Supplement: Intercomparison of MAX-DOAS vertical profile retrieval algorithms: studies on field data from the CINDI-2 campaign

Jan-Lukas Tirpitz[1], Udo Frieß[1], François Hendrick[2], Carlos Alberti[3,a], Marc Allaart[4], Arnoud Apituley[4], Alkis Bais[5], Steffen Beirle[6], Stijn Berkhout[7], Kristof Bognar[8], Tim Bösch[9], Ilya Bruchkouski[10], Alexander Cede[11,12], Ka Lok Chan[3,b], Mirjam den Hoed[4], Sebastian Donner[6], Theano Drosoglou[5], Caroline Fayt[2], Martina M. Friedrich[2], Arnoud Frumau[13], Lou Gast[7], Clio Gielen[2,c], Laura Gomez-Martín[14], Nan Hao[15], Arjan Hensen[13], Bas Henzing[13], Christian Hermans[2], Junli Jin[16], Karin Kreher[18], Jonas Kuhn[1,6], Johannes Lampel[1,19], Ang Li[20], Cheng Liu[21], Haoran Liu[21], Jianzhong Ma[17], Alexis Merlaud[2], Enno Peters[9,d], Gaia Pinardi[2], Ankie Piters[4], Ulrich Platt[1,6], Olga Puentedura[14], Andreas Richter[9], Stefan Schmitt[1], Elena Spinei[12,e], Deborah Stein Zweers[4], Kimberly Strong[8], Daan Swart[7], Frederik Tack[2], Martin Tiefengraber[11,22], René van der Hoff[7], Michel van Roozendael[2], Tim Vlemmix[4], Jan Vonk[7], Thomas Wagner[6], Yang Wang[6], Zhuoru Wang[15], Mark Wenig[3], Matthias Wiegner[3], Folkard Wittrock[9], Pinhua Xie[20], Chengzhi Xing[21], Jin Xu[20], Margarita Yela[14], Chengxin Zhang[21], and Xiaoyi Zhao[8,f]

[1]Institute of Environmental Physics, University of Heidelberg, Heidelberg, Germany
[2]Royal Belgian Institute for Space Aeronomy, Brussels, Belgium
[3]Meteorological Institute, Ludwig-Maximilians-Universität München, Munich, Germany
[4]Royal Netherlands Meteorological Institute (KNMI), De Bilt, The Netherlands
[5]Laboratory of Atmospheric Physics, Aristotle University of Thessaloniki, Thessaloniki, Greece
[6]Max Planck Institute for Chemistry, Mainz, Germany
[7]National Institute for Public Health and the Environment (RIVM), Bilthoven, The Netherlands
[8]Department of Physics, University of Toronto, Toronto, Canada
[9]Institute for Environmental Physics, University of Bremen, Bremen, Germany
[10]Belarusian State University, Minsk, Belarus
[11]LuftBlick Earth Observation Technologies, Mutters, Austria
[12]NASA-Goddard Space Flight Center, USA
[13]Netherlands Organisation for Applied Scientific Research (TNO), Utrecht, The Netherlands
[14]National Institute of Aerospatial Technology (INTA), Madrid, Spain
[15]Remote Sensing Technology Institute, German Aerospace Center (DLR), Oberpfaffenhofen, Germany
[16]Meteorological Observation Centre, China Meteorological Administration, Beijing, China
[17]Chinese Academy of Meteorology Science, China Meteorological Administration, Beijing, China
[18]BK Scientific GmbH, Mainz, Germany
[19]Airyx GmbH, Justus-von-Liebig-Straße 14, 69214 Eppelheim, Germany
[20]Anhui Institute of Optics and Fine Mechanics, Chinese Academy of Sciences, Hefei, China
[21]School of Earth and Space Sciences, University of Science and Technology of China, 230026, Hefei, China
[22]Department of Atmospheric and Cryospheric Sciences, University of Innsbruck, Innsbruck, Austria
[a]now at Institute of Meteorology and Climate Research (IMK-ASF), Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany
[b]now at Remote Sensing Technology Institute (IMF), German Aerospace Center (DLR), Oberpfaffenhofen, Germany
[c]now at Institute for Astronomy, KU Leuven, Belgium

[d]now at Institute for Protection of Maritime Infrastructures, Bremerhaven, Germany
[e]now at Virginia Polytechnic Institute and State University, Blacksburg, VA, USA
[f]now at Air Quality Research Division, Environment and Climate Change Canada, Canada

Correspondence: Jan-Lukas Tirpitz (jan-lukas.tirpitz@iup.uni-heidelberg.de)

S1 M^3 algorithm description

The M^3 algorithm developed at the Ludwig-Maximilians-University in Munich is an OEM algorithm with Newton Gauss optimization. The radiative transfer model LibRadTran (Mayer and Kylling, 2005) serves as forward model for the retrieval. LibRadTran provides several radiative transfer equation solvers which can handle both pseudo spherical and full spherical geometry. The Jacobian is calculated numerically using the finite difference method, while the box air mass factors for trace gas profile retrievals are calculated using the Monte Carlo module of LibRadTran (MYSTIC). The M^3 retrieval used in the CINDI-2 campaign is a modified version of the algorithm described in detail in Chan et al. (2019) with the iterative optimization of the *a priori* profile disabled. Aerosol extinction profiles are retrieved in the linear space.

S2 ~~Partial AOT correction~~

~~As described in the main text Sect. 3.4, ground-based MAX-DOAS observations have limited sensitivity to aerosol at higher altitudes and thus the profile is strongly biased by *a priori* assumptions here. In the case of OEM algorithms, this effect can be accounted for by applying what in the following is referred to by "partial AOT correction" (PAC). This Sect. demonstrates the impact of the PAC on the comparison between MAX-DOAS and sun photometer data. The left panel of Fig. **??** shows an example of an extreme case during the campaign from September 15^{th}, 15:00h. Shown are a ceilometer backscatter profile (, black) and the same profile smoothed by the MAX-DOAS median OEM averaging kernels for Aerosol UV and Aerosol Vis (, blue and green), respectively. In this particular case it is expected that a large fraction of the aerosol above 1 km altitude will not be detected by MAX-DOAS instruments, resulting in factors $f_\tau = \frac{\tau_s'}{\tau_s}$ of 0.67 and 0.78, for the UV and the Vis AOT, respectively. However, a part of the high-altitude aerosol appears to be shifted to lower altitudes here by the retrieval. The right panels in Fig. **??** show information for the UV and the Vis retrieval (2nd and 3rd columns) over the whole campaign. The mean values are $f_\tau = 0.81 \pm 0.16$ (0.9 ± 0.13) for the UV (Vis) AOTs. It shall be pointed out that for OEM algorithms the necessity for the PAC can be reduced by using improved *a priori* profiles and covariances (e.g. from climatologies and/or model data). Left panel: example for the smoothing of a ceilometer backscatter profile $\mathbf{x}$ (according to Eq. (9) in the main text) with particularly heavy aerosol load at high-altitudes retrieved in the UV and Vis, respectively. Right panel: distribution and impact of the correction factor $f_\tau = \tau_s'/\tau_s$ for the UV and the Vis retrieval. On the top, the distributions of f_τ are shown with the solid lines indicating the mean values. At the bottom the correlation plots between sun photometer and MAX-DOAS median AOTs are shown. Red circles represent sun photometer total AOTs A, other dots represent the partial AOT $\tau_s = f_\tau \cdot \tau_s'$.~~

S2 O_4 scaling factor

By some groups, the O_4 scaling factor SF is applied to the measured dSCDs before the profile inversion. Initial motivation for its application are previous MAX-DOAS retrieval studies (e.g Wagner et al., 2009; Clémer et al., 2010) which report on a significant mismatch between measured and simulated dSCDs and/ or between MAX-DOAS integrated aerosol extinction and simultaneously measured sun photometer AOT that could not yet be explained (Wagner et al., 2019; Ortega et al., 2016). Meanwhile a series of studies use an SF, empirically determined to values between 0.75 and 0.9.

As described in Sect. 2.1.3, in this study no scaling of O_4 measured dSCDs was applied, except for MPIC-mp0.8 (MAPA algorithm with $SF = 0.8$). For CINDI-2, a SF of 0.8 was observed to enhance the number of valid profiles retrieved by MAPA and to significantly improve the agreement to the sun photometer total AOT in particularly in the UV (Beirle et al. (2019) and within this study). However, as mentioned before, for MAPA (as a parametrized approach without *a priori* profile and AVKs), a PAC as described in Sect. 3.4, cannot be correctly applied and thus deviations to the sun photometer are expected. To further investigate the impact and necessity of the SF for CINDI-2 retrievals, also HEIPRO (as an OEM retrieval) was run with different SFs. In Fig. S1 the impacts of the SF and the PAC on the agreement between MAX-DOAS profiling results (of HEIPRO and MAPA) and the sun photometer are directly compared. Application of $SF = 0.8$ or the PAC, respectively, lead to a very similar improvement in the agreement (regarding RMS), while the application of both together results in a clear overcompensation. This suggests that the PAC and $SF = 0.8$ are equivalent to a large extent and that in the case of MAPA the SF is a way to at least partly account for high-altitude aerosol when it comes to retrieving total AOTs. On the other hand a closer look reveals that if only the PAC is applied, a systematic negative offset of ≈ -0.04 remains in the correlation (for both algorithms and also other participants, compare to main text Sect. 3.4). Indeed, the top row of Fig. S3 shows that for HEIPRO the best RMSD is observed for $SF = 0.92 \pm 0.02$ (UV). Regarding Aerosol Vis (Figure S2), the impact of high aerosol is smaller due to the enhanced vertical sensitivity range (see AVKs in main text Sect. 3.1), such that applying the same scaling factor $SF = 0.8$ as for the UV (without PAC) should already lead to an overcompensation. Indeed, this is observed for HEIPRO as well as MAPA. The bottom row of Fig. S3 shows that, in contrast to the UV spectral range, the best agreement between sun photometer and MAX-DOAS in the Vis is observed for an $SF > 1$.

The second indicator for the need of a scaling factor is a significant mismatch between modelled and measured dSCDs. Figure S4 shows, that for HEIPRO the application of an $SF < 1$ indeed improves the agreement between measured and modelled O_4 UV dSCDs by up to $35\,\%$ in RMSD under clear-sky conditions. Modelled dSCDs are systematically lower than the measured dSCDs in particular for higher elevation angles. ~~Finally~~Also, Wagner et al. (2009) reported ~~, that~~that, under low aerosol conditions, measured dSCDs sometimes even significantly exceed dSCDs modelled within an aerosol free atmosphere, where O_4 dSCDs are ~~expected to be~~ close to the largest possible (regarding clear-sky scenarios only). Therefore, the median measured dSCDs during CINDI-2 at low aerosol load ($\tau_s < 0.1$) were compared to a set of aerosol free modelled dSCDs, showing that this did not happen during CINDI-2 except for the two highest elevation angles $\alpha = 15°$ and $30°$.

Finally, as already stated by Beirle et al. (2019) applying $SF = 0.8$ to MAPA leads to an increased number of valid profiles (see Sect. S3), which again indicates that scaling brings the RTM closer to reality.

The conclusions drawn are as follows: Even without O_4 dSCD scaling, reasonable results and agreement with supporting observations are achieved (if a PAC is applied). In general, a scaling factor of 0.8 seems to be too small but might at least partly be used to account for high-altitude aerosol for algorithms, that cannot quantify their sensitivity or the assimilated *a priori* information. However, there are indications that a less extreme scaling ($0.8 < SF < 1.0$) might in general improve the retrieval. Finally, we think that for this study the prescribed $SF = 1.0$ is justified. Even though it might not be ideal, it is the most straightforward approach and yields reasonable and consistent results within the uncertainties introduced by other factors. To draw more concise conclusions, further studies similar to Wagner et al. (2019) are necessary.

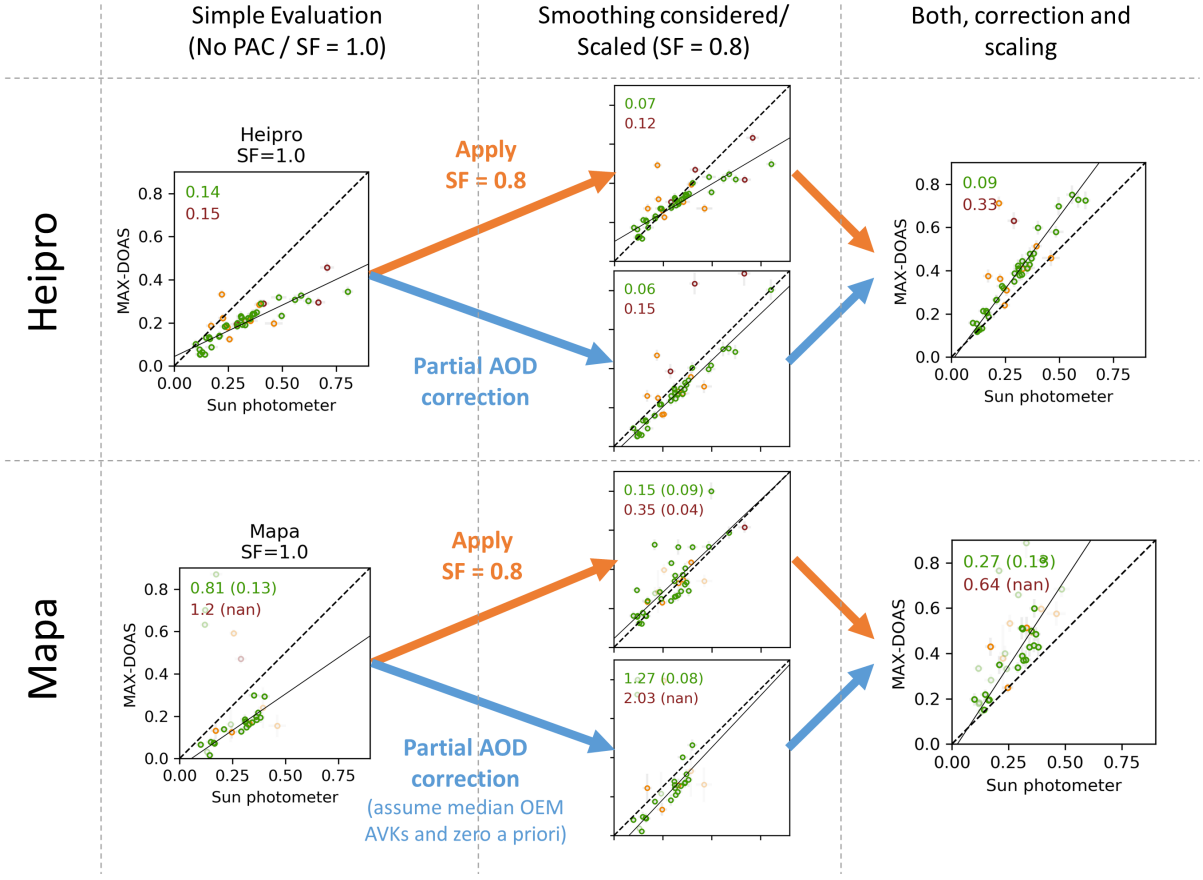

Figure S1. The impact of $SF = 0.8$ and PAC on the agreement between sun photometer and MAX-DOAS AOT (Aerosol UV) in the case of HEIPRO (OEM approach, top row) and MAPA (parametrized approach, bottom row) in a direct comparison. Axes limits and labels of the plots on the left apply for all plots in the figure. Left column: A standard retrieval with $SF = 1$ yields a clear underestimation of the sun photometer AOT for both algorithms. Middle column: Applying the PAC or $SF = 0.8$ leads to a significant improvement. Right column: Applying both leads to overcompensation. Note, that the PAC for MAPA incorporates vague assumptions (median AVKs from all OEM algorithms and $\boldsymbol{x_a} = \mathbf{0}$) and is therefore less meaningful.

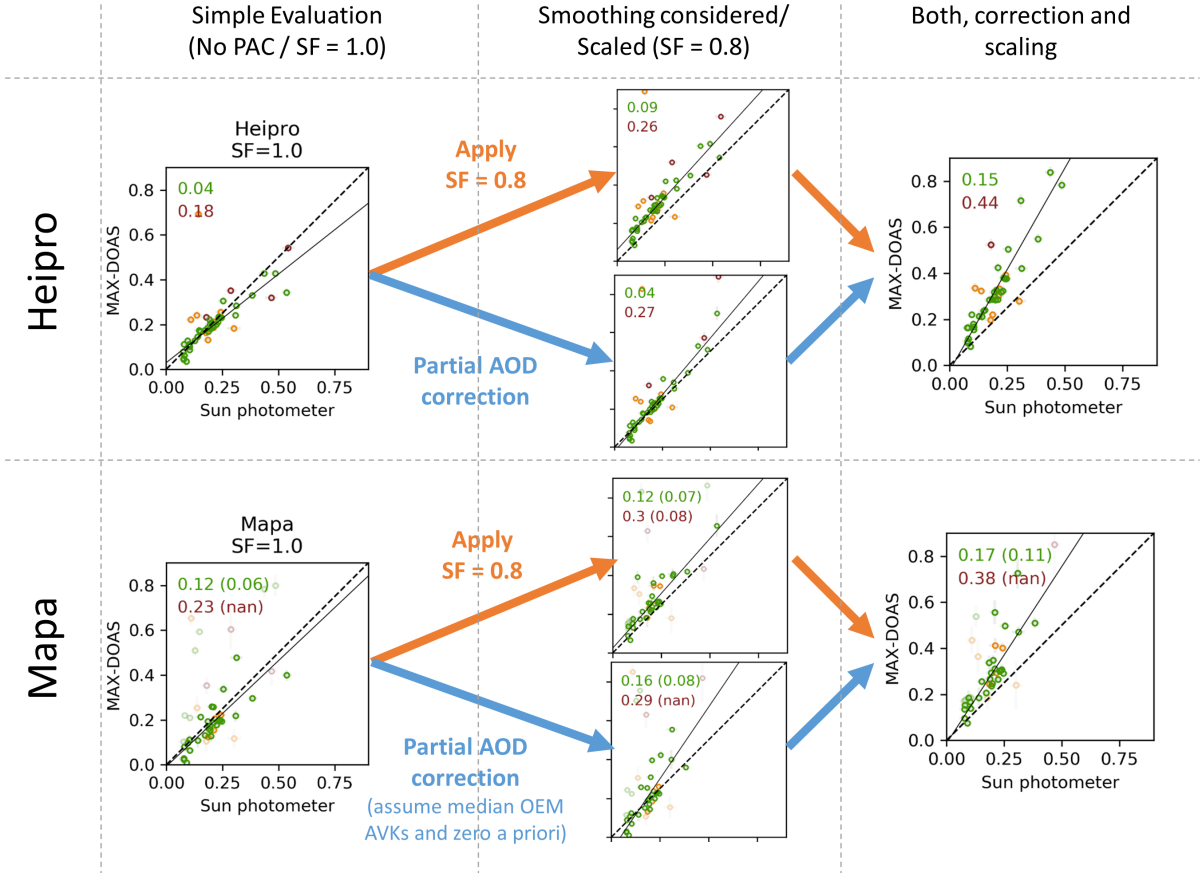

Figure S2. The impact of SF and PAC on Aerosol Vis results. Description of Fig. S1 applies. Due to the extended vertical sensitivity range of the MAX-DOAS observations (compare to main text Sect. 3.1), the effect of high-altitude aerosol is less significant. While this is accounted for in the PAC, the application of the same $SF = 0.8$ as for Aerosol UV leads to an overcompensation here already without additionally applying the PAC.

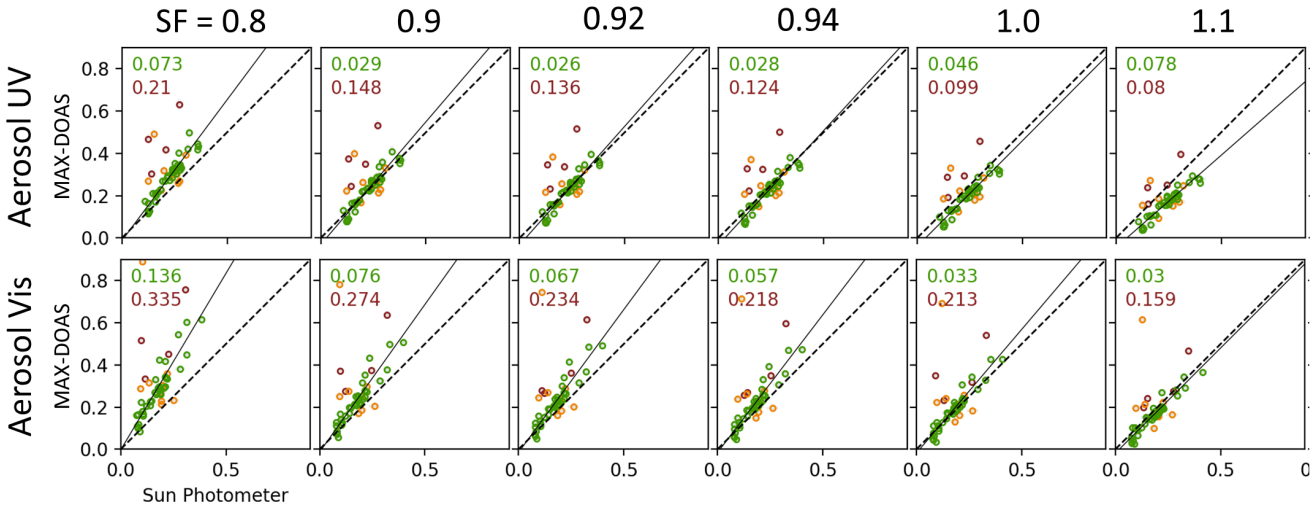

Figure S3. Correlation between the sun photometer AOT (PAC applied) and the HEIPRO algorithms for different SF for Aerosol UV (top row) and Aerosol Vis (bottom row). Symbol colours differentiate between clear-sky (green), and cloudy conditions (orange, red). Numbers in the upper left corners show the corresponding RMSD values. Interestingly, the best agreement for the two spectral ranges is found for different scaling factors, namely for $SF = 0.92 \pm 0.02$ (UV) and $SF > 1 \pm 0.02$ (Vis). The thin line depicts the linear fit for clear-sky data only.

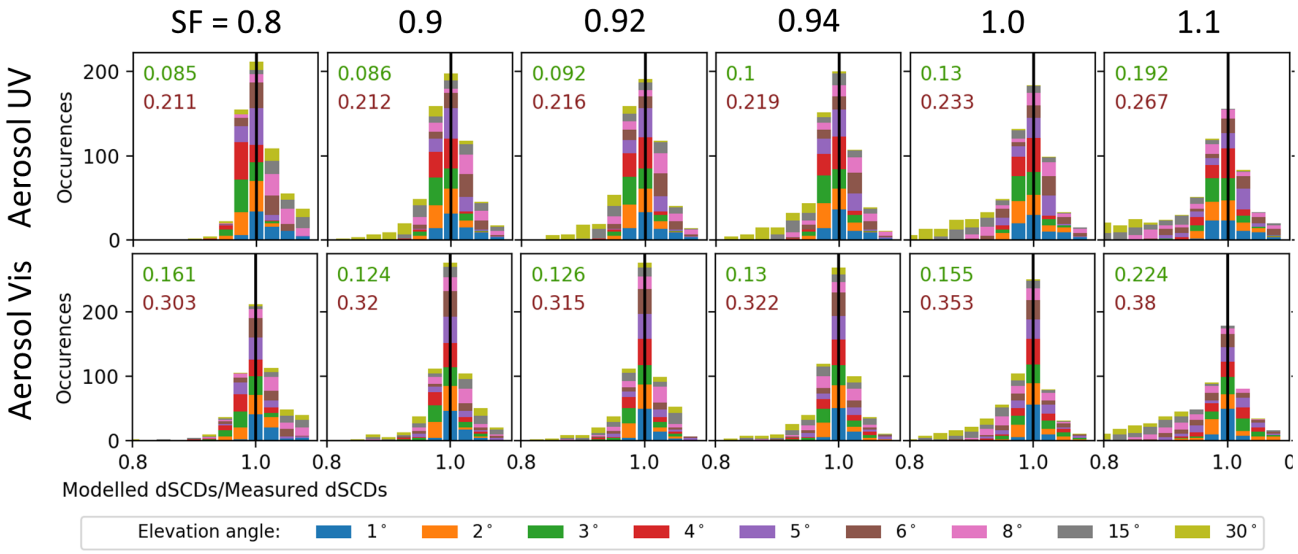

Figure S4. Agreement between modelled and measured dSCDs (depicted as histograms of the ratio) for HEIPRO with different SF for Aerosol UV (top row) and Aerosol Vis (bottom row). Here, only clear-sky and valid data are considered. Numbers in the upper left corners show the corresponding RMSD values. There is a clear tendency, that the agreement improves for $SF < 1$.

S3 Details on profile flagging

Participants were allowed to submit flags, giving them the opportunity to mark profiles as invalid. Four participants submitted flags based on the following criteria:

- BIRA/ bePRO: Profiles are considered valid if the retrieved degrees of freedom are > 1 and if the difference between measured and modelled dSCDs is smaller than $30\,\%$.
- INTA: Profiles are valid if DOFS > 1 and if the RMSD between measured and simulated dSCDs is smaller than 1.5 times the daily averaged RMS.
- KNMI: Profiles are invalid if the spread in the ensemble of solutions for AOT or the NO_2 or HCHO tropospheric column is larger than $15\,\%$ of the retrieved value, or if there are less than 5 out of the ensemble of 20 retrievals for which a solution was found.
- MAPA: Flagging is based on a row of criteria (e.g. the agreement of modelled and measured dSCDs, implausible results ~~,~~ or the consistency within the ensemble of possible solutions~~, ...~~) with carefully chosen thresholds. A detailed description can be found in Beirle et al. (2019).

Table S1 shows the statistics regarding the number of submitted profiles and the fraction of invalid profiles for all participants and species.

Table S1. Submission and flagging statistics.

		HCHO		NO_2 UV		NO_2 Vis		Aerosol UV		Aerosol Vis	
		Total	Valid [%]	Total	Valid [%]	Total	Valid [%]	Total	Valid [%]	Total	Valid [%]
bePRO	AUTH^o	170	100	170	100	-	-	170	100	-	-
	BIRA^o	170	93	170	93	170	87	170	93	170	88
	INTA^o	170	78	170	75	170	71	170	87	170	80
PRIAM	AIOFM^o_l	170	100	170	100	-	-	170	100	-	-
	BSU^o_l	170	100	170	100	170	100	170	100	170	100
	CMA^o_l	169	100	169	100	169	100	169	100	169	100
	MPIC^o_l	170	100	170	100	170	100	170	100	170	100
HEPRO	IUPHD^o_l	170	100	170	100	170	100	170	100	170	100
	UTOR^o_l	-	-	170	100	170	100	170	100	170	100
BOREAS	IUPB^o_l	170	100	170	100	170	100	170	100	170	100
M^3	LMU^o	170	100	170	100	170	100	170	100	170	100
MMF	BIRA^o_l	170	100	170	100	170	100	170	100	170	100
Realtime	NASA^a	170	100	170	100	170	100	-	-	170	100
MARK	KNMI^p	107	38	152	61	168	76	107	43	168	77
MAPA	MPIC-1.0^p	170	31	170	32	170	22	170	32	170	33
	MPIC-0.8^p	170	52	170	51	170	37	170	52	170	43

S4 Further details on supporting observations

S4.1 Aerosol extinction profiles

The available raw data from the ceilometer are attenuated backscatter coefficient profiles $\tilde{\beta}(h)$. For altitudes below 180 m, data are invalid due to insufficient overlap between sending and receiving telescope's FOVs of the instrument, thus $\tilde{\beta}(h < 180\,\mathrm{m})$ was set to $\tilde{\beta}(h = 180\,\mathrm{m})$. For profiles with simultaneously available sun photometer AOTs $\tau_s(\lambda)$, attenuated backscatter profiles $\tilde{\beta}(h)$ were converted to approximate backscatter profiles $\beta(h)$, applying

$$\beta(h) = \tilde{\beta}(h) \cdot \exp\left[2 \int_0^h \frac{\tau_s(\lambda = 1064\,\mathrm{nm})}{\int \tilde{\beta}(h')dh'} dh\right] \tag{1}$$

Extinction coefficient profiles $\alpha_\lambda(h)$ at the MAX-DOAS retrieval wavelengths were then obtained by scaling of $\beta(h)$ with the sun photometer AOT according to:

$$\alpha_\lambda(h) = \frac{\tau_s(\lambda)}{\int \beta(h')dh'} \cdot \beta(h) \tag{2}$$

Integrands with no specified limits in Eq. (1) and (2) indicate integration over the entire available profile. Values for $\tau_s(\lambda)$ at the desired wavelengths were derived according to Eq. (4) in the main text. In case of missing sun photometer data, for instance

due to clouds, the MAX-DOAS retrieved AOT was used instead of τ_s. In this case no attenuation correction (Eq. (1)) could be applied and the integration in Eq. (2) was performed over an averaging kernel smoothed profile (see main text Sect. 2.3.2), to take into account the blindness of MAX-DOAS instruments for higher aerosol layers.

The resulting extinction coefficient profiles at 360 nm could partly be validated with Raman lidar observations at 355 nm (the CESAR Water Vapor, Aerosol and Cloud lidar "CAELI", operated within the European Aerosol Research lidar Network (EARLINET, Bösenberg et al., 2003; Pappalardo et al., 2014) and described in detail in Apituley et al., 2009). Since for the Raman lidar there is not sufficient telescope FOV overlap for altitudes < 1 km to retrieve reliable extinction profiles this comparison is limited to the altitude range between 1 and 4 km. The average RMSD between scaled ceilometer and Raman lidar profiles is ≈ 0.03. Table S2 summarizes the instrument's properties. Figure S5 shows the available Raman lidar profiles in comparison to the ceilometer derived profiles scaled with the respective sun photometer and MAX-DOAS AOT respectively.

Table S2. Properties of the two lidar instruments.

Instrument	Raman lidar	Ceilometer
Data product	Aerosol extinction profile	Elastic backscatter profile
Operational wavelength	355 nm	1064 nm
Altitude range	1 to 10 km	0.2 to 15 km
Vertical resolution	7.5 m	10 m
Temporal resolution	30 s	12 s
Data coverage	5 profiles between 13.9. and 15.9.	Whole campaign

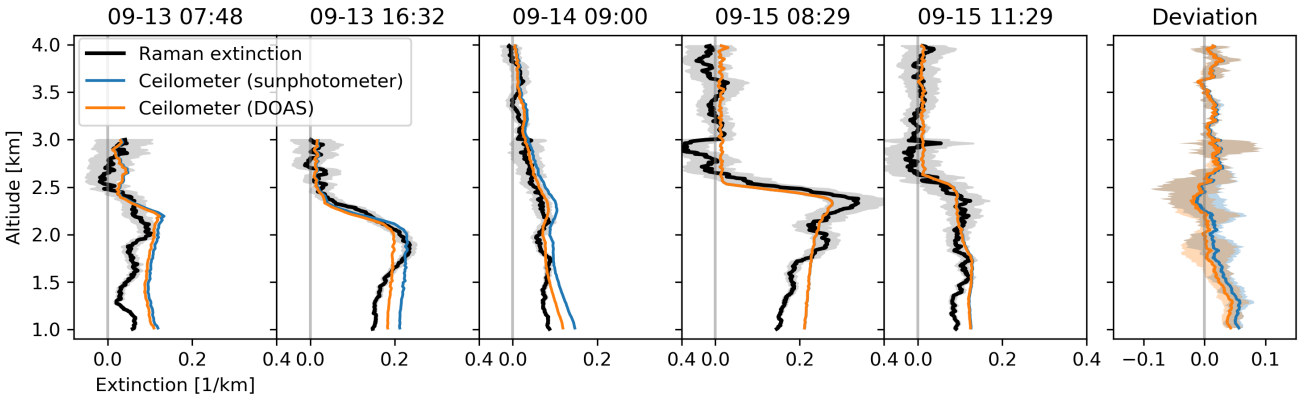

Figure S5. Comparison of the five available aerosol extinction profiles obtained from Raman lidar measurements (in black, with uncertainties indicated by the grey areas) with AOT scaled ceilometer backscatter profiles. Scaling was performed with AOTs from sun photometer (blue) and DOAS data (orange). In the last plot, lines show the mean deviation, whereas the borders of the couloured areas mark maximum and minimum deviation.

S4.2 Radiosonde flights

An overview of the radiosonde flights is given in Fig. S6. Example profiles are shown in the course of a comparison between lidar and radiosonde observations in Supplement S4.5. Another preprocessing step of the sonde profiles shall be mentioned: data quality was affected by calibrational drifting of the sensor, as it was optimized for low weight and cost rather than performance. Even though a calibration against the CE-DOAS was performed directly before each launch, most profiles showed a clear instrumental offset of about $(1-2)\times10^{10}\,\mathrm{molec\,cm^{-3}}$ in the free troposphere. The offset was subtracted and the profiles were subsequently rescaled to their initial surface concentration.

Table S3. Overview over the radiosonde sampling flights shown in this study.

Launch date	Flight time[a] [min]	Travel distance[a] [km]	Wind direction
9-13 08:42	10	7	SE
9-14 09:03	12	5	SE
9-14 13:06	14	4	SE
9-15 08:04	10	8	E
9-15 10:25	11	8	SE
9-21 07:57	12	10	SE
9-21 10:14	15	5	SE
9-25 06:59	17	7	S
9-25 09:29	12	18	S

[a] Only considering trajectory through the lowest 4 km of the atmosphere.

S4.3 NO_2 lidar

The NO_2 lidar provides profiles consisting of a series of altitude intervals or "boxes" with constant gas concentration between a lower and an upper altitude limit. The conversion to profiles on the MAX-DOAS 200 m grid is demonstrated in Fig. S7. First, the boxes were converted to a continuous profile by linearly interpolating over box overlaps or gaps, which was then averaged down to the 200 m MAX-DOAS retrieval grid resolution.

S4.4 Long path DOAS

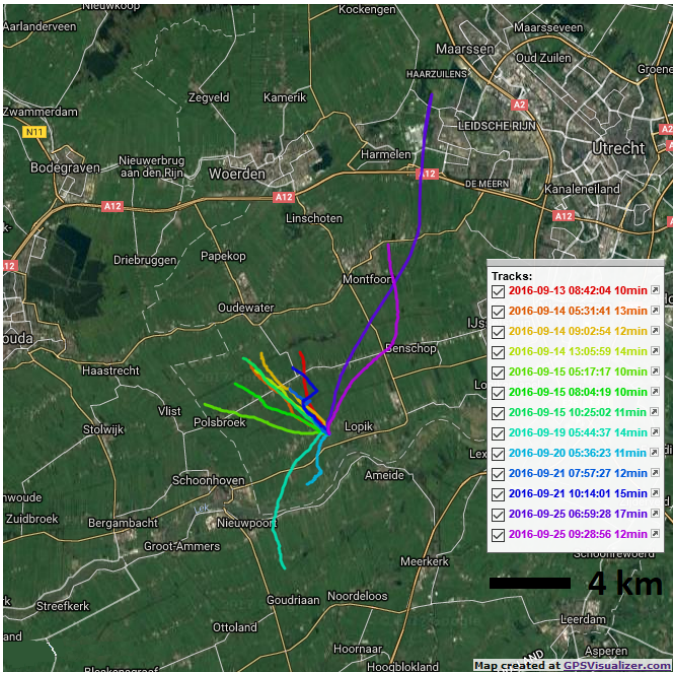

Figure S6. Sonde flight paths in the course of the campaign. Only data of the sonde ascent through the lowest 4 km of the atmosphere are shown. Some flights shown here are not included in the comparison as they were launched before 6:30 h. The "minute" values in the legend labels represent the flight time.

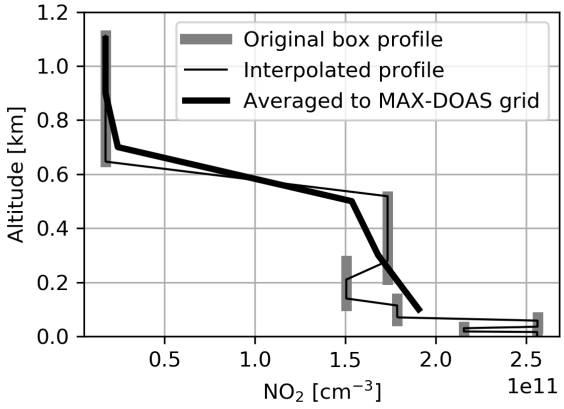

Figure S7. Regridding of an example NO_2 lidar box profile (27 September 2016, 11:00) to the MAX-DOAS 200 m vertical resolution. In a first step, gaps and overlaps within the box profile are linearly interpolated. The resulting profile is then averaged within the MAX-DOAS retrieval layers.

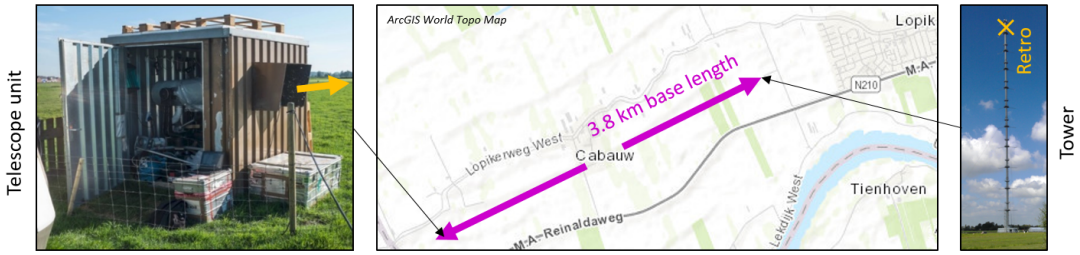

Figure S8. Setup of the LP-DOAS system. As shown on the map (Esri et al., 2018), the light sending and receiving telescope unit (left) was located at 3.8 km distance to the meteorological tower (right), resulting in a total light path of 7.6 km. There were several retroreflectors installed on the tower at different altitudes. However for this study, only the one at the very top (207 m altitude) was used to obtain the average gas in the lowest retrieval layer, extending from 0 to 200 m altitude.

S4.5 Consistency of supporting observations

The agreement of redundant supporting observations (processed as described in the main text Sect. 2.2) gives an impression of their reliability and/ or representativeness. In the case of NO_2 several observations of total vertical columns and surface concentration (note again, that throughout this paper "surface concentration" refers to the average concentration in the lowest retrieval layer) are available and compared in the correlation plots in Fig. S9 below. Corresponding time series plots are already shown in the main text in Fig. 17 and 20, respectively. Tables S4 and S5 show the RMSD (as observed) and σ (the expected deviation according to the specified measurement uncertainties) between each possible pair of observations. For the VCDs, the RMSD is close to σ or below. A maximum RMSD of 1.5 σ is found between NO_2 lidar and direct-sun DOAS. For the surface concentrations however, there seem to be systematic deviations which split the observations into two pairs: radiosonde and lidar observations agree well but are both systematically lower than LP-DOAS and tower measurements. MAX-DOAS UV agrees better with LP-DOAS and tower observations, while MAX-DOAS Vis agrees more with sonde and lidar. Between the LP-DOAS and the NO_2 lidar, an RMSD of more than 4 σ is observed. There are several potential explanations:

1. Biases are introduced due to data processing (temporal and spatial regridding, for instance for the lidar profiles described in Sect. S4.3).
2. Spatio-temporal variability of the real gas abundances
3. Imperfect estimates of the measurement uncertainties (in particular systematic deviations)

For the NO_2 lidar and the radiosondes, there are four simultaneously recorded NO_2 profiles available over the campaign (simultaneous in the sense that for a single MAX-DOAS profile timestamp, profiles from both systems are available according to the definitions in Sect. 2.2.3 in the main text). They are compared in the top row of Fig. S10. For the first situation, where good spatial and temporal overlap is given, there is mostly an agreement within the specified errors. In the case of bad temporal and/ or spatial overlap, strong deviations occur. For the 2nd and the 4th plot, there are several lidar profiles available, which

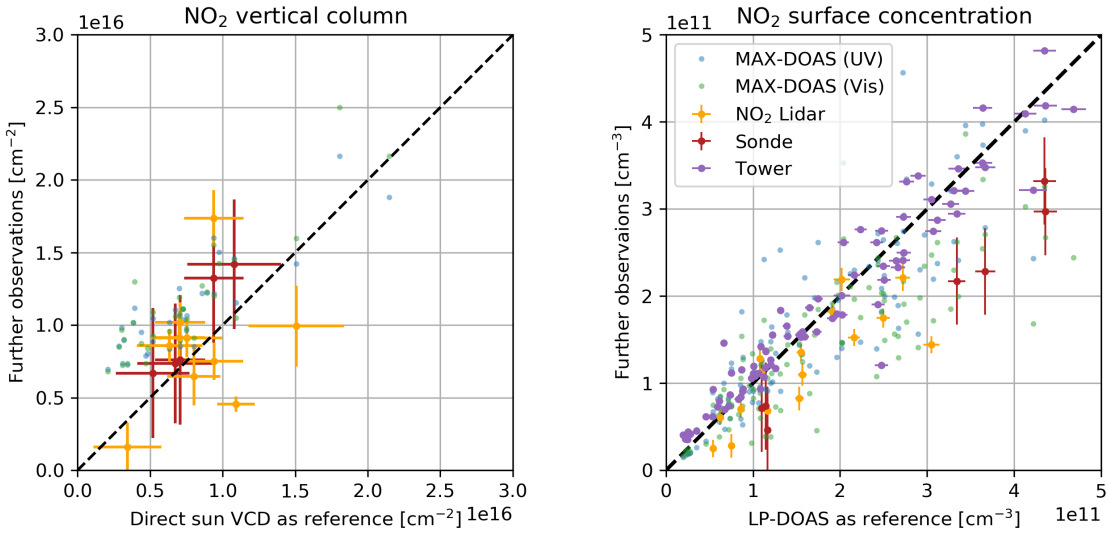

Figure S9. Comparison of redundant supporting observations of NO_2 VCDs (left panel) and surface concentration (right panel). MAX-DOAS retrieved values are plotted in the background. To improve visibility, tower measurement uncertainties (vertical error bars of typically $(6.0 \pm 0.5) \times 10^{10}\,\mathrm{molec\,cm^{-3}}$) are not shown.

are temporally closer to the radiosonde (however further away from the corresponding MAX-DOAS profile), which in contrast show very good agreement again. This shows that the real NO_2 profile varies strongly even on timescales of ≈ 30 minutes (see also (Peters et al., 2019)) and that improved synchronisation between MAX-DOAS and supporting observations should be considered for future campaigns.

Table S4. Comparison of redundant measurements of the NO_2 surface concentration (in $10^{11}\,\mathrm{molec\,cm^{-3}}$). For each pair of observations, the observed scatter (RMS) is compared to the specified uncertainty (σ).

	Tower in-situ (0.56)		Radiosonde (0.50)		NO_2-Lidar (0.13)	
	RMSD	σ	RMSD	σ	RMSD	σ
LP-DOAS (0.06)	0.32	0.56	1.01	0.51	0.57	0.13
NO_2-Lidar (0.13)	0.72	0.57	0.40	0.52	-	-
Radiosonde (0.50)	0.99	0.78	-	-	-	-

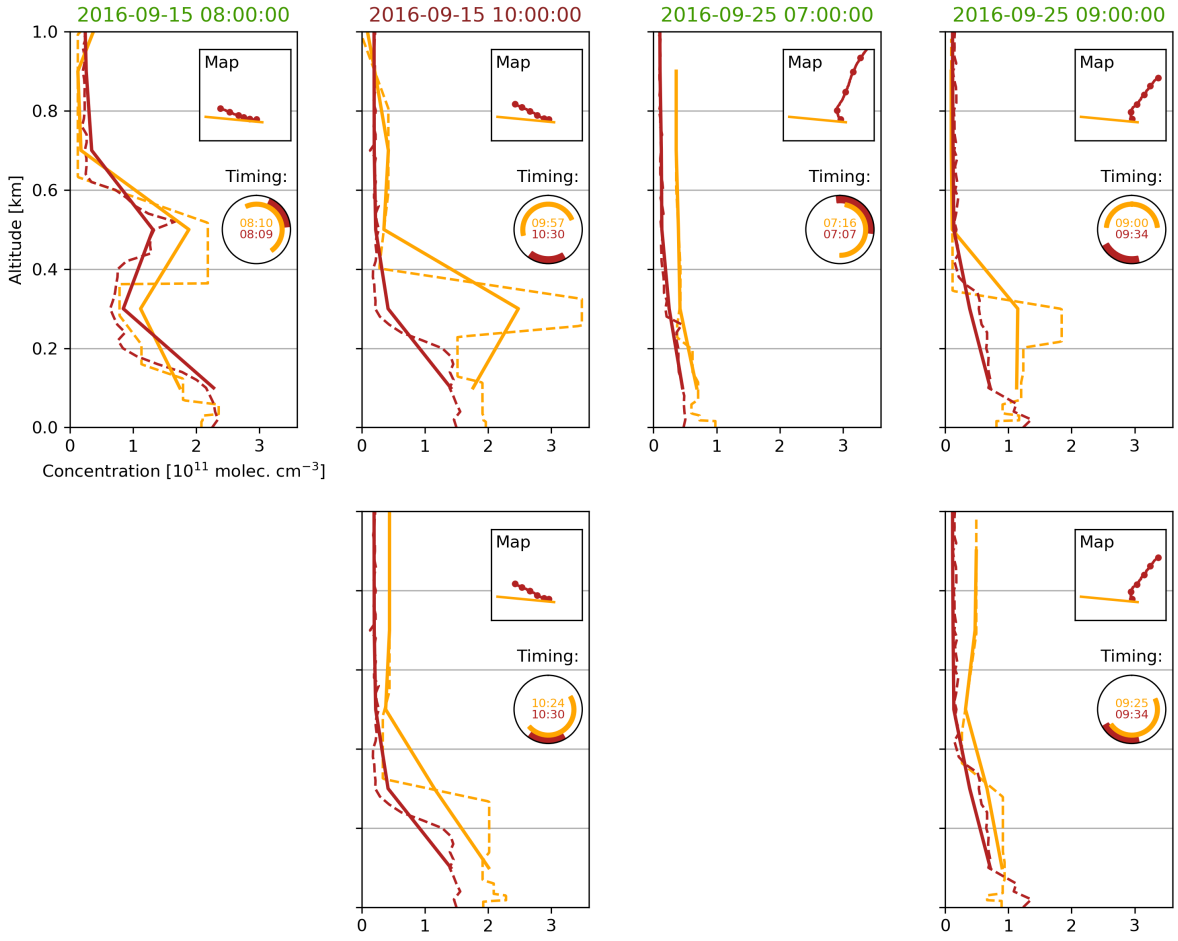

Figure S10. Top row: Comparison of NO_2 lidar (orange) and radiosonde (red) profiles, that were assigned to a common MAX-DOAS profile timestamp according to Sect. 2.2.3 in the main text. Dashed lines represent original instrument resolution while thick lines show the concentrations averaged to the MAX-DOAS altitude grid. The colour of the dates indicates the cloud conditions. The rectangular little subplots show a map (4 x 4 km) of the lidar line of sight and the sonde flight path. The dots on the sonde flight path mark the transitions between the retrieval layers. The polar plot (to be read like a clock) shows the temporal overlap between the two observations, together with the middle timestamps of each observation. Lower row: For the 2nd and 4th timestamp, there were lidar profiles available with improved temporal overlap (however, with a worse overlap with the corresponding MAX-DOAS profile).

Table S5. Comparison of redundant measurements of the NO_2 total columns (in $10^{16}\,\mathrm{molec\,cm^{-2}}$). For each pair of observation, the observed scatter (RMS) is compared to the specified uncertainty (σ).

	Radiosonde (0.44)		NO_2-Lidar (0.15)	
	RMSD	σ	RMSD	σ
Direct-sun DOAS (0.23)	0.24	0.51	0.40	0.26
NO_2-Lidar (0.15)	0.34	0.48	-	-

S5 MAX-DOAS viewing distance

Wagner and Beirle (2016) derived polynomial relationships between the "horizontal sensitivity range" (HSR, defined as the distance, at which the box airmass factors dropped to 1/e) and O_4 differential airmass factors (dAMF). Applying this approach to the CINDI-2 O_4 dAMFs yields the HSRs shown in Fig. S12. A constant vertical O_4 column of $1.19 \cdot 10^{43}\,\mathrm{molec^2\,cm^{-5}}$ was assumed. The HSR for the actually retrieved layers is more complicated and not assessed here, as information aspects (which elevation contributes to information on which layer), geometrical limitations and the atmospheric state (trace gas and aerosol layer height) would have to be taken into account. Depending on the conditions, HSRs vary between a few and tens of kilometres (as shown in Fig. S11) defining whether only air masses over rural areas and/or urban areas (Gouda at 15 km distance, Zoetermeer at 30 km distance and The Hague at 40 km distance. to the measurement site) are sampled. Further, depending on the wind, plumes of Utrecht, Rotterdam or Amsterdam might be sampled.

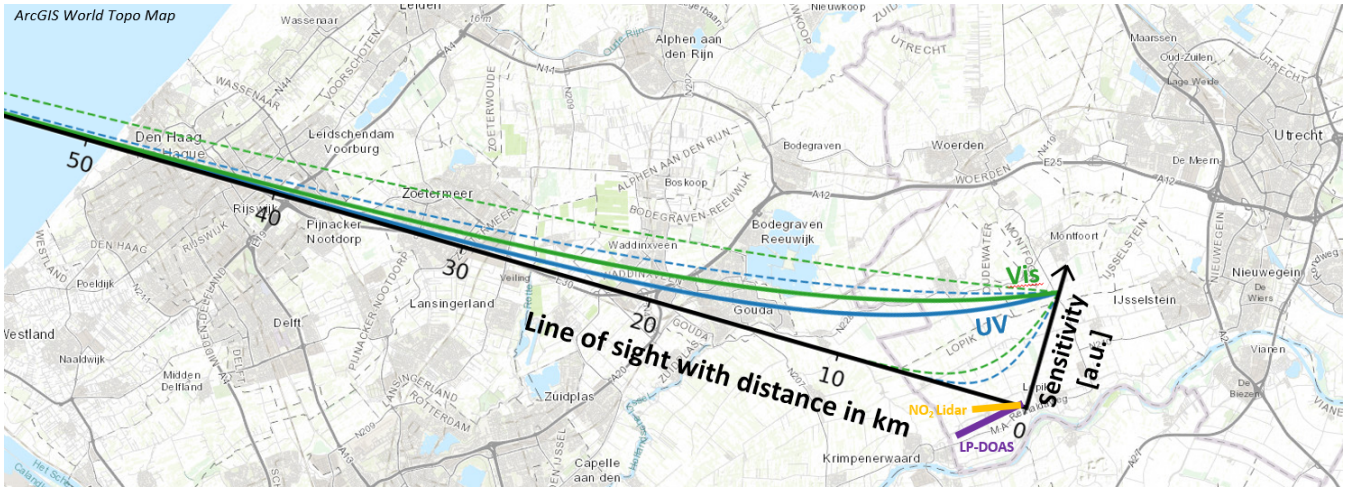

Figure S11. Line of sight of the MAX-DOAS instruments on a map (Esri et al., 2018). The coloured curves indicate the sensitivity for the ~~two extreme~~ average (~~shortest~~ solid), the minimum (dashed) and ~~longest~~maximum (dashed) viewing distances encountered during the campaign ~~. Dashed lines indicate~~ in the ~~1/e-length~~UV (blue) and Vis (green).

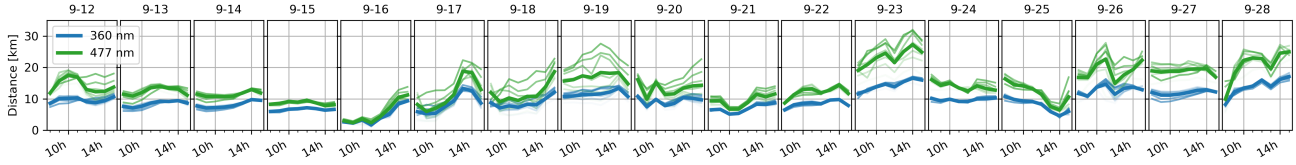

Figure S12. Viewing distance (HSR) of MAX-DOAS instruments during CINDI-2. It was calculated for different ~~elevations~~ elevation angles (1, 2, 3, 4, 5, 6 and 8° with increasing transparency of the curves) and the average value for UV and Vis (thick lines).

S6 ~~Spatio-temporal variability~~

~~For an impression of the impact of~~

S6 Spatio-temporal mismatch and variability

Given the MAX-DOAS horizontal sensitivity ranges determined in Sect. S5, approximate values for the spatio-temporal ~~variability on the comparisons, different~~ mismatch of MAX-DOAS and different supporting observations can be derived. They are given in Table S6. The potential impact of these mismatches can be demonstrated by means of the NO_2 surface concentration. The left panel of Fig. S13 shows observations of the NO_2 surface concentrations at their original temporal resolution Δt and integration time t_{int}~~are shown in the left panel of Fig. S13.~~ . The CE-DOAS as a point measurement with $\Delta t = t_{int} = 1\,\mathrm{min}$ shows very strong variability on short timescales. However, for the tower measurements (all in situ instruments in the tower vertically integrated as described in main text Sect. 2.2.5 at $\Delta t = 20\,\mathrm{min}$ and $t_{int} \approx 5\,\mathrm{min}$), the LP-DOAS

Table S6. Estimates for the average spatio-temporal mismatch of different supporting observations w.r.t. to the MAX-DOAS measurements. For the location of the MAX-DOAS observations the centers of mass of the horizontal sensitivity curves from Sect. S5 were used. For the location of sun photometer and direct-sun DOAS observations, the center of the lines of sight towards the sun up to 2 km atitude were considered.

Observation	Spatial mismatch [km]	Temporal mismatch [min]
Sun photometer	13	8
Ceilometer	11	0
Direct-sun DOAS	13	23
NO_2-Lidar	10	9
Radiosonde	6	13
LP-DOAS	10	6
In-situ in tower	11	0

($\Delta t = 32\,\mathrm{min}$, $t_{int} \approx 100\,\mathrm{s}$) there is already significant smoothing. The 1D-MAX-DOAS data was recorded by DLR (see Supplement S10), who retrieved profiles in the nominal azimuth direction (287°) more or less continuously ($\Delta t = 15\,\mathrm{min}$, $t_{int} \approx 10\,\mathrm{min}$). In all measurements there is significant variation on the sub-hour timescale. Further, spatial variability might be observed in the form of disagreement between UV and Vis observations of the 1D-MAX-DOAS as viewing distance and thus the sampled air volume changes between the two spectral ranges (see Supplement S5). To estimate the order of magnitude, the right panel of Fig. S13 shows a kind of autocorrelation of the total campaign time series of each observation. The RMSD between the original and a temporally shifted signal is calculated. ~~Note, that the~~ 1D MAX-DOAS Vis data is not ~~reliable as it features a lot of gaps . As described in main text Sect. 2.2, typical temporal shifts/ interpolation intervals between different observations during CINDI-2 are in the order of 30 minutes. RMSD values at 30 minutes shift are~~ shown, as multiple gaps in the data complicated the autocorrelation. Comparing this figure with values from Table S6 yields that spatio-temporal variability causes RMSD values of around $3.5 \times 10^{10}\,\mathrm{molec\,cm^{-3}}$ ~~and thus indeed of similar order as typical~~ in the NO_2 surface concentration, which is indeed of the order of the observed RMSD values in the NO_2 surface concentration comparisons within this study (approx. $5 \times 10^{10}\,\mathrm{molec\,cm^{-3}}$, compare to main text Fig. 22). ~~For an impression on~~

For another demonstration of the spatial variability, we refer to data from the IMPACT instrument (Peters et al., 2019), an imaging MAX-DOAS operated by IUP-Bremen (IUPB) which allows to perform elevation "scans" in different azimuth viewing directions in quick succession. During CINDI-2, the IMPACT performed full-azimuthal scans in 10° steps every 15 minutes. Figure S14 exemplarily shows the observed NO_2 Vis dSCDs at 4° elevation on the 20 September 2016 together with dSCDs measured by the IUPB standard MAX-DOAS instrument in the nominal azimuth direction (287°, compare main text Sect. 2.1). The red shaded area depicts the variation of the dSCD with azimuth viewing direction. In particular around local noon this variation is tremendous, exceeding a factor of five. Further investigation on this issue can be found in Peters et al. (2019).

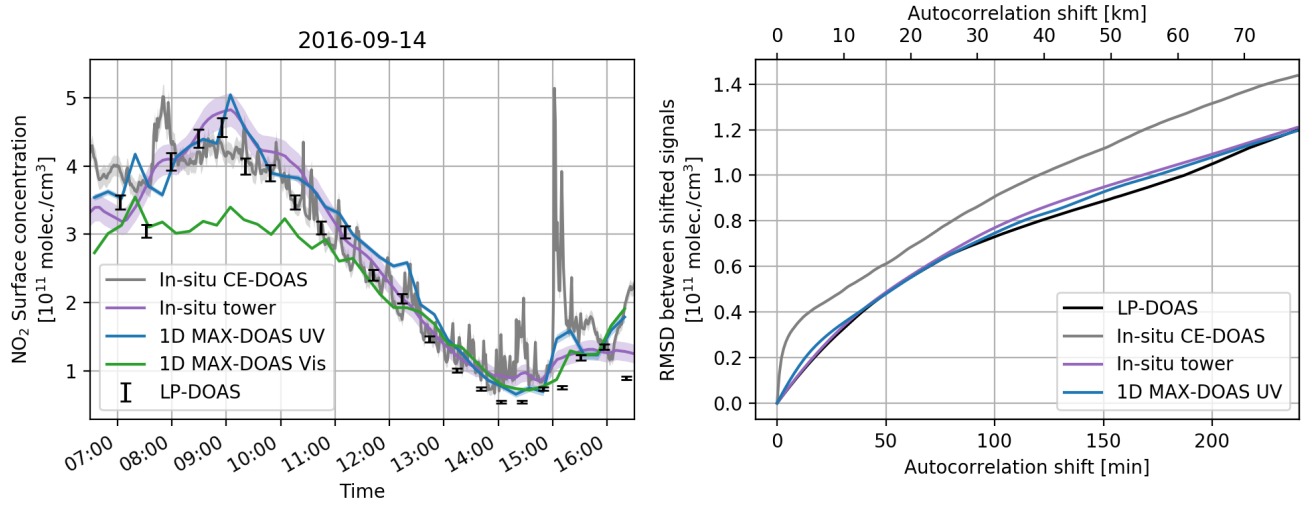

Figure S13. Left: Different observations of the NO_2 surface concentrations on 14 September 2016, each at its original temporal resolution to reveal short-term variations. Coloured areas behind the lines indicate the specified uncertainties. Right: RMSD values obtained from a kind of autocorrelation analysis over the whole campaign (night times excluded). For each observation, the RMSD between the original and a temporally shifted signal is calculated. The temporal shift (bottom horizontal axis) was varied between 0 and 4 hours. The temporal shift was roughly converted to its spatial equivalent by multiplication with the average observed wind speed in the surface layer ($\approx 5\,\mathrm{m/s}$), yielding the top horizontal axis.

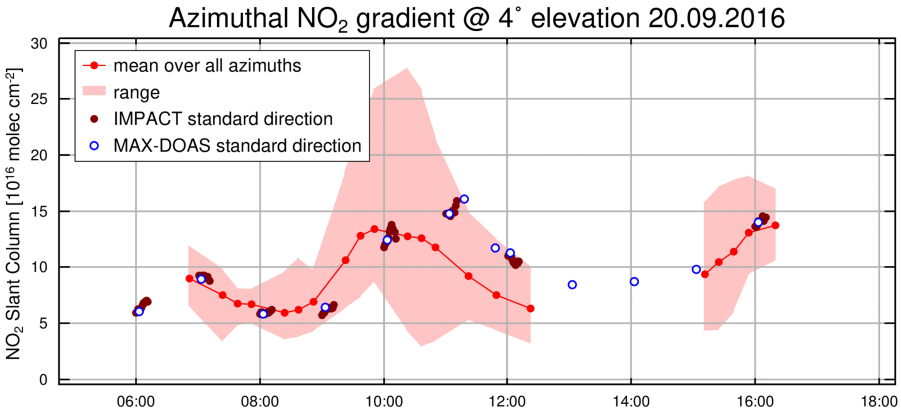

Figure S14. Variation in the NO_2 Vis dSCDs with different azimuth viewing directions at 4° elevation, as observed by the IMPACT imaging MAX-DOAS (Peters et al., 2019). Around local noon this variation is largest, exceeding a factor of 5. The time is UTC.

S7 Impact of the choice of pressure and temperature profiles for the RTMs

Pressure (p) and temperature (T) profiles used for the RTMs within this study are averaged sonde measurements performed in De Bilt by KNMI during September months of the years 2013-2015 (see main text Sect. 2.1.3). To estimate the effect of this approximation on the results, IUPHD/ HEIPRO retrieved an additional set of profiles, using p and T information from radiosondes launched at KNMI (De Bilt) during the campaign. Between one and three sondes were launched every day except on 16 September. For each profile inversion, the temporally closest sonde observation was used. Table S7 shows the difference in RMSD and Bias magnitude between these results and the "standard" results of IUPHD/ HEIPRO (that used the prescribed averaged p and T profiles from years before) relative to the average RMSDs and average Bias magnitude for all participants.

The impact on the dSCD comparison is less than 5% for both, RMSDs and Bias magnitudes. For AOTs, VCDs and surface concentrations, significant improvement ($> 10\,\%$ in RMSD) is only observed for HCHO surface concentrations ($17\,\%$) that contrasts with a deterioration for UV AOTs by $13\,\%$. The average improvement in RMSD for AOTs, VCDs and surface concentrations is $3.2\,\%$. The overall consistency between MAX-DOAS and supporting observations can thus be considered to remain similar, despite larger changes in some Bias magnitudes are observed (up to $51\,\%$ improvement for NO_2 Vis surface concentrations and up to $20\,\%$ deterioration for UV AOTs).

Table S7. The differences in RMSDs and Bias magnitudes for the IUPHD/ HEIPRO results arising from using daily p and T profiles, relative to the average RMSDs and Bias magnitudes assessed within the main study. Values are given for the comparisons of modelled and measured dSCDs ("dSCDs") and the comparisons against the supporting observations of AOTs, VCDs and surface concentrations as described in the main text. Minus signs indicate improvement. Only clear sky conditions were considered.

	dSCDs		AOT/VCD		Surface	
	ΔRMSD [%]	ΔBias [%]	ΔRMSD [%]	ΔBias [%]	ΔRMSD [%]	ΔBias [%]
HCHO	2.7	3.5	6.8	10.5	-17.4	-22.0
NO_2 UV	-0.7	-1.1	-2.7	-2.6	-3.5	8.7
NO_2 Vis	-0.7	-3.3	-0.8	-1.0	-2.8	-50.9
Aerosol UV	-0.7	0.7	12.5	20.2	-	-
Aerosol Vis	-0.2	2.1	-8.7	-40.1	-	-

S8 Further details on the comparison results

S8.1 AVKs of individual participants

Figures S35 to S39 show the averaging kernels (AVKs) and retrieved degrees of freedom of signal (DOFS) of each participant for aerosol UV. For explanation of colours and symbols please refer to main text Sect. 3.1. The DOFS values in brackets were calculated considering valid data only.

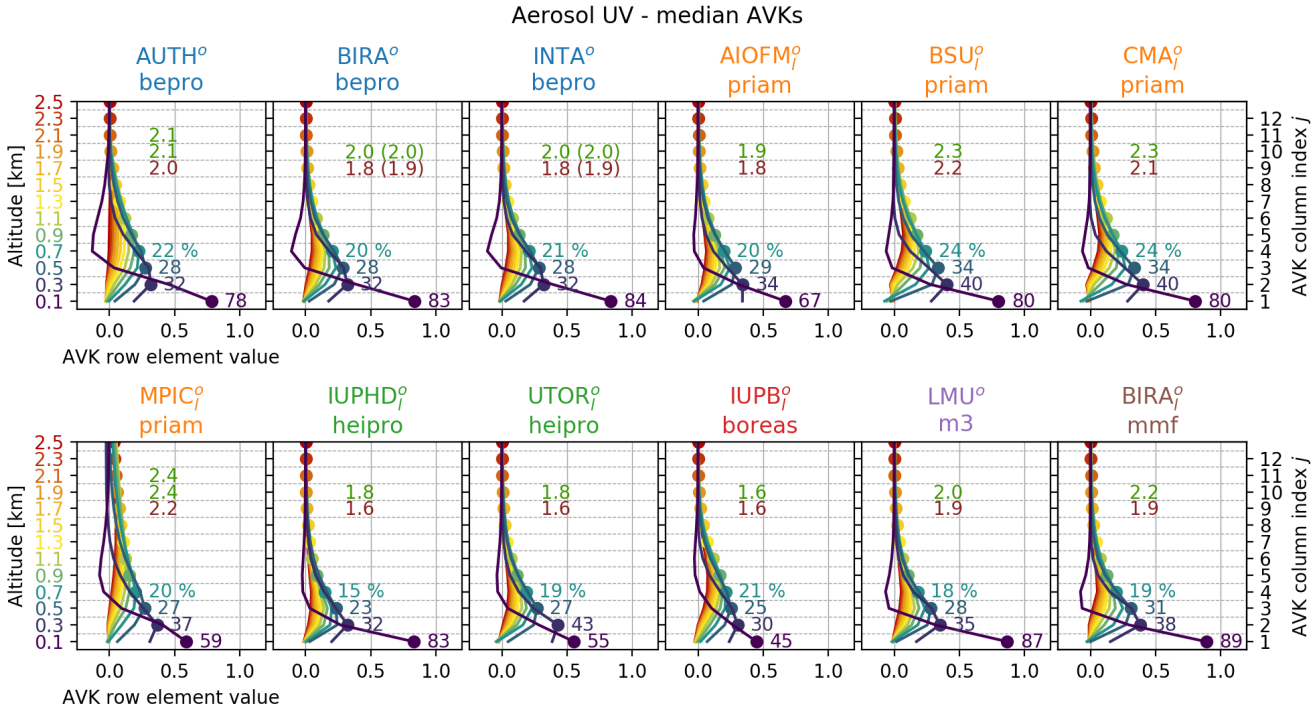

Figure S15. Mean averaging kernels for Aerosol UV for each participant. Coloured values at AVK peaks show the amount of retrieved information on the respective layer in percent. "DOFS" numbers are given for clear-sky (green) and cloudy (red) conditions. Values in brackets are DOFS including flagging.

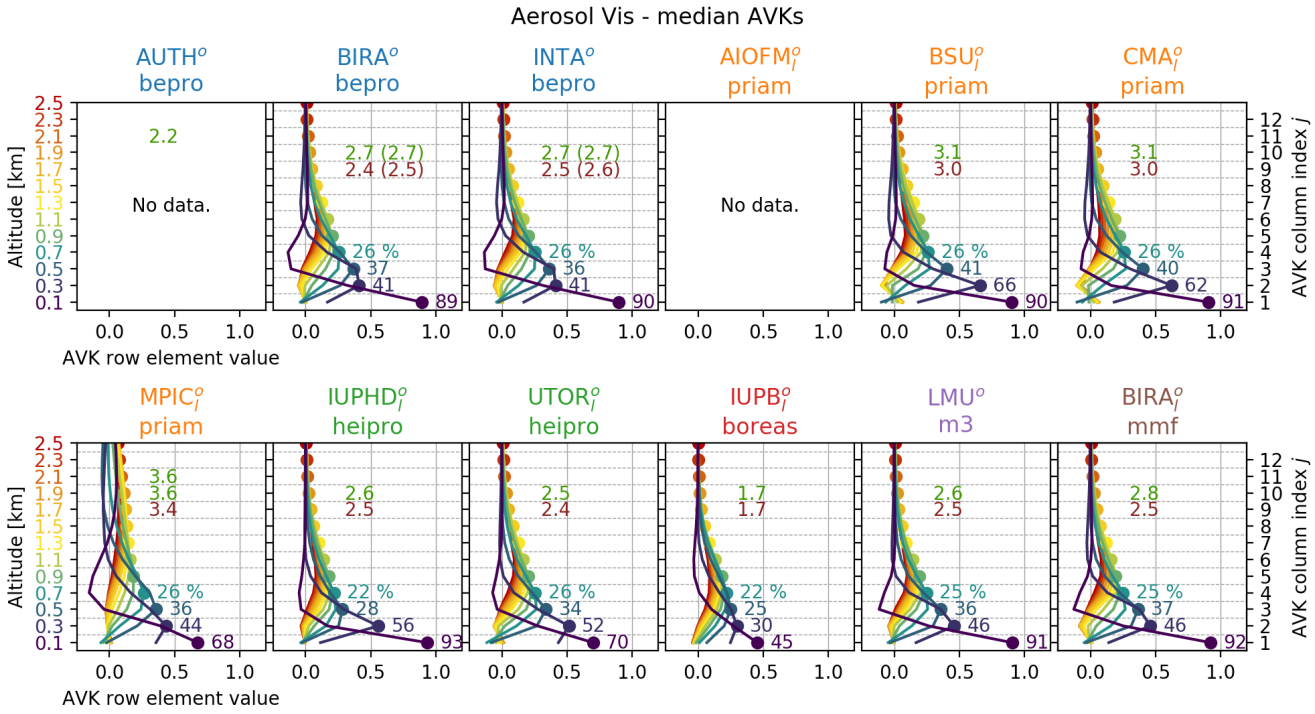

Figure S16. Mean averaging kernels for Aerosol Vis for each participant. Description of Fig. S35 applies.

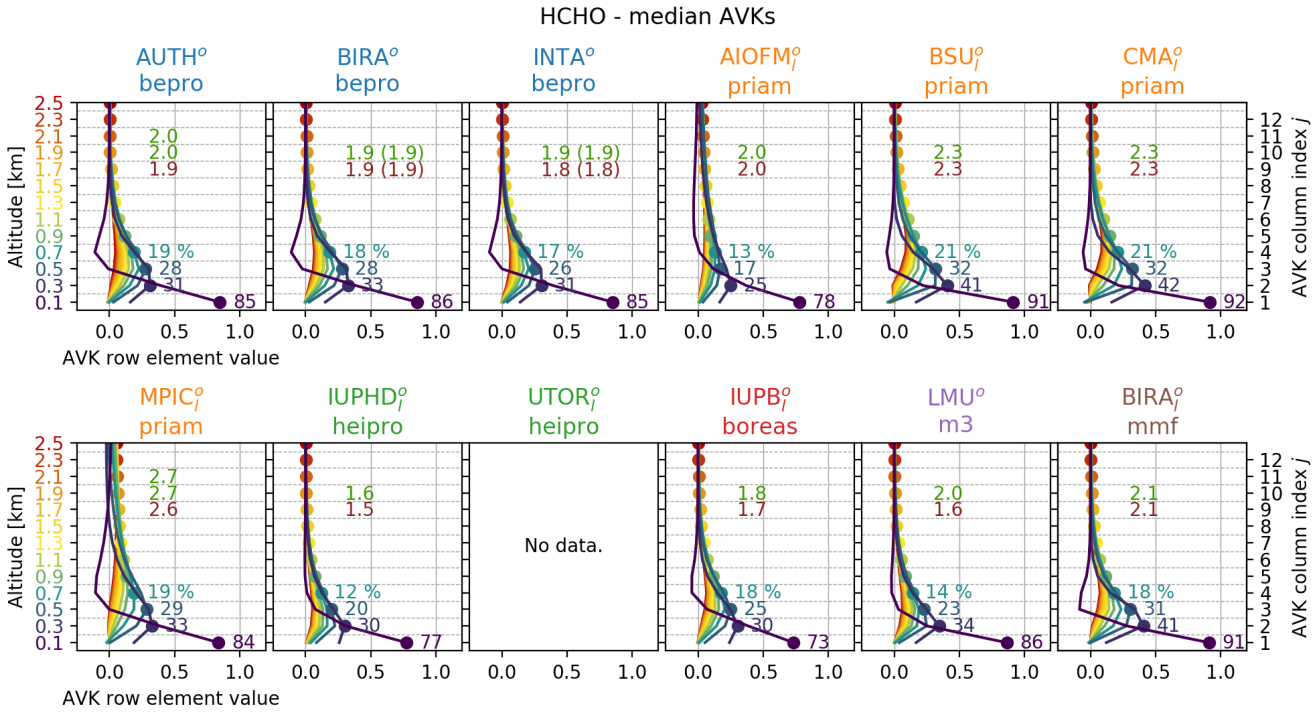

Figure S17. Mean averaging kernels for HCHO for each participant. Description of Fig. S35 applies.

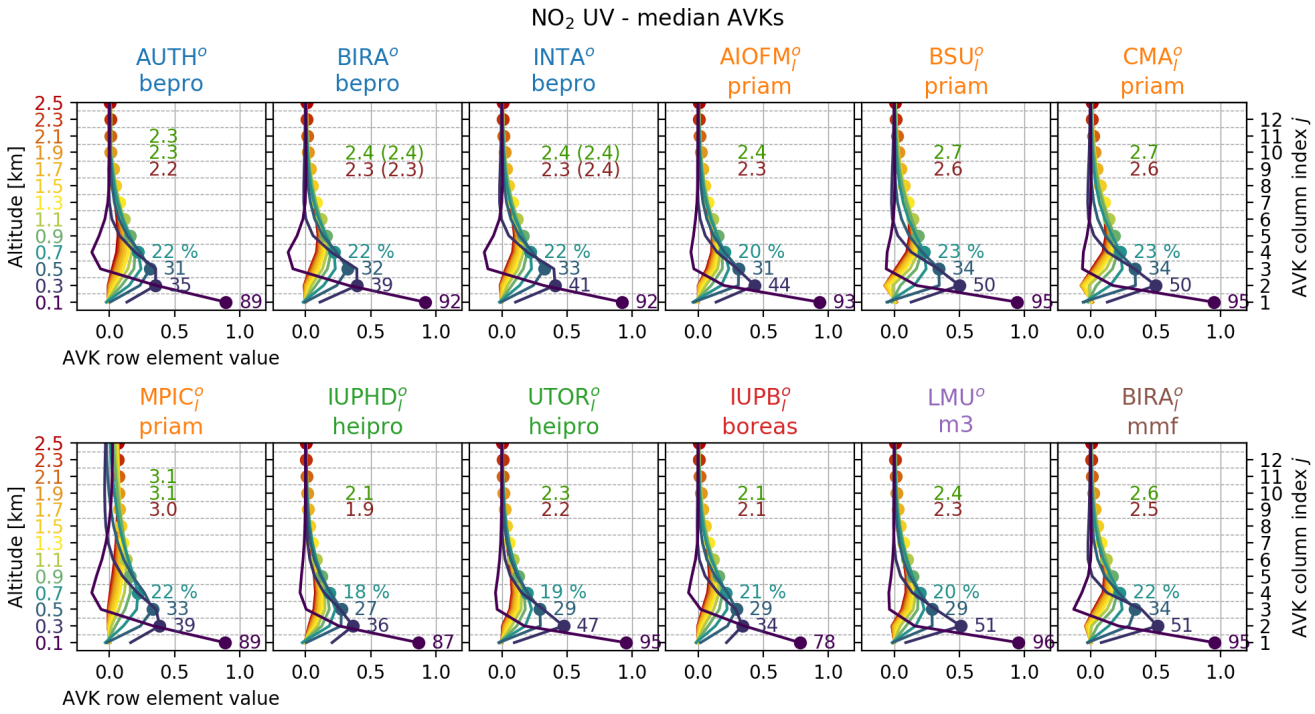

Figure S18. Mean averaging kernels for NO_2 UV for each participant. Description of Fig. S35 applies.

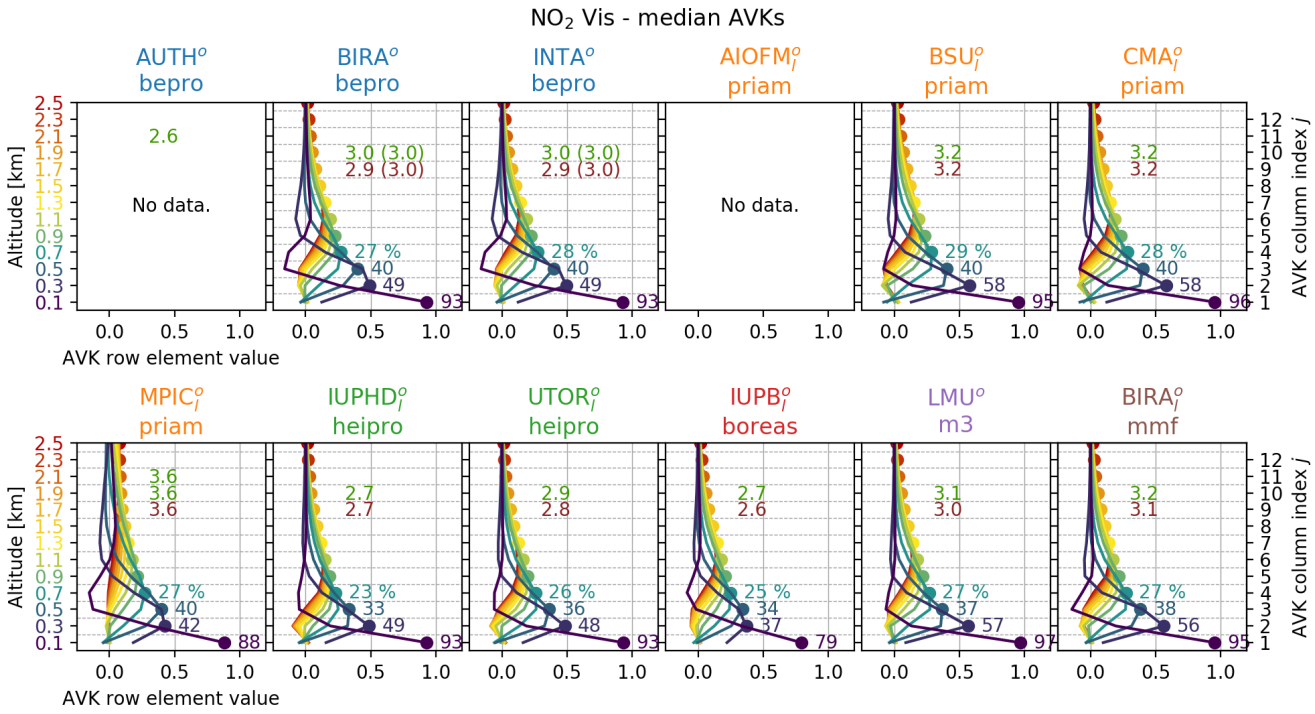

Figure S19. Mean averaging kernels for NO_2 Vis for each participant. Description of Fig. S35 applies.

S8.2 Profile deviation statistics

Figures S20 to S24 show statistics on the observed differences in the retrieved profiles for all five species. The plots on the left compare the retrieved profiles of individual participants $\boldsymbol{x}$ to the median MAX-DOAS profiles $\bar{\boldsymbol{x}}$. While the vertical axes represent altitude, the horizontal axes depicts the difference $\boldsymbol{x} - \bar{\boldsymbol{x}}$. The coloured boxes indicate the $25\,\% - 75\,\%$ percentile, whiskers are $5\,\% - 95\,\%$. Black dots indicate the mean value. For each layer there are boxplots for clear-sky (green) and cloudy conditions (red). Note that for aerosol there are two different horizontal axes defined for the two cloud conditions: the green scale at the bottom and the red scale on the top of each plot. Only valid data (flagged) was considered. For aerosol and NO_2 a plot on the very right shows statistics of the difference of supporting measurements $\boldsymbol{x_{anc}}$ (lidar/ radiosonde for NO_2, sun photometer scaled ceilometer for aerosol) to the median $\bar{\boldsymbol{x}}$, hence $\boldsymbol{x_{anc}} - \bar{\boldsymbol{x}}$. The numbers in all the plots show RMSD deviation of the three lowest (most sensitive) layers. Dashed lines indicate the median retrieval uncertainty as specified by the participants.

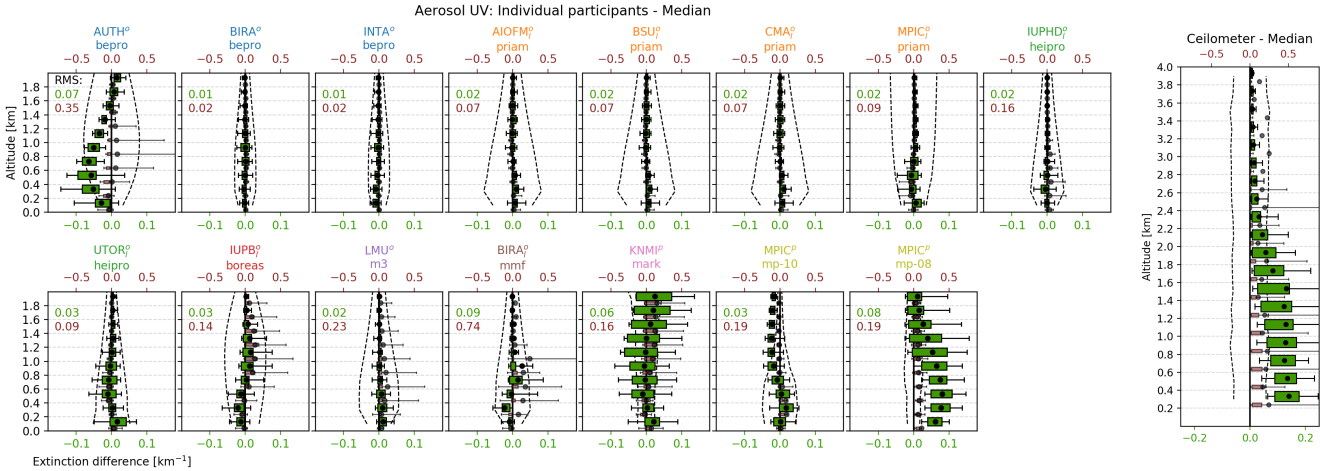

Figure S20. Left: ~~Flagged deviations for~~ Deviations of Aerosol UV profiles (valid only) of individual participants from the MAX-DOAS median profiles. Dots show the mean, boxes indicate the (25%-75%) percentile, and whiskers show (25%-75%) percentile. Green (red) box-whiskers represent clear (cloudy) conditions. Note, that there are different x-scales (on top and bottom of the plot) for different cloud conditions. The average standard deviations specified by the participants are indicated by the dashed lines. Right: Deviation of the AOT scaled ceilometer backscatter signal to the MAX-DOAS median profiles. The numbers in the plots indicate RMSD values for clear sky (green) and cloudy (red) conditions. Further details are given in the related text in Sect. S8.2

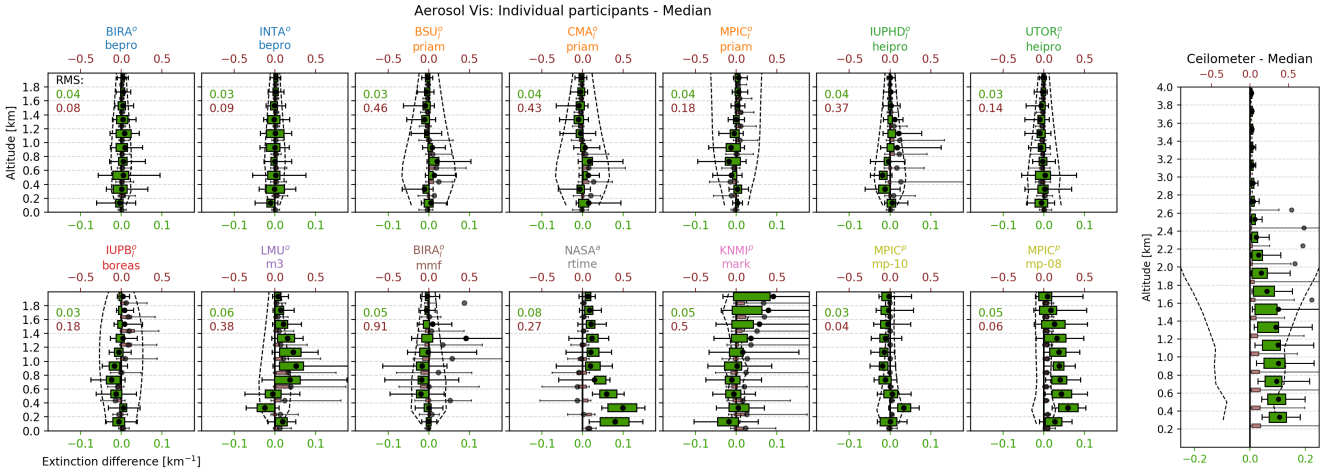

Figure S21. ~~Flagged deviations for~~ Deviations of Aerosol Vis profiles (valid only) of individual participants from the MAX-DOAS median profiles. Description of Fig. S20 applies.

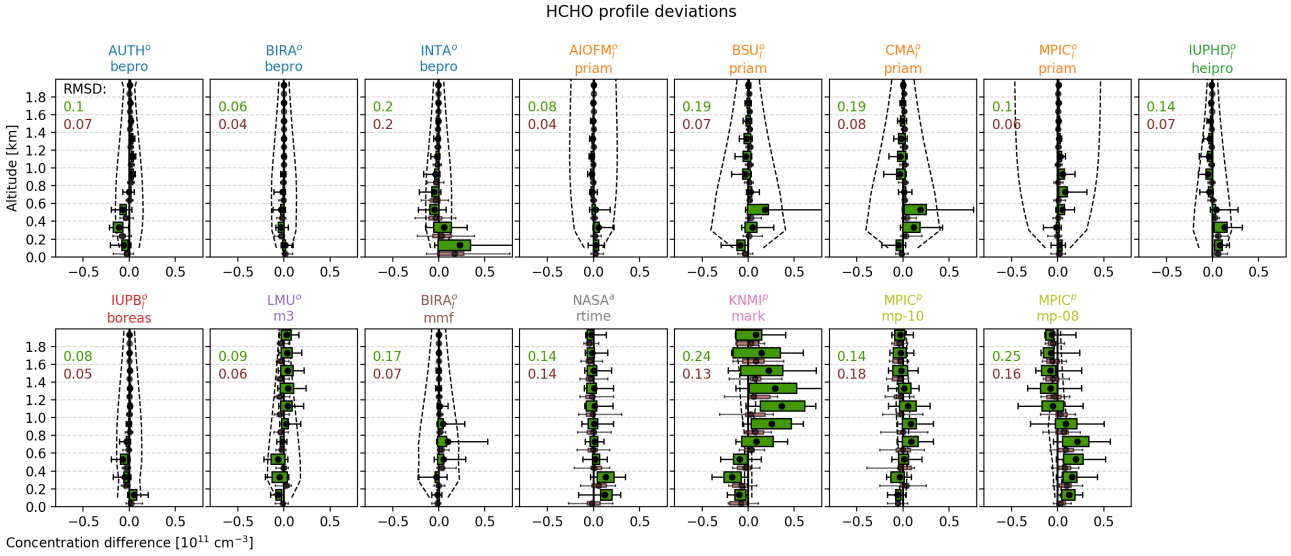

Figure S22. ~~Flagged deviations for~~ Deviations of HCHO profiles (valid only) of individual participants from the MAX-DOAS median profiles. The description of Fig. S20 applies but for HCHO, there is no independent reference profile available.

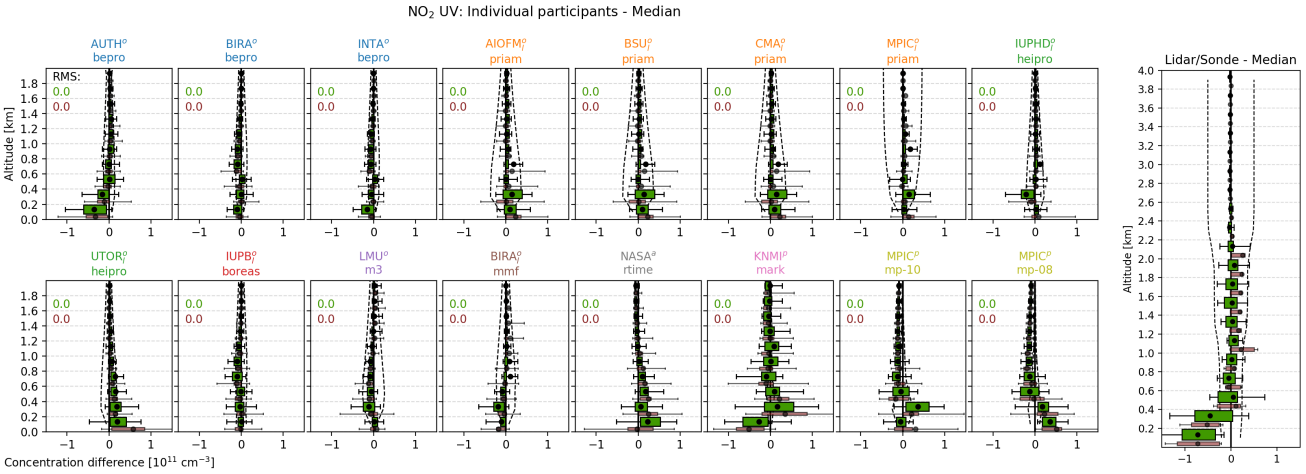

Figure S23. ~~Flagged deviations for~~ Deviations of NO_2 UV profiles (valid only) of individual participants from the MAX-DOAS median profiles. The description of Fig. S20 applies. On the right, deviation of the median retrieved profiles from the few available NO_2 lidar and sonde profiles are shown.

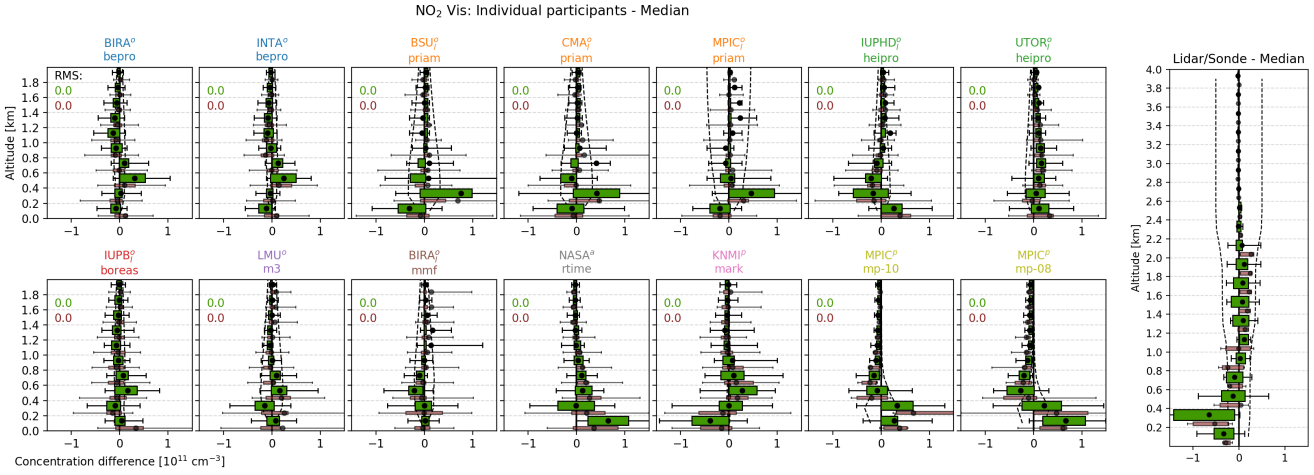

Figure S24. ~~Flagged deviations for~~ Deviations of NO_2 ~~UV~~ Vis profiles (valid only) of individual participants from the MAX-DOAS median profiles. The description of Fig. S23 applies.

S8.3 Correlation plots for AOTs, VCDs and surface concentrations

Figures S25 to S32 show the individual correlation plots for the comparisons of AOTs, VCDs and surface concentrations as summarized in Sect. 3.4, 3.5 and 3.6 in the main text, respectively. The colours indicate cloud conditions: clear-sky (green) and cloudy conditions (red). Transparent markers represent data points flagged as invalid. The small grey bars indicate uncertainties in the measurement. For the median values, the bars show the standard deviation among the participants (valid data only). Dashed lines in the correlation plots represent the ideal 1:1 line. Correlations were performed separately against MAX-DOAS median values and supporting observations. For AOTs a third correlation is shown for the partial AOT (with PAC applied).

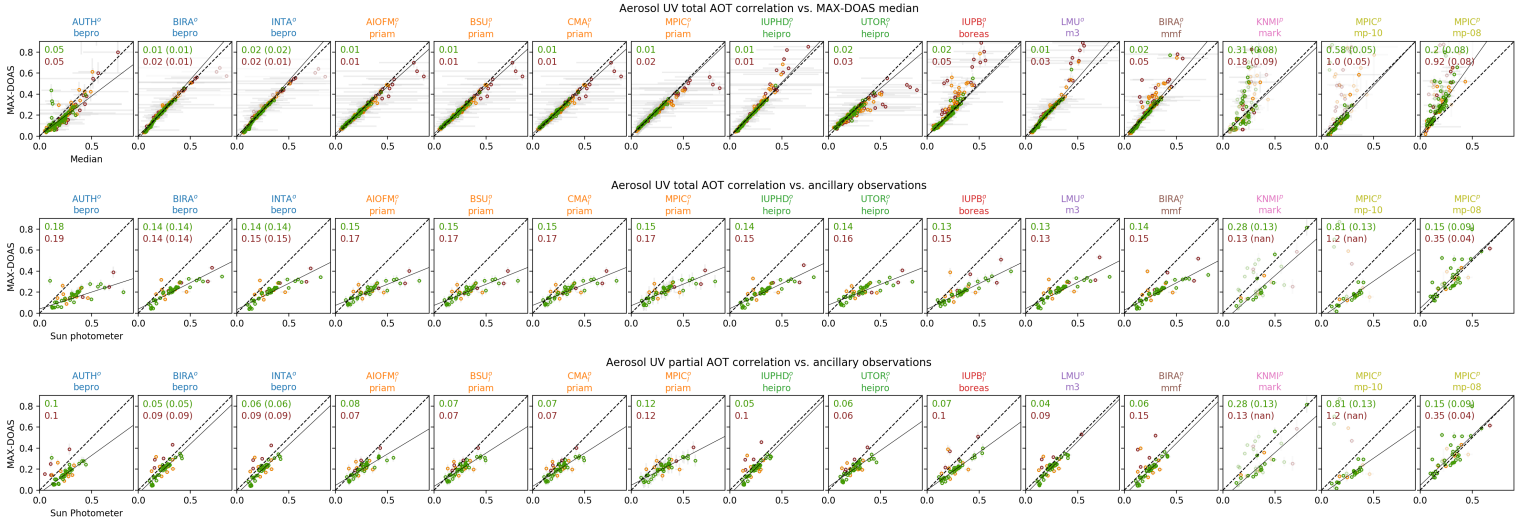

Figure S25. Correlation of Aerosol UV AOTs. Complementary to main text Sect. 3.4

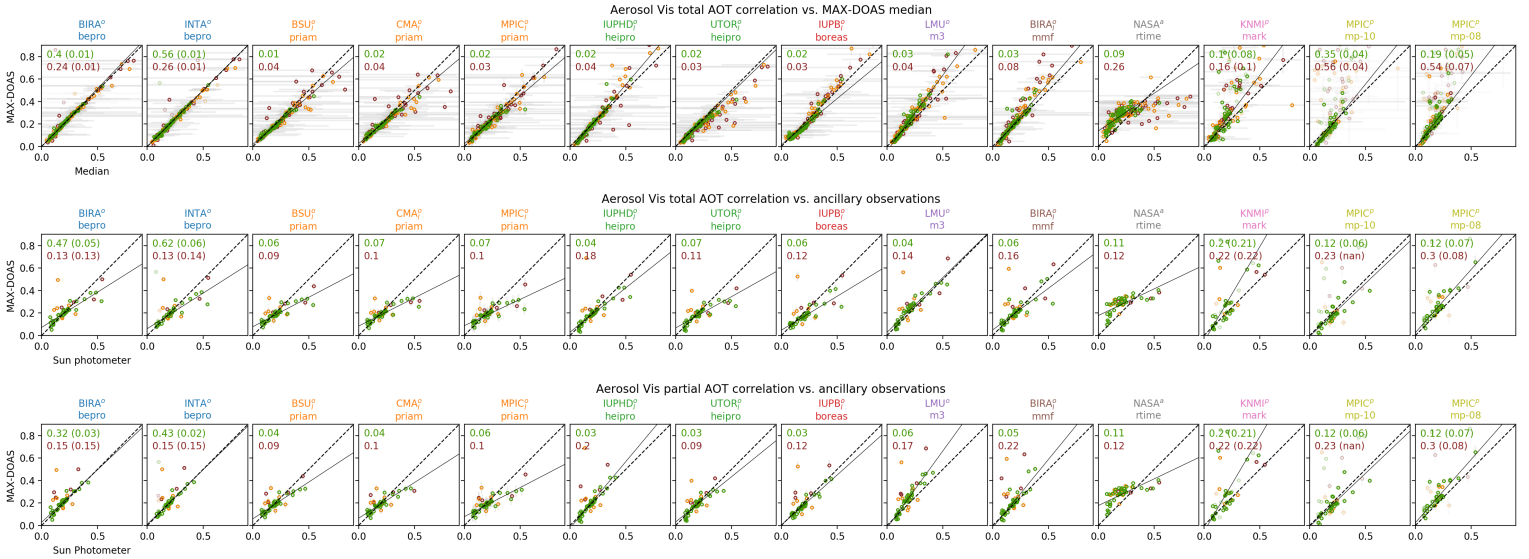

Figure S26. Correlation of Aerosol Vis AOTs. Complementary to main text Sect. 3.4

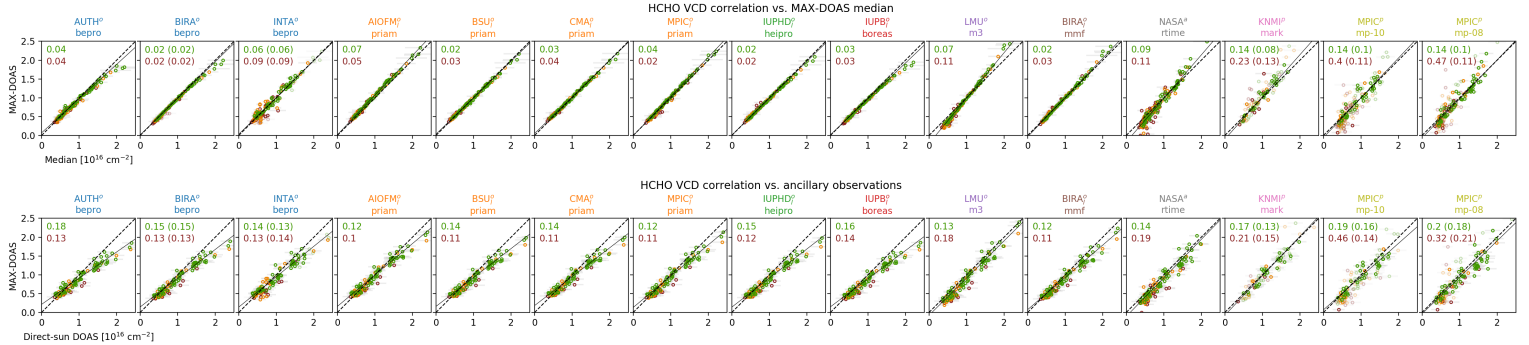

Figure S27. HCHO VCD correlation plots. Complementary to main text Sect. 3.5

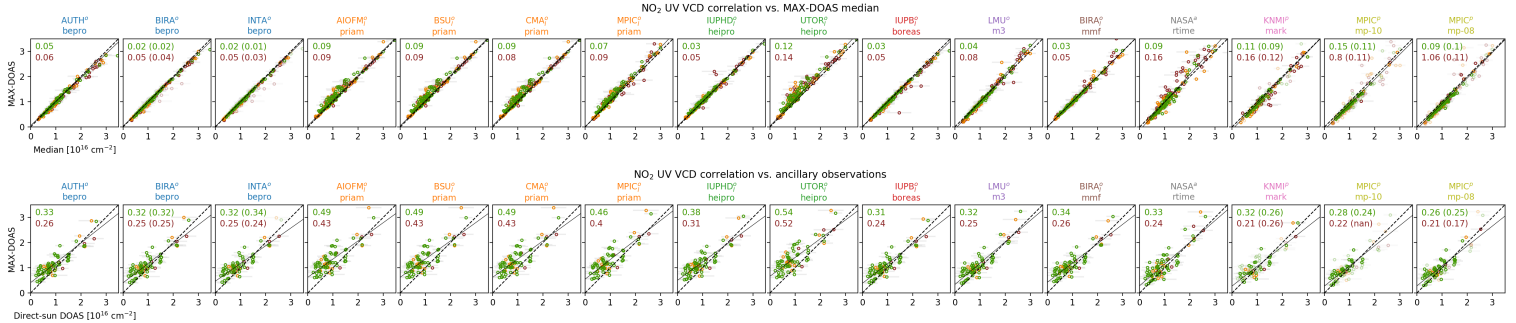

Figure S28. NO_2 UV VCD correlation plots. Complementary to main text Sect. 3.5

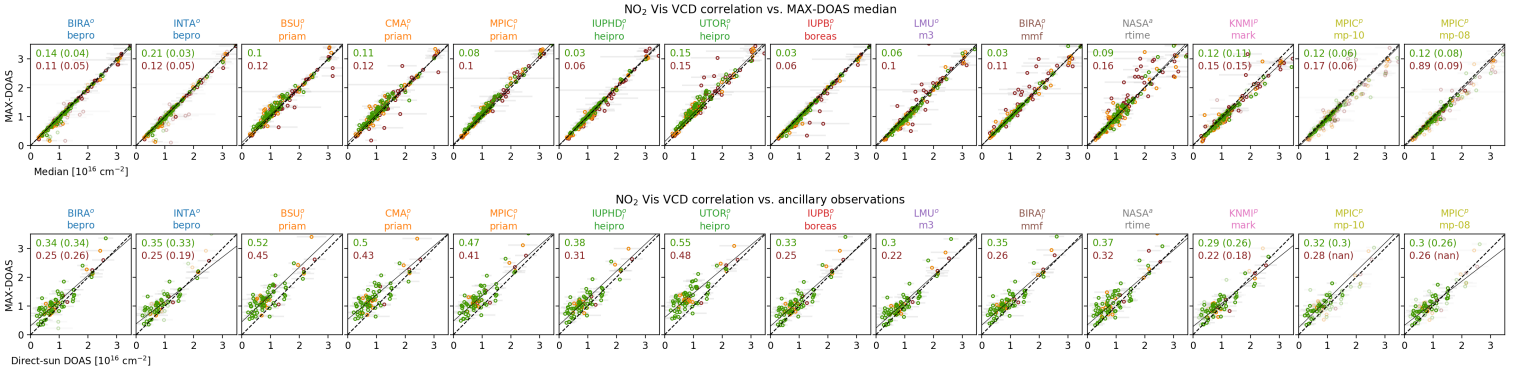

Figure S29. NO_2 Vis VCD correlation plots. Complementary to main text Sect. 3.5

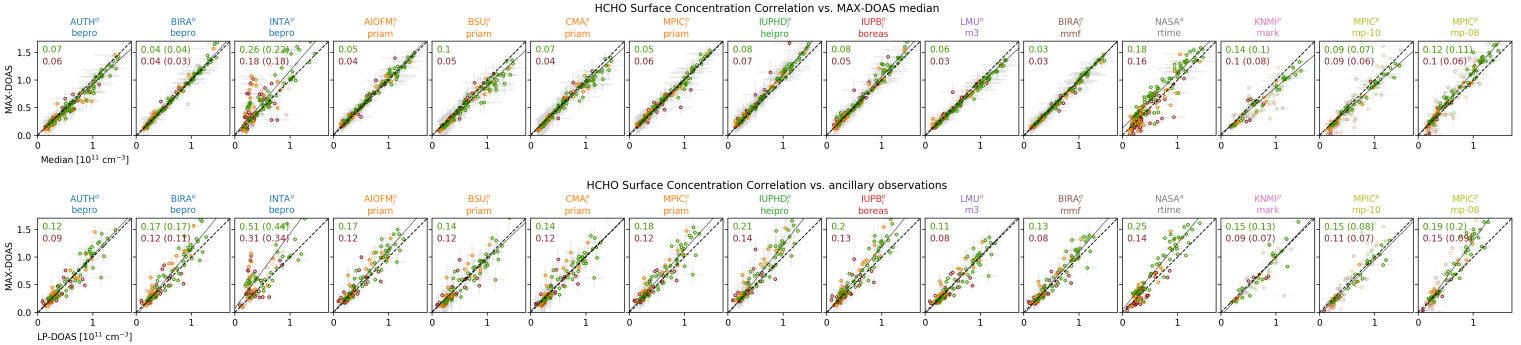

Figure S30. HCHO surface concentration correlation plots. Complementary to main text Sect. 3.6

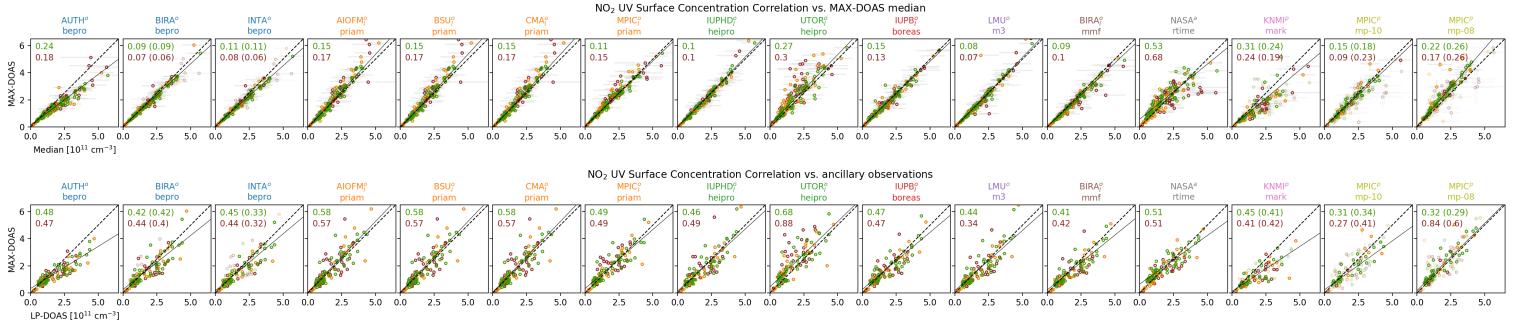

Figure S31. NO_2 UV surface concentration correlation plots. Complementary to main text Sect. 3.6

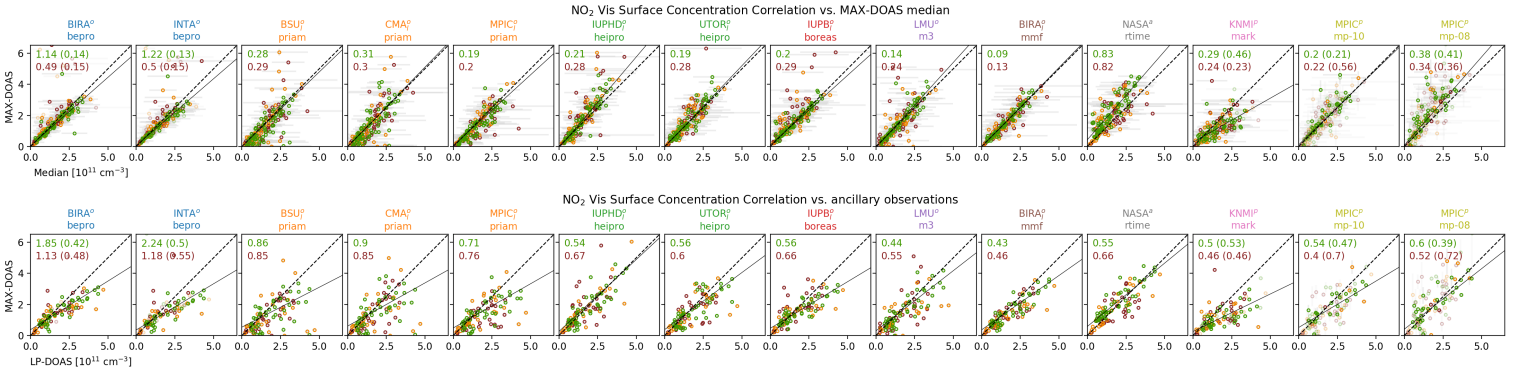

Figure S32. NO_2 Vis surface concentration correlation plots. Complementary to main text Sect. 3.6

S9 Impact of smoothing on surface concentration

For NO_2 the impact of smoothing effects on the surface concentration retrieved by OEM algorithms can be estimated from profiles of the NO_2 lidar and radiosondes. Each profile is smoothed according to Eq. (9) in the main text and the difference in surface concentration between the smoothed and the unsmoothed profile is calculated. ~~Figure~~ Fig. S33 shows histograms of the calculated differences. The standard deviation is about 5×10^9 molec cm^{-3} which is only about 10 % of the total average RMSD between MAX-DOAS and LP-DOAS observations. An estimate of the impact of smoothing on the retrieval results is actually provided by the OEM retrievals themselves as the "smoothing error". The specified smoothing errors are also indicated in Fig. S33 and ~~indeed slightly larger than~~ are similar to the standard deviation observed in in this test, meaning that for the surface layer they are well representative for the real impact of smoothing.

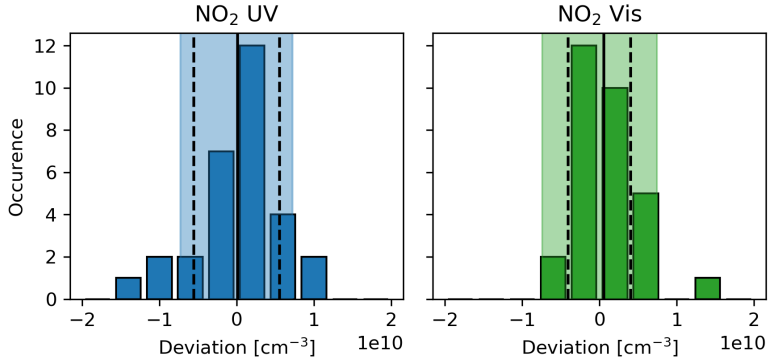

Figure S33. Histograms of the observed deviations in surface concentration between raw and smoothed lidar/ radiosonde NO_2 profiles. Solid and dashed lines indicate mean value and standard deviation, respectively. Coloured areas represent the median smoothing errors as specified by the OEM retrievals, which is in good agreement with the deviations obtained from the supporting NO_2 profiles.

S10 Participant's own dSCD comparison results

This section shows the comparison results for the case where each participant uses dSCDs measured with his own instrument. Evaluation and plots are fully equivalent to Sect. 3 in the main text. With DLR (German Aerospace Center, Oberpfaffenhofen, Germany, marked by blue squares) and USTC (University of Science and Technology of China, Hefei, China, marked by green squares), two other participants were included here, retrieving profiles with bePRO and HEIPRO, respectively. Gaps in the data are mostly related to instrument malfunction during the campaign. Further, not all instruments covered the spectral range to detect all desired species and the corresponding participants therefore do not appear in the respective plots.

S10.1 Information content

Information on the averaging kernels and DOFS. This section is equivalent to Sect. S10.1 in the main text and Sect. S8.1 in this supplement, respectively.

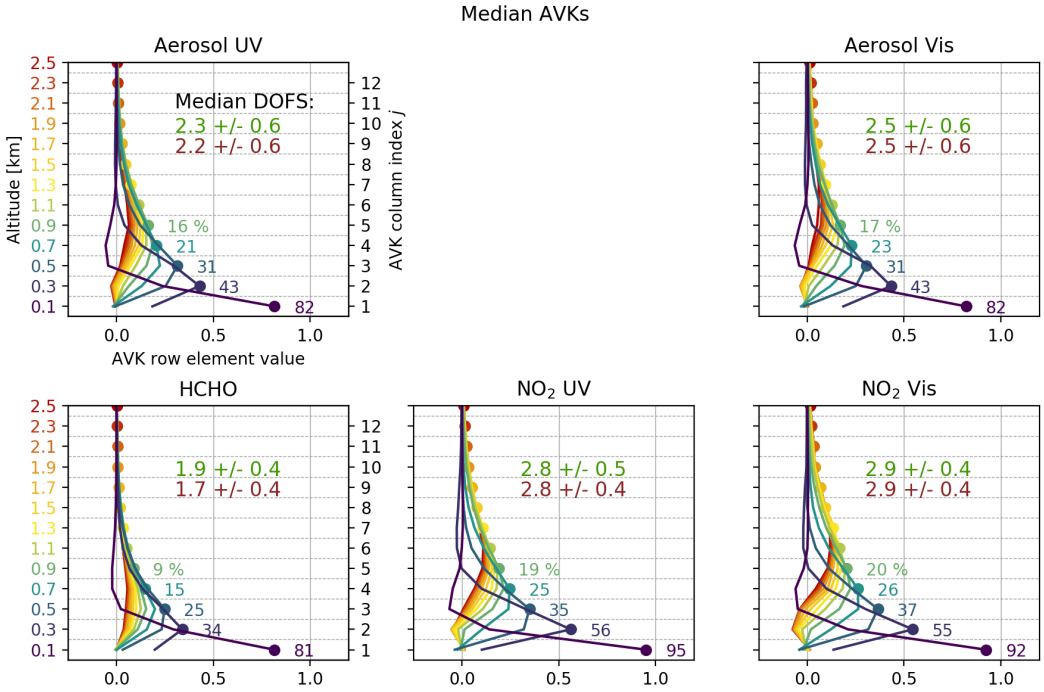

Figure S34. Average AVKs for the retrieved species (median over participants, mean over time). Each altitude and corresponding AVK line are associated with a colour, which is defined by the colour of the corresponding altitude-axis label. The dots mark the AVK diagonal elements. The number next to the dots show the exact value in percent, which corresponds to the amount of retrieved information on the respective layer. In the upper right of each panel, the DOFS (median among institutes, average over time) are given for clear-sky (green) and cloudy conditions (red).~~The vertical bars indicate the vertical resolution (the "spread" as defined by equation 3.23 in **?**) for the five lowest layers.~~

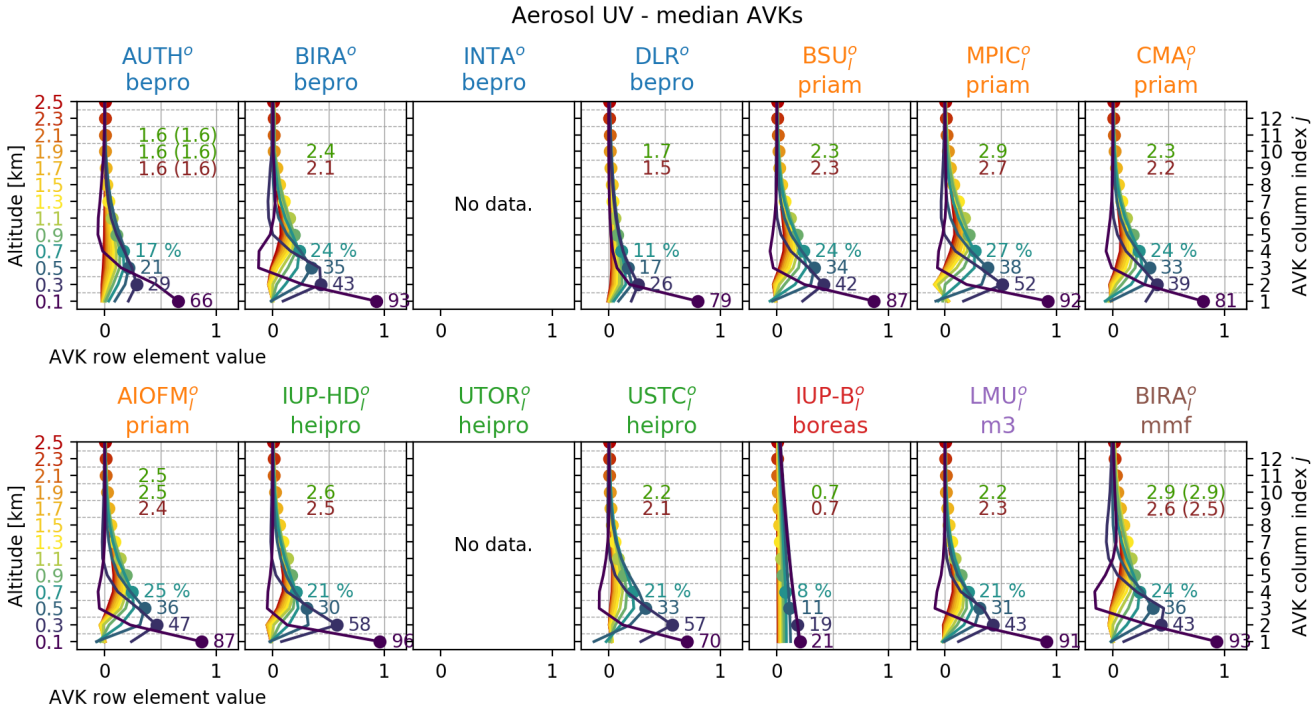

Figure S35. Mean averaging kernels for Aerosol UV for each participant retrieving from their own dSCDs. Coloured values at AVK peaks show the amount of retrieved information on the respective layer in percent. "DOFS" numbers are given for clear-sky (green) and cloudy (red) conditions. Values in brackets are DOFS including flagging.

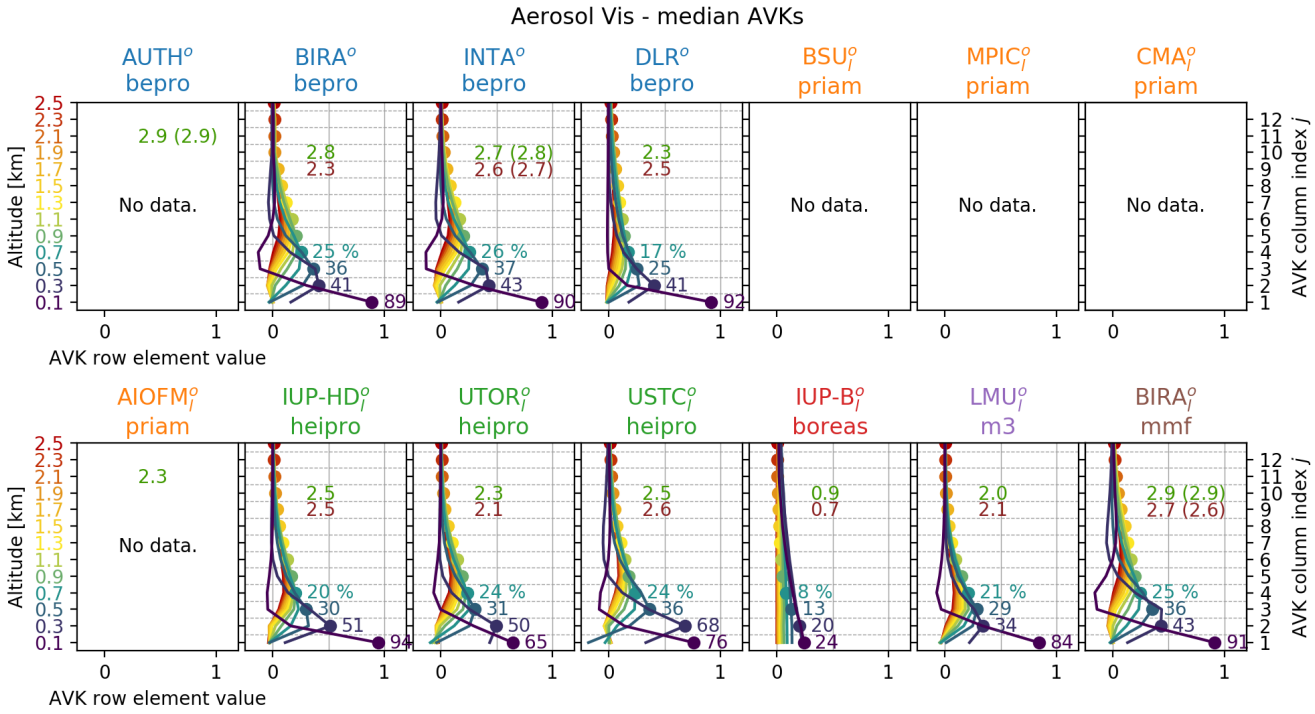

Figure S36. Mean averaging kernels for Aerosol Vis for each participant. Description of Fig. S35 applies.

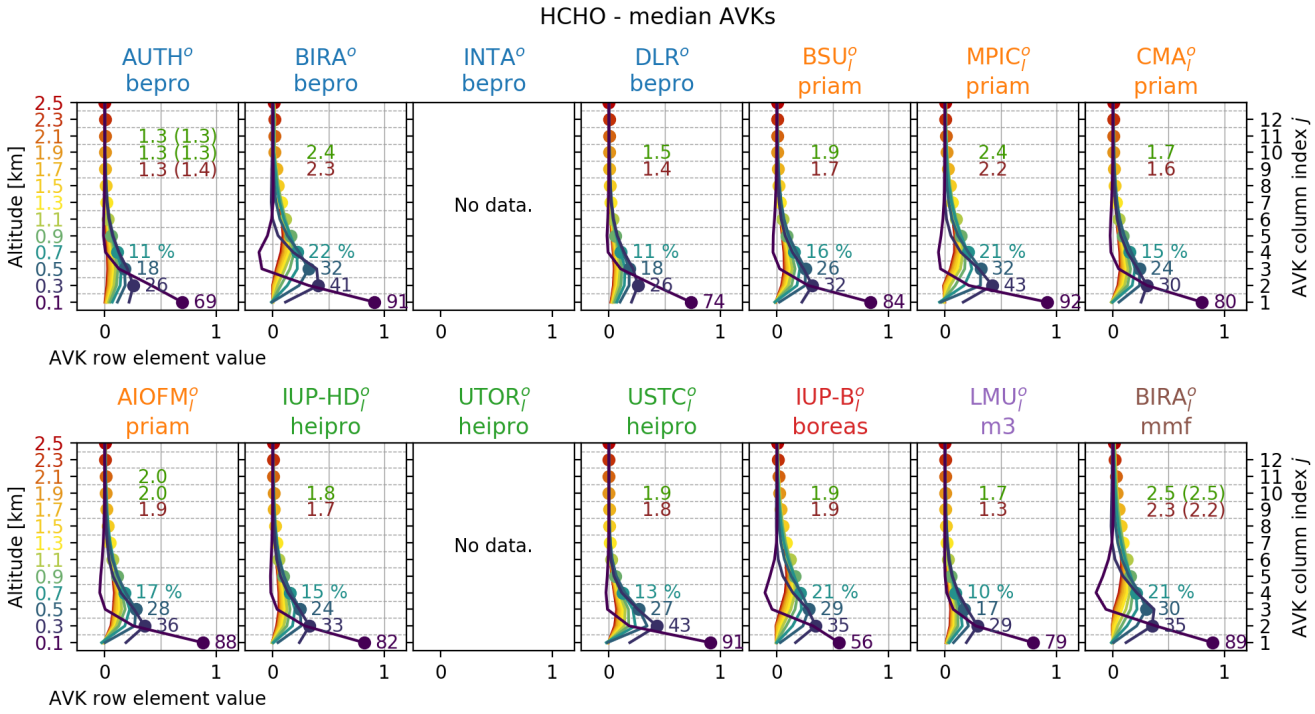

Figure S37. Mean averaging kernels for HCHO for each participant. Description of Fig. S35 applies.

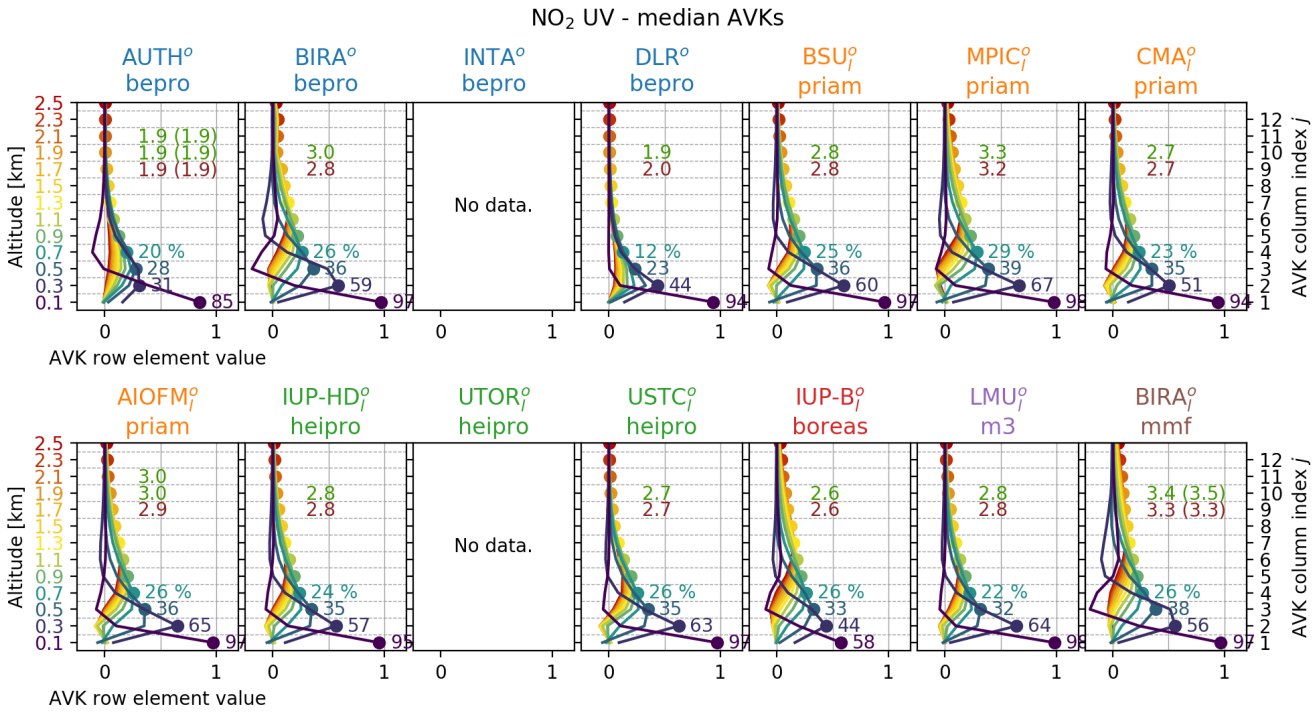

Figure S38. Mean averaging kernels for NO_2 UV for each participant. Description of Fig. S35 applies.

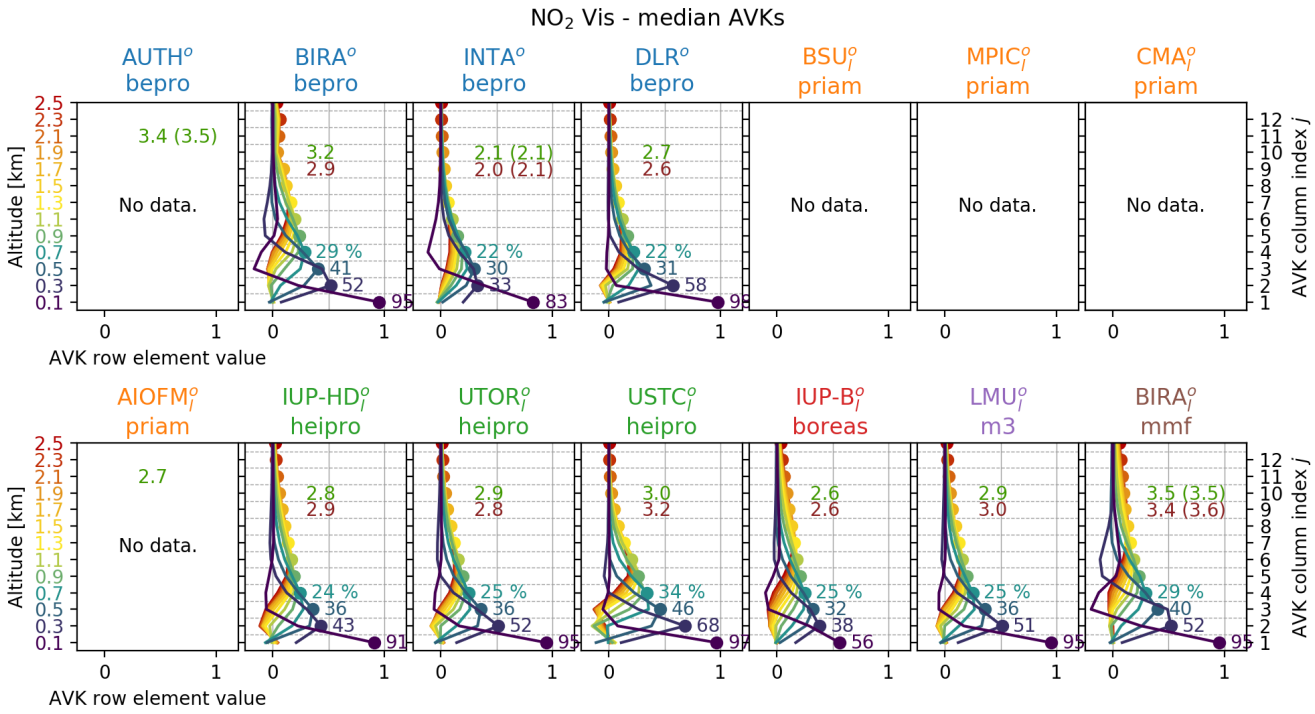

Figure S39. Mean averaging kernels for NO_2 Vis for each participant. Description of Fig. S35 applies.

S10.2 Overview plots

This section is equivalent to Sect. 3.2 in the main text.

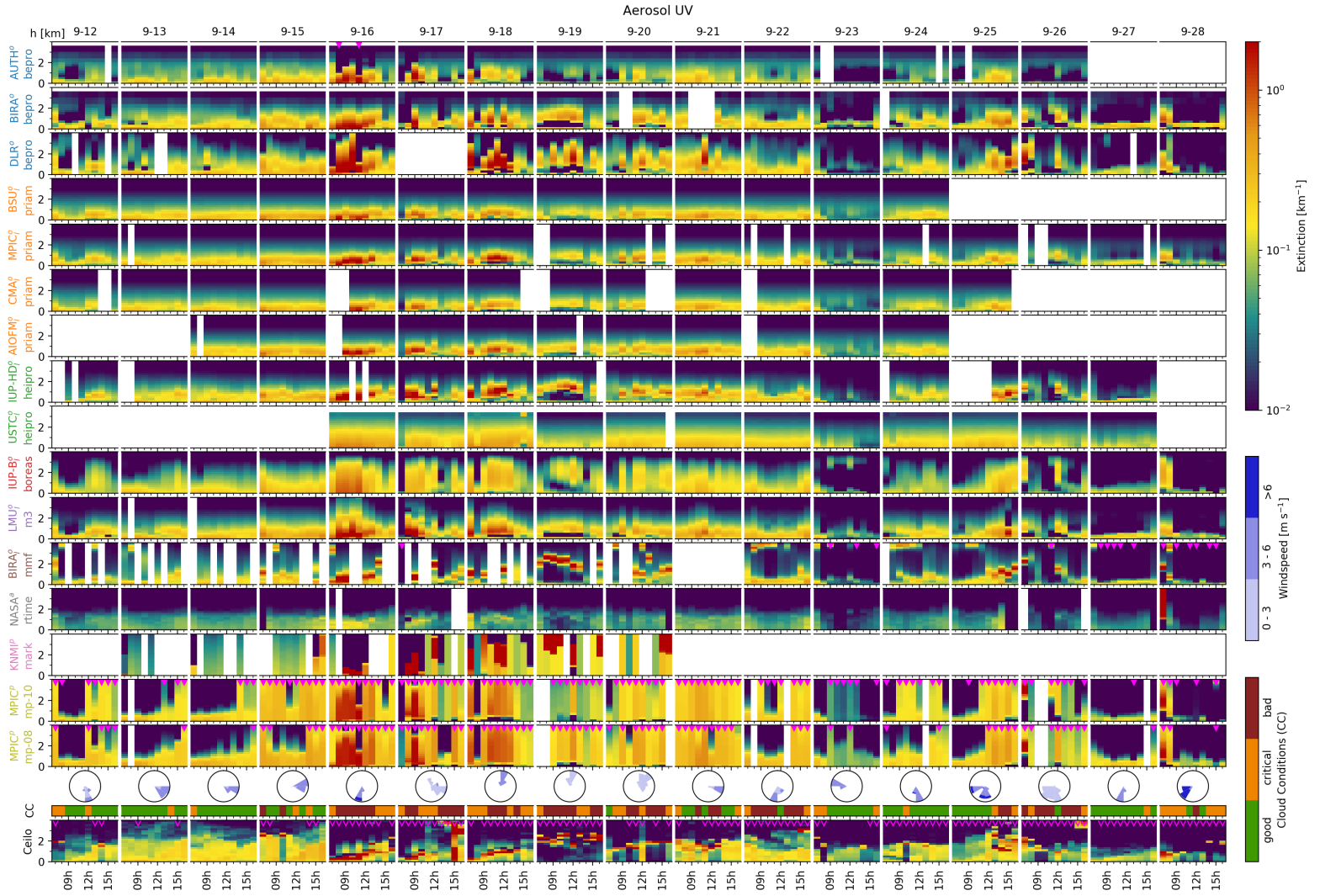

Figure S40. Aerosol UV extinction profiles retrieved from the participant's own dSCDs. The lowest row shows AOT scaled ceilometer backscatter profiles, calculated as described in Sect. S4.1. Backscatter profiles, which were scaled from MAX-DOAS AOTs (and which are therefore not fully independent) are marked by ~~red~~ pink triangles. Maximum extinction values reach $20\,\mathrm{km}^{-1}$, exceeding the colour scale.

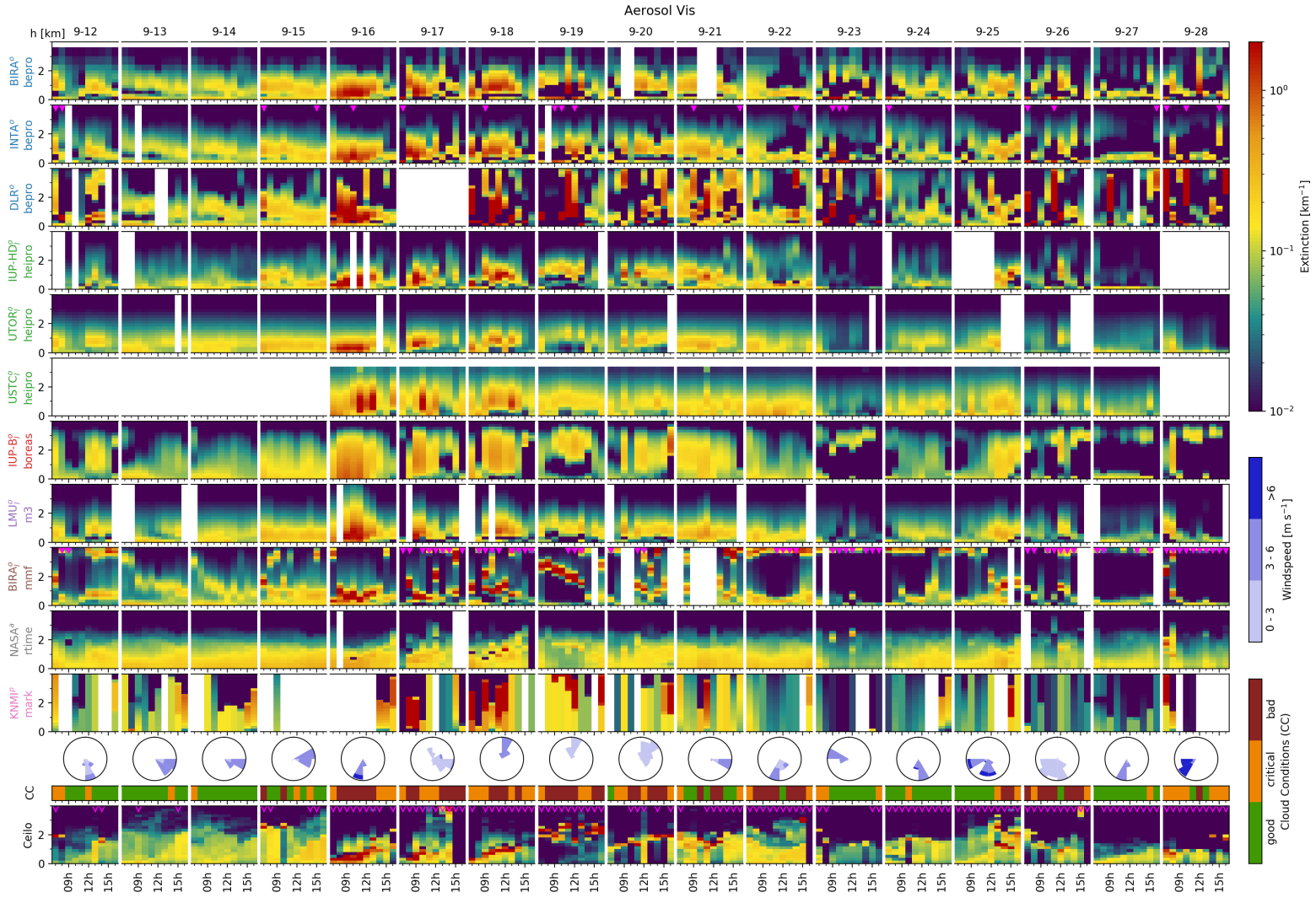

Figure S41. Aerosol Vis extinction profiles retrieved from the participant's own dSCDs. The lowest row shows AOT scaled ceilometer backscatter profiles, calculated as described in Sect. S4.1. Backscatter profiles, which were scaled from MAX-DOAS AOTs (and which are therefore not fully independent) are marked by ~~red~~ pink triangles. Maximum extinction values reach $20\,\mathrm{km}^{-1}$, exceeding the colour scale.

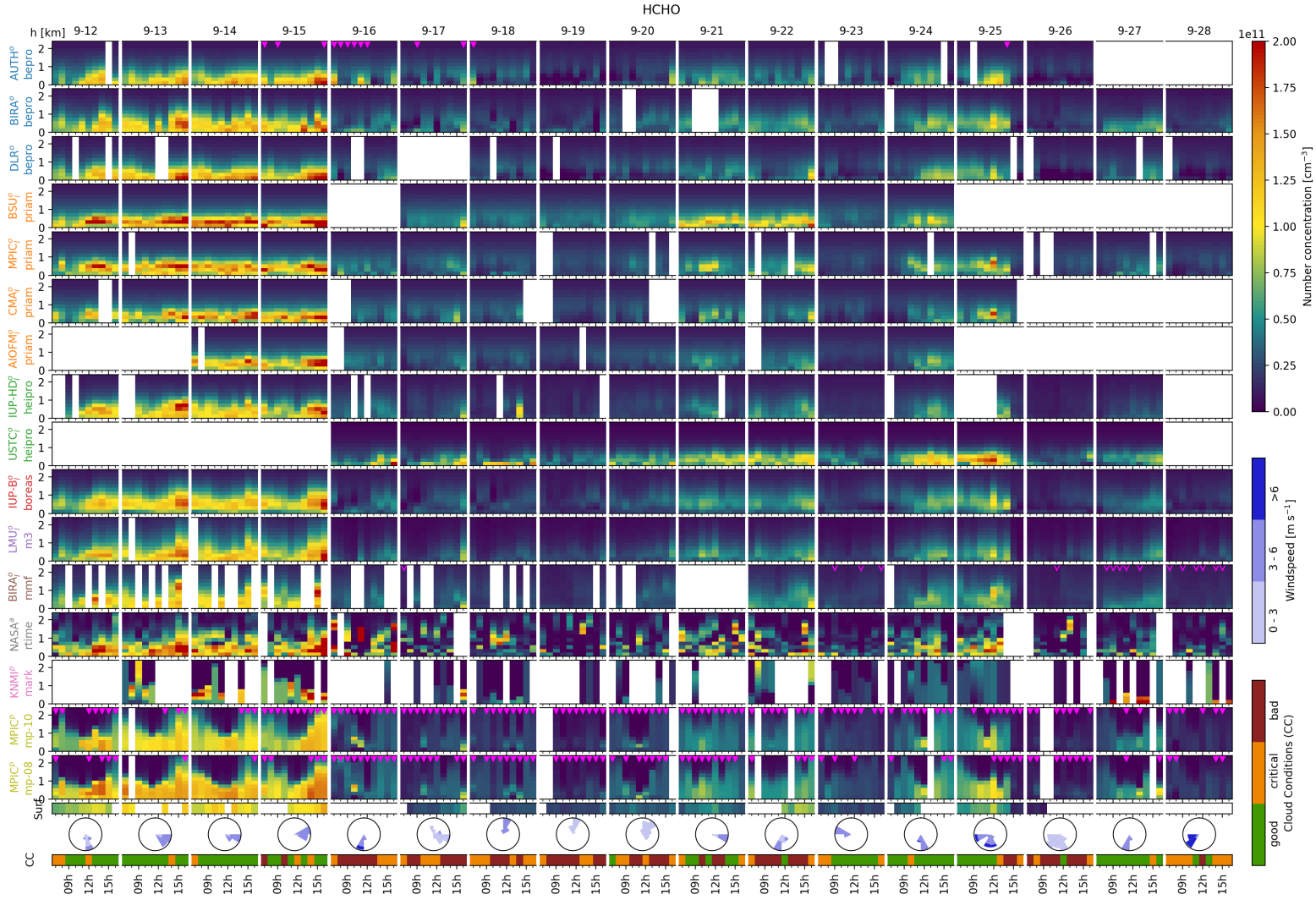

Figure S42. HCHO concentration profiles retrieved from the participant's own dSCDs. The "Surf"-row shows LP-DOAS surface concentrations.

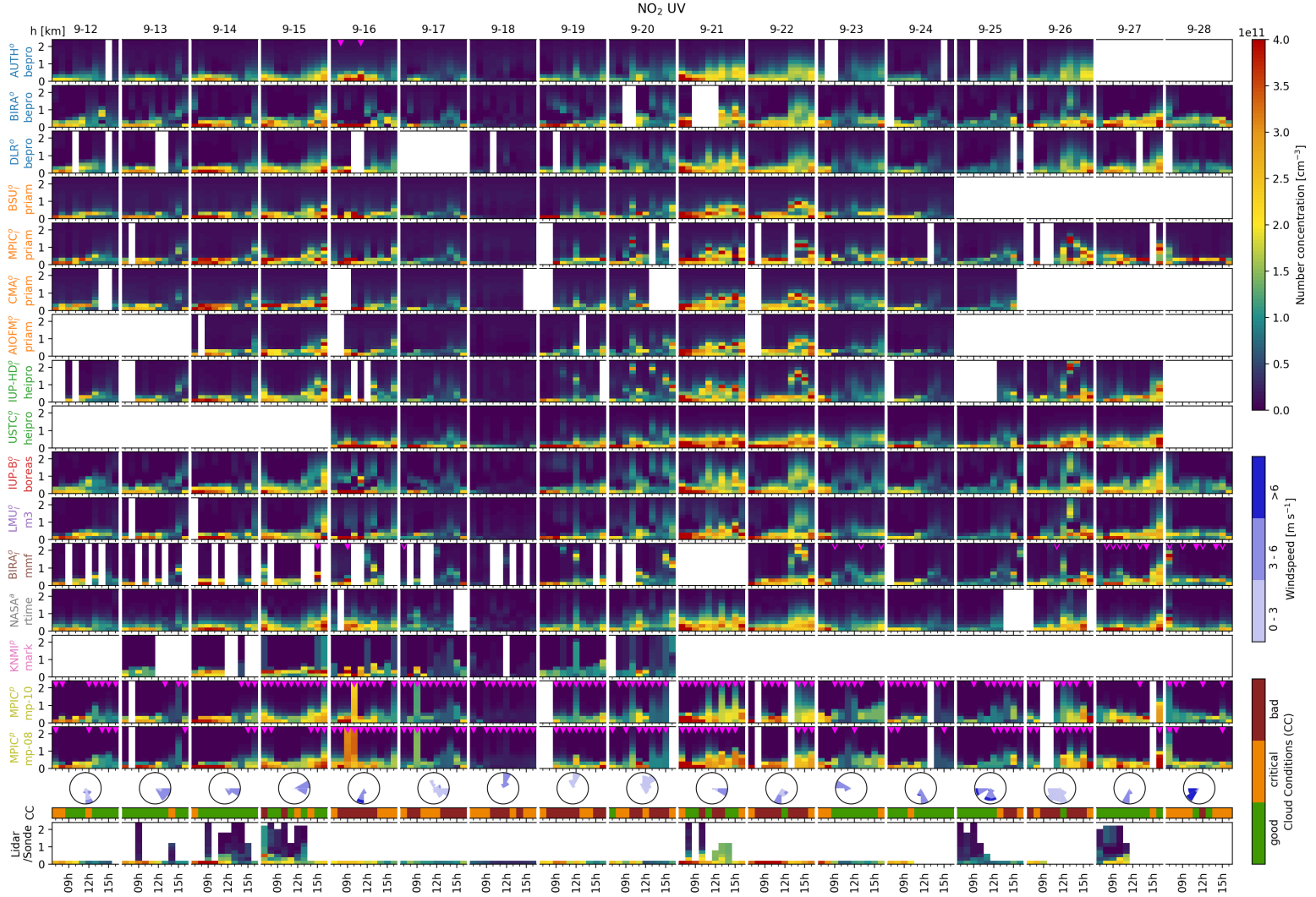

Figure S43. NO_2 UV concentration profiles retrieved from the participant's own dSCDs. The lowest row shows a combined dataset of NO_2 lidar, radiosonde, LP-DOAS and tower in-situ data. Redundant surface concentration measurements were averaged.

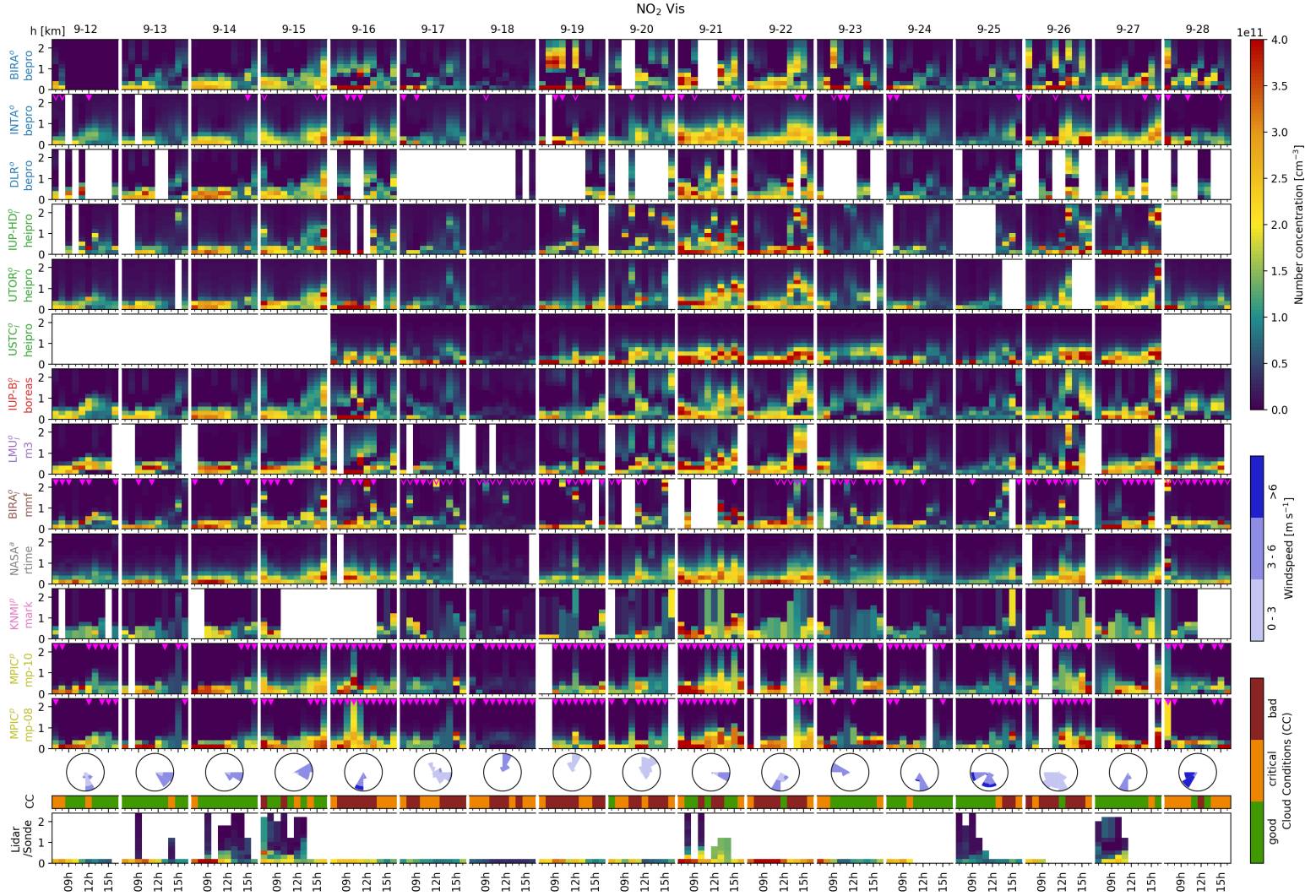

Figure S44. NO_2 Vis concentration profiles retrieved from the participant's own dSCDs. The lowest row shows a combined dataset of NO_2 lidar, radiosonde, LP-DOAS and tower in-situ data. Redundant surface concentration measurements were averaged.

S10.3 Modelled and measured dSCDs

This section is equivalent to Sect. S10.3 in the main text.

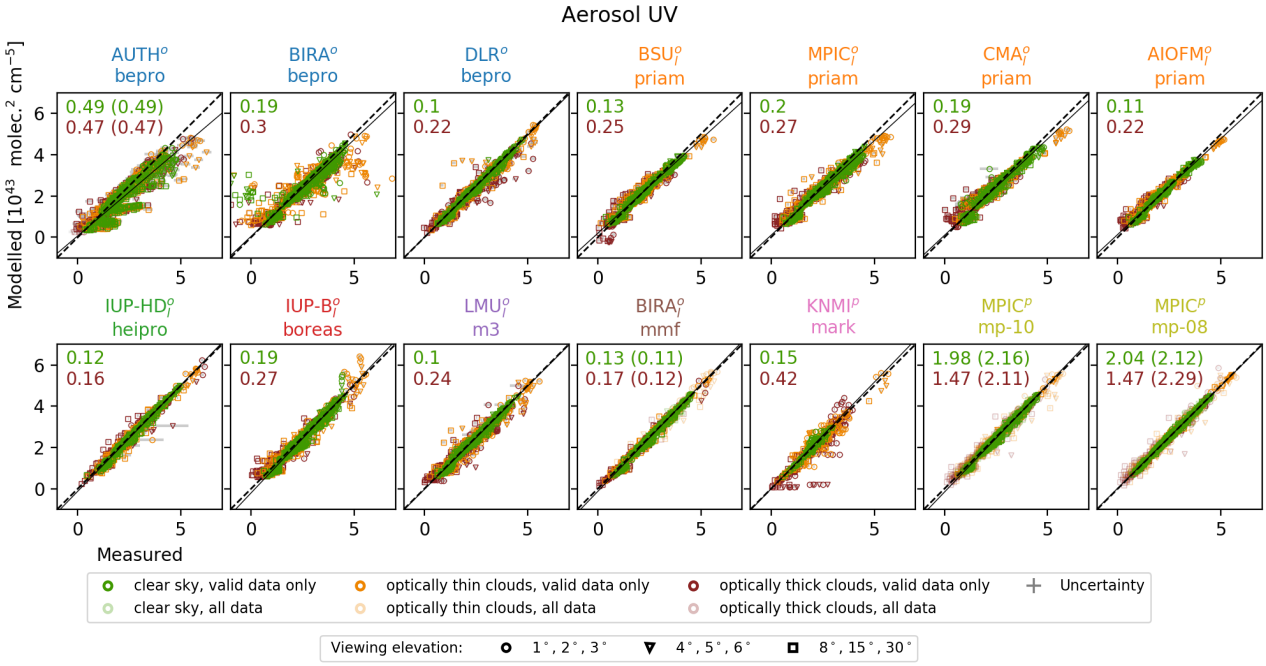

Figure S45. O_4 UV dSCD correlation when profiles are retrieved from the participant's own dSCDs. Marker colours and marker shapes indicate the cloud conditions and viewing elevation angles, respectively. Numbers represent the measurement error weighted RMSD between measured and modelled dSCDs for clear sky (green) and cloudy (red) conditions. Values in brackets were calculated only considering valid data.

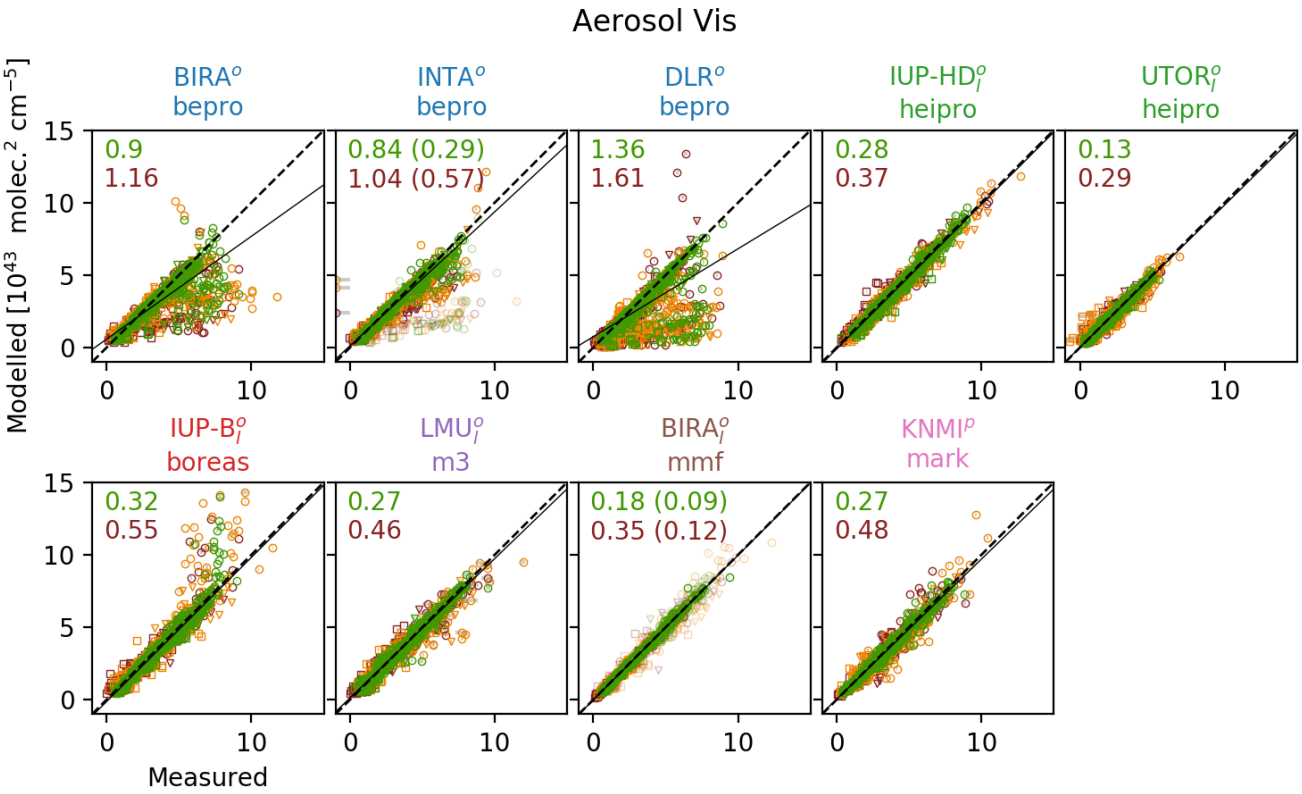

Figure S46. O_4 Vis dSCD correlation. Legends of Fig. S45 apply.

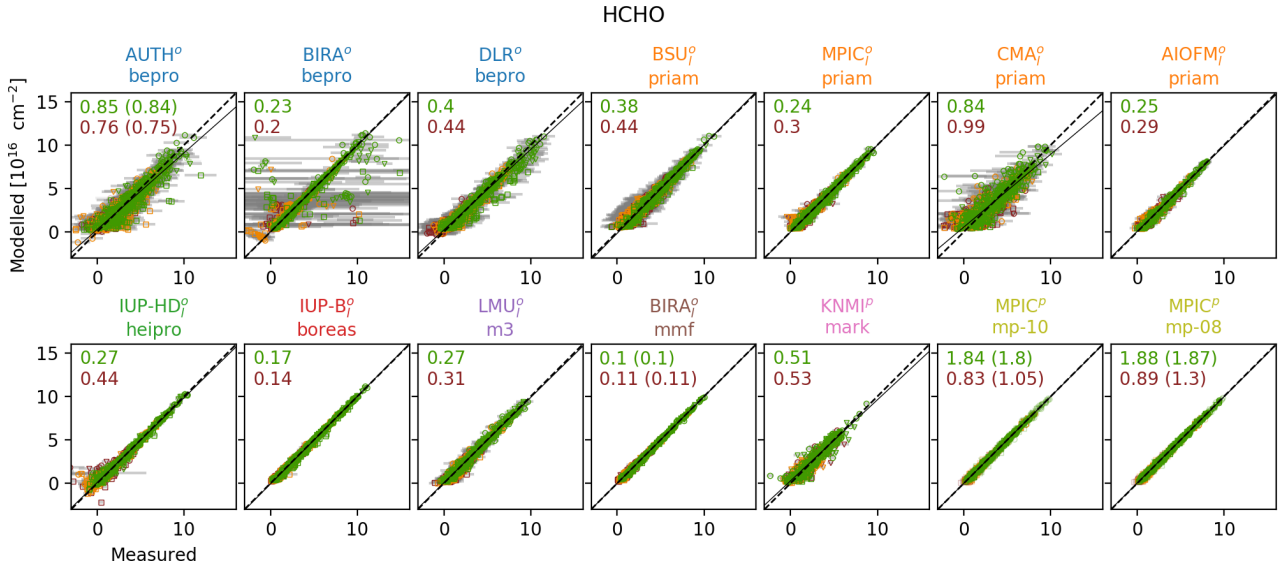

Figure S47. HCHO dSCD correlation. Legends of Fig. S45 apply.

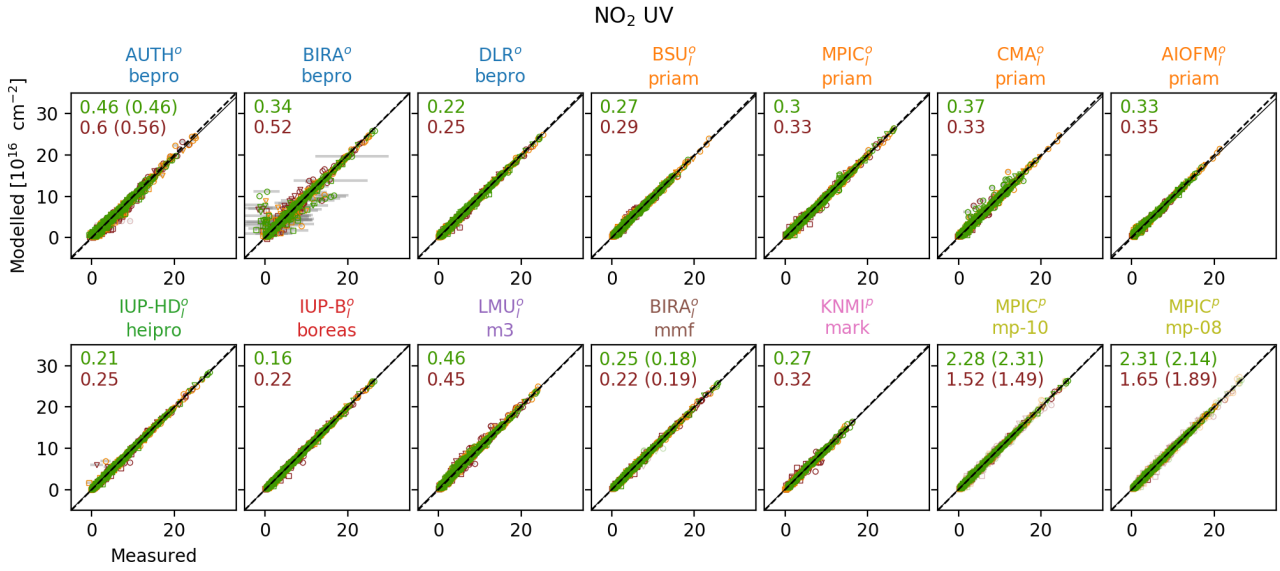

Figure S48. NO_2 UV dSCD correlation. Legends of Fig. S45 apply.

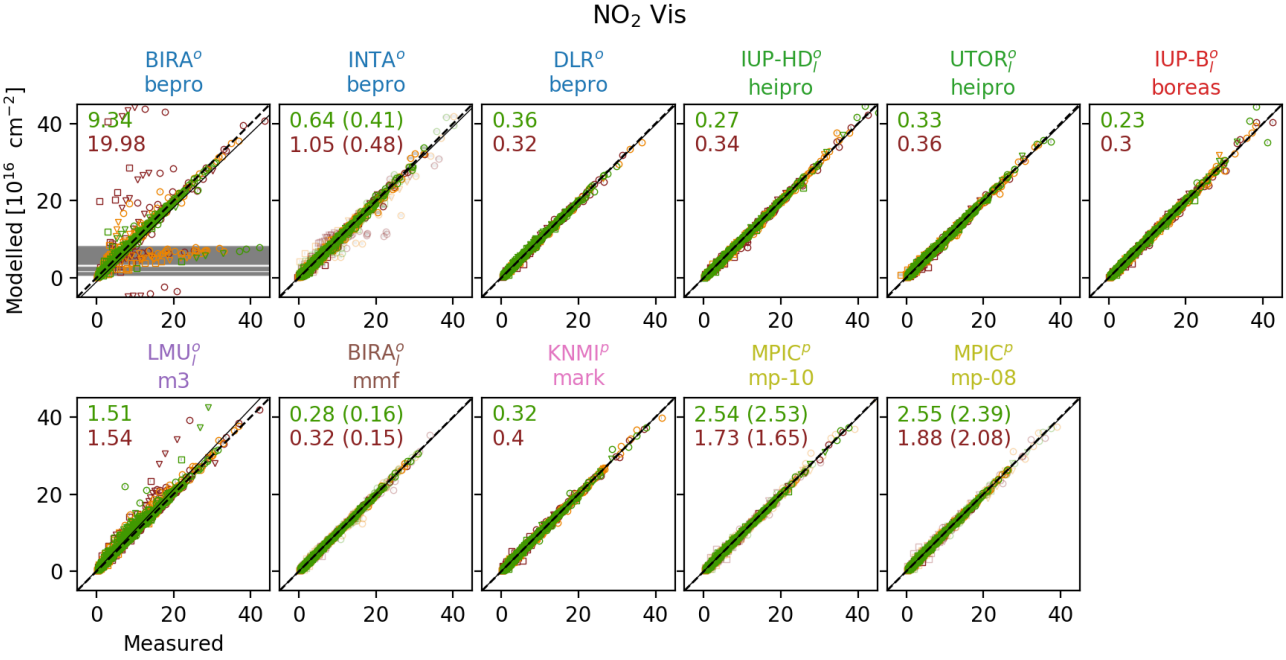

Figure S49. NO_2 Vis dSCD correlation. Legends of Fig. S45 apply.

S10.4 Aerosol optical thickness (AOT)

This section is equivalent to Sect. S10.4 in the main text.

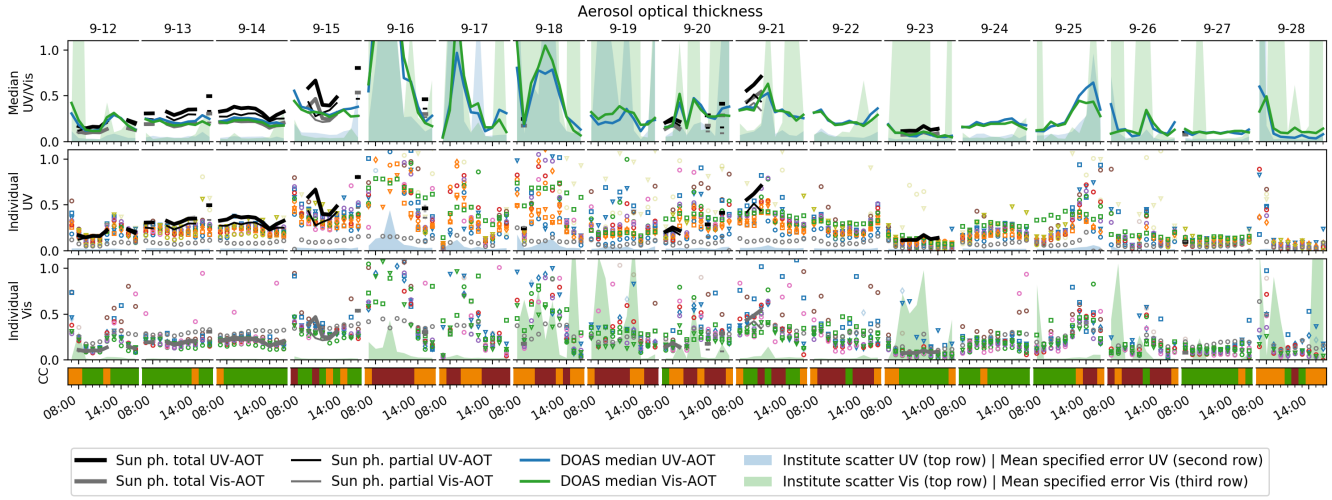

Figure S50. MAX-DOAS AOTs retrieved from the participant's own dSCDs in comparison to sun photometer data. Symbol and symbol colours are chosen according to Table 2 in the main text. Transparent symbols indicate data flagged as invalid. Top row: MAX-DOAS median results vs. the available supporting observations, according to the legend below the plot. ~~Hatched~~ The "institute scatter" areas ~~(starting at the top of the plot)~~ show the scattering ~~(standard deviation)~~ among the participants ~~(only~~ in terms of standard deviation with valid data considered ~~)~~only. Two lower rows: Comparison of the individual participants for the two spectral retrieval ranges. Here the coloured area is the average retrieval error, as specified by the participants.

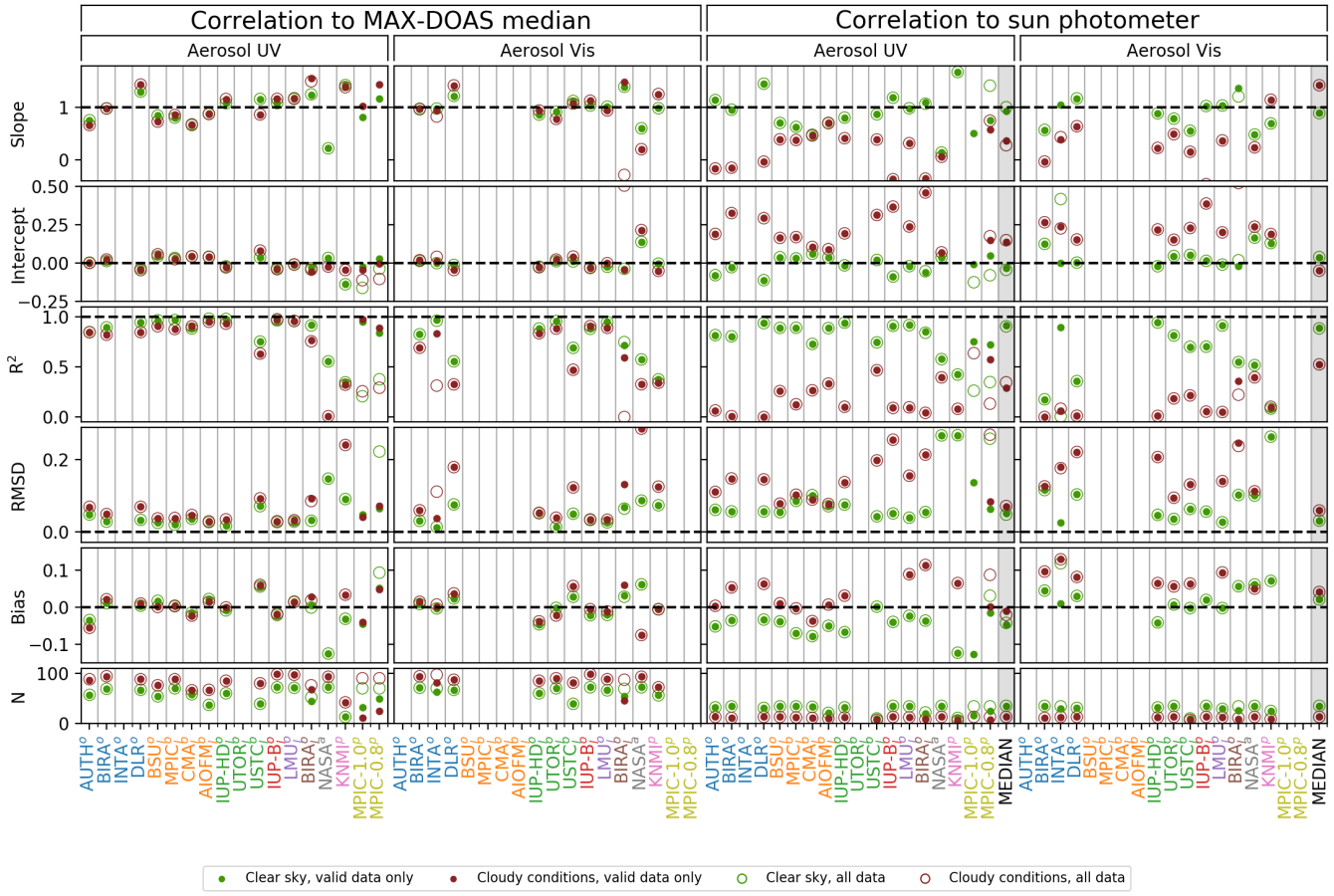

Figure S51. Correlation statistics for AOTs ~~retrieved from the participant's own dSCDs~~. The two left columns give an impression on the agreement among the institutes, as they show the correlation of the individual participant's retrieved AOT (ordinate of the underlying correlation plot) against the median (abscissa). The two right columns show the correlation against the sun photometer AOT (partial AOT in the case of OEM retrievals) instead of the median. Green and red symbols represent cloud-free and cloudy conditions, respectively. ~~Transparent symbols~~ Hollow circles represent values for all submitted data, ~~opaque symbols~~ the dots only consider data points flagged as valid. ~~The pies indicate, which fraction of~~ N is the ~~total~~ number of profiles ~~(170)~~ which contributed to the respective data points above. The total number of submitted profiles per participant and species were 170. On the right also the correlation between the MAX-DOAS median results and supporting observations are included (grey shaded columns).

S10.5 Trace gas vertical column densities (VCDs)

This section is equivalent to Sect. S10.5 in the main text.

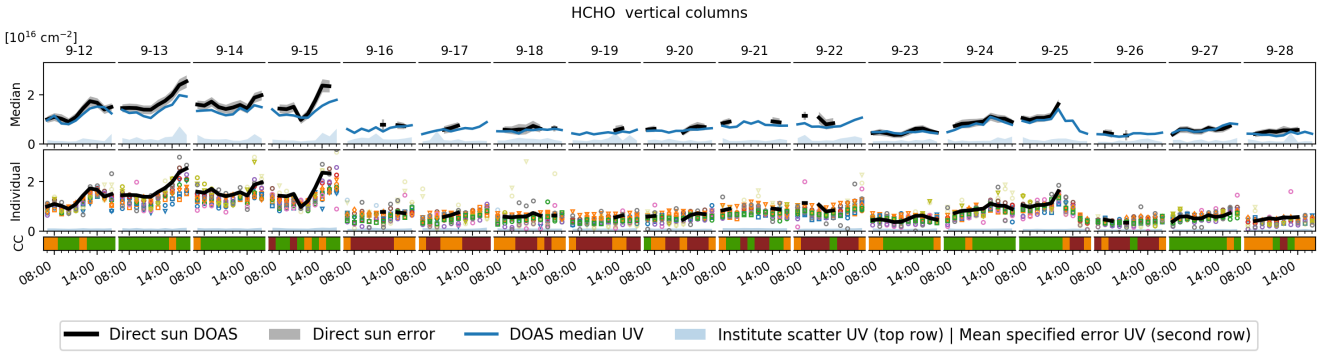

Figure S52. Comparison of MAX-DOAS HCHO VCDs retrieved from the participant's own dSCDs vs. direct-sun DOAS, NO_2 lidar and radiosonde. Descriptions of Fig. S50 apply.

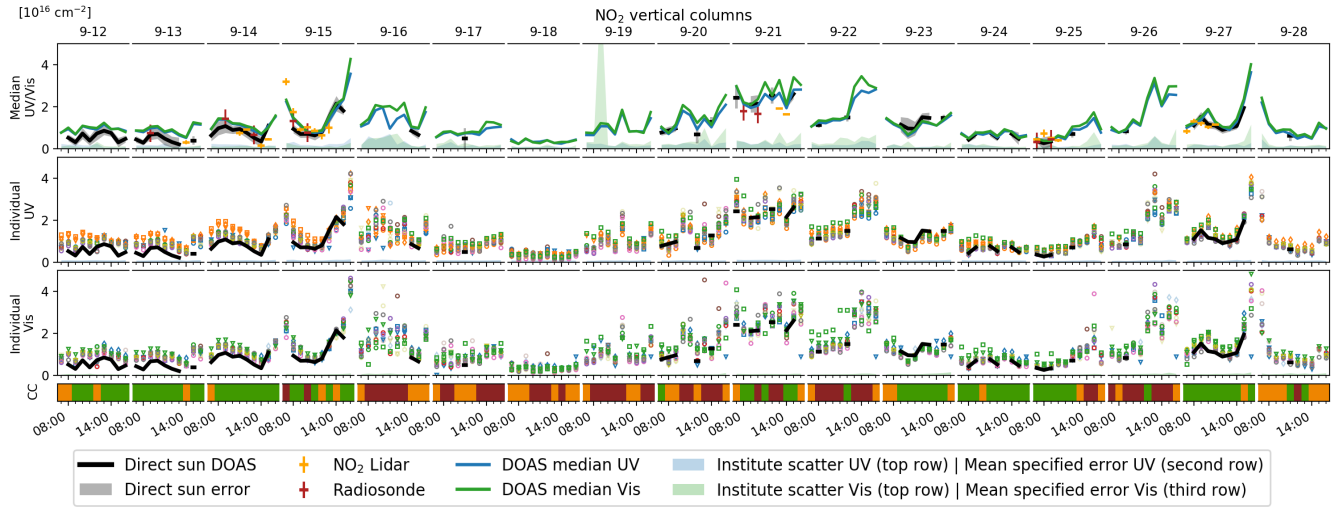

Figure S53. Comparison of MAX-DOAS NO_2 VCDs retrieved from the participant's own dSCDs vs. direct-sun DOAS. Descriptions of Fig. S50 apply.

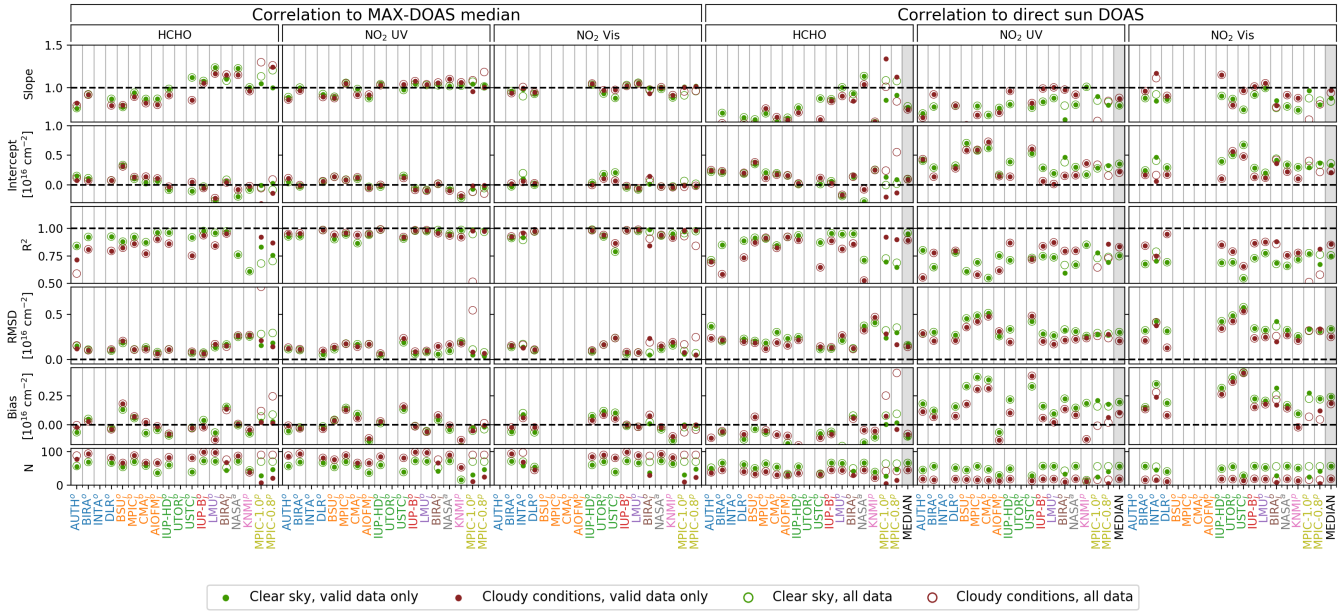

Figure S54. Correlation statistics of trace gas VCDs retrieved from the participant's own dSCDs. Legends and description of Fig. S51 apply.

S10.6 Trace gas surface concentrations

This section is equivalent to Sect. S10.6 in the main text.

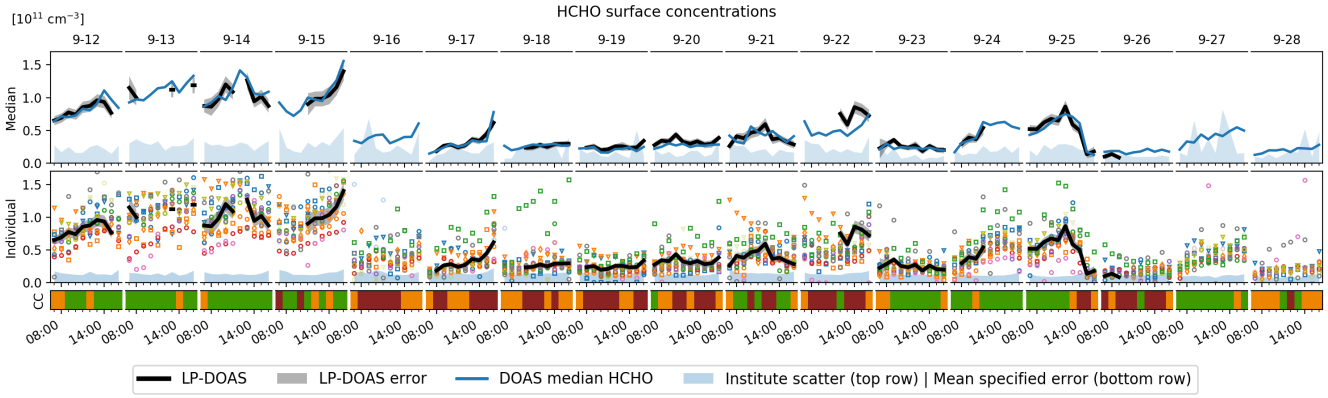

Figure S55. Comparison of MAX-DOAS HCHO surface concentrations retrieved from the participant's own dSCDs. Description of Fig. S50 applies.

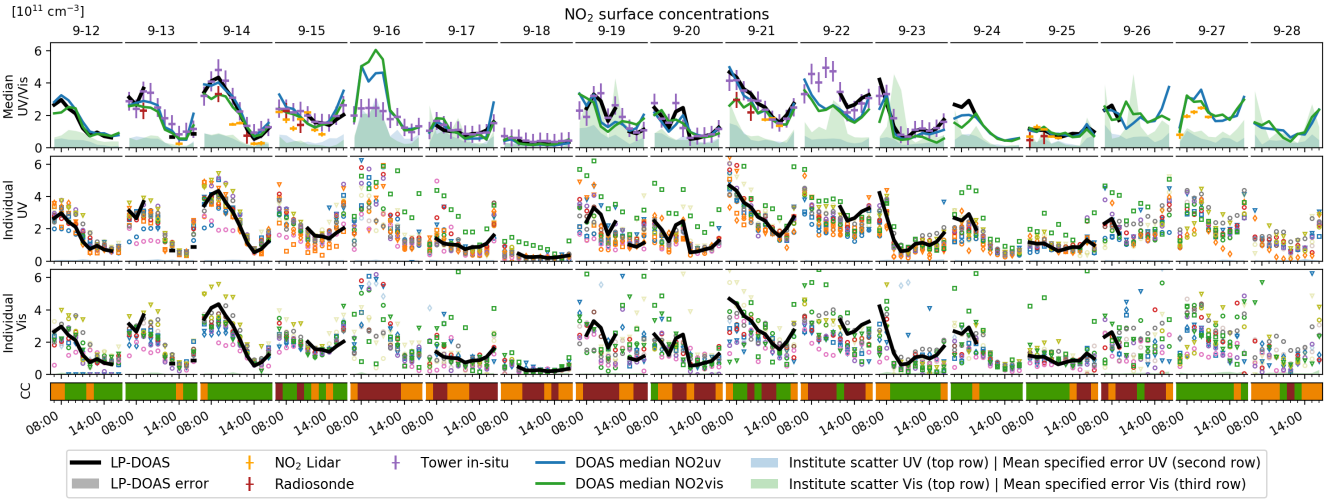

Figure S56. Comparison of MAX-DOAS NO_2 surface concentrations retrieved from the participant's own dSCDs. Description of Fig. S50 applies.

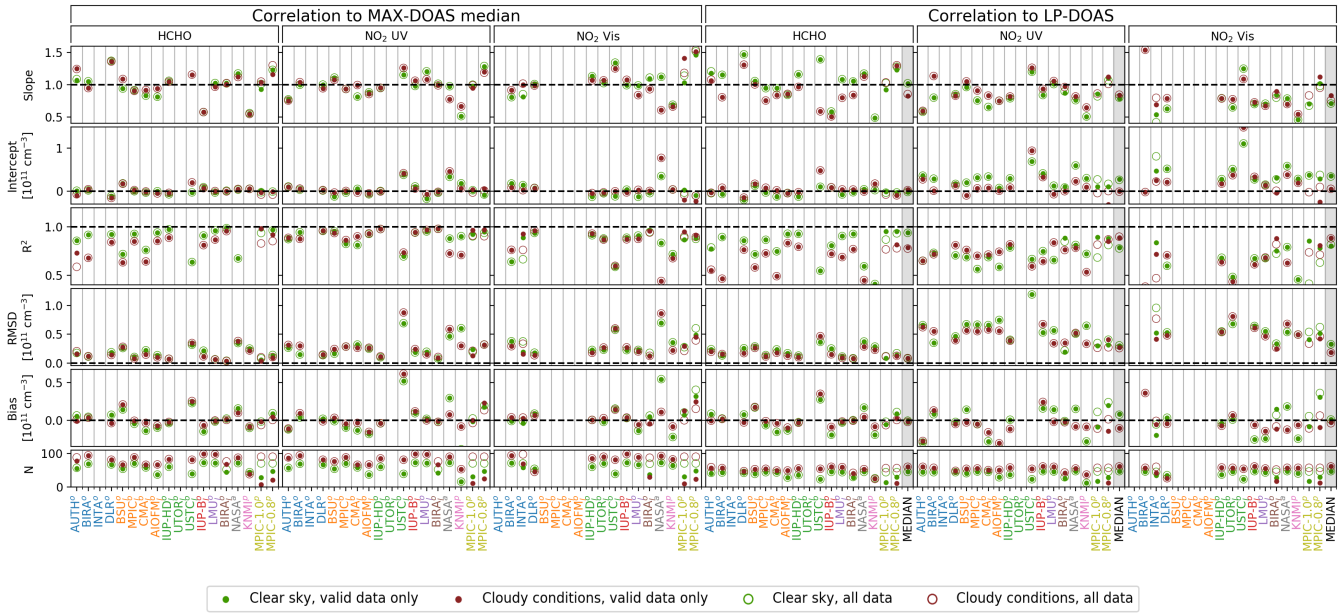

Figure S57. Correlation statistics of trace gas surface concentrations retrieved from the participant's own dSCDs. Basic description of Fig. S51 applies.

S10.7 ~~NO_2 UV-Vis comparison~~

~~This section is equivalent to Sect. ?? in the main text.~~

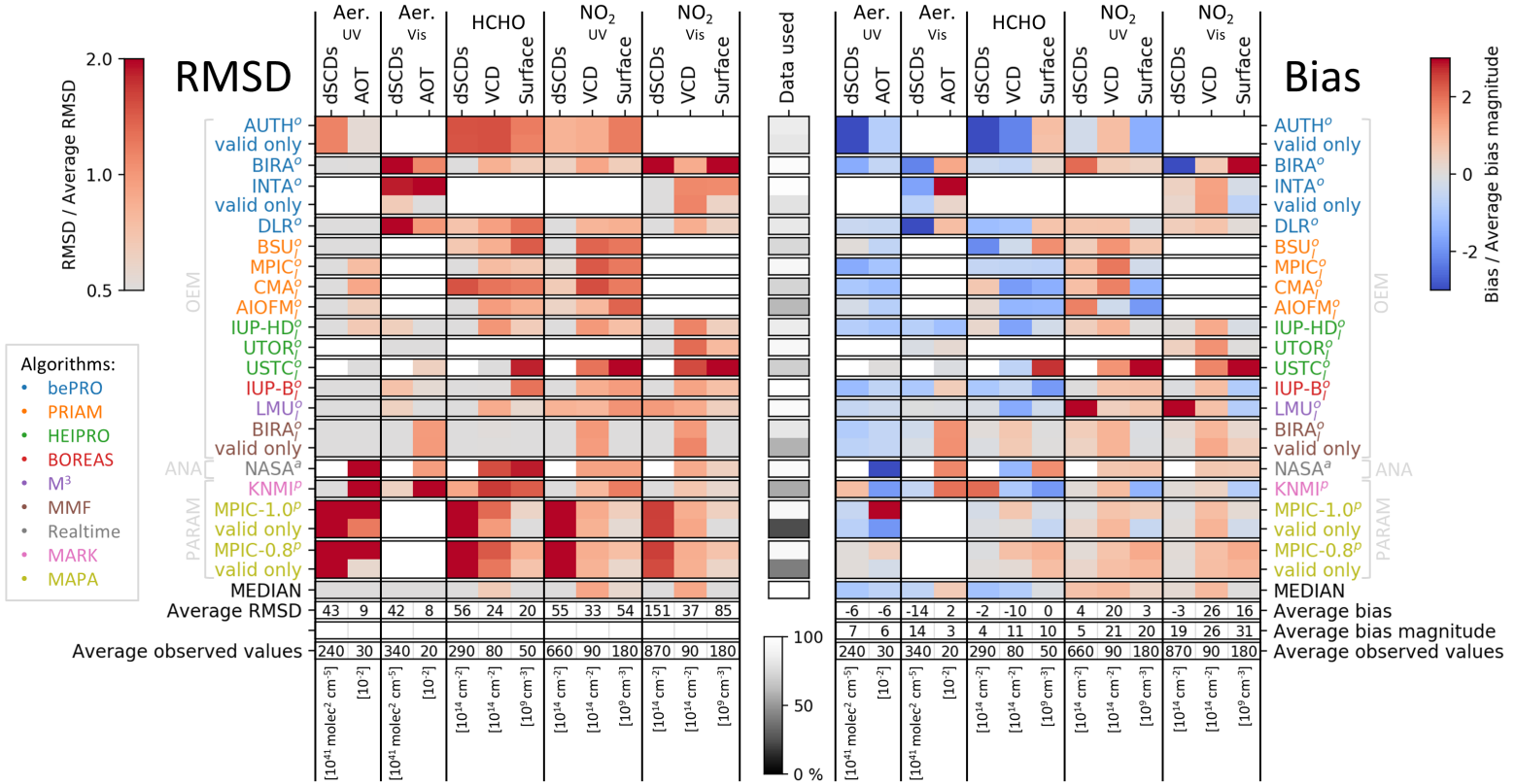

Figure S58. ~~Correlation~~ Summary of ~~MAX-DOAS VCDs retrieved from~~ the ~~participant's own dSCDs~~ comparisons in Sect. S10 for clear-sky conditions. Left panel shows RMSD, right panel shows Bias. Average values of RMSD (Bias) define the ~~UV and~~ colour scale of each column of the ~~Vis spectral range~~left (right) panel as indicated by the color bars on the top left (top right) of the figure. ~~Marker colours~~ Values of AOT, VCD and ~~marker shapes~~ surface concentration are given with respect to the corresponding supporting observations. White spaces indicate no data. Average observed values (bottom row) are rounded campaign averages of the ~~cloud conditions~~supporting observations. Average Bias and Average Bias magnitude values (third last and second last row of right panel) represent the averages over the signed and the absolute Bias values, respectively. The "data used"-column in the center indicates which fraction of the maximum number (170) of available profiles has been used. Participants who submitted flags are represented by two rows: one considering all data and one using only those flagged as valid ("valid only").

~~Correlation of MAX-DOAS surface concentrations retrieved from the participant's own dSCDs in the UV and the Vis spectral range. Legends of Fig. **??** apply.~~

References

Apituley, A., Wilson, K., Potma, C., Volten, H., and de Graaf, M.: Performance Assessment and Application of Caeli–A highperformance Raman lidar for diurnal profiling of Water Vapour, Aerosols and Clouds, in: Proceedings of the 8th International Symposium on Tropospheric Profiling, pp. 19–23, S06-O10-1-4, Delft/KNMI/RIVM Delft, Netherlands, 2009.

Beirle, S., Dörner, S., Donner, S., Remmers, J., Wang, Y., and Wagner, T.: The Mainz profile algorithm (MAPA), Atmospheric Measurement Techniques, 12, 1785–1806, https://doi.org/10.5194/amt-12-1785-2019, https://www.atmos-meas-tech.net/12/1785/2019/, 2019.

Bösenberg, J., Matthias, V., Amodeo, A., Amoiridis, V., Ansmann, A., Baldasano, J. M., Balin, I., Balis, D., Böckmann, C., Boselli, A., Carlsson, G., Chaikovsky, A., Chourdakis, G., Comeron, A., Tomasi, F. D., Eixmann, R., Freudenthaler, V., Giehl, H., Grigorov, I., Hagard, A., Iarlori, M., Kirsche, A., Kolarov, G., Komguem, L., S. Kreipl, W. K., Larcheveque, G., Linné, H., Matthey, R., Mattis, I., Mekler, A., Mironova, I., Mitev, V., Mona, L., Müller, D., Music, S., Nickovic, S., Pandolfi, M., Papayannis, A., Pappalardo, G., Pelon, J., Perez, C., Perrone, R., Persson, R., Resendes, D. P., Rizi, V., Rocadenbosch, F., Rodrigues, J. A., Sauvage, L., Schneidenbach, L., Schumacher, R., Shcherbakov, V., Simeonov, V., Sobolewski, P., Spinelli, N., Stachlewska, I., Stoyanov, D., Trickl, T., Tsaknakis, G., Vaughan, G., Wandinger, U., Wang, X., Wiegner, M., Zavrtanik, M., and Zerefos, C.: EARLINET: A European Aerosol Research Lidar Network to establish an aerosol climatology, Report 348, ISSN 0937-1060, 192 pp., Max-Planck-Institut für Meteorologie, 2003.

Chan, K. L., Wang, Z., Ding, A., Heue, K.-P., Shen, Y., Wang, J., Zhang, F., Hao, N., and Wenig, M.: MAX-DOAS measurements of tropospheric NO_2 and HCHO in Nanjing and the comparison to OMI observations, Atmospheric Chemistry and Physics Discussions, 2019, 1–25, https://doi.org/10.5194/acp-2018-1266, https://www.atmos-chem-phys-discuss.net/acp-2018-1266/, 2019.

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, Atmospheric Measurement Techniques, 3, 863–878, https://doi.org/10.5194/amt-3-863-2010, https://www.atmos-meas-tech.net/3/863/2010/, 2010.

Esri, EsriNL, Rijkswaterstaat, Intermap, NASA, NGA, Kadaster, U. ., Esri, HERE, Garmin, P, I., and METI: arcGIS World Topo Map, 2018.

Mayer, B. and Kylling, A.: Technical note: The libRadtran software package for radiative transfer calculations - description and examples of use, Atmospheric Chemistry and Physics, 5, 1855–1877, https://doi.org/10.5194/acp-5-1855-2005, https://www.atmos-chem-phys.net/5/1855/2005/, 2005.

Ortega, I., Berg, L. K., Ferrare, R. A., Hair, J. W., Hostetler, C. A., and Volkamer, R.: Elevated aerosol layers modify the O2–O2 absorption measured by ground-based MAX-DOAS, Journal of Quantitative Spectroscopy and Radiative Transfer, 176, 34 – 49, https://doi.org/https://doi.org/10.1016/j.jqsrt.2016.02.021, http://www.sciencedirect.com/science/article/pii/S0022407315301746, 2016.

Pappalardo, G., Amodeo, A., Apituley, A., Comeron, A., Freudenthaler, V., Linné, H., Ansmann, A., Bösenberg, J., D'Amico, G., Mattis, I., Mona, L., Wandinger, U., Amiridis, V., Alados-Arboledas, L., Nicolae, D., and Wiegner, M.: EARLINET: towards an advanced sustainable European aerosol lidar network, Atmospheric Measurement Techniques, 7, 2389–2409, https://doi.org/10.5194/amt-7-2389-2014, https://www.atmos-meas-tech.net/7/2389/2014/, 2014.

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, Atmospheric Measurement Techniques Discussions, 2019, 1–30, https://doi.org/10.5194/amt-2019-33, 2019.

~~Rodgers, C. D.: Inverse methods for atmospheric sounding : theory and practice, World Scientific Publishing, 2000.~~

Wagner, T. and Beirle, S.: Estimation of the horizontal sensitivity range from MAX-DOAS O_4 observations, Tech. rep., QA4ECV, http://www.qa4ecv.eu/sites/default/files, 2016.

Wagner, T., Deutschmann, T., and Platt, U.: Determination of aerosol properties from MAX-DOAS observations of the Ring effect, Atmospheric Measurement Techniques, 2, 495–512, https://doi.org/10.5194/amt-2-495-2009, https://www.atmos-meas-tech.net/2/495/2009/, 2009.

Wagner, T., Beirle, S., Benavent, N., Bösch, T., Chan, K. L., Donner, S., Dörner, S., Fayt, C., Frieß, U., García-Nieto, D., Gielen, C., González-Bartolome, D., Gomez, L., Hendrick, F., Henzing, B., Jin, J. L., Lampel, J., Ma, J., Mies, K., Navarro, M., Peters, E., Pinardi, G., Puentedura, O., Puķīte, J., Remmers, J., Richter, A., Saiz-Lopez, A., Shaiganfar, R., Sihler, H., Van Roozendael, M., Wang, Y., and Yela, M.: Is a scaling factor required to obtain closure between measured and modelled atmospheric O_4 absorptions? An assessment of uncertainties of measurements and radiative transfer simulations for 2 selected days during the MAD-CAT campaign, Atmospheric Measurement Techniques, 12, 2745–2817, https://doi.org/10.5194/amt-12-2745-2019, https://www.atmos-meas-tech.net/12/2745/2019/, 2019.