Fiducial Reference Measurements for Ground-Based DOAS Air-Quality Observations

FRM4DOAS

ESA Contract No. 4000118181/16/I-EF

Deliverable D10:

Processing System Validation

Date: 18/06/2019

Version: 1.6

Contributing authors:

Andreas Richter (IUPUB)
Tim Bösch (IUPUB)
Jan-Lukas Tirpiz (IUPHD)
Steffen Beirle (MPIC)
Thomas Wagner (MPIC)
Martina Friedrich (IASB-BIRA)
François Hendrick (IASB-BIRA)
Michel Van Roozendael (IASB-BIRA)

and all FRM_4DOAS partners

Content

1 Introduction 4

2 Approach taken to Processing System Validation 4

3 Validation using CINDI-2 data 5

3.1 Overview of results 6

3.2 Information content 12

3.3 DSCD retrievals 14

3.4 Validation of surface concentrations and columns 15

3.5 Profile statistics 18

3.6 Comparison with independent retrievals 20

4 Validation using IUP-UB retrievals 21

4.1 Comparison of tropospheric vertical columns with AMF approach 21

4.2 Comparison of NO_2 tropospheric vertical columns with BOREAS 23

4.3 Comparison of AOD with BOREAS 28

4.4 Comparison of surface NO_2 and extinction 33

4.5 Evaluation of uncertainties 38

4.6 Summary and conclusions 39

5 References 40

1 Introduction

The Fiducial Reference Measurements for Ground-Based DOAS Air-Quality Observations (FRM$_4$DOAS, see http://frm4doas.aeronomie.be/) is a 2-year ESA project which started in July 2016. It aims at further harmonization of MAXDOAS systems and data sets, through the

- specification of best practices for instrument operation
- demonstration of a centralised NRT (near-real-time/6-24h latency) processing system for MAXDOAS instruments operated within the international Network for the Detection of Atmospheric Composition Change (NDACC)
- establishment of links with other UV-Visible instrument networks, e.g. Pandonia Global Network (PGN)

While the project itself is limited to tropospheric and stratospheric NO_2 vertical profiles, total O_3 columns, and tropospheric HCHO profiles from a small number of stations, the aim is to collect and create the necessary information, guidelines and infrastructure which can be the basis for a network including many more MAXDOAS instruments and covering all MAXDOAS products.

In this document, the results of a preliminary validation and verification of the FRM4DOAS prototype processing system are reported.

2 Approach taken to Processing System Validation

Validation of the Processing System is complex and includes many different aspects, including validation of technical aspects such as timeliness of processing of data, consistency of formats, feedback to the users and proper reaction to error situations. These more technical points will not be covered in this document. So far they are only partly documented in the Prototype System Architecture document (D8) and Processing System Software documentation and will be further extended in the FRM4DOAS CCN 02.

The other extreme of validation would be the rigorous validation of the profiles retrieved by the system using independent profile information. Such a validation exercise is a major task and limited by availability of reference data. It is therefore not attempted in this report.

The approach taken here is to

1. Apply the processing system to data from a subset of instruments operating during the CINDI-2 campaign, evaluate the data for consistency, plausibility and coverage, and compare the results to validation data available from the campaign. This part is based on work performed in the framework of the CINDI-2 profile intercomparison exercise to be published in Tirpitz et al. (2019).
2. Apply the processing system to a full month of data of the Bremen MAX-DOAS instruments in Bremen, Athens and Ny-Alesund, to evaluate these data for plausibility, consistency and coverage, and to compare the results with those obtained from the standard data evaluation performed at IUP Bremen.

The subset of instruments to be used for the results of the CINDI-2 campaign is listed in Table 1. Instruments of different campaign performance where chosen to cover a broader range of input data quality. Instrument performances were assessed during the CINDI-2 semi-blind intercomparison exercise as described in Kreher et al. (2019). The instrument subset for this validation includes two double channel high signal-to-noise (SNR) scientific instruments (BIRA and IUP-Bremen), one scientific single channel standard SNR instrument (AUTH) and one commercial two channel Envimes instrument (DLR/USTC). For the analysis of this data set, the processing system was applied to the lv1 data submitted during the CINDI-2 campaign. The version of the processing system was validation_01.0.

For the data from the IUP-Bremen instruments, August 2018 was selected as representative month as it was characterised by few data gaps, overall favourable weather conditions and full availability of consistent NRT data produced by the processing system in version validation_01.0. Within the FRM4DOAS processing chain, profile inversion is performed with the OE-based MMF algorithm version validation_1.0 (Friedrich et al., 2019) and the parameter based MAPA algorithm v0.96 (Beirle et al., 2019).

All four options of the processing system (the optimal estimation retrieval MMF and three flavours of the parametric MAPA retrieval using a fixed O_4 scaling factor of 0.8, a fixed scaling factor of 1.0 and a fitted O_4 scaling factor were included. No O_4 scaling is applied in the current MMF implementation.

3 Validation using CINDI-2 data

The CINDI-2 campaign took place in Cabauw, the Netherlands in September 2016. MAX-DOAS data from all instruments are available for the time period September 12 – 28. Weather conditions were mostly sunny on many days including September 9, 12, 13, 23, 24, and 27, with varying levels of clouds on the remaining days.

Table 1: Legend of colours and symbols used in section 3

Algorithm	Institute	Code	Symbol
MAPA (0.8)	BIRA	BIR	●
	IUP-Bremen	IUP	▲
	AUTH	AUT	■
	DLR/USTC	DLR	▲
MAPA (1.0)	BIRA	BIR	●
	IUP-Bremen	IUP	▲
	AUTH	AUT	■
	DLR/USTC	DLR	▲
MAPA (free)	BIRA	BIR	●
	IUP-Bremen	IUP	▲
	AUTH	AUT	■
	DLR/USTC	DLR	▲
MMF	BIRA	BIR	●
	IUP-Bremen	IUP	▲
	AUTH	AUT	■
	DLR/USTC	DLR	▲

For the validation of the retrievals on MAX-DOAS data, different measurements are available, including

- Longpath-DOAS (LP-DOAS) measurements of NO_2 and HCHO for the surface concentrations
- Direct sun DOAS measurements for NO_2 tropospheric columns
- Sun-photometer measurements for AOD at different wavelengths

- Ceilometer data for relative extinction profiles
- NO_2 vertical profiles from the KNMI NO_2 sondes and the RIVM NO_2 lidar observations for a few short time periods during the campaign

Data from the four selected instruments was processed by the prototype processing system for NO_2, HCHO and aerosol in the UV and visible spectral ranges. For NO_2, only the visible retrievals were evaluated. As the AUTH instrument only has an UV channel, no NO_2 and visible aerosol data from that instrument are available. All data was then converted to the format required for the profile intercomparison study for CINDI-2, and run through the software and analysis package developed by J.-L. Tirpitz for this exercise.

One important part of the output of the processing system are quality flags for each profile. As the CINDI-2 intercomparison software does not allow for warnings, only for valid and in-valid profiles, the QA flags from the FRM4DOAS processor have been mapped onto the CINDI-2 flags by making all MMF retrievals valid unless an error was flagged while MAPA retrievals were flagged as valid only if neither errors, nor warnings were raised. This different treatment of the two retrieval types was recommended by the retrieval developers. In the following comparisons, often results are shown for both, the complete data set and only those data which were flagged as valid. In addition, sometimes cloudy and clear-sky scenarios are reported separately.

3.1 Overview of results

As a first step, an overview on the profiles retrieved by the 4 algorithms on the data from the 4 (3) instruments is given in Figure 1 - Figure 4. Where available, the lowest rows also show independent comparison data.

For aerosol retrievals at 360 nm, the following observations can be made:

1. On average, all three MAPA retrievals create vertical profiles with more aerosol in higher layers than MMF
2. MMF has a tendency to retrieve elevated layers, both in the presence and absence of clouds
3. In the presence of clouds, the elevated layers retrieved by MMF are often in similar altitudes as shown in the ceilometer data
4. From the visual impression, MAPA AOD appears to be larger than that retrieved by MMF.
5. Differences between MMF and the three MAPA variants are larger than difference between different MAPA versions
6. While results for the two high SNR instruments BIRA and IUP-UB are often similar, results for AUTH are often different (less vertical structure) while the DLR instrument is often in between
7. A large number of profiles are flagged as invalid, in particular for MAPA. Flagging between MAPA and MMF appears to be not consistent

For aerosol retrievals at 477 nm, similar observations are made with some differences. In particular

1. There are even more defined layers in MMF at visible wavelengths
2. The consistency between instruments is better in the vis
3. There are less profiles flagged as invalid

4. Flagging between UV and visible is not consistent

When comparing results for HCHO, the most apparent features are

1. The temporal evolution of HCHO over the campaign is similar in all retrievals
2. There are quite a few profiles where unrealistically large HCHO values are retrieved by MAPA, in particular if the O_4 factor is free to vary. However, all these cases are flagged as invalid.
3. As was the case for aerosols, MMF often retrieves elevated layers, which is not the case in any of the MAPA variants.
4. HCHO profiles retrieved from BIRA and IUP-UB data are quite consistent, while AUTH profiles seem to be confined more to the lowest layers. DLR profiles are somewhere in between.
5. In MAPA retrievals, many if not most profiles are flagged; for example, not a single HCHO profile of the AUTH instrument was flagged as valid.

Finally, for NO_2 the most important points are

1. Most profiles are confined to the lowest layers
2. There is much better overall consistency between different retrievals as for the other three products
3. There still are some failed profiles for MAPA free
4. Also for NO_2, a large fraction of the profiles is flagged as invalid
5. The DLR retrievals show consistently elevated layers for NO_2 for MMF during the first days

The unusual behaviour of the DLR profiles during the first 4 days could be related to instrument mispointing which was later corrected.

The reduced level of details in the AUTH data could be linked to larger noise and reduced information content as discussed in section 3.2.

The appearance of elevated layers in MMF results is indicative of profile oscillation as they are well known for OE retrievals if not strongly constrained.

The large number of flagged results is an important result which will be discussed again in section 3.5.

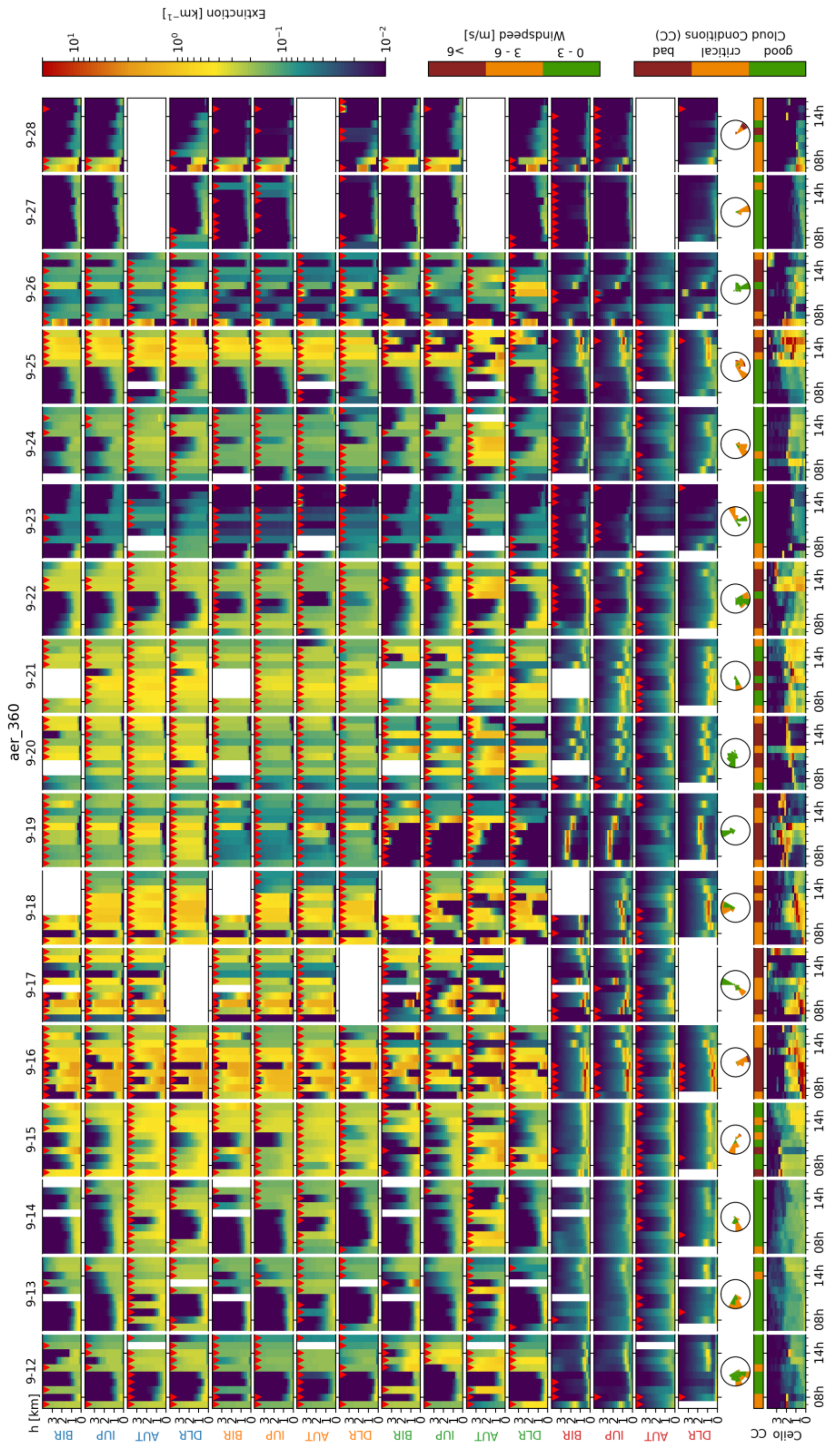

Figure 1: overview on aerosol retrieval results for 360 nm. Results for the 4 algorithms are shown from top to bottom (MAPA 0.8, MAPA 1.0, MAPA free, MMF, each for the four instruments providing UV radiances. The lowest three lines show wind direction, cloud conditions (derived from MAX-DOAS data with the MPIC cloud classification algorithm as described in Wagner et al., 2014) and ceilometer profiles (backscatter signal scaled with sun photometer or MAX-DOAS median AOD, respectively) for comparison. Red triangles indicate flagged results.

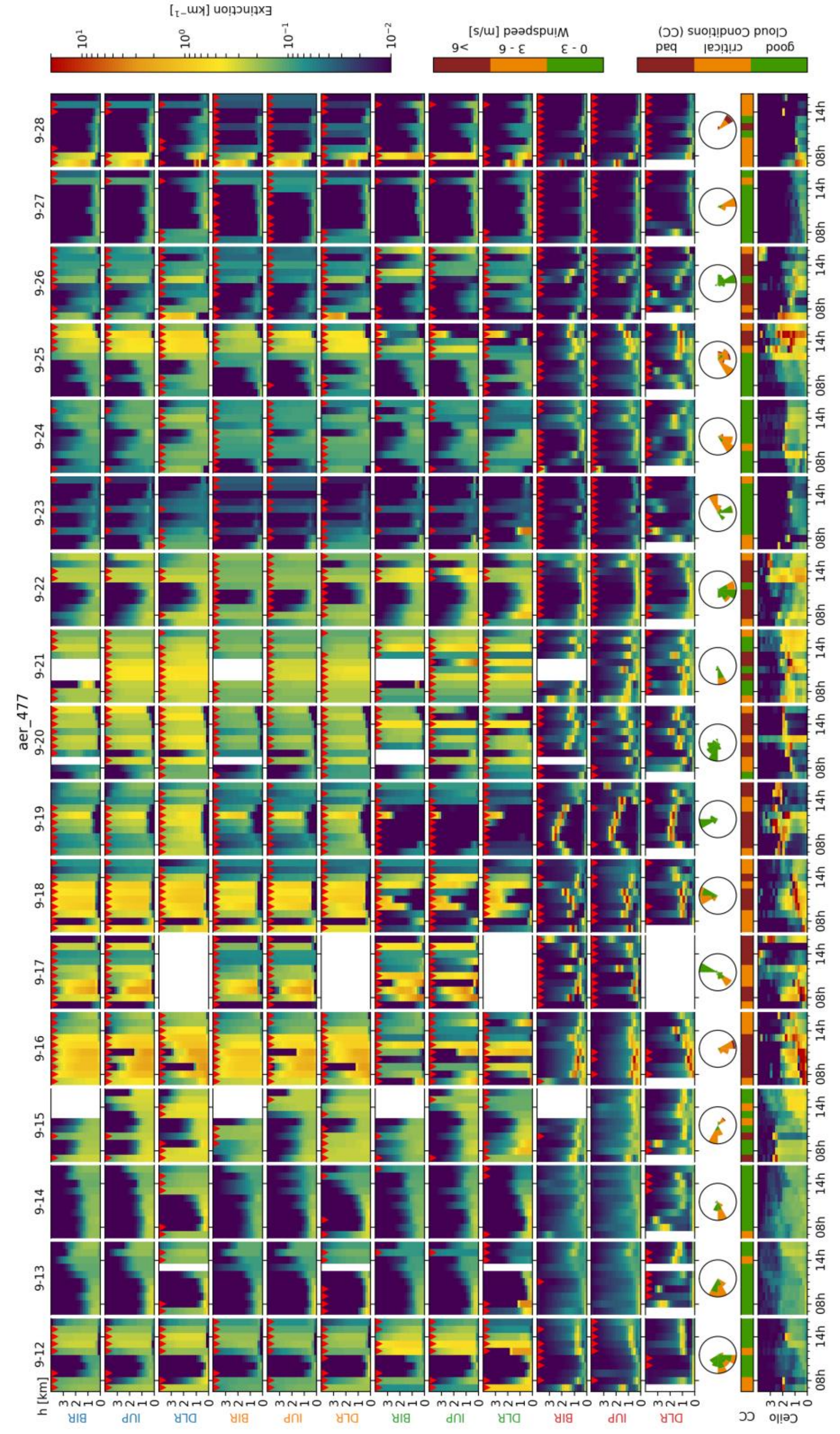

Figure 2: As Figure 1, but for aerosol retrievals at 477 nm

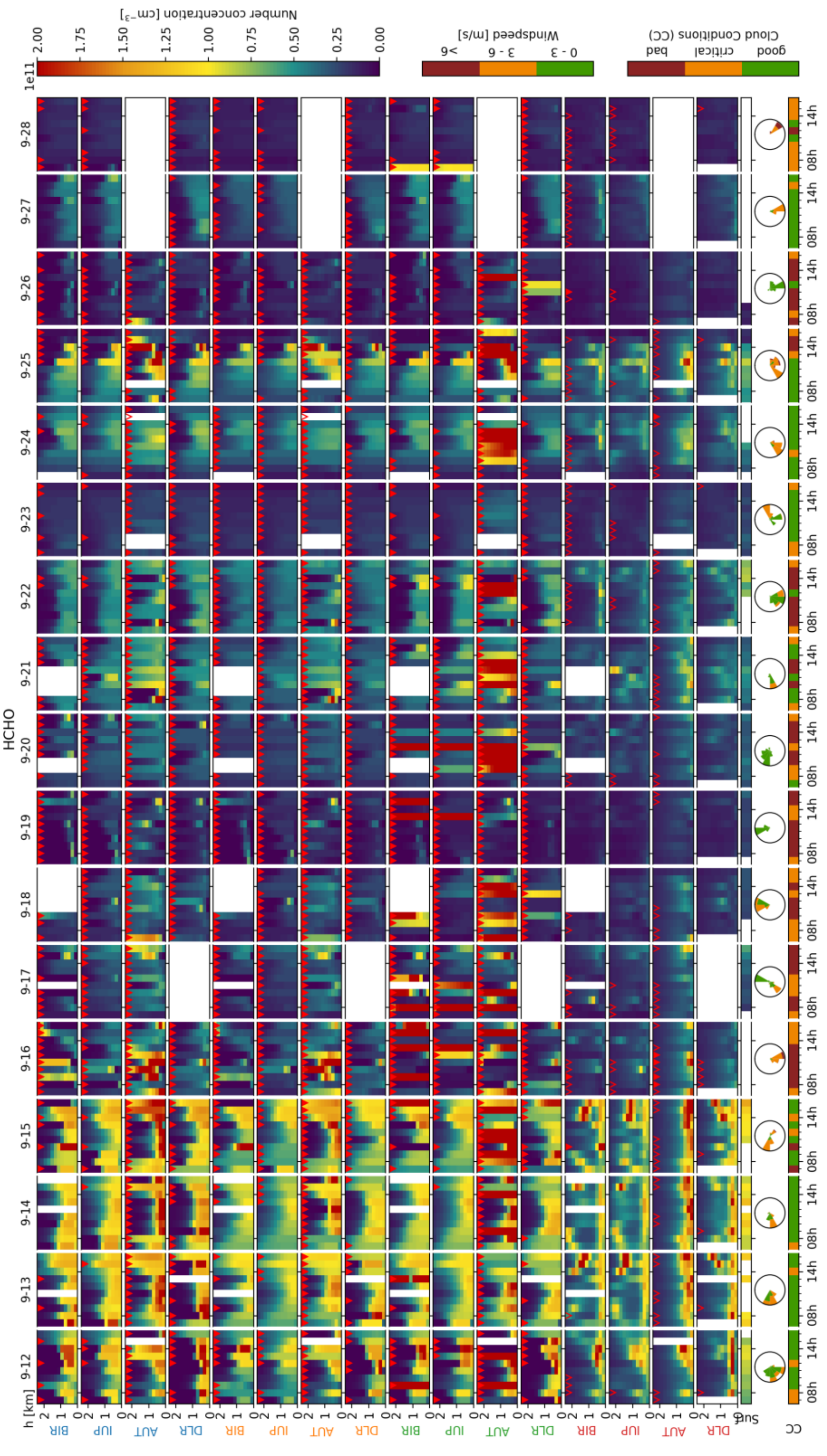

Figure 3: As Figure 1, but for HCHO retrievals. The line above the wind direction figures shows surface HCHO concentrations as observed by LP-DOAS. Faceless triangles indicate an invalid underlying aerosol retrieval.

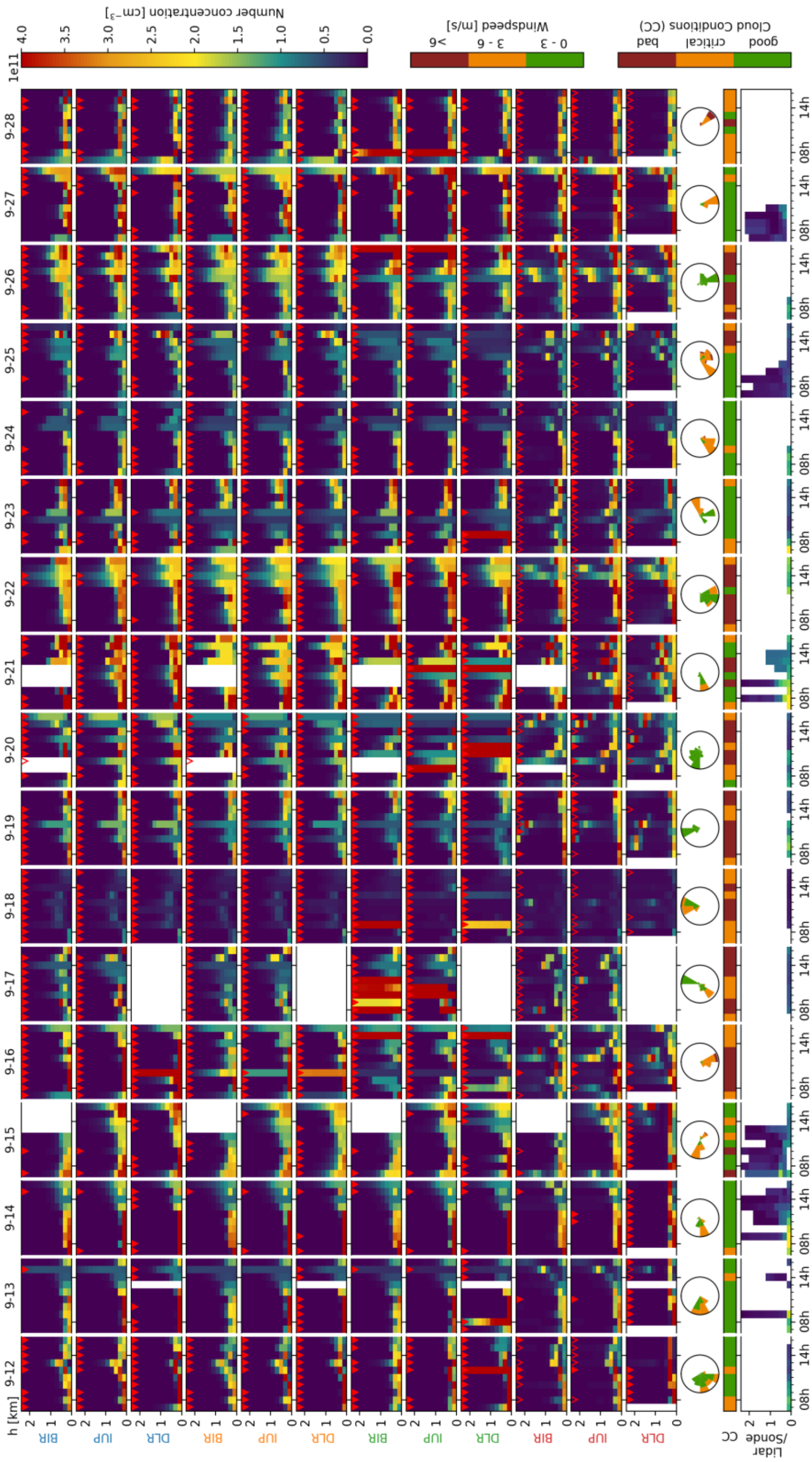

Figure 4: As Figure 1, but for NO_2 retrievals. The lowest line shows LP-DOAS surface concentrations as well as NO_2 sonde and NO_2 lidar profiles where available.

3.2 Information content

One of the advantages of optimal estimation retrievals is, that they provide detailed information on the information content of the results. Two important quantities in this context are (1) the Averaging Kernels (AVK) which represent the sensitivity of the retrieved profile in one altitude to changes in the real profile at any altitude and (2) the number of degrees of freedom (DOF), which basically is the number of independent pieces of information that the retrieval provides. In the output of the FRM4DOAS processor, this information is only available from MMF.

In Figure 5, median AVKs are shown for all four products and the 4(3) instruments. The main conclusions from these figures are that

1. During the CINDI-2 campaign, the averaging kernels peak in the lowest km for the UV retrievals and in the lowest 1.5 km for the visible retrievals,
2. Most of the information is very close to the surface and the gain in information and the vertical resolution drop quickly with altitude (AVKs quickly broaden and are less peaked)
3. The visible AVKs are nearly identical for the instruments in the visible spectral range while in the UV, the DLR instrument has less vertical information than BIRA and IUP-UB but more than AUTH which has clearly lower information content, presumably because of higher noise levels
4. There is a tendency for more information in clear-sky scenes and if flagging is applied, but the differences are relatively small
5. The number of degrees of freedom is larger for HCHO than for the aerosol retrievals and largest for NO_2.

In Figure 6, the temporal evolution of the degrees of freedom is shown over the course of the CINDI-2 campaign. As can be seen, the DOF remains surprisingly constant over time, even during days with full or broken cloud coverage. In the visible aerosol retrievals as well as in the NO_2 profiles, the profiles from the different instruments have very similar information content while for the UV retrievals, AUTH has lower information content for both aerosols and HCHO.

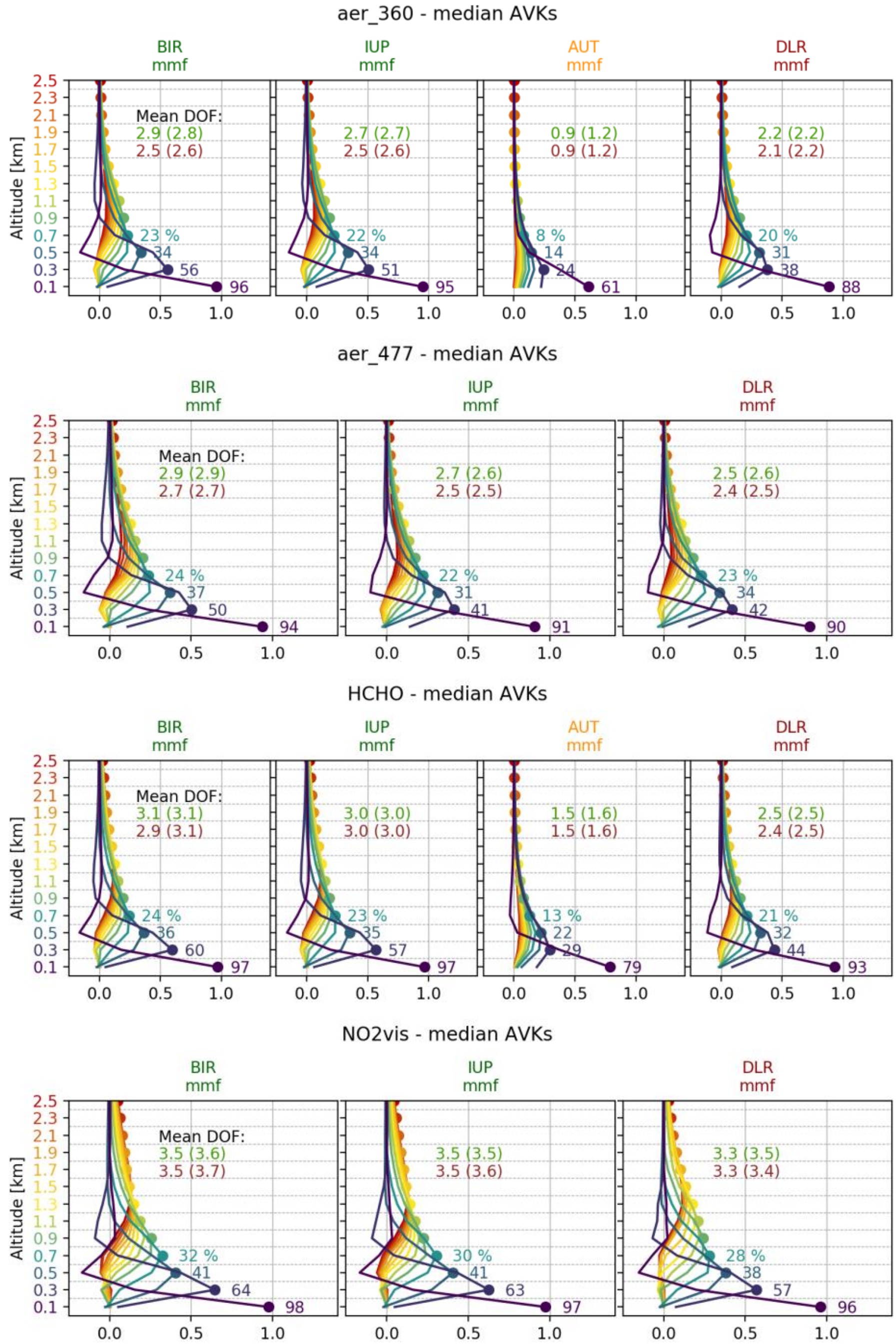

Figure 5: Median averaging kernels of the MMF profiles for aerosols at 360 nm and 477 nm, formaldehyde and NO_2 (from top to bottom). Also indicated are the degrees of freedom (DOF) for clear sky (green) and cloud scenes (red). Values in parenthesis are for valid data only.

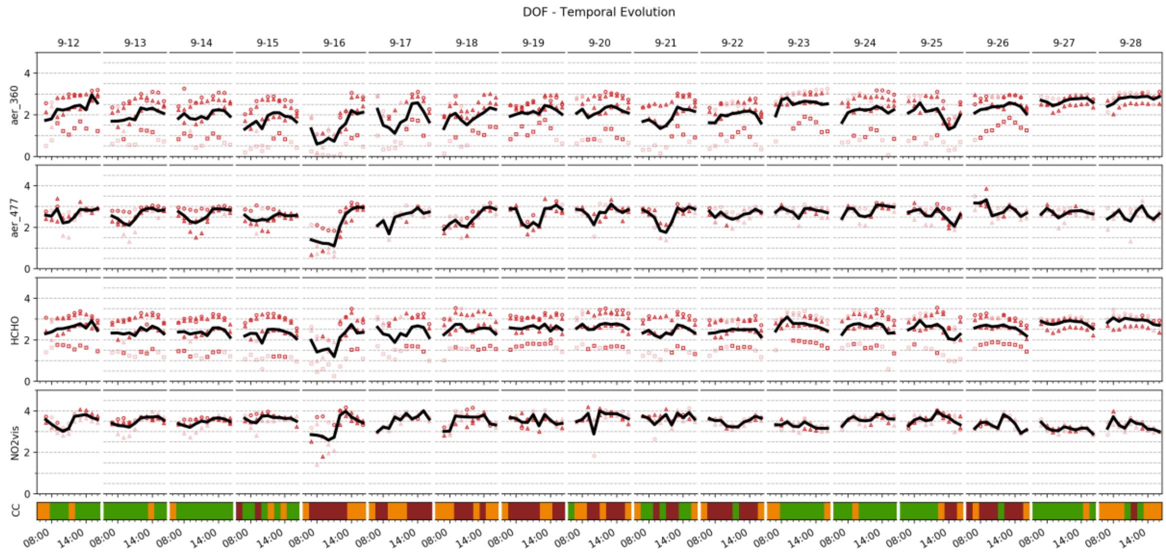

Figure 6: Temporal evolution of the MMF derived degrees of freedom during the CINDI-2 campaign. The lowest line indicates cloud conditions.

3.3 DSCD retrievals

One important check of the quality of a profile retrieval is the comparison of the measured and modelled differential slant column densities (DSCDs). This depends on many factors, including the uncertainty of the measurements, the accuracy of the forward model and the appropriateness of the retrieved vertical profiles of aerosol and trace gases.

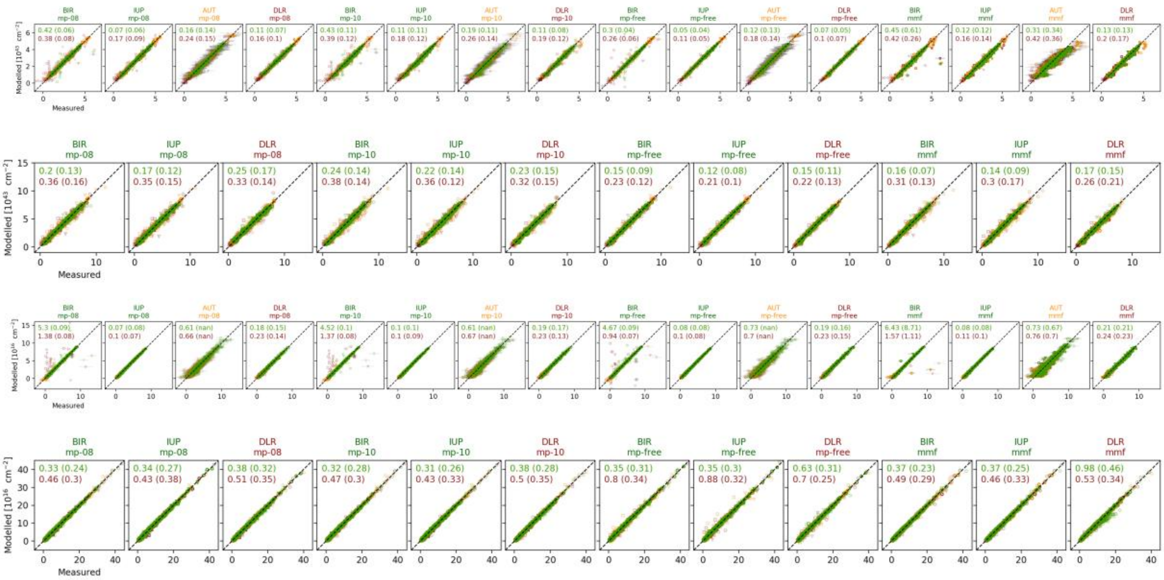

Figure 7: Correlations of measured and modelled DSCDs for O_4 in the UV, O_4 in the visible, HCHO, and NO_2. In each line, again results from all contributing instruments are shown, grouped by retrieval (MAPA 0.8, MAPA 1.0, MAPA free, and MMF). The numbers indicate the RMS of the DSCDs, for clear sky (green) and cloudy (red) conditions as well as for all data and in brackets results only for valid data.

In Figure 7, correlation plots are shown for all 4 products, all 4 retrieval and all 4(3) instruments. At first glance, mainly the different noise levels are apparent, resulting in more scatter in the UV than the visible, in particular for data from the AUTH instrument. There also are some clear outliers in the BIRA UV data (some even out of the plotting range) because of instrumental problems on some days. Comparing the four retrievals among each other, MAPA with fitting of the O_4 scaling parameter often produces the smallest scatter, with the exception of the NO_2 retrievals. This should probably be expected as it adds another free parameter to the fit which can adapt the modelled values better to the measured ones. This does however not automatically imply better profiles.

Comparing MMF and MAPA 1.0 results, there is a tendency for MMF correlations to be slightly poorer, probably because the solution is not driven only by minimizing the RMS of the differences as in MAPA but partly also by the fixed a priori which may or may not be a good fit to the real atmospheric situation.

Overall it should be stated that the measured DSCDs are reproduced very well by all retrievals. In the case of cloudy days, there is more scatter in unflagged data, arguably because of broken clouds or horizontal inhomogeneities.

3.4 Validation of surface concentrations and columns

In addition to the profile itself, there are two additional important outputs of a profile retrieval: the tropospheric vertical column / AOD and the surface concentration / extinction. Both can be compared to independent measurements, which makes them easier to validate than the profiles and also they are important for air quality applications and satellite data validation.

For the CINDI-2 campaign, surface concentrations of NO_2 and HCHO can be compared to results from continuous LP-DOAS measurements taken during the campaign in a measurement volume close to that observed by the MAX-DOAS instruments. However, it has to be kept in mind that even the lowest point in the retrieved profiles is sensitive to absorption in higher atmospheric layers (see Figure 5), meaning that MAX-DOAS instruments might see the real surface concentration smeared into layers above.

In Figure 8, the results are shown for both the median of the data sets and for individual data points. For HCHO, the median of the retrievals fits excellently to the LP-DOAS values throughout the campaign, regardless of weather conditions. This is surprising as HCHO absorption is small and HCHO profiles are usually subject to noise. Possible explanations are less vertical variability making the a priori used more representative of the real situation or less horizontal variability when compared to NO_2. When comparing individual results then MAPA 1.0 and MMF have a tendency to be lower than the other two MAPA versions on many days. Overall the scatter of individual results is not insignificant but mostly of the order of 30%.

For NO_2, there is more variability, and at least on some days, there is a clear underestimation of the surface NO_2 in the morning by the median of the MAX-DOAS retrievals, mainly driven by MAPA results. When looking at individual values, the scatter is large, partly more than 100% and no clear pattern of which algorithm or instrument yields lower or higher values can be identified. The large variability is a surprise given the much stronger absorption signal when compared to HCHO, but the confinement of the NO_2 to a shallow layer close to the surface makes retrieval of surface concentrations from MAX-DOAS measurements difficult. As is the case for HCHO, there is no clear indication for better results under clear-sky conditions.

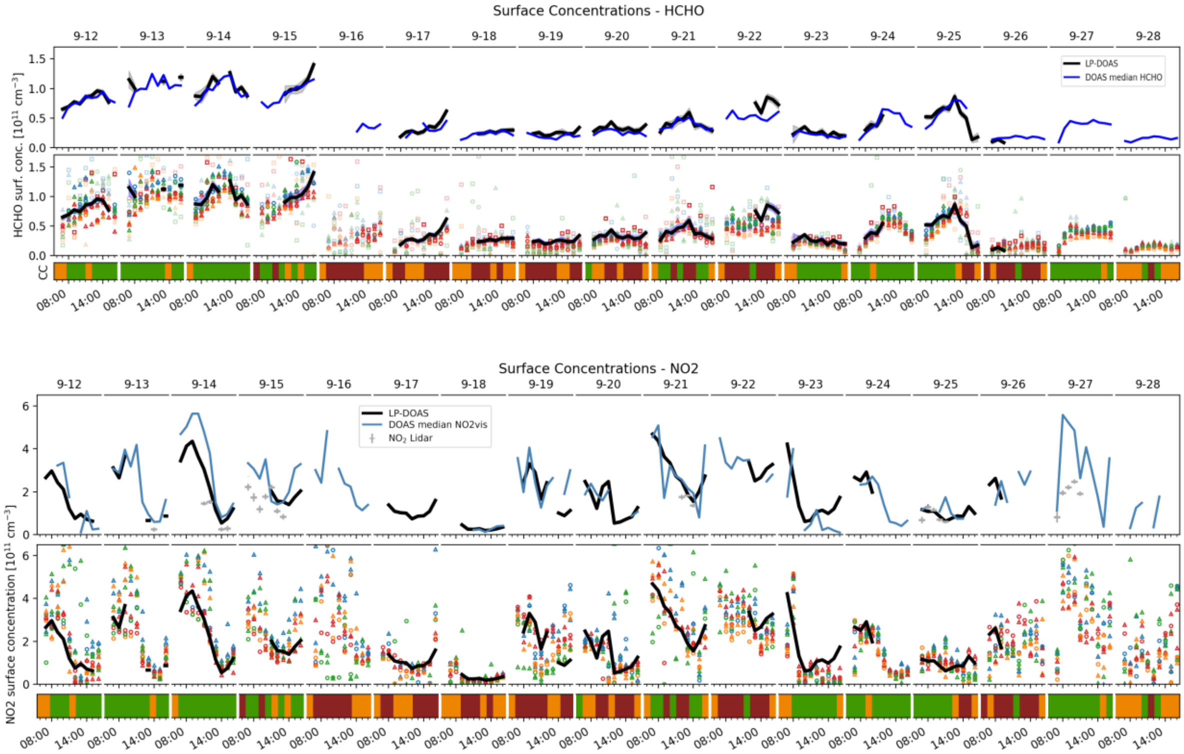

Figure 8: Comparison of retrieved surface concentrations of HCHO (top) and NO_2 (bottom) with LP-DOAS observations. The upper panel in each figure shows the comparison between LP-DOAS (black) and the median of the data (blue), the lower panel includes all individual values. Results from invalid profiles are shown in washed out colours. For NO_2, lidar data are included as grey crosses where available.

In addition to surface concentrations, also vertical tropospheric columns of NO_2 can be compared to independent measurements, in this case results from direct sun observations. The latter are insensitive to the vertical profile of NO_2 and can therefore be considered as much more accurate than vertical columns retrieved from MAX-DOAS data. The results are shown in Figure 9. Surprisingly, there is a clear overestimation of the NO_2 column by the MAX-DOAS retrievals during the first days which could be linked to the underestimation of the surface concentrations discussed above. Compared to the surface concentrations, the scatter in the vertical columns is much reduced, indicating that the column is better constrained retrieval quantity than the surface concentration. This supports the hypothesis, that deviations in the NO_2 surface concentrations are a result of shifting/smoothing of gas into higher layers.

Also shown in Figure 9 is a comparison between individual HCHO columns and the median values. This comparison is less instructive but there are unfortunately no independent HCHO columns available for comparison. As for NO_2, the scatter in the columns is much less than in the surface concentrations. During the cloudy parts of the campaign, the scatter increased but most of these retrievals are flagged as invalid.

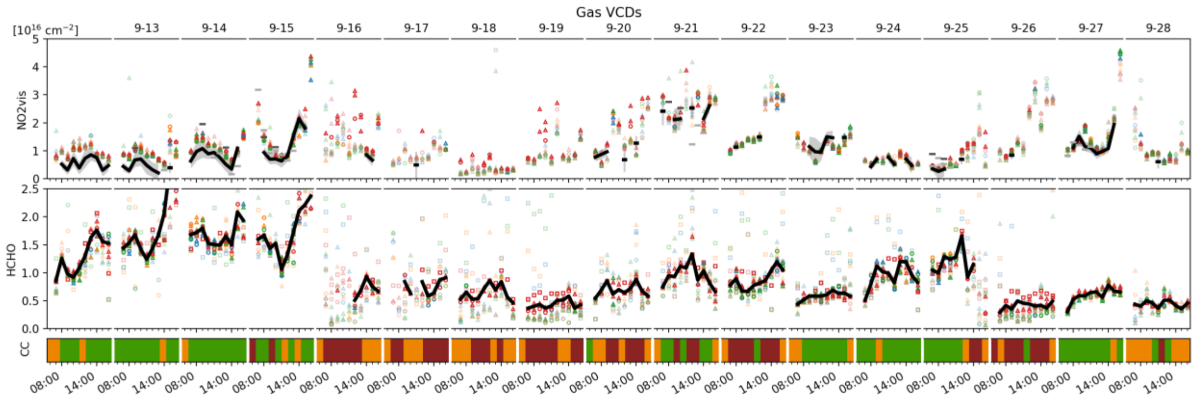

Figure 9: Comparison of all retrieved tropospheric columns of NO_2 and results from direct sun measurements (top) and comparisons of HCHO median values with all individual retrievals (bottom). The grey shaded area in the NO_2 panel indicates the uncertainty of the direct sun columns.

For the aerosol retrievals, results can be compared to AOD measurements from a CIMEL sun photometer, located at the MAX-DOAS measurement site. This comparison is shown in Figure 10 for both wavelengths and two different ways to calculate the AOD, total AOD and partial AOD. The latter accounts for the fact that MAX-DOAS measurements are sensitive only to the lowest part of the extinction profile (see Figure 5) and therefore combines the vertical extinction profile from the ceilometer with the averaging kernels of the measurements to compute that part of the sun-photometer observed AOD that is accessible to MAX-DOAS measurements. Details of this approach are given in Tirpitz et al. (2019).

As can be seen from Figure 10, MAX-DOAS AOD retrievals follow the temporal evolution of sun-photometer AOD well. In particular during the first days, agreement with the partial AOD is clearly better than with the full AOD as expected. When evaluating individual MAXDOAS retrievals, there is a clear tendency for MMF and MAPA 1.0 retrievals to underestimate sun-photometer AODs whereas the other two MAPA retrievals are closer to the validation data. This could be interpreted as support for the need of an O_4 scaling factor as for example discussed in Wagner et al. (2018).

As expected, AOD values vary widely during cloudy days, and no validation data is available as sun-photometer measurements are only possible during clear-sky periods.

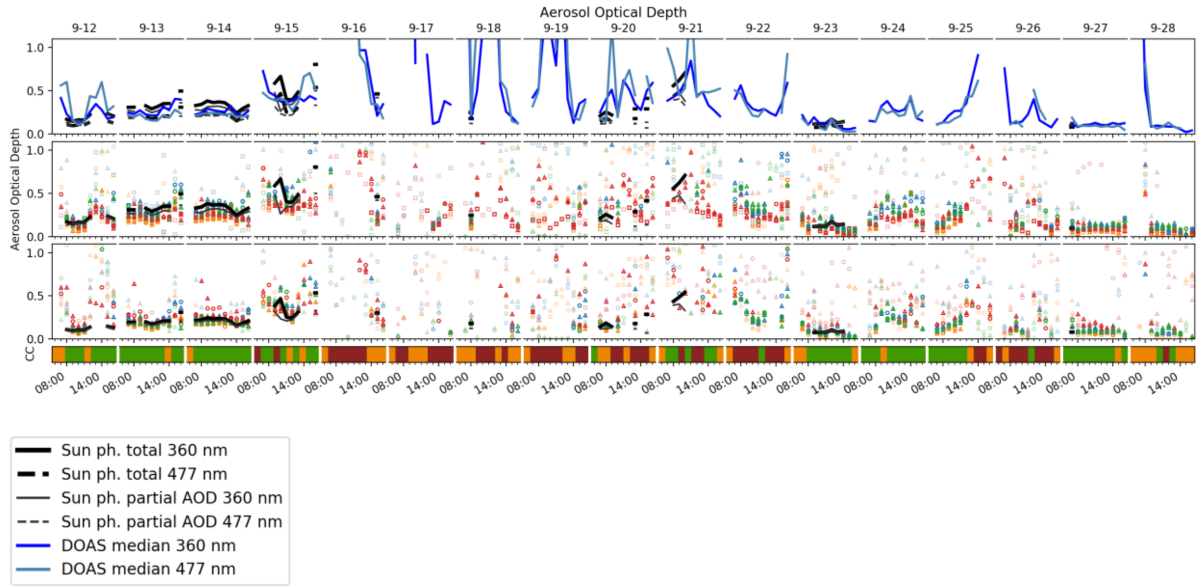

Figure 10: Comparison of MAX-DOAS and sun-photometer retrieved AOD during the campaign. The first row shows the comparison between medians and total and partial sun-photometer AOD, the second and third row show all results for the UV and the vis aerosol retrievals.

3.5 Profile statistics

In addition to the quality of the retrieved profiles, also the number of successful retrievals is an important parameter. This number depends on the robustness of the retrieval, the assumed uncertainties in the measurements but most importantly on the flagging applied. If the criteria for removing profiles are too weak, poor profiles will be included in the output which users will not appreciate. If filtering is too aggressive, only few profiles remain and users have nothing to work with.

In Table 2, a statistical evaluation is shown for the number of profiles retrieved for the different quantities, instruments and retrievals. As can be seen from the values shown, the percentage of successful retrievals varies strongly through the data set. The main observations are that

- Overall, roughly 50% of all profiles are flagged as valid.
- For the UV retrievals, MMF is much more generous than MAPA. For example, MAPA does not retrieve a single HCHO profile for the AUTH instrument while MMF reports 53% of successful retrievals. Also for the DLR instrument, much more MMF results are present than MAPA results.
- For MAPA, the number of successful retrievals is low if an O_4 scaling factor of 1.0 is used. The other two MAPA variants result in comparable numbers of profiles.

The differences are mainly linked to the way the flagging is implemented in the two different retrievals and therefore can potentially be tuned once more data is available. There is however also a basic difference in that MMF uses a priori information, which may result in reasonable profiles even if little information is in the measurements while MAPA has no such regularisation. This may also explain why MMF results were found to be of comparable quality as MAPA results in the previous sections although much less data is removed by flagging.

Table 2: Flagging statistics for all retrievals and all products. The maximum number of profiles is 170. The column *valid* reports the percentage of available profiles which is flagged as valid.

		Aer. UV		Aer. Vis		HCHO		NO_2 Vis	
		N	valid [%]	N	valid [%]	N	valid [%]	N	valid [%]
mp-08	BIR	153	52	162	49	153	48	161	41
	IUP	170	54	170	51	170	48	170	44
	AUT	146	8	-	-	145	0	-	-
	DLR	159	34	159	34	159	20	159	29
mp-10	BIR	153	28	162	25	153	21	161	22
	IUP	170	26	170	30	170	21	170	26
	AUT	146	5	-	-	145	0	-	-
	DLR	159	29	159	26	159	17	159	22
mp-free	BIR	153	50	161	45	153	45	161	39
	IUP	170	51	170	47	170	45	170	41
	AUT	145	13	-	-	145	0	-	-
	DLR	159	45	159	35	159	28	159	29
mmf	BIR	153	60	162	40	153	60	162	35
	IUP	170	84	170	50	170	84	170	41
	AUT	146	53	-	-	146	53	-	-
	DLR	143	92	143	54	143	92	143	40

It is interesting to investigate how clouds and flagging affect the comparison of retrieval results with independent data. Therefore, Table 3 and Table 4 give an overview on the RMS for slant columns, AOD, surface concentrations and VCDs for the different retrievals and instruments. Results are shown separately for clear sky and cloudy sky situations (left and right, respectively) and with (Table 4) and without (Table 3) flagging applied.

The main conclusions from these tables are that

- Results for clear sky and cloudy skies do not differ dramatically.
- That quality filtered data perform clearly better than unfiltered data.
- That the number of profiles available after filtering is much reduced.
- That there is no obvious “winner” between the algorithms based on this statistics.
- That the stronger filtering in the MAPA algorithm tends to result in smaller RMS after filtering, but at the price of lower number of profiles.

Table 3: Overview on RMS for different parameters and different subsets without flagging. Left: clear sky results, right: cloudy sky results. "Surf" is the comparison to the LP-DOAS surface concentration, "AOD" refers to the comparison to the sun photometer AOD. For the MMF algorithm, AODs were corrected with the averaging kernel as discussed in section 3.4. "Typical values" (last row) give average values observed during CINDI-2. Cell colours give a qualitative indication of problems (reddish).

		Aer. UV		Aer. Vis		HCHO		NO$_2$ Vis		
		dSCD	AOD	dSCD	AOD	dSCD	Surf	dSCD	VCD	Surf
		[10^{41} cm^{-2}]	[10^{-2}]	[10^{41} cm^{-2}]	[10^{-2}]	[10^{14} cm^{-2}]	[10^{9} cm^{-3}]	[10^{14} cm^{-2}]	[10^{14} cm^{-2}]	[10^{9} cm^{-3}]
mp-08	BIR	42	72	20	12	530	15	33	33	132
	IUP	7	10	17	23	7	12	34	36	109
	AUT	16	150	-	-	61	49	-	-	-
	DLR	11	96	25	58	18	27	38	34	133
mp-10	BIR	43	90	24	16	452	27	32	37	57
	IUP	11	95	22	13	10	18	31	39	73
	AUT	19	140	-	-	61	27	-	-	-
	DLR	11	75	23	82	19	18	38	35	108
mp-free	BIR	30	44	15	16	467	87	35	40	128
	IUP	5	38	12	17	8	22	35	33	204
	AUT	12	156	-	-	73	155	-	-	-
	DLR	7	44	15	108	19	20	63	1620	209
mmf	BIR	45	96	16	32	643	13	37	34	53
	IUP	12	25	14	28	8	14	37	38	53
	AUT	31	7	-	-	73	43	-	-	-
	DLR	13	9	17	68	21	16	98	42	82
Typical	values	240	30	340	20	270	50	890	90	180

		Aer. UV		Aer. Vis		HCHO		NO$_2$ Vis		
		dSCD	AOD	dSCD	AOD	dSCD	Surf	dSCD	VCD	Surf
		[10^{41} cm^{-2}]	[10^{-2}]	[10^{41} cm^{-2}]	[10^{-2}]	[10^{14} cm^{-2}]	[10^{9} cm^{-3}]	[10^{14} cm^{-2}]	[10^{14} cm^{-2}]	[10^{9} cm^{-3}]
mp-08	BIR	38	46	36	37	138	18	46	29	77
	IUP	17	43	35	45	10	13	43	29	88
	AUT	24	130	-	-	66	30	-	-	-
	DLR	16	205	33	78	23	20	51	28	98
mp-10	BIR	39	89	38	35	137	20	47	29	68
	IUP	18	107	36	22	10	15	43	30	80
	AUT	26	176	-	-	67	24	-	-	-
	DLR	19	169	32	138	23	17	50	30	93
mp-free	BIR	26	68	23	41	94	43	80	2958	1196
	IUP	11	46	21	62	10	40	88	3714	189
	AUT	18	138	-	-	70	104	-	-	-
	DLR	10	96	22	105	23	24	70	3381	95
mmf	BIR	42	75	31	998	157	13	49	26	61
	IUP	16	30	30	36	11	11	46	28	56
	AUT	42	16	-	-	76	26	-	-	-
	DLR	20	27	26	68	24	12	53	30	65
Typical	values	240	30	340	20	270	50	890	90	180

Table 4: As Table 3, but for quality filtered data. Red numbers indicate statistically critical values, calculated from less than 6 data points

		Aer. UV		Aer. Vis		HCHO		NO$_2$ Vis		
		dSCD	AOD	dSCD	AOD	dSCD	Surf	dSCD	VCD	Surf
		[10^{41} cm^{-2}]	[10^{-2}]	[10^{41} cm^{-2}]	[10^{-2}]	[10^{14} cm^{-2}]	[10^{9} cm^{-3}]	[10^{14} cm^{-2}]	[10^{14} cm^{-2}]	[10^{9} cm^{-3}]
mp-08	BIR	6	9	13	7	9	12	24	30	138
	IUP	6	8	12	6	8	11	27	34	99
	AUT	14	3	-	-	-	-	-	-	-
	DLR	7	6	17	7	15	24	32	29	145
mp-10	BIR	11	13	14	6	10	18	28	50	46
	IUP	11	13	14	6	10	21	26	46	71
	AUT	11	8	-	-	-	-	-	-	-
	DLR	8	10	15	5	17	20	28	28	86
mp-free	BIR	4	23	9	7	9	14	31	32	53
	IUP	4	9	8	5	8	14	30	34	70
	AUT	13	3	-	-	-	-	-	-	-
	DLR	5	8	11	4	16	15	31	31	116
mmf	BIR	61	6	7	14	871	15	23	41	50
	IUP	12	7	9	9	8	15	25	47	57
	AUT	34	7	-	-	67	30	-	-	-
	DLR	13	9	15	18	21	15	46	24	53
Typical	values	240	30	340	20	270	50	890	90	180

		Aer. UV		Aer. Vis		HCHO		NO$_2$ Vis		
		dSCD	AOD	dSCD	AOD	dSCD	Surf	dSCD	VCD	Surf
		[10^{41} cm^{-2}]	[10^{-2}]	[10^{41} cm^{-2}]	[10^{-2}]	[10^{14} cm^{-2}]	[10^{9} cm^{-3}]	[10^{14} cm^{-2}]	[10^{14} cm^{-2}]	[10^{9} cm^{-3}]
mp-08	BIR	8	4	16	6	8	8	30	15	115
	IUP	9	24	15	13	7	9	38	10	80
	AUT	15	-	-	-	-	-	-	-	-
	DLR	10	19	14	14	14	17	35	5	74
mp-10	BIR	12	14	14	6	8	14	30	19	111
	IUP	12	13	12	7	9	11	33	14	92
	AUT	14	15	-	-	-	-	-	-	-
	DLR	12	11	15	3	13	0	35	26	72
mp-free	BIR	6	7	12	4	7	15	34	16	115
	IUP	5	12	10	3	8	18	32	12	110
	AUT	14	10	-	-	-	-	-	-	-
	DLR	7	13	13	6	15	10	25	9	69
mmf	BIR	26	24	13	17	111	14	29	27	55
	IUP	14	19	17	37	10	11	33	26	53
	AUT	36	13	-	-	70	24	-	-	-
	DLR	17	27	21	72	23	13	34	23	66
Typical	values	240	30	340	20	270	50	890	90	180

3.6 Comparison with independent retrievals

One of the rationales of using a centralised processing system as demonstrated in FRM4DOAS is the hope that applying such a unified processing should result in more consistent results when applied to different instruments than if each research group uses their own retrieval. In order to check if this is indeed the case, results from the CINDI-2 profile intercomparison (Tirpitz et al., 2018) are confronted with the results obtained here, comparing the same instruments and quantities but in the case of the Tirpitz et al. (2018) study using the retrievals of the individual groups. In all cases, flagged data are used.

Overall, the RMS of the results from the individual retrievals is always of the order of the RMS of MMF or larger, and in the case of the extinction profile at 477 nm, it is much larger. MAPA results tend to have lower variability for the UV results (see also Table 4) but with much less valid profiles. While this improvement in RMS (and thus consistency) is expected, the results depend strongly on the quality flagging applied. As in the Tirpitz et al. (2018) study most groups did not provide any flagging, the larger scatter may at least in part result from the inclusion of poor profiles in the results, and do not necessarily imply that the results from the FRM4DOAS processor are more consistent than the individual retrievals. For this, a comparison with flagging would be needed and this information is currently not available.

4 Validation using IUP-UB retrievals

In addition to the comparison of CINDI-2 results, also one month of FRM4DOAS retrievals for the IUP-UB MAX-DOAS stations in Bremen, Ny-Alesund and Athens was evaluated. These data were produced in the automated NRT mode, where spectra files were transferred from the instruments at night and automatically ingested and processed in the FRM4DOAS demonstration processor. Data for August 2018 were selected as this month had few data gaps in the UB data and had many days with favourable observing conditions.

4.1 Comparison of tropospheric vertical columns with AMF approach

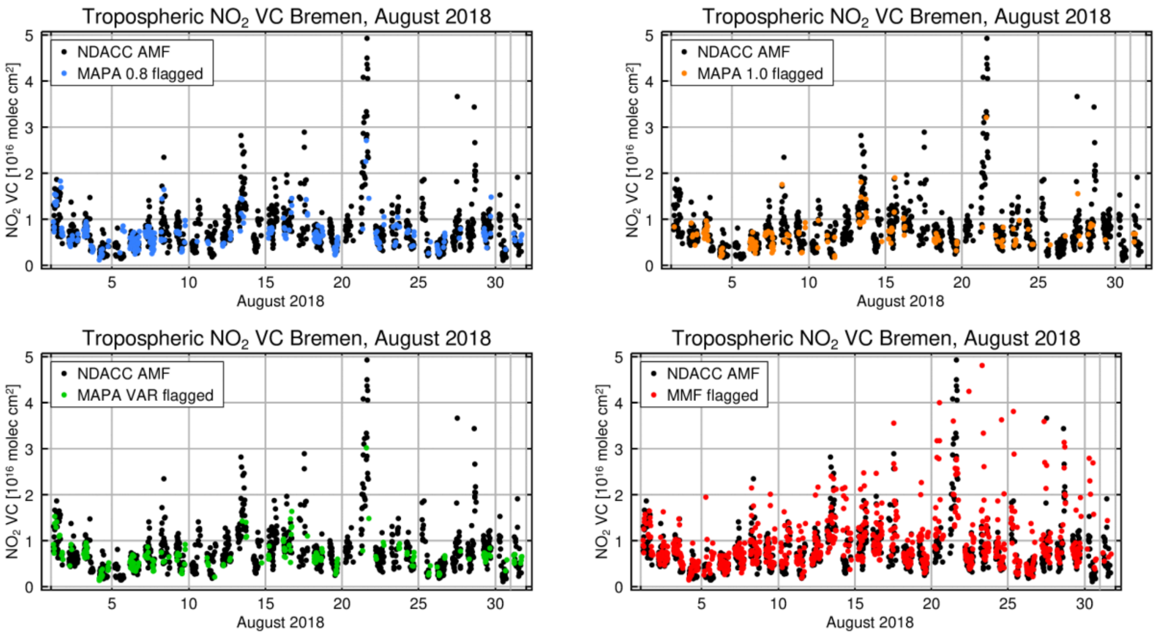

Figure 11: Time evolution of the tropospheric NO_2 vertical columns in Bremen for August 2018. Black dots are the NDACC evaluation, coloured symbols the results from the different FRM4DOAS retrievals. Some extreme outliers of the MMF retrieval are off scale and not visible.

In a first comparison, the time evolution of the FRM4DOAS retrieved vertical NO_2 columns is compared to that of the tropospheric columns submitted to the NDACC data base. The latter are based on a simple AMF approach applied to the 30° elevation measurements without profile retrieval. As climatological assumptions are applied to the vertical profile as well as the aerosol load, one would expect that the results are less accurate than the results from the profile inversions. On the other hand, the simple AMF retrieval is more robust and does not suffer from the instabilities sometimes found in profile retrievals.

As can be seen in Figure 11, the temporal evolution of the NDACC vertical columns is well reproduced by the integrated columns from the profile retrievals. More specifically

- MAPA results seem very close to NDACC columns but only a limited number of retrievals is flagged as valid.
- MMF has much more valid retrievals, but also shows some outliers, which do not match well the AMF determined vertical columns. There also are some extreme outliers which are off scale in the figure.
- All four retrievals do not reproduce the episodes of very high tropospheric columns found in the NDACC data on several days.

Another view on the same data is given in Figure 12, where the very good agreement of all three MAPA retrievals with the NDACC data is highlighted as well as the much larger number of valid retrievals and the larger scatter for MMF.

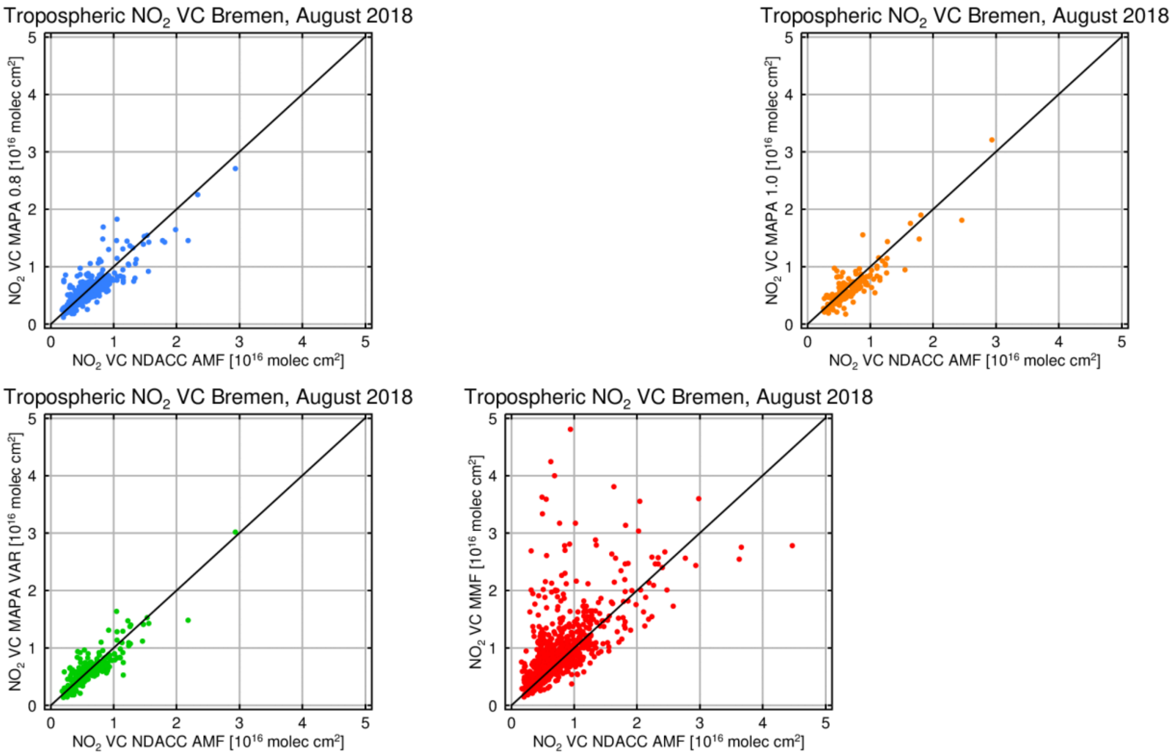

Figure 12: Scatter plots of tropospheric NO_2 vertical columns in Bremen from the NDACC evaluation (x-Axis) and the four FRM4DOAS retrievals (y-axis of different plots). All valid data from August 2018 is shown with the exception of a few extreme outliers in the MMF results which are off scale.

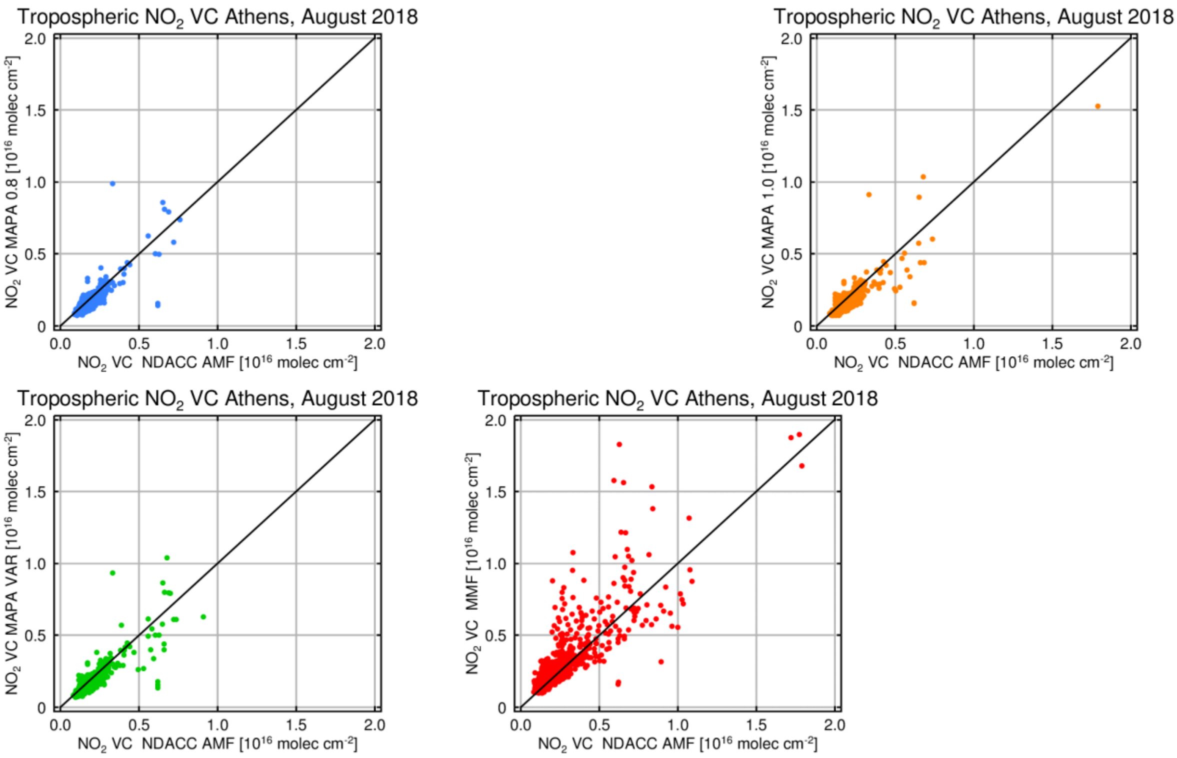

Figure 13: As Figure 12 but for Athens

Very similar results are found for measurements in Athens as shown in Figure 13. Again, MMF yields many more valid retrievals, but shows more scatter. Here, there also are some outliers in the MAPA data.

As already pointed out above, the NDACC retrievals use a simplified AMF and should not be considered as the “truth” for the vertical column. On the other hand, the very good agreement for most of the MAPA and a majority of the MMF results indicates that there is no systematic problem with any of the three approaches. The reduced number of valid retrievals and the missing profiling results for the high pollution situations could be linked to broken clouds, high aerosol loads or rapidly changing atmospheric situations which all impact the profiling retrievals more than the simple 30° retrievals.

As already discussed in section 3.5, the quality flagging in MAPA is much more restrictive and leads to a small number of valid retrievals with high quality, while MMF results in many more retrievals but with increased scatter. If a stricter screening would be applied to MMF data, the agreement with the NDACC data would probably be as good as for MAPA.

4.2 Comparison of NO_2 tropospheric vertical columns with BOREAS

In addition to the comparison with the standard NDACC retrievals, the FRM4DOAS results have also been compared to retrievals from the UB BOREAS algorithm (Bösch et al., 2019). The BOREAS retrieval is an Optimal Estimation based profiling retrieval which for the aerosol retrieval applies additional Tikhonov regularisation. It also differs from other retrievals in that it applies pre-scaling of the a priori profiles using a quick pre-analysis and can apply dynamic weights to the regularisation but the latter option is not used here. BOREAS is still a rather new retrieval and has only been tested in detail on synthetic and CINDI-2 data (see Frieß et al., 2019 and Tirpitz et al., 2019) as well as on Bremen data (Bösch, 2019).

As the first step, tropospheric NO_2 columns from all four FRM4DOAS retrievals are compared to results from BOREAS for data from the three UB stations in August 2018. The three stations differ in their characteristics with

- Bremen being a moderately polluted mid-latitude site.
- Athens being a polluted site with a complex topography and an instrument position on a hill close to the city at 350m altitude.
- Ny-Alesund being an Arctic clean air site with some local NO_2 pollution from cruise ships and a power generator which occasionally is blown into the viewing direction of the MAX-DOAS instrument. During summer, there is polar day with round the clock observations.

In Figure 14, the time evolution of the tropospheric NO_2 columns is shown for the three stations. Different scales are used for better representation of the variability. The main findings are that

- In Bremen, the overall evolution is similar between the retrievals, but MMF has a tendency of high outliers
- In Athens, there is very good agreement in the time evolution, but BOREAS results appear to be higher than the FRM4DOAS retrievals
- In Ny-Alesund, there is more scatter, and MMF has many unexpectedly high results. For the three MAPA algorithms, very few retrievals are flagged as successful.

How different the five algorithms treat flagging is summarised in Table 5 – MMF is very generous, while MAPA is very strict. BOREAS is somewhere in between. Close inspection of the data shows, that not only the number of retrievals flagged as successful varies, but also the times for which profiles are deemed valid. As a result, a direct comparison of NO_2 vertical columns between the retrievals leads to only a relatively small number of coincidences even for a full month of data.

Table 5: Number of successful NO_2 retrievals for August 2018 data from the three UB stations

Station	MMF	MAPA 1.0	MAPA 0.8	MAPA VAR	BOREAS
Bremen	803	155	201	163	311
Athens	1272	349	306	480	1217
Ny-Alesund	1694	31	68	67	906

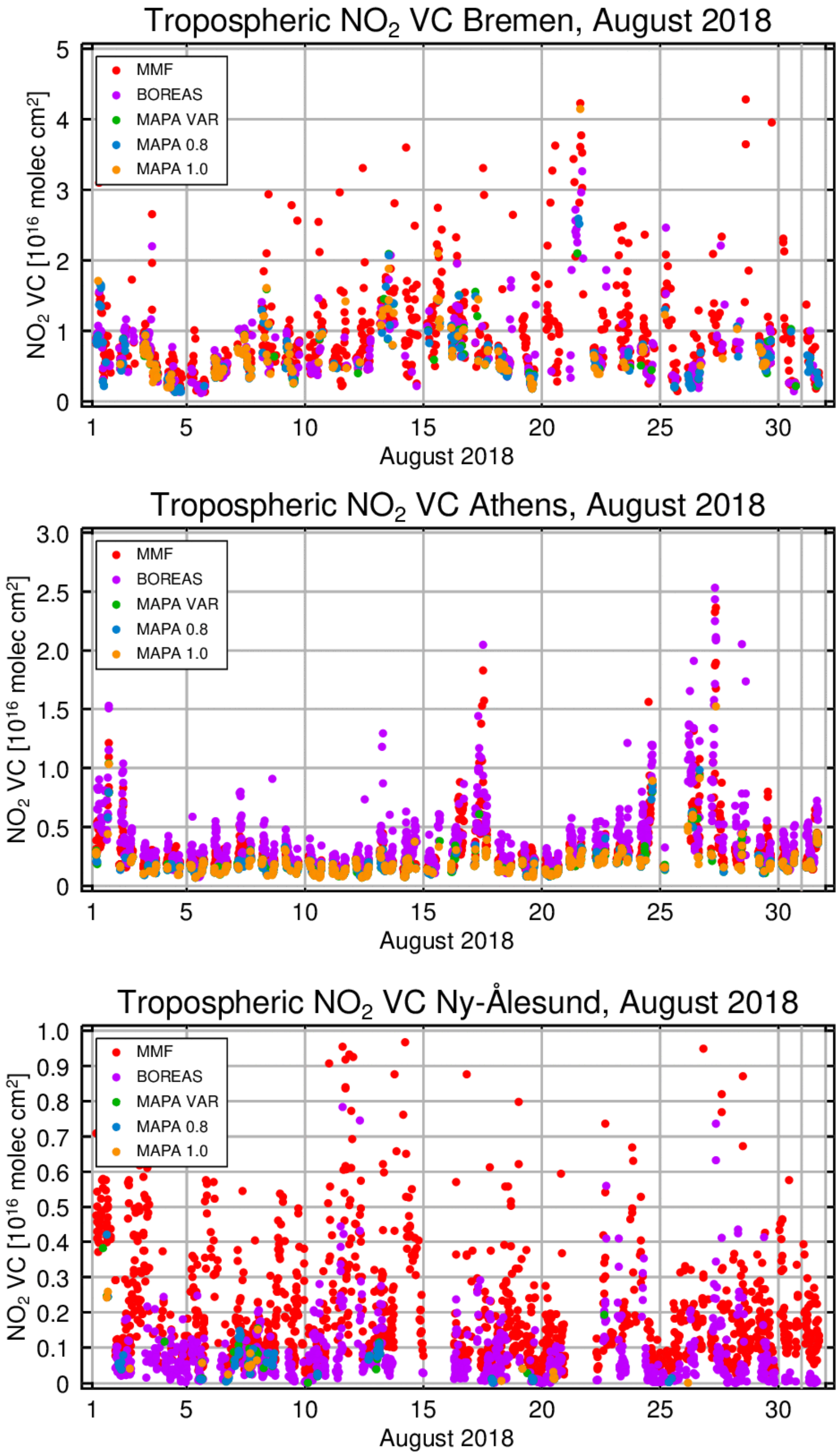

Figure 14: Comparison of tropospheric NO_2 vertical columns for the three UB stations in August 2018. All flagged data is shown with the exception of some MMF outliers, which are off scale. Please note that for Bremen, another viewing direction is shown than in Figure 11.

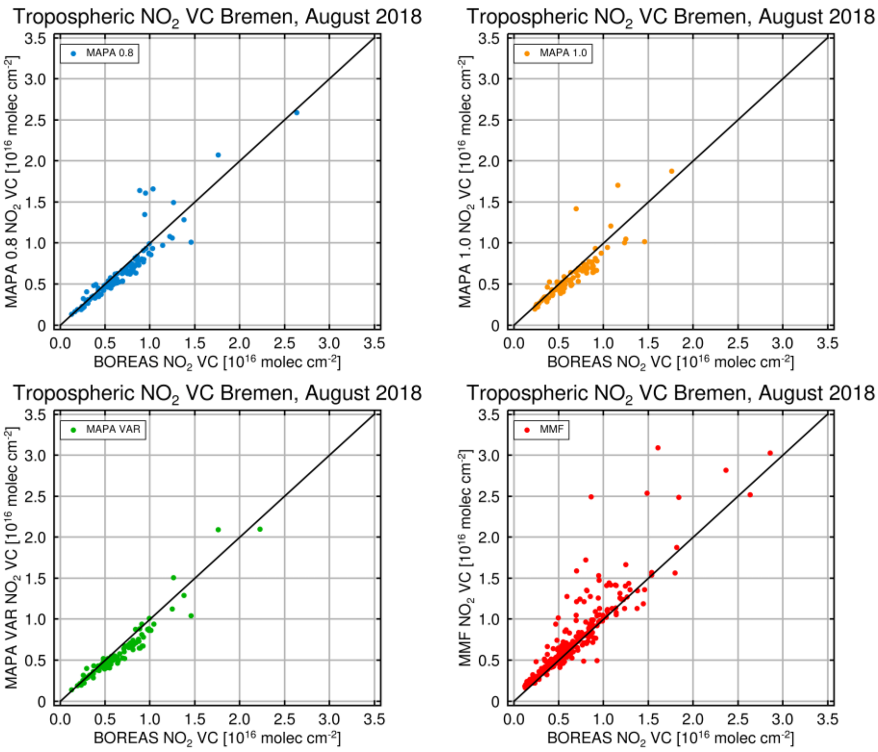

Figure 15: Correlation plots between tropospheric NO_2 columns for the four FRM4DOAS retrievals and the UB BOREAS results. All flagged data from August 2018 is used. Results of the linear regression are summarised in Table 6.

Comparisons of matching tropospheric NO_2 VC for all three stations are shown as scatter plots in Figure 15, Figure 16, and Figure 17, always with the BOREAS retrievals as x-axis. The results from a linear regression on these data are given in Table 6. It should be kept in mind that BOREAS is not the truth but just an independent retrieval on the same spectra.

The main conclusions from the comparison are

- For Bremen, the agreement of the tropospheric NO_2 columns is good, with MAPA yielding lower and MF higher columns than BOREAS. The scatter of MMF columns is larger, but also more values are provided.
- For Athens, the correlation is good for all three FRM4DOAS retrievals, but BOREAS columns appear to be systematically higher. This can be explained by the approach taken in BOREAS for the high altitude station: It is modelled as a flying instrument. This choice was made as clearly much of Athens' pollution is below the measurement altitude, and from the viewing point of the instrument, photons from below the station altitude will also contribute to the scatter light observed. Therefore, BOREAS profiles extend below the instrument's altitude, resulting in larger columns.
- In Ny-Alesund, very few MAPA retrievals are flagged as successful, but those agree very well with BOREAS: MMF on the other side is systematically much larger than BOREAS and MAPA and shows much scatter. This indicates that MMF still has problems for the Arctic clean air site Ny-Alesund. Sensitivity tests have shown that this feature is for a large part related to the use of a scale height for the a priori of 1.4km instead of the wanted 1.0 km. This

contributes to a "ghost" column at altitudes where there is no sensitivity of around 0.5 x 10^{15} molec/cm^2 for NO_2 (in the visible).

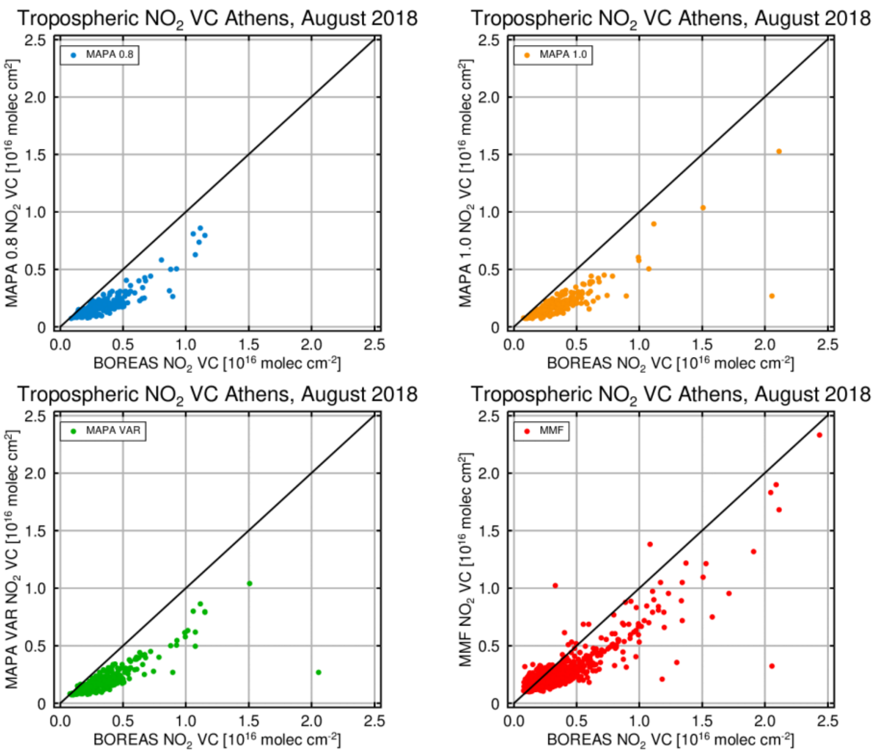

Figure 16: As Figure 15 but for Athens

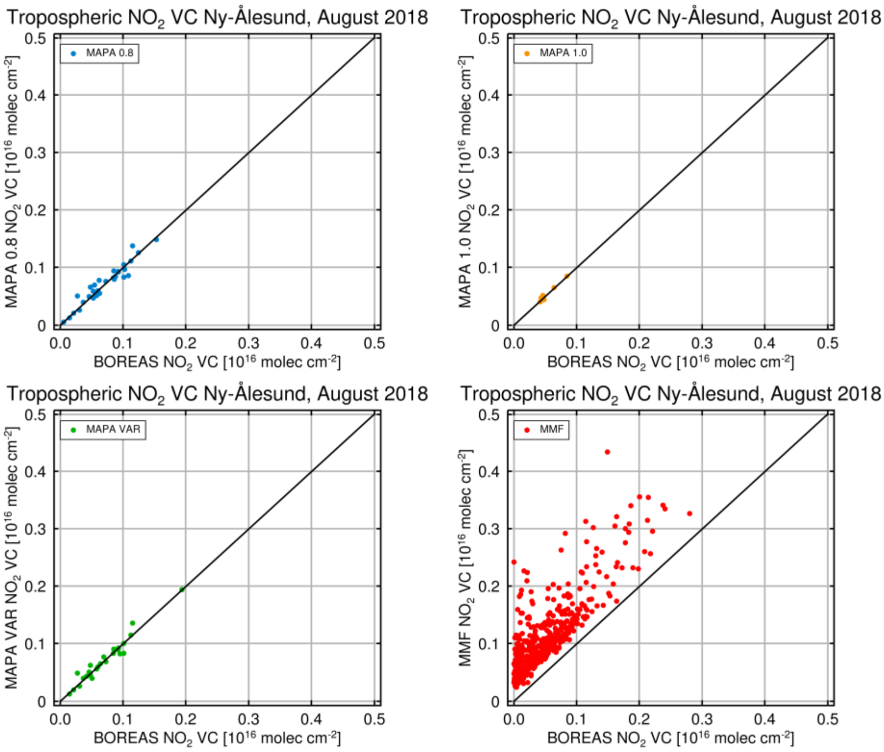

Figure 17: As Figure 15 but for Ny-Alesund

Table 6: Linear regression results of NO_2 retrievals for August 2018 relative to BOREAS results. The offset is given in units of molec cm^{-2}.

Station	MMF	MAPA 1.0	MAPA 0.8	MAPA VAR
Bremen correlation	0.88	0.87	0.91	0.96
slope	1.03	0.92	1.01	0.94
offset	5.1 x 10^{14}	-1.5 x 10^{14}	-4.8 x 10^{14}	-3.3 x 10^{14}
Athens correlation	0.89	0.84	0.89	0.86
slope	0.63	0.48	0.55	0.48
offset	3.7 x 10^{14}	2.3 x 10^{14}	1.1 x 10^{13}	2.6 x 10^{14}
Ny-Alesund correlation	0.84	0.98	0.95	0.96
slope	1.50	0.98	1.01	1.09
offset	3.9 x 10^{14}	1.8 x 10^{13}	-1.1 x 10^{13}	-8.5 x 10^{13}

4.3 Comparison of AOD with BOREAS

A similar comparison as for the NO_2 vertical columns can also be performed for the AOD from the five algorithms. For Bremen and Ny-Alesund, AOD in the FRM4DOAS algorithms is retrieved in the visible at 494nm; for Athens the spectral coverage is not sufficient and it is taken at 354nm.

In Figure 18, the time evolution of the retrieved AOD is shown for the FRM4DOAS retrievals and BOREAS for the three UB stations. The y-axis is limited to values up to 2 although much larger values are also found in the FRM4DOAS retrievals. BOREAS on the other hand flags all AODs larger than 1 as cloudy, removing many points. In general, it is not clear at which AOD one would speak of a cloud (or fog) rather than aerosols and haze, and whether or not cloudy scenes should be flagged. If a strict definition of aerosols is taken, only rather small AOD values are expected for Bremen, somewhat larger values for Athens and very low values for Ny-Alesund.

From Figure 18 it is clear, that AOD values vary strongly, with MMF showing the largest spread and including very high values while BOREAS cuts off at AOD of 1.0 and MAPA is mostly but not always limited to small values. In Bremen, little correlation between the different retrievals can be seen by eye within the scatter, while in Athens, the overall temporal evolution is picked up by all retrievals in a similar way. In Ny-Alesund, MMF values appear to be much larger than those of the other retrievals. The temporal evolution of MMF AOD is linked to clouds, and if also unflagged BOREAS data are included and the y-axis is extended, the similarity in the results from the two retrievals is more apparent (Figure 19).

Correlation plots for flagged AOD data are shown in Figure 20, Figure 21, and Figure 22. Compared to the NO_2 results, correlations are smaller and more variable also between the three MAPA retrievals. The main conclusions (see also Table 7) are that

- Correlation of MMF AOD data with BOREAS is low
- Correlation of MAPA AOD with BOREAS is higher but slopes vary showing both over- and underestimation

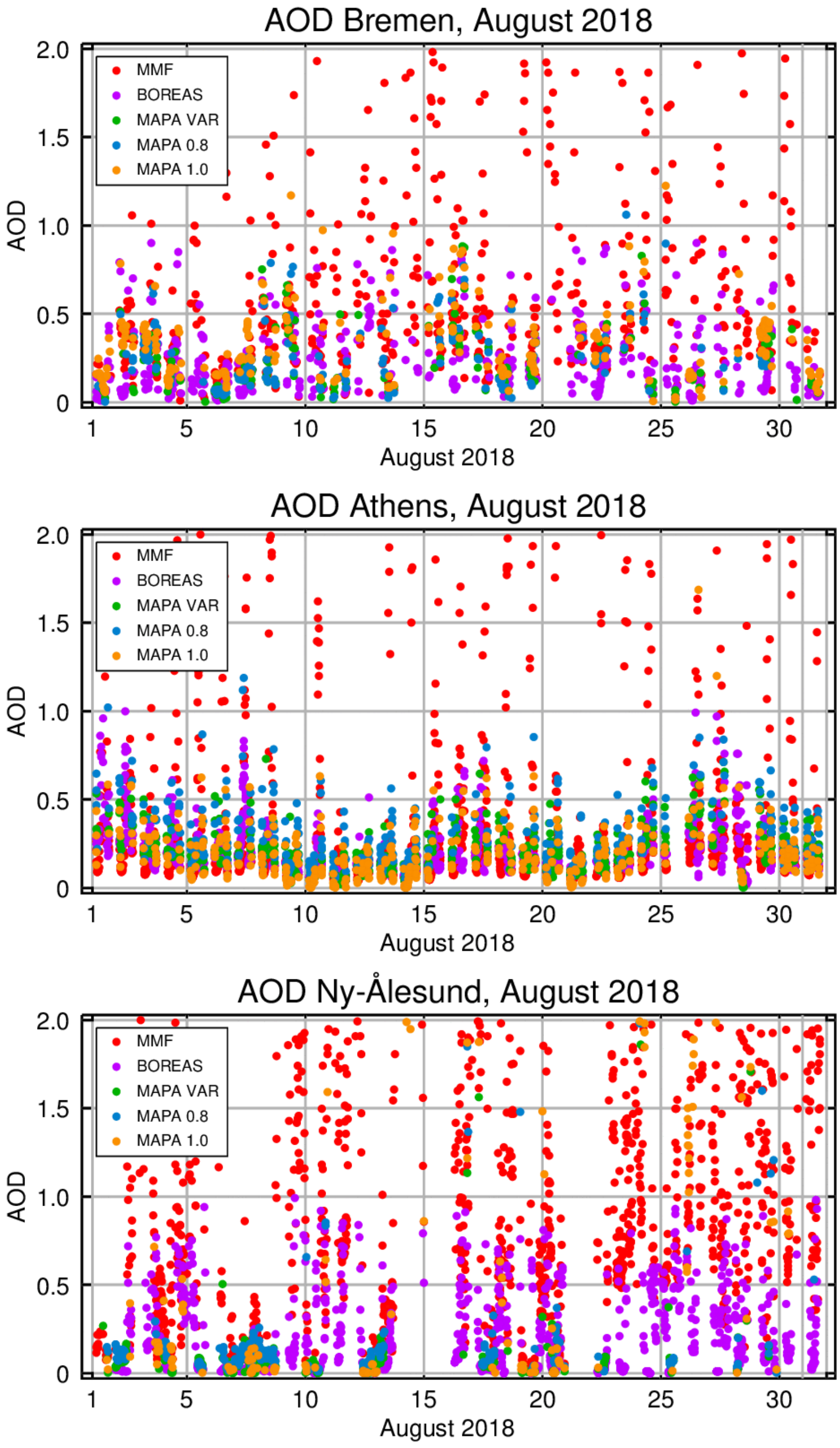

Figure 18: Comparison of AOD for the three UB stations in August 2018. Flagged data is shown and the y-axis is limited to an AOD of two although MMF reports also much larger values.

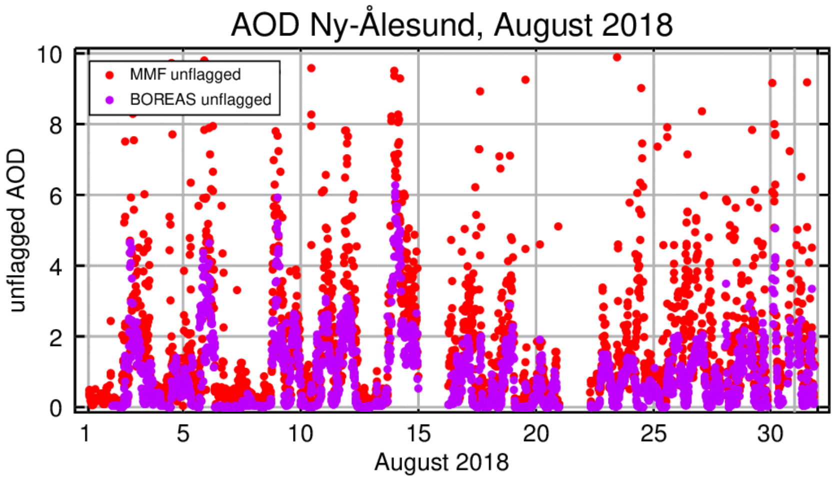

Figure 19: Unflagged AOD data from MMF and BOREAS for Ny-Alesund, August 2018. The y-axis is limited to an AOD of 10 although unflagged MMF data contain even larger values.

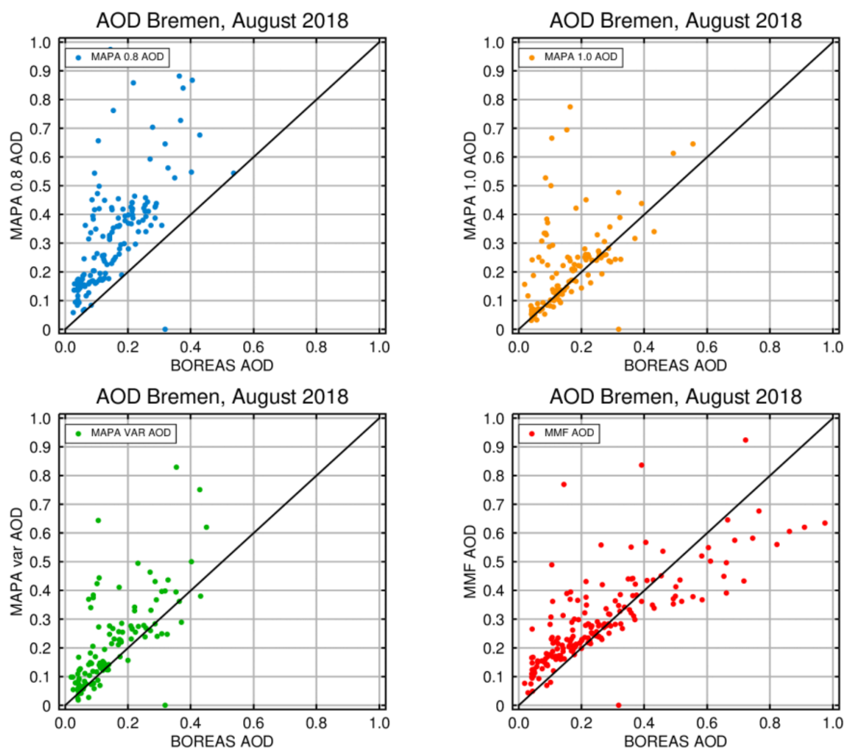

Figure 20: Correlation plots between AOD retrieved by the four FRM4DOAS retrievals and the UB BOREAS results. All flagged data from August 2018 is used. Results of the linear regression are summarised in Table 7.

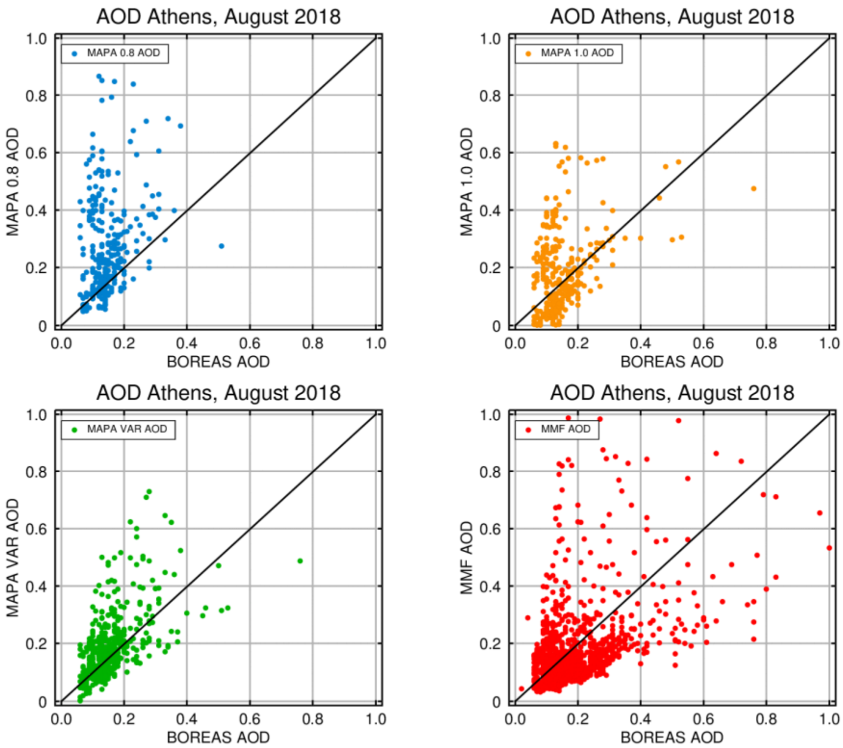

Figure 21: As Figure 20 but for Athens

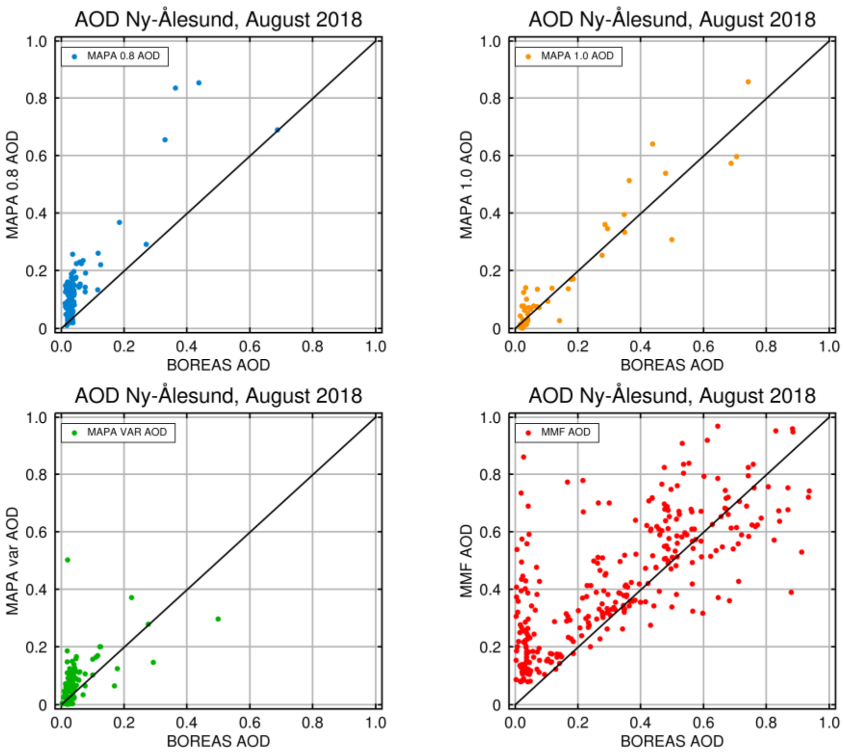

Figure 22: As Figure 20 but for Ny-Alesund

Table 7: Linear regression results of AOD retrievals for August 2018 relative to BOREAS results.

Station	MMF	MAPA 1.0	MAPA 0.8	MAPA VAR
Bremen correlation	0.34	0.54	0.71	0.72
slope	0.91	0.78	1.31	1.05
offset	0.12	0.09	0.12	0.05
Athens correlation	0.11	0.48	0.52	0.57
slope	0.74	0.84	1.36	0.83
offset	0.34	0.06	0.07	0.06
Ny-Alesund correlation	0.50	0.94	0.94	0.55
slope	0.65	0.99	1.55	0.65
offset	0.25	0.01	0.05	0.04

4.4 Comparison of surface NO_2 and extinction

As for vertical columns and AOD, also the surface NO_2 concentrations and surface extinctions retrieved by the FRM4DOAS system can be compared to the results from BOREAS.

The results for NO_2 in Bremen are shown in Figure 23 and Table 8. As can be seen, a good overall correlation is found but with significant scatter and slopes between 0.69 and 0.91, indicating systematically higher BOREAS retrievals in particular at high NO_2 concentrations. Nevertheless, the agreement for surface NO_2 between all 5 algorithms can be considered to be good in Bremen.

As another view of the results, comparisons between MAPA and MMF are plotted in Figure 24 for all three stations and summarised in Table 9. For NO_2 surface concentrations, good correlations are found in Bremen and Athens and reasonable correlation in Ny-Alesund. In general, the MAPA retrievals with variable O_4 factor show a larger scatter, but it has to be kept in mind that this retrieval also results in more values, some of which have larger uncertainties. Slopes vary between stations and between the different MAPA versions, but overall the agreement is satisfactory with systematic differences smaller than 20%. Interestingly, there is no clear pattern of which of the three MAPA retrievals shows the best agreement with MMF – this indicates that probably the sample size is too small and results are influenced by sampling rather than systematic effects. Comparisons for Bremen are tighter than the comparisons using BOREAS as reference, but this is at least in parts explained by the larger number of coinciding data points available in the comparison in Figure 23.

Surface extinctions are also well correlated between MMF and MAPA, again with larger scatter for the results with fitted O_4 factor. In contrast to NO_2 surface concentrations, there is a clear high bias for MAPA (or a low bias for MMF). Comparisons for Ny-Alesund are of limited values as there are only few valid points in the MAPA dataset and also outliers not shown in the figures dominate the results.

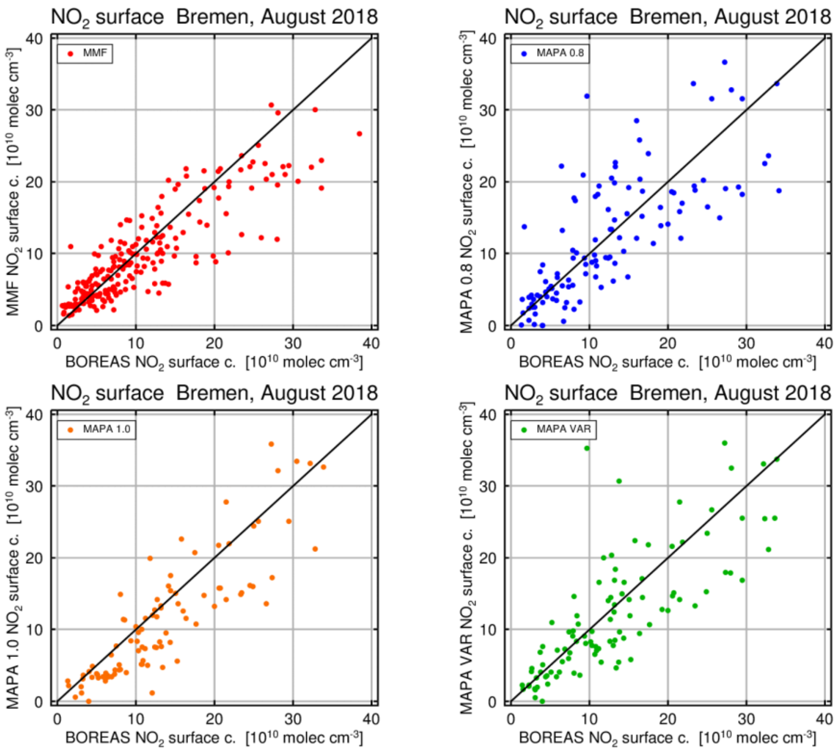

Figure 23: Comparison of MAPA and MMF surface NO_2 concentrations for Bremen, August 2018 relative to BOREAS

Table 8: Linear regression results of MMF and MAPA surface NO_2 retrievals for August 2018 in Bremen relative to BOREAS. Offsets are given in units of molec cm^{-3}.

Station	MMF	MAPA 1.0	MAPA 0.8	MAPA VAR
Bremen correlation	0.87	0.87	0.76	0.78
slope	0.69	0.91	0.79	0.80
offset	2.17×10^{10}	-1.05×10^{10}	2.52×10^{10}	1.55×10^{10}

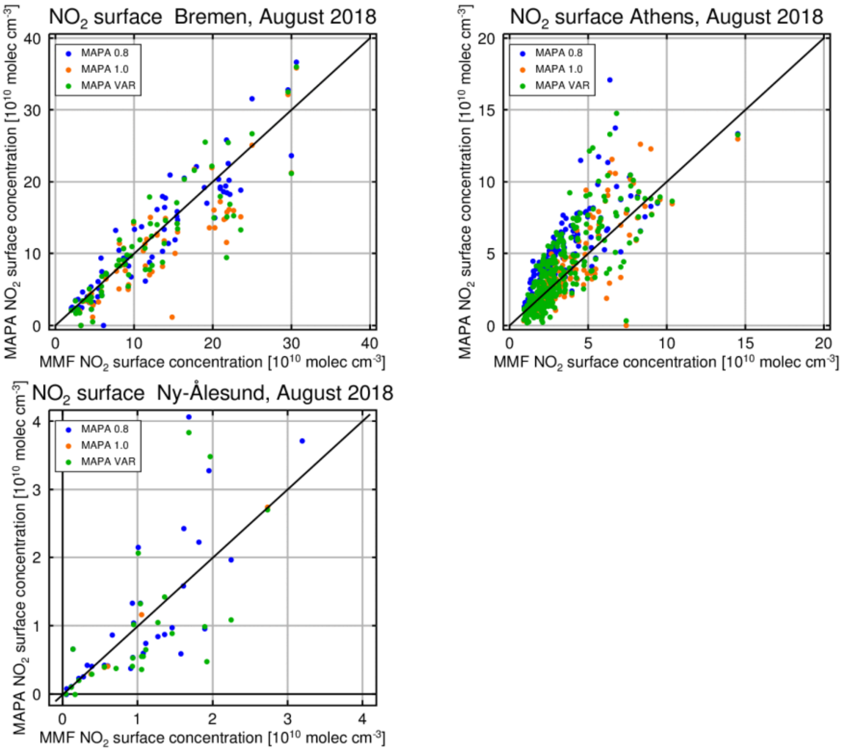

Figure 24: Comparison of MAPA and MMF surface NO_2 concentrations for August 2018.

Table 9: Linear regression results of MAPA surface NO_2 retrievals for August 2018 relative to MMF. Offset is given in molec cm^{-3}.

Station	MAPA 1.0	MAPA 0.8	MAPA VAR
Bremen correlation	0.88	0.93	0.89
slope	0.85	1.00	0.91
offset	-2.6 x10^9	-2.7x10^9	2.0 x10^9
Athens correlation	0.91	0.82	0.80
slope	1.07	1.22	1.13
offset	-3 x10^9	12 x10^9	0.7 x10^9
Ny-Alesund correlation	0.99	0.77	0.65
slope	1.06	1.11	0.92
offset	-1.3 x10^9	-0.5 x10^9	-0.2 x10^9

The comparison of MMF and MAPA results to BOREAS for surface extinction in Bremen are shown in Figure 25 and Table 10. As can be seen, a good overall correlation is found but slopes vary between 0.8 and 1.6. This is mainly the case because of additional points being included in the comparisons with MAPA 0.8 and MAPA 1.0 because of different flagging. Nevertheless, the agreement for surface extinction between all 5 algorithms can be considered to be reasonable in Bremen, with less scatter than for NO_2 but more variation in the slope.

As another view of the results, comparisons between MAPA and MMF are plotted in Figure 26 for all three stations and summarised in Table 11. Surface extinctions are well correlated between MMF and MAPA, with larger scatter for the results with fitted O_4 factor. In contrast to NO_2 surface concentrations, there is a clear high bias for MAPA (or a low bias for MMF). Comparisons for Ny-Alesund are of limited values as there are only few valid points in the MAPA dataset and also outliers not shown in the figures dominate the results.

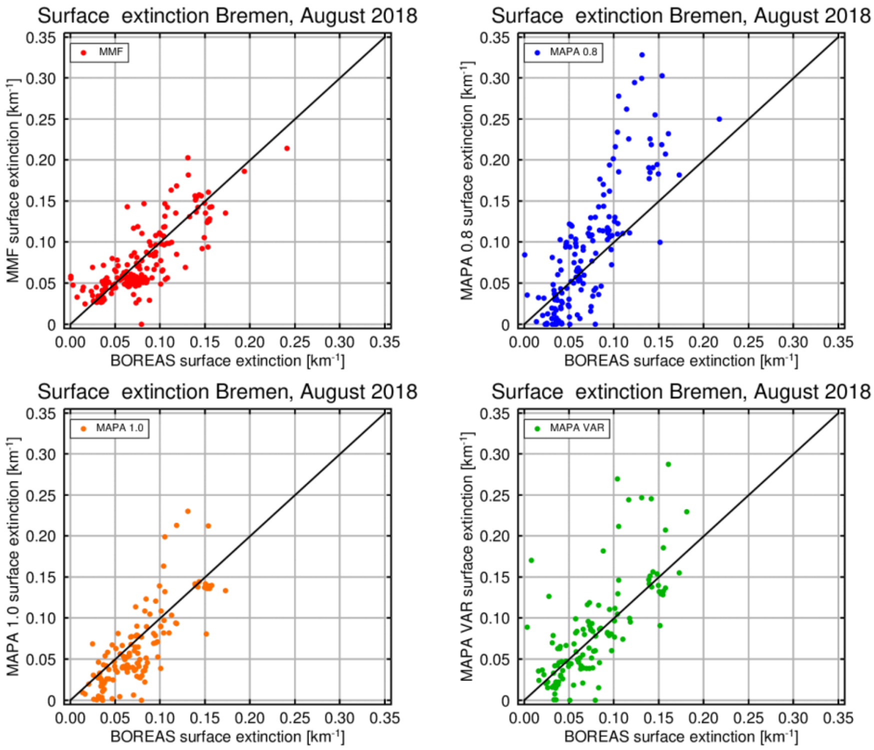

Figure 25: Comparison of MAPA and MMF surface extinctions for August 2018 in Bremen relative to BOREAS results

Table 10: Linear regression results of MMF and MAPA surface extinction retrievals for August 2018 in Bremen relative to BOREAS

Station	MMF	MAPA 1.0	MAPA 0.8	MAPA VAR
Bremen correlation	0.80	0.80	0.82	0.65
slope	0.82	1.07	1.63	1.24
offset	0.01	-0.02	-0.03	-0.01

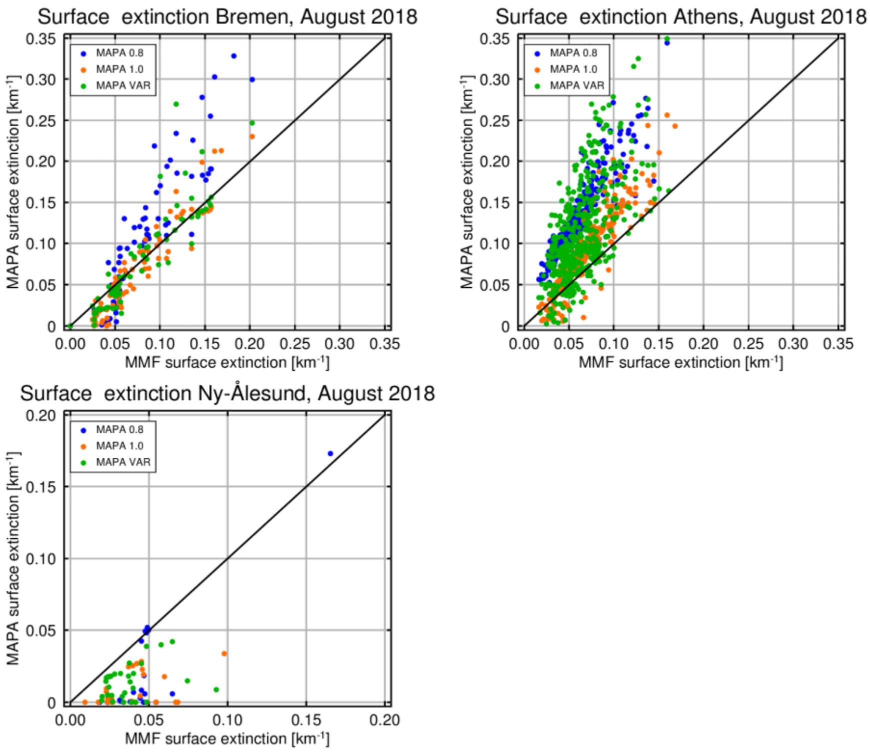

Figure 26: Comparison of MMF and MAPA surface extinction for August 2018.

Table 11: Linear regression results of MAPA surface extinction retrievals for August 2018 relative to MMF

Station	MAPA 1.0	MAPA 0.8	MAPA VAR
Bremen correlation	0.95	0.91	0.90
slope	1.2	1.76	1.22
offset	-0.02	-0.03	-0.02
Athens correlation	0.91	0.87	0.63
slope	1.47	1.66	1.48
offset	-0.01	0.03	0.02
Ny-Alesund correlation	0.43	0.87	0.31
slope	0.26	1.19	0.22
offset	0	-0.03	0

4.5 Evaluation of uncertainties

In addition to the values of vertical columns, surface concentrations, AOD and surface extinction, also the uncertainties reported by the algorithms are of interest as they inform the users about the expected validity of the results. Ideally, the reported uncertainties should be compared to the spread of the results relative to a (more accurate and precise) reference measurement. However, no such data are available for this evaluation. A comparison to results from BOREAS is unfortunately also not possible as the version of BOREAS used in this study does not yet report uncertainties on the columnar products. Therefore, a simple plausibility analysis is provided by comparing the reported error bars to the spread of values between the four different processors (counting the three MAPA versions independently). As an exemplary case, results for Bremen are presented in Figure 27.

As can immediately be seen from the figures, reported uncertainties are small for all algorithms and quantities. In particular MMF reports very low uncertainties. Zooming in on one arbitrarily selected day (August 16, 2018) shows that while with the exception of surface extinction, MAPA values from the three versions usually agree within their uncertainties, there hardly ever is overlap with MMF values within the combined uncertainties. Although this is certainly not a rigorous evaluation of the reported uncertainties, it is clear indication that uncertainties are underestimated by MMF, and in all probability also by MAPA.

Two possible explanations can be given for this clear underestimation of uncertainties:

1. The uncertainties reported by the DOAS fit could be underestimated, both because of neglecting systematic components of the error budget and because natural atmospheric variability is not considered while it can have a large impact on the results.
2. The uncertainties assumed in the profile retrieval are underestimated as the effects of variations in light path (horizontal inhomogeneities in absorber, aerosols and clouds) as well as atmospheric changes during the time needed for a complete viewing angle scan are not considered.

Comparison between results from the four algorithms run on the same data can of course only provide an estimate of the lower limit for the uncertainty as additional error sources such as clouds may have similar effects on the retrievals and therefore remain undetected.

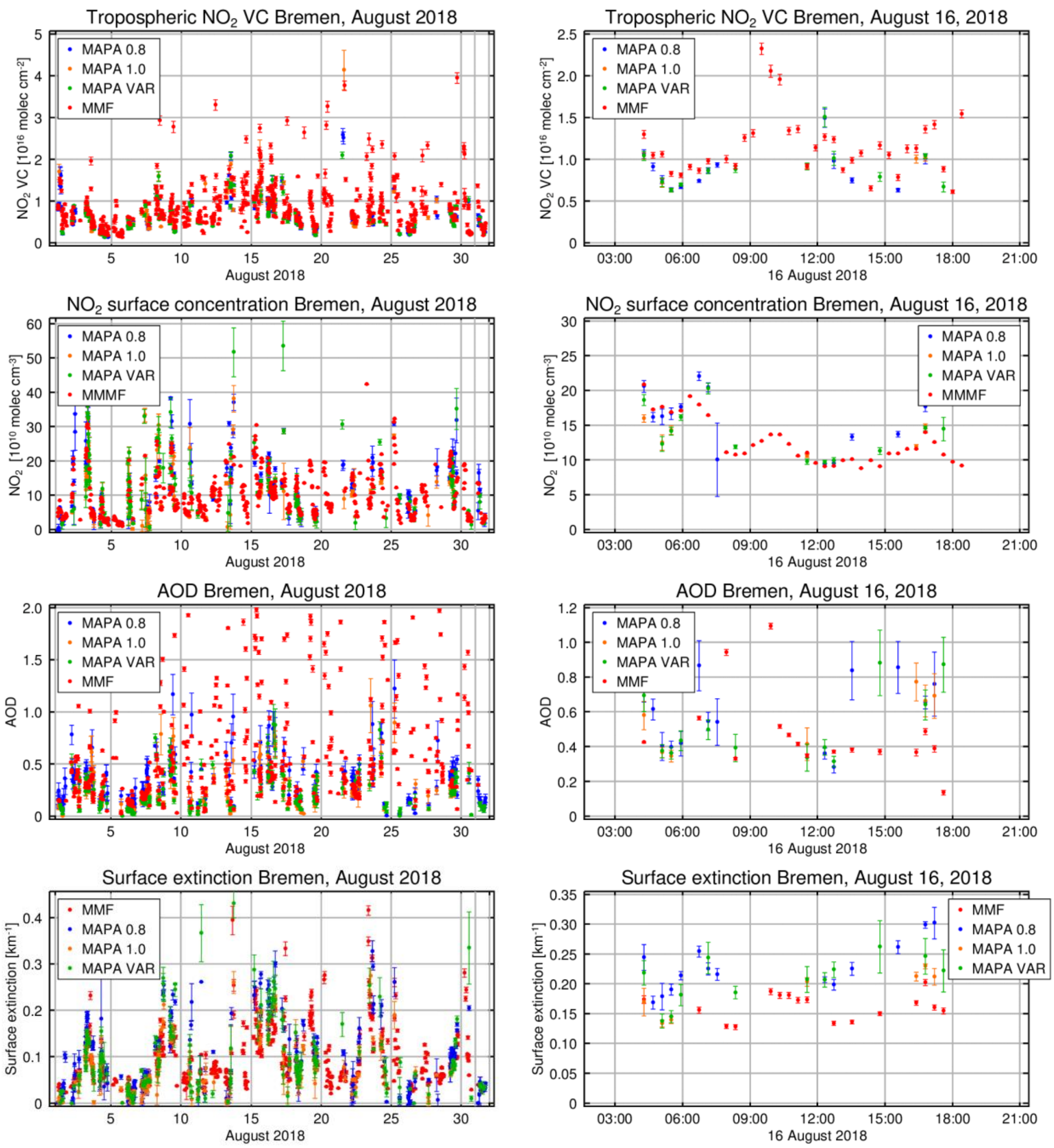

Figure 27: Retrieved values for all four algorithms and all four quantities for Bremen including error bars. The left column shows all data for August 2018, the right column a zoom on August 16, 2018. Only data without errors or warnings are shown.

4.6 Summary and conclusions

Profiles retrieved by the FRM4DOAS retrievals have been compared among each other and with the results from independent measurements during the CINDI-2 campaign. In addition, results for a month of NRT data in Bremen, Athens and Ny-Alesund have been compared to data retrieved by the University of Bremen BOREAS algorithm. The comparisons focus on the tropospheric columns, AOD and surface concentrations as profile comparison is difficult and no reliable reference data exist.

Overall, the FRM4DOAs profiling algorithms worked well on both the CINID-2 and the University of Bremen data sets. In particular for NO_2 columns, the agreement is very reasonable, both within the data set, with external validation data and with the independent BOREAS results. Surface concentrations of NO_2 and HCHO during CINDI-2 showed good agreement with independent data sets, as did the AOD retrievals. For the University of Bremen stations, AOD results are less consistent, probably because of the high probability of cloud contamination and the different treatment of these in the different algorithms.

While not in all cases the agreement between results of different retrievals is good, this is not a problem linked to the FRM4DOAS retrievals alone, but rather reflects the current state of the art in profile retrievals on MAXDOAS data. Because of the ill posed retrieval problem, different assumptions made in the algorithms are bound to lead to different results, and no clear decision can be taken on which approach is generally better unless a large and robust validation data set becomes available.

One of the most important aspects of MAXDOAS profiling is flagging of valid data, and clearly the algorithms take very different approaches here, leading to large differences in the number of profiles flagged as valid and also in surprisingly small overlap of the valid time series for individual stations.

Uncertainties reported in the products are low, in particular for MMF. This is in disagreement with the spread of values observed between different algorithms run on the same data set which could be seen as a lower estimate of the intrinsic uncertainty of the results.

For the implementation of the FRM4DOAS processing system, reasons for cases with large differences between results from the different algorithms should be investigated, flagging should be revisited and if possible harmonised more in order to come to more consistent results and uncertainties should be increased by adding additional terms.

5 References

Beirle, S., Dörner, S., Donner, S., Remmers, J., Wang, Y., and Wagner, T.: The Mainz profile algorithm (MAPA), Atmos. Meas. Tech., 12, 1785-1806, https://doi.org/10.5194/amt-12-1785-2019, 2019.

Bösch, T., Enhanced analyses of ground-based MAX_DOAS measurements – spatial inhomogeneities in the distribution of tropospheric nitrogen dioxide, *PhD thesis*, January 2019

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, https://doi.org/10.5194/amt-11-6833-2018, 2018.

Friedrich, M. M., Rivera, C., Stremme, W., Ojeda, Z., Arellano, J., Bezanilla, A., García-Reynoso, J. A., and Grutter, M.: NO_2 vertical profiles and column densities from MAX-DOAS measurements in Mexico City, Atmos. Meas. Tech., 12, 2545-2565, https://doi.org/10.5194/amt-12-2545-2019, 2019.

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.*, **12**, 2155-2181, https://doi.org/10.5194/amt-12-2155-2019, 2019.

Tirpitz, J.-L. et al., Comparison of CINDI-2 aerosol and trace gas profiles - Own dSCDs, to be submitted to AMT, 2019