Intercomparison of NO₂, O₄, O₃ and HCHO slant column measurements by MAX-DOAS and zenith-sky UV-Visible spectrometers during the CINDI-2 campaign

Karin Kreher[1], Michel Van Roozendael[2], Francois Hendrick[2], Arnoud Apituley[3], Ermioni Dimitropoulou[2], Udo Frieß[4], Andreas Richter[5], Thomas Wagner[6], Johannes Lampel[4,7], Nader Abuhassan[8], Li Ang[9], Monica Anguas[10], Alkis Bais[11], Nuria Benavent[10], Tim Bösch[5], Kristof Bognar[12], Alexander Borovski[13], Ilya Bruchkouski[14], Alexander Cede[8,15], Ka Lok Chan[16,26], Sebastian Donner[6], Theano Drosoglou[11], Caroline Fayt[2], Henning Finkenzeller[17], David Garcia-Nieto[10], Clio Gielen[2], Laura Gómez-Martín[18], Nan Hao[19], Bas Henzing[20], Jay R. Herman[8], Christian Hermans[2], Syedul Hoque[21], Hitoshi Irie[21], Junli Jin[22], Paul Johnston[23], Junaid Khayyam Butt[24], Fahim Khokhar[24], Theodore K. Koenig[17], Jonas Kuhn[4,6], Vinod Kumar[6,25], Cheng Liu[26], Jianzhong Ma[22], Alexis Merlaud[2], Abhishek K. Mishra[25], Moritz Müller[15,27], Monica Navarro-Comas[18], Mareike Ostendorf[5], Andrea Pazmino[28], Enno Peters[5,29], Gaia Pinardi[2], Manuel Pinharanda[28], Ankie Piters[3], Ulrich Platt[4], Oleg Postylyakov[13], Cristina Prados-Roman[18], Olga Puentedura[18], Richard Querel[23], Alfonso Saiz-Lopez[10], Anja Schönhardt[5], Stefan F. Schreier[30], André Seyler[5], Vinayak Sinha[25], Elena Spinei[8,31], Kimberly Strong[12], Frederik Tack[2], Xin Tian[9], Martin Tiefengraber[15,27], Jan-Lukas Tirpitz[4], Jeroen van Gent[2], Rainer Volkamer[17], Mihalis Vrekoussis[5,32,33], Shanshan Wang[10,34], Zhuoru Wang[35], Mark Wenig[16], Folkard Wittrock[5], Pinhua H. Xie[9], Jin Xu[9], Margarita Yela[18], Chengxin Zhang[26], Xiaoyi Zhao[12,36]

[1] BK Scientific, Mainz, Germany
[2] Royal Belgian Institute for Space Aeronomy, Brussels, Belgium
[3] Royal Netherlands Meteorological Institute, De Bilt, The Netherlands
[4] Institute of Environmental Physics, University of Heidelberg, Heidelberg, Germany
[5] Institute of Environmental Physics, University of Bremen, Bremen, Germany
[6] Max Planck Institute for Chemistry, Mainz, Germany
[7] Airyx GmbH, Justus-von-Liebig-Straße 14, 69214 Eppelheim, Germany
[8] NASA Goddard Space Flight Center, USA
[9] Anhui Institute of Optics and Fine Mechanics, Chinese Academy of Sciences, Hefei, China
[10] Department of Atmospheric Chemistry and Climate, Institute of Physical Chemistry Rocasolano, Madrid, Spain
[11] Laboratory of Atmospheric Physics, Aristotle University of Thessaloniki, Thessaloniki, Greece
[12] Department of Physics, University of Toronto, Toronto, Canada
[13] A. M. Obukhov Institute of Atmospheric Physics, Russian Academy of Sciences, Moscow, Russia
[14] Belarusian State University, Minsk, Belarus
[15] LuftBlick Earth Observation Technologies, Mutters, Austria
[16] Meteorologisches Institut, Ludwig-Maximilians-Universität München, Munich, Germany
[17] Department of Chemistry & Cooperative Institute for Research on Environmental Sciences (CIRES), University of Colorado, Boulder, USA
[18] National Institute for Aerospace Technology (INTA), Madrid, Spain
[19] European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT), Darmstadt, Germany
[20] Netherlands Organisation for Applied Scientific Research (TNO), the Netherlands
[21] Center for Environmental Remote Sensing, Chiba University, Chiba, Japan
[22] Meteorological Observation Center, China Meteorological Administration, Beijing, China & Chinese Academy of Meteorological Science, China Meteorological Administration, Beijing, China
[23] National Institute of Water and Atmospheric Research, Lauder, New Zealand
[24] National University of Sciences and Technology, Islamabad, Pakistan
[25] Department of Earth and Environmental Sciences, Indian Institute of Science Education and Research Mohali, Punjab, India
[26] School of Earth and Space Sciences, University of Science and Technology of China, Hefei, Anhui, China
[27] Department of Atmospheric and Cryospheric Sciences, University of Innsbruck, Innsbruck, Austria
[28] Laboratoire Atmosphères, Milieux, Observations Spatiales, Université de Versailles Saint-Quentin-en-Yvelines, Centre National de la Recherche Scientifique, Guyancourt, France
[29] now at Institute for the Protection of Maritime Infrastructures, German Aerospace Center (DLR), Bremerhaven, Germany
[30] Institute of Meteorology, University of Natural Resources and Life Sciences, Vienna, Austria
[31] Virginia Polytechnic Institute and State University, Blacksburg, VA, USA
[32] Center of Marine Environmental Sciences (MARUM), University of Bremen, Bremen, Germany

[33] Energy, Environment and Water Research Center (EEWRC), The Cyprus Institute, Nicosia, Cyprus
[34] Shanghai Key Laboratory of Atmospheric Particle Pollution and Prevention, Department of Environmental Science & Engineering, Fudan University, Shanghai, China
[35] Remote Sensing Technology Institute, German Aerospace Center (DLR), Oberpfaffenhofen, Germany
[36] now at Measurement and Analysis Research Section, Environment and Climate Change Canada, Toronto, M3H 5T4, Canada

Correspondence to: Karin Kreher (karin.kreher@bkscientific.eu)

Abstract.   In September 2016, 36 spectrometers from 24 institutes measured a number of key atmospheric pollutants for a period of 17 days during the Second Cabauw Intercomparison campaign for Nitrogen Dioxide measuring Instruments (CINDI-2) that took place at Cabauw, The Netherlands (51.97° N, 4.93° E). We report on the outcome of the formal semi-blind intercomparison exercise, which was held under the umbrella of the Network for the Detection of Atmospheric Composition Change (NDACC) and the European Space Agency (ESA). The three major goals of CINDI-2 were (1) to characterise and better understand the differences between a large number of Multi-AXis Differential Optical Absorption Spectroscopy (MAX-DOAS) and zenith-sky DOAS instruments and analysis methods, (2) to define a robust methodology for performance assessment of all participating instruments and (3) to contribute to a harmonisation of the measurement settings and retrieval methods. This, in turn, creates the capability to produce consistent high-quality ground-based data sets, which are an essential requirement to generate reliable long-term measurement time series suitable for trend analysis and satellite data validation.

The data products investigated during the semi-blind intercomparison are slant columns of nitrogen dioxide (NO_2), the oxygen collision complex (O_4) and ozone (O_3) measured in the UV and visible wavelength region, formaldehyde (HCHO) in the UV spectral region and NO_2 in an additional (smaller) wavelength range in the visible. The campaign design and implementation processes are discussed in detail including the measurement protocol, calibration procedures and slant column retrieval settings. Strong emphasis was put on the careful alignment and synchronisation of the measurement systems, resulting in a unique set of measurements made under highly comparable air mass conditions.

The CINDI-2 data sets were investigated using a regression analysis of the slant columns measured by each instrument and for each of the target data products. The slope and intercept of the regression analysis respectively quantify the mean systematic bias and offset of the individual data sets against the selected reference (which is obtained from the median of either all data sets or a subset), and the RMS error provides an estimate of the measurement noise or dispersion. These three criteria are examined and for each of the parameters and each of the data products, performance thresholds are set and applied to all the measurements. The approach presented here has been developed based on heritage from previous intercomparison exercises. It introduces a quantitative assessment of the consistency between all the participating instruments for the MAX-DOAS and zenith-sky DOAS techniques.

1   Introduction

Passive UV-visible spectroscopy using scattered sunlight as a light source provides one of the most effective methods for routine remote sensing of atmospheric trace gases from the ground. While zenith-sky observations have been used for several decades to monitor stratospheric gases such as NO_2, O_3, BrO and OClO (e.g. Noxon, 1975; Platt et al., 1979; Solomon et al., 1987; Pommereau and Goutail, 1988; Richter et al., 1999; Liley et al., 2000; Hendrick et al., 2011, Yela et al., 2017), measurements scanning the sky vertically at several elevation angles between horizon and zenith have been established more recently. In addition to total columns, the so-called MAX-DOAS (Multi-Axis Differential Optical Absorption Spectroscopy; Hönninger et al., 2004) technique also allows the derivation of vertically resolved information on a number of tropospheric species such as NO_2, HCHO, BrO, glyoxal, IO, HONO, SO_2, etc. (see e.g. Hönninger and Platt, 2002; Wittrock et al., 2004; Heckel et al., 2005; Lee et al., 2008, 2009; Sinreich et al., 2010; Frieß et al., 2011; Hendrick et al., 2014, Prados-Roman et al., 2018) as well as aerosols (see e.g. Wagner et al., 2004; Frieß et al., 2006; Clémer et al., 2010, Ortega et al., 2016). The number of MAX-DOAS instruments used worldwide has grown considerably in recent years notably in support of satellite validation (e.g. Wang et al., 2017a; Herman et al., 2018) and for urban pollution studies (e.g. Gratsea et al., 2016; Wang et al., 2017b) and this increase in deployment of MAX-DOAS instrumentation for tropospheric observations, together with the diversity of the designs and operation protocols, has created the need for regular formal intercomparisons which should include as many different instruments as possible.

In 2005 and 2006, two field campaigns were held at Cabauw, The Netherlands, involving MAX-DOAS instruments as part of DANDELIONS (Dutch Aerosol and Nitrogen Dioxide Experiments for vaLIdation of OMI and SCIAMACHY). This project was dedicated to the validation of satellite NO_2 measurements by the Ozone Monitoring Instrument (OMI) and SCIAMACHY (Scanning Imaging Absorption SpectroMeter for Atmospheric CartographY) and aerosol measurements by OMI and the Advanced Along-Track Scanning Radiometer (AATSR) (Brinksma et al., 2008). This was followed by the first Cabauw Intercomparison campaign for Nitrogen Dioxide measuring Instruments (CINDI) which was organised in 2009 under the auspices of the European Space Agency (ESA), the Network for the Detection of Atmospheric Composition Change (NDACC) and the European Union (EU) FP6 Global Earth Observation and MONitoring (GEOMON) project. This effort resulted in the first successful large-scale intercomparison of both MAX-DOAS and zenith-sky ground-based remote sensors of NO_2 and O_4 slant columns (Roscoe et al., 2010). Datasets of NO_2, aerosols and other air pollution components observed during CINDI were documented in a number of peer-reviewed articles (Piters et al., 2012; Roscoe et al., 2010; Pinardi et al., 2013; Zieger et al., 2011; Irie et al., 2011 and Frieß et al., 2016), providing an assessment of the performance of ground-based remote sensing instruments for the observation of NO_2, HCHO and aerosol. Recommendations were issued regarding the operation and calibration of the instruments, the retrieval settings, and the observation strategies for use in ground-based networks for air quality monitoring and satellite data validation. Several important findings were highlighted in view of preparing future campaigns, in particular (1) the need for accurate calibration and monitoring of the elevation angle of MAX-DOAS scanners and (2) for intercomparison purposes, the importance of synchronising measurements in time and space very accurately. The lack of such a synchronisation was indeed considered as being responsible for a large part of the scatter observed during CINDI (Roscoe et al., 2010), which limited the interpretation of the results.

Seven years after CINDI, a second campaign (CINDI-2) was undertaken at the same site (Cabauw Experimental Site for Atmospheric Research - CESAR) from 25 August until 7 October 2016. Its goal was to intercompare the new and extended generation of ground-based remote-sensing and in-situ air quality instruments. The interest of ESA in such intercalibration activities is motivated by the ongoing development of several UV-visible space missions targeting air quality monitoring such as the Copernicus Sentinel 5 Precursor (S5P) satellite launched in October 2017 and the future Copernicus Sentinel 4 and 5 satellites. The validation and ongoing support of measurements from such space missions is essential and requires dedicated ground-truth

measurement systems. Because tropospheric measurements from space-borne nadir UV-visible sensors show little or no vertical discrimination and inherently provide measurements of the total tropospheric amount, surface in-situ measurements are generally unsuitable for such a validation effort. Instead, validation requires a technique that can deliver column-integrated and vertically resolved information on the key tropospheric species measured by satellite instruments such as NO_2, HCHO and SO_2 with a horizontal representativeness compatible with the resolution of space measurements (e.g. 3.5x7 km^2 for S5P).

Hence, the specific goals of CINDI-2 were to support the creation of high-quality ground-based data sets as needed for long-term measurements, trend analysis and satellite data validation. To achieve this, it is essential to characterise the differences between a large number of MAX-DOAS and zenith-sky DOAS instruments and analysis methods and to assess the participating instruments in their ability to retrieve the same geophysical quantities (i.e. slant columns of NO_2, O_4, HCHO and O_3) when measured and processed in a controlled way (i.e. using a prescribed measurement protocol and retrieval settings). The design of the CINDI-2 campaign and the development of the measurement protocol, adhered to specifically during the official intercomparison phase, was based on the experience gained during the first CINDI campaign in 2009 as well as more recent projects and campaigns such as the MAD-CAT campaign in Mainz, Germany, in 2013 (e.g. Peters et al., 2017).

This paper is organized as follows. In Section 2, the campaign design is discussed including an overview of the participating groups and their instruments, and a discussion of the measurement protocol details. In Section 3, the results of the semi-blind slant column intercomparison are presented, and in Section 4, a systematic approach is proposed to quantitatively assess the performance of the participating instruments for the different target trace gas data products. Section 5 provides recommendations for observation networks and future intercomparison campaigns and Section 6 summarizes the campaign outcomes.

2   Intercomparison campaign design and measurement protocol

The CESAR site was accessible for the installation of the instruments from 25 August 2016 onwards, with the formal semi-blind intercomparison being held for 17 days from 12 – 28 September 2016. Here, we concentrate on this official intercomparison phase of the CINDI-2 campaign, and measurements and results are discussed for this time period only. A general description of the overall campaign including a more detailed discussion of the CESAR site and all ancillary measurements can be found in Apituley et al. (to be submitted to AMT, 2020). In short, the CESAR site at Cabauw is overall a rural site, with only a few pollution sources nearby, but the wider vicinity of Cabauw is densely populated, with the cities of Utrecht, Amsterdam, The Hague and Rotterdam less than 60 km away and a dense highway grid within 25 km, so that the site experiences recurring pollution events, e.g. such as from the daily morning and afternoon rush hours.

The MAX-DOAS instruments were also complemented with a suite of in-situ, profiling and mobile observations which are described in detail by Apituley et al. (to be submitted to AMT, 2020). In particular, a long-path DOAS measuring near surface mixing ratios of NO_2 and HCHO but also a range of other species such as HONO and SO_2 (see e.g. Merten et al, 2011, for a description of the technique) was operated at the CESAR site for the period of the campaign. Several mobile MAX-DOAS measurements were also made around Cabauw, and between Rotterdam and Utrecht (e.g. Merlaud, 2013) in addition to the static observations. NO_2 profiles were measured with NO_2 sondes (Sluis et al, 2010) and lidar (e.g. Volten et al., 2009), as well as through in-situ observations using the Cabauw meteorological tower. Extensive aerosol information was also gathered using Raman aerosol lidar and in situ samplers.

2.1 Instruments

Table 1 lists the groups and instruments that were included in the CINDI-2 semi-blind intercomparison, and an overview of the relevant instrumental details is given in Table 2. Among the 36 participating instruments, 17 were two-dimensional (2D) MAX-DOAS systems allowing for scans in both elevation and azimuth, 16 were one-dimensional (1D) MAX-DOAS systems performing elevation scans in one fixed azimuthal direction, one was an imaging DOAS instrument (Imaging MaPper for Atmospheric observaTions - IMPACT, Peters et al., 2019) for which only measurements in the common viewing direction were submitted, and the last two instruments were zenith-sky DOAS systems of the SAOZ (Système d'Analyse par Observation Zénithale) (Pommereau and Goutail, 1988) and most recent Mini SAOZ version. The complete technical specifications for each instrument can be found in Section 3 of the Supplement.

Instruments have been sorted into different categories. Custom-built systems refer to instruments developed by scientific organisations for their own research activities. Other categories denote commercial systems of various types. Pandora instruments (Herman et al., 2009) are being developed at NASA/LuftBlick, commercialised by the SciGlob company and deployed as part of the Pandonia Global Network (PGN) (http://pandonia.net/). EnviMes[1] MAX-DOAS instruments (Lampel et al., 2015) have been recently commercialised based on expertise developed at the University of Heidelberg. Mini-DOAS instruments (e.g. Hönninger et al., 2004; Bobrowski, 2005) are produced in Germany by Hoffmann GmbH (http://www.hmm.de/).

No particular guidelines were given concerning the spectral calibration of instruments, which means that participating groups were free to apply calibration steps of various levels of complexity. In addition to standard calibration procedures involving dark current and electronic offset corrections, wavelength registration and slit function determination, some groups performed more advanced pre-processing steps such as radiometric calibration, stray-light and interpixel variability correction or an explicit correction for detector response non-linearity, the latter being a known feature of Avantes spectrometers.

2.2 Campaign design

To allow for optimal synchronisation of the measurements, all the spectrometers participating in the semi-blind intercomparison exercise were installed in close proximity to each other on the remote-sensing site (RSS) of the CESAR station (see Figure 1 and Apituley et al., to be submitted to AMT, 2020). To achieve this, mobile units (similar to shipping containers) were temporarily installed for the campaign period.

The rationale behind this setup was to arrange the instruments in such a way to minimise ambiguity in air masses observed simultaneously by all spectrometers. This is essential for tropospheric NO_2 but also for aerosol and HCHO, since all these species can feature rapidly changing concentrations in both space and time. Considering the large number of systems that needed to be accommodated, two rows of containers were deployed with the bottom row being similar to the one deployed during the previous CINDI campaign. This bottom row of containers was predominantly used to host the 1D MAX-DOAS instruments and the two zenith-sky systems. A second row of containers was deployed on top of the first one, with the stacked double-containers providing additional height. All 2D MAX-DOAS systems were installed on the roof of the top-level containers allowing for more flexibility on the azimuth scan settings and avoiding any risk of interference with the 1D systems. All the 1D MAX-DOAS instruments used the same azimuth viewing direction of 287° (i.e. approximately WNW, with North (N) being 0° and East (E) 90° etc.) which was

---

[1] Now: SkySpec from Airyx GmbH (www.airyx.de)

already used during the first CINDI campaign since it provided an unobstructed view to the horizon. This direction was also one of the azimuth directions used by the 2D MAX-DOAS systems (see also discussion of the measurement protocol in Section 2.4).

In Sections 2.4 – 2.6, further procedures aiding the comparability of the MAX-DOAS measurements such as the overall
5   measurement protocol, elevation angle calibrations and slant column retrieval settings are discussed in more detail. Prescribing these procedures as strictly as possible was highlighted during previous campaigns as important (see in particular Roscoe et al., 2010) and the campaign design of CINDI-2 focused on implementing such recommendations.

2.3   Semi-blind intercomparison

10   As in previous intercomparison campaigns of the same type (see e.g. Vandaele et al., 2005; and Roscoe et al., 1999, 2010), a semi-blind intercomparison protocol was adopted. The CINDI-2 intercomparison exercise had three key objectives: (1) To characterise the differences between a large number of measurement systems and approaches, (2) to discuss the performance of the various types of instrument and define a robust methodology for performance assessment and (3) to provide guidelines to further harmonise the measurement settings and analysis methods. The adopted semi-blind intercomparison protocol was based on the following approach:

15   a)   The data acquisition schedule applied by the participants was strictly prescribed to coordinate the timing and geometry of each individual measurement as exactly as possible, so that the same air mass could be measured by all instruments with good synchronisation.

        b)   For each data product, a set of retrieval settings and parameters was prescribed (see Appendix A). These were mandatory for participation in the semi-blind exercise. The data analysis software, however, was not prescribed and the different
20   software types used by each institute are listed in Table 3.

        c)   All slant column data sets measured during the previous day were submitted to an independent campaign referee (K. Kreher) and her assistant (E. Dimitropoulou) every morning by 10:00 local time. At daily meetings in the afternoon (usually at 16:00), the results of the slant column comparison for measurements from the previous day were displayed anonymously, i.e. without any assignment to the different instruments. Basic analysis plots exploring the differences in the data sets
25   measured during the previous days were shown and discussed.

        d)   The referee notified instrument representatives if there was an obvious problem with their submitted data set so that this issue could be addressed and, if possible, corrected for the remainder of the campaign.

        e)   After the formal campaign had finished, all participants had about three weeks to undertake the analysis according to the prescribed measurement and analysis protocol (see Section 2.4), and the final slant column data sets had to be submitted
30   by 18 October 2016. After this date, any resubmissions were only accepted if the group could clearly state the reasons why the data set needed to be updated, e.g. if an error was found in the analysis and needed to be remedied. Further details on this process are given in Section 3.3 and Appendix B.

The semi-blind intercomparison exercise focused on a limited number of key data products of direct relevance for satellite validation and NDACC operational continuity. These data products are listed in Table 4. Depending on the specific characteristics of their
35   instrumentation, participants were free to submit all or only a subset of the data products.

2.4   Measurement protocol

As discussed above, it was recognised in previous intercomparison campaigns (see in particular Roscoe et al., 2010) that the achievable level of agreement between MAX-DOAS sensors is often limited by imperfect co-location and a lack of synchronisation. This problem is especially critical for tropospheric NO_2 comparisons, because of the large variability of this pollutant on very small
40   scales. However, it is also relevant for other gases such as HCHO, O_4, SO_2, glyoxal, etc. For this reason, it was decided to co-locate

all the MAX-DOAS instruments on the same observation platform (see Section 2.2) and additionally to impose a strict protocol on the timing of the spectral acquisition.

The baseline for all MAX-DOAS instruments was to point towards a fixed azimuth direction (287°) throughout the day. This direction was chosen because of the very close to obstruction-free line of sight towards the horizon. In addition, the 2D MAX-DOAS instruments performed azimuthal scans simultaneously according to a strict measurement schedule. The scheme described below was designed to ensure the maximum of synchronisation between the same type of instruments (e.g. azimuthal scans by 2D MAX-DOAS) but also between the different types of instruments (1D and 2D MAX-DOAS, and zenith-sky DOAS). A distinction was made between twilight (morning and evening) and daytime conditions, for which separate data acquisition protocols were prescribed. According to the geometry of the solar position during the campaign, the daytime period (excluding twilight) was defined to be from 6:00 to 16:45 UTC with 6:00 UTC corresponding to an SZA of approximately 83° - 87° and 16:45 UTC to an SZA of approximately 76° - 82°, depending on the exact date during the campaign.

To allow for an NDACC-type intercomparison of stratospheric measurements (e.g. Vandaele et al., 2005), zenith-sky twilight observations were also performed. The acquisition scheme for the dawn observations prescribed 39 measurements with a duration of 3 min each (integration time: 170s; overhead time: 10s), starting at 04:00:00 UTC and ending at 05:57:00 UTC. This sequence was followed by a 180s (3 min) interval allowing for a transition to the MAX-DOAS mode of which the first scans started at 06:00:00 UTC. For measurements at dusk, 40 acquisitions were recorded with a duration of 180s each starting at 16:45:00 UTC and ending at 18:45:00 UTC.

During daytime, the acquisition scheme for MAX-DOAS and zenith-sky systems included four sequences of 15 minutes per hourly slot starting at 06:00:00 UTC. Individual acquisitions (at one given angle) were set to one-minute-long in all cases. For 1D systems, the pointing azimuth direction was set to 287° with elevation angles of 1, 2, 3, 4, 5, 6, 8, 15, 30 and 90°. For 2D systems, the azimuth angles 45, 95, 135, 195, 245 and 355° were successively sampled in addition to the reference angle of 287°. In each azimuthal direction, four elevation angles (1, 3, 5, 15°) were scanned except for the reference azimuth of 287° where the same elevations as prescribed for the 1D MAX-DOAS systems were used. One zenith reference spectrum was recorded every 15 minutes, and for 2D systems or instruments equipped with a sun tracker, almucantar scans and/or direct-sun measurements were performed between the 10th and 15th minute of the sequence. For zenith-sky instruments, one-minute-long acquisitions were performed during the whole day from 06:00.00 UTC to 16:44:00 UTC.

Figure 2 (upper plot) provides an overview of the number of days each instrument was on duty during the intercomparison period. It also illustrates (lower plot) the accuracy with which the different groups were able to match the imposed measurement protocol. As can be seen, the instruments were in operation most of the time during the 17 days of the semi-blind period and most of them were able to follow the schedule to better than one minute. In comparison to past campaigns, the level of synchronisation was clearly improved which significantly reduced the need for smoothing or interpolating data in time. As a result, the impact of the atmospheric variability on the data comparisons could be reduced considerably, but not completely eliminated (see Section 3.7).

As discussed above, the measurement procedure was strict but in spite of this comprehensive protocol, there was still some freedom left on how to implement details of the acquisitions. E.g. for managing the acquisition time, most groups decided to move the telescope and gather the spectra within the prescribed one minute time period, while INTA (inta-17) gathered spectra for one minute and then moved the telescope. As a result, a time shift was accumulated when compared to other groups (see Figure 2). Chiba-9 also shows a noticeable time shift due to constraints in the acquisition software that prevented the strict implementation of the

protocol. In the case of niwa-30, the large time shift in the UV was due to instrument imposed alternating between measurements in the visible and UV wavelength regions (hence only one spectral range could be synchronised with the protocol).

Likewise, it must be noted that Pandora instruments also take separate measurements for the visible and the UV range, where a blocking filter is inserted in the optical path for the UV measurements in order to reduce spectral stray light. Therefore, a compromise had to be found in the time synchronization bracketing the requested measurement time. This is the reason for the systematic offsets for Pandoras in the bottom panel of Figure 2. Another consequence of this was that the total measurement time of Pandora instruments was about half the time of the other participating instruments, which affects the noise levels for Pandoras described throughout this paper to some extent.

2.5   Calibration of the MAX-DOAS elevation scans

Because of the importance of the elevation pointing accuracy for MAX-DOAS measurements at low elevation and as recommended after the first CINDI intercomparison (Roscoe et al., 2010), different calibration tests involving all the participating instruments were undertaken during both the warm-up and semi-blind intercomparison phases. Three different approaches were used:

1) On several evenings, MPIC (Max Planck Institute for Chemistry) installed an Opel car 1999 xenon lamp with a 17 cm diameter lens at a distance of 1280 m from the measurement site (angular lamp extension ~0.008°) in the main viewing azimuth direction (287°) of the MAX-DOAS instruments. It served as a common light source at long distance, and MAX-DOAS instruments recorded downward and upward scan spectra pointing towards the lamp.

2) A white stripe on a black target at known elevation close to the instruments was scanned.

3) Intensities were measured regularly during horizon scans (see Section 3.2 for details).

Additional calibration measurements using a near-distance lamp placed a few meters away from instruments were also performed by IUP-Heidelberg and several other groups. Overall, these calibration procedures allowed the pointing accuracy of the different instruments and their stability during the campaign to be fully characterized (see Donner et al. in review, 2019). As such they played an important role for the interpretation of the semi-blind intercomparison results (see Section 3.7).

2.6   Slant column retrieval settings

To minimise the sources of difference between measurements, a set of common retrieval settings and parameters was prescribed ahead of the campaign. The use of these settings was mandatory for participation in the semi-blind exercise. The detailed spectral retrieval settings imposed for each data product referenced in Table 1 are given in Appendix A. These settings were based on the NDACC protocol for UV-Vis measurements (http://www.ndaccdemo.org/data/protocols) as well as results from the first CINDI campaign (e.g. Pinardi et al., 2013), the MAD-CAT campaign (http://joseba.mpch-mainz.mpg.de/mad_analysis.htm) and the QA4ECV project (http://www.qa4ecv.eu/). Although not necessarily optimal, they represent a common baseline applicable to all data sets in a consistent way. Concerning the choice of the Fraunhofer reference spectrum, daily reference spectra obtained from the mean of all zenith-sky spectra acquired between 11:30:00 and 11:41:00 UTC were used. Slant columns retrieved against this reference spectrum are hereinafter referred to as differential slant column densities (dSCDs).

Note that additional retrievals were also performed using sequential reference spectra (zenith-sky observations taken close to the time of the respective horizon measurements). These data were, however, not included in the formal semi-blind intercomparison since they essentially lead to similar comparison results as the analyses using daily reference spectra. They were also not available from all groups. Moreover, the use of daily reference spectra presents the advantage of being directly applicable to twilight measurements and provides a better test of the instrumental stability over several hours of operation. As already noted in Section

2.1, the determination of the instrumental slit function and its eventual wavelength dependence was under the responsibility of the participating groups.

3 Semi-blind intercomparison results

3.1 Overview of slant column measurements and meteorological conditions

The meteorological conditions during CINDI-2 were exceptionally favourable for the location and season. The uppermost panel of Figure 3 shows the hourly sunshine duration and surface temperature records for the whole semi-blind intercomparison period (for more details, see Apituley et al., to be submitted to AMT, 2020). The first four days of the semi-blind phase were characterized by a clear sky with some haze in the morning and very high air temperatures for the season (>30°C), allowing for efficient formaldehyde production. The next seven days were cloudier with lower temperatures. The last six days of the semi-blind exercise were also characterized by mostly clear sky or occasionally broken cloud conditions.

All other panels of Figure 3 display the time variation of each of the dSCD data products included in the intercomparison, as measured by the IUP Bremen instrument, which had excellent data coverage throughout the campaign duration. Green lines represent zenith-sky measurements, red lines off-axis data at 30° elevation, and blue lines off-axis measurements up to 15° elevation. Results show a large variability of the NO_2, O_4 and HCHO tropospheric columns while ozone data display the expected regular diurnal pattern mainly due to the variation of the stratospheric light path during the ascent and descent of the sun. Due to the unusually favourable weather conditions, higher than expected values were observed for tropospheric HCHO while tropospheric NO_2 was at its lowest during the first Sunday (18 September) of the intercomparison campaign. The variability of the tropospheric trace gas content and the exceptionally large number of clear-sky sunny conditions were ideal for comparison purposes.

3.2 Horizon scans

Horizon scans, which consist of measuring the change in intensity when scanning the sky radiance across the horizon line, were systematically performed every day at noon during the semi-blind intercomparison period. Although difficult to calibrate absolutely because the horizon is generally not free of obstacles (e.g. trees, buildings or terrain height fluctuations), they provide a simple and valuable technique for monitoring the elevation pointing stability of MAX-DOAS instruments. Figure 4 shows an example of the variation of the intensity at 440 nm, as reported by the IUP-Bremen instrument (blue circles). Considering that the intensity measured as a function of the elevation angle yields the integral over the telescope's point spread function, measurements were fitted using an error function (Gaussian integral) according to equation 1:

$$S = A\left[erf\left(\frac{x-x_0}{B}\right) + 1\right] + C(x - x_0) + D \qquad \text{Eq. (1)}$$

where x is the elevation angle, and A, B, C and D are fitting parameters. The centre (x_0), also fitted, provides a measure of the horizon elevation.

The analytic derivative of equation 1 is a Gaussian curve of which the full width at half maximum (FWHM) is given by:

$$FWHM = 2\sqrt{\ln(2)}\, B \qquad \text{Eq. (2)}$$

We used this quantity to estimate the effective field of view (FOV) of the instrument (see Figure 4, red line).

Applying this fitting methodology, horizon scans delivered daily by each group were systematically analysed. Figure 5 presents an overview of the time evolution of the horizon elevation derived from each instrument (and their median values represented by red lines), all of them being measured in the visible wavelength range except for knmi-21. The same analysis was also performed at UV

wavelengths. A summary of the resulting median and 1σ standard deviation FOV derived from each instrument is presented in Figure 6.

The time series of horizon scans provide a useful assessment of the stability and precision of the elevation pointing devices used by the different instruments. In some cases, horizon scans allowed the identification of calibration biases, which could then be addressed by the instrument teams and corrected straight away. This is in particular the case for the dlrustc-13 and -14 instruments. Considering the effective field of view (FOV), a large variability between the instruments was identified. This generally reflects differences in the optical design of the different systems. However, horizon scans can also be influenced by atmospheric conditions and by perturbations of the light intensity at the horizon (e.g. due to fog, high aerosol loads or refraction at temperature inversions). Nevertheless, it is striking to note in Figure 6 that horizon elevations tend to be systematically higher at visible wavelengths than at UV ones. Likewise, FOVs measured in the UV tend to be wider than in the visible. This variation is larger than expected from typical chromatic aberration effects in telescope lenses. The reason for this behaviour is not fully understood but it is likely related to the wavelength dependence of the surface albedo, which may affect the horizon scan fitting process (for more details, see Donner et al. in review, 2019).

3.3   History of slant column data set revisions

As described in Section 2.3, semi-blind dSCD data sets had to be submitted by 18 October 2016, i.e. three weeks after the end of the formal intercomparison period. However, resubmissions were accepted after this date when a clear justification was provided for the change. The main motivation for accepting late revisions was to remedy well-identified mistakes. Details of the submitted revisions, including justifications for the changes and corresponding dates, are given in Appendix B.

3.4   Pre-processing of the slant column data

Before further processing, the dSCD measurements from all groups were checked to remove unphysical values and obvious outliers. For this purpose, the following filters were applied: 1) dSCD data exceeding 10 times the daily median values from the instrument were excluded, 2) data points with fitting RMS exceeding 4 times the daily median RMS were removed.

In addition, the results from the horizon scan analysis (see Section 3.2) were used to readjust the elevation angle of instruments presenting absolute elevation offsets larger than $1.5°$. This correction was performed assuming a reference horizon elevation of $0.1°$, as determined independently using lamp measurements performed at night combined with an analysis of terrain height variations (Donner et al. in review, 2019). The impact of this angular correction is illustrated in Figure 7 for NO_2 dSCD measurements, which are here represented in terms of their relative difference with respect to median values from a selection of the participating instruments (for more details see Section 3.5, and Figure 8). As can be seen, the large biases observed during the first few days of the campaign for some instruments were due to systematic mispointing effects well compensated by the correction. The impact of the correction is largest for NO_2, but it is also significant for other tropospheric species, in particular O_4. This again stresses the importance of accurately calibrating the elevation scanner of MAX-DOAS instruments.

3.5 Determination of reference comparison data sets

As in previous campaigns, the intercomparison of dSCD measurements was based on pre-selected reference data sets. In CINDI-2, these were based on the calculation of median dSCDs obtained from a selection of measurements presenting an acceptable agreement. Here, the selection of the reference groups, different for each data product, was performed after an initial regression analysis using the median of all data as reference. Only groups satisfying the performance criterion for the regression slopes were retained (see Section 4 and Table 4 for more details). The data sets included in the median references are displayed in Figure 8 for both MAX-DOAS and zenith-sky twilight data products. In the particular case of HCHO, the selection was performed through visual inspection of the dSCD comparisons. Only data sets displaying consistent behaviour at 30° elevation (the angle generally used to retrieve first guess total tropospheric columns using the geometrical approximation; see Hönninger and Platt, 2002) were retained for building the reference. This can be appreciated in Figure 9 where time series of the HCHO dSCDs measured by each group are compared to the reference values. As can be seen, many data sets display noisy and/or unphysical negative values and only the four selected groups (bira-4, iupb-18, mpic-28 and niwa-29) present mutually consistent values. Note that a similar approach was used for the selection of the HCHO dSCD reference in Pinardi et al. (2013).

3.6 Initial assessment of the overall agreement between measurement data sets

Tables 5 and 6 show the mean relative differences (in percent) from the reference dSCDs and their 1st sigma standard deviation for all participating instruments and, respectively, for all MAX-DOAS products and for all zenith-sky DOAS products. Extreme outliers (values exceeding percentile 97) are excluded from the analysis, as well as MAX-DOAS ozone measurements since these show very small off-axis enhancements (see Figure 8). Both tables provide an overall initial assessment of the intercomparison results indicating that for most data products (except HCHO), instruments generally agree within a few percent for the most relevant range of elevation angles of 1°-10° for MAX-DOAS data and for an SZA of 80°-93° for zenith-sky twilight data. One can also see that the overall agreement between instruments is better in the visible than in the UV spectral range.

For HCHO (last two columns of Table 5), the differences between the instruments are comparatively larger and, in some cases, extreme. However, restricting the analysis to the first four days of the measurement campaign (when the air temperature was warmer and the HCHO dSCDs higher) reduces discrepancies significantly and, although a higher spread remains compared to any of the other products, one can conclude that under such favourable conditions a large number of the participating instruments provide consistent HCHO dSCD measurements. For amoiap-2, however, the instrument was operated in different modes during different time periods with some modes being more advantages for the HCHO data analysis than others. The group found that when only HCHO data acquired during the optimal time period is used, the mean relative difference is substantially lower, approximately -16%. More details on the instrument and the different modes are provided in Borovski et al. 2017a and Borovski et al. 2017b.

The last row of Tables 5 and 6 shows the median values from the table entrees for each column. The median of the differences is by construction close to zero (but not exactly zero since the median reference values are derived from a selected subset of the participating instruments), while the median of the standard deviations provides an estimate of the most probable size of the deviations against the reference. For example, the median value for zenith-sky DOAS NO2uv shows the highest deviation from the reference when compared to the other zenith-sky DOAS products. For the MAX-DOAS data products, as expected, HCHO shows by far the highest deviation.

3.7   Regression analysis

The approach adopted for the CINDI-2 intercomparison follows from previous exercises, in particular the CINDI campaign (Roscoe et al., 2010) and previous NDACC intercomparisons (Vandaele et al., 2005; Roscoe et al., 1999). It is based on the systematic analysis of regression plots between individual measurements and corresponding median reference values (see Section 3.5). Assuming negligible uncertainties on the reference dSCDs, we use a simple linear least-squares regression method weighted by reported dSCD uncertainties. Owing to the strict measurement protocol imposed for the campaign, most measurement points could be compared one-to-one without the need for further interpolation or averaging. When interpolation was necessary, a simple linear procedure was used to bring measurements in line with the campaign protocol (see Section 2.4). This implies that, in comparison to previous similar exercises, sampling and mismatch errors (air mass co-location errors) could be reduced considerably, so that comparison noise and biases should more accurately reflect the intrinsic instrumental performances. This question is further investigated below.

Linear correlation plots between the dSCDs for each instrument and the median value of all the measurements were systematically generated for the complete semi-blind intercomparison time period for each data product, and for each elevation angle and azimuth viewing direction. This allowed identification of, e.g., whether a specific issue arose from particular observation geometries for one or several instruments. Concerning zenith-sky twilight analyses, zenith measurements were selected in a limited range of solar zenith angles (from 75° to 93°) representative of typical twilight measurements, similar to as performed within NDACC for stratospheric ozone and NO_2 monitoring (see e.g. Hendrick et al., 2011) where an SZA range from 86°-91° is used. Figure 10 and 11 show examples of the regression analysis for the case of MAX-DOAS NO_2 and O_4 measured in the visible spectral range. A more complete overview of the regression results obtained for all species can be found in the Supplement where the regression analysis is shown for all elevation angles and viewing directions. As can be seen, a tight correlation is observed for most of the participating instruments. The values for the slope (S), intercept (I), and the RMS calculated as part of the regression analysis are shown in each of the instrument panels. The slope and intercept parameters, respectively, quantify the mean systematic bias and offset of individual data sets against the median reference, while the RMS error provides an estimate of the measurement noise or dispersion.

A similar analysis is presented in Figure 12 for HCHO. Note the much larger relative noise obtained for this weak absorber, and the larger dispersion of the results. For this molecule, low-noise research grade instruments perform significantly better than other systems. A similar conclusion was reached in Pinardi et al. (2013) (see in particular Figure 18). Note however, that instruments equipped with compact Avantes spectrometers (e.g. the Pandora and EnviMes instruments) also provide good results despite a larger noise level.

It is interesting to further investigate the dSCD noise levels and their dependencies. Two approaches are generally used to characterize the random uncertainties of dSCD measurements. The first one consists of inspecting the dSCD uncertainties produced by the DOAS least-squares fitting procedure. Assuming normally distributed residuals, these uncertainties provide a good estimate of the random uncertainty due to instrument noise. Figure 13 (panel a) displays DOAS fit dSCD errors normalised to their median for the 12 data products investigated in this exercise for all instruments and all elevation angles. For each box, the bottom and top edges of the box indicate the 25th and 75th percentiles, respectively, while the whiskers extend to the most extreme data points. Median dSCD error values are given for reference on the upper x-axis. Next to the fitting errors, in the right panel of Figure 13, are represented the RMS residuals from regression analyses, normalised in the same way as the dSCD errors. Owing to the good synchronisation achieved during CINDI-2, these RMS values provide a good estimate of the comparison noise against median dSCD references. Assuming ideal comparison conditions (i.e. perfect co-location in time and space under stable atmospheric conditions),

one would expect these two independent estimates of random uncertainties to converge towards a common value. This happens to be approximately the case for HCHO and for most of the twilight (stratospheric) data products, except for the O3vis product. In contrast, however, regression noise values derived for NO_2 and O_4 dSCDs appear to be much larger than their corresponding fitting uncertainties, and in the case of the NO2vis product, the difference is most pronounced.

The results shown in Figure 13 indicate that despite the measurement synchronisation (to better than 1 minute) and the fact that all instruments were oriented and pointing towards the same air masses, the variability of the NO_2 and possibly aerosol or cloud features can be large enough to introduce a difference between the individual data sets in the comparison exceeding the measurement uncertainty by an order of magnitude. This means that in this intercomparison, atmospheric variability limits the reproducibility and

10   representativeness of individual MAX-DOAS measurements for species such as NO_2. Accordingly, it can be argued that for low-noise instruments the random uncertainty on tropospheric NO_2 dSCD measurements is by far dominated by atmospheric variability effects and the details of how this variability is smoothed out by the measurement system (in particular the FOV of the MAX-DOAS telescope and the integration time are key parameters). This also suggests that using DOAS fit errors as a measure of the dSCD error covariance (as often applied in MAX-DOAS profile inversion schemes, see e.g. Clémer et al., 2010; Vlemmix et al., 2015; Frieß et

15   al., 2018) is not appropriate especially for tropospheric NO_2 retrievals. Instead, a more representative estimate of the random error should be derived from the measured variability of the observed dSCD, with the DOAS fit uncertainties being a lower boundary for the measurement uncertainties at best. This issue has been further investigated in a recent publication by Bösch et al. (2018).

This interpretation is strongly corroborated by Figure 14, where the angular dependence of regression noise results is displayed (in

20   green) for the NO2vis, O4vis and HCHO products. As can be seen, the comparison noise on NO_2 dSCDs is largest at the lowest elevation angles and regularly decreases at larger elevations. This behaviour, which is less marked but also observed for O_4, is consistent with atmospheric variability effects since one expects that inhomogeneities of the tropospheric NO_2 field will affect more strongly observations at lowest elevation angles (which have strongest sensitivity to near-surface NO_2). In contrast, the HCHO comparison noise is virtually independent from the elevation angle and close in size to the fitting noise. Note that even at the highest

25   elevation of 30°, the comparison noise on NO_2 and O_4 dSCDs remains larger than the fitting noise, suggesting that atmospheric variability remains a dominant effect at all the angles used for profile inversion. Figure 15 displays results from the same analyses but restricted to reference data sets. Similar conclusions are reached for NO_2 and O_4. In the case of HCHO, the noise level drops considerably, which reflects the high sensitivity of instruments selected for building the HCHO reference. Interestingly, one can also see that regression RMS and fitting residuals now match almost perfectly (and at all elevation angles) meaning that for this

30   molecule most of the residual variance from regressions involving good instruments can be explained by instrument shot noise. Figure 16 displays results obtained when selecting Pandora instruments only. In comparison to other systems, Pandoras are characterised by a larger field of view (see Figure 6) which probably explains the smaller regression RMS observed for NO_2 and O_4 (likely due to a more efficient smoothing of the atmospheric variability).

Figure 17 provides a different view of the data set already presented in Figure 10, displaying the slope, intercept and RMS for the

35   NO_2 (visible) regression analysis graphically for all measurement days and viewing directions, and for several elevation angles (1°, 3°, 5°, 8°, 15°, and 30°). Similar plots have been generated for all the trace gas data products and are provided in Sections 1 and 2 of the Supplement. Note that for two instruments (chiba-9, amoiap-2), only one elevation angle from the above set is available due to technical reasons intrinsic to these instruments. The limits indicated with dashed lines are introduced and discussed further in Section 4. Figure 17 can be compared with similar figures in Roscoe et al. (2010) (Figure 6) and Pinardi et al. (2013) (Figure 7) allowing

40   results from both CINDI campaigns to be linked. It is interesting to note that although the range of variability on the slope and

intercept parameters was similar in both campaigns, the proportion of instruments matching the 5% limit on the slope was significantly improved in CINDI-2 indicating a general improvement of the overall consistency of the measurements.

As to be expected from well-calibrated instruments, the three regression parameters displayed in Figure 17 generally do not show any marked angular dependence. However, some data sets display larger deviations and sometimes also significant angular dependencies. For these cases, the lowest elevation angles often show the largest deviations (e.g. intercept and RMS for nasa-31 and dlrustc-13, slope for uto-36) but not always (e.g. RMS for cu-boulder-11 and slope for iupb-37). Although this certainly does not explain all discrepancies, it is interesting to note that, in many cases, the largest deviations are observed for instruments that did not supply (or could only partially supply) horizon scan information and therefore could not benefit from the angular correction applied in pre-processing (see Section 3.4).

4   Investigation of instrument performance

With MAX-DOAS-type instruments having gained popularity in recent years and their usage becoming more widespread, the need for a reliable and clearly documented assessment process is becoming more pressing. A semi-blind intercomparison campaign such as CINDI-2 provides the ideal conditions to obtain a data set for such a process and the opportunity to involve as many MAX-DOAS instruments as possible.

Three criteria based on the regression analysis discussed in Section 3.7 (slope, intercept and RMS) have been selected to assess the performance of each of the participating instruments with regard to the eight MAX-DOAS and four zenith-sky products. For each of these parameters, specific limits have been set for the performance evaluation as listed in Table 4. These were semi-empirically derived from a visual inspection of the distribution of the slope, intercept and RMS values for each of the eight CINDI-2 MAX-DOAS and four zenith-sky data products. The histograms and limits (indicated with red lines) for the eight MAX-DOAS data products are displayed in Figure 18 for the slope, intercept and RMS from the regression analysis. Note that the NO_2 and O_3 criteria were adapted from previous NDACC campaigns (see Introduction for further details). For other products, limits were set arbitrarily to capture the most probable values while excluding clear outliers. The limits were, however, chosen to exceed the median of the measurements (indicated with blue lines). The blue lines represent the percentiles 16 and 84 (84 only for RMS) and it can be seen that the certification criteria have been chosen to exceed these limits. One exception is HCHO, since for this product the difference between well performing and less well performing instruments was much larger than for the other products.

It must be acknowledged that the performance limits defined in this work (as in previous NDACC intercomparisons) are representative of the current state-of-the-art of the instrumentation, and to some extent also reflect the measurement conditions in Cabauw. Another campaign being performed e.g. in a cleaner or more stable site could lead to different values for the limits.

Figure 19 shows a summary of the same regression statistics previously discussed in Section 3.7 and displayed in Figure 10 and 15 but with all individual elevation angles added up resulting in one single value for each parameter, instrument and data product. This means that only 3 values are displayed for each instrument. The green shaded areas denote the limits defined in Figure 18 with all parameters falling within the limits being displayed as blue dots while values in red are not meeting the respective criterion. Note that not all of the 36 instruments measure NO2vis. For the slope of the NO2vis regression analysis shown in the top panel, two instruments (uto-36 and amoiap-2) fall outside the limit. One other instrument (iupb-37, the imaging instrument) is not meeting the criterion set for the intercept (see middle panel) and one (nasa-31) for the RMS (bottom panel). One of such summary plots has

been produced for each of the eight MAX-DOAS and four zenith-sky data products which can be individually viewed in Sections 1 and 2 of the Supplement.

To further summarize the outcome of the regression analysis and provide an overview of all 8 MAX-DOAS data products, Figure 20 displays the three selected parameters for all participating instruments. The performance is colour coded with regard to parameters falling inside the performance limit (green) or not meeting the criterion (orange). In exceptional cases where the slope or RMS is exceeding the threshold by more than a factor of 4, the performance is colour-coded in black. Just under one-third of all the instruments do meet all the criteria. Figure 21 shows the same summary for the four zenith-sky data products. In this case, more instruments meet all criteria and none of the products have any parameters which exceed any performance threshold by a factor of 4 or more.

Figure 22 further synthesizes all results into one overview plot. This assessment matrix shows the outcome for all 36 instruments, eight data products for MAX-DOAS and four data products for zenith-sky mode. Any box coloured with green denotes that all three assessment criteria for that instrument and data product have been fulfilled. Boxes marked with yellow and orange denote that one or two criteria, respectively, have not been met, while red means that all three criteria have not been met and black indicates that this data set has at least one extreme outlier. Additionally, both the reported dSCD regression RMS and the DOAS fit RMS are used to sort the data products accordingly, with the smallest median RMS being assigned the lowest number in each case.

The order in which the instruments are displayed in Figure 22 is identical to Figures 20 and 21 with the instruments being grouped into five different categories: Custom-built, Pandora, EnviMes, miniDOAS and SAOZ. Custom-built instruments are assembled in-house and often designed with specific research purposes in mind. This category displays the greatest diversity in performance, and it includes the highest performing instruments as well as the instrumentation with the biggest difficulties meeting the set criteria of the performance assessment. In some cases, this can be related to the level of experience of the research group involved in building the instrument and/or in operating the instrument and performing the data analysis.

The first seven custom-built instruments listed in Figure 22 meet all criteria for all measured MAX-DOAS data sets with the following three instruments also being close to fulfilling almost all criteria for most of the data. The last six instruments listed under the custom-built category, however, struggle to either meet two criteria or to meet all criteria for one of the measured data products. Additionally, HCHO or O3uv data sets measured by three of the instruments (csic-10, bsu-5 and iiserm-16) contain extreme outliers.

The seven Pandora and five EnviMes instruments show overall a more consistent picture. Four of the Pandoras are meeting all categories and two of the other Pandoras satisfy all but one of the criteria for one or two of the data products. Nasa-31, however, experienced problems during operation and had some dirt inside the head sensor which was moving around and blocking part of the instrument FOV as well as having a loose tracker shaft. This caused significantly reduced signal-to-noise and an increased pointing uncertainty (see the large error bar for this instrument in Figure 6) that had negative consequences for all data products analyzed in the campaign. These problems were detected during the campaign and an attempt was made to fix them. In spite of these issues, most criteria were still met. It should be noted though that the behaviour displayed by nasa-31 did not fully represent the observational capabilities of a Pandora, as clearly evidenced by results from other instruments of the same type. The EnviMes instruments performed overall well when measuring NO_2 but struggled more to fulfil all criteria for the HCHO and O3uv data sets apart from niwa-29 which satisfied all criteria for HCHO and O3uv while not satisfying one of the criteria for both of the O_4 data sets and one of the NO_2 data sets. Most of the six miniDOAS instruments measured NO_2 satisfactorily in all three wavelength ranges

and only failed to satisfy one criterion in the O_4 data sets. However, they experienced discernible difficulties when measuring HCHO and O_3 which includes all criteria failed and extreme outliers.

The zenith-sky twilight data set (rightmost four columns in Figure 22) show a consistent performance for all custom-built, Pandora, EnviMes and SAOZ type instruments and all four data products (apart from nasa-31, see discussion above) with in most cases (90%) all criteria satisfied and in just eight cases one criterion not satisfied. The performance of the miniDOAS instruments is for the zenith-sky data more variable with one instrument (cma-8) not satisfying any of the criteria for O3vis and another (nust-33) failing two out of three criteria for the NO2uv product. The two SAOZ instruments measure zenith-sky data only and either satisfy all criteria or just do not meet one of them.

The ranking provided in each of the individual boxes in Figure 22 is based on the dSCD regression RMS (first value) and the RMS calculated as part of the data fitting routine (second value), the instruments with the smallest RMS (i.e. the smallest measurement noise) being assigned the lowest number. Overall, the combined ranking reflects the performance assessment of the individual instruments, but there are a couple of noteworthy deviations. For example, the data products measured by auth-3 have very large numbers corresponding to a high RMS (high measurement noise in comparison to other systems) but at the same time meet almost all performance criteria. On the other hand, the data products measured by aiofm-1 have an excellent fit RMS rating corresponding to a very low measurement noise, while none of the data products satisfy all criteria and O3uv has at least one extreme outlier. This apparent inconsistency reflects the nature of the performance assessment methodology, which puts larger emphasis on the assessment of systematic biases on measured dSCDs than on the noise. We have also seen that the comparison noise in regression analyses is for some of the products (NO_2, O_4) dominated by atmospheric/observation geometry effects rather than by actual instrumental noise.

The performance matrix shown in Figure 22 can be used to assess the participating groups and their instruments regarding their capability to measure NO_2, O_3 and HCHO concentrations and aerosols (using O_4 measurements) at sufficiently high quality to allow reliable geophysical studies or satellite validation efforts. Based on their RMS rating and the fact that they meet all the other criteria as well, the top most four instruments listed in Figure 22 and the Luftblick 26 and 27 systems can be considered as the best performing instruments during CINDI-2. In addition to offering an instantaneous picture of the level of performance of the current international MAX-DOAS research community, these results also provide the background information needed for the formal assessment and certification of instruments contributing to the NDACC network.

5   Recommendations for network operation and future campaigns

The CINDI-2 intercomparison exercise included more target trace gas species, and more instruments and participants from many different institutes than previously attempted in any other UV-visible spectroscopy intercomparison exercise. This provided a logistical challenge, which was addressed by setting up a carefully managed campaign. Beyond the detailed consistency assessment documented in this work, several lessons were learned that are expected to be of benefit to measurements conducted at network sites.

1. The accuracy and stability of the MAX-DOAS elevation scans was found to be critical, especially for measurements at low elevation angles. Therefore, we recommend to regularly calibrate elevation scan devices using one of the methods described in Donner et al. (in review, 2019). Moreover, for instruments not equipped with an internal pointing verification system (e.g. digital inclinometer or self-calibrating sun-tracker) horizon scans should be regularly performed, ideally on a daily basis, in order to verify the long-term stability of the pointing elevation.

2. The degree of geometric and temporal synchronisation prescribed for the instruments has revealed that spatial and temporal variability in the atmosphere is significantly greater than the total effect of instrument-derived uncertainties. As a result, atmospheric variability limits the reproducibility and representativeness of individual MAX-DOAS measurements for species such as NO_2. For this molecule, we estimate that the variability has a spatial scale that is at least as fine as many tens to a few hundreds of meters. This order of magnitude is consistent with the horizontal distances sampled by the average FOV (1 degree) and the horizontal separation of the instrument telescopes. It implies that random error estimates on NO_2 dSCDs should account for atmospheric variability effects in addition to spectral fitting uncertainties. To a lesser extent, the same reasoning applies to O_4 dSCD measurements.

3. For high-quality HCHO measurements, radiance measurements should reach a signal-to-noise ratio of 1000 or better in the spectral range from 335 nm to 360 nm, corresponding to HCHO dSCD uncertainties of $5x10^{15}$ molec/cm^2 or better. At this level of random uncertainty (and in contrast to the NO_2 case), HCHO spectral fitting errors still dominate over atmospheric variability effects.

One also anticipates that future similar intercalibration campaigns will strongly benefit from the lessons learned during and after CINDI-2. As already pointed out, the campaign was successful in improving (1) the spatial and temporal synchronicity of the measurements, and (2) the characterisation of the pointing elevation accuracy from all instruments and their impact on the DOAS analysis results. Despite these achievements, a few critical points were identified that deserve more attention in future deployments.

1. The data acquisition protocol, which proved to be very useful for instrument synchronisation, was not fully adequate for monitoring the spatial variability in highly variable trace gases such as NO_2. As discussed in Section 3.7, results from CINDI-2 indicate that in spite of the improvement in measuring the same airmass, the variability in some of the trace gases can still be large enough to introduce noise which is clearly exceeding the measurement uncertainty, suggesting that using DOAS fit errors as a measure of the dSCD error covariance is not appropriate. A more representative estimate of the random error should be derived instead from the measured variability of the observed dSCDs (see e.g. Bösch et al., 2018). For future campaigns, we hence recommend to adopt a strategy combining full elevation scans suitable for profile inversion at one or two reference azimuths, and azimuth scans at one elevation for evaluation of the spatial variability in trace gas concentration.

2. Although the campaign had a strong focus on elevation scan calibration, other aspects of the instrument calibration were handled with far less attention. Results from the data analysis, however, indicated that some of the observed discrepancies were related to a lack of proper instrumental characterisation before the campaign (e.g. detector non-linearity or spectral stray-light), and it is likely that some of the remaining deviations are related to unresolved calibration issues. For future campaigns, a better strategy should be developed to improve the characterization of participating instruments in preparation for field deployment. This could, e.g. be organised in the form of a preparatory calibration campaign hosted by a suitably equipped lab. The focus of this exercise should be put on instrumental characteristics of major importance for DOAS-type instruments, i.e. in particular instrumental line shape, spectral stray-light, polarization response, detector response (dark-current and linearity), field of view of telescope, elevation scanner accuracy and reproducibility, and instrument throughput and sensitivity.

6   Conclusions

The CINDI-2 intercomparison campaign had a strong focus on synchronisation and collocation of the measurements as well as on the determination of the pointing accuracy, which altogether resulted in a reduction of the impact of atmospheric changes on the

intercomparison exercise in comparison to CINDI. While each participating institute used their own instrumentation and analysis software (Tables 2 and 3), specific measurement procedures and retrieval settings were prescribed and strictly adhered to.

This comprehensive measurement protocol was highly successful in synchronising the timing of the measurements between all the instruments (Figure 2). The different approaches applied to determine the pointing accuracy of the instruments and their stability during the campaign provided important information for monitoring the instrument performance (see Figure 6). Moreover, this information was used to correct the data analysis in cases where the measurements were compromised by pointing inaccuracies leading to further improvements in consistency (see e.g. Figure 7). The horizon scans, in particular, were useful for identifying calibration biases, which could be addressed and corrected for the remainder of the campaign. Based on the experiences made during CINDI-2, it is highly recommended to include horizon scans into the daily measurement routine at monitoring sites and for any future MAX-DOAS intercomparison exercise. The different methods for the elevation calibration used during the CINDI-2 campaign are discussed in more detail in Donner et al. (in review, 2019).

In line with previous intercomparisons, a regression analysis of the dSCDs measured by each instrument with a reference data set was performed (see Section 3.5 for details on how the reference data sets were derived) and a whole range of correlation plots between the dSCDs and the reference were generated in a systematic manner (Figures 10-12 and Sections 1 and 2 of the Supplement). The slope and intercept of the regression analysis respectively quantify the mean systematic bias and offset of the individual data sets against the reference, and the regression RMS error provides an estimate of the overall comparison noise (see e.g. Figure 17). These three performance criteria were further investigated, and for each of the parameters and data products, specific limits were set and applied to all the measurements (Table 4 & Figure 18). Figures 19-22 visualize the summary of the regression analysis and provide an overview of the performance of each of the instruments regarding the eight MAX-DOAS and four zenith-sky data products.

The general level of agreement achieved for the different data products is summarized in Table 7. The median bias against the reference data sets is generally low (<5% for most products) and comparison noise levels are of the order of 3-4 10^{15} molec./cm^2 for NO$_2$, 1.5 10^{42} molec.^{2}/cm^5 for O$_4$ and 1.0 10^{16} molec./cm^2 for HCHO. The table also lists the typical dSCD retrieval uncertainties that can be expected from high quality and standard instruments, respectively. These uncertainties are compatible with satellite validation requirements (for further details, e.g. see https://sentinels.copernicus.eu/web/sentinel/technical-guides/sentinel-5p/validation). The results summarized in Table 7 agree well with the mean relative differences and standard deviation from the reference listed for all participating instruments in Tables 5 and 6, which also show that most instruments agree within a few percent for all MAX-DOAS and twilight DOAS products (apart from HCHO and O$_3$).

This assessment process, undertaken as part of the CINDI-2 intercomparison campaign, provides the UV-visible absorption spectroscopy research community with guidelines and a procedure on how to assess the performance of MAX-DOAS and DOAS instruments, in particular for the inclusion into NDACC (see NDACC webpage for access to the UV/Vis Appendix describing these recommendations). It is expected that a similar level of consistency, as seen during CINDI-2, can be obtained in the field if recommended settings are implemented and used by each participant of the network. More control in this aspect of homogeneity can be obtained through centralized processing, which is the aim of the currently developed ESA FRM4DOAS project (see http://frm4doas.aeronomie.be/).

The semi-blind CINDI-2 intercomparison exercise, presented here, concludes with the comparison and assessment of the retrieved dSCDs of a limited number of mature data products (NO$_2$, O$_4$, O$_3$ and HCHO). However, additional species (e.g. HONO, glyoxal,

BrO, H₂O) were also measured during the campaign, some of them being the subject of ongoing studies to be published separately. In particular, the tropospheric ozone column retrieval has been investigated in depth (Wang et al., 2018) and a publication on HONO retrievals is under way (Wang et al., submitted to AMT, 2019) as a follow-up of the first HONO intercomparison during the MAD-CAT campaign (Wang et al., 2017c). In addition to dSCD measurements, the subsequent steps in MAX-DOAS retrievals, i.e. their

5 conversion into vertical column and profile information is also further investigated in a CINDI-2 profiling working group and as part of the ESA FRM₄DOAS project (Frieß et al., 2018; Tirpitz et al., submitted to AMT, 2019). Furthermore, other aspects of the campaign measurements are being further exploited, such as mobile car-DOAS observations, reference in-situ measurements and instruments' elevation pointing calibration (Donner et al. in review, 2019).

10 *Data availability.* The CINDI-2 campaign data are currently referenced on the following web site https://evdc.esa.int/campaigns/other-campaigns/

Author contributions. FH, MVR, AA, UF, AR, TW, JL and AP designed, planned and organised the CINDI-2 intercomparison campaign. KK was the referee for the intercomparison during the campaign and for the follow-up data analysis. ED contributed as

15 the assistant referee. All co-authors contributed to the campaign either as participants and instrument operators and/or by performing the data analysis, data quality control and data submission. MVR performed the intercomparison data analysis. KK and MVR interpreted the results and wrote the manuscript with feed-back and contributions from all other co-authors.

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

Acknowledgements. We gratefully acknowledge the KNMI staff at Cabauw for their excellent technical and infrastructure support during the campaign. CINDI-2 received funding from the Netherlands Space Office (NSO). Funding for this study was provided by ESA through the CINDI-2 (ESA Contract No.4000118533/16/I-Sbo) and FRM₄DOAS (ESA Contract No. 4000118181/16/I-EF) projects and partly within the EU 7ᵗʰ Framework Programme QA₄ECV project (Grant Agreement no. 607405). The BOKU MAX-

25 DOAS instrument was funded and the participation of S.F. Schreier was supported by the Austrian Science Fund (FWF): I 2296-N29. The participation of the University of Toronto team was supported by the Canadian Space Agency (through the AVATARS project) and the Natural Sciences and Engineering Research Council (through the PAHA project). The instrument was primarily funded by the Canada Foundation for Innovation and is usually operated at the Polar Environment Atmospheric Research Laboratory (PEARL) by the Canadian Network for the Detection of Atmospheric Change (CANDAC). Funding for CISC was provided by the

30 UVAS ("Ultraviolet and Visible Atmospheric Sounder") project SEOSAT/INGENIO, ESP2015-71299-R, MINECO-FEDER, UE. The activities of the IUP Heidelberg were supported by the DFG project RAPSODI (grant No. PL 193/17-1). SAOZ and Mini-SAOZ instruments are supported by the Centre National de la Recherche Scientifique (CNRS) and the Centre National d'Etudes Spatiales (CNES). INTA acknowledges support from the National funding projects HELADO (CTM2013-41311-P) and AVATAR (CGL2014-55230-R). AMOIAP acknowledges support from the Russian Science Foundation (grant 16-17-10275) and Russian

35 Foundation for Basic Research (grants 16-05-01062 and 18-35-00682). K.L. Chan received trans-national access funding from ACTRIS-2 (H2020 Grant Agreement N° 654109). R. Volkamer acknowledges funding from NASA's Atmospheric Composition Program (NASA-16-NUP2016-0001), and US National Science Foundation (award AGS-1620530). H. Finkenzeller is recipient of a NASA graduate fellowship. M. Vrekoussis acknowledges support from the University of Bremen and the DFG Research Center/Cluster of Excellence "The Ocean in the Earth System-MARUM". Financial support through the University of Bremen

40 Institutional Strategy in the framework of the DFG Excellence Initiative is gratefully acknowledged for A. Schönhardt. Pandora instrument deployment was supported by Luftblick through ESA Pandonia Project and NASA Pandora Project at Goddard Space Flight Center under NASA Headquarters' Tropospheric Composition Program.

7   References

Apituley, A., et al.: Second Cabauw Intercomparison of Nitrogen Dioxide Measuring Instruments (CINDI-2) – Campaign Overview, to be submitted to AMT, 2020.

5   Bobrowski, N.: Volcanic Gas Studies by Multi Axis Differential Optical Absorption Spectroscopy, Ph.D. thesis, University of Heidelberg, Germany, 2005.

Bösch, T., Rozanov, V., Richter, A., Peters, E., Rozanov, A., Wittrock, F., Merlaud, A., Lampel, J., Schmitt, S., de Haij, M., Berkhout, S., Henzing, B., Apituley, A., den Hoed, M., Vonk, J., Tiefengraber, M., Müller, M., and Burrows, J. P.: BOREAS – a new MAX-DOAS profile retrieval algorithm for aerosols and trace gases, Atmos. Meas. Tech., 11, 6833-6859,
10   https://doi.org/10.5194/amt-11-6833-2018, 2018.

Borovski, A., Elokhov, A., Postylyakov, O., and Bruchkouski, I., Study of different operational modes of the IAP 2-port-DOAS instrument for investigation of atmospheric trace gases during CINDI-2 campaign, Proc. SPIE 10424, 104240Y, 2017a, doi: 10.1117/12.2278234.

Borovski, A., Postylyakov, O., Elokhov, A., and Bruchkovski, I., Study of different operational modes of the IAP 2-port-DOAS
15   instrument for atmospheric trace gases investigation during CINDI-2 campaign basing on residual noise analysis, Proc. SPIE 10466, 104662Z, 2017b, doi: 10.1117/12.2285798.

Brinksma, E. J., Pinardi, G., Volten, H., Braak, R., Richter, A., Schönhardt, A., Van Roozendael, M., Fayt, C., Hermans, C., Dirksen, R. J., Vlemmix, T., Berkhout, A. J. C., Swart, D. P. J., Oetjen, H., Wittrock, F., Wagner, T., Ibrahim, O. W., de Leeuw, G., Moerman, M., Curier, R. L., Celarier, E. A., Cede, A., Knap, W. H., Veefkind, J. P., Eskes, H. J., Allaart, M., Rothe, R.,
20   Piters, A. J. M., and Levelt, P. F.: The 2005 and 2006 DANDELIONS NO$_2$ and aerosol intercomparison campaigns, J. Geophys. Res., 113, D16S46, doi:10.1029/2007JD008808, 2008.

Clémer, K., Van Roozendael, M., Fayt, C., Hendrick, F., Hermans, C., Pinardi, G., Spurr, R., Wang, P., and De Mazière, M .: Multiple wavelength retrieval of tropospheric aerosol optical properties from MAXDOAS measurements in Beijing, Atmos. Meas. Tech., 3, 863-878, 2010.

25   Donner, S., Kuhn, J., Van Roozendael, M., Bais, A., Beirle, S., Bösch, T., Bognar, K., Bruchkousky, I., Chan, K. L., Drosoglou, T., Fayt, C., Frieß, U., Hendrick, F., Hermans, C., Jin, J., Li, A., Ma, J., Peters, E., Pinardi, G., Richter, A., Schreier, S. F., Seyler, A., Strong, K., Tirpitz, J.-L., Wang, Y., Xie, P., Xu, J., Zhao, X., and Wagner, T.: Evaluating different methods for elevation calibration of MAX-DOAS instruments during the CINDI-2 campaign, Atmos. Meas. Tech. Discuss., https://doi.org/10.5194/amt-2019-115, in review, 2019.

30   Fleischmann, O. C., Hartmann, M., Burrows, J. P., and Orphal, J.: New ultraviolet absorption cross-sections of BrO at atmospheric temperatures measured by time-windowing Fourier transform spectroscopy, Journal of Photochemistry and Photobiology A: Chemistry, 168, 117–132, 2004.

Frieß, U., Monks, P., Remedios, J., Rozanov, A., Sinreich, R., Wagner, T., and Platt, U.: MAXDOAS O$_4$ measurements: a new technique to derive information on atmospheric aerosols: 2. Modeling studies, J. Geophys. Res., 111, D14203,
35   doi:10.1029/2005JD006618, 2006.

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. Journal of Geophysical Research, 116, D00R04. https://doi.org/10.1029/2011JD015938, 2011.

Frieß, U., Klein Baltink, H., Beirle, S., Clémer, K., Hendrick, F., Henzing, B., Irie, H., de Leeuw, G., Li, A., Moerman, M. M., van Roozendael, M., Shaiganfar, R., Wagner, T., Wang, Y., Xie, P., Yilmaz, S., and Zieger, P.: Intercomparison of aerosol extinction profiles retrieved from MAX-DOAS measurements, Atmos. Meas. Tech., 9, 3205-3222, https://doi.org/10.5194/amt-9-3205-2016, 2016.

5   Frieß, U., Beirle, S., Alvarado Bonilla, L., Bösch, T., Friedrich, M. M., Hendrick, F., Piters, A., Richter, A., van Roozendael, M., Rozanov, V. V., Spinei, E., Tirpitz, J.-L., Vlemmix, T., Wagner, T., and Wang, Y.: Intercomparison of MAX-DOAS Vertical Profile Retrieval Algorithms: Studies using Synthetic Data, Atmos. Meas. Tech. Discuss., https://doi.org/10.5194/amt-2018-423, in review, 2018.

Gratsea, M., Vrekoussis, M., Richter, A., Wittrock, F., Schönhardt, A., Burrows, J., Kazadzis, S., N. Mihalopoulos, N., 10   Gerasopoulos, E., Slant column MAX-DOAS measurements of nitrogen dioxide, formaldehyde, glyoxal and oxygen dimer in the urban environment of Athens. Atmospheric Environment, 135, 118–131. https://doi.org/10.1016/j.atmosenv.2016.03.048, 2016.

Heckel, A., Richter, A., Tarsu, T., Wittrock, F., Hak, C., Pundt, I., Junkermann, W., and Burrows, J. P.: MAX-DOAS measurements of formaldehyde in the Po-Valley, Atmos. Chem. Phys., 5, 909-918, https://doi.org/10.5194/acp-5-909-2005, 2005.

15   Hendrick, F., Pommereau, J.-P., Goutail, F., Evans, R.D., Ionov, D., Pazmino, A., Kyrö, E., Held, G., Eriksen, P., Dorokhov, V., Gil, M., and Van Roozendael, M.: NDACC/SAOZ UV-visible total ozone measurements: improved retrieval and comparison with correlative ground-based and satellite Observations, Atmos. Chem. Phys., 11, 5975–5995, doi:10.5194/acp-11-5975-2011, 2011.

Hendrick, F., Müller, J.-F., Clémer, K., Wang, P., De Mazière, M., Fayt, C., Gielen, C., Hermans, C., Ma, J. Z., Pinardi, G., 20   Stavrakou, T., Vlemmix, T., and Van Roozendael, M.: Four years of ground-based MAX-DOAS observations of HONO and NO_2 in the Beijing area, Atmos. Chem. Phys., 14, 765–781, doi:10.5194/acp-14-765-2014, 2014.

Herman, J., A. Cede, E. Spinei, G. Mount, M. Tzortziou, and N. Abuhassan: NO_2 column amounts from ground-based Pandora and MFDOAS spectrometers using the direct-sun DOAS technique: Intercomparisons and application to OMI validation, J. Geophys. Res., 114, D13307, doi:10.1029/2009JD011848, 2009.

25   Herman, J., Spinei, E., Fried, A., Kim, J., Kim, J., Kim, W., Cede, A., Abuhassan, N., Segal-Rozenhaimer, M.: NO_2 and HCHO measurements in Korea from 2012 to 2016 from Pandora spectrometer instruments compared with OMI retrievals and with aircraft measurements during the KORUS-AQ campaign. Atmospheric Measurement Techniques, 11(8), 4583–4603, https://doi.org/10.5194/amt-11-4583-2018, 2018.

Hönninger G. and Platt, U.: Observations of BrO and its vertical distribution during surface ozone depletion at Alert, Atmos. 30   Environ., 36, 2481–2489, 2002.

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.

Irie, H., Takashima, H., Kanaya, Y., Boersma, K. F., Gast, L., Wittrock, F., Brunner, D., Zhou, Y., and Van Roozendael, M.: Eight-component retrievals from ground-based MAX-DOAS observations, Atmos. Meas. Tech., 4, 1027–1044, doi:10.5194/amt-4-35   1027-2011, 2011.

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.

Lee, C., Richter, A., Lee, H., Kim, Y. J., Burrows, J. P., Lee, Y. G., and Choi, B. C.: Impact of transport of sulfur dioxide from the Asian continent on the air quality over Korea during May 2005. Atmospheric Environment, 42(7), 1461–1475. 40   https://doi.org/10.1016/j.atmosenv.2007.11.006, 2008.

Lee, H., Irie, H., Kim, Y. J., Noh, Y., Lee, C., Kim, Y., and Chun, K. J.: Retrieval of Aerosol Extinction in the Lower Troposphere Based on UV MAX-DOAS Measurements. Aerosol Science and Technology, 43(5), 502–509. https://doi.org/10.1080/02786820902769691, 2009.

Liley, J., Johnston, P., McKenzie, R., Thomas, A., and Boyd, I.: Stratospheric NO_2 variations from a long time series at Lauder, New Zealand, J. Geophys. Res., 105, 11 633–11 640, 2000.

Meller, R. and Moortgat, G. K.: Temperature dependence of the absorption cross sections of formaldehyde between 223 and 323 K in the wavelength range 225–375 nm, Journal of Geophysical Research: Atmospheres, 105, 7089–7101 2156–2202, 2000.

Merlaud, A.: Development and use of compact instruments for tropospheric investigations based on optical spectroscopy from mobile platforms, Presses univ. de Louvain, 2013.

Merten, A., Tschritter, J., and Platt, U.: Design of differential optical absorption spectroscopy long-path telescopes based on fiber optics, Appl. Opt, 10, 738–754, http://ao.osa.org/abstract.cfm?URI=ao-50-5-738, 2011.

Noxon, J. F.: Nitrogen dioxide in the stratosphere and troposphere measured by ground-based absorption spectroscopy, Science,189, 547–549, 1975.

Ortega, I., Coburn, S., Berg, L.K., Lantz, K., Michalsky, J., Ferrare, R., Hair, J., Hostetler, C. and Volkamer, R.: The CU 2D-MAX-DOAS instrument - part 2: Raman Scattering Probability Measurements and Retrieval of Aerosol Optical Properties, Atmos. Meas. Techn., 9, 3893-3910, doi:10.5194/amt-2015-385, 2016.

Peters, E., Pinardi, G., Seyler, A., Richter, A., Wittrock, F., Bösch, T., Van Roozendael, M., Hendrick, F., Drosoglou, T., Bais, A. F., Kanaya, Y., Zhao, X., Strong, K., Lampel, J., Volkamer, R., Koenig, T., Ortega, I., Puentedura, O., Navarro-Comas, M., Gómez, L., Yela González, M., Piters, A., Remmers, J., Wang, Y., Wagner, T., Wang, S., Saiz-Lopez, A., García-Nieto, D., Cuevas, C. A., Benavent, N., Querel, R., Johnston, P., Postylyakov, O., Borovski, A., Elokhov, A., Bruchkouski, I., Liu, H., Liu, C., Hong, Q., Rivera, C., Grutter, M., Stremme, W., Khokhar, M. F., Khayyam, J., and Burrows, J. P.: Investigating differences in DOAS retrieval codes using MAD-CAT campaign data, Atmos. Meas. Tech., 10, 955-978, https://doi.org/10.5194/amt-10-955-2017, 2017.

Peters, E., Ostendorf, M., Bösch, T., Seyler, A., Schönhardt, A., Schreier, S.F., Henzing, J.S., Wittrock, F., Richter, A., Vrekoussis, M. and Burrows, J.P.: Full-azimuthal imaging-DOAS observations of NO_2 and O_4 during CINDI-2, Atmos. Meas. Tech., 12, 4171–4190, https://doi.org/10.5194/amt-12-4171-2019, 2019.

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., Piters, 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, https://doi.org/10.5194/amt-6-167-2013, 2013.

Piters, 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., Perner, D., and Patz, H. W.: Simultaneous Measurement of Atmospheric CH_2O, O_3, and NO_2 by Differential Optical Absorption, J. Geophys. Res., 84, 6329–6335, 1979.

Pommereau, J.-P. and Goutail, F.: O_3 and NO_2 ground-based measurements by visible spectrometry during Arctic winter and spring 1988, Geophys. Res. Lett., 15, 891–894, 1988.

Prados-Roman, C., Gómez-Martín, L., Puentedura, O., Navarro-Comas, M., Iglesias, J., de Migo, J. R., Pérez, M., Ochoa, H., Barlasina, M. E., Carbajal, G. and Yela., M.: Reactive bromine in the low troposphere of Antarctica: estimations at two research sites, Atmos. Chem. Phys., 18, 8549-8570, https://doi.org/10.5194/acp-18-8549-2018, 2018.

Puķīte, 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, https://doi.org/10.5194/amt-3-631-2010, 2010.

Richter, A., Eisinger, M., Ladstätter-Weißenmayer, A.,Wittrock, F., and Burrows, J. P.: DOAS zenith-sky observations: 2.Seasonal variations of BrO over Bremen (53°N) 1994–1995, J. Atmos.Chem., 32, 83–99, 1999.

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. B., 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 Tornkvist, K., Heese, B., Jones, R. L., Aliwell, S. R., Freshwater, R. A., Slant Column Measurements of O_3 and NO_2 During the NDSC Intercomparison of Zenith-Sky UV-Visible Spectrometers in June 1996. J. Atmos. Chem., 32, 281–314, 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., 5 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., Piters, A., and Kim, Y. J.: Intercomparison of slant column measurements of NO_2 and O_4 by MAX-DOAS and zenith-sky UV and visible spectrometers, Atmos. Meas. Tech., 3, 1629–1646, 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.: HITEMP, the high-temperature molecular spectroscopic database, J. Quant. Spectrosc. Ra., 111, 2139–2150, https://doi.org/10.1016/j.jqsrt.2010.05.001, 2010.

Serdyuchenko, A., Gorshelev, V., Weber, M., Chehade, W., and Burrows, J.: High spectral resolution ozone absorption cross-sections, Atmospheric Measurement Techniques, 7, 1867–8548, 2014.

Sinreich, R., Coburn, S., Dix, B., and Volkamer, R.: Ship-based detection of glyoxal over the remote tropical Pacific Ocean, Atmos. Chem. Phys., 10, 11359-11371, https://doi.org/10.5194/acp-10-11359-2010, 2010.

Sluis, W. W., Allaart, M. A. F., Piters, A. J. M., and Gast, L. F. L.: The development of a nitrogen dioxide sonde, Atmos. Meas. Tech., 3, 1753–1762, https://doi.org/10.5194/amt-3-1753-2010, 2010.

Solomon, S., A. L. Schmeltekopf, and R. W. Sanders, On the interpretation of zenith sky absorption-measurements, J. Geophys. Res., 92, 8311–8319, 1987.

Thalman, R. and Volkamer, R.: Temperature dependent absorption cross-sections of O 2–O 2 collision pairs between 340 and 630 nm and at atmospherically relevant pressure, Physical chemistry chemical physics, 15, 15 371–15 381, 2013.

Tirpitz, J.-L, Frieß, U., Hendrick, F., Alberti, C., Allaart, M., Apituley, A., Bais, A. Beirle, S., Berkhout, S., Bognar, K., Bösch, T., Bruchkouski, I., Cede, A., Chan, K.L., den Hoed, M., Donner, S., Drosoglou, T., Fayt, C., Friedrich, M.M., Frumau, A., Gast, L., Gielen, C., Gomez-Martín, L., Hao, N., Hensen, A., Henzing, B., Hermans, C., Jin, J., Kreher, K., Kuhn, J., Lampel, J., Li, A., Liu, C., Liu, H., Ma, J., Merlaud, A., Peters, E., Pinardi, G., Piters, A., Platt, U., Puentedura, O., Richter, A., Schmitt, S., Spinei, E., Stein Zweers, D., Strong, K., Swart, D., Tack, F., Tiefengraber, M., van der Hoff, R., van Roozendael, M., Vlemmix, T., Vonk, J., Wagner, T., Wang, Y., Wang, Z., Wenig, M., Wiegner, M., Wittrock, F., Xie, P., Xing, C., Xu, J., Yela, M., Zhang, C., and Zhao, X.: Intercomparison of MAX-DOAS vertical profile retrieval algorithms: studies on field data from the CINDI-2 campaign, submitted to AMT, 2019.

Vandaele, A. C., Hermans, C., Simon, P. C., Carleer, M., Colin, R., Fally, S., Merienne, M.-F., Jenouvrier, A., and Coquart, B.: Measurements of the NO$_2$ absorption cross-section from 42 000 cm-1 to 10 000 cm-1 (238–1000 nm) at 220 K and 294 K, Journal of Quantitative Spectroscopy and Radiative Transfer, 59, 171–184, 1998.

Vandaele, A. C., et al.: An intercomparison campaign of ground-based UV-visible measurements of NO$_2$, BrO, and OClO slant columns: Methods of analysis and results for NO$_2$, Journal of Geophysical Research, 110(D8), D08305, https://doi.org/10.1029/2004JD005423, 2005.

Vlemmix, T., Hendrick, F., Pinardi, G., De Smedt, I., Fayt, C., Hermans, C., Piters, A., Wang, P., Levelt, P., and Van Roozendael, M.: MAX-DOAS observations of aerosols, formaldehyde and nitrogen dioxide in the Beijing area: comparison of two profile retrieval approaches, Atmos. Meas. Tech., 8, 941–963, doi:10.5194/amt-8-941-2015, 2015.

Volten, H., E. J. Brinksma, A. J. C. Berkhout, J. Hains, J. B. Bergwerff, G. R. Van der Hoff, A. Apituley, R. J. Dirksen, S. Calabretta-Jongen, and D. P. J. Swart, NO$_2$ lidar profile measurements for satellite interpretation and validation, J. Geophys. Res., 114, D24301, doi:10.1029/2009JD012441, 2009.

Wagner, T., Dix, B., von Friedeburg, C., Frieß, U., Sanghavi, S., Sinreich, R., and Platt, U.: MAX-DOAS O$_4$ measurements: a new technique to derive information on atmospheric aerosols – principles and information content, J. Geophys. Res., 109, D22205, doi:10.1029/2004JD004904, 2004.

Wang, Y., Beirle, S., Lampel, J., Koukouli, M., De Smedt, I., Theys, N., Li, A., Wu, D., Xie, P., Liu, C., Van Roozendael, M., Stavrakou, T., Müller, J.-F., and Wagner, T.: Validation of OMI, GOME-2A and GOME-2B tropospheric NO$_2$, SO$_2$ and HCHO products using MAX-DOAS observations from 2011 to 2014 in Wuxi, China: investigation of the effects of priori profiles and aerosols on the satellite products, Atmos. Chem. Phys., 17, 5007-5033, https://doi.org/10.5194/acp-17-5007-2017, 2017a.

Wang, Y., Lampel, J., Xie, P., Beirle, S., Li, A., Wu, D., and Wagner, T.: Ground-based MAX-DOAS observations of tropospheric aerosols, NO$_2$, SO$_2$ and HCHO in Wuxi, China, from 2011 to 2014, Atmos. Chem. Phys., 17, 2189-2215, https://doi.org/10.5194/acp-17-2189-2017, 2017b.

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., Puentedura, O., Van Roozendael, M., 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, sensitivity studies on spectral analysis settings, and error budget, Atmos. Meas. Tech., 10, 3719-3742, https://doi.org/10.5194/amt-10-3719-2017, 2017c.

Wang, Y., J. Puķīte, T. Wagner, S. Donner, S. Beirle, A. Hilboll, M. Vrekoussis, A. Richter, A. Apituley, A. Piters, M. Allaart, H. Eskes, A. Frumau, M. Van Roozendael, J. Lampel, U. Platt, S. Schmitt, D. Swart, and J. Vonk, Vertical profiles of tropospheric

ozone from MAX-DOAS measurements during the CINDI-2 campaign: Part 1— Development of a new retrieval algorithm. Journal of Geophysical Research: Atmospheres, 123. https://doi. org/10.1029/2018JD028647, 2018.

Wang, Y., Apituley, A., Bais, A., Beirle, S., Benavent, N., Borovski, A., Bruchkouski, I., Chan, K.L., Donner, S., Drosoglou, T., Finkenzeller, H., Friedrich, M.M., Frieß, U., Garcia-Nieto, D., Gómez-Martín, L., Hendrick, F., Hilboll, A., Jin, J., Johnston, P., Koenig, T.K., Kreher, K., Kumar, V., Kyuberis, A., Lampel, J., Liu, C., Liu, H., Ma, J., Polyansky, O.L., Postylyakov, O., Querel, R., Saiz-López, A., Schmitt, S., Tian, X., Tirpitz, J.L., Van Roozendael, M., Volkamer, R., Wang, Z., Xie, P., Xing, C., Xu, J., Yela, M., Zhang, C., and Wagner, T.: Inter-comparison of MAX-DOAS measurements of tropospheric HONO slant column densities and vertical profiles during the CINDI-2 Campaign, submitted to Atmos. Meas. Tech., 2019.

Wittrock, F., Oetjen, H., Richter, A., Fietkau, S., Medeke, T., Rozanov, A., and Burrows, J. P.: MAX-DOAS measurements of atmospheric trace gases in Ny-Alesund – Radiative transfer ˚ studies and their application, Atmos. Chem. Phys., 4, 955–966, doi:10.5194/acp-4-955-2004, 2004.

Yela, M., Gil-Ojeda, M., Navarro-Comas, M., Gonzalez-Bartolomé, D., Puentedura, O., Funke, B., Iglesias, J., Rodríguez, S., García, O., Ochoa, H., and Deferrari, G.: Hemispheric asymmetry in stratospheric NO_2 trends, Atmos. Chem. Phys., 17, 13373-13389, https://doi.org/10.5194/acp-17-13373-2017, 2017.

Zieger, P., Weingartner, E., Henzing, J., Moerman, M., de Leeuw, G., Mikkilä, J., Ehn, M., Petäjä, T., Clémer, K., van Roozendael, M., Yilmaz, S., Frieß, U., Irie, H., Wagner, T., Shaiganfar, R., Beirle, S., Apituley, A., Wilson, K., and Baltensperger, U.: Comparison of ambient aerosol extinction coefficients obtained from in-situ, MAX-DOAS and LIDAR measurements at Cabauw, Atmos. Chem. Phys., 11, 2603–2624, doi:10.5194/acp-11-2603-2011, 2011.

Table 1: List of participating groups and corresponding instrument IDs in alphabetical order according to their acronym.

Institute	Country	Acronym	Instrument ID
Anhui Institute of Optics and Fine Mechanics	China	AIOFM	aiofm-1
A. M. Obukhov Institute of Atmospheric Physics, Russian Academy of Sciences, Moscow	Russia	AMOIAP	amoiap-2
Aristotle University of Thessaloniki	Greece	AUTH	auth-3
Royal Belgian Institute for Space Aeronomy	Belgium	BIRA-IASB	bira-4
University of Natural Resources and Life Sciences, Vienna	Austria	BOKU	boku-6
Belarusian State University	Belarus	BSU	bsu-5
Chiba University	Japan	CHIBA	chiba-9
China Meteorological Administration	China	CMA	cma-7, cma-8
Spanish National Research Council	Spain	CSIC	csic-10
University of Colorado	USA	CU-Boulder	cu-boulder-11, cu-boulder-12
Deutsches Zentrum für Luft- und Raumfahrt/ University of Science and Technology of China	Germany/ China	DLR-USTC	dlrustc-13, dlrustc-14
Indian Institute of Science Education and Research Mohali	India	IISERM	iiserm-16
National Institute for Aerospace Technology	Spain	INTA	inta-17
University of Bremen	Germany	IUP-Bremen	iupb-18, iupb-37
University of Heidelberg	Germany	IUP-Heidelberg	iuph-19
Royal Netherlands Meteorological Institute	The Netherlands	KNMI	knmi-21, knmi-22, knmi-23
Laboratoire Atmosphères, Milieux, Observations Spatiales	France	LATMOS	latmos-24, latmos-25
Ludwig-Maximilians-Universität München	Germany	LMU-MIM	lmumim-35
LuftBlick Earth Observation Technologies	Austria	Luftblick	luftblick-26, luftblick-27, luftblick-260, luftblick-270
Max Planck Institute for Chemistry, Mainz	Germany	MPIC	mpic-28
NASA Goddard Space Flight Center	USA	NASA	nasa-31, nasa-32
National Institute of Water and Atmospheric Research	New Zealand	NIWA	niwa-29, niwa-30
National University of Sciences and Technology	Pakistan	NUST	nust-33
University of Toronto	Canada	UTO	uto-36

Table 2: Overview of the main characteristics of the instruments taking part in the semi-blind intercomparison campaign. The table lists the type, specific ID and model name for each participating instrument (columns 1-3). In columns 4-5, it also specifies whether instruments could take azimuthal scans (ASc) and/or be operated in direct-sun mode (DS). The spectral range, spectral resolution and field of view (FOV) are summarized in columns 6-8. Note that the FOV given in column 8 is the value provided as part of the instrument specification which may differ from the effective FOV shown in Figure 6. Light coupling (column 9) denotes whether spectrometers were fed by means of optical fibers (F) or using a telescope or lens directly coupled to the entrance slit (D). The detector type is specified in column 10 either as charge-coupled device (CCD) or linear array (LinArr) and the detector temperature is listed in column 11.

Instrument type	Instrument ID	Instrument name	ASc	DS	Spectral range (nm)	Spectral res. (nm)	FOV (°)	Light coupl.	Detector type	Detector T (°C)
Custom-built MAX-DOAS	bira-4	2D MAX-DOAS	y	y	300-390/ 400-560	0.37/ 0.58	1.0/ 0.5	F	CCD	-50/ -50
	iupb-18	2D MAX-DOAS	y	n	305-390/ 406-579	0.5/ 0.85	1.0	F	CCD	-35/ -30
	boku-6	2D MAX-DOAS	y	n	419-553	0.8	0.8	F	CCD	-60
	cu-boulder-11	2D MAX-DOAS	y	y	325-470/ 430-680	0.7/ 1.2	0.7	F	CCD	-30
	cu-boulder-12	1D MAX-DOAS	n	n	300-465/ 380-490	0.8/ 0.5	0.7	F	CCD	-30/ 0
	inta-17	RASAS-III	y	n	420-540	0.55	1.0	F	CCD	-30
	mpic-28	Tube MAX-DOAS	n	n	305-464	0.6	0.7	F	CCD	20
	niwa-30	ACTON275 MAX-DOAS	n	n	290-363/ 400-460	0.54	0.5	F	CCD	-20
	uto-36	2D MAX-DOAS	y	y	340-560	0.75	0.62	F	CCD	-70
	auth-3	Phaethon	y	y	300-450	0.4	1.0	F	CCD	5
	aiofm-1	2D MAX-DOAS	y	n	290-380	0.4	0.2	F	CCD	-30
	chiba-9	CHIBA-U MAX-DOAS	n	n	310-515	0.4	<1	F	CCD	ambient T
	csic-10	1D MAX-DOAS	n	n	300-500	0.5	0.7	F	CCD	-70
	amoiap-2	2-port DOAS	n	n	315-385/ 395-465/ 420-490	0.4	0.3	F	CCD	-40
	bsu-5	MARSB	n	n	300-500	0.4	0.2-1.0	D	CCD	-40
	iupb-37	Imaging-DOAS	y	n	420-500	0.8	1.2	F	CCD	-30
Pandora	knmi-23	Pandora-1S	y	y	290-530	0.6	1.5	F	CCD	20
	luftblick-26	Pandora-2S	y	y	280-540	0.6	1.5	F	CCD	20
	luftblick-260	Pandora-2S	y	y	380-900	1.1	1.5	F	CCD	20
	luftblick-27	Pandora-2S	y	y	280-540	0.6	1.5	F	CCD	20
	luftblick-270	Pandora-2S	y	y	380-900	1.1	1.5	F	CCD	20
	nasa-31	Pandora-1S	y	y	280-540	0.6	1.6	F	CCD	20
	nasa-32	Pandora-1S	y	y	280-540	0.6	1.6	F	CCD	20
EnviMes	iuph-19	2D EnviMes	y	y	300-460/ 440-580	0.6/ 0.5	<0.5	F	CCD	20

	dlrustc-13	1D EnviMes	n	n	300-460/ 450-600	0.6/ 0.6	0.4	F	CCD	20
	dlrustc-14	1D EnviMes	n	n	300-460/ 450-600	0.6/ 0.6	0.4	F	CCD	20
	niwa-29	1D EnviMes	n	n	305-460/ 410-550	0.6	<0.5	F	CCD	20
	lmumim-35	2D EnviMes	y	n	300-460/ 450-600	0.6/ 0.9	0.4	F	CCD	20
Mini-DOAS	cma-7	Mini-DOAS-UV	n	n	300-450	0.7	0.8	F	LinArr	ambient T
Hoffmann GmbH	cma-8	Mini-DOAS-Vis	n	n	400-710	1.6	0.8	F	LinArr	ambient T
	iiserm-16	Mini-DOAS-UV	n	n	316-466	0.7	0.7	F	CCD	-10.4
	knmi-21	Mini-DOAS-UV	n	n	290-443	0.6	0.45	F	LinArr	-5
	knmi-22	Mini-DOAS-Vis	n	n	400-600	0.5	0.4	F	LinArr	-5
	nust-33	Mini-DOAS-UV	n	n	320-465	0.7	1.2	F	CCD	ambient T
SAOZ	latmos-24	SAOZ	n	n	270-640	1.3	20	D	LinArr	ambient T
	latmos-25	Mini-SAOZ	n	n	270-820	0.7	8	F	CCD	ambient T

Table 3: Overview of analysis software used by each of the participating institutes.

Data analysis software	Institute acronym
QDOAS	AUTH, BIRA-IASB, CSIC, CU-Boulder, LMU-MIM
QDOAS & WinDOAS	AIOFM, NUST
QDOAS & in-house developed software	UTO
DOASIS	DLR-USTC, IUP-Heidelberg
DOASIS & WinDOAS	IISERM,
DOASIS & in-house developed software (STRATO)	NIWA
WinDOAS	CMA, MPIC
WinDOAS & in-house developed software	BSU
Blick Software Suite	LuftBlick, NASA
Blick Software Suite & in-house developed software	KNMI
NLIN	BOKU, IUP-Bremen
LANA	INTA
SAOZ SAM v5.9 & Mini SAOZ in-house developed software	LATMOS
JM2 (Japanese MAX-DOAS profile retrieval algorithm, version 2)	CHIBA
Andor Solis & in-house developed software	AMOIAP

Table 4: Data products included in the semi-blind intercomparison exercise and wavelength intervals selected for the analysis. Performance limits on bias (deviation from unity slope), offset and RMS of dSCD linear regressions are also listed for each of the eight data products.

Data product	Spectral interval (nm)	Bias (%)	Offset (molec/cm²)	RMS (molec/cm²)
NO2vis	425 – 490	5	$1.5 \ 10^{15}$	$8.0 \ 10^{15}$
NO2visSmall	411 – 445	5	$1.5 \ 10^{15}$	$8.0 \ 10^{15}$
NO2uv	338 – 370	6	$2.0 \ 10^{15}$	$1.0 \ 10^{16}$
O4vis*	425 – 490	5	$0.7 \ 10^{42}$	$3.0 \ 10^{42}$
O4uv*	338 – 370	6	$0.8 \ 10^{42}$	$3.0 \ 10^{42}$
HCHO	336.5 – 359	10	$5.0 \ 10^{15}$	$1.0 \ 10^{16}$
O3vis	450 – 520	4	$0.2 \ 10^{18}$	$1.0 \ 10^{18}$
O3uv	320 – 340	4	$1.0 \ 10^{18}$	$4.0 \ 10^{18}$

*Note: the units for O_4 are $molec^2/cm^5$

Table 5: Mean relative difference from the reference and standard deviation (in percent) for all participating instruments and MAX-DOAS data products (apart from ozone). The last column provides the values for HCHO when only considering measurements made during the first 4 days of the campaign period (12-15 Sep. 2016).

Instrument ID	NO2vis	NO2visSmall	NO2uv	O4vis	O4uv	HCHO	HCHO (12-15/09)
bira-4	-0.0 (2.0)	0.7 (2.0)	1.7 (2.1)	0.6 (2.0)	1.0 (1.7)	5.2 (6.9)	1.0 (2.9)
iupb-18	-2.2 (2.7)	-1.2 (2.4)	0.1 (2.2)	-0.7 (2.2)	-1.2 (2.5)	-2.9 (6.4)	0.0 (3.6)
boku-6	0.7 (2.6)	--	--	0.3 (2.0)	--	--	--
cu-boulder-11	0.9 (4.9)	-1.8 (4.3)	-3.7 (5.1)	-0.7 (3.2)	-0.4 (3.3)	-19.9 (32.0)	-7.1 (11.7)
cu-boulder-12	-3.9 (1.5)	-0.6 (1.6)	-0.6 (2.9)	-0.7 (1.6)	-0.2 (4.7)	--	--
inta-17	0.7 (2.6)	--	--	-0.2 (2.6)	--	--	--
mpic-28	--	1.4 (2.1)	3.4 (3.3)	--	0.9 (2.2)	-0.2 (14.5)	-4.0 (5.4)
niwa-30	-2.6 (2.3)	--	-0.2 (10.0)	-0.1 (2.5)	1.1 (6.5)	-24.5 (36.1)	-11.5 (7.7)
uto-36	-6.4 (3.2)	-5.0 (3.1)	--	-3.6 (3.1)	--	--	--
auth-3	--	-2.4 (3.4)	-3.4 (8.2)	--	0.5 (8.5)	7.9 (62.1)	16.3 (26.3)
aiofm-1	--	--	-15.8 (5.3)	--	-7.3 (5.1)	18.2 (54.7)	-0.2 (16.3)
chiba-9	-2.3 (3.4)	-1.3 (3.6)	1.0 (4.0)	6.5 (6.8)	10.6 (4.1)	0.1 (24.0)	-2.6 (13.3)
csic-10	--	--	-17.7 (12.5)	--	0.5 (8.4)	-131.5 (164.8)	--
amoiap-2	-7.3 (3.3)	-7.9 (3.2)	-6.3 (9.9)	-0.8 (8.5)	-10.7 (8.0)	-70.5 (80.0)	-31.7 (12.1)
bsu-5	--	--	-6.5 (6.5)	--	-5.0 (5.1)	33.3 (90.5)	13.2 (22.9)
iupb-37	3.3 (6.8)	--	--	-4.2 (7.0)	--	--	--
knmi-23	1.9 (2.3)	2.8 (2.3)	3.3 (6.8)	1.3 (1.5)	4.2 (4.2)	-12.3 (47.1)	-12.2 (17.9)
luftblick-26	-0.4 (1.4)	-0.4 (1.3)	0.6 (2.6)	-0.0 (1.3)	0.6 (3.0)	-17.6 (32.5)	-11.9 (16.7)
luftblick-260	3.4 (2.1)	2.8 (2.3)	--	-0.3 (1.5)	--	--	--
luftblick-27	-1.3 (1.8)	-1.0 (1.6)	-0.5 (2.8)	0.8 (1.4)	-1.0 (2.7)	-12.6 (28.0)	-9.0 (13.4)
luftblick-270	-0.5 (1.7)	0.7 (2.0)	--	-0.6 (1.3)	--	--	--
nasa-31	1.1 (6.2)	1.0 (5.9)	1.2 (5.7)	-0.1 (4.2)	-1.0 (5.1)	-21.5 (38.0)	-11.4 (15.7)
nasa-32	0.5 (1.7)	0.2 (1.7)	-0.2 (3.0)	1.0 (1.5)	-0.5 (3.1)	-10.6 (30.6)	-7.4 (9.6)
iuph-19	--	-2.1 (3.0)	-1.0 (3.2)	--	-1.2 (3.0)	-32.1 (28.8)	-14.2 (7.9)
dlrustc-13	-3.9 (3.7)	-3.1 (3.5)	-4.2 (3.8)	-3.1 (2.4)	0.8 (2.3)	-42.6 (42.0)	-14.1 (8.1)
dlrustc-14	-1.3 (3.0)	-0.4 (2.7)	-0.1 (2.7)	-1.5 (2.3)	1.7 (2.0)	-57.5 (60.0)	-17.7 (9.9)
niwa-29	-6.5 (12.0)	-5.1 (13.3)	-4.0 (14.8)	-0.2 (4.0)	3.8 (6.2)	-10.5 (15.8)	--
lmumim-35	2.1 (4.4)	1.2 (4.1)	-0.4 (3.7)	7.1 (7.8)	-3.9 (3.0)	-9.0 (22.5)	-8.5 (8.3)
cma-7	--	-1.5 (5.4)	-2.1 (5.4)	--	1.7 (5.4)	-26.2 (35.5)	-20.7 (13.8)
cma-8	-4.0 (4.1)	--	--	0.7 (7.8)	--	--	--
iiserm-16	--	1.2 (5.0)	-0.1 (8.8)	--	8.7 (7.0)	-111.5 (80.1)	-59.1 (24.1)
knmi-21	--	--	-4.6 (5.0)	--	2.7 (4.4)	4.9 (60.0)	0.4 (17.6)
knmi-22	-1.5 (4.9)	--	--	-2.5 (4.6)	--	--	--
nust-33	--	6.7 (6.1)	4.3 (9.2)	--	-22.6 (6.8)	48.3 (73.7)	--
latmos-24	--	--	--	--	--	--	--
latmos-25	--	--	--	--	--	--	--
Median from all instruments	**-0.9 (2.8)**	**-0.5 (3.1)**	**-0.4 (5.1)**	**-0.2 (2.4)**	**0.5 (4.3)**	**-12.3 (36.1)**	**-8.7 (12.7)**

Table 6: Mean relative difference from the reference and standard deviation (in percent) for all participating instruments and zenith-sky DOAS data products.

Instrument ID	NO2vis	NO2visSmall	NO2uv	O3vis
bira-4	0.4 (1.0)	0.5 (1.2)	0.9 (2.4)	0.2 (1.0)
iupb-18	0.8 (1.1)	0.8 (1.4)	4.1 (3.0)	0.2 (0.4)
boku-6	2.0 (1.0)	--	--	0.7 (0.7)
cu-boulder-11	3.3 (2.7)	1.3 (2.4)	-3.6 (7.8)	0.5 (1.1)
cu-boulder-12	-0.6 (2.2)	-0.2 (3.1)	-16.5 (21.5)	--
inta-17	1.4 (1.6)	--	--	-0.5 (0.7)
mpic-28	--	0.5 (3.1)	6.3 (6.1)	--
niwa-30	-0.1 (2.8)	--	1.7 (14.4)	--
uto-36	-1.0 (3.4)	-1.6 (2.8)	--	-6.7 (2.4)
auth-3	--	2.1 (3.6)	1.8 (16.5)	--
aiofm-1	--	--	-1.7 (17.5)	--
chiba-9	1.0 (6.0)	5.3 (6.3)	3.2 (16.1)	--
csic-10	--	--	-14.3 (28.1)	--
amoiap-2	0.9 (3.1)	0.0 (3.1)	13.9 (9.3)	--
bsu-5	--	--	1.8 (10.7)	--
iupb-37	4.8 (10.2)	--	--	--
knmi-23	0.3 (1.8)	1.4 (1.7)	2.6 (12.7)	-0.5 (1.4)
luftblick-26	-1.4 (1.5)	-0.2 (1.4)	-0.5 (4.5)	-1.3 (0.7)
luftblick-260	0.5 (1.2)	-0.5 (2.8)	--	-2.6 (5.1)
luftblick-27	-1.7 (1.5)	-1.7 (1.8)	-2.7 (4.7)	-0.4 (0.7)
luftblick-270	-2.5 (1.5)	-1.0 (3.5)	--	-2.5 (5.1)
nasa-31	-1.9 (2.3)	-0.3 (2.3)	6.3 (14.0)	-2.3 (1.3)
nasa-32	-1.1 (1.9)	-0.9 (2.0)	-2.6 (6.0)	-1.3 (0.9)
iuph-19	--	-1.1 (1.7)	-0.2 (4.4)	--
dlrustc-13	-0.8 (2.1)	0.4 (3.0)	-2.2 (5.2)	0.5 (1.7)
dlrustc-14	-2.6 (2.0)	-0.9 (2.1)	-1.6 (4.5)	-5.7 (2.3)
niwa-29	1.3 (6.0)	1.8 (7.9)	-5.2 (8.2)	-0.0 (3.0)
lmumim-35	-3.9 (3.2)	-5.0 (1.9)	-3.8 (4.2)	1.1 (9.7)
cma-7	--	-2.4 (5.1)	1.9 (8.5)	--
cma-8	-2.1 (3.7)	--	--	11.6 (7.4)
iiserm-16	--	4.0 (2.7)	5.6 (13.1)	--
knmi-21	--	--	-19.3 (12.4)	--
knmi-22	1.0 (4.8)	--	--	--
nust-33	--	9.0 (3.7)	22.1 (11.8)	--
latmos-24	-9.2 (6.1)	--	--	3.1 (2.7)
latmos-25	-2.5 (3.7)	--	--	1.0 (1.8)
Median from all instruments	**-0.4 (2.3)**	**-0.1 (2.7)**	**0.3 (8.9)**	**-0.2 (1.5)**

Table 7: Summary of the level of agreement obtained for dSCD measurements during CINDI-2 and typical uncertainties achieved by high quality and standard instruments for the different data products.

| Data product | Median agreement level between instruments | | Median dSCD fit error (molec/cm^2) | |
	Bias (%)	RMS (molec/cm^2)	High quality instruments	Standard instruments
NO2vis	3	$3 \ 10^{15}$	$2 \ 10^{14}$	$7 \ 10^{14}$
NO2visSmall	3	$3.5 \ 10^{15}$	$2 \ 10^{14}$	$5 \ 10^{14}$
NO2uv	3	$4 \ 10^{15}$	$6 \ 10^{14}$	$1.6 \ 10^{15}$
O4vis[*]	2	$1.5 \ 10^{42}$	$1.5 \ 10^{41}$	$3 \ 10^{41}$
O4uv[*]	2	$1.5 \ 10^{42}$	$3 \ 10^{41}$	$8 \ 10^{41}$
HCHO	8	$1 \ 10^{16}$	$3 \ 10^{15}$	$8 \ 10^{15}$
O3vis	2	$6 \ 10^{17}$	$3 \ 10^{17}$	$3 \ 10^{17}$
O3uv	4	$1.6 \ 10^{17}$	$1.3 \ 10^{16}$	$6 \ 10^{16}$

[*] Note: the units for O_4 are molec^2/cm^5

Figure 1: Picture of the CINDI-2 container layout at the main campaign site showing the organisation of the MAX-DOAS instruments on two superposed rows of mobile units (similar to shipping containers).
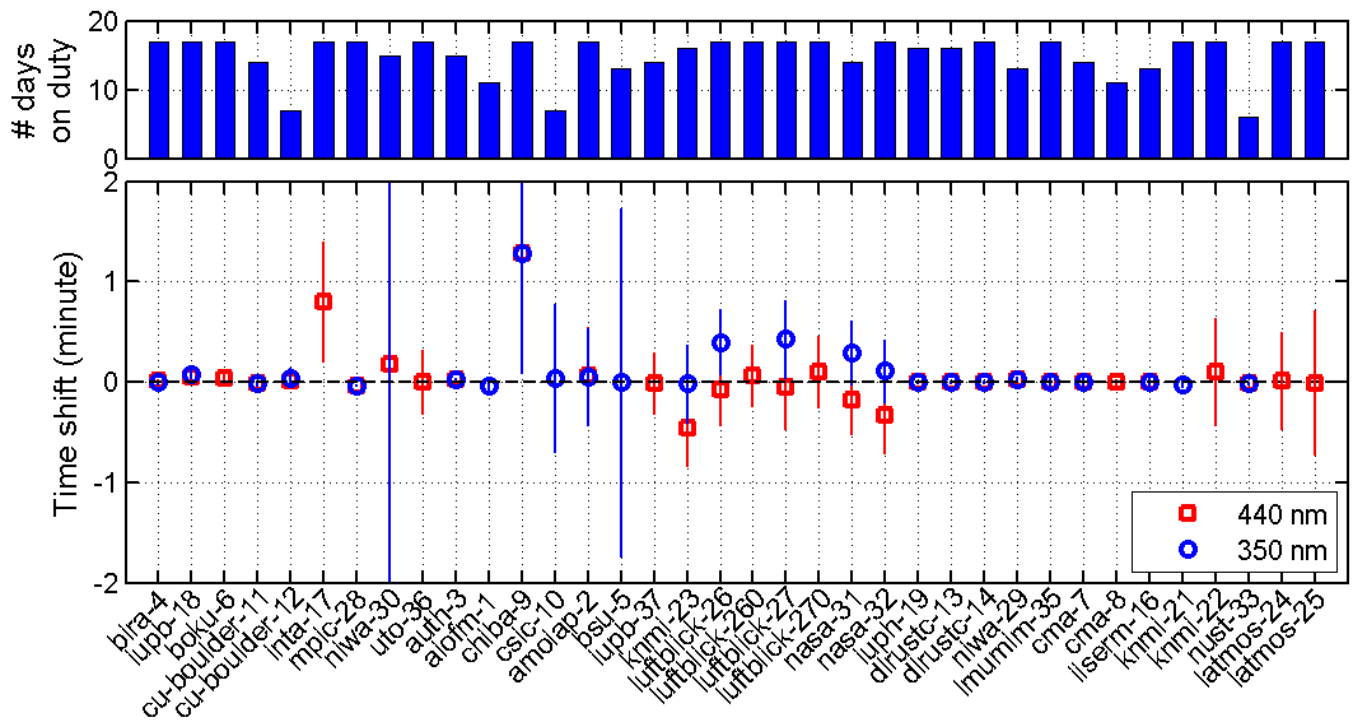
Figure 2: Top panel: The number of days when instruments were on duty during the 17-day intercomparison period. **Bottom Panel:** The mean and standard deviation of the time deviations (in decimal minutes) observed in the MAX-DOAS measurements as reported by each participating group with respect to the measurement schedule defined for the campaign. Note that the instruments are listed in order of how they are categorised, and this is further explained in Section 4.
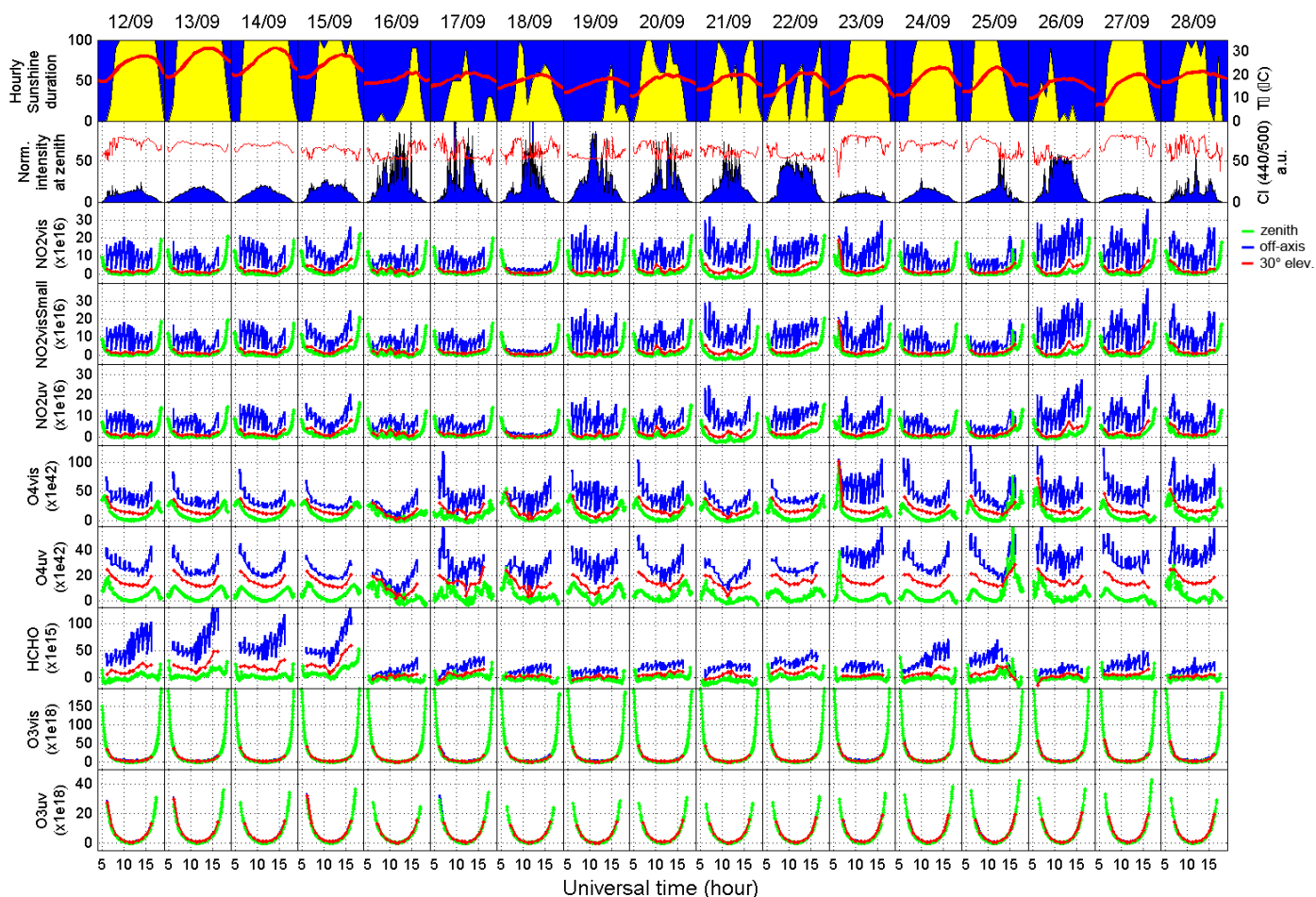
Figure 3: Hourly sunshine duration (yellow area) and temperature at the surface (red line) during the intensive campaign (topmost panel), the intensity measured in the zenith and the colour index (2ⁿᵈ panel from top), and the variability of the various trace gas slant column measurements performed during the semi-blind intercomparison exercise (all other panels). Slant column data measured at the main azimuth viewing direction (287°) with the IUP Bremen instrument (iupb-18) are shown. Green lines and symbols represent zenith-sky measurements, red lines and symbols off-axis data at 30° elevation, and blue lines off-axis measurements up to 15° elevation.
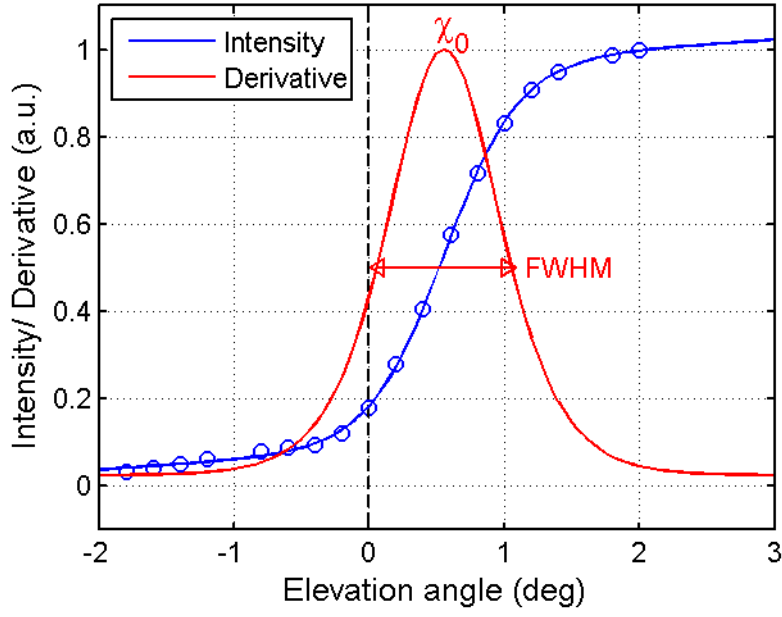
Figure 4: Horizon scans measured by IUP-Bremen on 14 September 2016 in the visible wavelength range. The blue circles display the intensity at 440 nm plotted as a function of the elevation angle reported by the instrument. Measured points are fitted by least-squares minimisation using an error function (blue line) allowing to estimate the horizon elevation (χ_0) and effective field of view (FWHM) (see Section 3.2). The corresponding Gaussian curve (analytical derivative of the fitted blue curve) is represented in red.
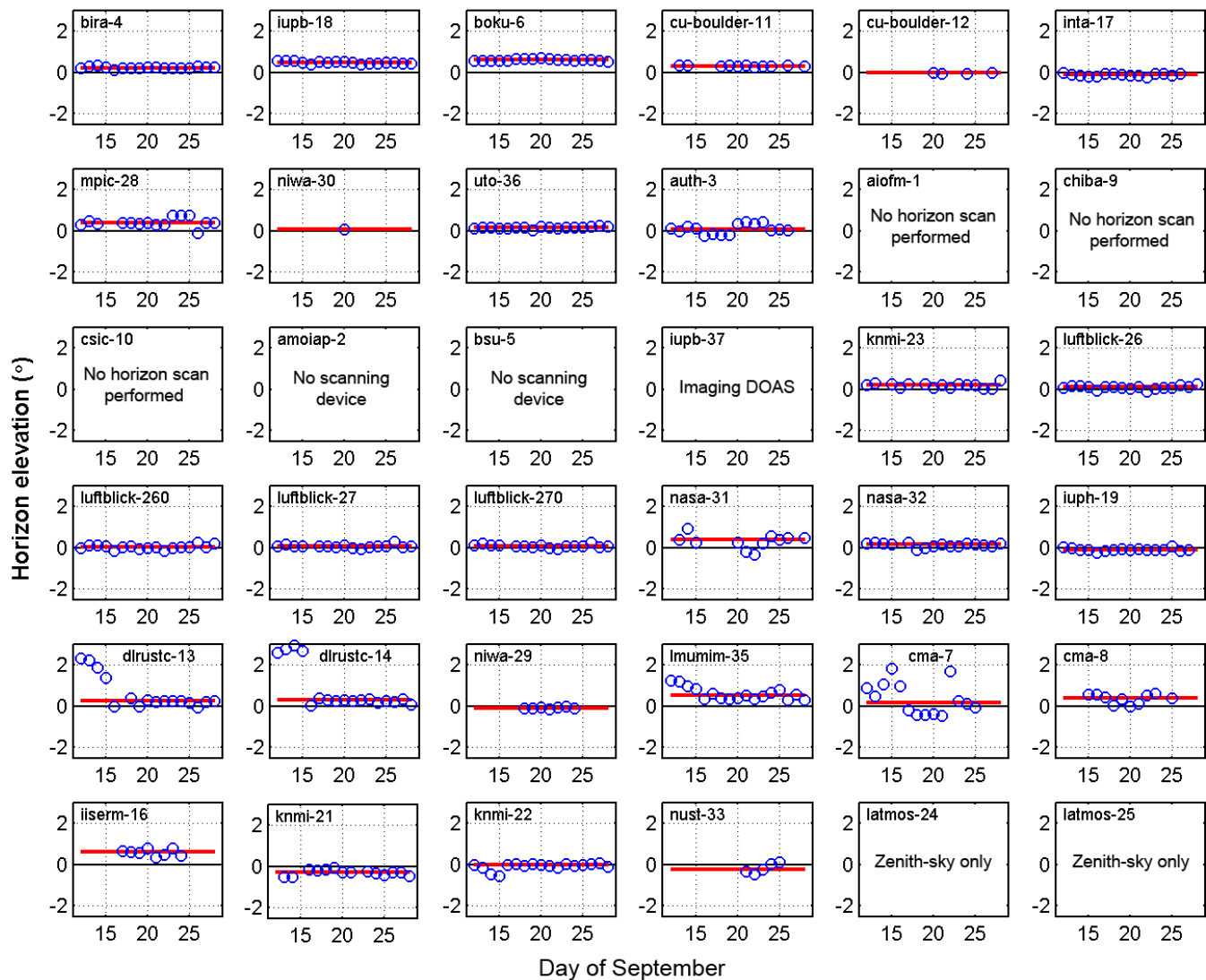
Figure 5: Time series of horizon elevation values (blue circles) derived from daily horizon scans performed with each instrument during the intercomparison period in the visible wavelength range (except for knmi-21). When no data is available for the horizon scan analysis, a short explanation is given. The red lines indicate the median values.
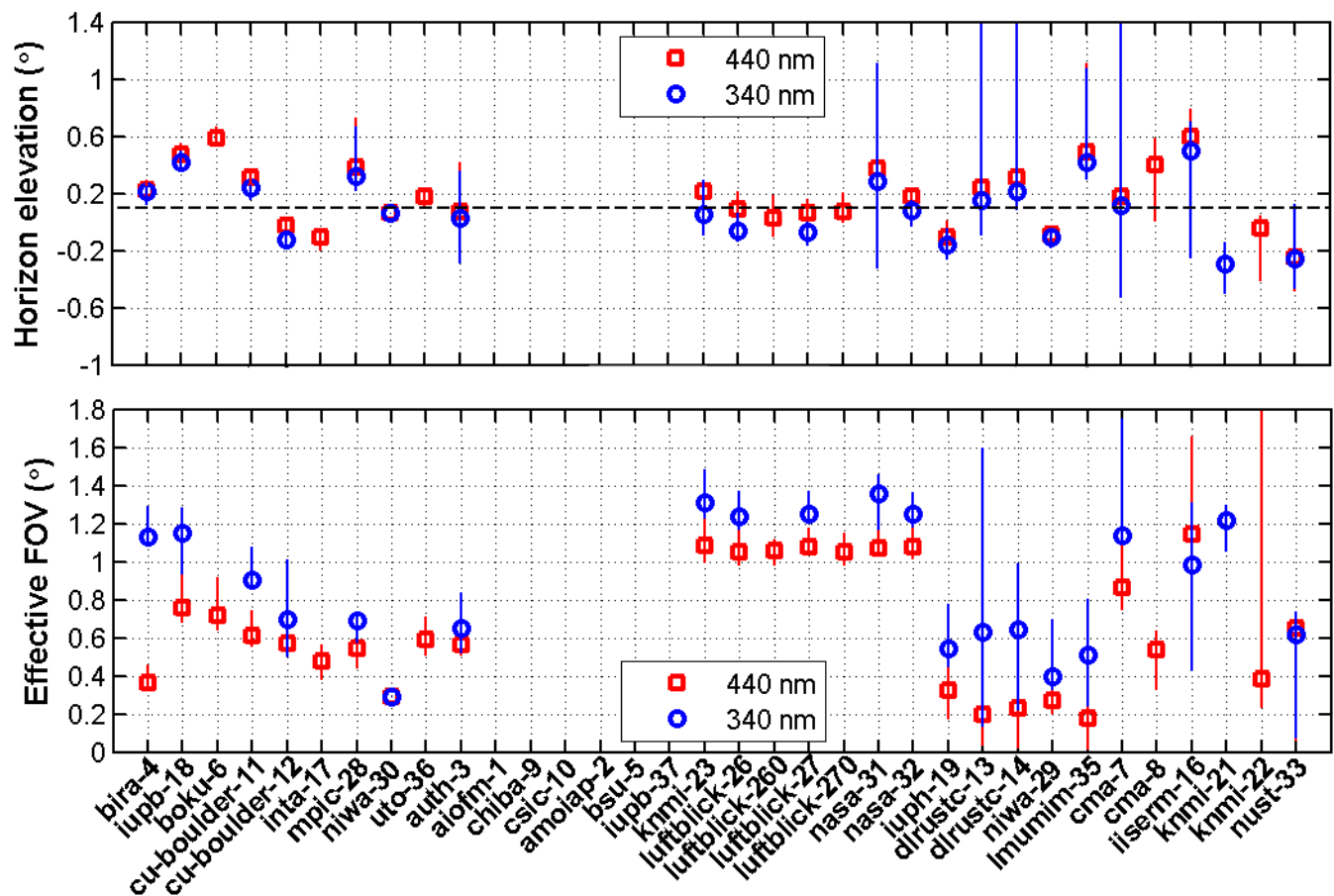
Figure 6: Summary of the average horizon elevation (top panel) and of the field of view (bottom panel), resulting from the horizon scans performed at 340nm and 440nm. Symbols represent median values and the vertical bars 10th and 90th percentiles.
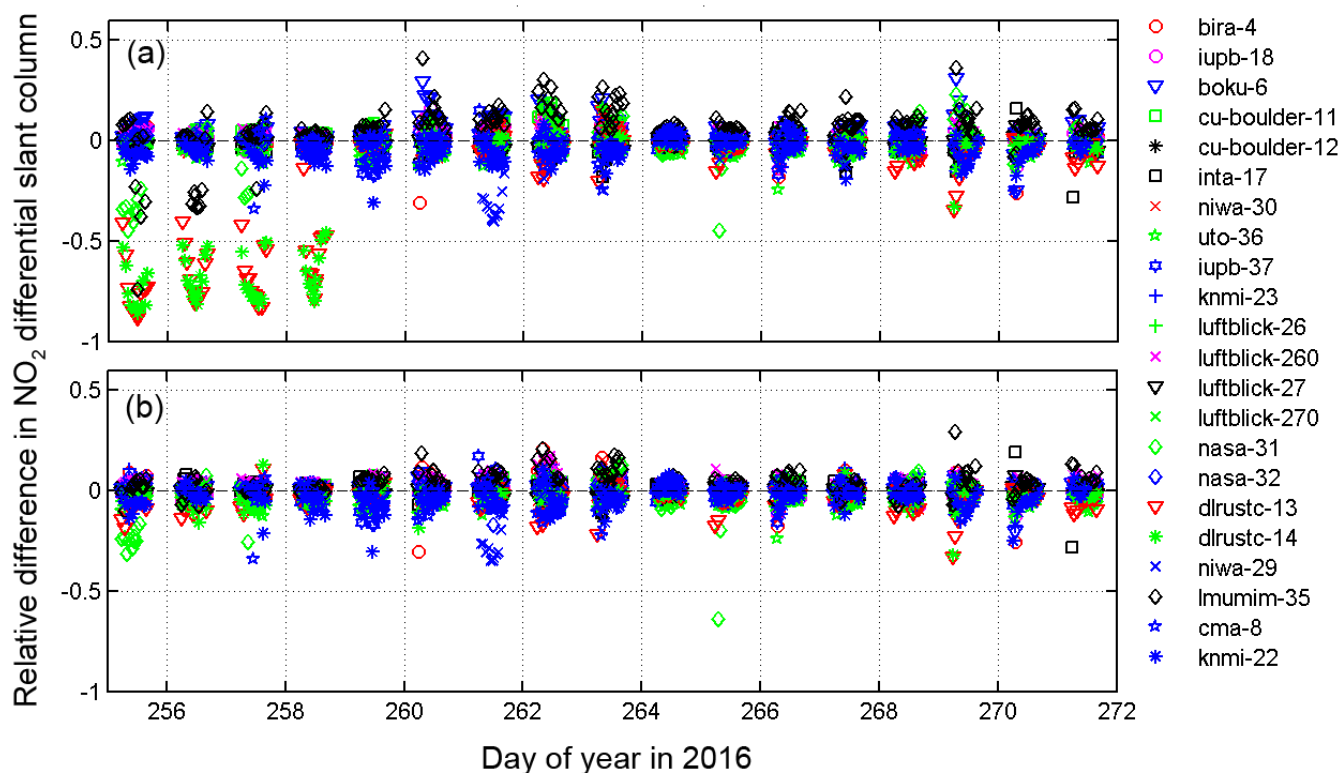
Figure 7: Relative differences of NO₂ dSCDs (in the visible wavelength region) with respect to the median from all instruments measured during the whole semi-blind intercomparison phase for the 287° azimuthal direction and 1° elevation angle. (a) Results before correction for elevation offsets, (b) same results after correction for elevation offsets derived from horizon scans. Colours and symbols represent different instruments.
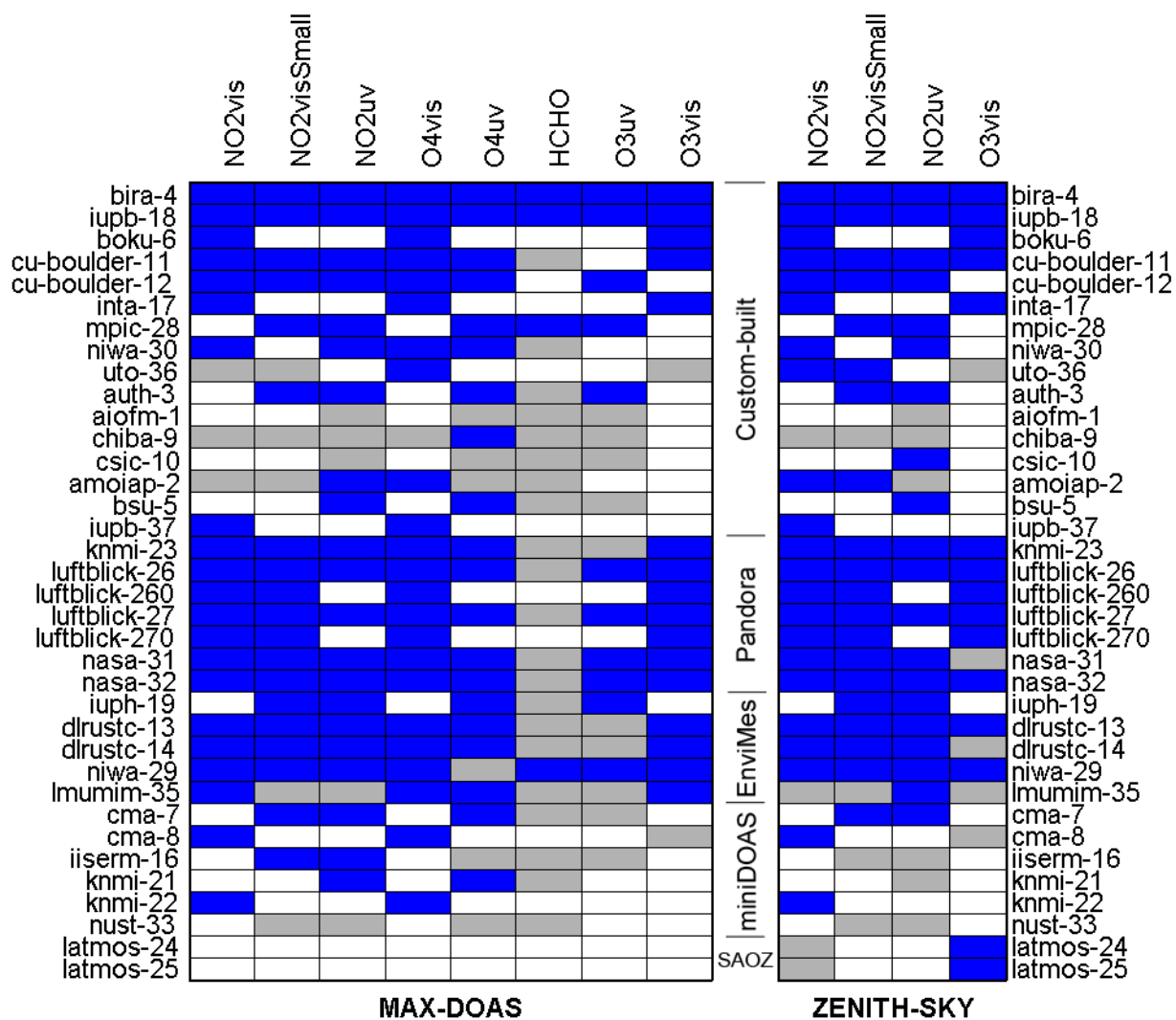
Figure 8: Instrument data sets selected to build the median MAX-DOAS reference (left panel) and zenith-sky (right panel) data sets. Blue marks the data sets included in the median while grey marks the data sets not included and white the ones not available. Note that the instruments are grouped according to their specific design as Custom-built, Pandora, EnviMes, miniDOAS or SAOZ.
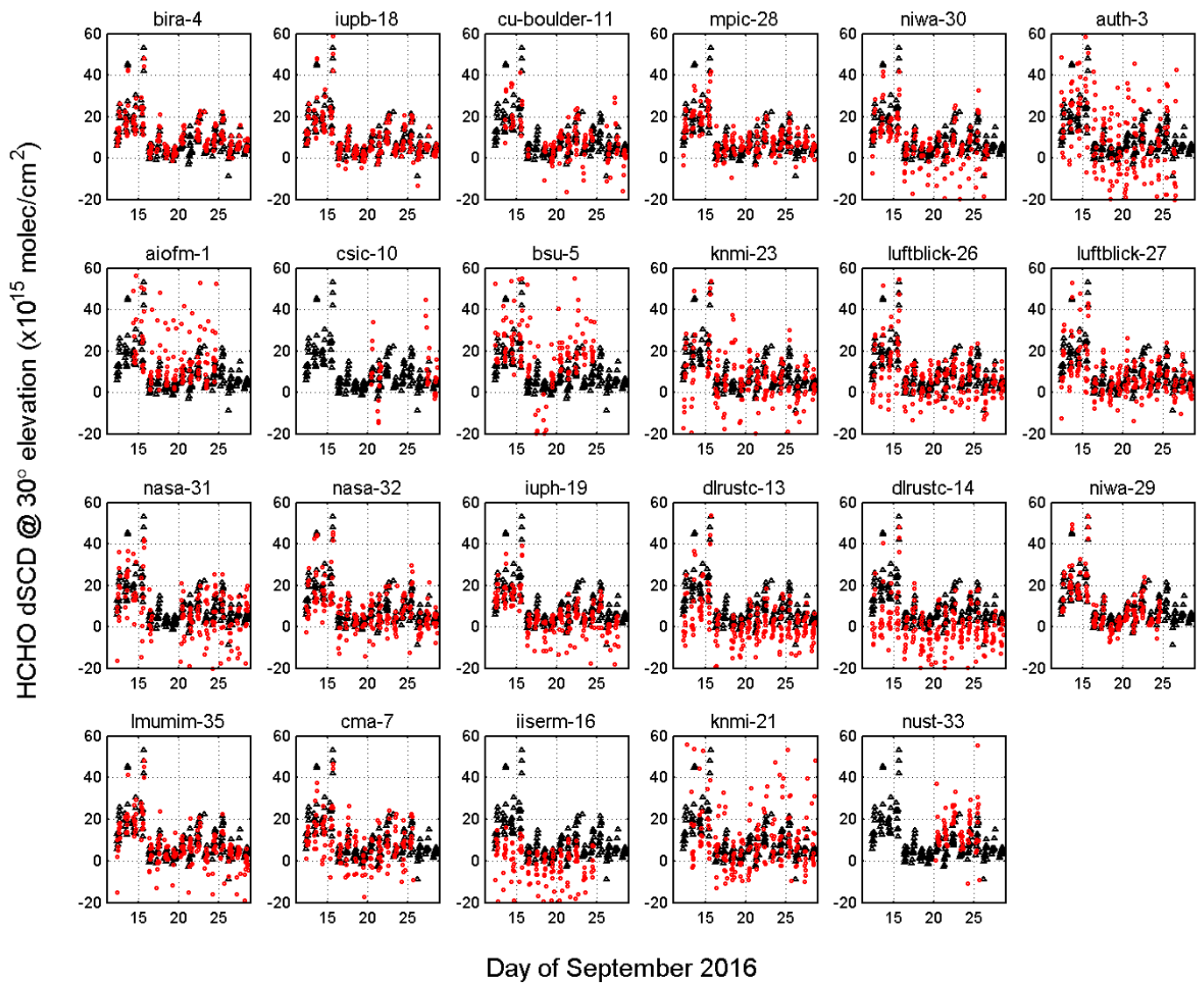
Figure 9: Comparison of HCHO dSCDs retrieved by each group at 30° elevation (red dots), and median values (black triangles). Only the four data sets (bira-4, iupb-18, mpic-28 and niwa-29) showing consistent values and a comparatively low noise level were selected for the calculation of the HCHO median.
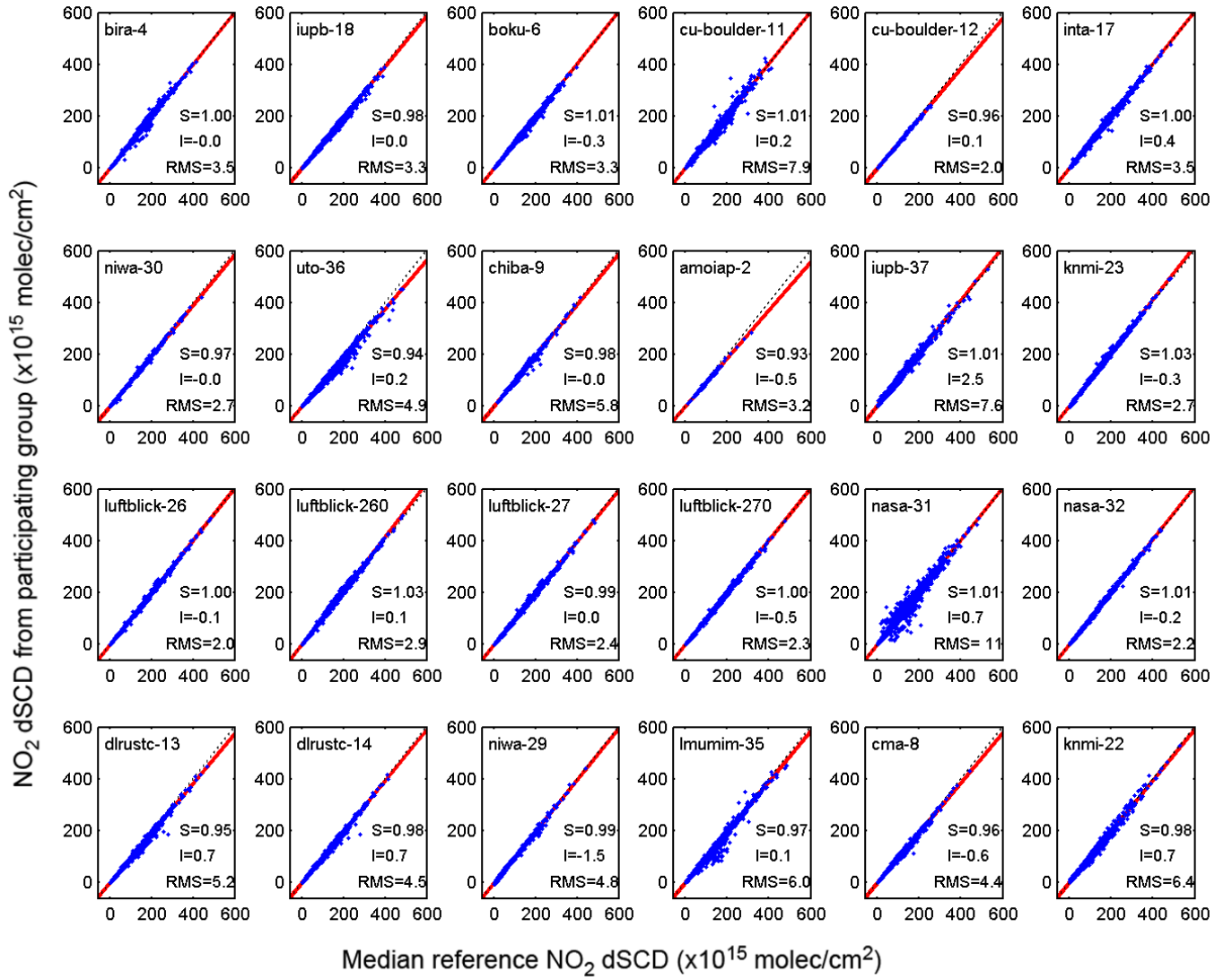
Figure 10: Regression analysis for NO₂ dSCDs (measured in the visible wavelength region) for each instrument which was measuring NO₂ in this wavelength region plotted against median values for the whole semi-blind phase, including all viewing and azimuth angles (blue crosses). The linear regression line is displayed as red line, the 1-to-1 line as a reference as dotted line. Instruments are identified with their affiliation and instrument ID number. S is the slope of the regression, I the intercept and RMS the root-mean-square of the regression residuals.
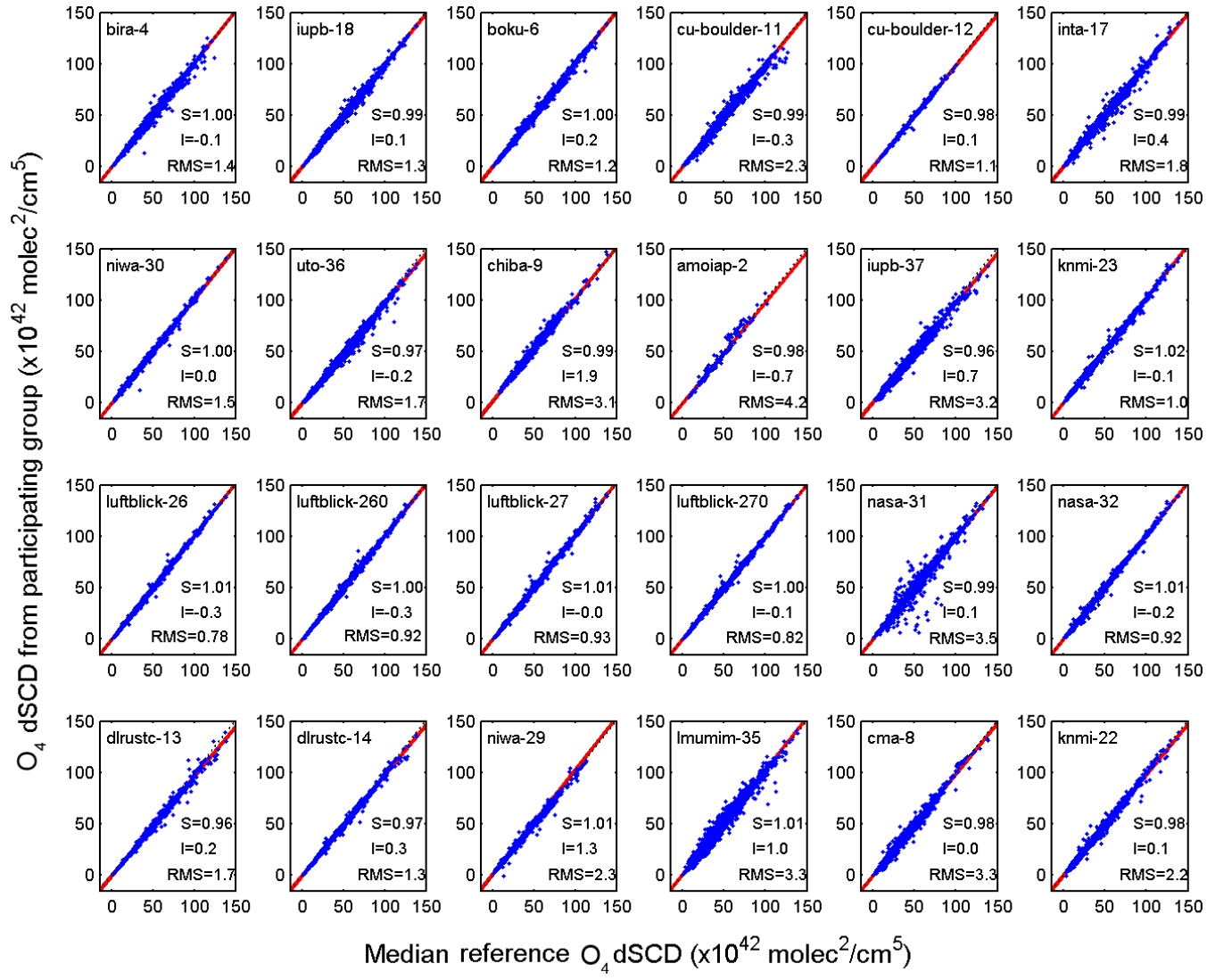
Figure 11: Same as Figure 10 but for O₄ dSCDs measured in the visible wavelength range.
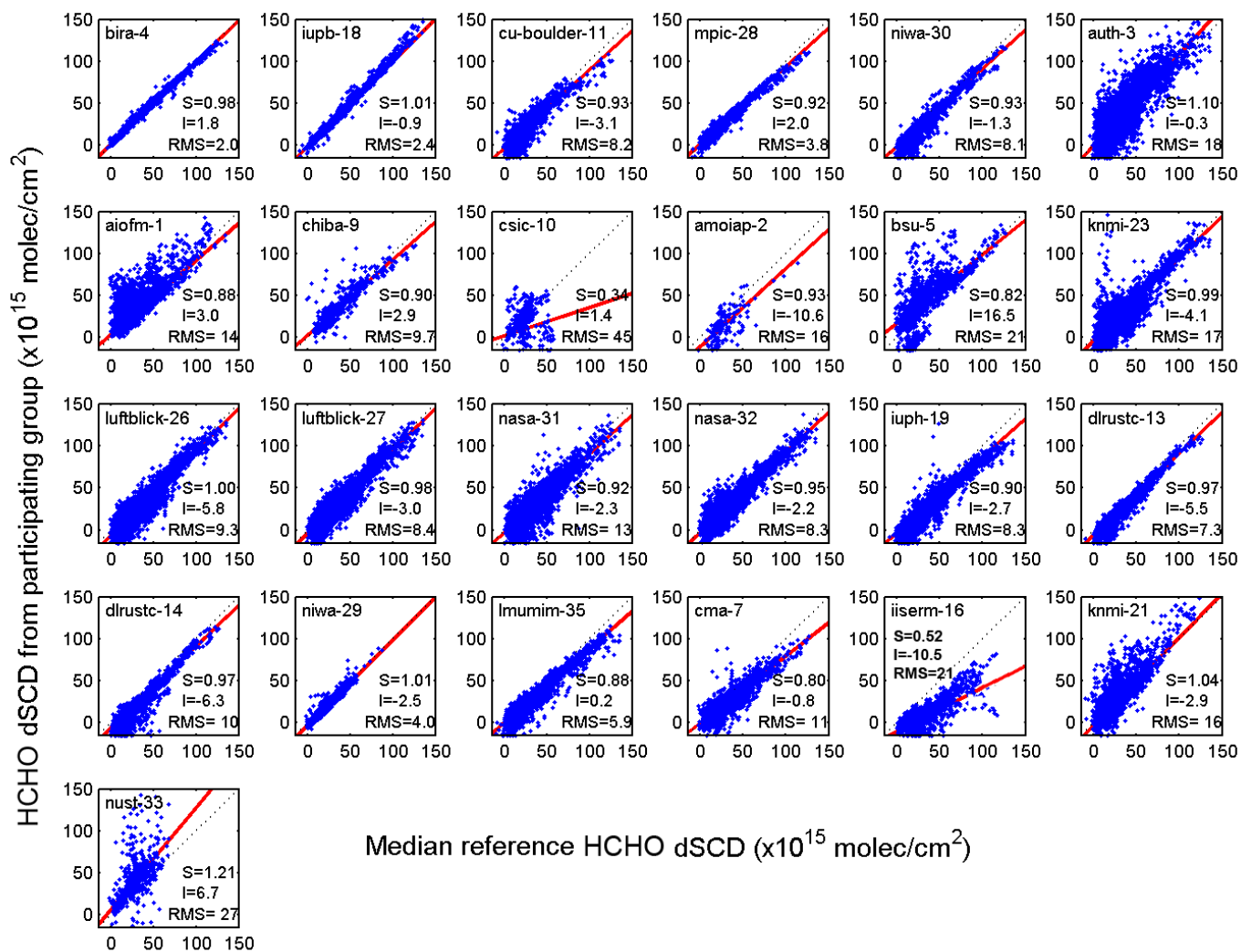
Figure 12: Same as Figure 10 but for HCHO dSCDs.
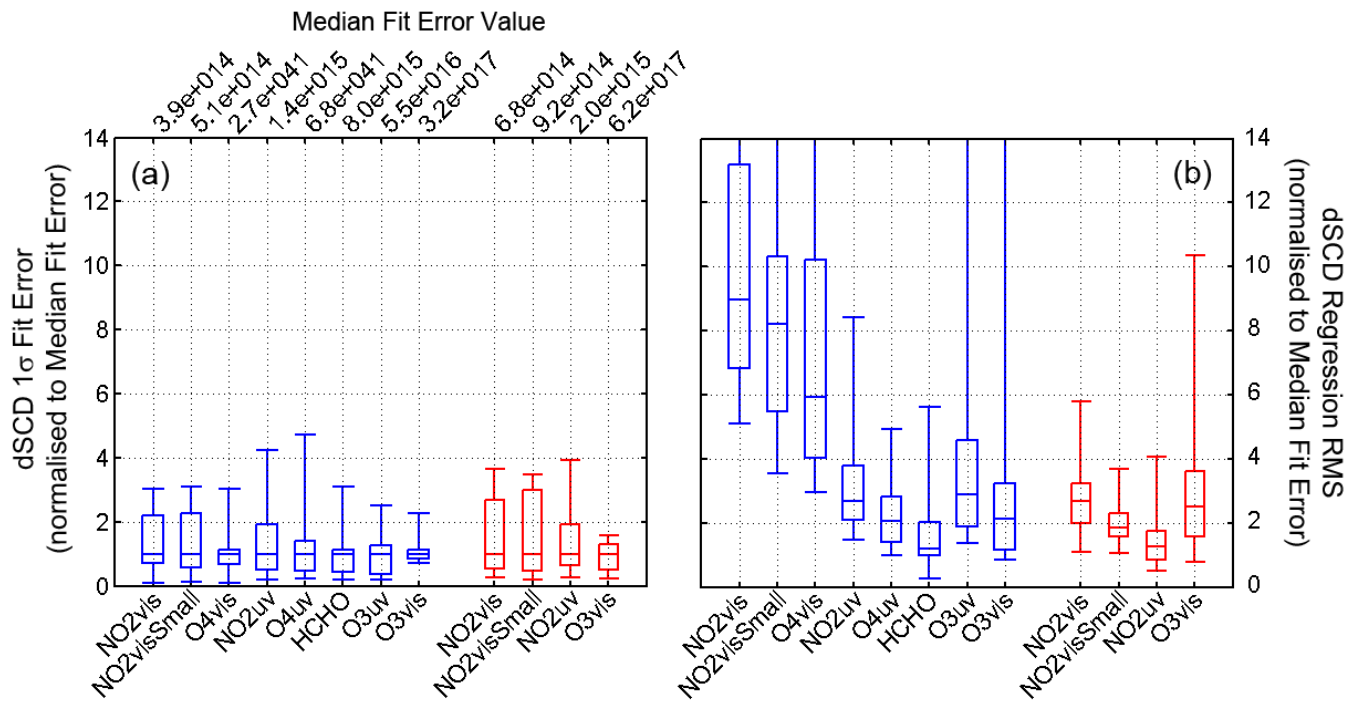
Figure 13: (a) Box-and-whisker plot of the 1 sigma fit error of the dSCDs for the 12 data products, for all instruments and for all elevation angles. MAX-DOAS products are represented in blue and zenith-sky twilight in red. **(b)** Box-and-whisker plot of the RMS from dSCD regression analyses, again for the 12 data products under investigation.
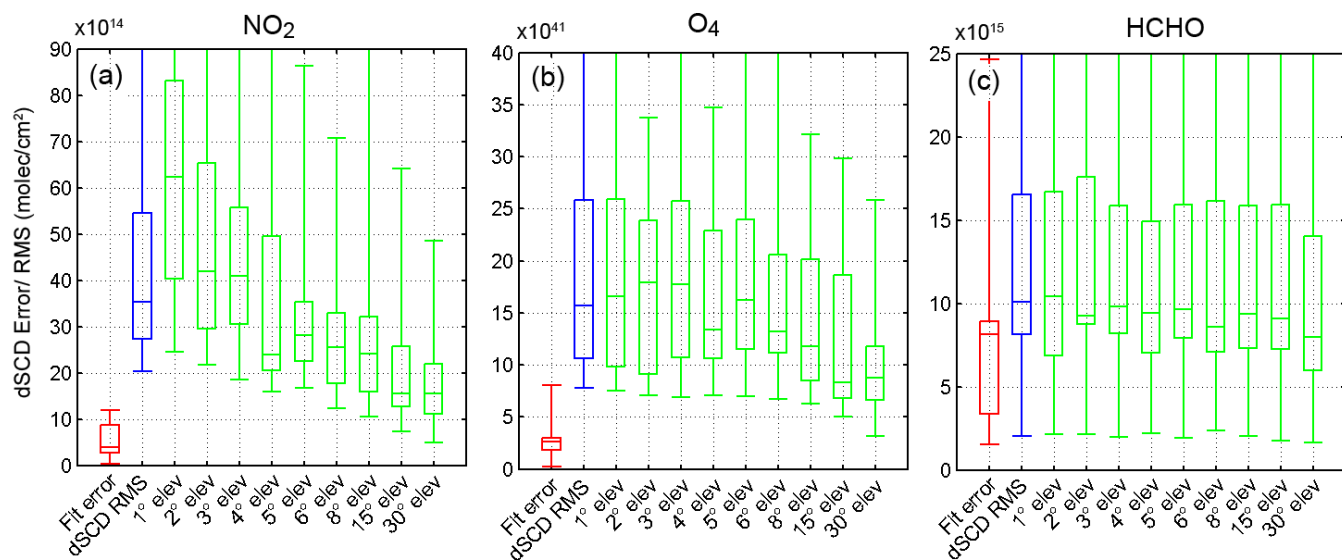
Figure 14: (a) Box-and-whisker plot of the 1 sigma dSCD fit error (red), the regression RMS for all elevation angles (blue) and RMS from dSCD regression analyses sorted as a function of the elevation angle (green) for NO2 in the visible wavelength range. (b) Same as (a) but for O4 (visible). (c) Same as (a) but for HCHO.
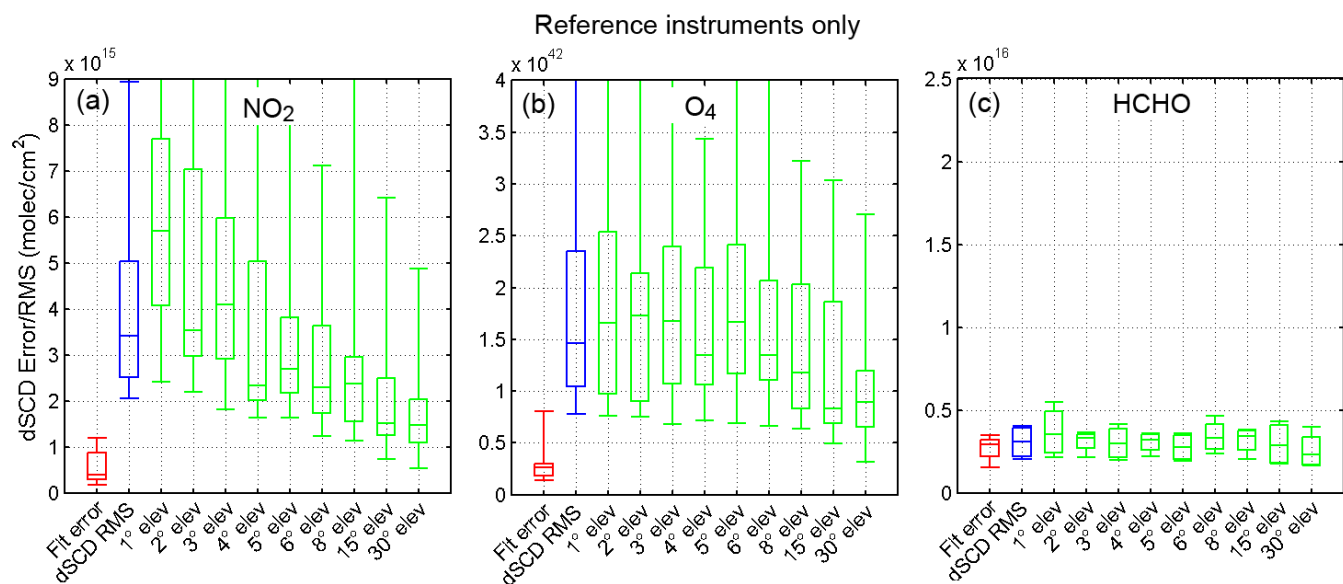
Figure 15: same as Figure 14, but for reference instruments only.
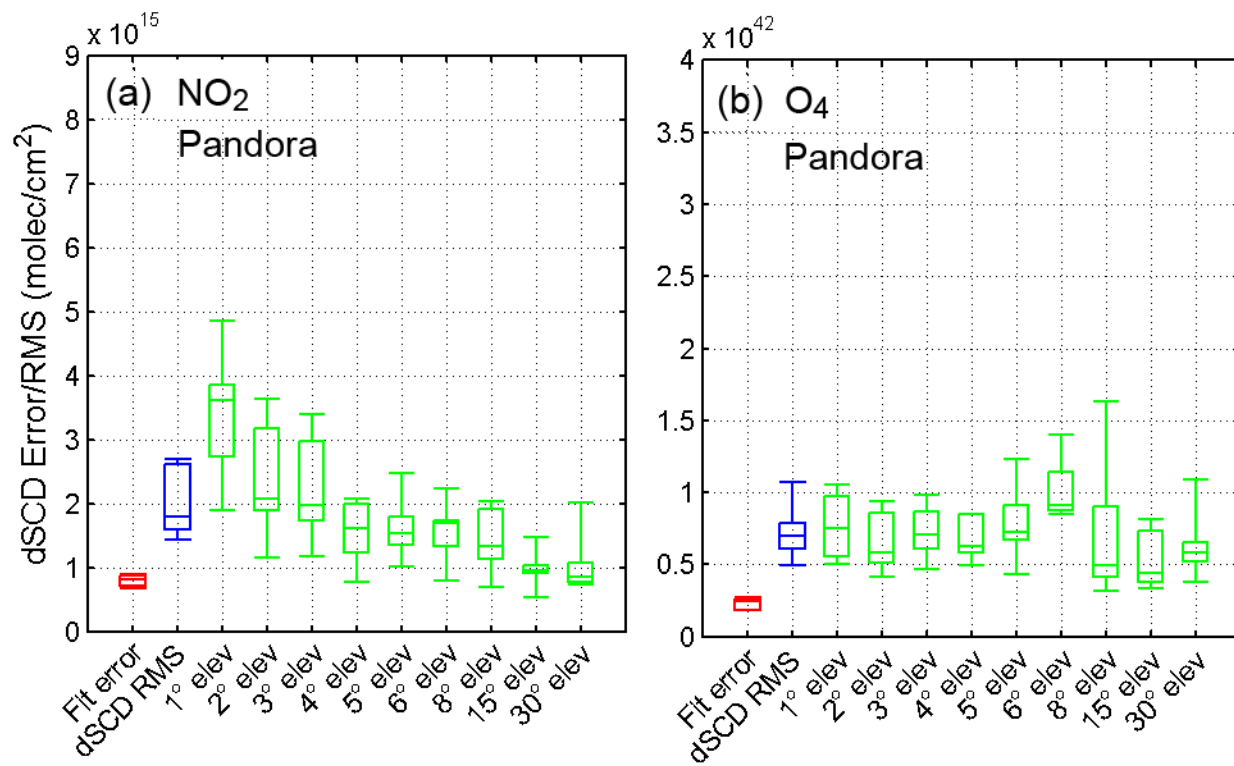
Figure 16: same as Figure 14, but for Pandora instruments only.
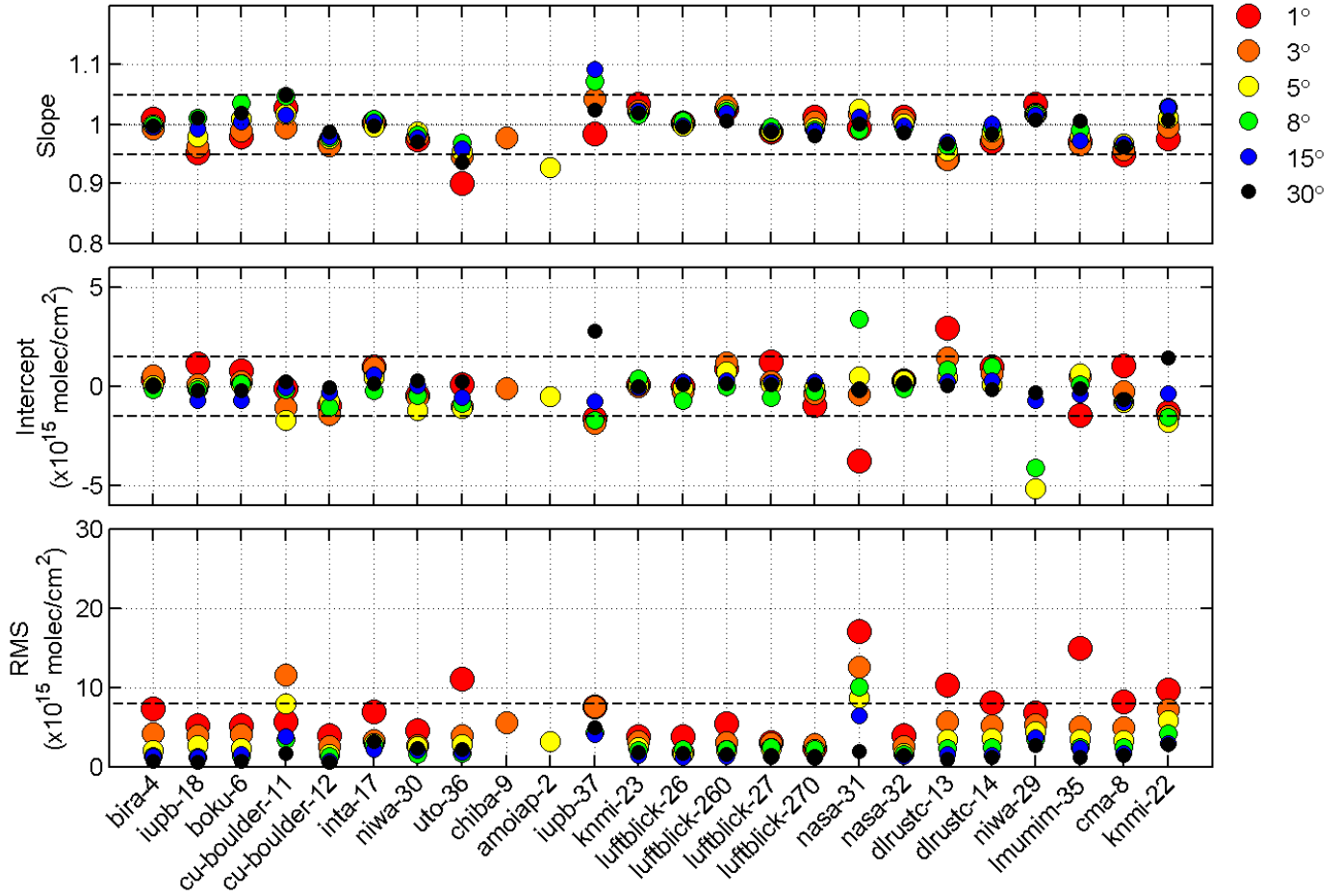
Figure 17: Slope, intercept and RMS of regression plots against the median dSCD reference, for each of the 24 instruments measuring NO₂ in the visible wavelength range (as shown in Figure 10). The values are colour-coded corresponding to the elevation angles (1º to 30º). Apart from a couple of exceptions (chiba-9, amoiap-2), most instruments are measuring the whole range of elevation angles. The dashed lines indicate the limits when comparing the values of the parameters for the different instruments with the aim to identify outliers in a more objective way. Section 4 and Figure 18 explain in more detail how the actual values of the limits were selected , and the values are listed in Table 4.
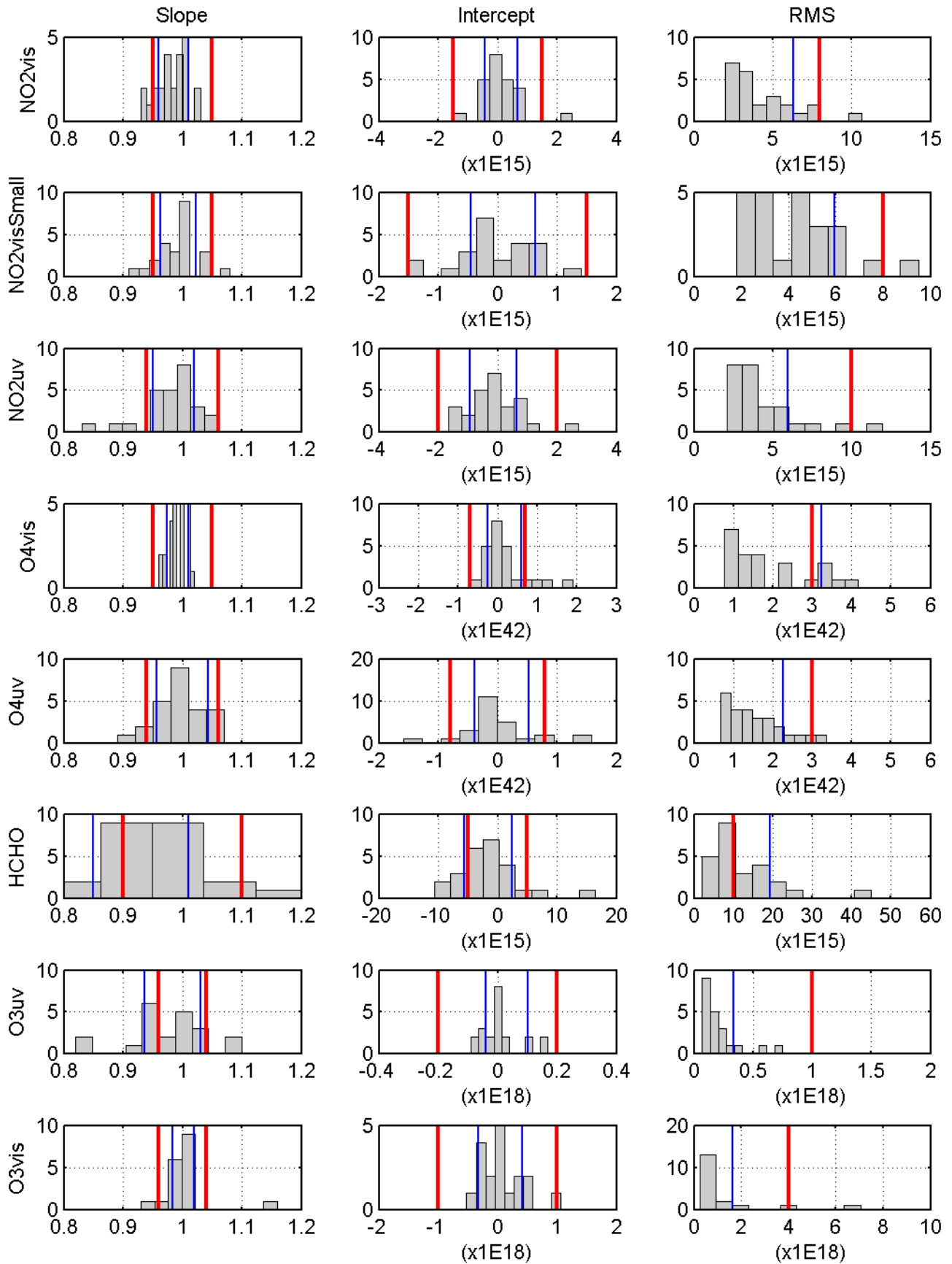
Figure 18: Limits for the assessment criteria for the eight MAX-DOAS data sets shown by red lines. The blue lines represent the percentiles 16 and 84 (84 only for RMS), together with histograms of the slope being displayed in the left column of panels, the intercept in the middle and the RMS in the rightmost panels (see also Table 4).
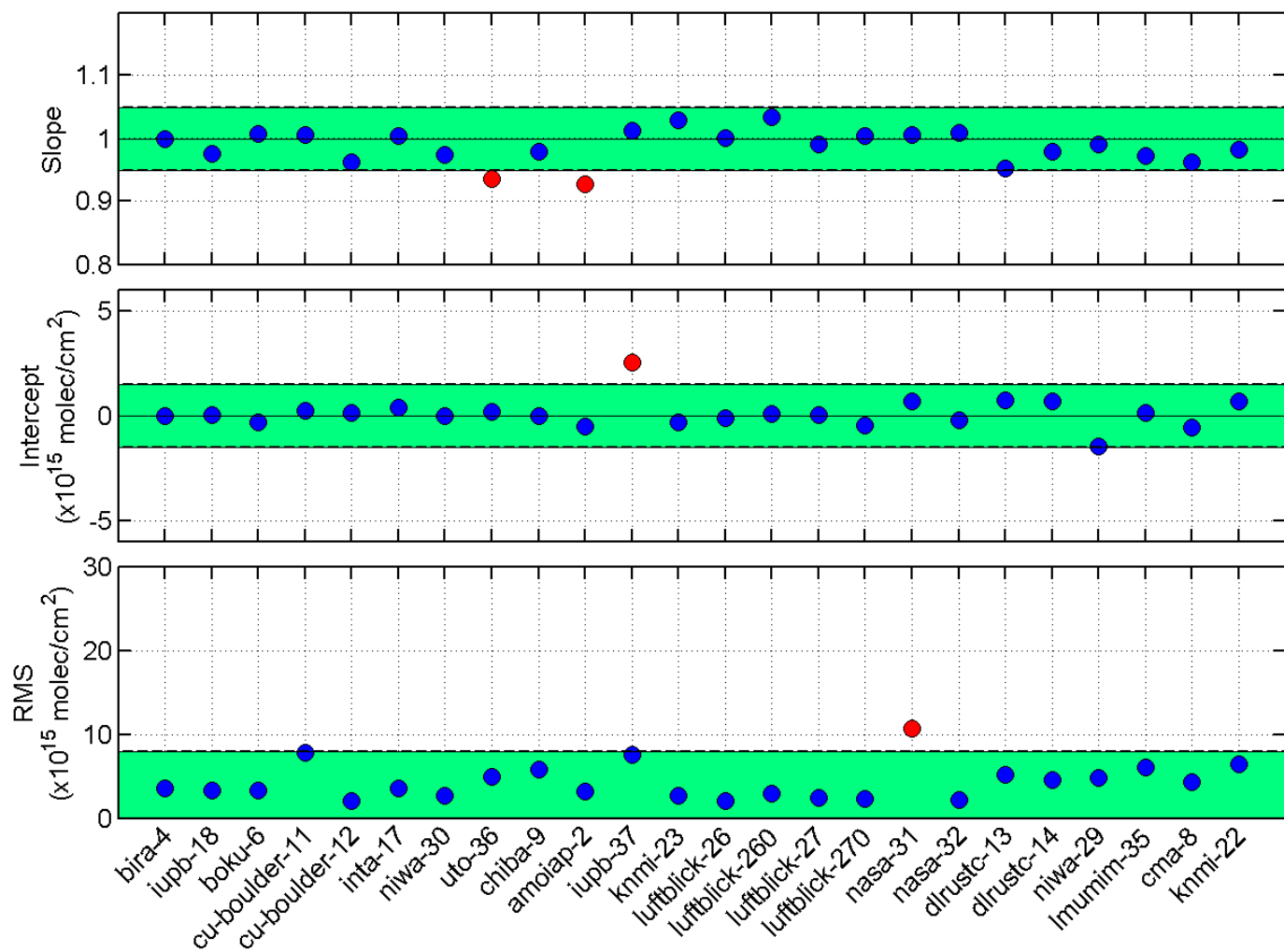
Figure 19: Summary of the NO₂ visible regression statistic shown in Figure 10. The slope, intercept and RMS values are displayed in the top, middle and bottom panel, respectively, for all measurement days, all viewing direction and all elevation angles. The green shading indicates the limits as defined in Table 4 and Figure 18 for NO2vis; the values falling within these limits are plotted in blue, the ones outside the limits in red.
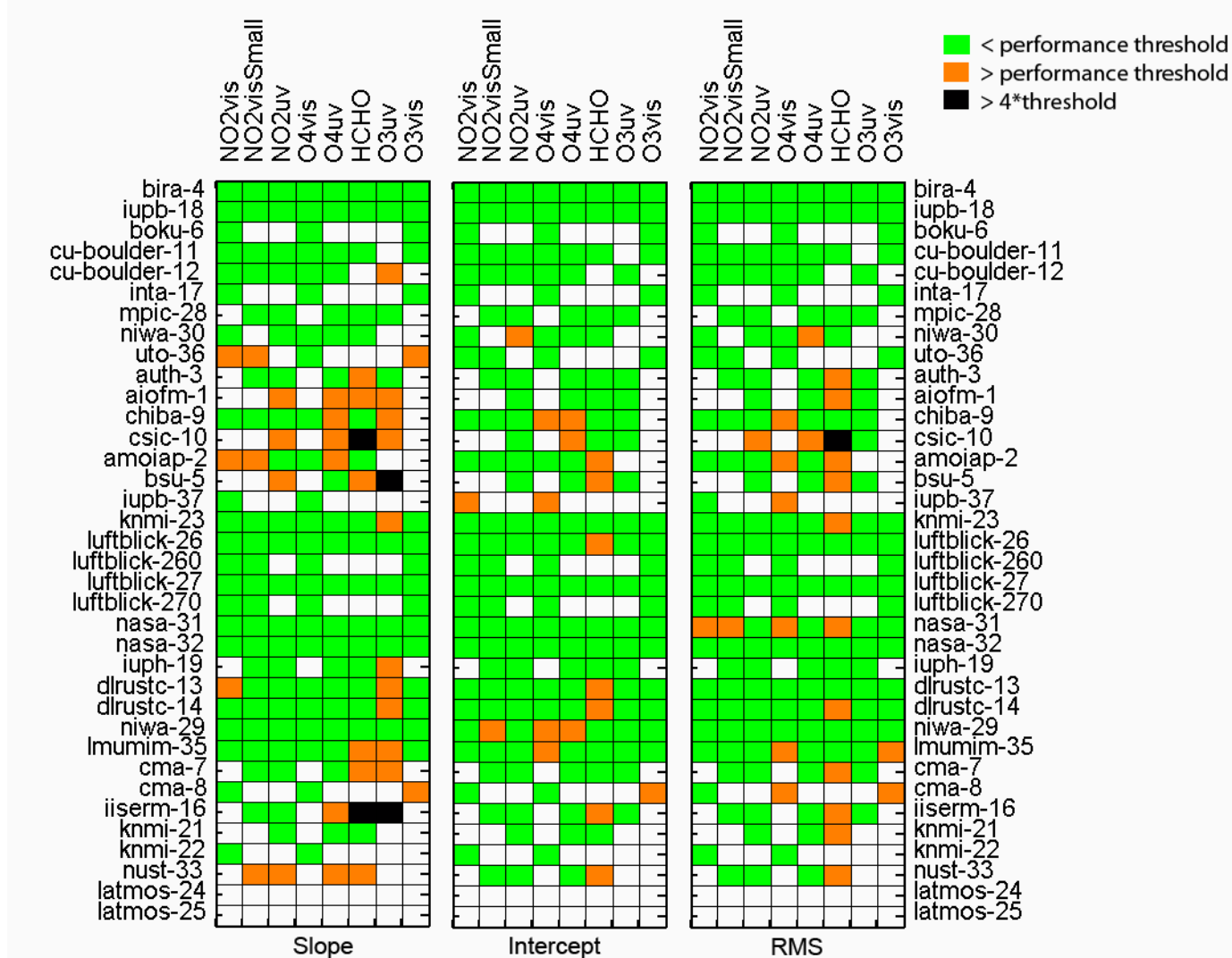
Figure 20: Overview of performance results for the slope, intercept and RMS from the regression analysis displayed for all participating instruments and MAX-DOAS data products. Colour coding denotes if each of the parameters is within the set criteria (green), if the performance threshold is exceeded (orange), or if it is exceeded by more than a factor of 4 (black).
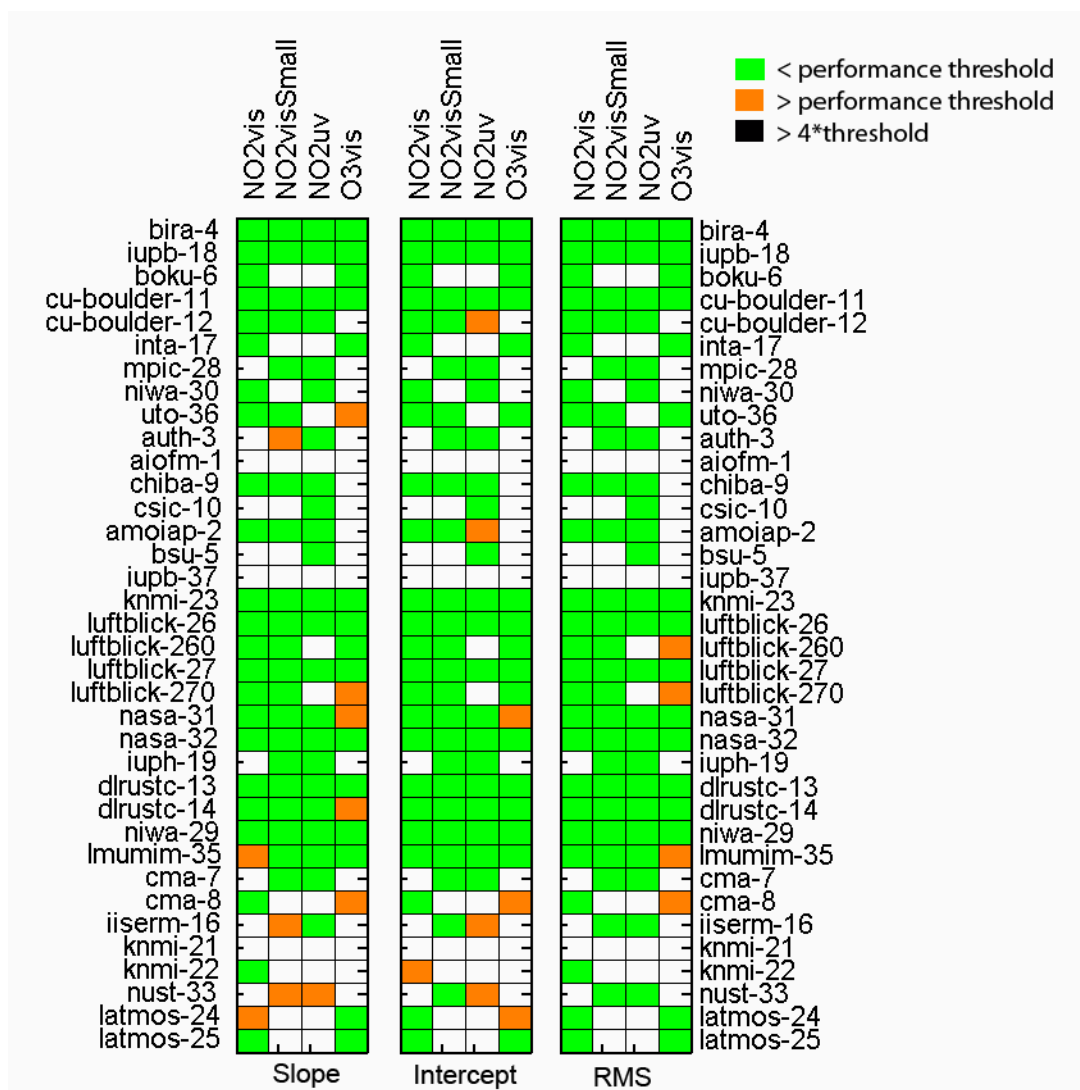
Figure 21: Same as Figure 20 but for the zenith-sky products.
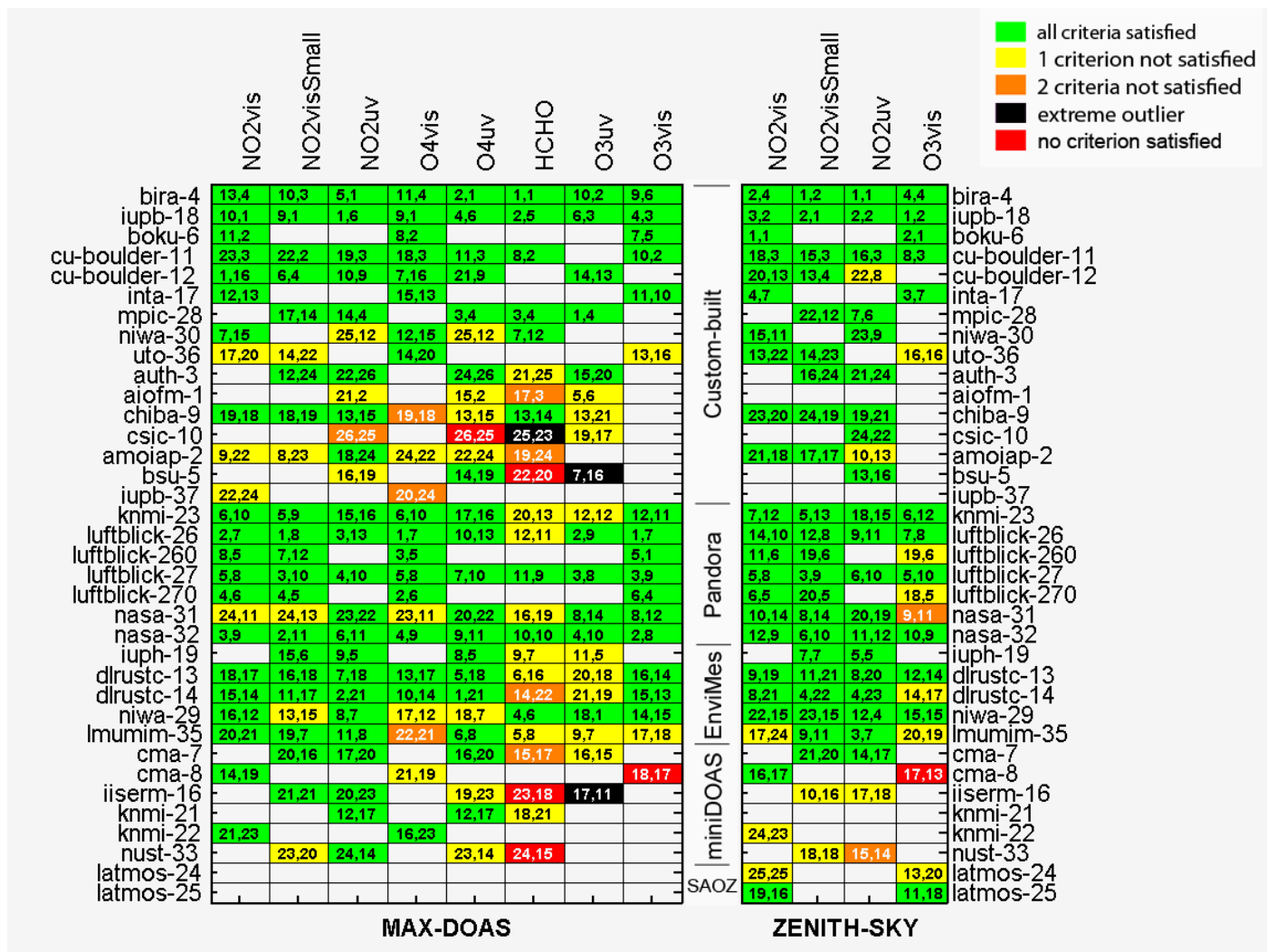
Figure 22: Assessment matrix for all 36 instruments and eight data products for MAX-DOAS, and four data products for zenith-sky mode. Green indicates that all three assessment criteria have been fulfilled, yellow means that one criterion is not satisfied, orange means two are not, red means all three criteria have not been met and black indicates that this data set has at least one extreme outlier. White indicates when data sets were not measured. The two numbers in each box indicate the rating for each product and instrument according to the dSCD regression RMS (first value) and the RMS calculated as part of the data fitting routine (second value). The instruments with the smallest RMS are denoted with the smallest numbers. Note that the instruments are grouped according to their specific design as Custom-built, Pandora, EnviMes, miniDOAS or SAOZ.

Appendix A: DOAS retrieval settings

For each data product, a set of retrieval settings and parameters was prescribed. The use of these settings was mandatory for participation in the semi-blind intercomparison. The tables below summarize the details of the DOAS retrieval configurations used for each data product. The referenced absorption cross-section files are available from the FRM4DOAS website (http://frm4doas.aeronomie.be/index.php/documents).

Table A1: DOAS settings for NO_2 and O_4 (VIS range)

Wavelength range	425-490 nm
Fraunhofer reference spectra	Noon zenith spectra averaged between 11:30:00 and 11:40:00 UT
Cross-sections:	
NO_2 (294 K)	Vandaele et al. (1998) with I_0 correction (SCD of 10^{17} molecules/cm^2) File: no2_294K_vanDaele.xs
NO_2 (220 K)	Pre-orthogonalized Vandaele et al. (1998) with I_0 correction (SCD of 10^{17} molecules/cm^2) File: no2a_220p294K_vanDaele_425-490nm.xs
O_3 (223 K)	Serdyuchenko et al. (2014) with I_0 correction (SCD of 10^{20} molecules/cm^2) File: o3_223K_SDY_air.xs
O_4 (293 K)	Thalman and Volkamer (2013) File: o4_thalman_volkamer_293K_inAir.xs
H_2O	HITEMP (Rothman et al., 2010) File: H2O_HITEMP_2010_390-700_296K_1013mbar_air.xs
Ring	RING_QDOAS_SAO2010 File: Ring_QDOAScalc_HighResSAO2010_Norm.xs
Polynomial degree	Order 5 (6 coefficients)
Intensity off-set	Constant

Table A2: DOAS settings for NO_2 and O_4 (alternative VIS range)

Wavelength range	411-445 nm
Fraunhofer reference spectra	Noon zenith spectra averaged between 11:30:00 and 11:40:00 UT
Cross-sections:	
NO_2 (294 K)	Vandaele et al. (1998) with I_0 correction (SCD of 10^{17} molecules/cm^2) File: no2_294K_vanDaele.xs
NO_2 (220 K)	Pre-orthogonalized Vandaele et al. (1998) with I_0 correction (SCD of 10^{17} molecules/cm^2) File: no2a_220p294K_vanDaele_425-490nm
O_3 (223 K)	Serdyuchenko et al. (2014) with I_0 correction (SCD of 10^{20} molecules/cm^2) File: o3_223K_SDY_air.xs
O_4 (293 K)	Thalman and Volkamer (2013)

	File: o4_thalman_volkamer_293K_inAir.xs
H₂O	HITEMP (Rothman et al., 2010) File: H2O_HITEMP_2010_390-700_296K_1013mbar_air.xs
Ring	RING_QDOAS_SAO2010 File: Ring_QDOAScalc_HighResSAO2010_Norm.xs
Polynomial degree	Order 4 (5 coefficients)
Intensity off-set	Constant

Table A3: DOAS settings for NO₂ and O₄ (UV range)

Wavelength range	338-370 nm
Fraunhofer reference spectra	Noon zenith spectra averaged between 11:30:00 and 11:40:00
Cross-sections:	
NO₂ (294 K)	Vandaele et al. (1998) with I_0 correction (SCD of 10^{17} molecules/cm^2) File: no2_294K_vanDaele.xs
NO₂ (220 K)	Pre-orthogonalized Vandaele et al. (1998) with I_0 correction (SCD of 10^{17} molecules/cm^2) File: no2a_220p294K_vanDaele_338-370nm.xs
O₃ (223 K)	Serdyuchenko et al. (2014) with I_0 correction (SCD of 10^{20} molecules/cm^2) File: o3_223K_SDY_air.xs
O₃ (243 K)	Pre-orthogonalized Serdyuchenko et al. (2014) with I_0 correction (SCD of 10^{20} molecules/cm^2) File: o3a_243p223K_SDY_338-370nm.xs
O₄ (293 K)	Thalman and Volkamer (2013) File: o4_thalman_volkamer_293K_inAir.xs
HCHO (297 K)	Meller and Moortgat (2000) File: hcho_297K_Meller.xs
BrO (223 K)	Fleischmann et al. (2004) File: bro_223K_Fleischmann.xs
Ring	RING_QDOAS_SAO2010 File: Ring_QDOAScalc_HighResSAO2010_Norm.xs
Polynomial degree	Order 5 (6 coefficients)
Intensity off-set	Constant

Table A4: DOAS settings for HCHO

Wavelength range	336.5-359 nm
Fraunhofer reference spectra	Noon zenith spectra averaged between 11:30:00 and 11:40:00 UT
Cross-sections:	

HCHO (297 K)	Meller and Moortgat (2000) File: hcho_297K_Meller.xs
NO$_2$ (294 K)	Vandaele et al. (1998) with I$_0$ correction (SCD of 10^{17} molecules/cm^2) File: no2_294K_vanDaele.xs
O$_3$ (223 K)	Serdyuchenko et al. (2014) with I$_0$ correction (SCD of 10^{20} molecules/cm^2) File: o3_223K_SDY_air.xs
O$_3$ (243 K)	Pre-orthogonalized Serdyuchenko et al. (2014) with I$_0$ correction (SCD of 10^{20} molecules/cm^2) File: o3a_243p223K_SDY_324-359nm.xs
O$_4$ (293 K)	Thalman and Volkamer (2013) File: o4_thalman_volkamer_293K_inAir.xs
BrO (223 K)	Fleischmann et al. (2004) File: bro_223K_Fleischmann.xs
Ring	RING_QDOAS_SAO2010 File: Ring_QDOAScalc_HighResSAO2010_Norm.xs
Polynomial degree	Order 5 (6 coefficients)
Intensity off-set	Order 1

Table A5: DOAS settings ozone in the Chappuis band

Wavelength range	450-520 nm
Fraunhofer reference spectra	Noon zenith spectra averaged between 11:30:00 and 11:40:00 UT
Cross-sections:	
O$_3$ (223 K)	Serdyuchenko et al. (2014) with I$_0$ correction (SCD of 10^{20} molecules/cm^2) File: o3_223K_SDY_air.xs
O$_3$ (293 K)	Pre-orthogonalized Serdyuchenko et al. (2014) with I$_0$ correction (SCD of 10^{20} molecules/cm^2) File: o3a_293p223K_SDY_450-550nm.xs
NO$_2$ (294 K)	Vandaele et al. (1998) with I$_0$ correction (SCD of 10^{17} molecules/cm^2) File: no2_294K_vanDaele.xs
NO$_2$ (220 K)	Pre-orthogonalized Vandaele et al. (1998) with I$_0$ correction (SCD of 10^{17} molecules/cm^2) File: no2a_220p294K_vanDaele_450-550nm.xs
O$_4$ (296 K)	Thalman and Volkamer (2013) File: o4_thalman_volkamer_293K_inAir.xs
H$_2$O	HITEMP (Rothman et al., 2010) File: H2O_HITEMP_2010_390-700_296K_1013mbar_air.xs
Ring	RING_QDOAS_SAO2010 File: Ring_QDOAScalc_HighResSAO2010_Norm.xs
Polynomial degree	Order 5 (6 coefficients)
Intensity off-set	Order 1

Table A6: DOAS settings ozone in the Huggins band

Wavelength range	320-340 nm
Fraunhofer reference spectra	Noon zenith spectra averaged between 11:30:00 and 11:40:00 UT
Cross-sections:	
O_3 (223 K)	Serdyuchenko et al. (2014) with I_0 correction (SCD of 10^{20} molecules/cm^2) File: o3_223K_SDY_air.xs
O_3 (293 K)	Pre-orthogonalized Serdyuchenko et al. (2014) with I_0 correction (SCD of 10^{20} molecules/cm^2) File: o3a_293p223K_SDY_320-340nm.xs
O_3	Non-linear correction terms (Puķīte et al., 2010) Files: o3_SDY_Pukite1_320-340nm.xs and o3_SDY_Pukite2_320-340nm.xs
NO_2 (294 K)	Vandaele et al. (1998) with I_0 correction (SCD of 10^{17} molecules/cm^2) File: no2_294K_vanDaele.xs
HCHO (297 K)	Meller and Moortgat (2000) File: hcho_297K_Meller.xs
Ring	RING_QDOAS_SAO2010 File: Ring_QDOAScalc_HighResSAO2010_Norm.xs
Polynomial degree	Order 3 (4 coefficients)
Intensity off-set	Order 1

Appendix B: History of slant column data set revisions

This appendix provides a history of the slant column data set resubmissions accepted after the formal deadline for participation to the semi-blind intercomparison (18 October 2016). The main motivation for accepting late revisions was to remedy well-identified mistakes. Details of the submitted revisions, including justifications for the changes and corresponding dates, are listed below.

AIOFM (aiofm-1)

Data files resubmitted on 16 October 2017 with two additional corrections applied, which were: (1) A dark current correction and (2) a wavelength shift that needed to be applied with respect to the reference spectrum. The O3uv data set was also resubmitted in September 2019 because an incorrect ozone cross-section was used previously for the data analysis.

AUTH (auth-3)

Data files resubmitted on 17 March 2017. These were corrected for a systematic wavelength shift of the measured spectra.

BIRA-IASB (bira-4)

Revised data submitted on 28 February 2017, with small changes summarized as follows: (1) A correction of an error affecting the dark current subtraction in the UV channel (affecting HCHO, NO2uv, O4uv and O3uv, mostly at large SZA) and (2) an optimisation of the filtering scheme were applied. For the visible products, all measurement points having RMS values exceeding 5 times the daily median RMS calculated in hourly bins were excluded. The same procedure was also applied to the UV products with any data values exceeding 4 times the median being excluded. This approach was found sufficient to exclude outliers due to an electronic instability in the UV channel.

CHIBA (chiba-9)

Data files resubmitted on 11 January 2018, with additional stray-light corrections applied to the measured spectra. This correction was derived as part of the wavelength calibration procedure. Considering the nominal spectral range 310 to 525 nm, 11 discrete wavelength regions (316 ± 5, 336 ± 5, 344 ± 5, 358 ± 5, 374 ± 5, 384 ± 5, 395 ± 5, 410 ± 5, 431 ± 5, 486 ± 10, and 518 ± 5 nm) were selected and analysed. In each spectral window, the spectrum was fitted using an iterative inversion method. The measurement vector consisted of the intensities measured by the MAX-DOAS instrument. The components of the state vector were set to the wavelength shift, the FWHM for the left part of an asymmetric Gaussian instrument line shape (FWHM1), the FWHM for the right part (FWHM2), and the differential slant column (dSCD) of significant absorbers (O_3, NO_2) in the analysed wavelength region. In addition, a scaling polynomial and a constant offset term (or stray-light correction term) were included in the state vector to scale the high-resolution solar spectrum data to the intensities measured by MAX-DOAS.

CMA (cam-7, cma-8)

Revised data files were resubmitted on 26 September 2016 for CMA-7 (UV and VisSmall range) and CMA-8 (Vis range). Periods with bad motor connection were filtered out in the resubmitted data. Additionally, fitting of the wavelength shift between measurement spectrum and reference spectrum was added in the revised processing.

CU-Boulder (CU-boulder-11, CU-boulder-12)

Revised data files were submitted for all gases on 4 March 2017. For CU-boulder-11, the resubmitted data were filtered for periods with bad motor connection (when instrument operated in 1D or in zenith geometry), and one corrupt file was corrected. For CU-boulder-12, revised files were only submitted for gases analysed in the UV wavelengths range. Resubmitted data accounted for a

time-dependent etalon identified on the UV spectrometer and fitted as a pseudo absorber with independent shift and stretch. This approach captured the errant signal effectively at longer wavelength but was less effective at shorter wavelengths; no HCHO data were reported. The source of the etalon has since been eliminated.

INTA (inta-17)

Revised data files submitted on 14 February 2017, due to one change in their data analysis routine: The inverse of the actual measurement was used as offset instead of the inverse of the reference spectrum leading to a smaller uncertainty and improved retrievals of the sunrise and sunset slant columns. This change mainly affects twilight data but for consistency the complete data set was reanalysed.

KNMI (knmi-21, knmi-22)

Data files resubmitted on 27 January 2017 with the following corrections: (1) Fitting of the wavelength shift between measurement spectrum and reference spectrum was previously omitted and had to be added. (2) For knmi-22: Due to an instable tripod, the logged angles can only be trusted when the horizon measurements show a consistent horizon from day to day (<0.5° difference). The measurements during all other periods have been filtered out.

LATMOS (latmos-25)

Data files resubmitted data on 4 April 2018, because the data files had to be corrected for detector non-linearity effects that were identified after the campaign. The detector is a Hamamatsu CCD 2048x16 type S11071-1104. The non-linearity of this detector was measured and corrected applying the procedure described in AvaSpec-DLL Manual V9.7.0.0 (p71-73). A stable light source (Xe lamp, VG9 filter and diffuser) was used to measure spectra at different integration times between 50 ms and 1830 ms. The maximum level of the elementary spectrum varies from 400 to 16000 counts. The correlation between the flux (count/s) and the number of counts of an elementary spectrum at several pixels was fitted by a polynomial of degree 7 and this curve was then used to correct raw data as recommended by Avantes.

LMU-MIM (lmumim-35)

Data files resubmitted with two corrections applied on 24 March 2018: (1) The spectra were re-analyzed with a correction for detector non-linearity and the analysis was updated by using offset and dark current spectra. The latter spectra were measured after the CINDI-2 campaign and also corrected for detector non-linearity. (2) The DOAS fit was performed using QDOAS to fit the instrument slit function while for the originally submitted data set a fixed instrument slit function measured with a Hg lamp was used.

LuftBlick/NASA (knmi-23, luftblick-26, 27, 260, 270, nasa-31, 32)

Revised data sets submitted on 4 October 2017. Pandora data during CINDI-2 were processed using BlickP, the native Pandonia Global Network (PGN) software. BlickP allows for fitting of molecular absorption cross-sections of a specific species represented in terms of constant, or linear, or quadratic functions of temperature. Orthogonalization of cross-sections is not allowed. Pandora NO_2 and O_3 slant columns had to be recalculated to "simulate" the case, where cross-sections of the same gas at different temperatures are used in the fitting. In addition, measurements at azimuth angles 95° and 135° at elevation angle of 1° were eliminated due to obstruction. There was also a mistake in the intensity calibration correction in the original submission.

NIWA (niwa-30)

Data files resubmitted for NO_2 in the visible and UV, and for HCHO on 27 March 2017. The data were reprocessed to include a test that detects any bad timing on a spectrum and removes the results for that spectrum. This occasional fault was likely due to last minute logging program changes to enable the one available spectrometer to switch wavelength between the visible and UV regions every quarter hour.

NUST (nust-33)

Data files resubmitted on 10 February 2017, after exploring the relatively larger RMS values. A misalignment of elevation angles was noticed in the analyses due to malfunctioning of the Peltier controller unit and loose gear of the stepper motor. On 15 Sep. 2016, as the instrument was replaced with a new instrument No. 15306 (where a problem with the slit was identified and was adjusted). The new instrument was functioning properly, but there was no lamp experiment to adjust the azimuth direction until 19 Sep 2017. Systematic high RMS values are observed for all elevation angles in the retrieved NO2visSmall (411- 445 nm) and HCHO DSCDs for the period of 12 – 17 Sep. 2016. Finally, on 19 Sep. 2016, a lamp experiment was performed, and the data showed a relatively large improvement in RMS values from 20 Sep. 2016 onward. After extensive check and quality control, the retrieved slant columns were only submitted for a limited number of days.