



Validation of tropospheric NO₂ columns from
OMI and SCIAMACHY with MAX-DOAS ground-based
measurements in Tai' an and De Bilt

Author: Qingyuan Gan

Supervisor: Folkert Boersma

Examiner: Maarten Krol

NOVEMBER 2, 2018

WAGENINGEN UNIVERSITY AND RESEARCH CENTRE
qingyuan.gan@wur.nl

Abstract: In this study, three satellite NO₂ retrievals: QA4ECV OMI, DOMINO v2 and SCIAMACHY (GOME2-A) are validated with MAX-DOAS measurements in China (Tai' an) and in The Netherlands (De Bilt) in June 2006 and July 2014. Through this validation, we can quantify the NO₂ retrievals uncertainty of these three satellites and have a better understanding on their performance at different regions. The results indicate that QA4ECV OMI is an improvement over DOMINO v2. QA4ECV OMI NO₂ retrievals have the smallest absolute mean bias Tai' an: 6%; De Bilt: 17%) among all three satellite products in both Tai' an and De Bilt. For DOMINO v2, the absolute mean bias is slightly higher compare to QA4ECV OMI (Tai' an: 7%; De Bilt: 26%). In addition, from the observed differences between MAX-DOAS and satellites, all three satellite products were in better agreement with MAX-DOAS in Tai' an than in De Bilt. This can be explained because De Bilt is close to a strong NO₂ source (the city of Utrecht and nearby highways) that influenced the MAX-DOAS uncertainties. SCIAMACHY and GOME2-A could not be properly validated with MAX-DOAS within a short duration such as one month because there are not enough samples can be found with a satisfactory quality.

Key words: QA4ECV OMI, DOMINO v2, SCIAMACHY, GOME2-A, MAX-DOAS, NO₂ retrievals, validation.

Contents

1. Introduction	3
2. Methodology	6
2.1. Satellite NO ₂ retrievals.....	6
2.2. Data selection.....	7
2.3. Data correction.....	9
2.4. Statistical analysis.....	11
3. Results	12
3.1. Tai' an.....	12
3.1.1. Satellite products' performance validated with MAX-DOAS	12
3.1.2. Uncertainties and probability distribution.....	20
3.2. De Bilt.....	23
3.2.1. Satellite products' performance validated with MAX-DOAS	23
3.2.2. Uncertainties and probability distribution.....	34
3.3 Comparison of the Tai' an and De Bilt cases	36
4. Conclusion and recommendation	39
References	41

1. Introduction

The measurement of trace gases in the atmosphere is of great significance when monitoring air quality and for understanding the effects of radiative forcing and its impact on climate. At present, the total amount of pollutant emissions and their impact on the local, regional and even global scales are still uncertain (Han et al., 2011; Park et al., 2014; Han et al., 2015). Nitrogen oxides ($\text{NO}_x = \text{NO} + \text{NO}_2$) participate in chemical reactions that lead to the formation of ozone in the troposphere and are responsible for smog and acid rain. As a main production of fuels combustion, an increasing amount of NO_x gases was produced in the boundary layer by industrial processes, power generation, transport and biomass burning over pollution hotspots in twentieth century (Kanter, David, 2018). NO_2 is one of nitrogen oxides, which is also considered as an important atmospheric pollutant affecting human health. Short-term exposure to NO_2 can cause airway responses, and long-term exposure to NO_2 can cause dysfunction of the human immune system and respiratory tract infections (US Environmental Protection Agency, 2003).

Satellite observation is widely considered one of the important tools for quantitatively evaluating the long-term spatial and temporal distribution characteristics of atmospheric composition (Beirle et al., 2003). The characteristics of emissions and changes in China, especially in north China, have been studied extensively using satellite data for NO_2 (Richter et al., 2005; van der A et al., 2006; Wang et al., 2012; Hilboll et al., 2013). However, the retrieval of satellite data is affected by several error sources, which can lead to increased uncertainty, bias and deviation, especially in areas close to the ground with high pollution levels. Considering that NO_2 is an important trace gas and NO_2 vertical density columns (VCDs) are dominated by their amounts in the planetary boundary layer (over 90% over highly polluted areas; Ma et al., 2006), comprehensive ground measurements of NO_2 tropospheric VCDs are essential for evaluating and validating satellite retrievals (Theys et al., 2015).

Multi-axis differential optical absorption spectroscopy (MAX-DOAS) is a surface remote sensing technology that uses scattered sunlight measured from different, mostly tilted, elevation angles to obtain relatively high sensitivity to trace gases and aerosols' VCDs and profiles in the lower atmosphere (Honniger et al., 2004b; Wagner et al., 2004; Wittrock et al., 2004; Pikel'naya et al., 2007; Theys et al., 2007; Platt and Stutz, 2008). MAX-DOAS has been widely used around the world to obtain column information of NO_2 and some other pollutants (Vlemmix et al., 2010). In many cases, strong agreement has been achieved between independent satellite data sets, ground measurements, model calculations and other remote sensing instruments (Valks et al., 2011). However, the difference between MAX-DOAS and satellite NO_2 measurements was found to be particularly strong at highly

polluted areas such as Beijing megacity (Ma et al., 2013). In Ma's study, compared with the MAX-DOAS measurements, Scanning Imaging Absorption Spectrometer for Atmospheric Chartography (SCIAMACHY) and Ozone Monitoring Instrument (OMI) (DOMINO v2.0 and DOMINO v1.02) satellite retrievals underestimated the tropospheric NO₂ VCDs over Beijing by a magnitude of 43% and 26–38% respectively. These differences mainly related to two effects commonly referred to as the gradient smoothing effect and the aerosol shielding effect (Chen et al., 2009; Ma et al., 2013; Lin et al., 2014; Chan et al., 2015). The satellite retrievals could underestimate the tropospheric NO₂ VCDs that partly caused by the strong spatial gradient of NO₂ near NO₂ sources, which are not resolved by the satellite observations (Chen et al., 2009). In addition, the aerosol shielding effect also could play a role. The aerosols can scatter or absorb solar radiation, thus affecting the radiation field in retrieving both NO₂ and clouds. This aerosol effect results in an unclarity in the net effect in the retrieved NO₂ columns (Lin et al., 2014). There are relatively fewer studies investigating smaller polluted cities such as Tai'an (north-east China) and De Bilt (central Netherlands) and comparing validation results of satellites using MAX-DOAS measurements over Eastern Asia and Western Europe.

Previous studies often used the conventional data set DOMINO v2 as OMI NO₂ satellite retrievals (Boersma et al., 2011) to validate with MAX-DOAS measurements. However, for this paper, we also used the new NO₂ data sets from the QA4ECV-project (www.qa4ecv.eu) to do the validation with MAX-DOAS measurements. For different resolutions from OMI, we used SCIAMACHY and Global Ozone Monitoring Experiment-2 (GOME 2-A; launched on METOP A) as coarser retrievals in the validation. The first main research question (and two sub-questions) for this paper is as follows:

1. What is the performance of satellite measurements of NO₂ (OMI, SCIAMACHY and GOME2-A) in Tai'an and De Bilt?
 - a. What is the difference between the DOMINO v2 and QA4ECV OMI data sets?
 - b. Can different instrument resolutions be properly validated with MAX-DOAS measurements? If so, how?

Alternatively, although MAX-DOAS observations are almost unaffected by gradient smoothing effect and aerosol shielding effect, Beirle et al. (2003) have demonstrated that European countries have a weekly cycle in NO_x emissions, which leads to a large variation in NO₂ emissions between weekdays and weekends, while China follows no such trend. Furthermore, Vlemmix (2015) has revealed MAX-DOAS NO₂ VCDs measurements can be influenced by wind coming from strong NO₂ sources. Therefore, the second research question is as follows:

2. Do satellite measurements perform similarly in Tai'an and De Bilt? If

not, what are the possible reasons for differences?

Section 2 will provide information on the criteria and methods needed to answer these questions, including the selection of representative data in satellite products, the smoothing of the spatial mismatch between MAX-DOAS measurements and satellite data and the analysis of satellite uncertainties. In section 3, we will analyse the correlation and uncertainties of satellite products and MAX-DOAS observations in two sites—that is, the quantitative validation—and then investigate wind-direction impact, weekly cycles and the pixel centre distance effect at the De Bilt site. Conclusions will be provided in section 4.

2. Methodology

2.1. Satellite NO₂ retrievals

The satellite data in this paper were retrieved from SCIAMACHY, OMI and GOME2-A. SCIAMACHY was launched in 2002 on ENVISAT (Bovensmann et al., 1999) and was stopped in 2012. The instrument has a 30×60 km² footprint with a 10:00-hr average local overpass time. OMI was launched on AURA in 2004 (Levelt & Noordhoek, 2002) and has a 13×24 km² footprint at nadir and on average a 13:40-hr local overpass time at northern hemisphere mid-latitude. The small pixel size enables OMI to look in between the clouds, which is important for retrieving tropospheric information both underneath and above cloud levels. The first GOME2 instrument was launched on METOP-A in 2006 (EUMETSAT, 2008); it has a 40×80 km² footprint and a 9:30-hr average local overpass time.

We compared two OMI products, QA4ECV and DOMINO v2. According to Boersma et al., 2018, the QA4ECV algorithm is improved based on DOMINO v2 algorithm. To obtain tropospheric NO₂ VCDs, the DOMINO v2 (Boersma et al., 2011) and QA4ECV (Boersma et al., 2018) algorithms require three steps: (1) obtain the total NO₂ slant column density (SCD; N_s) from the OMI reflectance spectra; (2) separate the stratospheric ($N_{s, strat}$) and tropospheric NO₂ SCD ($N_s - N_{s, strat}$) from the total NO₂ SCD; (3) convert the tropospheric NO₂ SCD to vertical column density (VCD; $N_{v, trop}$) by dividing by the tropospheric air mass factor (AMF; M_{trop}). The retrieval equation can be expressed as

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

However, there are several differences between the DOMINO v2 and QA4ECV algorithm settings. Table 1 demonstrates the general differences between the two algorithms.

Table 1

The General Differences between QA4ECV and DOMINO v2

	QA4ECV	DOMINO v2
Spectral fitting	QDOAS/NLIN 405 - 465 nm	OMNO2A v1 405 - 465 nm
	Intensity offset correction, liquid water, O ₂ -O ₂ (Irie, et al., 2018)	Only NO ₂ , O ₃ , H ₂ O, Ring, polynomial (Bucsela, et al., 2006)
	Optimized wavelength calibration	
Stratospheric correction	TM5 1° × 1°, nudging to HNO ₃ :O ₃ climatologies from ODIN, major speed-up, versatile code (Irie, Hitoshi, et al., 2018)	TM4 3° × 2°, nudging to HNO ₃ :O ₃ ratios from UARS + O3MSR climatologies (Boersma et al., 2004)
AMF LUT	More reference points (437.5 nm) 173 × 14 × 26 × 10 × 11 × 16 (Müller et al., 2016)	Fewer reference points (439 nm) 24 × 10 × 10 × 10 × 13 × 16 (Palmer et al., 2001; Boersma et al., 2004)
	Sphericity corrected	Pseudospherical
Surface albedo	OMI LER 5-year	OMI LER 3-year
Clouds	improved O ₂ -O ₂	O ₂ -O ₂
A priori profile	TM5-MP 1° × 1°	TM4 3° × 2°
Terrain height	Pixel average, GMTED2010	Centre of DEM_3KM

As the MAX-DOAS instrument was only in Tai' an for 1 month in June 2006, we only had data for that month to compare to the satellite data. For De Bilt, to make a representative comparison to Tai' an, we chose July 2014, the same season and for the same amount of time as in Tai' an. Because SCIAMACHY only had data until 2012, we chose GOME2-A as a lower resolution sample in the De Bilt case to compare with the relatively higher resolution satellite OMI. In general, we evaluated two different approaches (DOMINO v2 and QA4ECV): OMI NO₂ VCDs as high-resolution samples and SCIAMACHY NO₂ VCDs as low-resolution samples for Tai' an, and OMI NO₂ VCDs as high-resolution samples and GOME2-A NO₂ VCDs as low-resolution samples for De Bilt.

2.2. Data selection

MAX-DOAS measurements were recorded at fixed locations (36.16 ° N, 117.15 ° E; De Bilt site: 52.10 ° N, 5.18 ° E); however, as the location of satellite retrievals is dependent on the satellite's orbit and exact viewing geometry relative to the surface, satellite retrievals are not recorded at the exact same location of the MAX-DOAS instrument. Moreover, the MAX-DOAS instrument and satellite may sample different air masses. MAX-DOAS measurements and satellite retrievals are also not synchronized in time. The mismatch in both spatial and

temporal dimensions leads to an unavoidable mismatch between MAX-DOAS measurements and satellite retrievals. Due to this mismatch, filters that select relatively representative samples from both satellite retrievals and MAX-DOAS measurements are required. In this section, the filters that select data from three satellite products and MAX-DOAS measurements will be introduced. There are three schemes based on different resolutions of each satellite: the OMI scheme, the SCIAMACHY scheme and the GOME2 scheme. By applying these filters, we can select the representative data needed to validate the QA4ECV OMI, QA4ECV SCIAMACHY and DOMINO v2 tropospheric NO₂ VCDs with the MAX-DOAS VCDs. The OMI scheme is for the high-resolution satellite products QA4ECV OMI and DOMINO v2; it consists of five limitations: (1) satellite pixel centre within 20 km to MAX-DOAS site; (2) satellite retrieval taken within 30 min of the MAX-DOAS measurement; (3) satellite retrieval recorded when cloud radiance fraction lower than 0.5; (4) satellite pixel area smaller than 700 km²; (5) satellite retrieval recorded when cloud pressure lower than 875 hpa. The SCIAMACHY scheme and the GOME2 scheme are applied to QA4ECV SCIAMACHY and GOME2-A, respectively. For these two lower resolution satellite products, we need to relax the restriction of the distance between the MAX-DOAS instrument and the pixel centre to 40 km and 60 km, respectively. Moreover, because we only ran a 1-month validation for both study areas, to find as many relatively representative samples in QA4ECV SCIAMACHY and GOME2-A validation as possible, we excluded the pixel area limitation for these two satellite products.

In consideration of the spatial dimensions of MAX-DOAS (± 10 km) and OMI (the nadir pixels have a 13- to 24-km footprint) footprints (Boersma et al., 2018), OMI NO₂ tropospheric VCDs measured with a pixel centre within 20 km of the location of the MAX-DOAS instrument in the study site were selected for QA4ECV OMI and DOMINO v2 to reduce the spatial mismatch. Due to the coarser resolution, far fewer SCIAMACHY pixels were retrieved near the study site. The coincidence criteria limit the spatial representativeness mismatches between MAX-DOAS and satellite products. According to the suggestions in the Product Specification Document for QA4ECV OMI data (QA4ECV Deliverable D4.6), the satellite retrieval can be strongly influenced by a cloud-screening effect: if clouds are present, it is more difficult to detect the NO₂ underneath the clouds. Therefore, we also included mostly clear skies condition as another criterion: satellite NO₂ column measurements were only used when the cloud radiance fraction was lower than 0.5.

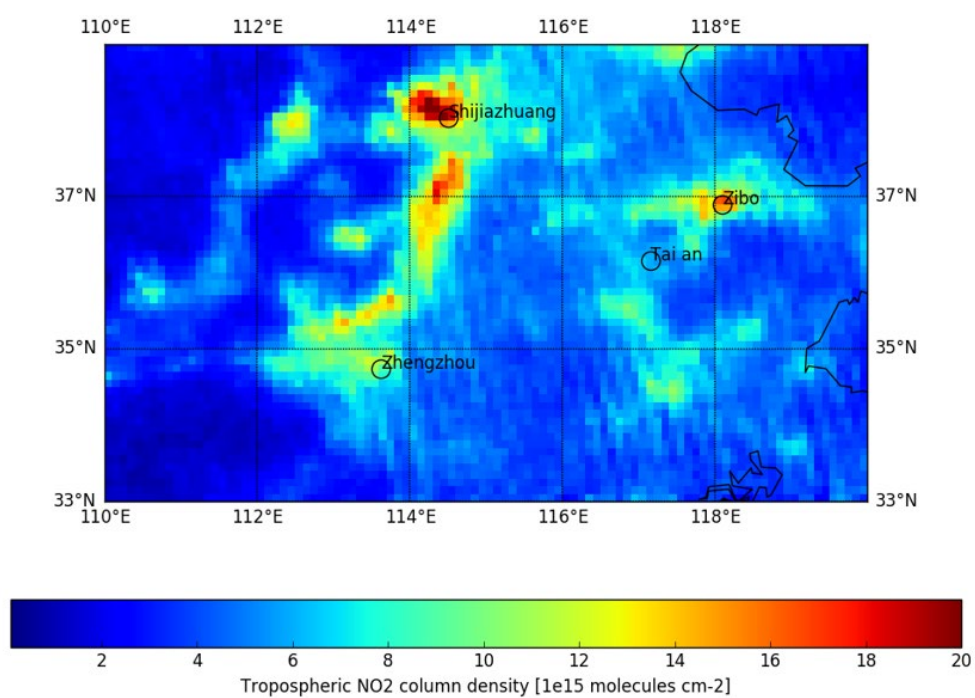
In polluted areas like Tai' an and De Bilt, aerosol concentration are often high in the lower atmosphere. A low aerosol layer can be interpreted as low-hanging clouds by satellites. To reduce aerosol influence on satellite retrieval quality, the satellite NO₂ column measurements were also limited to effective cloud pressure < 875 hpa. To ensure the satellite and MAX-DOAS instrument are measured at least partly same air mass, we also excluded the pixel when the pixel area was < 700 km² in high-resolution cases (QA4ECV OMI and DOMINO v2).

2.3. Data correction

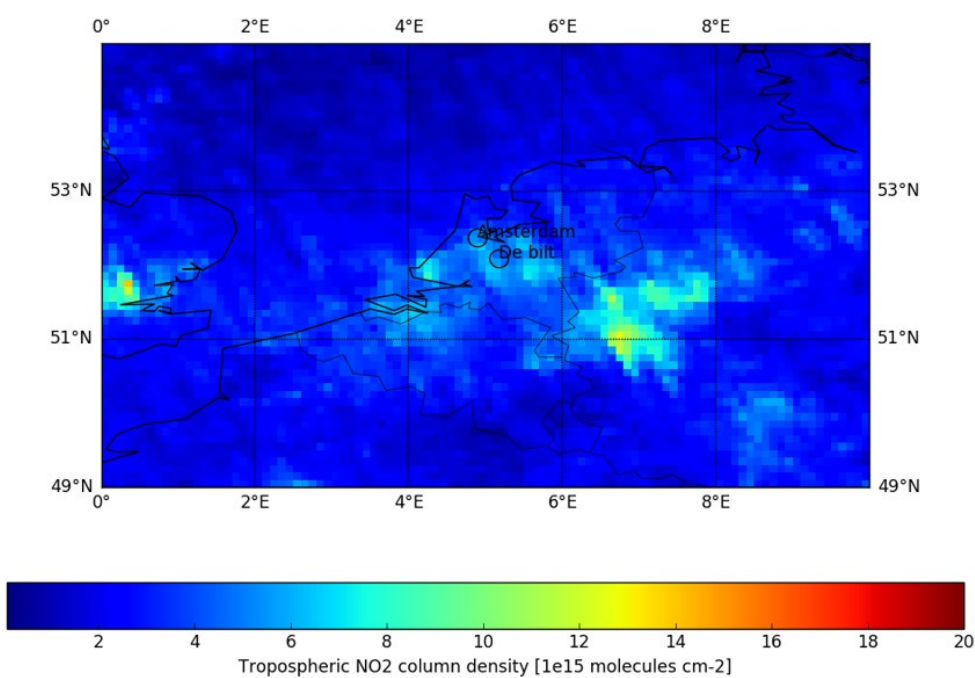
Previous studies (e.g., Drosoglou et al., 2017) have revealed a large divergence between MAX-DOAS and satellite NO₂ VCDs in heavily polluted areas. Spatial inhomogeneity in the NO₂ field near the MAX-DOAS site is one of possible causes of such divergence. To have an overview of systematic spatial discrepancies between the location of the MAX-DOAS instrument and the satellite pixels, we calculated the OMI QA4ECV monthly mean spatial tropospheric NO₂ distribution (Figure 1). From Figure 1, we see that the total amount of monthly mean tropospheric NO₂ VCDs over Tai' an (June 2006) and De Bilt (July 2014) is similar (around 6.0×10^{15} molec/cm²). At Tai' an, the north and east sides have lower tropospheric NO₂ VCDs compared to the south and west sides; and for De Bilt, tropospheric NO₂ VCDs were also lower in the north and east sides that month. That means if a satellite pixel is measured to the north-east of either site, it is likely have lower bias than MAX-DOAS measurements, and if it is measured at south-west side, the situation is opposite. The spatial inhomogeneity in the NO₂ field is needed to consider in order to avoid attributing the bias to satellite errors. To smooth this spatial inhomogeneity in the NO₂ field, we applied as a correction factor the ratio of monthly mean tropospheric NO₂ VCDs at MAX-DOAS instrument location ($\overline{N_{V,T}}$) and monthly mean tropospheric NO₂ VCDs at the location of individual satellite pixels ($\overline{N_{V,P}}$). By multiplying the individual satellite tropospheric NO₂ VCDs ($N_{V,P}$) with this correction factor, we obtained a more representative satellite tropospheric NO₂ VCDs at the MAX-DOAS instrument location ($N_{V,T}$; Irie, et al., 2018):

$$N_{V,T} = \frac{\overline{N_{V,T}}}{\overline{N_{V,P}}} \cdot N_{V,P} \quad (2)$$

It should be noted that we used the QA4ECV OMI monthly mean tropospheric NO₂ VCDs for all the satellite products' correction factor calculations due to data availability. This could result in some bias in the application of the correction factor for DOMINOv2, SCIAMACHY and GOME2-A.



(a)



(b)

Figure 1. Monthly mean of QA4ECV tropospheric NO₂ VCDs for (a) Tai'an in June 2006 and (b) De Bilt in July 2014. The resolution is $0.1^\circ \times 0.1^\circ$. Only pixels measured under mostly clear-

sky conditions were used to compute the average.

2.4. Statistical analysis

To validate the satellite retrievals, we made a comparison of co-located, coincident satellite and MAX-DOAS measurements. As part of the validation effort, statistical evaluations of the discrepancies, such as regression analyses, were made. For the regression analyses, each corrected satellite NO₂ column was paired with a MAX-DOAS measurement that met the criteria described in section 2.2, which allowed us to evaluate the correlation between satellite retrieval and MAX-DOAS tropospheric NO₂ column density. Furthermore, to understand the differences between satellite and MAX-DOAS tropospheric NO₂ column density, the mean bias and root mean squared deviation (RMSD) for each of the satellite products at the two sites were calculated. In the traditional, standard ordinary least squares regression, discrepancies between the MAX-DOAS and satellite columns are minimized exclusively in the ‘vertical’ distance (along the y axis). We used a reduced major axis (RMA) regression analysis to minimize the discrepancies. The RMA regression assumes that there are errors in both y and x . What is minimized is the product of the ‘ y -distance’ and ‘ x -distance’ of observations to the line. Therefore, the total error is the area of the triangle between the observation and the minimized line. The RMA regression is suitable when the symmetrical relationship between two variables is required. The mean bias was calculated as:

$$MB = \frac{1}{n} \sum_{i=1}^n (NO2_{satellite,i} - NO2_{MAX-DOAS,i}) \quad (3)$$

The root mean squared deviation was calculated by:

$$RMSD = \sqrt{\frac{1}{n} \sum_{i=1}^n (NO2_{MAX-DOAS,i} - NO2_{OMI,i})^2} \quad (4)$$

By analysing the differences between satellite and MAX-DOAS tropospheric NO₂ VCDs, we can also assess the uncertainties of the satellite retrievals. This is based on the MAX-DOAS uncertainty (10%) (Boersma et al., 2011) and the well-established uncertainty of the satellite NO₂ VCD (Boersma, et al 2004, 2011). Assuming that the retrieval uncertainties between satellite and MAX-DOAS are independent and follow a normal distribution, the distribution of differences between satellite and MAX-DOAS is expected as a Gaussian form characterized by width (Boersma et al., 2018):

$$\sigma = (\sigma_o^2 + \sigma_{MD}^2 + \sigma_R^2)^{1/2} \quad (5)$$

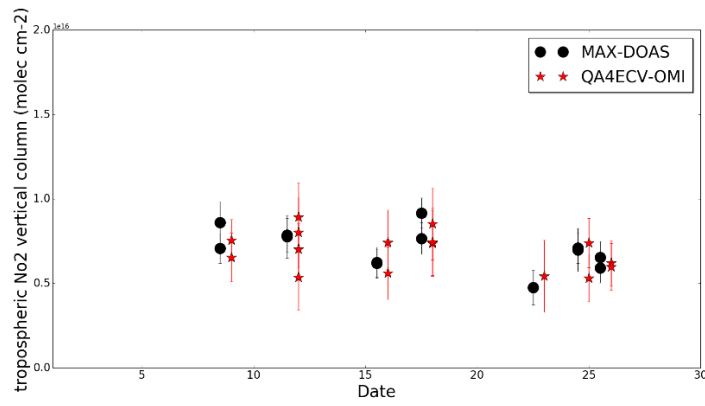
In this equation, σ_o is the uncertainty recorded in the satellite data files for tropospheric vertical NO₂ VCDs, σ_{MD} is the uncertainty recorded in MAX-DOAS data files for MAX-DOAS NO₂ VCDs and σ_R is the uncertainty from spatiotemporal mismatches between satellite and ground measurement. We tried many combinations of the spatiotemporal filters to minimize σ_R . For instance, Irie et al. (2012) have suggested limiting the spatial difference between the satellite pixel and MAX-DOAS measurement to within 0.5° (i.e., latitude and longitude difference between satellite pixel centre and MAX-DOAS site lower than 0.5°); based on that, they found a relatively small σ_R ($\pm 14\%$). In our case, we made a stricter filter for spatial difference to further reduce the mismatch error σ_R . However, we should still expect contributions from remaining mismatches between the samples, since the MAX-DOAS and satellite footprint will never sample the exact same air mass.

3. Results

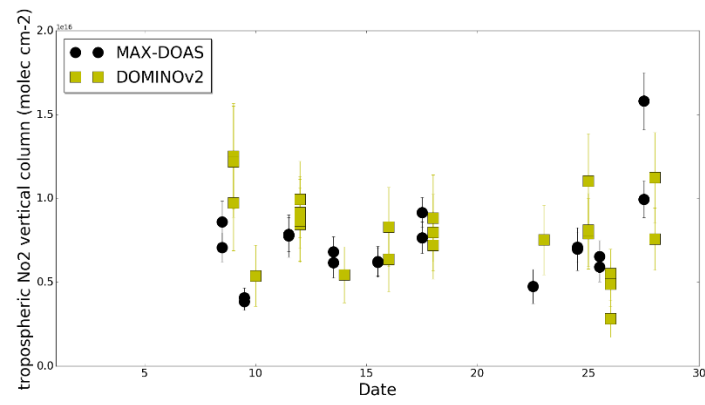
3.1. Tai'an

3.1.1. Satellite products' performance validated with MAX-DOAS

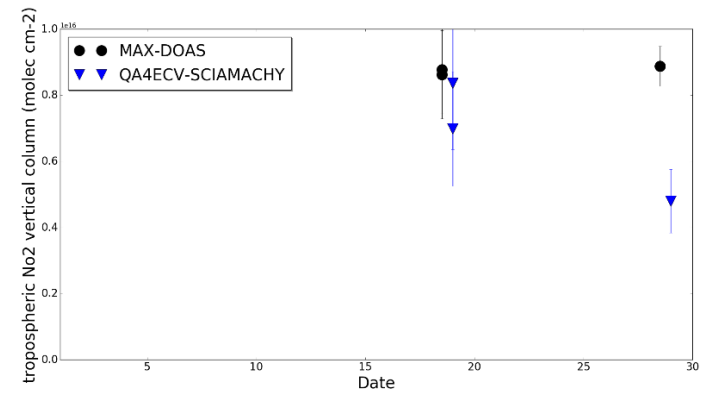
Under the criteria described in the previous section, for QA4ECV OMI, we find 31 pixels matching 13 independent MAX-DOAS measurements selected over 7 different days in Tai'an; for DOMINO v2, we find 45 pixels matching 19 independent MAX-DOAS measurements selected over 10 different days; and for SCIAMACHY, we find five pixels matching with three independent MAX-DOAS measurements selected over 2 different days (Figure 2).



(a)



(b)

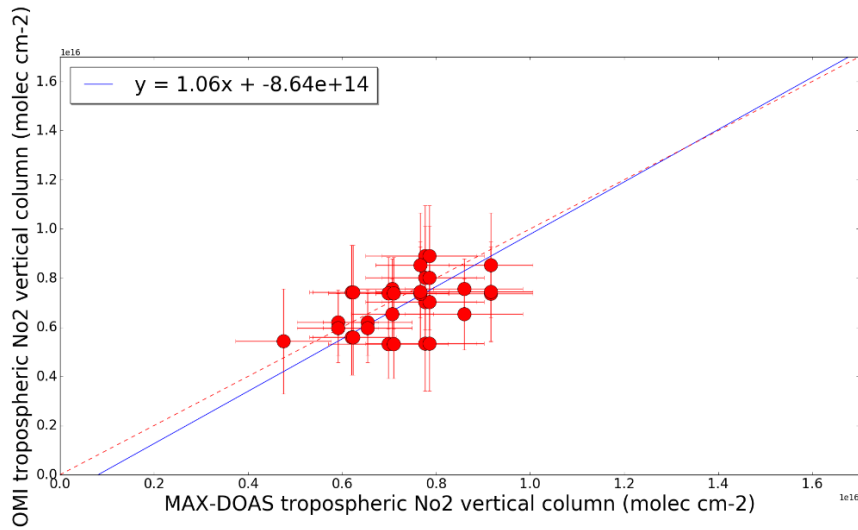


(c)

Figure 2. (a) depicts 31 QA4ECV OMI pixels matching 13 MAX-DOAS measurements over 7 different days in June 2006. (b) depicts 45 DOMINO v2 pixels matching 19 MAX-DOAS measurements over 10 different days in June 2006. (c) depicts five QA4ECV

SCIAMACHY pixels matching three MAX-DOAS measurements over 2 different days in June 2006. The OMI scheme was used for QA4ECV OMI and DOMINO v2, and the SCIAMACHY scheme was used for QA4ECV SCIAMACHY. See section 2 for a description of these schemes. To set them apart from the satellite samples, MAX-DOAS samples in Figure 2 have been moved to 12 hours earlier.

Figure 3 displays the scatterplot of satellite products and MAX-DOAS tropospheric NO₂ VCDs for Tai'an. The relationship between QA4ECV OMI (y) and MAX-DOAS NO₂ VCDs (x) can be expressed as $y = -0.86 \times 10^{15} \text{ molec/cm}^2 + 1.06x$ ($R^2 = 0.06$, $n = 31$). For DOMINO v2 (y) and MAX-DOAS (x), the relationship can be expressed as $y = -0.07 \times 10^{15} \text{ molec/cm}^2 + 1.08x$ ($R^2 = 0.16$, $n = 45$). Due to fewer retained SCIAMACHY samples, a scatterplot illustrating the relationship between SCIAMACHY NO₂ VCDs and MAX-DOAS VCDs would not make sense; instead, we have provided a table for illustration (Table 2).



(a)

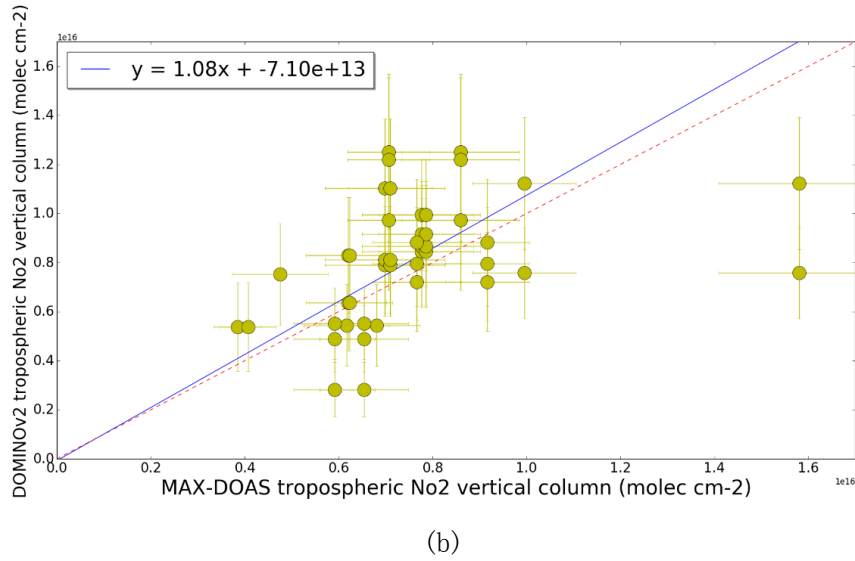


Figure 3. Scatterplot of (a) QA4ECV OMI versus MAX-DOAS tropospheric NO₂ VCDs for Tai'an (China) in June 2006 ($R^2 = 0.06$, $n = 31$), (b) DOMINO v2 OMI versus MAX-DOAS tropospheric NO₂ VCDs for Tai'an (China) in June 2006 ($R^2 = 0.16$, $n = 45$). The blue line in (a) and (b) indicates the result of a reduced major axis regression to the data. The OMI scheme was used. Satellite samples were corrected by applied Eq (2).

Table 2

SCIAMACHY Versus MAX-DOAS Tropospheric NO₂ VCDs for Tai'an (China) in June 2006

QA4ECV SCIAMACHY (molec/cm ²)	MAX-DOAS (molec/cm ²)	Difference (molec/cm ²)	Relative difference (molec/cm ²)
6.98E+15	8.77E+15	-1.79E+15	-20%
8.37E+15	8.77E+15	-4.01E+14	-5%
6.98E+15	8.62E+15	-1.64E+15	-19%
8.37E+15	8.62E+15	-2.50E+14	-3%
4.80E+15	8.88E+15	-4.08E+15	-46%

Note. $n = 5$; SCIAMACHY scheme.

Table 3 demonstrates the mean bias and RMSD between three satellite products and MAX-DOAS measurements. We found a mean bias of -0.39×10^{15} (-6%) molec/cm² for QA4ECV OMI tropospheric NO₂ VCDs. This means that QA4ECV OMI NO₂ tropospheric VCDs tended to be slightly lower than the MAX-DOAS measurements in June 2006 in

Tai'an. The mean bias for DOMINO v2 is 0.55×10^{15} (7%) molec/cm², which means that DOMINO v2 NO₂ tropospheric VCDs tended to be higher than MAX-DOAS measurements in June 2006 in Tai'an. For SCIAMACHY, the mean bias is -1.63×10^{15} (-23%) molec/cm².

Table 3

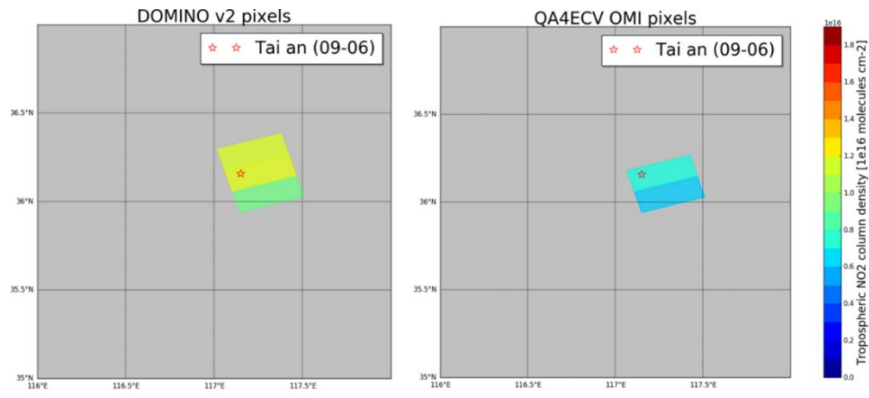
Mean Bias, Root Mean Squared Deviation, Mean Correction Factor and R^2 of QA4ECV OMI, DOMINO v2 and SCIAMACHY Versus MAX-DOAS in Tai'an. "Corrected row" is calculated by applying Eq (2).

	QA4ECV OMI	DOMINO v2	QA4ECV SCIAMACHY
Corrected mean bias (molec/cm ²)	-0.39E+15	0.55E+15	-1.63E+15
Uncorrected mean bias (molec/cm ²)	-0.47E+15	0.55E+15	-2.00E+15
Corrected RMSD (molec/cm ²)	1.13E+15	2.57 E+15	2.13E+15
Uncorrected RMSD (molec/cm ²)	1.19E+15	2.46 E+15	2.53E+15
Corrected relative bias (%)	-6%	7%	-23%
Uncorrected relative bias (%)	-7%	7%	-28%
Mean correction factor	1.01	0.99	1.07
Corrected R^2 (-)	0.06	0.16	-
Uncorrected R^2 (-)	0.06	0.15	-

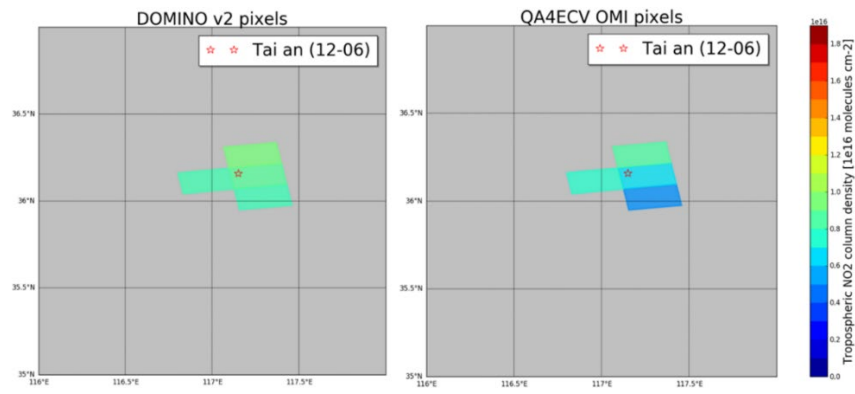
The RMSD for the three satellites' tropospheric VCDs versus MAX-DOAS measurements are 1.13×10^{15} molec/cm² (17%), 2.57×10^{15} molec/cm² (32%) and 2.13×10^{15} molec/cm² (30%), respectively. This indicates that the deviations are substantially smaller for QA4ECV OMI than for DOMINO v2 and QA4ECV SCIAMACHY. The scatterplot and RMSD illustrate that QA4ECV OMI has a smaller deviation with MAX-DOAS than DOMINO v2, reflecting either the retrieval improvements in the algorithm or the stricter flagging scheme (Boersma et al., 2018). Due to coarser resolution compared to the other two products, SCIAMACHY pixels have a larger pixel area and are generally farther away from the MAX-DOAS site than the OMI pixels (see Figures 4 and 5). That might be the reason that QA4ECV SCIAMACHY has a larger mean bias and RSMD compared to QA4ECV OMI. In addition, for the 1-month duration, we only found three pixels that met the criteria, which is not likely to be representative of the overall retrieval performance.

As can be inferred from Table 3, the mean correction factor for the three satellite products all stay close to 1, and modifications do not exceed 1×10^{15} molec/cm² (< 15% of Tai' an column). In other words, when the distance difference between the satellite pixel centre and the MAX-DOAS site is smaller than 20 km (or 40 km for SCIAMACHY), the impact of inhomogeneity in the NO₂ field around the site is still limited. It is notable that when we did not apply the correction, the deviation between QA4ECV (both OMI and SCIAMACHY) and MAX-DOAS increased from 1.13×10^{15} molec/cm² to 1.19×10^{15} molec/cm² and from 2.13×10^{15} molec/cm² to 2.53×10^{15} molec/cm², respectively. We expected that the increase in deviation of QA4ECV would lead to a decrease in the correlation between satellite retrievals and MAX-DOAS measurements; in fact, the correlation between QA4ECV OMI and MAX-DOAS remained very small. The correlation coefficient (R) of QA4ECV OMI increased as deviation also increased. This reversed change in correlation factor also happened when we applied the correction method to the DOMINO data.

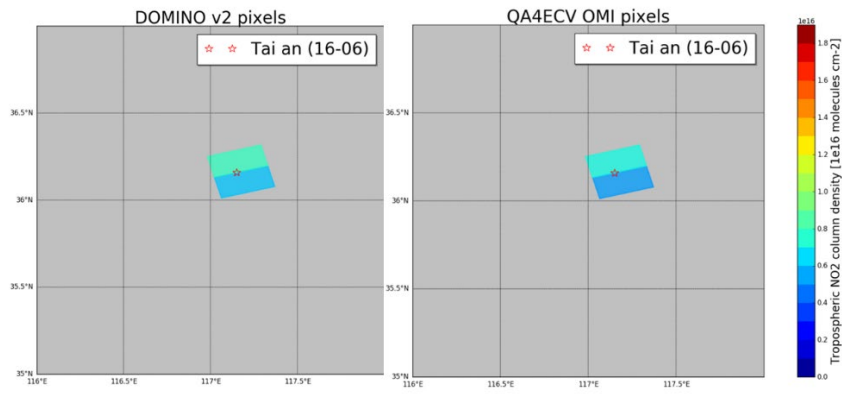
Absent the correction factor, the RMSD of DOMINO decreased to 2.46×10^{15} molec/cm², and R^2 increased to 0.18. This may be due to the fact that the correction method only corrects for mean spatial mismatch and does not consider individual, day-to-day meteorological conditions. For example, wind speed and direction are important factors that can influence NO₂ gradients at the local scale in only a short time, which may well be averaged out in Figure 1. Therefore, the correction method could minimize the smoothing effect in satellite pixels according to monthly mean NO₂ distribution around the MAX-DOAS site. The low R^2 is because the total range of values is small, and the individual uncertainties are large. The difference decreases in deviation for QA4ECV OMI, and SCIAMACHY versus MAX-DOAS indicates that when the pixel resolution is high, and the spatial filter is strict, the correction is not very important. However, as the resolution becomes coarser and the spatial filter looser, the correction may be more important for validation.



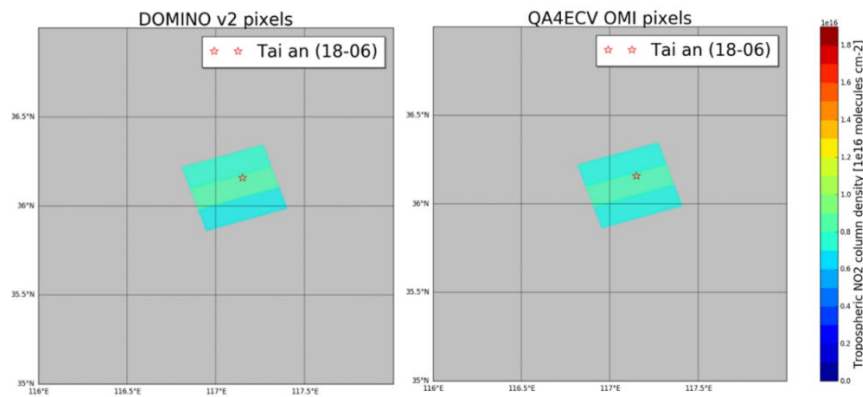
(a)



(b)



(c)



(d)

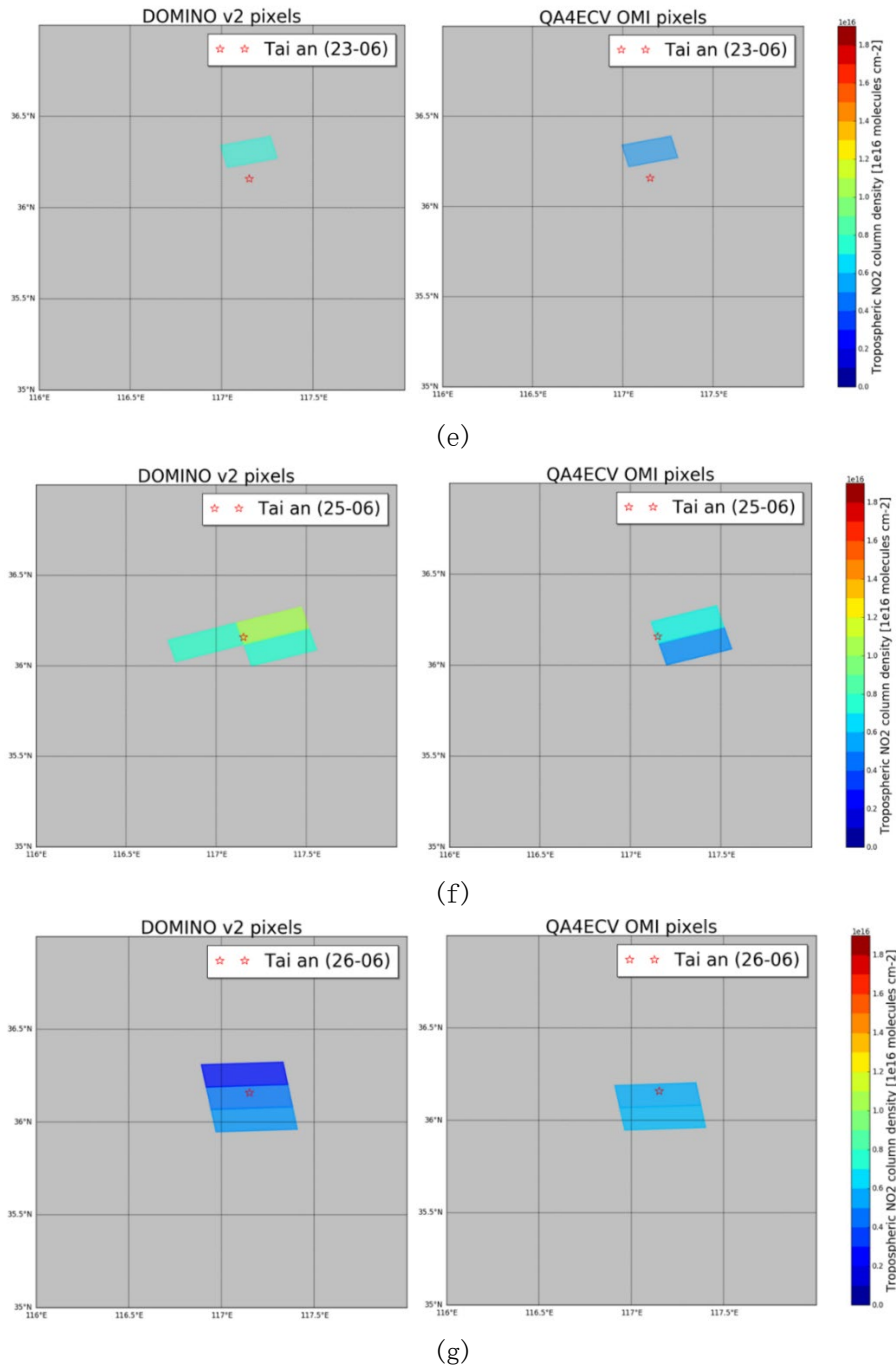


Figure 4. Corrected OMI DOMINO v2 (left) and QA4ECV (right) pixels in the same day: (a) 9

June 2006, (b) 12 June 2006, (c) 16 June 2006, (d) 18 June 2006, (e) 23 June 2006, (f) 25

June 2006, (g) 26 June 2006.

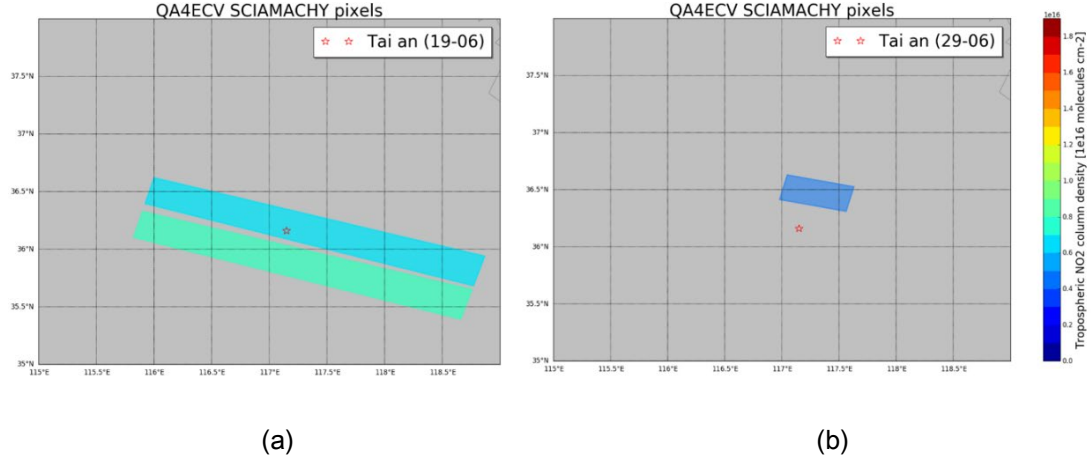


Figure 5. Corrected QA4ECV SCIAMACHY pixels in (a) 19 June 2006 and (b) 29 June 2006.

The SCIAMACHY pixels are remarkably large due to those are the back-scan pixel.

3.1.2. Uncertainties and probability distribution

To investigate the differences between these three satellite products and MAX-DOAS measurements further, we evaluate the uncertainties reported in the three satellite products and their observed differences, as demonstrated in Table 4.

Table 4

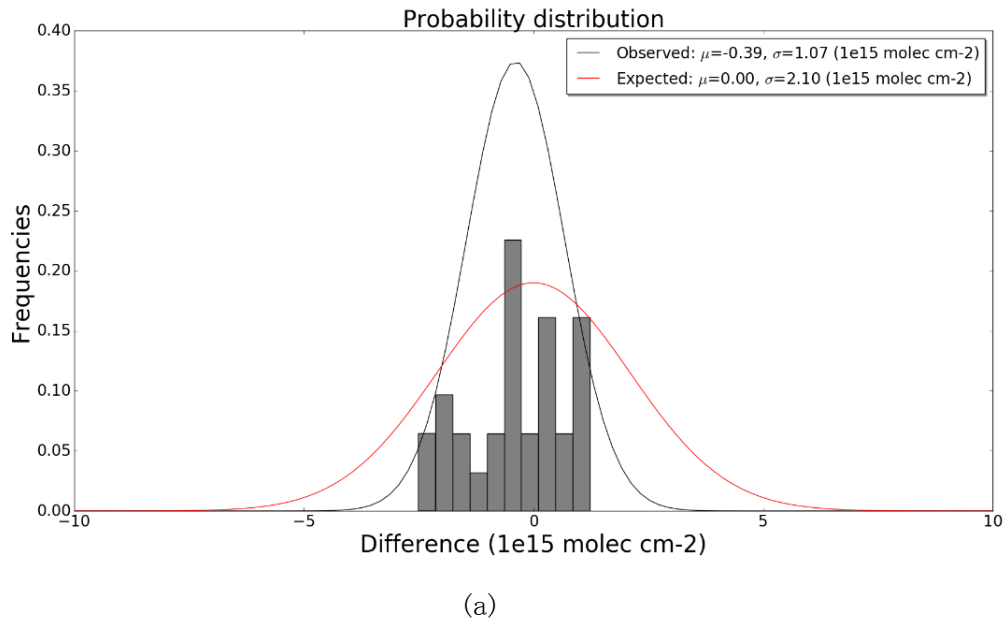
Expected Differences and Observed Differences of QA4ECV OMI, DOMINO v2 and QA4ECV

SCIAMACHY in Tai'an

	Expected difference s QA4ECV OMI (<i>n</i> = 31)	Observed difference s QA4ECV OMI (<i>n</i> = 31)	Expected difference s DOMINO v2 (<i>n</i> = 45)	Observed difference s DOMINO v2 (<i>n</i> = 45)	Expected difference s QA4ECV SCIAMACHY (<i>n</i> = 5)	Observed difference s QA4ECV SCIAMACHY (<i>n</i> = 5)
σ (molec/cm ²)	2.09E+15	1.07E+15	2.55E+15	2.51E+15	2.13E+15	1.37E+15
σ_O (molec/cm ²)	1.71E+15		2.21E+15		1.69E+15	
σ_{MD} (molec/cm ²)	1.00E+15		0.99E+15		1.08E+15	
σ_R (molec/cm ²)	0.69E+15		0.81E+15		0.71E+15	

Note. σ_O is the mean uncertainty reported in satellite products, σ_{MD} is the mean uncertainty reported in MAX-DOAS measurements and σ_R is considered to be a 10% contribution from mismatches.

As Table 4 demonstrates, the observed discrepancies are smaller than the expected differences between the satellite measurements and the MAX-DOAS measurements based on reported uncertainties in the data. The expected differences in DOMINO v2 products are close to the observed differences. For QA4ECV OMI and SCIAMACHY, the observed differences are 49% and 35% smaller than the expected differences. This indicate that the QA4ECV algorithm did improved compared to DOMINO v2 algorithm, and the expected differences of QA4ECV are generally conservative at Tai'an. The probability distribution of the three satellite products is illustrated in Figure 6.



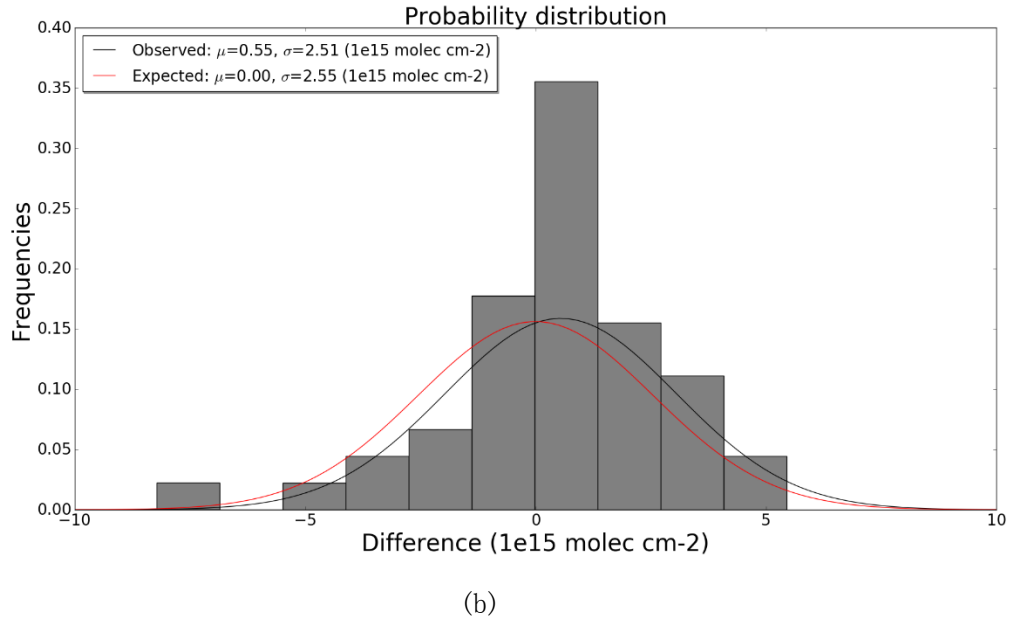


Figure 6. Probability distribution of QA4ECV OMI (a), DOMINO v2 (b) in Tai'an. The histogram represents the distribution of observed differences, the black line indicates the distribution of the observed differences and the red line indicates the Gaussian form distribution of the expected differences.

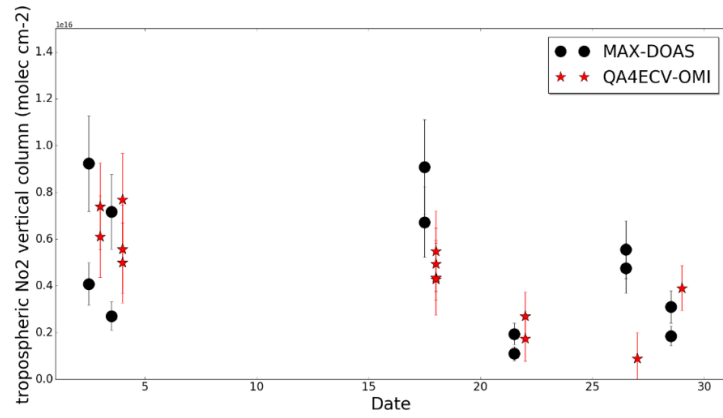
Because the retrievals uncertainties between OMI and QA4ECV are assumed to be independent and follow a normal distribution, we expect the distribution of the differences between OMI and MAX-DOAS can be expressed as a Gaussian form. Figure 6 compares the distribution of differences predicted from the Gaussian function based on the uncertainties reported in the satellite products and MAX-DOAS data files and a 10% mismatch error to the observed differences from individual pairs of satellite products and MAX-DOAS NO₂ column values. The differences between the QA4ECV OMI and MAX-DOAS NO₂ columns are more narrowly distributed than expected from the algorithm. This figure agrees with Table 4, where QA4ECV OMI observed differences are much smaller than expected differences based on algorithm uncertainties estimates. For DOMINO v2, the observed differences are close to the expected ones; this leads to two similar distributions, where the distribution of expected differences is slightly tighter than that of observed differences. This results in the observed differences of QA4ECV and DOMINO v2 are mostly distributed around zero difference. This means QA4ECV and DOMINO v2 measurements tend to be slightly higher than MAX-DOAS measurements. However, since there are only five samples in SCIAMACHY measurements that match the criteria, the representativeness of this finding is questionable.

3.2. De Bilt

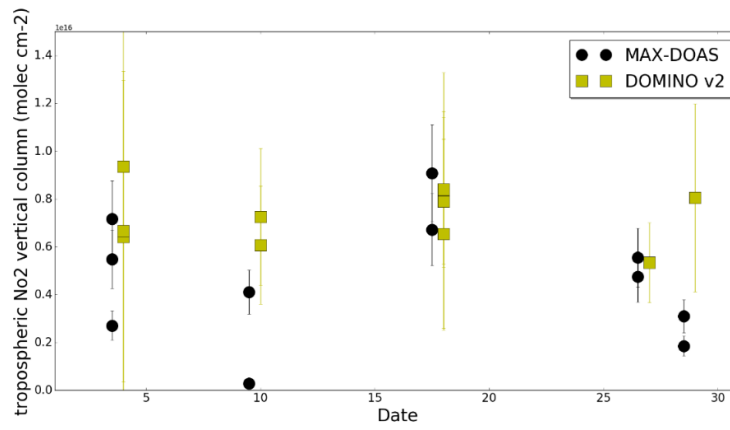
3.2.1. Satellite products' performance validated with MAX-DOAS

We used the same criteria for De Bilt in July 2014 that we had used for Tai' an. In this case, we found 26 QA4ECV OMI pixels matching with 12 MAX-DOAS measurements over 6 different days; for DOMINO v2, there were 24 pixels matching with 11 MAX-DOAS measurements over 5 different days. Because the pixels of GOME2-A are much larger than the pixels of the other two products ($80 \times 40 \text{ km}^2$), we removed the 700 km^2 limitation for pixel area and increased the limitation for distance between pixel centre to MAX-DOAS site from 20 km to 60 km (only applied for GOME2-A). We found 23 GOME2-A pixels matching with 18 MAX-DOAS measurements over 6 different days (Figure 7). From the matching plots, we found MAX-DOAS measurements were highly variable on some days. Within a period of 30 min, for example, on 3 July 2014, in the QA4ECV OMI plot (Figure 7 (a)), the NO_2 column changed from 0.4 to $0.9 \times 10^{16} \text{ molec/cm}^2$, possibly because the MAX-DOAS measurements at the De Bilt site can be strongly influence by wind direction (Vlemmix et al., 2015). The map (Figure 8) depicts a major city (Utrecht) to the west of the De Bilt site. To the south, there is the A28 highway, which is a main traffic source nearby. To the north and east, however, there are many green areas, and the regional road north of the De Bilt site (N237) has much less traffic compared to A28. Therefore, with wind coming from the western or southern direction, the MAX-DOAS measurements might detect a high value of NO_2 VCDs. And if the wind direction or speed changes between two samples on the same day, NO_2 values may fluctuate rapidly.

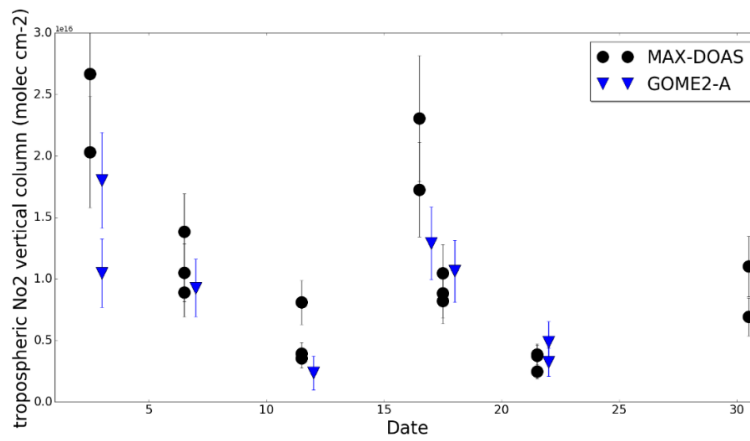
As Table 5 demonstrates, on 3 July, the MAX-DOAS tropospheric NO_2 VCDs jumped from $4.08 \times 10^{15} \text{ molec/cm}^2$ to $9.23 \times 10^{15} \text{ molec/cm}^2$ within 30 min when the hourly mean wind direction was from the south. On 29 July, the difference between two MAX-DOAS samples within 30 min was relatively small ($1.25 \times 10^{15} \text{ molec/cm}^2$) when the wind was coming from the north and north-east. This is consistent with Vlemmix et al. (2015), who found that MAX-DOAS tropospheric NO_2 VCD density measurements at De Bilt were influenced by wind direction. The rest of July 2014 followed a similar pattern, except for 27 July. On that day, the wind was coming from the west, but the difference between the two samples was only $0.8 \times 10^{15} \text{ molec/cm}^2$. This is possibly related to the weekly cycle with tropospheric NO_2 VCDs dropping significantly during the weekend in the European countries. Therefore, Utrecht to the west is a weaker NO_2 source on the weekend, when the wind coming from the direction of Utrecht has less influence than it does on weekdays.



(a)



(b)



(c)

Figure 7. (a) depicts 26 QA4ECV OMI pixels matching with 12 MAX-DOAS measurements over 6 different days in July 2014. (b) depicts 24 DOMINO v2 pixels

matching with 11 MAX-DOAS measurements over 5 different days in July 2014. (c) depicts 23 GOME2-A pixels matching with 18 MAX-DOAS measurements over 6 different days in July 2014. For QA4ECV OMI and DOMINO v2, the OMI scheme was used. For GOME2-A, the GOME2 scheme was used. For visuality reason, MAX-DOAS samples in the Figure 7 was moved to 12 hours earlier.



Figure 8. Map of the De Bilt site, where the MAX-DOAS instrument is located. Source: www.google.nl/maps/place/KNMI/@52.0790769,5.1489501,12.67z/data=!4m5!3m4!1s0x47c668f26e2ea921:0xb13f15f6cb77230!8m2!3d52.1015441!4d5.1779992?hl=en.

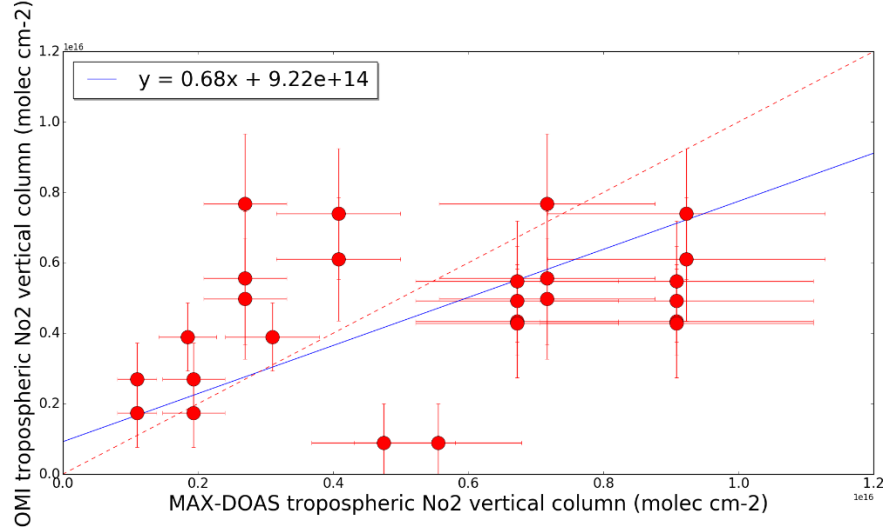
Table 5

MAX-DOAS Samples in July 2014 With Hourly Wind Direction at 12:00–13:00 Each Day

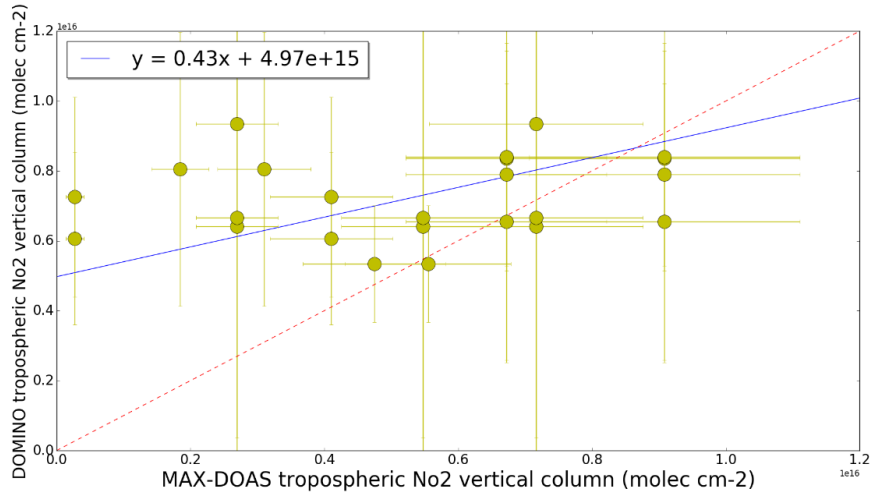
2014/7/						
Date	2014/7/ 3 (THU)	2014/7/ 4 (FRI)	2014/7/ 18 (FRI)	2014/7/ 2 (TUE)	2014/7/ 2 (SUN)	2014/7/ 2 (TUE)
MAX-DOAS time (sample1)	12:58:26	12:11:16	12:10:11	11:49:27	12:00:59	11:46:38

MAX-DOAS time (sample2)	13:27:19	12:38:19	12:36:04	12:18:01	12:29:12	12:13:01
MAX-DOAS NO2 column (sample1) (molec/cm ²)	4.08E+15	2.70E+15	6.72E+15	1.94E+15	5.55E+15	1.85E+15
MAX-DOAS NO2 column (sample2) (molec/cm ²)	9.23E+15	7.16E+15	9.08E+15	1.10E+15	4.75E+15	3.10E+15
Hourly mean wind direction (12:00– 13:00)	S	S	ENE	NNW	W	NNE
Hourly mean wind speed (km/h) (12:00–13:00)	12.5	14.5	11	13	9	9
Differences between MAX-DOAS samples (molec/cm ²)	5.15E+15	4.46E+15	2.36E+15	8.40E+14	8.00E+14	1.25E+15

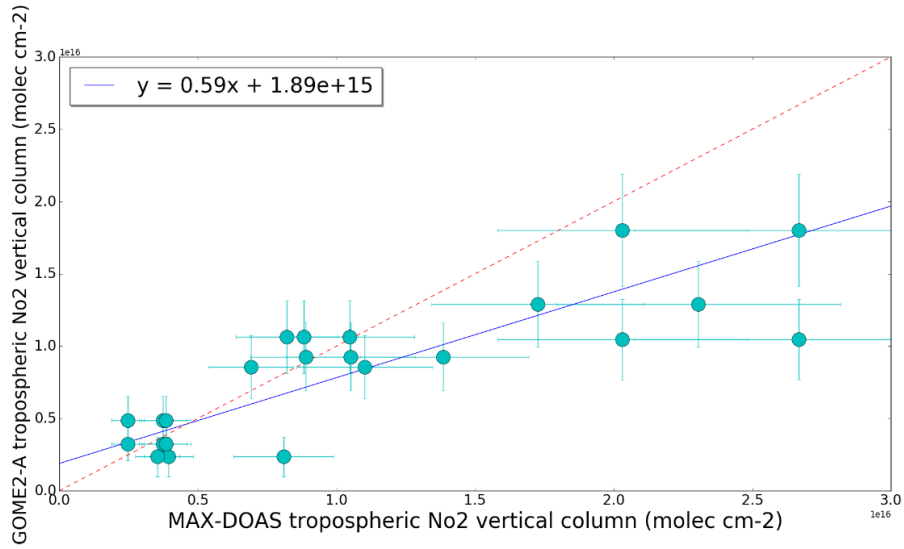
The correlation between these three satellite measurements and the MAX-DOAS measurements in De Bilt is illustrated in Figure 9. Each satellite pixel is paired to one MAX-DOAS measurement that met the criteria. With De Bilt, using the same method as for Tai' an, the relationship between QA4ECV OMI (y) and MAX-DOAS NO₂ VCDs (x) can be expressed as $y = 0.92 \times 10^{15} \text{ molec/cm}^2 + 0.68x$ ($R^2 = 0.65$, $n = 26$). The relationship between DOMINO v2 (y) and MAX-DOAS (x) can be expressed as $y = 4.97 \times 10^{15} \text{ molec/cm}^2 + 0.43x$ ($R^2 = 0.01$, $n = 24$). The relationship between GOME2-A(y) and MAX-DOAS (x), can be expressed as $y = 1.89 \times 10^{15} \text{ molec/cm}^2 + 0.59x$ ($R^2 = 0.25$, $n = 23$).



(a)



(b)



(c)

Figure 9. Scatterplot of (a) QA4ECV OMI versus MAX-DOAS tropospheric NO₂ VCDs for De Bilt (Netherlands) in July 2014 ($R^2 = 0.65$, $n = 26$), (b) DOMINO v2 OMI versus MAX-DOAS tropospheric NO₂ VCDs for De Bilt (Netherlands) in July 2014 ($R^2 = 0.01$, $n = 24$) and (c) GOME2-A versus MAX-DOAS tropospheric NO₂ VCDs for De Bilt (Netherlands) in July 2014 ($R^2 = 0.25$, $n = 23$). The blue line indicates the result of a reduced major axis regression to the data. In (a) and (b), the OMI scheme was used. In (c), the GOME2 scheme was used. Satellite samples were corrected by applied Eq (2).

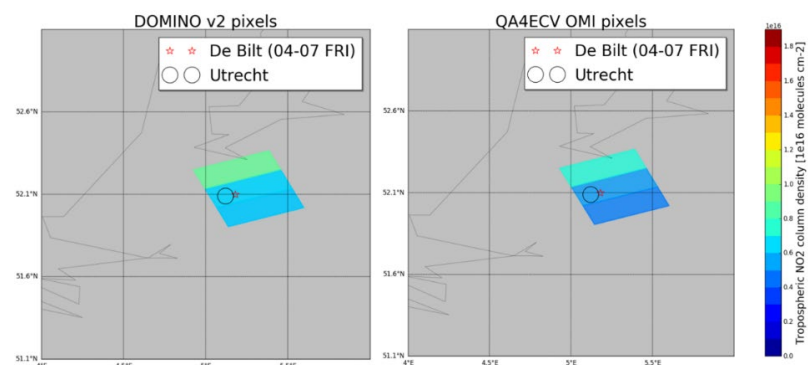
Table 6 demonstrates the RMSD and mean bias between the three satellite products and the MAX-DOAS measurements in De Bilt, with and without correction. When corrected, QA4ECV OMI and DOMINO v2 performed differently. For QA4ECV OMI, the RMSD increased from 2.65×10^{15} molec/cm² to 2.85×10^{15} molec/cm²; however, the mean bias and R^2 improved slightly. For DOMINO v2, the RMSD increased from 2.77×10^{15} molec/cm² to 3.27×10^{15} molec/cm² with the correction, and the mean bias increased from 1.08×10^{15} molec/cm² to 1.92×10^{15} molec/cm². The R^2 dropped from 0.13 to 0.01, meaning that after correction there was no correlation left between the satellite products and the MAX-DOAS measurements. From the relative bias, we can see the correction works well for all three satellite products. The bias become smaller after satellite samples corrected. We also find that the differences between satellite data and MAX-DOAS data are quite high, which is in line with what we have found in matching plots. The wind coming from the nearby NO₂ source could be one of the reasons that the differences between satellite products and MAX-DOAS measurements are higher than what we found in Tai' an. De Bilt is near a highway and the city of Utrecht, but for Tai' an, as Figure 1 has illustrated, the situation is much smoother.

Table 6

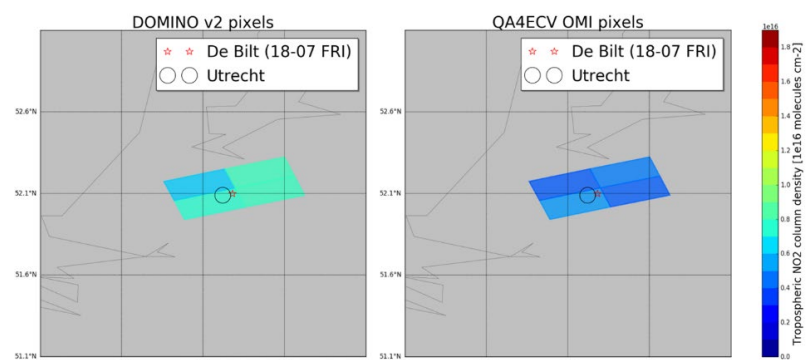
Mean Bias, Root Mean Squared Deviation, Mean Correction Factor and R^2 of QA4ECV OMI, DOMINO v2 and GOME2-A Versus MAX-DOAS in De Bilt. "Corrected row" is calculated by applying Eq (2).

	QA4ECV OMI	DOMINO v2	GOME2-A
Corrected mean bias (molec/cm ²)	-0.80E+15	1.92E+15	-2.51E+15
Uncorrected mean bias (molec/cm ²)	-0.90 E+15	2.77 E+15	-4.4 E+15
Corrected RMSD (molec/cm ²)	2.85E+15	3.27E+15	5.31E+15
Uncorrected RMSD (molec/cm ²)	2.65E+15	2.77 E+15	6.86 E+15
Corrected relative bias (%)	-17%	30%	30%
Uncorrected relative bias (%)	-19%	43%	69%
Mean correction factor	1.07	1.15	1.32
Corrected R^2 (-)	0.65	0.01	0.25
Uncorrected R^2 (-)	0.63	0.13	0.19

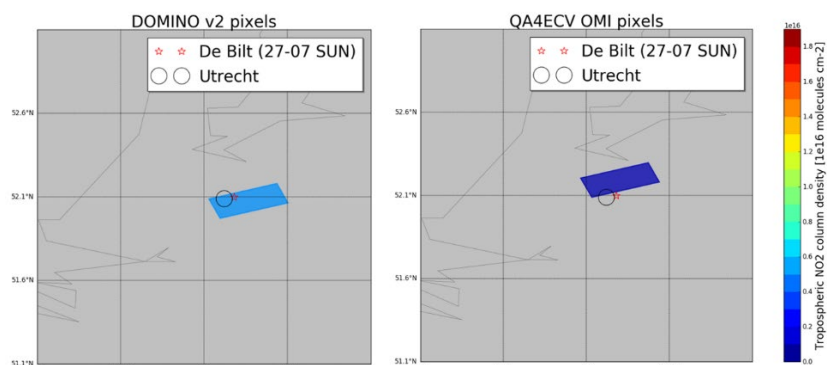
We selected the days that both QA4ECV and DOMINO v2 had pixels matching with MAX-DOAS measurements (Figure 8). DOMINO VCDs tended to be higher than QA4ECV, in line with Figure 7 and Table 5. Figure 10 demonstrates that the De Bilt site and the city of Utrecht city are small compared to the pixel size. If most areas within the pixel are less polluted than the MAX-DOAS site, the high emission value of the MAX-DOAS site could be averaged into a relatively small value. Therefore, the satellite data might underestimate the VCDs when the ground observation site is near strong emission sources, due to the satellite being unable to capture the spatial gradient that could be detected by the MAX-DOAS site.



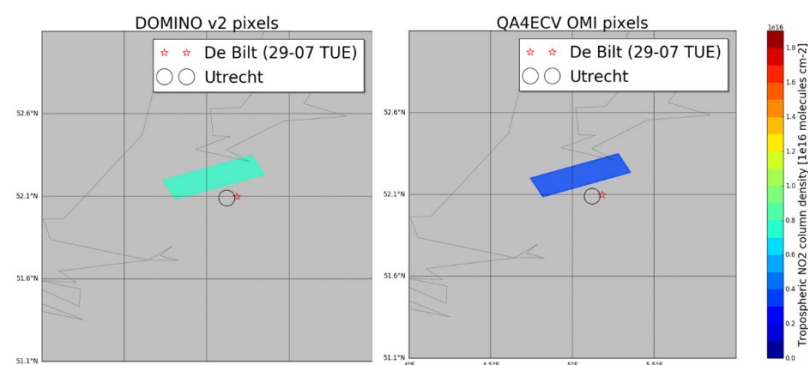
(a)



(b)



(c)



(d)

Figure 10. Corrected DOMINO v2 (left) and QA4ECV OMI (right) pixels matched with MAX-DOAS measurements in 04-07 (FRI) (a), 18-07 (FRI) (b), 27-07 (SUN)(c), 29-07 (TUE) (d) 2014

at De Bilt.

For coarse resolution instruments, for example, as in the case of GOME2-A, the pixels are large, and the distance limitation that restricts the distance from pixel centre to target location loosens to within 60 km; in this case, De Bilt itself is often not included in the pixel area (Figure 11). To make sure that the satellite and the MAX-DOAS instrument measurements at least partly overlap in their air masses, the distance limitation needs to be modified to a smaller value. This modification will require a longer period of sampling time to be able to select more data to conduct a representative statistical analysis. In this case, we only compared GOME2-A with MAX-DOAS measurements within 1 month, and there were simply not sufficient matches for strict criteria for such a coarse resolution satellite like GOME2-A. This problem also exists in the Tai' an SCIAMACHY case, when the pixel size was coarse and made it difficult to select enough representative data in 1 month.

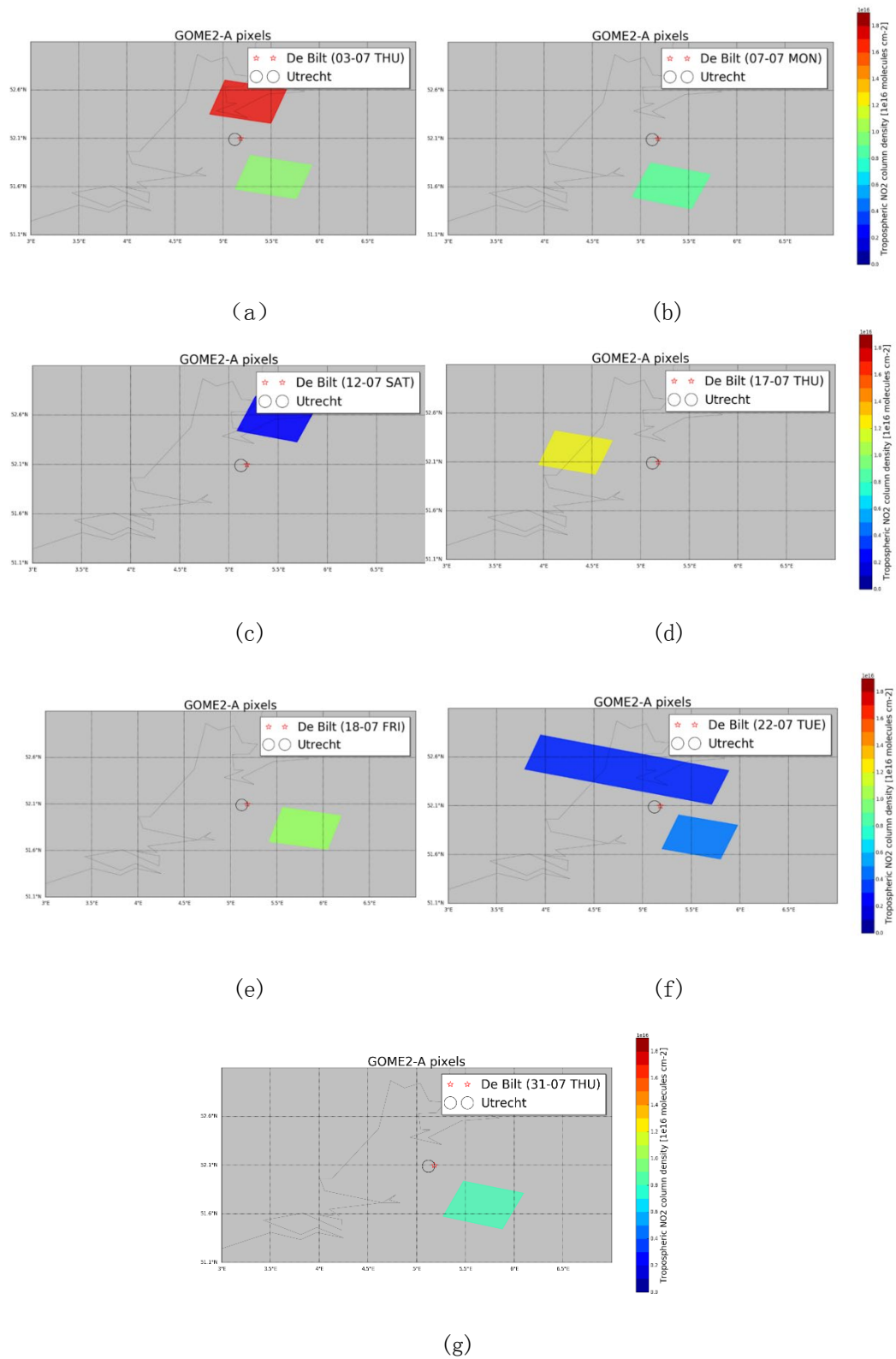


Figure 11. Corrected GOME2-A pixels matched with MAX-DOAS measurements in 03-07

(THU) (a), 07-07 (MON) (b), 12-07 (SAT) (c), 17-07 (THU) (d), 18-07 (FRI) (e), 22-07 (TUE)

(f), 31-07 (THU) (g) 2014 at the De Bilt site.

3.2.2. Uncertainties and probability distribution

Table 7 demonstrates the expected and observed differences of QA4ECV OMI, DOMINO v2 and GOME2-A in the case of De Bilt.

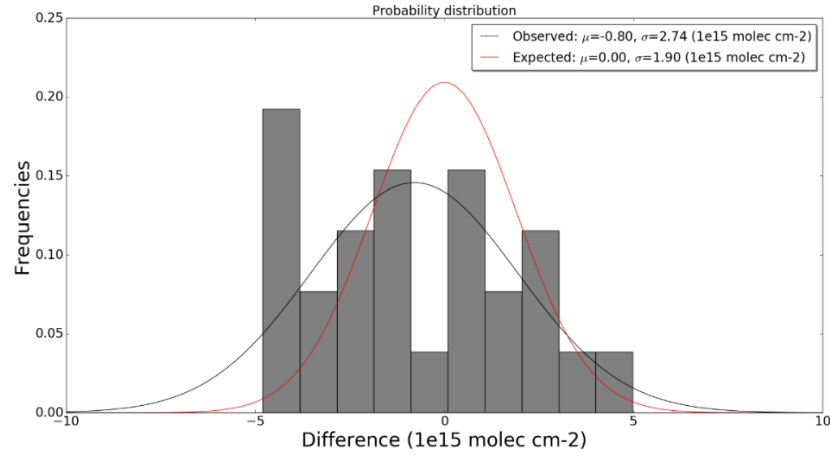
Table 7

Expected Differences and Observed Differences of QA4ECV OMI, DOMINO v2 and QA4ECV
SCIAMACHY in De Bilt

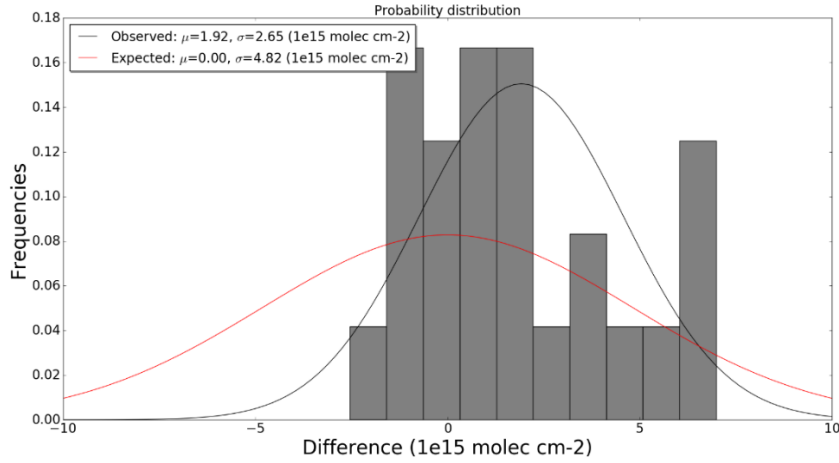
	Expected difference s QA4ECV OMI (<i>n</i> = 26)	Observed difference s QA4ECV OMI (<i>n</i> = 26)	Expected difference s DOMINO v2 (<i>n</i> = 24)	Observed difference s DOMINO v2 (<i>n</i> = 24)	Expected difference s GOME2-A (<i>n</i> = 23)	Observed difference s GOME2-A (<i>n</i> = 23)
σ (molec/cm ²)	1.90E+15	2.74E+15	4.82E+15	2.65E+15	3.31E+15	4.68E+15
σ_O (molec/cm ²)	1.07E+15		4.66E+15		2.38E+15	
σ_{MD} (molec/cm ²)	1.51E+15		1.04E+15		2.21E+15	
σ_R (molec/cm ²)	0.49E+15		0.64E+15		0.77E+15	

Note. σ_O is the mean uncertainty reported in satellite products, σ_{MD} is the mean uncertainty reported in MAX-DOAS measurements and σ_R is considered to be a 10% contribution from mismatches.

In the case of De Bilt, the QA4ECV OMI and GOME2-A algorithm underestimate uncertainties by 44% and 41%, respectively. However, the DOMINO v2 algorithm overestimates uncertainties by 45%. The probability distribution of the three satellite products is illustrated in Figure 12.



(a)



(b)

Figure 12. Probability distribution of (a) QA4ECV OMI and (b) DOMINO v2 in De Bilt. The histogram represents the observed differences distribution, the black line indicates the distribution of the observed differences and the red line indicates the Gaussian form distribution of the expected differences.

Figure 12 compares the distribution of differences predicted from the Gaussian function based on the uncertainties reported in the satellite products and MAX-DOAS data files and a 10% mismatch error to the observed differences from individual pairs of satellite products and MAX-DOAS NO₂ column values. The probability distribution of QA4ECV OMI, DOMINO v2 and GOME2-A is in line with Table 6. It appears that the QA4ECV OMI and GOME2-A reported uncertainties are small compared to the observed discrepancies. Nevertheless, QA4ECV OMI has the

smallest mean bias, which is distributed near zero (-0.8×10^{15} molec/cm²), and due to GOME2-A pixels are not even at De Bilt, GOME-2A has the worst mean bias (-2.51×10^{15} molec/cm²). The mean bias of DOMINO v2 is 1.92×10^{15} molec/cm². Figure 12 illustrates that at De Bilt, only DOMINO v2 products performed better than expected. This is possibly due to DOMINO v2 having a relatively large mean uncertainty coming from the satellite data files, which leads to a large expected deviation. However, the results of both site show that QA4ECV OMI have lower bias than DOMINO v2, and QA4ECV OMI captures the variability of NO₂ VCDs at De Bilt much better than DOMINO v2. These two results both indicate that QA4ECV is an improvement over DOMINO v2. Because the coarse resolution of GOME2-A, the GOME2-A pixels selected by GOME2 scheme are all too far from De Bilt, which makes the comparison meaningless, therefore GOME2-A probability distribution is not showing in figure 12.

3.3 Comparison of the Tai'an and De Bilt cases

Tai'an and De Bilt both are quite polluted areas. We expected similar performance of satellite products in these two sites. However, the satellite products seem to perform better (lower bias? Higher R? lower RMSD - be specific) at Tai'an than at De Bilt in general. Table 8 summarises the deviations and uncertainties of two OMI products for the two sites as reported in section 3.1 and 3.2.

Table 8

Monthly Mean, Mean Bias, RMSD and Uncertainties of QA4ECV and DOMINO v2 for Tai'an and De Bilt.

	QA4ECV (Tai'an)	DOMINO v2 (Tai'an)	QA4ECV (De Bilt)	DOMINO v2 (De Bilt)
MAX-DOAS monthly mean (molec/cm ²)	7.06E+15	7.29E+15	4.77E+15	4.83E+15
Satellite monthly mean (molec/cm ²)	6.93E+15	8.14E+15	4.62E+15	7.33E+15
Mean bias (molec/cm ²)	-0.39E+15 (-6%)	0.55E+15 (7%)	-0.80E+15 (-17%)	1.92E+15 (26%)
RMSD (molec/cm ²)	1.13E+15 (16%)	2.57E+15 (32%)	2.85E+15 (62%)	3.27E+15 (26%)
Expected differences (molec/cm ²)	2.09E+15 (30%)	2.55E+15 (31%)	1.90E+15 (41%)	4.82E+15 (66%)

Observed differences (molec/cm ²)	1.07E+15 (15%)	2.51E+15 (31%)	2.74E+15 (59%)	2.65E+15 (36%)
---	-------------------	-------------------	-------------------	-------------------

In general, QA4ECV performs better than DOMINO v2 in both Tai' an and De Bilt cases. In Table 8 we can see that compared to MAX-DOAS, QA4ECV tends to be lower and DOMINO v2 tends to be higher on average. Also, QA4ECV has less absolute bias and has smaller RMSD than DOMINO v2. It appears that QA4ECV is an improvement over DOMINO v2 for both Tai' an and De Bilt, in line with the substantial improvement of all retrieval sub-steps in QA4ECV relative to DOMINO v2 (Zara et al., 2018; Boersma et al., 2018; Lorente et al., 2018).

On the other hand, each satellite product performs better at Tai' an than at De Bilt. For QA4ECV, the mean bias and RMSD at Tai' an are smaller than at De Bilt. Although the total amount of mean bias and RMSD at De Bilt are relatively small, they are still double of those in Tai' an. This indicates two possible reasons (1) QA4ECV retrievals are more uncertain at De Bilt compared to Tai' an; (2) MAX-DOAS measurements are more uncertain at De Bilt compared to Tai' an, or both. From theoretical considerations we expected the differences to be of the same order of magnitude (Table 8). In fact, the results indeed show that the expected differences of QA4ECV for the two sites are similar (2×10^{15} molec/cm², 30–41%, Table 8). This indicates that the uncertainties of QA4ECV NO₂ for Tai' an and De Bilt are similar. However, the observed differences at De Bilt are almost three times higher than at Tai' an. One reason for the larger differences (higher RMSD) at De Bilt could be that NO₂ columns are much more variable in time than at Tai' an (compare Figure 2 (a) to Figure 7 (a)). The reason for the high temporal variability at De Bilt is likely the rapid changes of wind direction and the vicinity of a strong NO₂ source: the city of Utrecht, located about 4 km to the West of De Bilt. This is also supported by the values for σ_{MD} (Table 4 and Table 7), which are lower at Tai' an than at De Bilt. σ_{MD} is the mean uncertainty of each independent MAX-DOAS measurement that has satellite pixels match with it.

DOMINO v2 also has a smaller mean bias and RMSD at Tai' an compared to De Bilt. Unlike QA4ECV, the observed differences at these two sites are similar. However, the expected differences at De Bilt are almost double compared to the expected differences at Tai' an. This indicates that DOMINO v2 retrievals are unstable at De Bilt compared to Tai' an. Table 4 and 7 demonstrated that at De Bilt, DOMINO v2 σ_o is more than 2 times higher than at Tai' an, which is also supported by DOMINO v2 σ_o . One possible reason for this is the high AMF structural uncertainty, which is mostly caused by substantial differences in the a priori trace gas profiles, surface albedo and cloud parameters. The different cloud and aerosol correction approaches can result in significant substantial AMF differences in polluted areas (Lorente, et al, 2017).

The representativeness of the validation of coarser resolution satellite products (SCIAMACHY, GOME-2) has remained doubtful. If the same distance limitation as OMI is applied, no pixel will be found in the measured month. Increasing the distance limitation for coarser resolution satellite products could result in much higher uncertainty simply due to the pixels are far away from the MAX-DOAS site.

4. Conclusion and recommendation

To assess the performance of QA4ECV OMI, DOMINO v2 and SCIAMACHY (GOME2-A) at Tai'an (June 2006) and De Bilt (July 2014), we compared these satellite tropospheric NO₂ VCDs to MAX-DOAS tropospheric NO₂ VCDs at these two sites. Through this comparison, we quantified these satellite tropospheric NO₂ retrievals quality. Five criteria have been imposed in the samples' selection process to ensure good mutual representativeness of samples. For QA4ECV, we found 31 pixels matching with 13 MAX-DOAS measurements over 7 days at Tai'an and 26 pixels matching with 12 MAX-DOAS measurements over 6 different days at De Bilt. Statistical analyses between satellite products and MAX-DOAS suggest that QA4ECV OMI performs best at both Tai'an and De Bilt. In Tai'an, the mean bias and uncertainties between QA4ECV OMI and MAX-DOAS are smaller than in De Bilt (Tai'an mean bias: -6%, uncertainties: 17%; De Bilt mean bias: -17%, uncertainties: 59%). The results of DOMINO v2 also revealed that it performs better at Tai'an than at De Bilt (Tai'an mean bias: 7%, uncertainties: 32%; De Bilt mean bias: 27%, uncertainties: 37%). It can be concluded that QA4ECV OMI slightly underestimates NO₂ VCDs and DOMINO v2 tends to overestimate NO₂ VCDs. Over all, QA4ECV OMI agrees better with MAX-DOAS than DOMINO v2 both in Tai'an and in De Bilt in terms of mean bias, and more reliable on captures the variability of NO₂ VCDs.

To investigate the apparent instability at De Bilt, this study found that when the wind blowing from a strong NO₂ source in the weekdays, the NO₂ VCDs of MAX-DOAS are likely to be more variable within one-hour duration. The higher deviations between MAX-DOAS and satellite tropospheric NO₂ VCDs at De Bilt indicate that the validation result could suffer from strong spatio-temporal NO₂ gradients caused by local meteorological conditions and strong NO_x source in the neighbourhood. Thus, we conclude that QA4ECV and DOMINO v2 performed better at Tai'an than at De Bilt mainly due to the De Bilt site being influenced by a significant fluctuation in NO₂ levels. SCIAMACHY and GOME2-A, due to coarse resolution, could not be properly validated with MAX-DOAS within a period of only one month because there will be no or too few samples if we stick to strict distance criterion. A loosened distance criterion leads to larger distances between satellite pixels and the MAX-DOAS site, which makes a direct comparison impossible. Therefore, we can conclude that coarse resolution satellites such as SCIAMACHY and GOME2-A are not suitable for short-term validation under a strict distance limitation (<20 km).

During this study, we found few deficiencies that could be improved. Therefore, based on these deficiencies, we have three main recommendations for further

studies related to QA4ECV OMI, DOMINO v2, SCIAMACHY and GOME2-A validation with MAX-DOAS:

- (1) In this study, due to data availability, we only validated for one month at both site. This forced us to loosen the criteria for SCIAMACHY and GOME2-A simply because if we stick to same criteria of OMI, there will be no sample meeting the criteria. In the further studies, if one wants to compare OMI, SCIAMACHY and GOME2-A to MAX-DOAS, we recommend to use at least whole year data set to ensure enough samples can be selected under strict criteria;
- (2) Because MAX-DOAS could be influenced by strong NO_2 source nearby, we recommend further researchers choose the station while take into account the NO_2 gradient around the station during the measured period as the MAX-DOAS site.
- (3) For this study, we used QA4ECV OMI NO_2 monthly mean data to calculate correction factor for all satellite samples due to data availability. This leads to inaccuracy in corrected results of satellite products except QA4ECV OMI. Therefore, in further studies, we recommend to use corresponding monthly mean data to calculate correction factor for each satellite products.

References

- Beirle, S., et al. "Weekly cycle of NO₂ by GOME measurements: A signature of anthropogenic sources." *Atmospheric Chemistry and Physics* 3.6 (2003): 2225-2232.
- Bovensmann, H., Burrows, J. P., Buchwitz, M., Frerick, J., Noël, S., Rozanov, V. V., Chance, K. V., and Goede, A. P. H.: SCIAMACHY: Mission Objectives and Measurement Modes, *J. Atmos. Sci.*, 56(2), 127–150, 1999.
- Boersma, K. F., H. J. Eskes, and E. J. Brinksma. "Error analysis for tropospheric NO₂ retrieval from space." *Journal of Geophysical Research: Atmospheres* 109.D4 (2004).
- Boersma, K. F., et al. "An improved tropospheric NO₂ column retrieval algorithm for the Ozone Monitoring Instrument." *Atmospheric Measurement Techniques* 4.9 (2011): 1905-1928.
- 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 Roozendaal, M., Wagner, T., Maasakkers, J. D., van der A, R. J., Nightingale, J., De Rudder, A., Irie, H., Pinardi, G., Lambert, J.-C., and Compernolle, S.: Improving algorithms and uncertainty estimates for satellite NO₂ retrievals: Results from the Quality Assurance for Essential Climate Variables (QA4ECV) project, *Atmos. Meas. Tech. Discuss.*, <https://doi.org/10.5194/amt-2018-200>, in review, 2018.
- Bucsela, Eric J., et al. "Algorithm for NO₂/vertical column retrieval from the ozone monitoring instrument." *IEEE Transactions on Geoscience and remote sensing* 44.5 (2006): 1245-1258.
- Chan, K.L., Hartl, A., Lam, Y.F., Xie, P.H., Liu, W.Q., Cheung, H.M., Lampel, J., Pöhlner, D., Li, A., Xu, J., 2015. Observations of tropospheric NO₂ using ground based MAXDOAS and OMI measurements during the Shanghai World Expo 2010. *Atmos. Environ.* 119, 45e58. <http://dx.doi.org/10.1016/j.atmosenv.2015.08.041>.
- Chen, D., Zhou, B., Beirle, S., Chen, L.M., Wagner, T., 2009a. Tropospheric NO₂ column densities deduced from zenith-sky DOAS measurements in Shanghai, China, and their application to satellite validation. *Atmos. Chem. Phys.* 9, 3641e3662.
- Drosoglou, Theano, et al. "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." *Atmospheric Chemistry and Physics* 17.9 (2017): 5829-5849.
- EUMETSAT: GOME-2 Products Guide, <http://oiswww.eumetsat.org/WEBOPS/eps-pg/GOME-2/GOME2-PG-index.htm>, 2008.
- Honninger, G., Friedeburg, C. v., Platt, U., 2004b. Multi axis differential optical absorption

spectroscopy (MAX-DOAS). *Atmos. Chem. Phys.* 4, 231e254.

Han, K.M., Lee, C.K., Lee, J., Kim, J., Song, C.H., 2011. A comparison study between model-predicted and OMI-retrieved tropospheric NO₂ VCDs over the Korean peninsula. *Atmos. Environ.* 45 (17), 2962 – 2971.

Han, K.M., Lee, S., Chang, L.S., Song, C.H., 2015. A comparison study between CMAQsimulated and OMI-retrieved NO₂ VCDs over East Asia for evaluation of NO_x emission fluxes of INTEX-B, CAPSS, and REAS inventories. *Atmos. Chem. Phys.* 15, 1913 – 1938.

Irie, H., Takashima, H., Kanaya, Y., Boersma, K. F., Gast, L., Wittrock, F., Brunner, D., Zhou, Y., and Van Roozendaal, M.: Eight-component retrievals from ground-based MAXDOAS observations, *Atmos. Meas. Tech.*, 4, 1027–1044, <https://doi.org/10.5194/amt-4-1027-2011>, 2011.

Kanter, David R. "Nitrogen pollution: a key building block for addressing climate change." *Climatic Change* 147.1-2 (2018): 11-21.

Ma, J.Z., Richter, A., Burrows, J.P., Nüß, H., van Aardenne, J.A., 2006. Comparison of model-simulated tropospheric NO₂ over China with GOME-satellite data. *Atmos. Environ.* 40 (4), 593e604. <http://dx.doi.org/10.1016/j.atmosenv.2005.09.029>.

Ma, J.Z., Beirle, S., Jin, J., Shaiganfar, R., Yan, P., Wagner, T., 2013. Tropospheric NO₂ VCDs densities over Beijing: results of the first three-years of groundbased MAX-DOAS measurements (2008-2011) and satellite validation. *Atmos. Chem. Phys.* 13 (11), 1547e1567.

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

Levelt, P. F. and Noordhoek, R.: OMI Algorithm Theoretical Basis Document Volume I: OMI Instrument, Level 0-1b Processor, Calibration & Operations, Tech. Rep. ATBD-OMI-01, Version 1.1, August 2002.

Lin, J.T., Martin, R.V., Boersma, K.F., Sneep, M., Stammes, P., Spurr, R., Wang, P., Van Roozendaal, M., Cl_{emer}, K., Irie, H., 2014. Retrieving tropospheric nitrogen dioxide from the Ozone Monitoring Instrument: effects of aerosols, surface reflectance anisotropy, and vertical profile of nitrogen dioxide. *Atmos. Chem. Phys.* 14 (3), 1441e1461.

Lorente, Alba, et al. "Structural uncertainty in air mass factor calculation for NO₂ and HCHO satellite retrievals." (2017).

Lorente, A., Boersma, K. F., Stammes, P., Tilstra, L. G., Richter, A., Yu, H., ... & Muller, J. P. (2018). The importance of surface reflectance anisotropy for cloud and NO₂ retrievals from GOME-2 and OMI. *Atmospheric Measurement Techniques*, 11(7), 4509-4529.

Palmer, Paul I., et al. "Air mass factor formulation for spectroscopic measurements from satellites: Application to formaldehyde retrievals from the Global Ozone Monitoring Experiment." *Journal of Geophysical Research: Atmospheres* 106.D13 (2001): 14539-14550.

Pikelnaya, O., Hurlock, Stephen C., Trick, S., Stutz, J., 2007. Intercomparison of multi-axis and long-path differential optical absorption spectroscopy measurements in the marine boundary layer. *J. Geophys. Res.* 112 (D10S01) [http:// dx.doi.org/10.1029/2006JD007727](http://dx.doi.org/10.1029/2006JD007727).

Platt, U., Stutz, J., 2008. *Differential Absorption Spectroscopy, Principles and Applications*. Springer, Berlin.

Park, R.S., Lee, S.J., Shin, S.K., Song, C.H., 2014. Contribution of ammonium nitrate to aerosol optical depth and direct radiative forcing by aerosols over East Asia. *Atmos. Chem. Phys.* 14 (4), 2185 – 2201.

Richter, A., Burrows, J.P., Nuss, H., Granier, C., Niemeier, U., 2005. Increase in tropospheric nitrogen dioxide over China observed from space. *Nature* 437 (7055), 129e132.

Theys, N., De, S.I., Gent, J., Danckaert, T., Wang, T., Hendrick, F., Stavrou, T., Bauduin, S., Clarisse, L., Li, C., 2015. Sulfur dioxide vertical column DOAS retrievals from the Ozone Monitoring Instrument: global observations and comparison to ground-based and satellite data. *J. Geophys. Res. Atmos.* 120 [http:// dx.doi.org/10.1002/2014JD022657](http://dx.doi.org/10.1002/2014JD022657).

US Environmental Protection Agency, 2003. *National Air Quality and Emissions Trends Report 2003*. Research Triangle Park, NC

Valks, P., Pinardi, G., Richter, A., Lambert, J.C., Hao, N., Loyola, D., Van Roozendael, M., Emmadi, S., 2011. Operational total and tropospheric NO₂ column retrieval for GOME-2. *Atmos. Meas. Tech.* 4 (7), 1491e1514. <http://dx.doi.org/10.5194/amt-4-1491-2011>.

van der A, R.J., Peters, D.H.M.U., Eskes, H.J., Boersma, K.F., Van Roozendael, M., De Smedt, I., Kelder, H.M., 2006. Detection of the trend and seasonal variation in tropospheric NO₂ over China. *J. Geophys. Res.* 111 (D12317), 1125e1132. [http:// dx.doi.org/10.1029/2005JD006594](http://dx.doi.org/10.1029/2005JD006594).

van der A, R.J., Eskes, H.J., Boersma, K.F., Van Noije, T.P.C., Van Roozendael, M., De Smedt, I., Peters, D.H.M.U., Meijer, E.W., 2008. Trends, seasonal variability and dominant NO_x source derived from a ten-year record of NO₂ measured from space. *J. Geophys. Res.*

113 (D04302) <http://dx.doi.org/10.1029/2007JD009021>.

Vandaele, A.C., Simon, P.C., Guilmot, J.M., Carleer, M., Colin, R., 1994. SO₂ absorption cross section measurement in the UV using a Fourier transform spectrometer. *J. Geophys. Res.* 99, 25599e25605. <http://dx.doi.org/10.1029/94JD02187>.

Vandaele, A.C., Hermans, C., Simon, P.C., Carleer, M., Colin, R., Fally, S., 1998. Measurement of the NO₂ absorption cross section from 42000cm⁻¹ to 10000cm⁻¹ (238-1000 nm) at 220K and 294K. *J. Quant. Spectrosc. Ra.* 59, 171e184.

Vlemmix, T., Piters, A., Stammes, P., Wang, P., Levelt, P., 2010. Retrieval of tropospheric NO₂ using the MAX-DOAS method combined with relative intensity measurements for aerosol correction. *Atmos. Meas. Tech.* 3 (5), 1287e1305. <http://dx.doi.org/10.5194/amt-3-1287-2010>.

Wagner, T., Dix, B., Friedeburg, C.v., Frieß, U., Sanghavi, S., Sinreich, R., Platt, U., 2004. MAX-DOAS O₄ measurements: a new technique to derive information on atmospheric aerosols. *Principles and information content. J. Geophys. Res.* 109 (D22205) <http://dx.doi.org/10.1029/2004JD004904>.

Wittrock, F., Oetjen, H., Richter, A., Fietkau, S., Medeke, T., Rozanov, A., Burrows, J.P., 2004. MAX-DOAS measurements of atmospheric trace gases in Ny-lesund - radiative transfer studies and their application. *Atmos. Chem. Phys.* 4 (4), 955e966.

Wagner, T., Deutschmann, T., Platt, U., 2009. Determination of aerosol properties from MAX-DOAS observations of the Ring effect. *Atmos. Meas. Tech.* 2, 725e779. <http://dx.doi.org/10.5194/amt-2-495-2009>.

Wagner, T., S. Beirle, T. Brauers, et al., 2011. Inversion of tropospheric profiles of aerosol extinction and HCHO and NO₂ mixing ratios from MAX-DOAS observations in Milano during the summer of 2003 and comparison with independent data sets. *Atmos. Meas. Tech.* 4, 2685e2715. <http://dx.doi.org/10.5194/amt-4-2685-2011>.

Wang, S.W., Zhang, Q., Streets, D.G., He, K.B., Martin, R.V., Lamsal, L.N., Chen, D., Lei, Y., Lu, Z., 2012. Growth in NO_x emissions from power plants in China: bottom-up estimates and satellite observations. *Atmos. Chem. Phys.* 12 (10), 4429e4447.

Wang, T., Hendrick, F., Wang, P., et al., 2014. Evaluation of tropospheric SO₂ retrieved from MAX-DOAS measurements in Xianghe, China. *Atmos. Chem. Phys.* 14 (5), 6501e6536.

Wittrock, F., Oetjen, H., Richter, A., Fietkau, S., Medeke, T., Rozanov, A., Burrows, J.P., 2004. MAX-DOAS measurements of atmospheric trace gases in Ny-lesund - radiative transfer studies and their application. *Atmos. Chem. Phys.* 4 (4), 955e966.

Zara, M., Boersma, K. F., De Smedt, I., Richter, A., Peters, E., Van Geffen, J. H. G. M., ... & Lamsal, L. N. (2018). Improved slant column density retrieval of nitrogen dioxide and formaldehyde for OMI and GOME-2A from QA4ECV: intercomparison, uncertainty