



Investigating differences in DOAS retrieval codes using MAD-CAT campaign data

Enno Peters¹, Gaia Pinardi², André Seyler¹, Andreas Richter¹, Folkard Wittrock¹, Tim Bösch¹, Michel Van Roozendaal², François Hendrick², Theano Drosoglou³, Alkiviadis F. Bais³, Yugo Kanaya⁴, Xiaoyi Zhao⁵, Kimberly Strong⁵, Johannes Lampel^{6,11}, Rainer Volkamer^{7,8}, Theodore Koenig^{7,8}, Ivan Ortega^{7,8,a}, Olga Puentedura⁹, Mónica Navarro-Comas⁹, Laura Gómez⁹, Margarita Yela González⁹, Ankie Piters¹⁰, Julia Remmers¹¹, Yang Wang¹¹, Thomas Wagner¹¹, Shanshan Wang^{12,13}, Alfonso Saiz-Lopez¹², David García-Nieto¹², Carlos A. Cuevas¹², Nuria Benavent¹², Richard Querel¹⁴, Paul Johnston¹⁴, Oleg Postlyakov¹⁵, Alexander Borovski¹⁵, Alexander Elokhov¹⁵, Ilya Bruchkouski¹⁶, Haoran Liu¹⁷, Cheng Liu^{17,18,19}, Qianqian Hong¹⁹, Claudia Rivera²⁰, Michel Grutter²¹, Wolfgang Stremme²¹, M. Fahim Khokhar²², Junaid Khayyam²², and John P. Burrows¹

¹Institute of Environmental Physics, University of Bremen, Bremen, Germany

²Royal Belgian Institute for Space Aeronomy (BIRA-IASB), Brussels, Belgium

³Aristotle University of Thessaloniki, Thessaloniki, Greece

⁴Japan Agency for Marine-Earth Science and Technology (JAMSTEC), Yokohama, Japan

⁵Department of Physics, University of Toronto, Ontario, Canada

⁶Institute of Environmental Physics, University of Heidelberg, Heidelberg, Germany

⁷Department of Chemistry and Biochemistry, University of Colorado, Boulder, CO, USA

⁸Cooperative Institute for Research in Environmental Sciences (CIRES), University of Colorado, Boulder, CO, USA

⁹National Institute for Aerospace technology, INTA, Madrid, Spain

¹⁰Royal Netherlands Meteorological Institute (KNMI), De Bilt, the Netherlands

¹¹Max Planck Institute for Chemistry, Mainz, Germany

¹²Department of Atmospheric Chemistry and Climate, Institute of Physical Chemistry Rocasolano, CSIC, Madrid, Spain

¹³Shanghai Key Laboratory of Atmospheric Particle Pollution and Prevention (LAP³), Department of Environmental Science & Engineering, Fudan University, Shanghai, China

¹⁴National Institute of Water and Atmospheric Research (NIWA), Lauder, New Zealand

¹⁵A. M. Obukhov Institute of Atmospheric Physics, Russian Academy of Sciences, Moscow, Russia

¹⁶National Ozone Monitoring Research and Education Center BSU (NOMREC BSU), Belarusian State University (BSU), Minsk, Belarus

¹⁷School of Earth and Space Sciences, University of Science and Technology of China, Hefei, 230026, China

¹⁸CAS Center for Excellence in Regional Atmospheric Environment, Xiamen, 361021, China

¹⁹Key Lab of Environmental Optics & Technology, Anhui Institute of Optics and Fine Mechanics, Chinese Academy of Sciences, Hefei, 230031, China

²⁰Facultad de Química, Universidad Nacional Autónoma de México, Mexico City, Mexico

²¹Centro de Ciencias de la Atmósfera, Universidad Nacional Autónoma de México, Mexico City, Mexico

²²Institute of Environmental Sciences and Engineering (IESE), National University of Sciences and Technology (NUST) Islamabad, Islamabad, Pakistan

^anow at: National Center for Atmospheric Research (NCAR), Boulder, CO, USA

Correspondence to: Enno Peters (enno.peters@iup.physik.uni-bremen.de)

Received: 28 October 2016 – Discussion started: 11 November 2016

Revised: 31 January 2017 – Accepted: 10 February 2017 – Published: 10 March 2017

Abstract. The differential optical absorption spectroscopy (DOAS) method is a well-known remote sensing technique that is nowadays widely used for measurements of atmospheric trace gases, creating the need for harmonization and characterization efforts. In this study, an intercomparison exercise of DOAS retrieval codes from 17 international groups is presented, focusing on NO₂ slant columns. The study is based on data collected by one instrument during the Multi-Axis DOAS Comparison campaign for Aerosols and Trace gases (MAD-CAT) in Mainz, Germany, in summer 2013. As data from the same instrument are used by all groups, the results are free of biases due to instrumental differences, which is in contrast to previous intercomparison exercises.

While in general an excellent correlation of NO₂ slant columns between groups of > 99.98 % (noon reference fits) and > 99.2 % (sequential reference fits) for all elevation angles is found, differences between individual retrievals are as large as 8 % for NO₂ slant columns and 100 % for rms residuals in small elevation angles above the horizon.

Comprehensive sensitivity studies revealed that absolute slant column differences result predominantly from the choice of the reference spectrum while relative differences originate from the numerical approach for solving the DOAS equation as well as the treatment of the slit function. Furthermore, differences in the implementation of the intensity offset correction were found to produce disagreements for measurements close to sunrise (8–10 % for NO₂, 80 % for rms residual). The largest effect of ≈ 8 % difference in NO₂ was found to arise from the reference treatment; in particular for fits using a sequential reference. In terms of rms fit residual, the reference treatment has only a minor impact. In contrast, the wavelength calibration as well as the intensity offset correction were found to have the largest impact (up to 80 %) on rms residual while having only a minor impact on retrieved NO₂ slant columns.

1 Introduction

In this study, the consistency of differential optical absorption spectroscopy (DOAS) retrievals of tropospheric nitrogen dioxide (NO₂) from ground-based scattered light observations is evaluated. NO₂ is released into the atmosphere predominantly in the form of NO as a result of combustion processes at high temperatures. Through reaction of NO with ozone (O₃), NO₂ is rapidly produced and (during the day) back-converted into NO by photolysis. Therefore, nitrogen oxides are often discussed in terms of NO_x (NO + NO₂ = NO_x). NO₂ is a key species in the formation of tropospheric ozone and a prominent pollutant in the troposphere, causing (together with aerosols) the typical brownish colour of polluted air. In addition, it is harmful for lung tissue and a powerful oxidant. In the lower troposphere, the lifetime of NO₂ is short (several hours) due to reaction with OH and

photodissociation; thus it is mostly found close to its sources, making it a good tracer of local pollution. Anthropogenic sources such as burning fossil fuel in industry, power generation and traffic as well as biogenic sources including bush and forest fires contribute to the tropospheric NO_x loading. In addition, NO₂ is released from soil microbial processes and lightning events (Lee et al., 1997). As a result, high NO₂ amounts are mostly observed above industrialized and urban areas, traffic routes and over bush fires.

Using its characteristic absorption bands in the UV and visible spectral range, NO₂ has been successfully measured for many years using the DOAS technique (Brewer, 1973; Noxon, 1975; Platt, 1994), both from the ground and from space (e.g. Richter et al., 2005; Beirle et al., 2011). In addition, airborne and car-based measurements (e.g. Schönhardt et al., 2015; Shaiganfar et al., 2011), have been taken to close the gap between ground-based observations providing continuous temporal, but poor spatial resolution and satellite measurements which offer global observations, but only up to one measurement per day above each location. Using ground-based multi-axis (MAX)-DOAS measurements at different elevation angles, a more accurate vertical column (VC) can be retrieved with information on the vertical distribution of NO₂ and other trace gases in the troposphere (e.g. Hönninger et al., 2004; Wittrock, 2006; Frieß et al., 2011; Wagner et al., 2011).

As result of the viewing geometry, ground-based MAX-DOAS observations are most sensitive to the lowest layers of the troposphere. Here they provide high sensitivity and low relative measurement errors. Averaging over longer integration times can further reduce statistical noise. Another advantage is that MAX-DOAS stations can be operated automatically and usually require little maintenance. MAX-DOAS measurements were therefore taken in remote regions for the investigation of background concentrations and in many locations for the validation of satellite observations (e.g. Takashima et al., 2012; Peters et al., 2012).

There currently exists a variety of different instruments and retrieval codes designed to perform and analyse MAX-DOAS measurements. While the basic approaches are similar, differences exist which can potentially lead to inconsistencies in measurements from different instruments and retrievals. As this is an important limitation for the use of MAX-DOAS measurements in a global observing system, the Network for the Detection of Atmospheric Composition Change NDACC (formerly known as NDSC) has organized several intercomparison campaigns aimed at comparing instruments, retrievals and uncertainties of a wide range of DOAS instruments. The first of these campaigns were focused entirely on zenith-sky observations for stratospheric absorbers (Hofmann et al., 1995; Vaughan et al., 1997; Roscoe et al., 1999) but later also included other viewing directions (Vandaele et al., 2005). The Cabauw Intercomparison of Nitrogen Dioxide measuring Instruments (CINDI) campaign in 2009 in Cabauw, the Netherlands

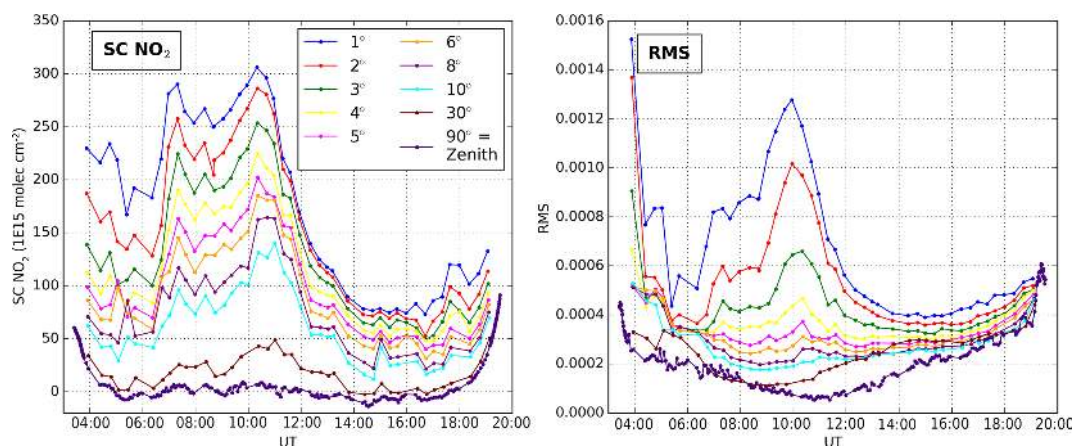


Figure 1. NO_2 slant columns and fit residual root mean square (rms) obtained from the IUPB retrieval code for the intercomparison day used in this study. Different elevation angles are colour coded. The fit settings correspond to v1 settings as described in Sect. 2.4 (Table 1).

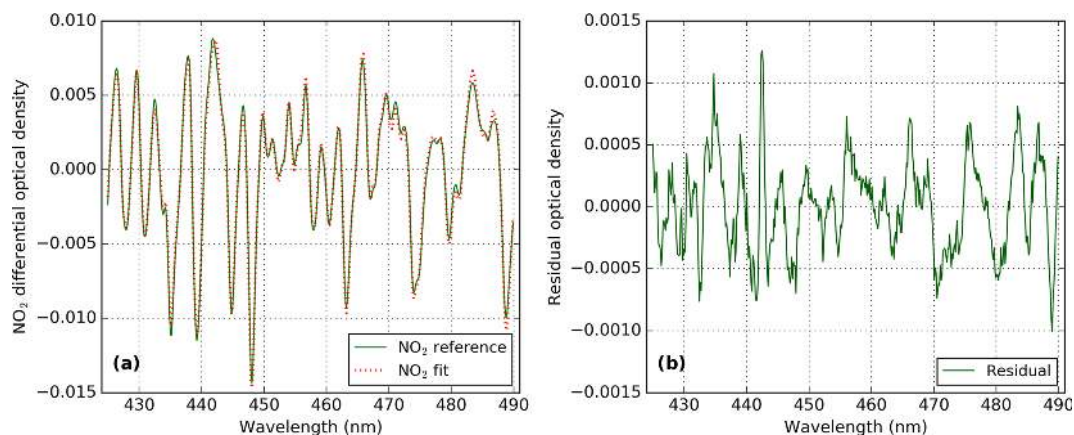


Figure 2. (a) Example of the fitted NO_2 differential optical density in 2° elevation angle at 14:40 UT corresponding to a slant column of $\approx 7\text{E}16 \text{ molec cm}^{-2}$ (compare Fig. 1). The solid line is the NO_2 differential optical density (= differential cross section, i.e. the cross section after subtraction of a fitted polynomial, multiplied by the retrieved slant column) and the dashed line is the solid line plus the fit residual. The NO_2 differential optical density is much larger than the residual, which is explicitly shown in (b).

was the first to also have a focus on MAX-DOAS observations of tropospheric species (Roscoe et al., 2010; Peters et al., 2012; Pinardi et al., 2013) and has recently been followed by the CINDI-2 campaign carried out in 2016, also in Cabauw. However, these intercomparison exercises concentrated mostly on results originating from different retrieval codes and instruments, and separation of instrument and retrieval effects was not easily possible. In some cases, synthetic spectra were used to intercompare retrieval algorithms, and while such tests can highlight differences between retrieval approaches, they give no insight into the way different retrieval codes deal with instrumental effects such as intensity offsets, resolution changes or spectral drifts.

The present study was carried out in the framework of the European FP7 project QA4ECV (Quality Assurance for Essential Climate Variables), which aims at providing quality assurance for satellite-derived ECVs such as NO_2 , HCHO

and CO by characterizing the uncertainty budgets through uncertainty analysis and error propagation but also by validation with external data sets. In this context, ground-based MAX-DOAS measurements can play an important role, and harmonization of the retrieval approaches and quality assurance for the reference measurements are needed. The results of the analysis in the QA4ECV project are expected to contribute to the harmonization of data from MAX-DOAS instruments, in particular for the ongoing integration of such measurements in the NDACC network.

The work reported here overcomes limitations from previous studies by using real measurements originating from a single instrument. This facilitates the study of the agreement between different retrieval codes on real data without instrumental biases between results from different groups. The intercomparison was carried out on spectra recorded by the University of Bremen (IUPB) instrument during

the Multi-Axis DOAS Comparison campaign for Aerosols and Trace gases (MAD-CAT) carried out in Mainz, Germany, in summer 2013 (http://joseba.mpch-mainz.mpg.de/mad_cat.htm). Data were distributed to 17 international groups working on ground-based DOAS applications. Each group analysed these spectra using their own DOAS retrieval code but prescribed fit settings for fitting window, cross sections, polynomial and offset correction. Resulting slant columns from all groups were then compared to IUPB results chosen as the arbitrary reference, evaluating the level of agreement and systematic differences and investigating their algorithmic origins. With this set-up, nearly all sources of disagreement were removed, and only those differences between retrieval codes were investigated, which are not usually reported.

This intercomparison exercise concentrates on 1 day (18 June 2013) of the MAD-CAT campaign, which had the best weather and viewing conditions. As an example, Fig. 1 shows NO₂ slant columns (left) and fit rms residual (right) retrieved with the IUPB software. It can clearly be seen that NO₂ slant columns measured at different elevation angles are separated as a result of differences in the light path. It should be mentioned that NO₂ slant columns are relatively large as a result of anthropogenic pollution at the densely populated and industrialized measurement location and thus the findings of this study correspond to polluted urban environments. However, these are normally the ones of interest for NO₂ observations. Figure 1 also shows that the rms of the fit residuals (in the following simply denoted rms) separates with elevation angles as well. In addition, the shape of the rms in small elevation angles is very similar to that of NO₂ slant columns, indicating that predominantly NO₂-related effects such as the wavelength dependence of the NO₂ AMF (which was not included in the fit shown here) limit the fit quality (note that this study aims not at finding best NO₂ fit settings but studying disagreements originating from different retrieval codes). This is supported by the observation that the shape of NO₂ slant columns is not seen in the rms for larger elevation angles as effects related to the NO₂ absorption are smaller and other effects limiting the fit quality dominate. In addition, the rms is very large in the early morning for small elevations, which is discussed in detail later, and was found to result from pointing towards sunrise. However, in general NO₂ is retrieved very accurately as Fig. 2 demonstrates, showing the NO₂ optical density and fit residual for a measurement in 2° elevation angle (this is a typical example fit).

The paper is structured in the following way: Sect. 2 provides details about the MAD-CAT campaign, measurements and the NO₂ intercomparison exercise as well as participating retrieval codes. Intercomparison results between groups are presented in Sect. 3. Sensitivity studies of non-harmonized retrieval aspects are described in Sect. 4, attempting to reproduce observed differences between groups and evaluating the impact on NO₂ slant columns and DOAS fit quality. The paper ends with conclusions and recom-

mendations for better harmonization of ground-based MAX-DOAS measurements.

2 Measurements

2.1 DOAS technique

The DOAS technique is based on Beer–Lambert law, which describes the attenuation of light passing through a medium. Here, it is applied to measurements of scattered sunlight. The spectral attenuation caused by scattering is smooth in wavelength (e.g. λ^{-4} -dependence for Rayleigh scattering) while molecular absorption often has structured spectra. In DOAS, the total spectral attenuation is therefore split into a high-frequency part consisting of the (high-frequency components of) trace gas absorptions and a low-frequency part that accounts for elastic scattering, which is described by a low-order polynomial, also compensating for intensity changes, e.g. caused by clouds. In addition, the effect of inelastic scattering, which is predominantly due to rotational Raman scattering, known as the Ring effect (Shefov, 1959; Grainger and Ring, 1962) is accounted for by a pseudo cross section (e.g. Vountas et al., 1998). Similarly, intensity offsets mostly resulting from stray light within the spectrometer are accounted for by pseudo cross sections σ_{off} (see Sect. 4.3 for more details). In total, the optical depth τ is approximated by

$$\tau = \ln\left(\frac{I_0}{I}\right) = \sum_i \sigma_i \cdot \text{SC}_i + \sigma_{\text{Ring}} \cdot \text{SC}_{\text{Ring}} + \sigma_{\text{off}} \cdot \text{SC}_{\text{off}} + \sum_p a_p \lambda^p + r, \quad (1)$$

where the first sum is over all i absorbers having cross section σ_i , the polynomial degree is p , and the residual term r contains the remaining (uncompensated) optical depth. An important quantity used within this study to identify and evaluate differences between DOAS retrieval codes is the root mean square of the fit residual r (for simplicity denoted rms in the context of this study), which is a measure of the fit quality. In Eq. (1), known as the DOAS equation, all quantities with the exception of a_p depend on wavelength λ . For tropospheric absorbers, the spectrum I is normally taken at small elevation angles above the horizon where the tropospheric light path is large. The reference spectrum I_0 is normally a zenith spectrum either at small solar zenith angle (SZA), which is called a noon reference fit in the following, or close in time to the measured spectrum I , in the following called a sequential reference fit.

The DOAS equation is usually an overdetermined problem (τ is measured at more wavelengths λ than unknowns exist in Eq. 1) and solved by means of a least-squares fit (see also Sect. 4.5), i.e. minimizing the residual term. The resulting fit factors are the polynomial coefficients a_p and the so-called slant columns SC_i , which are the quantities of interest. The slant column is the integrated concentration ρ_i of absorber i

along the effective light path s (for simplicity we use the SC also for the fit coefficients of the Ring and offset spectra):

$$SC_i = \int \rho_i ds. \quad (2)$$

Note that this is a simplification as normally an ensemble of different light paths contribute to the measurement. In extreme cases, the absorption optical depth can become a non-linear function of the trace gas concentration. However, this is usually only of importance for satellite limb measurements and is discussed in detail, for example, in Pukite and Wagner (2016). A comprehensive discussion of the DOAS technique can be found, for example, in Hönninger et al. (2004); Platt and Stutz (2008).

2.2 The MAD-CAT campaign

The Multi-Axis DOAS Comparison campaign for Aerosols and Trace gases (MAD-CAT) was carried out in Mainz, Germany, in summer 2013. During the intensive phase of the campaign (7 June to 6 July 2013), 11 groups deployed MAX-DOAS instruments on the roof of the MAX-Planck Institute for Chemistry at the Mainz University campus. Being located in the densely populated Rhine-Main region, the observations are dominated by anthropogenic pollution, predominantly by NO_2 (in the visible). The main azimuthal viewing direction was 51° (from the north), pointing roughly in the direction of the city of Frankfurt ≈ 30 km away. Series of vertical scans comprising elevation angles of 0, 1, 2, 3, 4, 5, 6, 8, 10 and 30° were performed. In each direction, single measurements of varying exposure time depending on illumination were integrated for 20 s (for the IUPB instrument, other instruments used different integration times). Between vertical scans, multiple zenith measurements were taken (see Fig. 1 for an example of resulting NO_2 slant columns). Detailed information about MAD-CAT can be also found at http://joseba.mpch-mainz.mpg.de/mad_cat.htm and the first campaign data focusing on range-resolved distributions of NO_2 measured at different wavelengths were published by Ortega et al. (2015). Furthermore, Lampel et al. (2015) demonstrated the presence of vibrational Raman scattering on N_2 molecules in spectra measured during MAD-CAT. In addition, publications based on MAD-CAT data are in preparation focusing on HCHO (Pinardi, 2017), HONO (Wang et al., 2016) and CHOCHO (Ortega et al., 2017).

2.3 The IUPB MAX-DOAS instrument

The IUPB MAX-DOAS instrument deployed in Mainz during the MAD-CAT campaign is a two-channel CCD-spectrometer system measuring in the UV and visible. Within this exercise, only data from the visible spectrometer are used as NO_2 is best retrieved in this spectral range. The spectrometer is an ANDOR Shamrock 303i, covering a spectral range from 399 to 536 nm at a resolution of ≈ 0.7 nm. The spec-

trometer was actively temperature stabilized to 35°C . Spectra were recorded with a charge-coupled device (CCD) of the ANDOR iDUS 420 type with 1024×255 pixels ($26 \times 26 \mu\text{m}$ each), leading to a spectral sampling of 7–8 pixels nm^{-1} .

Light was collected by a telescope unit mounted on a commercial ENEO VPT-501 pan-tilt head allowing pointing in any viewing direction. Photons entering the telescope through a fused silica window were focused by a lens on an optical fibre bundle. The instrument's field of view (FOV) was $\approx 1.2^\circ$. The Y-shaped optical fibre bundle (length 20 m) connecting the telescope to both spectrometers consists of $2 \times 38 = 76$ single fibres, minimizing polarization effects. A video camera inside the telescope housing takes snapshots of every recorded spectrum for scene documentation, and a mercury-cadmium (HgCd) line lamp allows for wavelength calibration measurements. Dark current and slit function measurements are taken every night. The same instrumental set-up has been used in previous campaigns, e.g. CINDI and TransBrom (Roscoe et al., 2010; Peters et al., 2012).

2.4 Intercomparison exercise

Spectra recorded by the IUPB MAX-DOAS instrument on 18 June 2013 during MAD-CAT were distributed to partners. It is worth mentioning that this intercomparison exercise was not restricted to groups participating in MAD-CAT, as common observations were provided. 18 June 2013 was selected as having good viewing and weather conditions. However, temperature stabilization of the IUPB spectrometer was problematic due to unusually hot weather which, due to a lack of air conditioning on that day, led to the instrument overheating. As a result, spectral stability was not as good as usual, providing the opportunity to investigate how different retrieval codes deal with potential spectral shifts during the day. The provided data comprised observations at several elevation angles in the main azimuthal viewing direction of 51° (see Fig. 1 for an example). In addition, common trace gas cross sections as well as DOAS fit settings were distributed as summarized in Table 1, which are the common fit settings agreed on during the MAD-CAT campaign. All groups then performed DOAS fits using these prescribed settings on IUPB spectra using their own retrieval software. The influence of different retrievals was then analysed, focusing on the impact on NO_2 columns. A brief description of retrieval codes is given in the following. The participating groups (with abbreviations used within this study) and retrieval codes they used are summarized in Table 2.

2.4.1 NLIN

NLIN (Richter, 1997) was originally developed at IUPB for the analysis of ground-based measurements but over time extended and also used for airborne and satellite data sets. The DOAS matrix is inverted using a singular value decompo-

Table 1. Summary of fit settings used for the NO₂ intercomparison. These fit settings were agreed on during the MAD-CAT campaign and can be found also at http://joseba.mpch-mainz.mpg.de/mad_analysis.htm.

Fit	Reference	Window	Cross sections	Intensity offset	Polynomial
v1	noon	425–490	1, 2, 4, 5, 6, 7	constant (0th order)	5 (6 coefs)
v1a	sequential	425–490	1, 2, 4, 5, 6, 7	constant (0th order)	5 (6 coefs)
v2	noon	411–445	1, 3, 4, 5, 6, 7	constant (0th order)	4 (5 coefs)
v2a	sequential	411–445	1, 3, 4, 5, 6, 7	constant (0th order)	4 (5 coefs)
Cross sections					
1	NO ₂ at 298 K (Vandaele et al., 1996), I_0 -correction using 1×10^{17} molec cm ⁻²				
2	NO ₂ at 220 K orthogonalized to 298 K within 425–490 nm				
3	NO ₂ at 220 K orthogonalized to 298 K within 411–445 nm				
4	O ₃ at 223 K (Bogumil et al., 2003)				
5	O ₄ Hermans et al., unpublished, http://spectrolab.aeronomie.be/o2.htm				
6	H ₂ O, HITEMP (Rothman et al., 2010)				
7	Ring, NDSC2003 (Chance and Spurr, 1997)				

Table 2. Summary of participating institutes (with abbreviation) and retrieval codes used by each group.

Abbr.	Retrieval(s)	Institute
IUPB	NLIN	Institute for Environmental Physics, University of Bremen, Germany
AUTH	QDOAS	Aristotle University of Thessaloniki, Greece
BIRA	QDOAS	Belgian Institute for Space Aeronomy, Brussels, Belgium
JAMSTEC	QDOAS	Japan Agency for Marine-Earth Science and Technology, Japan
Toronto	QDOAS	Department of Physics, University of Toronto, Ontario, Canada
IUPHD	DOASIS	Institute of Environmental Physics, University of Heidelberg, Germany
Boulder	QDOAS	University of Colorado, Boulder, USA
KNMI	KMDOAS	Royal Netherlands Meteorological Institute, De Bilt, The Netherlands
INTA	LANA	National Institute for Aerospace technology, Madrid, Spain
MPIC	MDOAS, WinDOAS	Max Planck Institute for Chemistry, Mainz, Germany
CSIC	QDOAS	Department of Atmospheric Chemistry and Climate, Institute of Physical Chemistry Rocasolano, CSIC, Madrid, Spain
NIWA	STRATO	National Institute of Water and Atmospheric Research, Lauder, New Zealand
IAP	RS.DOAS	A. M. Obukhov Institute of Atmospheric Physics, Russian Academy of Sciences, Moscow, Russia
BSU	WinDOAS	Belarusian State University, Minsk, Belarus
USTC	QDOAS	University of Science and Technology, Hefei, China
UNAM	QDOAS	National Autonomous University of Mexico, Mexico
NUST	QDOAS	Institute of environmental sciences and engineering (IESE), National University of Sciences and Technology, Islamabad, Pakistan

sition (SVD) which is comprised in an iterative non-linear fit for calibration of the wavelength axis using a Marquardt–Levenberg fit.

2.4.2 QDOAS

QDOAS (Dankaert et al., 2013) is the multi-platform (Windows, Unix/Linux and Mac) successor of WinDOAS. A coupled linear/non-linear least squares (NLLS) algorithm (Marquardt–Levenberg fitting and SVD decomposition) is used to solve the DOAS equation including a wavelength calibration module. During this operation, the slit function can be characterized in addition to the wavelength registra-

tion of the measured spectra. QDOAS is widely used within the DOAS community and has been applied in different versions for the present exercise: AUTH (version 2.109.3), JAMSTEC (version 2.109), University of Toronto (version 2.109), CSIC (QDOAS windows 2.110.1 beta 20151123), BIRA (version 2.110.1), USTC (version 2.109.4), UNAM (version 2.109.3), NUST (version 2.111), CU Boulder (version 2.110). These various versions do not differ in the core implementation of the DOAS analysis, but can feature slightly different capabilities in terms of data handling or specific retrieval functionalities (see <http://uv-vis.aeronomie.be/software/QDOAS/LastChanges.php>).

2.4.3 LANA

LANA is a two-step iterative algorithm developed at INTA and used since 1994. In the first step, cross sections and spectra are positioned and in the second step the linear equations system is solved using a Gauss–Jordan procedure. For this exercise, the intensity offset correction was based on the reference spectrum I_0 .

2.4.4 MDOAS

MDOAS is a Matlab DOAS code developed at MPIC, but calibration and convolution have been performed in WinDOAS. For sequential reference fits, two versions, MPIC_MDa and MPIC_MDb, are included in this exercise, which used a different treatment of the Ring spectrum.

2.4.5 KMDOAS

KMDOAS was recently (2013–2014) developed at KNMI and verified using QDOAS. It is written in Python using standard modules (matplotlib, numpy, scipy, pandas).

2.4.6 DOASIS

DOASIS has been used by IUPHD in its version 3.2.4595.39926. A detailed explanation can be found in (Kraus, 2006). In the version submitted here, an additional shift of cross sections with respect to the optical depth was allowed and a saturation correction was implemented.

2.4.7 STRATO

The NIWA-Strato package was originally (1980s) used for processing zenith DOAS spectra, but extended as needed to handle MAX-DOAS measurements. The fitting code uses least squares or optional SVD fitting inside iterations applying shift, stretch and (optional) offset, for minimum residual. In contrast to other groups, an internal equidistant wavelength grid is applied using the average interpixel spacing.

2.4.8 RS.DOAS

RS.DOAS (Ivanov et al., 2012; Borovski et al., 2014; Postlyakov et al., 2014; Postlyakov and Borovski, 2016) was developed by IAP. The DOAS inversion is performed by LU decomposition. The wavelength calibration uses several subwindows, and for each of them a non-linear shift and stretch fit including all trace gases (resulting from the linear DOAS fit) is applied. From a set of shifts obtained for the subwindows a 2nd-order polynomial approximation of the shift dependency on pixel number is constructed. For convoluting cross sections the QDOAS software (offline version 2.0 at 5 March 2012) was used.

2.4.9 WinDOAS

WinDOAS (Fayt and Van Roozendaal, 2001) is the precursor of QDOAS. It has been used by BSU as well as MPIC for this intercomparison exercise (the MPIC submission is denoted MPIC_WD, the reference has been selected by hand), but only for noon reference fits.

3 Intercomparison results

3.1 Noon reference, 425–490 nm fit window (v1 fit parameters)

Differences between groups for the 425–490 nm fit using a noon reference (v1 fit settings; see Table 1) are shown in Fig. 3 for individual measurements at an elevation angle of 2° . Small elevation angles above the horizon are associated with long tropospheric light paths and are therefore important for the detection of tropospheric absorbers. Differences shown in Fig. 3 are relative to IUPB results. For the objective of identifying retrieval-code-specific effects, the use of a single retrieval code as a reference seems advantageous in comparison to using the mean of all retrieval codes which would average over all retrieval-specific features. Note that this does not exclude IUPB from the intercomparison as problems with the IUPB retrieval would be easily detected, leading to the same systematic patterns in all lines shown in Fig. 3.

Absolute differences (institute-IUPB) and relative differences (absolute difference/IUPB) of NO_2 slant columns and fit rms are shown in Fig. 3a–d. In general, NO_2 slant column differences are in the range of $\pm 2\text{--}3 \times 10^{15}$ molec cm^{-2} or $< 2\%$. This is about a factor of 2–3 larger than NO_2 slant column errors, which are typically $< 1 \times 10^{15}$ molec cm^{-2} or $< 0.6\%$ for 2° elevation. A clearly enhanced disagreement is observed for the first data point in the INTA time series as well as for MPIC_WD. The latter could be linked to the reference spectrum, which in this case is not the one with the smallest sun zenith angle (SZA), while the outlier in the INTA time series was identified to arise from different implementations of the intensity offset correction (see Sect. 4.3). Note that these NO_2 differences for individual measurements are much smaller than the variability (diurnal cycle) of NO_2 and thus almost invisible in Fig. 3e where no differences but absolute NO_2 slant columns from each groups (including IUPB in black) are plotted.

Interestingly, many groups show a smooth behaviour (constant offset) in absolute NO_2 differences (Fig. 3a), which is mostly an effect of the choice of the reference (see Sect. 4.1), while relative differences (Fig. 3c) reflect the shape of NO_2 slant columns in Fig. 3e (smaller slant columns lead to larger relative differences and vice versa). However, some groups show a smooth line, not for absolute differences but for relative differences. Thus, two types of disagreements are observed, (1) constant in absolute differences and (2) constant

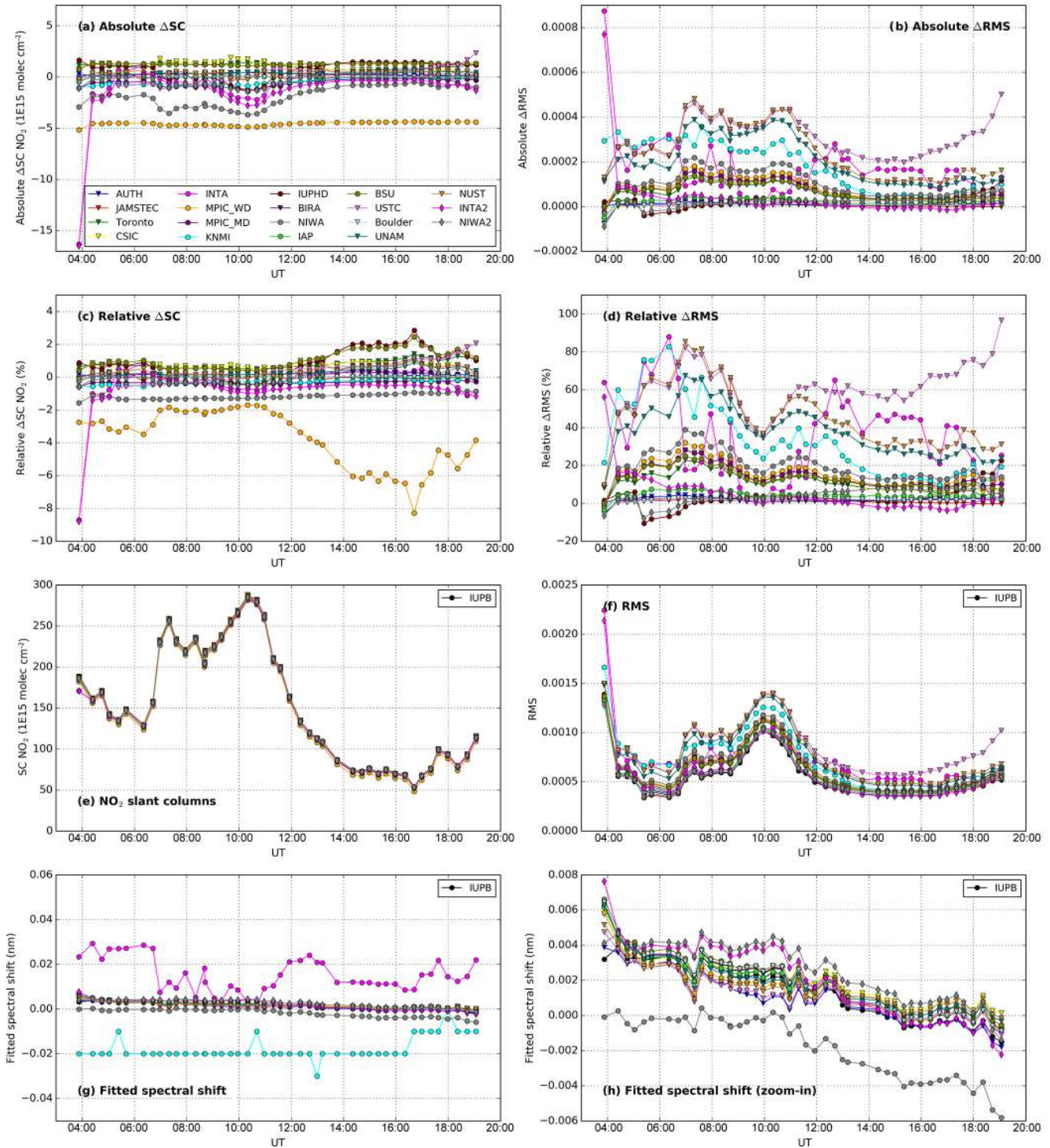


Figure 3. Results from v1 fit settings in 2° elevation angle as a function of time. **(a, b)** Absolute NO_2 slant column and rms differences, **(c, d)** relative NO_2 slant column and rms differences, **(e)** NO_2 slant columns, **(f)** rms, **(g, h)** fitted spectral shift (**h** is a zoom-in of **g** without INTA and KNMI lines, but including the reanalysed INTA2 and NIWA2 lines). Panels **(a)–(d)** are differences with respect to IUPB, panels **(e)–(h)** are absolute results and IUPB is explicitly shown in black. Triangular markers indicate groups using QDOAS, circular markers indicate groups using independent software, and diamonds indicate groups that corrected faults in their code during the course of this study and submitted a reanalysed data set.

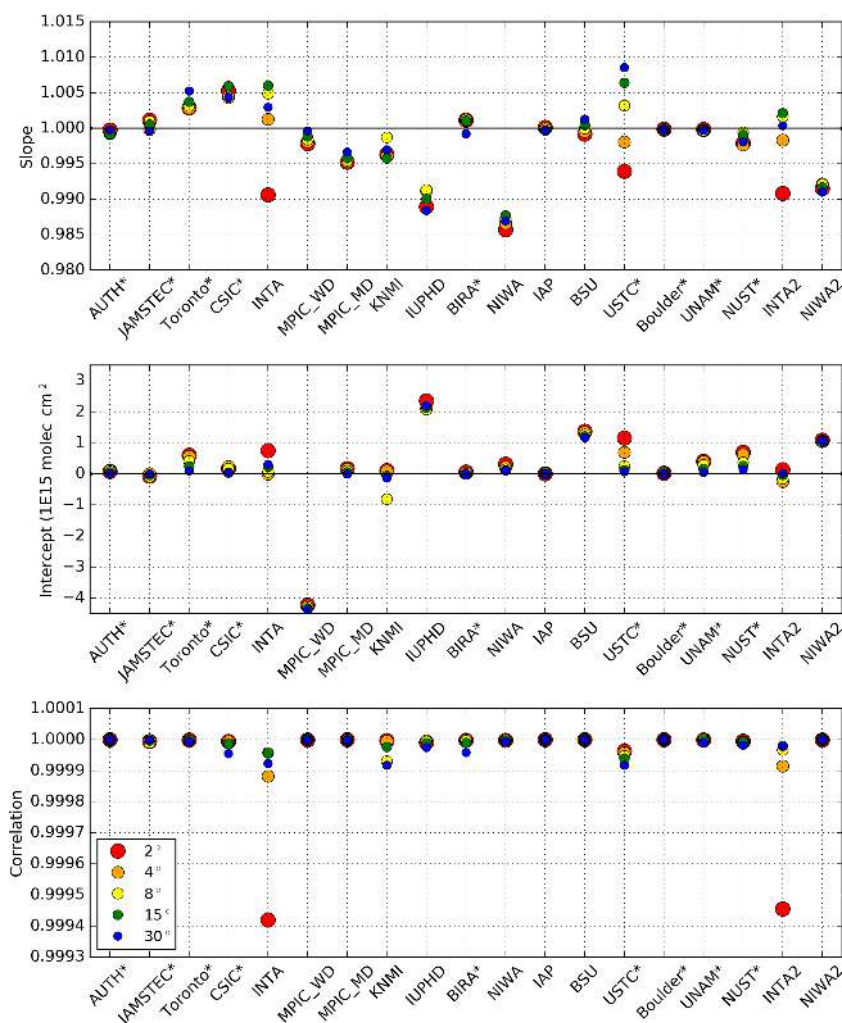


Figure 4. Linear regression results (slope, intercept, correlation) for different elevation angles (with respect to IUPB) for fit settings v1 (noon reference). Groups using QDOAS (in different versions) are indicated by a star (*).

in relative differences. These two types are linked to differences in retrieval codes, which are investigated in detail in Sect. 4.

Absolute rms differences (with respect to IUPB) in Fig. 3b show the same shape as NO_2 slant columns. This is because at small elevations the rms itself reflects the shape of the NO_2 , which was already demonstrated and discussed in Fig. 1. A better measure for the identification of differences between retrieval codes is thus the relative rms disagreement shown in Fig. 3d. Interestingly, the first data point for INTA, showing a large disagreement with IUPB and other groups in NO_2 slant columns is prominent in absolute rms differences as well, but not in relative rms differences. The reason is that the rms in the morning is very large (Fig. 3f, compare also to Fig. 1) and thus decreases the relative difference. Remarkably, relative rms differences are found up to 80–100 %, which is substantially more than NO_2 slant column differences (only a few percent).

The IUPB spectra provided were wavelength precalibrated using nightly HgCd line lamp measurements as explained in Sect. 2.3. However, the DOAS fit quality can be improved (rms reduced) by applying a postcalibration. In addition, the night-time calibration can change during the day as a result of temperature drifts. This is accounted for by DOAS retrieval codes in terms of a non-linear shift fit (for a more detailed discussion see Sect. 4.6). Figure 3g and h (being a zoomed-in image of g) show the reported shift resulting from the wavelength calibration in participating retrieval codes. Note that absolute values for the shift are shown here and the IUPB result is explicitly included. For an ideal spectrometer without drifts, only a very small shift would be expected and it would be caused by a non-commutivity of convolution and DOAS polynomial, which is known as the tilt effect and typically in the order of less than 1–2 pm, depending on the instrument resolution (Sioris et al., 2003; Lampel et al., 2017). The shift retrieved here is larger than that and driven by the overheat-

ing of the system on this day. In general, time series of shifts shown in Fig. 3g and h agree well in shape. A small offset is observed for NIWA, who successfully found and corrected a fault in their retrieval code (related to the intensity offset calculation) in the course of this exercise. The DOAS fit was then repeated using the corrected retrieval code and the result, called NIWA2, is included in Fig. 3. Apparently, NIWA2 shows a much better comparison than other groups (most obvious in the fitted spectral shift) and performs better because the rms in Fig. 3d is much smaller than before. Similarly, INTA fixed a problem in their retrieval code related to the wavelength calibration. The reanalysed INTA data are denoted INTA2 in Fig. 3. The INTA2 shift is comparable to other groups, while large disagreements were seen before, and the rms is largely reduced in comparison to the first INTA submission. Notably, almost no differences between INTA and INTA2 are observed in terms of NO₂ slant columns; in particular the first data point is still outlying. This indicates (1) that the wavelength calibration has a large impact on the fit rms but only a minor impact on the NO₂ slant columns, and (2) that the disagreement of the first data point is not related to the wavelength calibration. Despite INTA and NIWA, only KNMI is fitting a rather different shift (and provides the shift only in 0.01 nm resolution which leads to the displayed discrete steps). This is caused by different definitions of the shift (see Sect. 4.6): KNMI is fitting the shift of the optical depth relative to the cross sections while the other groups are fitting the shift of I relative to the reference spectrum I_0 . As a result, the shift of KNMI and other groups cannot be expected to match.

The agreement between groups, including other elevation angles, has been quantified by linear regression analyses to correlation plots (of each group relative to IUPB results, not shown). Resulting slopes, offsets and correlation coefficients are summarized in Fig. 4 and colour-coded for different elevation angles. As expected, the correlation coefficient is > 99.98 % with the only exception of the INTA (and INTA2) 2° elevation which is predominantly caused by the outlier already seen in Fig. 3. The slope ranges between 0.985 and 1.01, the offset between -4 and 2.5×10^{15} molec cm⁻². Apart from USTC and INTA, no large separation of the slope and offset with elevation angle is observed. An important observation is that groups using the same retrieval code (QDOAS) do not necessarily show the same systematic behaviour in Fig. 4, implying that the influence of remaining fit parameters that are different from the harmonized general settings in Table 1 is still larger than the effect of the specific retrieval software used. For example, the best agreement in terms of slope, offset and correlation coefficient is found between IUPB, AUTH, IAP and CU Boulder, which all use different retrieval codes.

3.2 Other fit parameters

The agreement between groups was investigated in the same way when using a sequential reference (denoted v1a fit setting; see Table 1) instead of a noon reference spectrum (v1 fit). A sequential reference is often preferred if tropospheric absorbers are of interest as stratospheric effects are removed to a large extent.

Figure 5 is similar to Fig. 3, but some groups are missing here (e.g. IAP) as they only provided noon reference fits. The range of disagreements for individual NO₂ slant columns is up to ≈ 8 % and therefore larger than NO₂ disagreements using a noon reference (v1 fit). In addition, neither absolute nor relative differences between groups are smooth (Fig. 5a and c). The reason is that, in contrast to the noon reference, a different (zenith) I_0 spectrum is used for every scan and thus the impact of details of the implementation of the reference changes from scan to scan. Different implementations of the reference spectrum are further investigated in Sect. 4.1 and were found to be the major reason for NO₂ disagreements between groups. The outlier (first data point) seen for v1 (noon) fit settings is present for v1a as well but it is not prominent here as fluctuations for the above-mentioned reason are of same magnitude.

In terms of rms, differences between groups are comparable to v1 results and as large as ≈ 80 %.

Compared to the noon reference fit, correlations from linear regressions (not shown) are smaller, especially for the 30° elevation angle, mainly resulting from the larger disagreements as explained above (but also partly expected as slant columns are smaller using a sequential reference). Correlations are still > 99.2 % for 30° elevation and even > 99.8 % for smaller elevations.

In addition to the 425–490 nm fit (v1 and v1a), a smaller fit window used within the MAD-CAT campaign was inter-compared (v2 and v2a using a spectral range of 411–445 nm; see Table 1). Results of these fits are not explicitly shown as mainly confirming observations and findings above. Typical correlation coefficients (as well as offsets and slopes) are summarized in Table 3 for all fits. A very small tendency of better agreement between groups if using the larger fit window is seen. Although this should not be overinterpreted, it could be caused by more information being present in the large fit window and therefore more accurate results (fewer statistical fluctuations).

4 Sensitivity studies of non-harmonized retrieval aspects

The observed differences between the harmonized MAD-CAT retrieval results prompted the team to systematically investigate effects of the non-harmonized aspects of the retrievals, with the aim to derive a key set of best practice recommendations. A survey suggested five potential

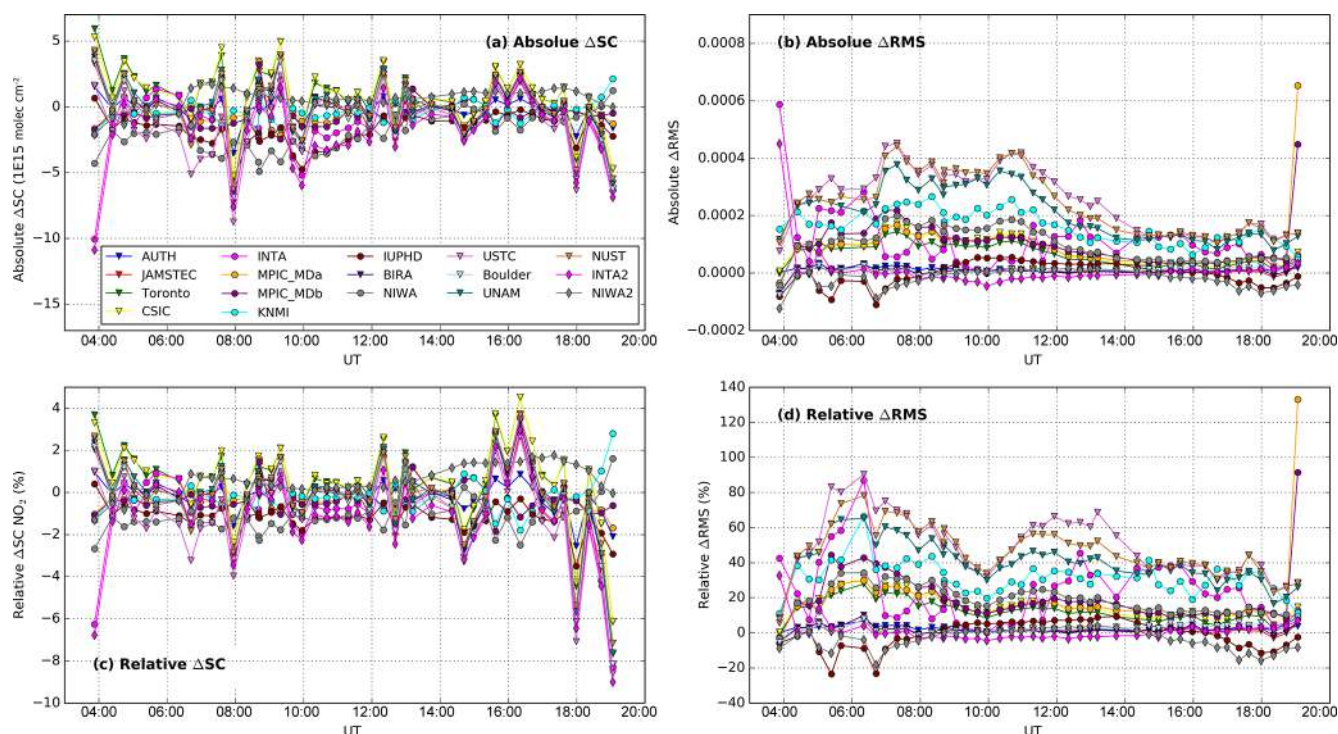


Figure 5. Results from v1a fit settings (sequential reference spectrum) on 2° elevation angle data as a function of time. (a, b) Absolute NO_2 slant column and rms differences, (c, d) relative NO_2 slant column and rms differences. All differences are relative to IUPB. Triangular markers indicate groups using QDOAS, circular markers indicate groups using independent software, and diamonds indicate groups that corrected faults in their code during the course of this study and submitted a reanalysed data set.

Table 3. Correlations, slopes and offsets from linear regressions on NO_2 slant columns between each group and IUPB (neglecting the INTA 2° outlier).

Fit	Correlation (%)	Slope ($1 \times 10^{15} \text{ molec cm}^{-2}$)	Offset
v1	> 99.98	0.985 to 1.01	−4 to 3
v1a	> 99.2	0.96 to 1.01	−1.5 to 1
v2	> 99.94	0.985 to 1.005	−4 to 3
v2a	> 99.2	0.96 to 1.01	−2 to 1

sources of differences in the DOAS fit results: (1) the selection/calculation of the reference spectrum, predominantly for sequential reference fits, (2) treatment of the slit function, (3) the intensity offset correction, (4) differences in the numerical calculation of the DOAS fit (linear fit) and (5) differences in the wavelength calibration (non-linear fit). In the following, tests using the same retrieval code (IUPB) have been performed to characterize the impact of each of the above-mentioned systematic differences.

4.1 Effect of the reference spectrum

In the IUPB spectra provided to intercomparison partners, two different zenith spectra at the same smallest SZA were

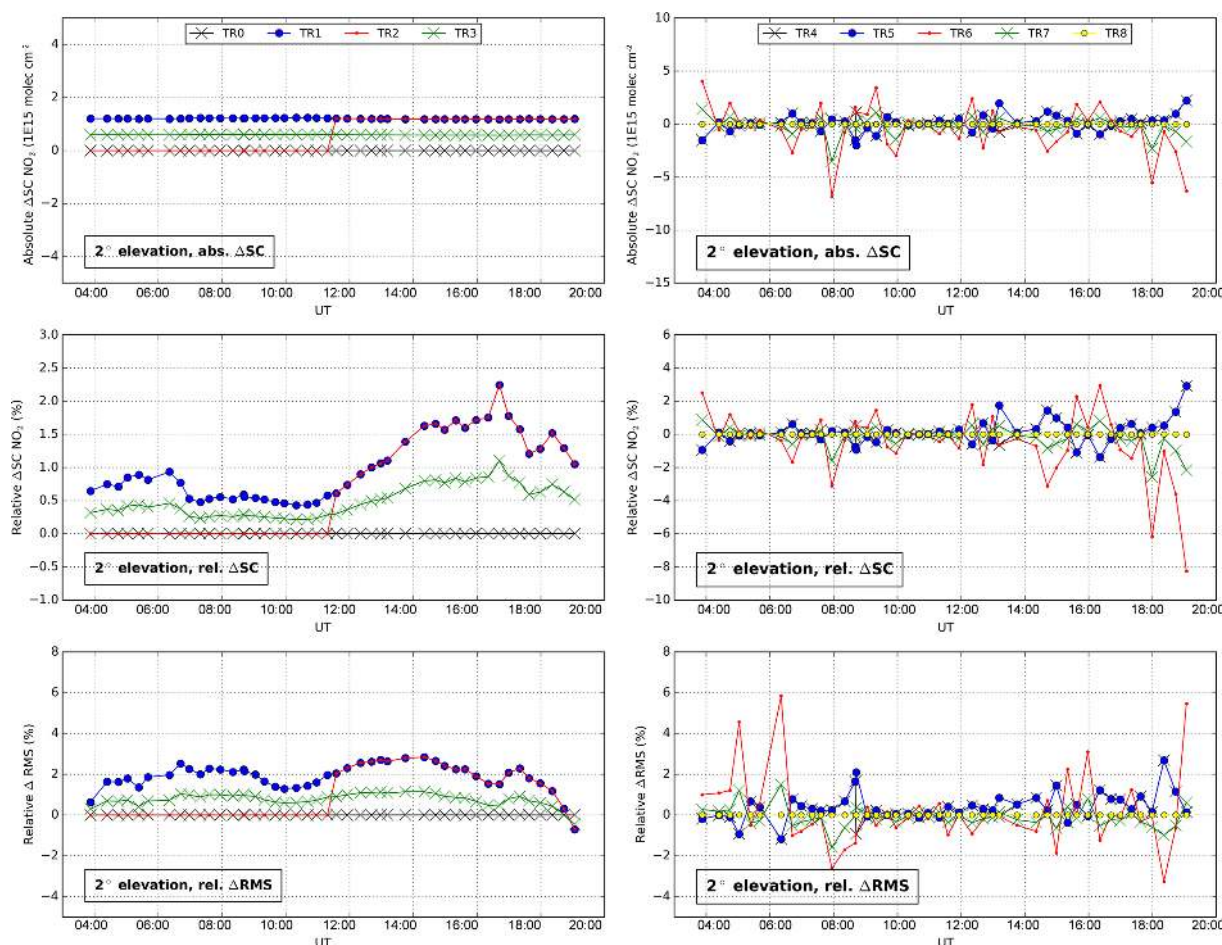
reported, which is of course non-physical. Actually, the second zenith spectrum is the one with the smallest SZA, but for rounding reasons (the SZA was a 4-digit number in the input data provided), both spectra had the same SZA. Consequently, for the noon reference fits v1 and v2 within this intercomparison exercise, four options exist: (1) taking the first zenith spectrum of smallest SZA, (2) taking the second one, (3) taking the first for the morning and the second for the afternoon (i.e. always taking the closest in time) and (4) taking the average of both spectra. Similarly, different options exist for calculating the sequential references for fits v1a and v2a: (1) always taking the zenith spectrum closest in time, (2) always taking the last zenith spectrum before the actual measurement, (3) always taking the next zenith spectrum after the measurement, (4) taking the average of the two (before and after) and (5) interpolating the two zenith spectra to the time of the actual measurement.

All different options for noon and sequential reference fits were evaluated using the IUPB retrieval code NLIN (Table 4). Figure 6 shows the resulting absolute and relative slant column differences (top and middle) as well as relative rms differences (bottom) for noon reference (left) and sequential reference (right) with respect to v1 or v1a fit results.

Taking another noon reference spectrum results in a constant offset in absolute NO_2 differences (Fig. 6, top left).

Table 4. Tests performed to study the influence of different reference spectra.

Test	Reference (fit setting)	Remarks
TR0	noon (v1)	first spectrum of smallest SZA
TR1	noon (v1)	second spectrum of smallest SZA
TR2	noon (v1)	first for morning, second for afternoon
TR3	noon (v1)	average of both
TR4	sequential (v1a)	closest zenith in time
TR5	sequential (v1a)	closest before the scan
TR6	sequential (v1a)	closest after the scan
TR7	sequential (v1a)	average of before and after
TR8	sequential (v1a)	interpolation of before and after to time of measurement

**Figure 6.** Different test results (see Table 4) in 2° elevation angle as a function of time. Top panel shows absolute NO_2 SC differences, middle shows relative SC differences, bottom shows relative rms differences. Left is for noon reference (differences are with respect to IUPB v1 fit results), right is for sequential references (differences are with respect to IUPB v1a fit results).

Test TR0 (using the first spectrum) yields the same results as the IUPB v1 fit from Sect. 3, which uses the first zenith spectrum as well. In contrast, using the second zenith spectrum as a reference (TR1) results in a constant offset of 1.5×10^{15} molec cm^{-2} (0.5–2.5 % in relative differences, depending on the actual NO_2 slant column), because the second

zenith spectrum apparently had a smaller NO_2 content. The change of the NO_2 content could be related to changes of the atmospheric NO_2 amount or the atmospheric light path. In terms of rms, TR1 is up to 3 % larger (Fig. 6, bottom left). The main reason for this is probably the larger NO_2 slant columns as associated effects like the wavelength depen-

dence of the NO₂ slant column (Pukite et al., 2010) were not compensated in this intercomparison exercise as discussed in Sect. 1.

Test TR2 yields results which are identical to TR0 am and TR1 pm values. This is not seen for any groups in Fig. 3 above; i.e. this option is apparently not present in any retrieval code. Not surprisingly, TR3 (averaging both zenith spectra) yields results which are between TR0 and TR2.

In contrast to noon reference tests, sequential references show no smooth behaviour, neither for absolute nor for relative differences (Fig. 6, right). The reason is that for each vertical scanning sequence (from which only the 2° elevation is shown here and in Fig. 3) another reference spectrum is used and consequently reference-related differences between groups are also changing from scan to scan. Note that almost no difference can be seen in Fig. 6 between TR4 and TR5 as the 2° elevation angle is shown here and consequently the closest zenith measurement in time is normally the one before the scan as IUPB measurements proceed from low to large elevations. TR8 (interpolation to the measurement time) resembles the sequential reference treatment normally implemented in the IUPB code; thus the TR8 line is zero. For TR6 and TR7, absolute and relative differences (up to 8 %) are remarkably similar to observed differences between groups using v1a fit settings, both in shape and in absolute values (compare to Fig. 5).

To conclude, the exact treatment of reference spectra is the major reason for observed NO₂ discrepancies between groups in Sect. 3, causing differences of up to 8 %. Unfortunately, no clear recommendation can be derived from relative rms differences in Fig. 6 (bottom right) as all lines scatter around zero. However, the relative difference in rms can be as large as 6 % and in general, using a single zenith reference spectrum before or after the scan tends to produce larger rms. However, while the reference treatment explains the majority of NO₂ disagreements, it cannot explain the large rms differences (up to 100 %) between groups.

4.2 Slit function treatment

The slit function distributed to intercomparison participants originated from HgCd line lamp measurements made in the night before 18 June 2013. It was preprocessed in terms of subtraction of the (also measured) dark signal, centred and provided on an equidistant 0.1 nm grid.

While some groups used this slit function as is, other retrieval codes include a further online processing of the slit function, e.g. by fitting line parameters. In order to evaluate resulting differences, trace gas cross sections have been convolved offline (before the fit) using different treatments of the slit function as summarized in Table 5. Again, the IUPB retrieval code NLIN has been used then performing the DOAS fit.

Examples of different treatments of the slit function are shown in Fig. 7. The original slit function is displayed in

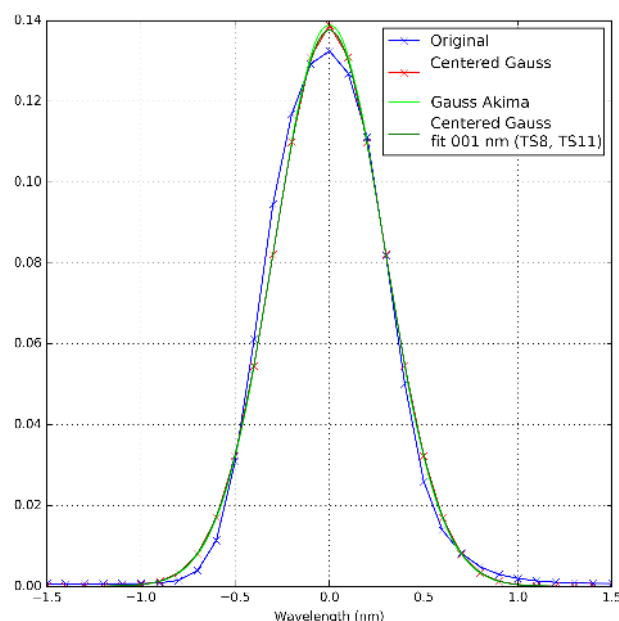


Figure 7. Examples of different slit function treatments. Blue is the measured slit function (from HgCd line at ≈ 480 nm). Red is the fit of a Gaussian shape with $\mu = 0$. Light green is the Akima interpolation of the red line to a finer 0.01 nm grid. Dark green is the interpolation first (linearly) to 0.01 nm and fit of a Gaussian shape afterwards.

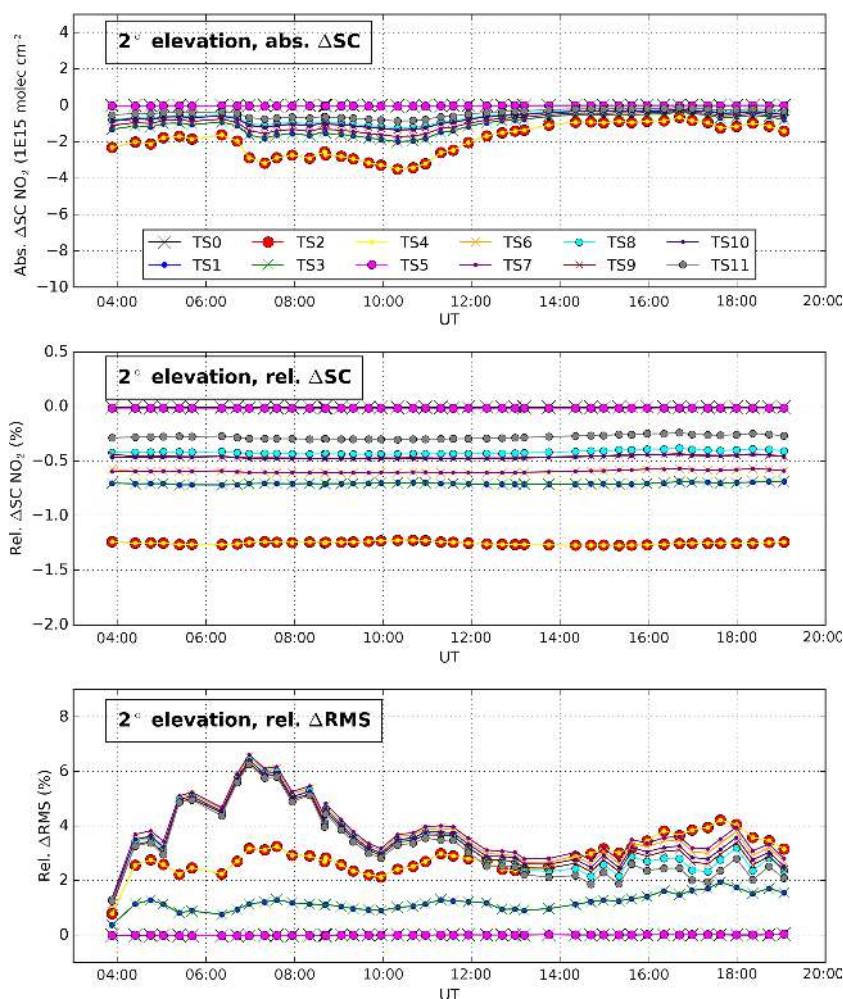
blue. If fitting a Gaussian shape to it, the maximum is not exactly centred around zero, presumably caused by a small asymmetry. The slit function after fitting a Gaussian shape and forcing centring (i.e. $\mu = 0$) is shown in red (as used in TS7 and TS10). Furthermore, performing a discrete convolution of cross sections requires the same wavelength sampling; i.e. the slit function has to be interpolated to the cross section grid, which was 0.01 nm. Here, an Akima interpolation has been applied (green line). However, if the original slit function is first linearly interpolated to the required 0.01 nm grid and then a Gaussian shape is fitted, a slightly different result is obtained, shown in dark green (as used in TS8 and TS11). Note that often not a discrete convolution but a (faster) convolution using Fourier transformation is implemented in DOAS retrieval codes (see below).

Figure 8 shows resulting absolute and relative differences in NO₂ slant columns (top and middle) as well as relative differences of the fit rms (bottom) with respect to the IUPB v1 fit (without I0-correction as this was not applied to the slit function test fits). The reference IUPB v1 fit uses an online convolution of cross sections using FFT and a further geometrical centring of the slit function is applied.

No difference between the reference fit, TS0 and TS5 is observed, neither for NO₂ nor for the rms. In TS0, the same slit function treatment is applied as in the reference fit, but using discrete convolution instead of FFT, meaning that no difference arises due to the method of convolution. In ad-

Table 5. Different treatments of the slit function.

Test	Sub. offset	Geom. centring	Cut-off value	Fit line parameters
TS0	no	yes	0.0	no
TS1	no	no	0.001	no
TS2	yes	no	0.001	no
TS3	no	yes	0.001	no
TS4	yes	yes	0.001	no
TS5	no	no	0.0	no
TS6	no	no	0.001	Gaussian
TS7	no	no	0.001	Gaussian ($\mu = 0$)
TS8	no	no	0.001	Gaussian ($\mu = 0$) fitted on 0.01 nm grid
TS9	no	no	0.0	Gaussian
TS10	no	no	0.0	Gaussian ($\mu = 0$)
TS11	no	no	0.0	Gaussian ($\mu = 0$) fitted on 0.01 nm grid

**Figure 8.** NO_2 slant columns and rms differences with respect to IUPB v1 fit results in 2° elevation angle for different treatments of the slit function.

dition, TS5 and TS0 are identical tests except that TS0 applies an explicit geometrical centring (i.e. centring the area) of the slit function. As this has no visible effect on NO_2 slant

columns and rms, the centring during preprocessing was already sufficient (or the non-linear shift fit was compensating for small deficits in the slit function centring).

Table 6. Tests performed for different implementations of the intensity offset correction (and I_0 correction). A 0th-order offset correction means applying a constant term only, a 1st-order correction means applying a constant and slope term.

Test	Offset order	Offset approach	I_0 -correction
TI0	0th	$1/I$	yes
TI1	0th	$1/I_0$	yes
TI2	0th	Eq. (4) using I	yes
TI3	0th	Eq. (4) using I_0	yes
TI4	None	No offset correction	yes
TI5	1st	$1/I, \lambda/I$	yes
TI6	0th	$1/I$, additionally Ring $\times \lambda$	yes
TI7	0th	$1/I$ (simple approach)	no

The largest impact on NO_2 (up to -1.3%) is seen for TS2 and TS4 which both subtract the smallest value from the slit function (i.e. forcing the slit function to be zero for the smallest value measured). This is usually not performed in DOAS fits and only advisable if instrumental stray light is a large problem or if the dark signal drifts which is normally not the case in state-of-the-art CCD detectors that are cooled and temperature-stabilized using Peltier elements.

In contrast, the largest effect in terms of rms (up to 6.5%) occurs not for TS2 and TS4, but for all tests using a fitted Gaussian. Interestingly, the Gaussian tests (TS6–TS11) show almost no difference among them in terms of rms in the morning, but split up towards the evening. This might have to do with decreasing NO_2 slant columns in the afternoon or changes of the slit function during the day. In terms of NO_2 , all tests using Gaussian slit functions yield smaller NO_2 slant columns of 0.2 – 0.7% .

In addition to the basic Gaussian line parameters fitted here, some retrieval codes offer the possibility of fitting more sophisticated line shape parameters while taking into account potential asymmetry much better (the slit function in Fig. 7 shows indeed slight asymmetry). However, the exact implementations differ and were therefore not reproduced in the tests performed here, which can then be regarded as two extreme scenarios, namely (1) using the (original, slightly asymmetric) slit function as it is, or (2) fitting basic line parameters (a more sophisticated fit taking into account the slit function's slight asymmetry would clearly lead to a result in between these scenarios).

An important finding is that all performed tests lead to constant offsets in relative (and not absolute) NO_2 differences. In addition, all tests yield larger rms than the reference fit using the measured slit function.

4.3 Intensity offset correction

Photons may hit the CCD detector at locations not corresponding to their wavelength (e.g. through scattering on mir-

rors, surfaces etc. inside the spectrometer) which produces an intensity offset, also called stray light. In addition, other effects such as changes in the dark current can lead to intensity offsets and the vibrational Raman scattering (VRS) is known to produce spectral effects that are very similar to intensity offsets (Peters et al., 2014; Lampel et al., 2015). In the DOAS fit, pseudo cross sections are usually included in order to compensate for intensity offsets. If the measured spectrum I is superimposed by a constant intensity C (which is the most simple assumption), the optical depth reads

$$\begin{aligned}
 -\tau &= \ln\left(\frac{I+C}{I_0}\right) = \ln\left(\frac{I}{I_0}\right) + \ln\left(1 + \frac{C}{I}\right) \\
 &\approx \ln\left(\frac{I}{I_0}\right) + \frac{C}{I},
 \end{aligned} \quad (3)$$

with the Taylor expansion $\ln(1+x) \approx x$. Thus, in a first approximation the intensity offset causes an additive term of optical depth that is proportional to $1/I$, which is often used as a pseudo absorber (and showing large similarities to the Ring cross section as this compensates a filling-in of Fraunhofer lines).

However, often more sophisticated approaches are used. For example, the IUPB retrieval code NLIN allows either the simple implementation of $\sigma_{\text{off}} = 1/I$ or

$$\sigma_{\text{off}} = \ln\left(\frac{I+C \cdot I_{\text{max}}}{I}\right), \quad (4)$$

omitting the Taylor expansion in Eq. (3) and superimposing I by a certain constant C of the maximum intensity within the fit interval. In addition, higher correction terms are also sometimes used, assuming that, not only does a constant superimposes the spectrum, but a contribution also changes with wavelength (in which case often λ/I is included in the DOAS fit). Furthermore, sometimes the offset is not included in the linear DOAS fit, but fitted non-linearly (this is not included in tests performed here).

Different implementations summarized in Table 6 were tested in order to evaluate the influence of the intensity offset correction. Figure 9 shows resulting absolute and relative differences of NO_2 slant columns and rms values. Again, the reference for these differences is the IUPB v1 (noon) fit.

No difference in NO_2 or rms is seen between TI0 and the v1 fit as both fits use the same simple approach of $\sigma_{\text{off}} = 1/I$. In contrast, TI1 uses $1/I_0$. In terms of NO_2 differences, the TI1 line slightly follows the shape of total fit rms and NO_2 slant columns (compare to Fig. 3). Interestingly, while the disagreement of TI1 with respect to the reference fit is on average $\approx 2\%$ for NO_2 , the first data point is clearly off by almost 10% for NO_2 and 60% for rms. This agrees perfectly with the observed outlier in the INTA analysis (compare to Fig. 3); i.e. the reason for this disagreement could be identified as a different offset implementation (which has been verified by INTA). The SZA and the sun azimuth angle (SAA)

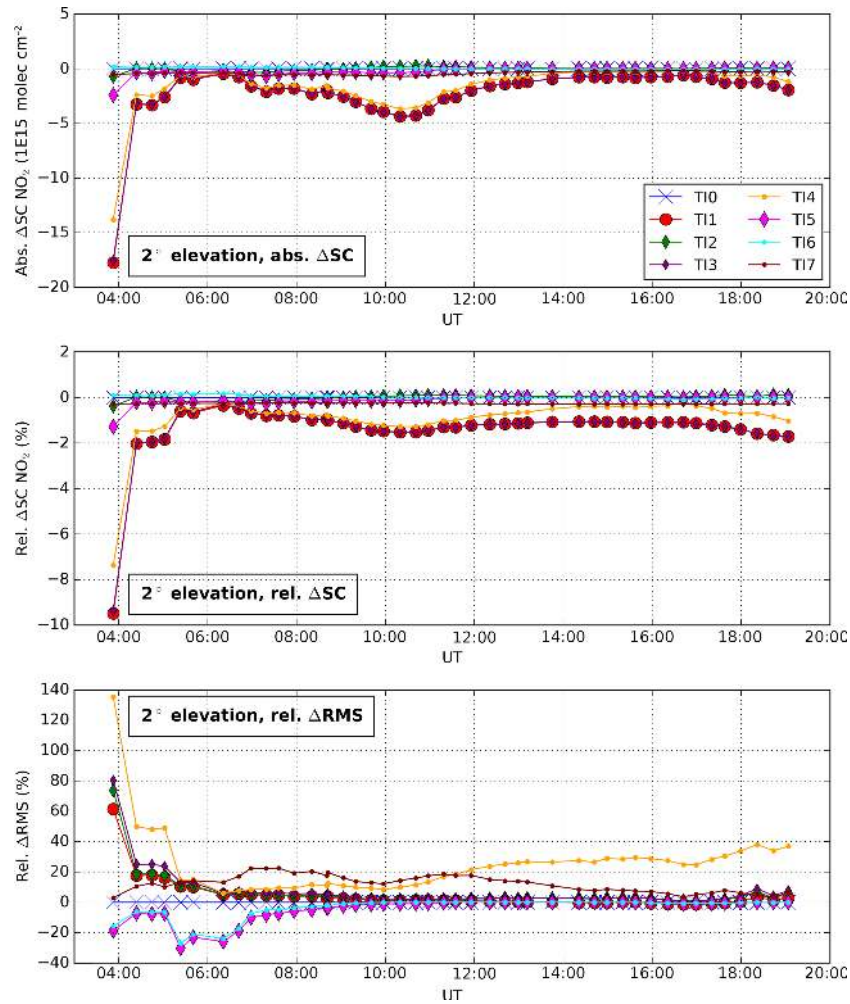


Figure 9. Absolute (top) and relative (middle) NO_2 slant column differences and relative rms differences (bottom) with respect to IUPB v1 fit results in 2° elevation angle resulting from different implementations of the intensity offset correction (and I_0 correction).

of this measurement are ≈ 90 and 58° (from the north), respectively, while the instrument's elevation angle is 2° and the azimuthal viewing direction 51° ; i.e. the instrument was pointing close to the rising sun. Enhanced stray light in the spectrometer (caused by the large contribution of photons at longer wavelengths while observing the red sky during sunrise) seems plausible and using $1/I$ is a better choice for compensation. It was verified (not shown) that the fit coefficient of the offset (also called the offset slant column) is particularly large not only in this but also in adjacent measurements and a colour index indicated that these spectra are indeed more reddish. However, these measurements could also be affected by direct sunlight in the instrument, which is known to increase rms (e.g. due to polarization issues). Figure 3f demonstrates that the respective measurement is affected by a very large rms (potentially caused by a combination of direct light and stray light).

TI2 and TI3 are more sophisticated approaches (Eq. 4) based on either I or I_0 . However, the resulting lines largely

follow TI0 and TI1 (simple approaches). Unexpectedly, both TI2 and TI3 lead to larger rms values for the sunrise measurement in the morning where the simple approach performs better.

The total effect of using an intensity offset compensation is evaluated by TI4 which includes no offset correction. The resulting effect on NO_2 is almost the same as TI1 and TI3 (based on I_0) leading to the clear recommendation of using an offset compensation based on I instead of I_0 , which is supported by largely increased rms values of TI4 (Fig. 9 bottom).

An intensity offset correction of 1st-order (i.e. a term varying linearly with λ in addition to a constant term) was tested in TI5, which is in practice often used not only for intensity offsets but also for compensation of the wavelength-dependence of the Ring slant column. The resulting NO_2 differences with respect to the reference fit (or TL0) are small with the exception of the first data point that is slightly off. In terms of rms, TI5 leads to improvements of $\approx 20\%$ in the

morning. Interestingly, the TI5 rms line shows some similarities to the IUPHD-IUPB line in Fig. 3d. However, no 1st-order offset was included in the IUPHD fit; i.e. the offset implementation is not causing the observed similar shape in the morning (and the reason remains unclear). This is supported by increasing IUPHD rms values in the evening in Fig. 3d, which are not present in TI5.

Fit TI6 includes again only a 0th-order intensity offset, but a pseudo cross section accounting for the wavelength-dependence of the Ring slant column was added (the Ring cross section was multiplied by λ and orthogonalized against the original cross section). The resulting rms is indeed almost identical to TI5 (using a 1st-order intensity offset) while the NO_2 is identical to TI0; i.e. the outlying first data point which is still partially present in TI5 disappeared in TI6. Thus, using a pseudo-cross section for compensation of the Ring wavelength-dependence seems preferable compared to using a 1st-order intensity offset correction.

4.4 I_0 -correction

In addition to the intensity offset tests, the effect of inclusion of an I_0 -correction was evaluated in TI7 (which is identical to TI0 except for the I_0 -correction; see Table 6). The respective line is shown in addition to the intensity offset investigations in Fig. 9. The I_0 -effect addresses the problem that the limited instrument's resolution can cause an incomplete removal of Fraunhofer structures in the vicinity of strong narrow-banded absorption bands (Johnston, 1996; Wagner et al., 2001; Aliwell et al., 2002).

Only a very small constant offset in relative NO_2 slant column differences is obtained in TI7 ($\approx 0.25\%$, which is almost invisible in Fig. 9, middle). In terms of rms, exclusion of the I_0 -correction leads to an increased rms of up to 20%, which is comparable to different treatments of the intensity offset correction. Thus, including an I_0 -correction is recommended in polluted environments such as the MAD-CAT site. It should be noted that the first data point is not an outlier in TI7; i.e. it is not sensitive to the I_0 -correction.

4.5 Numerical methods (linear DOAS inversion)

The DOAS equation (Eq. 1) is a linear inverse problem

$$\mathbf{Ax} = \mathbf{b}, \quad (5)$$

with the vector \mathbf{x} (size n) containing the n trace gas slant columns and polynomial coefficients of interest, the vector $\mathbf{b} = \ln(\frac{I_0}{I})$ (size m) containing the measured optical depths at m wavelengths and the $m \times n$ DOAS matrix \mathbf{A} with columns consisting of absorption cross sections and polynomial terms ($1, \lambda, \lambda^2$, etc.) for the m wavelengths.

Different numerical methods exist to solve Eq. (5) for \mathbf{x} . As \mathbf{A} is non-square, no inverse exists. However, most retrieval codes calculate a pseudo-inverse \mathbf{A}^{-1} (almost) fulfilling $\mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$ (identity matrix) using a singular value de-

composition (SVD) and obtain the slant columns of interest by $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$. This method is frequently recommended for solving overdetermined linear inverse problems in terms of least squares (see e.g. Press, 1989).

However, after multiplying Eq. (5) with \mathbf{A}^T , the matrix $\mathbf{A}^T\mathbf{A}$ is quadratic and can be decomposed into an upper and a lower triangular matrix, \mathbf{L} and \mathbf{U} . The linear inverse problem

$$\mathbf{A}^T\mathbf{Ax} = \mathbf{LUx} = \mathbf{A}^T\mathbf{b} \quad (6)$$

can then be solved by a forward substitution, obtaining \mathbf{y} from $\mathbf{L}\mathbf{y} = \mathbf{A}^T\mathbf{b}$, and a backward substitution, obtaining \mathbf{x} from $\mathbf{U}\mathbf{x} = \mathbf{y}$.

Furthermore, $\mathbf{A}^T\mathbf{A}$ can also be directly inverted, normally by LU decomposition as well. The vector of slant columns is then obtained by $\mathbf{x} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{b}$. As the inversion normally takes many more computational steps, this method is known to be subject to round-off errors and is therefore not recommended (Press, 1989).

The influence of these different numerical methods on resulting slant columns could not be easily tested with the IUPB retrieval code and was thus evaluated in a Python script solving the DOAS equation using the same input (spectra, cross sections). Python (which is a well-established programming language in scientific computing) provides numerous different routines within its numpy and scipy packages (based on different subroutines from the LAPACK package, <http://www.netlib.org/lapack/>) that were tested for solving the DOAS equation. All performed tests are summarized in Table 7. Again, differences of NO_2 slant columns and fit rms have been calculated with respect to the IUPB v1 fit results (i.e. using NLIN). In order to restrict differences to the influence of numerical approaches only, the same slit function treatment as the IUPB retrieval code NLIN was applied in the Python script and the same (noon) reference spectrum was used. In addition, no further wavelength calibration, i.e. no non-linear shift and squeeze fit, was performed (neither in the Python script nor in the NLIN reference fit used here) and no I_0 -correction was included. The test results are shown in Fig. 10, again for the 2° elevation angle.

Results of test TL0 appear to be identical to the reference fit of the IUPB retrieval code, both using a SVD for inversion of the DOAS matrix. However, very small differences exist between TL0 and the reference fits (too small to be seen on the scale of Fig. 10), which are $< 0.006\%$ for NO_2 slant columns and $< 0.07\%$ for rms. These tiny disagreements can be attributed to numerical differences in programming languages.

TL1 and TL2 yield identical NO_2 slant columns (both are using an LU decomposition) which differ from SVD results by up to 0.7%. However, the rms from TL2 was found to be an order of magnitude larger compared to the other tests and is therefore not shown in Fig. 10 (bottom). This is most likely due to problems mentioned above; i.e. this finding is in agreement with common recommendations in textbooks. Interestingly, both TL1 (and TL2) NO_2 and rms lines (which

Table 7. Different methods tested for solving the linear DOAS equation $\mathbf{Ax} = \mathbf{b}$.

Test	Retrieval code	Spectral grid	Method	Remarks
Reference	NLIN	I_0	Pseudo-inverse of \mathbf{A} using SVD	following Press (1989)
TL0	Python	I_0	Pseudo-inverse of \mathbf{A} using SVD	
TL1	Python	I_0	Solving quadratic $\mathbf{A}^T \mathbf{A} \mathbf{x} = \mathbf{A}^T \mathbf{b}$ using LU decomposition	
TL2	Python	I_0	Invert $\mathbf{A}^T \mathbf{A}$ using LU decomp. and multiply with $\mathbf{A}^T \mathbf{b}$	
TL3	Python	0.01 nm	same as TL0	linear interpolation to 0.01 nm
TL4	Python	0.01 nm	same as TL0	cubic spline interpolation to 0.01 nm

are differences with respect to IUPB) are similar in shape to the total fit rms (compare to Fig. 3f). Thus, when the rms increases, SVD inversion and LU decomposition lead to larger disagreements, both in rms and NO_2 . As the SVD yields smaller rms values, it seems to be preferable, although the obtained improvement is only about 2.5 %.

Numerical differences may be obtained when performing the linear DOAS fit (Eq. 5) on another wavelength grid. Changes of the grid potentially arise from the wavelength calibration. Some retrieval codes (e.g. NIWA) also use an internal, equidistant wavelength grid. To test the effect of changes in the wavelength grid, the TL0 fit was repeated on an equidistant 0.01 nm grid (i.e. I , I_0 and cross sections were interpolated to 0.01 nm before solving Eq. 5). TL3 and TL4 are identical to TL0, but a linear interpolation was applied in TL3, while TL4 uses a cubic spline interpolation to 0.01 nm. Apparently, this results in a constant offset in relative NO_2 differences, which is seen most clearly in the TL3 line in Fig. 10 (middle). The resulting constant shift is $\approx 0.4\%$ for TL3, but only $\approx 0.02\%$ for TL4 meaning that the type of interpolation to the equidistant grid is of importance and the spline interpolation (not surprisingly) seems to resemble the spectrum better than a linear interpolation. However, using different wavelength grids might, for example, explain some of the observed differences between IUPB and NIWA, which were found to be constant in relative differences as well. In terms of rms, the computation on an equidistant 0.01 nm grid using linear interpolation behaves even a bit better on average (up to 1 %). However, no recommendation can be drawn from this, as discussed above.

4.6 Non-linear wavelength calibration

As mentioned in Sect. 2.3, the spectra provided from the IUPB instrument were precalibrated using nightly HgCd line lamp measurements, which provide accuracies better than 0.1 nm. However, usually a postcalibration is included in DOAS retrieval codes in order to increase the fit quality (reducing rms). This wavelength calibration is imple-

mented in different ways in participating retrieval codes. Most groups calibrate the reference spectrum I_0 to a high-resolution Fraunhofer atlas, apply the resulting calibration to all measured spectra and allow in addition a shift and squeeze between I and I_0 in order to compensate spectral shifts of the spectrometer during the day, e.g. caused by temperature changes (KNMI uses a slightly different definition of the shift as discussed below). This non-linear shift and squeeze fit of the wavelength axis is mostly implemented in an iterative scheme together with the linear DOAS fit on $\ln(I_0/I)$. However, codes differ, for example in whether trace gases are included in the shift and squeeze fit of I_0 to the high-resolution Fraunhofer atlas. Sometimes also a higher-order calibration is allowed or several subwindows are used in order to differently characterize different parts of the spectra.

Table 8 summarizes tests performed to investigate the impact of different wavelength calibration approaches using the IUPB retrieval code NLIN. In addition, some tests were performed with the Python script from Sect. 4.5, which has therefore been extended to perform the non-linear shift fit as not all tests could be easily implemented in the comprehensive NLIN software. Note that, in contrast to the shift, the squeeze has been excluded from the intercomparison as it was found to always be 1.0 for measurements shown here. Also, QDOAS-specific implementations were not tested here.

Figure 11 shows the resulting impact on NO_2 and rms as well as the fitted shift between I and I_0 (not present in all tests). NO_2 and rms are again differences relative to the IUPB v1 fit results (without I_0 -correction as this was not implemented in the Python routine). As in Fig. 3, the shift in Fig. 11d is no difference, and for comparison the shift from the reference fit is shown explicitly in black. The (fixed) shift retrieved from the non-linear fit of I_0 to the Fraunhofer atlas is -0.035 nm. This is roughly a factor of 10 larger than the fitted shift between I and I_0 shown in Fig. 11d. It is interesting to note that the shift is not zero around noon (time of the reference spectrum), indicating correlations between shift fit and other effects (predominantly intensity offset correction)

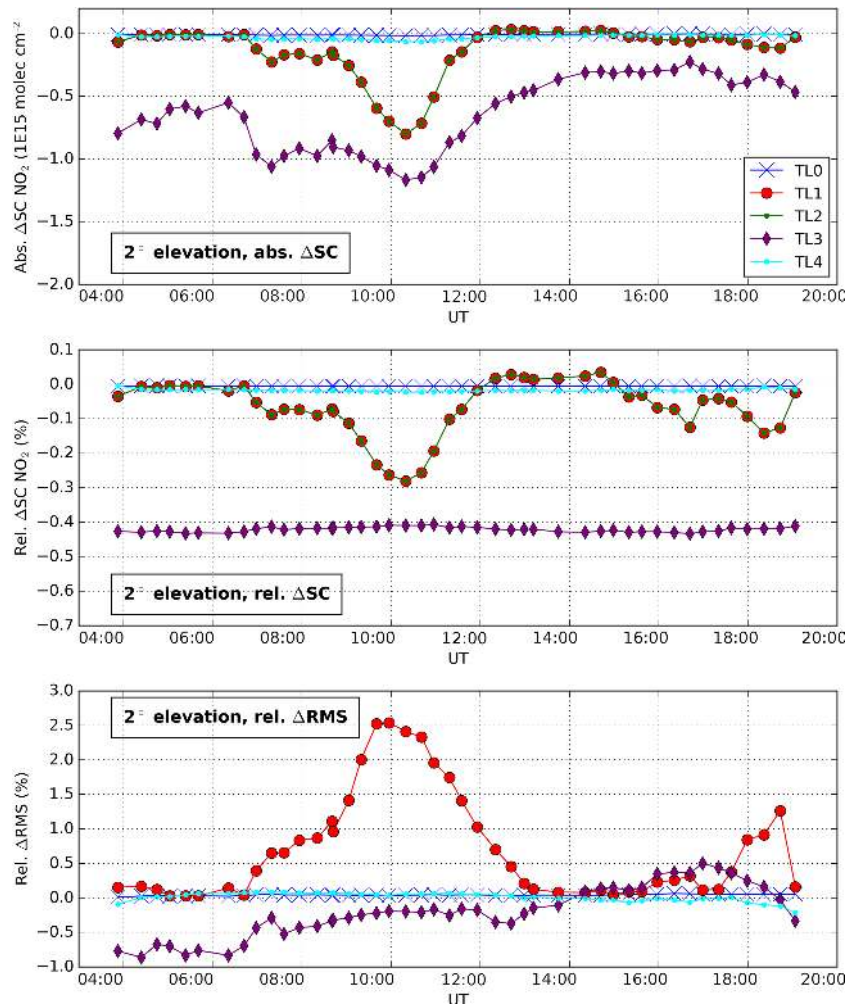


Figure 10. Absolute (top) and relative (middle) NO₂ slant column differences and relative rms differences (bottom) with respect to IUPB v1 results (only linear fit) in 2° elevation angle resulting from different numerical methods solving the linear DOAS equation.

Table 8. Different wavelength calibration approaches evaluated.

Test	Retrieval code	Shift I_0 to Fraunhofer atlas	Shift I to I_0	Remarks
Reference	NLIN_D	yes	yes	alternating scheme, v1 fit settings
TW0	NLIN_D	no	no	linear DOAS fit only
TW1	NLIN_D	no	yes	
TW2	NLIN_D	yes	no	
TW3	Python	yes	yes	same a TW0
TW4	Python	yes	yes	same a TW3, but linear interpolation
TW5	Python	yes	yes	same a TW3, all trace gases included in Fraunhofer I_0 fit

and also indicating the presence of the tilt effect (Lampel et al., 2017).

The most extreme test is TW0, which excludes both the calibration of I_0 to the Fraunhofer atlas as well as the shift between I and I_0 . When omitting both calibration steps, the

rms is largely enhanced by up to 80 %, peaking in the morning and decreasing towards noon with a second smaller maximum around 12:00 UT. Interestingly, a very similar shape is seen in the rms differences of KNMI, NUST, UNAM, USTC and INTA (but not in INTA2 where a fault in the wavelength

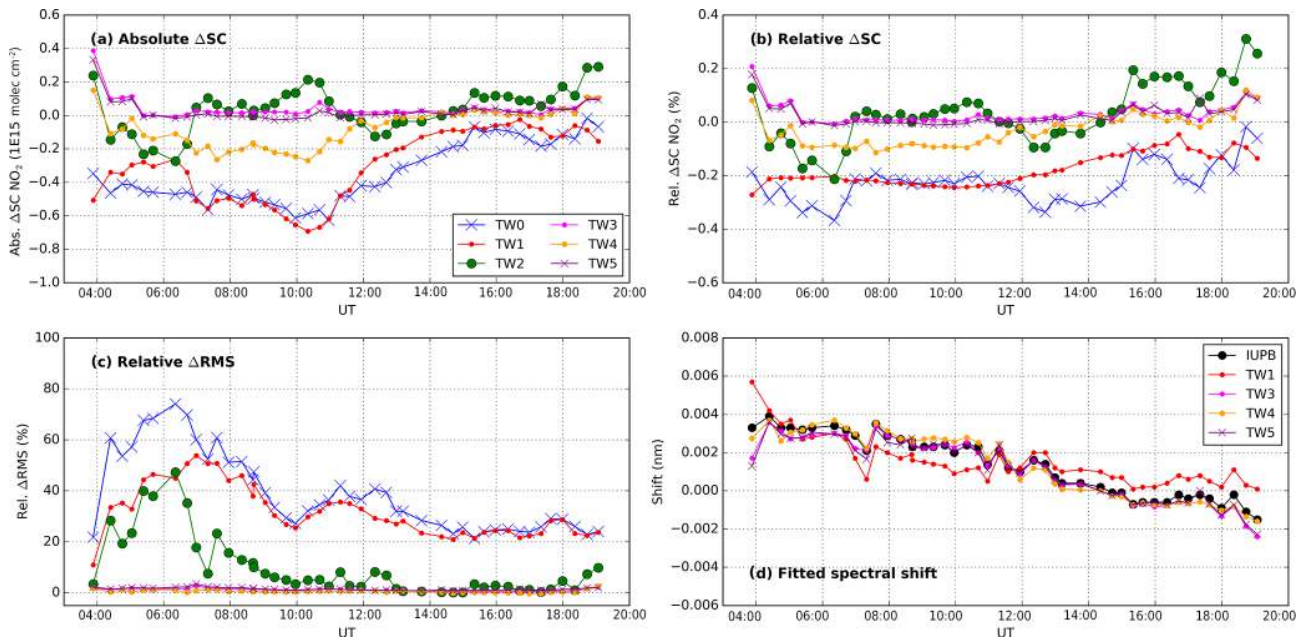


Figure 11. Absolute (a) and relative (b) NO_2 slant columns differences and relative rms differences (c) in 2° elevation angle resulting from different wavelength calibration approaches (Table 8). Corresponding shifts between I and I_0 resulting from non-linear fits are shown in (d).

registration module was corrected) in Fig. 3d. Although all of these groups are performing a wavelength calibration, TW0 indicates that differences in the calibration procedure are causing most of the disagreements between groups in terms of rms. This is in contrast to NO_2 where changes of only 0.4 % are obtained from TW0.

TW1 still excludes the absolute calibration to the Fraunhofer atlas, but includes the shift fit between I and I_0 . As seen before, the impact on NO_2 is very small ($\approx 0.4\%$), but absolute NO_2 differences of TW1 reflect the shape of total rms and NO_2 slant columns (compare to Fig. 3); i.e. the relative differences are smooth in shape. The rms is similar in shape to TW0, but the morning maximum is slightly later at 07:00 UT; the noon maximum is around 11:00 UT and a small maximum in the evening occurs at 18:00 UT. This shape is similar to the relative rms of NUST, UNAM and NIWA in Fig. 3, but the absolute numbers are different. However, this behaviour indicates that differences in the fitted Fraunhofer shift are partially responsible for observed differences between these and other groups. Interestingly, the fitted shift between I and I_0 of TW1 in Fig. 11d is very similar to corresponding values of the reference fit, which is because the missing Fraunhofer shift is a different effect than the shift between I and I_0 .

In contrast to TW1, TW2 includes the Fraunhofer shift fit, but excludes the shift between I and I_0 . As this is the larger effect (-0.035 nm compared to only ≈ 0.004 nm), the rms is much smaller than in TW1 with a single maximum (up to 50 %) in the early morning at 06:00 UT. The rms time series shape is similar to the KNMI line in Fig. 3. The reason

is a different definition of the shift in the KNMI retrieval: while in most retrievals I is shifted relative to I_0 , KNMI calculates the optical depth $\tau = \ln(I_0/I)$ without any shifts but then allows a shift of all cross sections relative to τ . This is in first order compensating the effect of the Fraunhofer shift but neglecting potential shifts between I and I_0 (in this case Fraunhofer lines would not cancel out completely in the optical depth τ). As a result, the KNMI approach is similar (but not identical) to TW2. The fit quality following the KNMI approach is expected to be better using a sequential reference as the temperature drift of the spectrometer is much smaller then. It has to be mentioned that the change of NO_2 in TW2 is small (0.3 %) compared to the change in fit rms (50 %).

TW3 uses the same wavelength calibration treatment as the IUPB reference fit, but was run in another programming code (Python) to evaluate how much difference is caused by use of another programming code and numerical issues. The fitted shift is mostly identical to the reference fit except for the early morning and late evening when the NO_2 also shows very slight differences. The largest disagreement of NO_2 is 0.2 % for the first measurement of the day that was affected by large stray light effects (and potentially direct light) and thus most likely indicating cross-correlations between shift fit and intensity offset correction. The resulting rms is almost identical to the reference fit (Fig. 11c).

TW4 is the same as TW3 but the spline interpolation (calculating I at spectral points of I_0 during the non-linear shift fit) is replaced by a simple linear interpolation. The fitted shift is changed slightly and NO_2 differences are of almost the same order than when not performing any shift fit at all,

Table 9. Summary of performed sensitivity tests and associated impacts on NO₂ slant columns and rms.

Reason for disagreement	ΔNO_2 (%)	Δrms (%)	Remarks
Reference treatment (noon)	2.5	3	Produces constant absolute NO ₂ SC offsets.
Reference treatment (seq.)	8	6	
Slit function treatment	1.3	6.5	Produces constant relative NO ₂ SC offsets.
Intensity offset correction	2 (typically) 10 (outlier)	20 (typically) 60 (outlier)	
I_0 correction	0.25	20	Produces constant relative NO ₂ SC offsets.
Numerical methods (for linear DOAS fit)	0.3 (0.7)	2.5	Produces constant relative NO ₂ SC offsets. Disagreements increase with rms.
Wavelength calibration (non-linear shift fit)	0.4	up to 80	

demonstrating that the shift fit has a negligible impact on NO₂. In contrast, it has a large impact on rms, where the marginally different methods between TW3 and TW4 produce the same rms while the effect of excluding the shift completely leads to largely enhanced rms.

In the Fraunhofer calibration of the reference fit using NLIN (non-linear shift fit of I_0) as well as in TW3 all trace gas absorptions are omitted; i.e. an iterative scheme between shift fit and DOAS fit comprising only a polynomial of order 4 is applied. In contrast, all trace gas absorptions are included in the Fraunhofer calibration in TW5. As a result, the rms of the DOAS fit between the Fraunhofer spectrum and I_0 is reduced by a factor of 2 and the non-linearly fitted shift is -0.031 nm instead of -0.035 nm in the reference fit. However, this only has a marginal influence on NO₂, rms and fitted shift between I and I_0 in Fig. 11 and consequently explains none of the observed differences between groups.

Notably, all shifts (even TW1) in Fig. 11d show the same general shapes that are retrieved by most groups in Fig. 3. Only shifts of KNMI clearly differ, for the reasons mentioned above. Rms shapes in Fig. 11c suggest that differences in the wavelength calibration are the major reason for observed rms disagreements between groups in Fig. 3.

5 Summary and conclusions

An intercomparison of DOAS retrieval codes using measured spectra from the same instrument during the MAD-CAT campaign and harmonized fit settings was performed. Excellent agreement was found between different DOAS fit algorithms from 17 international groups. In some of the retrieval codes, faults were identified and corrected in the course of the study, leading to even better agreements. For noon reference fits, the correlation in terms of NO₂ slant columns was found to be larger ($> 99.98\%$) than for sequential references ($> 99.2\%$), which is caused by different implementations of the sequential reference.

Despite the excellent overall correlation, for individual measurements in low elevations differences of up to 8% in

resulting NO₂ slant columns were observed, which is up to 2–3 times larger than corresponding typical NO₂ slant column errors. In terms of fit rms, large differences of up to $\approx 100\%$ were found.

Interestingly, groups using the same retrieval code (QDOAS groups) do not always produce results showing the same systematic behaviour, which is the result of different options users select – other than the harmonization settings agreed on. This has some impact for the interpretation of other studies as the extent of prescribed fit settings in this study is comparable to intercomparison campaigns like CINDI or CINDI-2. Consequently, groups participating in those campaigns will provide intrinsic differences in their results due to non-harmonized (detailed) settings even if using the same retrieval code, which has to be expected in the same range as observed here.

Comprehensive sensitivity studies systematically investigating effects of the non-harmonized retrieval aspects were performed in order to (1) attribute observed differences between groups to certain sources and (2) evaluate the impact of each of these sources on NO₂ slant columns and fit rms. For this purpose, five causes were identified in a survey of participating retrieval codes. Typical impacts on NO₂ slant columns and rms are summarized in Table 9.

In general, the wavelength calibration and the intensity offset correction were found to produce the majority of observed rms differences, but have a negligible impact on NO₂ slant columns ($< 0.4\%$ or $< 2\%$ except for the first measurement of the day affected by stray light and possibly direct light in the telescope). In contrast, the reference selection explains the majority of observed NO₂ slant column differences between groups while having a minor impact on the rms. Thus, if harmonization of NO₂ slant columns is of interest, the reference treatment needs to be harmonized (otherwise in small elevation angles differences of up to 8% – and even more for larger elevation angles due to decreasing absolute slant columns – have to be expected predominantly in the case of using a sequential reference) while for rms reduction/harmonization, the offset intensity correction and the wavelength calibration need to be harmonized.

In terms of NO₂, two types of disagreements have been observed, which are (1) constant in absolute differences or (2) constant in relative differences. The latter was found to arise from the numerical approach used for solving the DOAS equation as well as the treatment of the slit function while the choice of the reference spectrum causes absolute differences.

The best practices aiming at improvement of the fit quality and harmonization between MAX-DOAS retrievals derived from this study are as follows.

1. Reference treatment: using averaged or interpolated sequential reference spectra matches the atmospheric conditions at the measurement time better and was found to produce slightly smaller rms.
2. Slit function: using a measured slit function gave a better performance than fitting line parameters in the data set used here. However, slit function measurements then have to be taken regularly (e.g. daily to monitor possible instrument changes).
3. Intensity offset: an approach based on I instead of I_0 is recommended. Surprisingly, the simple approach performs better for measurements pointing close to sunrise but this could be just a coincidence in this data set. The inclusion of an additional Ring spectrum multiplied by wavelength is preferred over adding a linear term to the offset as this mostly compensates the wavelength-dependence of the Ring slant column.
4. Numerical approaches: using an SVD provides more stability and produces slightly smaller rms than LU decomposition.
5. Wavelength calibration: although HgCd line lamp calibration measurements lead to absolute accuracies (in this case) of ≈ 0.03 nm, a Fraunhofer shift fit reduces the rms by up to 40–50%. For the I_0 fit with respect to the Fraunhofer spectrum, the inclusion of all trace gases showed no advantage over the inclusion of only a polynomial. Temperature instabilities of the spectrometer produced shifts of I relative to I_0 changing over the day. Compensation of this effect within DOAS retrieval codes further improves the rms by up to 40–50% when using a noon reference spectrum.

6 Data availability

The data (spectra and common cross sections) used in this study are available from the cited references and directly from the authors upon request.

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

Acknowledgements. We thank the Max-Planck Institute for Chemistry, Mainz, for hosting and organizing the MAD-CAT campaign. This study has been performed in the context and with financial support of the EU-QA4ECV project. CU Boulder furthermore acknowledges B. K. Dix for support. The work of the IAP group was supported by the Russian Science Foundation by grant 14-47-00049. JAMSTEC was supported by Coordination Funds for Promoting AeroSpace Utilization, MEXT, Japan.

The article processing charges for this open-access publication were covered by the University of Bremen.

Edited by: H. Maring

Reviewed by: two anonymous referees

References

- Aliwell, S. R., Van Roozendaal, M., Johnston, P. V., Richter, A., Wagner, T., Arlander, D. W., Burrows, J. P., Fish, D. J., Jones, R. L., Tornkvist, K. K., Lambert, J. C., Pfeilsticker, K., and Pundt, I.: Analysis for BrO in zenith-sky spectra: An intercomparison exercise for analysis improvement, *J. Geophys. Res.-Atmos.*, 107, doi:10.1029/2001JD000329, 2002.
- Beirle, S., Boersma, K. F., Platt, U., Lawrence, M. G., and Wagner, T.: Megacity Emissions and Lifetimes of Nitrogen Oxides Probed from Space, *Science*, 333, 1737–1739, doi:10.1126/science.1207824, 2011.
- Bogumil, K., Orphal, J., Homann, T., Voigt, S., Spietz, P., Fleischmann, O. C., Vogel, A., Hartmann, M., Kromminga, H., Bovensmann, H., Frerick, J., and Burrows, J. P.: Measurements of molecular absorption spectra with the SCIAMACHY pre-flight model: instrument characterization and reference data for atmospheric remote-sensing in the 230–2380 nm region, *J. Photochem. Photobiol. A*, 157, 167–184, 2003.
- Borovski, A. N., Dzhola, A. V., Elokho, A. S., Grechko, E. I., Kanaya, Y., and Postlyakov, O. V.: First measurements of formaldehyde integral content in the atmosphere using MAX-DOAS in the Moscow region, *Int. J. Remote Sens.*, 35, 5609–5627, 2014.
- Brewer, A. W.: A Replacement for the Dobson Spectrophotometer?, *Pure Appl. Geophys.*, 106–108, 919–927, 1973.
- Chance, K. V. and Spurr, R. J. D.: Ring effect studies: Rayleigh scattering, including molecular parameters for rotational Raman scattering, and the Fraunhofer spectrum, *Appl. Opt.*, 36, 5224–5230, 1997.
- Dankaert, T., Fayt, C., Van Roozendaal, M., De Smedt, I., Letocart, V., Merlaud, A., and Pinardi, G.: QDOAS Software user manual, Version 2.106, available at: <http://uv-vis.aeronomie.be/software/QDOAS> (last access: 7 March 2017), 2013.
- Fayt, C. and Van Roozendaal, M.: WinDOAS 2.1 Software User Manual, available at: <http://uv-vis.aeronomie.be/software/WinDOAS/> (last access: 7 March 2017), 2001.
- Frieß, U., Sihler, H., Sander, R., Pöhler, D., Yilmaz, S., and Platt, U.: The vertical distribution of BrO and aerosols in the Arctic: Measurements by active and passive differential optical absorption spectroscopy, *J. Geophys. Res.*, 116, D00R04, doi:10.1029/2011JD015938, 2011.

- Grainger, J. F. and Ring, J.: Anomalous Fraunhofer Line Profiles, *Nature*, 193, 762, doi:10.1038/193762a0, 1962.
- Hofmann, D. J., Bonasoni, P., De Maziere, M., Evangelisti, F., Giovanelli, G., Goldman, A., Goutail, F., Harder, J., Jakoubek, R., Johnston, P. V., Kerr, J. E., Matthews, W. A., McElroy, C. T., McKenzie, R. L., Mount, G., Platt, U., Pommereau, J.-P., Sarkissian, A., Simon, P., Solomon, S., Stutz, J., Thomas, A., Van Roozendael, M., and Wu, E.: Intercomparison of UV/visible spectrometers for measurements of stratospheric NO₂ for the Network for the Detection of Stratospheric Change, *J. Geophys. Res.-Atmos.*, 100, 16765–16791, doi:10.1029/95JD00620, 1995.
- Hönninger, G., von Friedeburg, C., and Platt, U.: Multi axis differential optical absorption spectroscopy (MAX-DOAS), *Atmos. Chem. Phys.*, 4, 231–254, doi:10.5194/acp-4-231-2004, 2004.
- Ivanov, V. A., Elokhov, A. S., and Postlyakov, O. V.: On the possibility of estimating the volume of NO₂ emissions in cities using zenith spectral observations of diffuse solar radiation near 450 nm, *Atmospheric and Oceanic Optics*, 25, 434–439, doi:10.1134/S1024856012060061, 2012.
- Johnston, P. V.: Studies on the IO-effect, unpublished manuscript, 1996.
- Kraus, S.: DOASIS – A Framework Design for DOAS, Dissertation, University of Heidelberg, 2006.
- Lampel, J., Frieß, U., and Platt, U.: The impact of vibrational Raman scattering of air on DOAS measurements of atmospheric trace gases, *Atmos. Meas. Tech.*, 8, 3767–3787, doi:10.5194/amt-8-3767-2015, 2015.
- Lampel, J., Wang, Y., Beirle, S., Sihler, H., Pukite, J., Hilboll, A., Wagner, T., and Platt, U.: The tilt-effect in passive DOAS observations, *Atmos. Meas. Tech.*, in preparation, 2017.
- Lee, D. S., Köhler, I., Grobler, E., Rohrer, F., Sausen, R., Gallardo-Klenner, L., Olivier, J. G. J., Dentener, F. J., and Bouwman, A. F.: Estimations of Global NO_x Emissions and their Uncertainties, *Atmos. Environ.*, 31, 1735–1749, 1997.
- Noxon, J. F.: Nitrogen-Dioxide in Stratosphere and Troposphere Measured by Ground-Based Absorption Spectroscopy, *Science*, 189, 547–549, 1975.
- 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, doi:10.5194/amt-8-2371-2015, 2015.
- Ortega, I., Lampel, J., Wagner, T., Van Roozendael, M., Richter, A., Sinha, V., Xie, P., Volkamer, R., and the Glyoxal MAD-CAT Team: Inter-comparison of glyoxal retrievals from MAX-DOAS during the MAD-CAT campaign, *Atmos. Meas. Tech.*, in preparation, 2017.
- Peters, E., Wittrock, F., Großmann, K., Frieß, U., Richter, A., and Burrows, J. P.: Formaldehyde and nitrogen dioxide over the remote western Pacific Ocean: SCIAMACHY and GOME-2 validation using ship-based MAX-DOAS observations, *Atmos. Chem. Phys.*, 12, 11179–11197, doi:10.5194/acp-12-11179-2012, 2012.
- Peters, E., Wittrock, F., Richter, A., Alvarado, L. M. A., Rozanov, V. V., and Burrows, J. P.: Liquid water absorption and scattering effects in DOAS retrievals over oceans, *Atmos. Meas. Tech.*, 7, 4203–4221, doi:10.5194/amt-7-4203-2014, 2014.
- Pinardi, G.: Harmonization of NO₂ and HCHO MAXDOAS retrievals as part of the QA4ECV project, *Atmos. Meas. Tech.*, in preparation, 2017.
- Pinardi, G., Van Roozendael, M., Abuhassan, N., Adams, C., Cede, A., Clémer, K., Fayt, C., Frieß, U., Gil, M., Herman, J., Hermans, C., Hendrick, F., Irie, H., Merlaud, A., Navarro Comas, M., Peters, E., Piter, A. J. M., Puentedura, O., Richter, A., Schönhardt, A., Shaiganfar, R., Spinei, E., Strong, K., Takashima, H., Vrekoussis, M., Wagner, T., Wittrock, F., and Yilmaz, S.: MAX-DOAS formaldehyde slant column measurements during CINDI: intercomparison and analysis improvement, *Atmos. Meas. Tech.*, 6, 167–185, doi:10.5194/amt-6-167-2013, 2013.
- Piter, A. J. M., Boersma, K. F., Kroon, M., Hains, J. C., Van Roozendael, M., Wittrock, F., Abuhassan, N., Adams, C., Akrami, M., Allaart, M. A. F., Apituley, A., Beirle, S., Bergwerff, J. B., Berkhout, A. J. C., Brunner, D., Cede, A., Chong, J., Clémer, K., Fayt, C., Frieß, U., Gast, L. F. L., Gil-Ojeda, M., Goutail, F., Graves, R., Griesfeller, A., Großmann, K., Hemerijckx, G., Hendrick, F., Henzing, B., Herman, J., Hermans, C., Hoexum, M., van der Hoff, G. R., Irie, H., Johnston, P. V., Kanaya, Y., Kim, Y. J., Klein Baltink, H., Kreher, K., de Leeuw, G., Leigh, R., Merlaud, A., Moerman, M. M., Monks, P. S., Mount, G. H., Navarro-Comas, M., Oetjen, H., Pazmino, A., Perez-Camacho, M., Peters, E., du Piesanie, A., Pinardi, G., Puentedura, O., Richter, A., Roscoe, H. K., Schönhardt, A., Schwarzenbach, B., Shaiganfar, R., Sluis, W., Spinei, E., Stolk, A. P., Strong, K., Swart, D. P. J., Takashima, H., Vlemmix, T., Vrekoussis, M., Wagner, T., Whyte, C., Wilson, K. M., Yela, M., Yilmaz, S., Zieger, P., and Zhou, Y.: The Cabauw Intercomparison campaign for Nitrogen Dioxide measuring Instruments (CINDI): design, execution, and early results, *Atmos. Meas. Tech.*, 5, 457–485, doi:10.5194/amt-5-457-2012, 2012.
- Platt, U.: Differential optical absorption spectroscopy (DOAS), *Chem. Anal. Series*, 127, 27–83, 1994.
- Platt, U. and Stutz, J.: *Differential Optical Absorption Spectroscopy: Principles and Applications*, Springer Verlag, 2008.
- Postlyakov, O. and Borovski, A.: Measurement of formaldehyde total content in troposphere using DOAS technique: improvements in version 1.3a of IAP retrieval algorithm, in: *SPIE 9876*, doi:10.1117/12.2229231, 2016.
- Postlyakov, O., Borovski, A., Dzhola, A., Elokhov, A. S., Grechko, E. I., and Kanaya, Y.: Measurements of formaldehyde total content in troposphere using DOAS technique in Moscow Region: preliminary results of 3 year observations, in: *SPIE 9242*, doi:10.1117/12.2069824, 2014.
- Press, W. H.: *Numerical Recipes in Pascal (First Edition): The Art of Scientific Computing*, Bd. 1, Cambridge University Press, 1989.
- Pukite, J. and Wagner, T.: Quantification and parametrization of non-linearity effects by higher-order sensitivity terms in scattered light differential optical absorption spectroscopy, *Atmos. Meas. Tech.*, 9, 2147–2177, doi:10.5194/amt-9-2147-2016, 2016.
- Pukite, J., Kühl, S., Deutschmann, T., Platt, U., and Wagner, T.: Extending differential optical absorption spectroscopy for limb measurements in the UV, *Atmos. Meas. Tech.*, 3, 631–653, doi:10.5194/amt-3-631-2010, 2010.
- Richter, A.: *Absorptionsspektroskopische Messungen stratosphaerischer Spurengase ueber Bremen*, 53 N, PhD thesis, University of Bremen, Bremen, 1997.

- Richter, A., Burrows, J. P., Nüß, H., Granier, C., and Niemeier, U.: Increase in tropospheric nitrogen dioxide over China observed from space, *Nature*, 437, 129–132, 2005.
- Roscoe, H. K., Johnston, P. V., Van Roozendael, M., Richter, A., Sarkissian, A., Roscoe, J., Preston, K. E., Lambert, J. C., Hermans, C., Decuyper, W., Dzienus, S., Winterrath, T., Burrows, J., Goutail, F., Pommereau, J. P., D'Almeida, E., Hottier, J., Coureul, C., Didier, R., Pundt, I., Bartlett, L. M., McElroy, C. T., Kerr, J. E., Elokhov, A., Giovanelli, G., Ravegnani, F., Premuda, M., Kostadinov, I., Erle, F., Wagner, T., Pfeilsticker, K., Kenntner, M., Marquard, L. C., Gil, M., Puentedura, O., Yela, M., Arlander, D. W., Kastad Hoiskar, B. A., Tellefsen, C. W., Karlsen Tørnkvist, K., Heese, B., Jones, R. L., Aliwell, S. R., and Freshwater, R. A.: Slant column measurements of O₃ and NO₂ during the NDSC intercomparison of zenith-sky UV-visible spectrometers in June 1996, *J. Atmos. Chem.*, 32, 281–314, doi:10.1023/A:1006111216966, 1999.
- Roscoe, H. K., Van Roozendael, M., Fayt, C., du Piesanie, A., Abuhassan, N., Adams, C., Akrami, M., Cede, A., Chong, J., Clémer, K., Friess, U., Gil Ojeda, M., Goutail, F., Graves, R., Griesfeller, A., Grossmann, K., Hemerijckx, G., Hendrick, F., Herman, J., Hermans, C., Irie, H., Johnston, P. V., Kanaya, Y., Kreher, K., Leigh, R., Merlaud, A., Mount, G. H., Navarro, M., Oetjen, H., Pazmino, A., Perez-Camacho, M., Peters, E., Pinardi, G., Puentedura, O., Richter, A., Schönhardt, A., Shaiganfar, R., Spinei, E., Strong, K., Takashima, H., Vlemmix, T., Vrekoussis, M., Wagner, T., Wittrock, F., Yela, M., Yilmaz, S., Boersma, F., Hains, J., Kroon, M., PETERS, A., and Kim, Y. J.: Intercomparison of slant column measurements of NO₂ and O₄ by MAX-DOAS and zenith-sky UV and visible spectrometers, *Atmos. Meas. Tech.*, 3, 1629–1646, doi:10.5194/amt-3-1629-2010, 2010.
- Rothman, L. S., Gordon, I. E., Barber, R. J., Dothe, H., Gamache, R. R., Goldman, A., Perevalov, V. I., Tashkun, S. A., and Tennyson, J.: HITRAN, the high-temperature molecular spectroscopic database, *J. Quant. Spectrosc. Ra.*, 111, 2139–2150, doi:10.1016/j.jqsrt.2010.05.001, 2010.
- Schönhardt, A., Altube, P., Gerilowski, K., Krautwurst, S., Hartmann, J., Meier, A. C., Richter, A., and Burrows, J. P.: A wide field-of-view imaging DOAS instrument for two-dimensional trace gas mapping from aircraft, *Atmos. Meas. Tech.*, 8, 5113–5131, doi:10.5194/amt-8-5113-2015, 2015.
- Shaiganfar, R., Beirle, S., Sharma, M., Chauhan, A., Singh, R. P., and Wagner, T.: Estimation of NO_x emissions from Delhi using Car MAX-DOAS observations and comparison with OMI satellite data, *Atmos. Chem. Phys.*, 11, 10871–10887, doi:10.5194/acp-11-10871-2011, 2011.
- Shefov, N.: Intensivnosti nokotorykh emissiy sumerochnogo i nochnogo neba (Intensities of some Emissions of the Twilight and Night Sky), *Spectral, electrophotometrical and radar researches of aurorae and airglow, IGY program, section IV*, 1, 25 pp., 1959.
- Sioris, C. E., Haley, C. S., McLinden, C. A., von Savigny, C., McDade, I. C., McConnell, J. C., Evans, W. F. J., Lloyd, N. D., Llewellyn, E. J., Chance, K. V., Kurosu, T. P., Murtagh, D., Frisk, U., Pfeilsticker, K., Bösch, H., Weidner, F., Strong, K., Stegman, J., and Megie, G.: Stratospheric profiles of nitrogen dioxide observed by Optical Spectrograph and Infrared Imager System on the Odin satellite, *J. Geophys. Res.-Atmos.*, 108, 4215, doi:10.1029/2002JD002672, 2003.
- Takashima, H., Irie, H., Kanaya, Y., and Syamsudin, F.: NO₂ observations over the western Pacific and Indian Ocean by MAX-DOAS on *Kaiyo*, a Japanese research vessel, *Atmos. Meas. Tech.*, 5, 2351–2360, doi:10.5194/amt-5-2351-2012, 2012.
- Vandaele, A. C., Hermans, C., Simon, P. C., Van Roozendael, M., Guilmot, J. M., Carleer, M., and Colin, R.: Fourier Transform Measurement of NO₂ Absorption Cross-Section in the Visible Range at Room Temperature, *J. Atmos. Chem.*, 25, 289–305, 1996.
- Vandaele, A. C., Fayt, C., Hendrick, F., Hermans, C., Humbled, F., Van Roozendael, M., Gil, M., Navarro, M., Puentedura, O., Yela, M., Braathen, G., Stebel, K., Tørnkvist, 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, doi:10.1029/2004jd005423, 2005.
- Vaughan, G., Roscoe, H. K., Bartlett, L. M., O'Connor, F. M., Sarkissian, A., Van Roozendael, M., Lambert, J.-C., Simon, P. C., Karlsen, K., Kastad, B. A., Oiskar, H., Fish, D. J., Jones, R. L., Freshwater, R. A., Pommereau, J.-P., Goutail, F., Andersen, S. B., Drew, D. G., Hughes, P. A., Moore, D., Mellqvist, J., Hegels, E., Klupfel, T., Erle, F., Pfeilsticker, K., and Platt, U.: An intercomparison of ground-based UV-visible sensors of ozone and NO₂, *J. Geophys. Res.*, 102, 1411–1422, doi:10.1029/96JD00515, 1997.
- Vountas, M., Rozanov, V. V., and Burrows, J. P.: Ring effect: Impact of Rotational Raman scattering on radiative transfer in earth's atmosphere, *J. Quant. Spectrosc. Ra.*, 60, 943–961, 1998.
- Wagner, T., Chance, K., Frieß, U., Gil, M., Goutail, F., Hönninger, G., Johnston, P. V., Karlsen-Tørnkvist, K., Kostadinov, I., Leser, H., Petritoli, A., Richter, A., Van Roozendael, M., and Platt, U.: Correction of the Ring effect and I0-effect for DOAS observations of scattered sunlight, in: *Proc. of the 1st DOAS Workshop, Heidelberg, Germany*, 2001.
- Wagner, T., Beirle, S., Brauers, T., Deutschmann, T., Frieß, U., Hak, C., Halla, J. D., Heue, K. P., Junkermann, W., Li, X., Platt, U., and Pundt-Gruber, I.: 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, 268–2715, doi:10.5194/amt-4-2685-2011, 2011.
- Wang, Y., Beirle, S., Hendrick, F., Hilboll, A., Jin, J., Kyuberis, A. A., Lampel, J., Li, A., Luo, Y., Lodi, L., Ma, J., Navarro, M., Ortega, I., Peters, E., Polyansky, O. L., Remmers, J., Richter, A., Rodriguez, O. P., Roozendael, M. V., Seyler, A., Tennyson, J., Volkamer, R., Xie, P., Zobov, N. F., and Wagner, T.: MAX-DOAS measurements of HONO slant column densities during the MAD-CAT Campaign: inter-comparison and sensitivity studies on spectral analysis settings, *Atmos. Meas. Tech. Discuss.*, doi:10.5194/amt-2016-387, in review, 2017.
- Wittrock, F.: The retrieval of oxygenated volatile organic compounds by remote sensing techniques, PhD thesis, Bremen, 2006.