regional-scale models) is recommended.

When the averaging kernel is applied, it is recommended that the satellite a priori shape component is removed from the uncertainty budget (Boersma et al., 2018). This component was tentatively assigned 10% of the VCD value. This only leads to a modest reduction in the combined uncertainty in Fig. 13 (compare the non-hatched and hatched pink bars), as the dominant contribution to the OMI AMF uncertainty component is related to surface albedo rather than profile shape. For example, the combined uncertainty at Uccle decreases from 2.8 to 2.7 Pmolec. cm⁻². However,

the smoothing operation decreases the RMSD at Uccle by about 2 Pmolec. cm⁻², which strongly suggests that the current 10% uncertainty assignment is an underestimate.

4 Conclusion

In this work, stratospheric and tropospheric NO₂ VCDs of the QA4ECV OMI 1.1 data product are validated using ground-based NDACC ZSL-DOAS data and MAX-DOAS data respectively. Two MAX-DOAS processing methods are used: the NDACC bePRO profile retrieval and the harmonized QA4ECV column retrieval.

Quality screening according to the data product provider's recommendations is an essential step before the satellite product can be used. However, users (e.g. developers of L3-type temporally averaged data) should be aware that this leads to a preference of cloud-free scenes for tropospheric VCD and, therefore, to a negative sampling bias, especially at polluted sites (a strong reduction in mean VCD from 24 to 15 Pmolec. cm⁻² at Xianghe, and a reduction of a few peta molecules per square centimetre (Pmolec. cm⁻²) at Nairobi, Bremen, Thessaloniki and De Bilt/Cabauw). This sampling bias is reduced at De Bilt and Bremen by relaxing the lower bound filter on $\frac{M_{\text{trop}}}{M_{\text{geo}}}$ from 0.2 to 0.05.

The QA4ECV OMI stratospheric NO₂ VCD has a small (mostly wintertime) bias with respect to the ZSL-DOAS measurements of the order of -0.2 ± 0.06 Pmolec. cm⁻² (5%–10%) and a dispersion of 0.2–1 Pmolec. cm⁻², with good representation of the seasonal cycle.

QA4ECV OMI tropospheric NO₂ VCD is negatively biased vs. the MAX-DOAS data. This is not unique to this data product: the same conclusion is reached for NASA's OMI OMNO2 data product and for several other tropospheric NO₂ data products in the literature. The overall discrepancy exceeds the combined ex ante uncertainty of satellite and MAX-DOAS data. This is a contrasting conclusion to that of Boersma et al. (2018), who states that uncertainties seem to be overestimated, although their findings were for a single site over a 1-month time period (Tai'an, China, June 2006).

We studied a wide range of potential error sources of the discrepancy in tropospheric VCD between satellite and MAX-DOAS data. An overview is provided in Table 3.

At several sites, the MAX-DOAS instrument is located close (within satellite pixel distance) to an emission source; therefore, the horizontal smoothing difference error explains (part of) the bias, but there are also a few cases (OHP, Cabauw and Xianghe) where this does not hold. Sampling difference errors were found to be either minor (temporal and horizontal), or they were found to contribute in the opposite direction (vertical).

Measurement/retrieval error in satellite and MAX-DOAS data are other potential sources of discrepancy. Errors in the satellite total SCD and stratospheric SCD do not contribute much, leaving errors in satellite tropospheric AMF or MAX-

Table 3. Overview of discrepancies and error sources studied in this work. (MXD refers to MAX-DOAS.)

Contribution	Description
Full discrepancy e_{total}	Negative bias ranging from $-0.9 \text{ Pmolec. cm}^{-2}$ (OHP) to $-4 \text{ Pmolec. cm}^{-2}$ (Mainz and Thessaloniki). RMSD ranging from 2 (OHP and Bujumbura) to $8 \text{ Pmolec. cm}^{-2}$ (Xianghe). RMSD dominated by bias in Bujumbura and Thessaloniki, by seasonal cycle dispersion in Nairobi, and by residual dispersion otherwise.
Comparison errors	
Temporal sampling diff. error $e_{\Delta t}$	Mitigated by averaging MXD within a 1 h interval. No systematic component. Impact on dispersion ^a is $\leq 0.1 \text{ Pmolec. cm}^{-2}$ (low pollution) to 0.1 to $0.5 \text{ Pmolec. cm}^{-2}$ (high pollution).
Horizontal sampling diff. error $e_{\Delta r}$	Mitigated by excluding ground pixels not covering the site. Systematic component between 0 and $-0.6 \text{ Pmolec. cm}^{-2}$. Impact on dispersion ^a is $\leq 0.1 \text{ Pmolec. cm}^{-2}$ (low pollution) to $\leq 0.6 \text{ Pmolec. cm}^{-2}$ (high pollution).
Vertical sampling diff. error $e_{\Delta z, \text{ surface level}}$	The alignment of the satellite a priori profile to the MXD surface level using the method from Zhou et al. (2009) changes the bias by $\leq 0.3 \text{ Pmolec. cm}^{-2}$. Bujumbura's complicated orography might lead to a higher bias. In Athens, the MXD's location on a hill is a likely source of positive bias
Vertical sampling diff. error $e_{\Delta z, \text{ top grid level}}$	The MXD VCD is restricted to the lower troposphere. The correction is estimated from the satellite upper tropospheric a priori profile which increases the bias.
Horizontal smoothing diff. error e_{S_r}	Qualitatively assessed. Contributes to bias in Nairobi, Thessaloniki and Mainz, but does not contribute (significantly) to the bias in OHP, Cabauw or Xianghe. For other sites, the results are mixed.
Measurement/retrieval errors	
OMI total SCD error $e_{\text{SAT},s}$	Impact of the noise term on the dispersion ^a is $\leq 0.1 \text{ Pmolec. cm}^{-2}$ for low pollution sites and negligible for high pollution sites.
OMI strat. SCD error $e_{\text{SAT},s,\text{strat}}$	Bias in the stratospheric VCD of $\sim -0.2 \text{ Pmolec. cm}^{-2}$ translates (via $\frac{M_{\text{strat}}}{M_{\text{trop}}}$) to $\sim +0.6 \text{ Pmolec. cm}^{-2}$ in tropospheric VCD.
OMI trop. AMF error $e_{\text{SAT},M_{\text{trop}}}$	Error of between 32 % and 42 % (Lorente et al., 2017), dominated by the choice of a priori profile, cloud parameters and surface albedo. This could explain (most or all) of the discrepancy at Uccle, Cabauw/De Bilt and Xianghe.
Error due to cloud or aerosol (OMI or MXD)	Strong reduction in bias and/or dispersion due to stricter filtering for Uccle, Mainz, Cabauw and Xianghe. Simulations (Ma et al., 2013; Jin et al., 2016) indicate that cloud or aerosol can cause a factor of 2 underestimation for satellite data and up to a 20 % overestimation for MXD data.
Error due to vertical smoothing	Only assessed with bePRO MXD at Uccle and Xianghe, applying a priori harmonization and smoothing. The mean difference decreases from -3 to $-1 \text{ Pmolec. cm}^{-2}$, and the median difference decreases from -2 to $0 \text{ Pmolec. cm}^{-2}$. The RMSD shows a small reduction.

^a "Impact on dispersion" refers to the potential reduction in the standard deviation of $N_{v,\text{trop},\text{SAT}} - N_{v,\text{trop},\text{REF}}$ if the estimated standard deviation due to this particular error source was subtracted in quadrature.

DOAS data as candidate error sources. Part of the discrepancy is caused by errors in either the satellite or MAX-DOAS measurement induced by (low) clouds and/or aerosol (e.g. at the Mainz and Xianghe sites). According to radiative transfer simulations (Ma et al., 2013; Jin et al., 2016), these effects impact the satellite tropospheric NO₂ VCD measurements (factor of ~ 2 decrease) more than the MAX-DOAS measurements (overestimation of 20 % at most). Moreover, the nonuniform vertical sensitivity of OMI and the uncertainty in the a priori profile shape contributes to the discrepancy, as

shown here with the QA4ECV OMI vs. bePRO MAX-DOAS comparison. This is in agreement with the work of Lorente et al. (2017), who showed that the uncertainty in the retrieval method (due to inter-team retrieval setting differences; shorthand methodological uncertainty) in the tropospheric AMF is dominated by differences in the a priori profile, cloud parameters and surface albedo. Moreover, using this uncertainty estimate for the AMF instead of the ex ante, one can explain the SAT-REF tropospheric VCD discrepancies for three sites (Uccle, Cabauw/De Bilt and Xianghe). For these three sites,

consistency can also be reached by filtering cloud parameters more strictly.

Finally, for some of the discrepancies there is no straightforward explanation. An example of this is the negative bias at OHP in wintertime. This is possibly related to a lower tropospheric AMF in wintertime, as the planetary boundary layer is shallower and the SZA is higher. As a result, comparisons become more sensitive to factors such as errors in the profile shape. Another example of the unexplained discrepancy is the negative bias at Nairobi, even when focusing on the months from December to March when tropospheric VCD values measured by MAX-DOAS are relatively low.

The potential impact of the horizontal smoothing difference error was analysed in a rather qualitative way in this work. Analysis using “Observing System Simulation Experiments” at a fine spatial resolution (Verhoelst et al., 2015) or other experimental set-ups (e.g. sensors measuring in multiple azimuth directions; Brinksmas et al., 2008; Ortega et al., 2015) can improve on this.

The inter-team harmonization of MAX-DOAS data within the QA4ECV project is an important step forward for satellite validation, although some issues remain regarding factors such as the harmonization of reported uncertainties. The FRM4DOAS project (<http://frm4doas.aeronomie.be>, last access: 22 April 2020) funded by the European Space Agency (ESA) should improve upon this with the development of the first central processing system for MAX-DOAS measurements built on state-of-the-art retrieval algorithms and corresponding settings.

The availability of an ex ante uncertainty for each measurement as well as its decomposition in source components greatly facilitates the validation. However, information on how individual measurement uncertainties should be combined is incomplete in the satellite and MAX-DOAS data files. This limits the ability to check certain things, such as if the respective overall bias, dispersion or seasonal cycle of the bias are within expectations; in this work, we only checked the consistency of the overall discrepancy (expressed as the RMSD) with the combined total uncertainty. It is recommended that information on the systematic/random nature and error correlation is included in the satellite data product.

The ex ante uncertainty for each pixel in the QA4ECV NO₂ satellite data product is likely underestimated. A solution for this could be to explicitly account for the methodological uncertainty on the AMF in a similar fashion to the process carried out for the QA4ECV HCHO data product (De Smedt et al., 2018). Alternatively, the uncertainty component due to the profile shape in the OMI product could be increased, as tests in this work show that the current 10 % assignment is an underestimate. The QA4ECV NO₂ recommended filter on the AMF ratio can be made less restrictive (e.g. 0.05 lower bound), reducing data loss and sampling bias without compromising the comparisons with MAX-DOAS. Furthermore, the replacement of the coarsely resolved TM5 NO₂ profiles with high spatial resolution profiles from re-

gional air quality analyses (e.g. CAMS regional, <http://www.regional.atmosphere.copernicus.eu>, last access: 20 September 2019) would be very helpful to bridge part of the gap between MAX-DOAS and OMI.

Code and data availability. The QA4ECV OMI NO₂ data are available from <http://www.qa4ecv.eu> (last access: 20 April 2020), under “ECV data” (Boersma et al., 2017a, <https://doi.org/10.21944/qa4ecv-no2-omi-v1.1>). The OMNO2 data are publicly available from the NASA Goddard Earth Sciences (GES) Data and Information Services Center public website: https://disc.gsfc.nasa.gov/datasets/OMNO2_V003/summary/ (last access: 22 September 2019) (Krotkov et al., 2019, <https://doi.org/10.5067/Aura/OMI/DATA2017>). The ZSL-DOAS data and bePRO MAX-DOAS, as part of the Network for the Detection of Atmospheric Composition Change (NDACC), are publicly available (see <http://www.ndacc.org>, last access: 22 April 2020). The QA4ECV MAX-DOAS data are available at http://uv-vis.aeronomie.be/groundbased/QA4ECV_MAXDOAS/index.php (last access: 20 April 2020); it is mandatory to contact the instrument principal investigators regarding any use of the data. The AERONET AOD data are available at <https://aeronet.gsfc.nasa.gov> (last access: 22 September 2019). Sentinel-5P NO₂ RPRO (reprocessed) and OFFL (offline) data from 1 February 2000 to 1 February 2002 can be obtained from the Sentinel-5P Pre-Operations Data Hub (<https://s5phub.copernicus.eu/dhus/#/home>, last access: 22 September 2019) (Copernicus Sentinel-5P, 2018, <https://doi.org/10.5270/S5P-s4ljg54>).

Part of the validation processing was performed using the HARP data harmonization toolset (© s[&t], the Netherlands), which is available at <https://github.com/stcorp/harp> (last access: 22 April 2020) under the BSD 3-Clause “New” or “Revised” Licence.

Supplement. The supplement related to this article is available online at: <https://doi.org/10.5194/acp-20-8017-2020-supplement>.

Author contributions. SC coordinated the paper and carried out the validation analysis. TV carried out the stratospheric VCD validation analysis. GP contributed insights into the tropospheric VCD validation analysis. DH, AK, JCL and TV contributed validation expertise. JG, SN and BR created software tools for validation. JG performed general data collection and format harmonization. FH coordinated the creation of the QA4ECV MAX-DOAS improved data sets. AB, JPB, FH, AP, JR, AR, MVR and TW are principal investigators for the QA4ECV MAX-DOAS measurements. FH and MVR are principal investigators for the bePRO MAX-DOAS measurements. FG, AP and JPP are principal investigators for the SAOZ ZSL-DOAS measurements. FB, HE, AR, IDS, AL, JvG, EP, MVR and TW are the authors of the QA4ECV NO₂ OMI data set. JCL is the coordinator of this research. All authors reviewed and commented on the paper.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. This research was carried out in the framework of the EU FP7 “Quality Assurance for Essential Climate Variables” (QA4ECV; grant no. 5 607405) project with support from the EU H2020 “Gap Analysis for Integrated Atmospheric ECV CLimate Monitoring” (GAIA-CLIM; Ares (2014)3708963/grant no. 640276) project. The authors are particularly grateful to QA4ECV for the generation of harmonized OMI and MAX-DOAS data sets within a rigorous quality assurance framework and to GAIA-CLIM WP3 (Comparison Error Budget Closure) for dedicated metrology support. Several validation analysis tools were funded by the Belgian Science Policy Office (BELSPO) and ESA through the ACROSAT PRODEX-10 project. We are grateful to Marina Zara (KNMI) for clarification regarding the different QA4ECV NO₂ OMI uncertainty fields. Ground-based ZSL-DOAS data and other MAX-DOAS data used in this publication were obtained as part of the Network for the Detection of Atmospheric Composition Change (NDACC) and are publicly available (see <http://www.ndacc.org>, last access: 22 April 2020). NASA OMNO₂ data were obtained from NASA’s Earth Observing System Data and Information System (EOSDIS). We are grateful to Nickolay A. Krotkov and Lok Nath Lamsal (NASA/GSFC) for clarification regarding the NASA OMNO₂ data product. We are also grateful to Trisisevgeni Stavrakou (BIRA-IASB) for fruitful discussions on tropospheric NO₂ chemistry. The European Commission is further acknowledged for having supported cross-fertilization meetings among FP7 (CLIP-C, QA4ECV, ERACLIM-2, EUCLEIA, EUPORIAS and UERRA) and H2020 (GAIA-CLIM and FIDUCEO) climate service-related projects. Regarding the AERONET data, we thank the principal investigators Rachel Akimana, Vassilis Amiridis, Meinrat Andreae, Alkiviadis Bais, Philippe Goloub, J. S. Bas Henzing, Christian Hermans, Eughne Ndenzako, Pierre Nzohabonayo, Michel Van Roozendael, Ucai Wang, Xiangao Xia and their staff for establishing and maintaining the eight AERONET sites used in this investigation. Sentinel-5 Precursor NO₂ data were used in this work. Sentinel-5 Precursor is a European Space Agency (ESA) mission on behalf of the European Commission (EC). The TROPOMI payload is a joint development by ESA and the Netherlands Space Office (NSO). The Sentinel-5 Precursor ground-segment development has been funded by ESA with national contributions from the Netherlands, Germany and Belgium.

Financial support. This research has been supported by the European Commission, FP7 (QA4ECV (grant no. 607405)) and Horizon 2020 (GAIA-CLIM (grant no. 640276)), and by the Belgian Federal Science Policy Office (BELSPO), and ESA (ProDEX-10 ACROSAT grant).

Review statement. This paper was edited by Ralf Sussmann and reviewed by two anonymous referees.

References

- 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 NO₂ from nadir-viewing satellites by weighted convolution, *Atmos. Meas. Tech.*, 9, 2753–2779, <https://doi.org/10.5194/amt-9-2753-2016>, 2016.
- Boersma, K. F., Eskes, H. J., and Brinksma, E. J.: Error analysis for tropospheric NO₂ retrieval from space, *J. Geophys. Res.*, 109, D04311, <https://doi.org/10.1029/2003jd003962>, 2004.
- 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., Eskes, H., Richter, A., De Smedt, I., Lorente, A., Beirle, S., Van Geffen, J., Peters, E., Van Roozendael, M., and Wagner, T.: QA4ECV NO₂ tropospheric and stratospheric vertical column data from OMI (Version 1.1) [Data set], Royal Netherlands Meteorological Institute (KNMI), <https://doi.org/10.21944/qa4ecv-no2-omi-v1.1> (last access: 20 April 2020), 2017a.
- Boersma, K. F., van Geffen, J., Eskes, H., van der A, R., De Smedt, I., Van Roozendael, 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 NO₂ ECV precursor product, techreport QA4ECV Deliverable D4.6, KNMI, available at: <http://www.qa4ecv.eu/sites/default/files/D4.6.pdf> (last access: 20 April 2020), 2017b.
- Boersma, K. F., Eskes, H. J., Richter, A., De Smedt, I., Lorente, A., Beirle, S., van Geffen, J. H. G. M., Zara, M., Peters, E., Van Roozendael, M., Wagner, T., Maasakkers, J. D., van der A, R. J., Nightingale, J., De Rudder, A., Irie, H., Pinardi, G., Lambert, J.-C., and Compernelle, S. C.: Improving algorithms and uncertainty estimates for satellite NO₂ retrievals: results from the quality assurance for the essential climate variables (QA4ECV) project, *Atmos. Meas. Tech.*, 11, 6651–6678, <https://doi.org/10.5194/amt-11-6651-2018>, 2018.
- Bognar, K., Zhao, X., Strong, K., Boone, C., Bourassa, A., Deegenstein, D., Drummond, J., Duff, A., Goutail, F., Griffin, D., Jeffery, P., Lutsch, E., Manney, G., McElroy, C., McLinden, C., Millán, L., Pazmino, A., Sioris, C., Walker, K., and Zou, J.: Updated validation of ACE and OSIRIS ozone and NO₂ measurements in the Arctic using ground-based instruments at Eureka, Canada, *J. Quant. Spectrosc. Ra.*, 238, 106571, <https://doi.org/10.1016/j.jqsrt.2019.07.014>, 2019.
- Brinksma, E. J., Pinardi, G., Volten, H., Braak, R., Richter, A., Schönhardt, A., van Roozendael, M., Fayt, C., Hermans, C., Dirksen, R. J., Vlemmix, T., Berkhout, A. J. C., Swart, D. P. J., Oetjen, H., Wittrock, F., Wagner, T., Ibrahim, O. W., de Leeuw, G., Moerman, M., Curier, R. L., Celarier, E. A., Cede, A., Knap, W. H., Veefkind, J. P., Eskes, H. J., Allaart, M., Rothe, R., PETERS, A. J. M., and Levelt, P. F.: The 2005 and 2006 DANDELIONS NO₂ and aerosol intercomparison campaigns, *J. Geophys. Res.-Atmos.*, 113, D16S46, <https://doi.org/10.1029/2007JD008808>, 2008.
- Bucsela, E. J., Krotkov, N. A., Celarier, E. A., Lamsal, L. N., Swartz, W. H., Bhartia, P. K., Boersma, K. F., Veefkind, J. P., Gleason, J. F., and Pickering, K. E.: A new stratospheric and

- tropospheric NO₂ retrieval algorithm for nadir-viewing satellite instruments: applications to OMI, *Atmos. Meas. Tech.*, 6, 2607–2626, <https://doi.org/10.5194/amt-6-2607-2013>, 2013.
- Bucsel, E. J., Celarier, E. A., Gleason, J. L., Krotkov, N. A., Lamsal, L. N., Marchenko, S. V., and Swartz, W. H.: OMNO2 README Document. Data Product Version 3.0, Tech. rep., NASA /Goddard Space Flight Center, available at: https://acdisc.gesdisc.eosdis.nasa.gov/data/Aura_OMI_Level3/OMNO2d.003/doc/README.OMNO2.pdf (last access: 22 September 2019), 2016.
- Celarier, E. A., Brinksma, E. J., Gleason, J. F., Veefkind, J. P., Cede, A., Herman, J. R., Ionov, D., Goutail, F., Pommereau, J.-P., Lambert, J.-C., van Roozendaal, M., Pinardi, G., Wittrock, F., Schönhardt, A., Richter, A., Ibrahim, O. W., Wagner, T., Bojkov, B., Mount, G., Spinei, E., Chen, C. M., Pongetti, T. J., Sander, S. P., Bucsel, E. J., Wenig, M. O., Swart, D. P. J., Volten, H., Kroon, M., and Levelt, P. F.: Validation of Ozone Monitoring Instrument nitrogen dioxide columns, *J. Geophys. Res.*, 113, D15S15, <https://doi.org/10.1029/2007jd008908>, 2008.
- Chan, K., Hartl, A., Lam, Y., Xie, P., Liu, W., Cheung, H., Lampel, J., Pöhler, D., Li, A., Xu, J., Zhou, H., Ning, Z., and Wenig, M.: Observations of tropospheric NO₂ using ground based MAX-DOAS and OMI measurements during the Shanghai World Expo 2010, *Atmos. Environ.*, 119, 45–58, <https://doi.org/10.1016/j.atmosenv.2015.08.041>, 2015.
- Chen, D., Zhou, B., Beirle, S., Chen, L. M., and Wagner, T.: Tropospheric NO₂ column densities deduced from zenith-sky DOAS measurements in Shanghai, China, and their application to satellite validation, *Atmos. Chem. Phys.*, 9, 3641–3662, <https://doi.org/10.5194/acp-9-3641-2009>, 2009.
- Chipperfield, M. P.: Multiannual simulations with a three-dimensional chemical transport model, *J. Geophys. Res.*, 104, 1781–1805, <https://doi.org/10.1029/98jd02597>, 1999.
- Clémer, K., Van Roozendaal, M., Fayt, C., Hendrick, F., Hermans, C., Pinardi, G., Spurr, R., Wang, P., and De Mazière, M.: Multiple wavelength retrieval of tropospheric aerosol optical properties from MAXDOAS measurements in Beijing, *Atmos. Meas. Tech.*, 3, 863–878, <https://doi.org/10.5194/amt-3-863-2010>, 2010.
- Compernelle, S. and Lambert, J.-C.: Standard terms and definitions applicable to the quality assurance of Essential Climate Variable data records, Tech. rep., Royal Belgian Institute for Space Aeronomy, <https://doi.org/10.18758/71021041>, 2017.
- Compernelle, S., Lambert, J.-C., and Niemeijer, S.: Prototype QA/Validation Service for Atmospheric ECV Precursors : Detailed Processing Model – Version 2, QA4ECV report Deliverable D2.5, Royal Belgian Institute for Space Aeronomy, available at: http://www.qa4ecv.eu/sites/default/files/QA4ECV_BIRA-IASB_D-2-5_AVS-DPMv2_20160623.pdf (last access: 20 April 2020), 2016.
- Compernelle, S., Lambert, J.-C., Verhoelst, T., Granville, J., Hubert, D., Keppens, A., Niemeijer, S., Rino, B., Pinardi, G., Beirle, S., Boersma, F., Clerbaux, C., Coheur, P., Smedt, I. D., Eskes, H., George, M., Hendrick, F., Lorente, A., Nightingale, J., Peters, E., Richter, A., van Geffen, J., Roozendaal, M. V., Wagner, T., and Yu, H.: Quality assessment of QA4ECV climate data records of atmospheric composition: terminology, methodology and application to tropospheric NO₂, HCHO and CO from the GOME-2, IASI and OMI satellites, in: Proceedings for the 2018 EU-METSAT Meteorological Satellite Conference, 17–21 September 2018, Tallinn, Estonia, 2018.
- Copernicus Sentinel-5P (processed by ESA): TROPOMI Level 2 Nitrogen Dioxide total column products, Version 01, European Space Agency, <https://doi.org/10.5270/S5P-s4ljg54> (last access: 22 September 2019), 2018.
- Crutzen, P.: The influence of nitrogen oxides on the atmospheric ozone content, *Q. J. Roy. Meteor. Soc.*, 96, 320–325, <https://doi.org/10.1002/qj.49709640815>, 1970.
- Delmas, R., Serça, D., and Jambert, C.: Global inventory of NO_x sources, *Nutr. Cycl. Agroecosys.*, 48, 51–60, <https://doi.org/10.1023/A:1009793806086>, 1997.
- De Mazière, M., Thompson, A. M., Kurylo, M. J., Wild, J. D., Bernhard, G., Blumenstock, T., Braathen, G. O., Hannigan, J. W., Lambert, J.-C., Leblanc, T., McGe, T. J., Nedoluha, G., Petropavlovskikh, I., Seckmeyer, G., Simon, P. C., Steinbrecht, W., and Strahan, S. E.: The Network for the Detection of Atmospheric Composition Change (NDACC): history, status and perspectives, *Atmos. Chem. Phys.*, 18, 4935–4964, <https://doi.org/10.5194/acp-18-4935-2018>, 2018.
- De Smedt, I., Theys, N., Yu, H., Danckaert, T., Lerot, C., Compernelle, S., Van Roozendaal, M., Richter, A., Hilboll, A., Peters, E., Pedergrana, M., Loyola, D., Beirle, S., Wagner, T., Eskes, H., van Geffen, J., Boersma, K. F., and Veefkind, P.: Algorithm theoretical baseline for formaldehyde retrievals from S5P TROPOMI and from the QA4ECV project, *Atmos. Meas. Tech.*, 11, 2395–2426, <https://doi.org/10.5194/amt-11-2395-2018>, 2018.
- Dirksen, R. J., Boersma, K. F., Eskes, H. J., Ionov, D. V., Bucsel, E. J., Levelt, P. F., and Kelder, H. M.: Evaluation of stratospheric NO₂ retrieved from the Ozone Monitoring Instrument: Intercomparison, diurnal cycle, and trending, *J. Geophys. Res.*, 116, D08305, <https://doi.org/10.1029/2010jd014943>, 2011.
- Drosoglou, T., Bais, A. F., Zyrichidou, I., Kouremeti, N., Poupkou, A., Liora, N., Giannaros, C., Koukouli, M. E., Balis, D., and Melas, D.: Comparisons of ground-based tropospheric NO₂ 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.
- Drosoglou, T., Koukouli, M. E., Kouremeti, N., Bais, A. F., Zyrichidou, I., Balis, D., van der A, R. J., Xu, J., and Li, A.: MAX-DOAS NO₂ observations over Guangzhou, China; ground-based and satellite comparisons, *Atmos. Meas. Tech.*, 11, 2239–2255, <https://doi.org/10.5194/amt-11-2239-2018>, 2018.
- Errera, Q. and Fonteyn, D.: Four-dimensional variational chemical assimilation of CRISTA stratospheric measurements, *J. Geophys. Res.*, 106, 12253–12265, 2001.
- 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.
- European Environment Agency: Air quality in Europe – 2018 report, Tech. Rep. No 12/2018, European Environment Agency, Publications Office of the European Union, Luxembourg, 2018.
- Frieß, U., Beirle, S., Alvarado Bonilla, L., Bösch, T., Friedrich, M. M., Hendrick, F., PETERS, A., Richter, A., van Roozendaal, M., Rozanov, V. V., Spinei, E., Tirpitz, J.-L., Vlemmix, T., Wagner, T., and Wang, Y.: Intercomparison of MAX-DOAS vertical profile retrieval algorithms: studies using synthetic data, At-

- mos. Meas. Tech., 12, 2155–2181, <https://doi.org/10.5194/amt-12-2155-2019>, 2019.
- GCOS: The Global Observing System for Climate: Implementation Needs. GCOS 2016 Implementation Plan, Tech. Rep. GCOS-200, The Global Climate Observing System, available at: https://library.wmo.int/opac/doc_num.php?explnum_id=3417 (last access: 23 September 2019), 2016.
- Gielen, C., Hendrick, F., Pinardi, G., De Smedt, I., Fayt, C., Hermans, C., Stavrakou, T., Bauwens, M., Müller, J.-F., Ndenzako, E., Nzohabonayo, P., Akimana, R., Niyonzima, S., Van Roozendael, M., and De Mazière, M.: Characterisation of Central-African aerosol and trace-gas emissions based on MAX-DOAS measurements and model simulations over Bujumbura, Burundi, Atmos. Chem. Phys. Discuss., <https://doi.org/10.5194/acp-2016-1104>, in review, 2017.
- Giles, D. M., Sinyuk, A., Sorokin, M. G., Schafer, J. S., Smirnov, A., Slutsker, I., Eck, T. F., Holben, B. N., Lewis, J. R., Campbell, J. R., Welton, E. J., Korkin, S. V., and Lyapustin, A. I.: Advancements in the Aerosol Robotic Network (AERONET) Version 3 database – automated near-real-time quality control algorithm with improved cloud screening for Sun photometer aerosol optical depth (AOD) measurements, Atmos. Meas. Tech., 12, 169–209, <https://doi.org/10.5194/amt-12-169-2019>, 2019.
- Goldberg, D. L., Lamsal, L. N., Loughner, C. P., Swartz, W. H., Lu, Z., and Streets, D. G.: A high-resolution and observationally constrained OMI NO₂ satellite retrieval, Atmos. Chem. Phys., 17, 11403–11421, <https://doi.org/10.5194/acp-17-11403-2017>, 2017.
- Gothenburg: The 1999 Gothenburg Protocol (part of the Convention on Long-Range Transboundary Air Pollution), available at: <http://www.unece.org/environmental-policy/conventions/air/guidance-documents-and-other-methodological-materials/gothenburg-protocol.html> (last access: 23 September 2019), 1999.
- Hendrick, F., Barret, B., Van Roozendael, M., Boesch, H., Butz, A., De Mazière, M., Goutail, F., Hermans, C., Lambert, J.-C., Pfeilsticker, K., and Pommereau, J.-P.: Retrieval of nitrogen dioxide stratospheric profiles from ground-based zenith-sky UV-visible observations: validation of the technique through correlative comparisons, Atmos. Chem. Phys., 4, 2091–2106, <https://doi.org/10.5194/acp-4-2091-2004>, 2004.
- Hendrick, F., Mahieu, E., Bodeker, G. E., Boersma, K. F., Chipperfield, M. P., De Mazière, M., De Smedt, I., Demoulin, P., Fayt, C., Hermans, C., Kreher, K., Lejeune, B., Pinardi, G., Servais, C., Stübi, R., van der A, R., Vernier, J.-P., and Van Roozendael, M.: Analysis of stratospheric NO₂ trends above Jungfraujoch using ground-based UV-visible, FTIR, and satellite nadir observations, Atmos. Chem. Phys., 12, 8851–8864, <https://doi.org/10.5194/acp-12-8851-2012>, 2012.
- Hendrick, F., Müller, J.-F., Clémer, K., Wang, P., De Mazière, M., Fayt, C., Gielen, C., Hermans, C., Ma, J. Z., Pinardi, G., Stavrakou, T., Vlemmix, T., and Van Roozendael, M.: Four years of ground-based MAX-DOAS observations of HONO and NO₂ in the Beijing area, Atmos. Chem. Phys., 14, 765–781, <https://doi.org/10.5194/acp-14-765-2014>, 2014.
- Hendrick, F., Dils, B., Gielen, C., Langerock, B., Pinardi, G., De Mazière, M., Van Roozendael, M., Peters, E., Richter, A., PETERS, A., Beirle, S., Wagner, T., Drosoglou, T., Bais, A., Wang, S., Cuevas, C., and Saiz-Lopez, A.: Historical record of independent reference data for NO₂, HCHO, and CO, techreport QA4ECV Deliverable D3.8, Belgian Institute for Space Aeronomy, available at: http://www.qa4ecv.eu/sites/default/files/QA4ECV_D3.8_v1.0_web.pdf (last access: 20 April 2020), 2016.
- Hendrick, F., Dils, B., Langerock, B., Pinardi, G., Roozendael, M. V., Seyler, A., Peters, F. W. E., Richter, A., PETERS, A., Drosoglou, T., Bais, A., Wagner, T., and Dönnner, S.: Report on independent validation of atmospheric reference data sets., techreport QA4ECV Deliverable D3.10, Belgian Institute for Space Aeronomy, available at: http://www.qa4ecv.eu/sites/default/files/QA4ECV_D3.10_v2.pdf (last access: 20 April 2020), 2018.
- Heue, K.-P., Richter, A., Bruns, M., Burrows, J. P., v. Friedeburg, C., Platt, U., Pundt, I., Wang, P., and Wagner, T.: Validation of SCIAMACHY tropospheric NO₂-columns with AMAX-DOAS measurements, Atmos. Chem. Phys., 5, 1039–1051, <https://doi.org/10.5194/acp-5-1039-2005>, 2005.
- 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, Environ. Health, 12, 43, <https://doi.org/10.1186/1476-069X-12-43>, 2013.
- Hovila, J., Schmidt, A., Valks, P., Tuinder, O., van Versendaal, R., Jøynch-Sørensen, H., Koukouli, M., Garane, K., Delcloo, A., Pinardi, G., Langerock, B., Steinbrecht, W., George, M., Clerbaux, C., Astoreca, R., Hurtmans, D., Coheur, P.-F., and Vicente, C.: EUMETSAT AC SAF Operations report, Issue 1/2018 rev. 2, Reporting period: January–June 2018, Tech. Rep. SAF/AC/FMI/OPS/RP/001, EUMETSAT Satellite Application Facility on Atmospheric Composition Monitoring, available at: https://acsaf.org/docs/or/AC_SAF_Operations_Report_1-2018.pdf (last access: 21 April 2020), 2018.
- Huijnen, V., Williams, J., van Weele, M., van Noije, T., Krol, M., Dentener, F., Segers, A., Houweling, S., Peters, W., de Laat, J., Boersma, F., Bergamaschi, P., van Velthoven, P., Le Sager, P., Eskes, H., Alkemade, F., Scheele, R., Nédélec, P., and Pätz, H.-W.: The global chemistry transport model TM5: description and evaluation of the tropospheric chemistry version 3.0, Geosci. Model Dev., 3, 445–473, <https://doi.org/10.5194/gmd-3-445-2010>, 2010.
- Ionov, D. V., Timofeyev, Y. M., Sinyakov, V. P., Semenov, V. K., Goutail, F., Pommereau, J.-P., Bucsele, E. J., Celarier, E. A., and Kroon, M.: Ground-based validation of EOS-Aura OMI NO₂ vertical column data in the midlatitude mountain ranges of Tien Shan (Kyrgyzstan) and Alps (France), J. Geophys. Res., 113, D15S08, <https://doi.org/10.1029/2007jd008659>, 2008.
- Irie, H., Takashima, H., Kanaya, Y., Boersma, K. F., Gast, L., Wittrock, F., Brunner, D., Zhou, Y., and Van Roozendael, M.: Eight-component retrievals from ground-based MAX-DOAS observations, Atmos. Meas. Tech., 4, 1027–1044, <https://doi.org/10.5194/amt-4-1027-2011>, 2011.
- Jin, J., Ma, J., Lin, W., Zhao, H., Shaiganfar, R., Beirle, S., and Wagner, T.: MAX-DOAS measurements and satellite validation of tropospheric NO₂ and SO₂ vertical column densities at a rural site of North China, Atmos. Environ., 133, 12–25, <https://doi.org/10.1016/j.atmosenv.2016.03.031>, 2016.
- Joint Committee for Guides in Metrology (JCGM): Evaluation of measurement data – Guide to the expression of uncertainty in measurement, Tech. rep., JCGM, available

- at: http://www.bipm.org/utis/common/documents/jcgm/JCGM_100_2008_E.pdf (last access: 24 April 2020), 2008.
- Joint Committee for Guides in Metrology (JCGM): International Vocabulary of Metrology – Basic and General Concepts and Associated Terms, Tech. rep., JCGM, available at: http://www.bipm.org/utis/common/documents/jcgm/JCGM_200_2012.pdf (last access: 24 April 2020), 2012.
- Kanaya, Y., Irie, H., Takashima, H., Iwabuchi, H., Akimoto, H., Sudo, K., Gu, M., Chong, J., Kim, Y. J., Lee, H., Li, A., Si, F., Xu, J., Xie, P.-H., Liu, W.-Q., Dzhola, A., Postlyakov, O., Ivanov, V., Grechko, E., Terpugova, S., and Panchenko, M.: Long-term MAX-DOAS network observations of NO₂ in Russia and Asia (MADRAS) during the period 2007–2012: instrumentation, elucidation of climatology, and comparisons with OMI satellite observations and global model simulations, *Atmos. Chem. Phys.*, 14, 7909–7927, <https://doi.org/10.5194/acp-14-7909-2014>, 2014.
- Keppens, A., Lambert, J.-C., Granville, J., Miles, G., Siddans, R., van Peet, J. C. A., van der A, R. J., Hubert, D., Verhoelst, T., Delcloo, A., Godin-Beekmann, S., Kivi, R., Stübi, R., and Zehner, C.: Round-robin evaluation of nadir ozone profile retrievals: methodology and application to MetOp-A GOME-2, *Atmos. Meas. Tech.*, 8, 2093–2120, <https://doi.org/10.5194/amt-8-2093-2015>, 2015.
- Keppens, A., Compernelle, S., Verhoelst, T., Hubert, D., and Lambert, J.-C.: Harmonization and comparison of vertically resolved atmospheric state observations: methods, effects, and uncertainty budget, *Atmos. Meas. Tech.*, 12, 4379–4391, <https://doi.org/10.5194/amt-12-4379-2019>, 2019.
- 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.*, 113, D18308, <https://doi.org/10.1029/2008jd010290>, 2008.
- Krotkov, N. A., Lamsal, L. N., Celarier, E. A., Swartz, W. H., Marchenko, S. V., Bucsela, E. J., Chan, K. L., Wenig, M., and Zara, M.: The version 3 OMI NO₂ standard product, *Atmos. Meas. Tech.*, 10, 3133–3149, <https://doi.org/10.5194/amt-10-3133-2017>, 2017.
- Krotkov, N. A., Lamsal, L. N., Marchenko, S. V., Celarier, E. A., Bucsela, E. J., Swartz, W. H., Joiner, J., and the OMI core team: OMI/Aura Nitrogen Dioxide (NO₂) Total and Tropospheric Column 1-orbit L2 Swath 13 × 24 km V003, Goddard Earth Sciences Data and Information Services Center (GES DISC), Greenbelt, MD, USA, <https://doi.org/10.5067/Aura/OMI/DATA2017>, last access: 22 September 2019.
- Lambert, J.-C., Van Roozendaal, M., Granville, J., Gérard, P., Simon, P., Claude, H., and Staehelin, J.: Comparison of the GOME ozone and NO₂ total amounts at mid-latitude with ground-based zenith-sky measurements, in: Atmospheric Ozone, Proceedings of the XVIII Quadrennial Ozone Symposium, L'Aquila, Italy, 12–21 September 1996.
- Levelt, P. F., van den Oord, G. H. J., Dobber, M. R., Malkki, A., Huib Visser, Johan de Vries, Stammes, P., Lundell, J. O. V., and Saari, H.: The ozone monitoring instrument, *IEEE Trans. Geosci. Remote Sens.*, 44, 1093–1101, <https://doi.org/10.1109/TGRS.2006.872333>, 2006.
- Liu, S., Valks, P., Pinardi, G., De Smedt, I., Yu, H., Beirle, S., and Richter, A.: An improved total and tropospheric NO₂ column retrieval for GOME-2, *Atmos. Meas. Tech.*, 12, 1029–1057, <https://doi.org/10.5194/amt-12-1029-2019>, 2019.
- Loew, A., Bell, W., Brocca, L., Bulgin, C. E., Burdanowitz, J., Calbet, X., Donner, R. V., Ghent, D., Gruber, A., Kaminski, T., Kinzel, J., Klepp, C., Lambert, J.-C., Schaeppman-Strub, G., Schröder, M., and Verhoelst, T.: Validation practices for satellite-based Earth observation data across communities, *Rev. Geophys.*, 55, 2017RG000562, <https://doi.org/10.1002/2017RG000562>, 2017.
- Lorente, A., Folkert Boersma, K., Yu, H., Dörner, S., Hilboll, A., Richter, A., Liu, M., Lamsal, L. N., Barkley, M., De Smedt, I., Van 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 NO₂ and HCHO satellite retrievals, *Atmos. Meas. Tech.*, 10, 759–782, <https://doi.org/10.5194/amt-10-759-2017>, 2017.
- Lorente Delgado, A.: From photon paths to pollution plumes: better radiative transfer calculations to monitor NO_x emissions with OMI and TROPOMI, Wu thesis, 9789463439329, Wageningen University, Wageningen, available at: <http://edepot.wur.nl/474563> (last access: 20 April 2020), 2019.
- Ma, J. Z., Beirle, S., Jin, J. L., Shaiganfar, R., Yan, P., and Wagner, T.: Tropospheric NO₂ vertical column densities over Beijing: results of the first three years of ground-based MAX-DOAS measurements (2008–2011) and satellite validation, *Atmos. Chem. Phys.*, 13, 1547–1567, <https://doi.org/10.5194/acp-13-1547-2013>, 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, <https://doi.org/10.1002/2014JD022913>, 2015.
- Mayer, B. and Kylling, A.: Technical note: The libRadtran software package for radiative transfer calculations – description and examples of use, *Atmos. Chem. Phys.*, 5, 1855–1877, <https://doi.org/10.5194/acp-5-1855-2005>, 2005.
- Myhre, G., Shindell, D., Bréon, F.-M., Collins, W., Fuglestedt, J., Huang, J., Koch, D., Lamarque, J.-F., Lee, D., Mendoza, B., Nakajima, T., Robock, A., Stephens, G., Takemura, T., and Zhang, H.: Anthropogenic and Natural Radiative Forcing, in: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, edited by: Stocker, T., Qin, D., Plattner, G.-K., Tignor, M., Allen, S., Boschung, J., Nauels, A., Xia, Y., Bex, V., and Midgley, P., chap. 8, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2013.
- 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 Sens.*, 10, 1254, <https://doi.org/10.3390/rs10081254>, 2018.
- Noxon, J. F.: Stratospheric NO₂: 2. Global behavior, *J. Geophys. Res.-Oceans*, 84, 5067–5076, <https://doi.org/10.1029/JC084iC08p05067>, 1979.
- Ortega, I., Koenig, T., Sinreich, R., Thomson, D., and Volkamer, R.: The CU 2-D-MAX-DOAS instrument – Part 1: Retrieval of

- 3-D distributions of NO₂ and azimuth-dependent OVOC ratios, *Atmos. Meas. Tech.*, 8, 2371–2395, <https://doi.org/10.5194/amt-8-2371-2015>, 2015.
- Peters, E., Pinardi, G., Seyler, A., Richter, A., Wittrock, F., Bösch, T., Van Roozendaal, M., Hendrick, F., Drosoglou, T., Bais, A. F., Kanaya, Y., Zhao, X., Strong, K., Lampel, J., Volkamer, R., Koenig, T., Ortega, I., Puentedura, O., Navarro-Comas, M., Gómez, L., Yela González, M., PETERS, A., Remmers, J., Wang, Y., Wagner, T., Wang, S., Saiz-Lopez, A., García-Nieto, D., Cuevas, C. A., Benavent, N., Querel, R., Johnston, P., Postlyakov, O., Borovski, A., Elokhov, A., Bruchkouski, I., Liu, H., Liu, C., Hong, Q., Rivera, C., Grutter, M., Stremme, W., Khokhar, M. F., Khayyam, J., and Burrows, J. P.: Investigating differences in DOAS retrieval codes using MAD-CAT campaign data, *Atmos. Meas. Tech.*, 10, 955–978, <https://doi.org/10.5194/amt-10-955-2017>, 2017.
- Petrutoli, A., Bonasoni, P., Giovanelli, G., Ravegnani, F., Kostadinov, I., Bortoli, D., Weiss, A., Schaub, D., Richter, A., and Fortezza, F.: First comparison between ground-based and satellite-borne measurements of tropospheric nitrogen dioxide in the Po basin, *J. Geophys. Res.-Atmos.*, 109, D15307, <https://doi.org/10.1029/2004JD004547>, 2004.
- Pinardi, G., Van Roozendaal, M., Lambert, J.-C., Granville, J., Hendrick, F., Tack, F., Yu, H., Cede, A., Kanaya, Y., Irie, H., Goutail, F., Pommereau, J.-P., Pazmino, A., Wittrock, F., Richter, A., Wagner, T., Gu, M., Remmers, J., Friess, U., Vlemmix, T., PETERS, A., Hao, N., Tiefengraber, M., Herman, J., Abuhassan, N., Bais, A., Kouremeti, N., Hovila, J., Holla, R., Chong, J., Postlyakov, O., and Ma, J.: GOME-2 total and tropospheric NO₂ validation based on zenith-sky, direct-sun and multi-axis doas network observations, in: EUMETSAT Conference, 22–26 September 2014, Geneva, Switzerland, 2014.
- Pinardi, G., Van Roozendaal, M., Hendrick, F., Theys, N., Abuhasan, N., Bais, A., Boersma, F., Cede, A., Chong, J., Donner, S., Drosoglou, T., Frieß, U., Granville, J., Herman, J. R., Eskes, H., Holla, R., Hovila, J., Irie, H., Kanaya, Y., Karagiozidis, D., Kouremeti, N., Lambert, J.-C., Ma, J., Peters, E., PETERS, A., Postlyakov, O., Richter, A., Remmers, J., Takashima, H., Tiefengraber, M., Valks, P., Vlemmix, T., Wagner, T., and Wittrock, F.: Validation of tropospheric NO₂ column measurements of GOME-2A and OMI using MAX-DOAS and direct sun network observations, *Atmos. Meas. Tech. Discuss.*, <https://doi.org/10.5194/amt-2020-76>, in review, 2020.
- Platt, U. and Stutz, J.: *Differential Optical Absorption Spectroscopy: Principles and Applications*, Springer, <https://doi.org/10.1007/978-3-540-75776-4>, 2008.
- Pommereau, J. and Goutail, F.: O₃ and NO₂ ground-based measurements by visible spectrometry during Arctic winter and spring 1988, *Geophys. Res. Lett.*, 15, 891–894, <https://doi.org/10.1029/GL015i008p00891>, 1988.
- QA4EO: Quality Assurance Framework for Earth Observation - The Guide, Tech. rep., available at: http://qa4eo.org/docs/QA4EO_guide.pdf, last access: 23 September 2019.
- Richter, A., Godin, S., Gomez, L., Hendrick, F., Hocke, K., Lange-rock, B., van Roozendaal, M., and Wagner, T.: EC FP7 NORS Technical Note D4.4 – Spatial Representativeness of NORS observations, Tech. rep., Institute of Environmental Physics, University of Bremen, Bremen, 2013a.
- Richter, A., Weber, M., Burrows, J., Lambert, J.-C., and van Gijssel, A.: Validation strategy for satellite observations of tropospheric reactive gases, *Ann. Geophys.*, 56, <https://doi.org/10.4401/ag-6335>, 2013b.
- Richter, A., Bais, A., Dils, B., Gielen, C., Hendrick, F., Pinardi, G., Peters, E., PETERS, A., Remmers, J., Wagner, T., Wang, S., and Wang, Y.: Quality indicators on uncertainties and representativity of atmospheric reference data, techreport QA4ECV Deliverable D3.9, Belgian Institute for Space Aeronomy, available at: <http://www.qa4ecv.eu/sites/default/files/D3.9.pdf> (last access: 20 April 2020), 2016.
- Rino, B., Niemeijer, S., Compennolle, S., and Lambert, J.-C.: Prototype QA/Validation service for Atmosphere ECVs: Web based prototype, QA4ECV report Deliverable D2.6, s[&t] Corporation, available at: http://www.qa4ecv.eu/sites/default/files/QA4ECV_D-2-6_final.pdf (last access: 20 April 2020), 2017.
- Rodgers, C. D.: *Inverse Methods for Atmospheric Sounding*, vol. 2 of Series on Atmospheric, Oceanic and Planetary Physics, World Scientific, Singapore, 2000.
- Rodgers, C. D. and Connor, B. J.: Intercomparison of remote sounding instruments, *J. Geophys. Res.*, 108, 4116, <https://doi.org/10.1029/2002JD002299>, 2003.
- Schaub, D., Boersma, K. F., Kaiser, J. W., Weiss, A. K., Folini, D., Eskes, H. J., and Buchmann, B.: Comparison of GOME tropospheric NO₂ columns with NO₂ profiles deduced from ground-based in situ measurements, *Atmos. Chem. Phys.*, 6, 3211–3229, <https://doi.org/10.5194/acp-6-3211-2006>, 2006.
- Seinfeld, J. H. and Pandis, S. N.: *Atmospheric Chemistry and Physics – From air pollution to climate change*, first edn., John Wiley, & Sons, Hoboken, New Jersey, 1997.
- 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*, 326, 716–718, <https://doi.org/10.1126/science.1174760>, 2009.
- Sillman, S., Logan, J. A., and Wofsy, S. C.: The sensitivity of ozone to nitrogen oxides and hydrocarbons in regional ozone episodes, *J. Geophys. Res.-Atmos.*, 95, 1837–1851, <https://doi.org/10.1029/JD095iD02p01837>, 1990.
- Solomon, S., Schmeltekopf, A. L., and Sanders, R. W.: On the interpretation of zenith sky absorption measurements, *J. Geophys. Res.*, 92, 8311–8319, <https://doi.org/10.1029/JD092iD07p08311>, 1987.
- Spurr, R.: LIDORT and VLIDORT: Linearized pseudo-spherical scalar and vector discrete ordinate radiative transfer models for use in remote sensing retrieval problems, in: *Light Scattering Reviews 3: Light Scattering and Reflection*, edited by: Kokhanovsky, A., pp. 229–275, Springer Berlin Heidelberg, Berlin, Heidelberg, https://doi.org/10.1007/978-3-540-48546-9_7, 2008.
- Strahan, S., Douglass, A., and Newman, P.: The contributions of chemistry and transport to low arctic ozone in March 2011 derived from Aura MLS observations, *J. Geophys. Res.*, 118, 1563–1576, <https://doi.org/10.1002/jgrd.50181>, 2013.
- Tirpitz, J.-L., Frieß, U., Hendrick, F., Alberti, C., Allaart, M., Apituley, A., Bais, A., Beirle, S., Berkhout, S., Bognar, K., Bösch, T., Bruchkouski, I., Cede, A., Chan, K. L., den Hoed, M., Donner, S., Drosoglou, T., Fayt, C., Friedrich, M. M., Frumau, A., Gast, L., Gielen, C., Gomez-Martín, L., Hao, N., Hensen, A., Henzing, B., Hermans, C., Jin, J., Kreher, K., Kuhn, J., Lampel, J., Li, A., Liu,

- C., Liu, H., Ma, J., Merlaud, A., Peters, E., Pinardi, G., Piter, A., Platt, U., Puentedura, O., Richter, A., Schmitt, S., Spinei, E., Stein Zweers, D., Strong, K., Swart, D., Tack, F., Tiefengraber, M., van der Hoff, R., van Roozendaal, M., Vlemmix, T., Vonk, J., Wagner, T., Wang, Y., Wang, Z., Wenig, M., Wiegner, M., Wittrock, F., Xie, P., Xing, C., Xu, J., Yela, M., Zhang, C., and Zhao, X.: Intercomparison of MAX-DOAS vertical profile retrieval algorithms: studies on field data from the CINDI-2 campaign, *Atmos. Meas. Tech. Discuss.*, <https://doi.org/10.5194/amt-2019-456>, in review, 2020.
- Vandaele, A. C., Fayt, C., Hendrick, F., Hermans, C., Humbled, F., Van Roozendaal, M., Gil, M., Navarro, M., Puentedura, O., Yela, M., Braathen, G., Stebel, K., Tørnkqvist, K., Johnston, P., Kreher, K., Goutail, F., Mieville, A., Pommereau, J.-P., Khaikine, S., Richter, A., Oetjen, H., Wittrock, F., Bugarski, S., Frieß, U., Pfeilsticker, K., Sinreich, R., Wagner, T., Corlett, G., and Leigh, R.: An intercomparison campaign of ground-based UV-visible measurements of NO₂, BrO, and OClO slant columns: Methods of analysis and results for NO₂, *J. Geophys. Res.-Atmos.*, 110, D08305, <https://doi.org/10.1029/2004JD005423>, 2005.
- Veefkind, J. P., de Haan, J. F., Sneep, M., and Levelt, P. F.: Improvements to the OMI O₂-O₂ operational cloud algorithm and comparisons with ground-based radar-lidar observations, *Atmos. Meas. Tech.*, 9, 6035–6049, <https://doi.org/10.5194/amt-9-6035-2016>, 2016.
- Verhoelst, T. and Lambert, J. C.: Generic metrology aspects of an atmospheric composition measurement and of data comparisons. EC Horizon2020 GAIA-CLIM technical Report / Deliverable D3.2, Tech. rep., BIRA-IASB, available at: <http://www.gaia-clim.eu/page/deliverables> (last access: 20 April 2020), 2016.
- Verhoelst, T., Granville, J., Hendrick, F., Köhler, U., Lerot, C., Pommereau, J.-P., Redondas, A., Van Roozendaal, M., and Lambert, J.-C.: Metrology of ground-based satellite validation: co-location mismatch and smoothing issues of total ozone comparisons, *Atmos. Meas. Tech.*, 8, 5039–5062, <https://doi.org/10.5194/amt-8-5039-2015>, 2015.
- Vlemmix, T., Hendrick, F., Pinardi, G., De Smedt, I., Fayt, C., Hermans, C., Piter, A., Wang, P., Levelt, P., and Van Roozendaal, M.: MAX-DOAS observations of aerosols, formaldehyde and nitrogen dioxide in the Beijing area: comparison of two profile retrieval approaches, *Atmos. Meas. Tech.*, 8, 941–963, <https://doi.org/10.5194/amt-8-941-2015>, 2015.
- von Clarmann, T.: Validation of remotely sensed profiles of atmospheric state variables: strategies and terminology, *Atmos. Chem. Phys.*, 6, 4311–4320, <https://doi.org/10.5194/acp-6-4311-2006>, 2006.
- Wang, Y., Beirle, S., Lampel, J., Koukouli, M., De Smedt, I., Theys, N., Li, A., Wu, D., Xie, P., Liu, C., Van Roozendaal, M., Stavrou, T., Müller, J.-F., and Wagner, T.: Validation of OMI, GOME-2A and GOME-2B tropospheric NO₂, SO₂ and HCHO products using MAX-DOAS observations from 2011 to 2014 in Wuxi, China: investigation of the effects of priori profiles and aerosols on the satellite products, *Atmos. Chem. Phys.*, 17, 5007–5033, <https://doi.org/10.5194/acp-17-5007-2017>, 2017.
- Wenig, M. O., Cede, A. M., Bucsela, E. J., Celarier, E. A., Boersma, K. F., Veefkind, J. P., Brinksma, E. J., Gleason, J. F., and Herman, J. R.: Validation of OMI tropospheric NO₂ column densities using direct-Sun mode Brewer measurements at NASA Goddard Space Flight Center, *J. Geophys. Res.*, 113, D16S45, <https://doi.org/10.1029/2007jd008988>, 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.
- World Health Organization: Review of evidence on health aspects of air pollution – REVIHAAP Project, Tech. rep., World Health Organization, Copenhagen, Denmark, available at: <http://www.euro.who.int/> (last access: 23 September 2019), 2013.
- Zara, M., Boersma, K. F., van Geffen, J., and Eskes, H.: An improved temperature correction for OMI NO₂ slant column densities from the 405–465 nm fitting window – TN-OMIE-KNMI-982, Tech. rep., KNMI, De Bilt, the Netherlands, available at: <https://kfolkertboersma.files.wordpress.com/2019/09/tn-omie-knmi-982.pdf> (last access: 20 April 2020), 2017.
- 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.
- Zhou, Y., Brunner, D., Boersma, K. F., Dirksen, R., and Wang, P.: An improved tropospheric NO₂ retrieval for OMI observations in the vicinity of mountainous terrain, *Atmos. Meas. Tech.*, 2, 401–416, <https://doi.org/10.5194/amt-2-401-2009>, 2009.