FRM4DOAS Profile Retrieval Intercomparison

WP1300: MAX-DOAS Algorithm Selection

Version 1.0, 24.01.2017 – Prepared by Udo Frieß

WP Manager: Udo Frieß, HEID
WP Members: HEID, BIRA-IASB, IUP, MPIC, KNMI, NASA

Related document(s):

FRM4DOAS Slant Column Intercomparison v3 20161205.pdf

1. Introduction

The objective of FRM4DOAS WP1300 is to review the strengths and weaknesses of the MAXDOAS profile retrieval algorithms available in the scientific community in order to jointly define a community algorithm that will be implemented as baseline in the centralized processing system. Target species for the retrieval of tropospheric vertical profiles are HCHO and NO_2. In addition to trace gas profiles, the ability of the retrievals to reconstruct the aerosol profile needs to be assessed since aerosol profiles retrieved from MAX-DOAS O_4 measurements represent an important intermediate data product which serves as input for the trace gas profile inversion.

The definition of the community algorithm will be based on intercomparison exercises using synthetic slant column densities. The ability of the algorithms to reconstruct vertical profiles will assessed based on the agreement between initial and retrieved quantities including tropospheric columns, surface concentrations and parameters describing the profile shapes.

This document describes the settings for the profile retrieval intercomparison exercise. Profile retrievals will be performed using a reference dataset of differential slant columns (dSCDs) of HCHO, NO_2 and O_4, which is calculated using the median of an ensemble of simulated SCDs from the FRM4DOAS participants for a large variety of atmospheric scenarios as described in the *FRM4DOAS Slant Column Intercomparison v3 20161205.pdf* document.

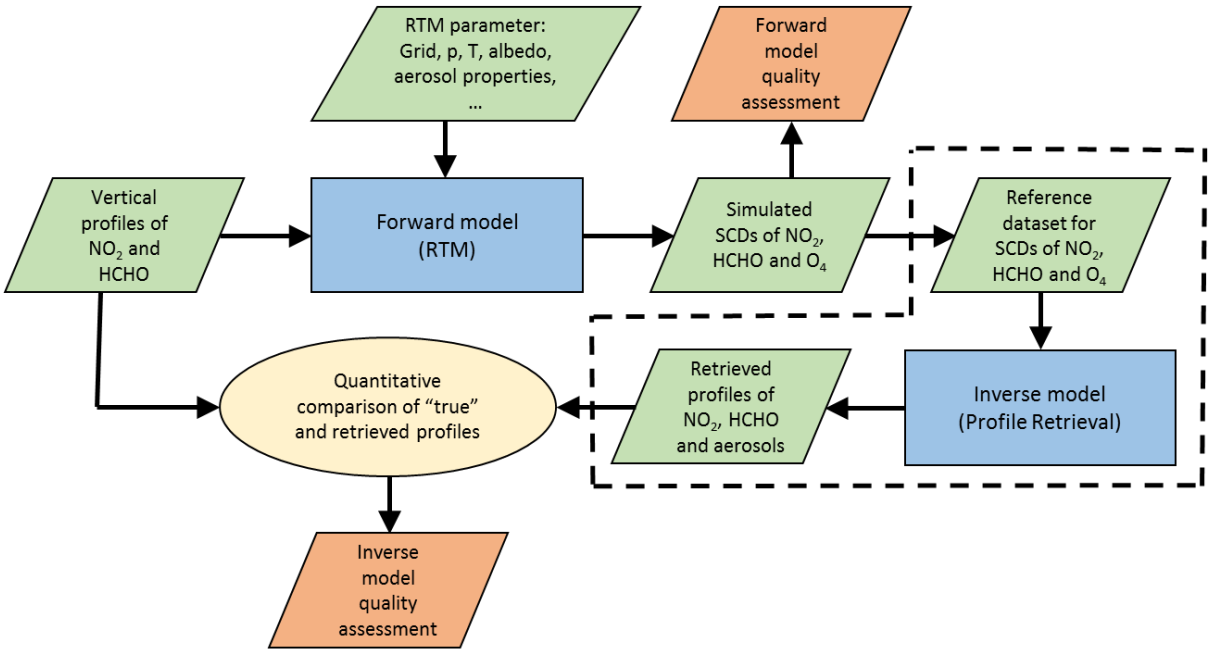

Figure 1: Flow diagram depicting the strategy for the retrieval algorithm intercomparison. The dashed line surrounds the sub-task described in this document.

2. Available retrieval algorithms and RTMs

The retrieval algorithms available within the FRM4DOAS team, together with the underlying RTM, are listed in Table 1.

Participant	Method	Algorithm	Forward Model
BIRA-IASB	OEM	BePro	LIDORT
IUPUB	OEM	Bream	SCIATRAN
MPIC	Parametrised		McArtim
MPIC	OEM	Priam	SCIATRAN
UHEID	OEM	HEIPRO	SCIATRAN
KNMI	Parametrised		DAK
NASA	OEM		?

Table 1: List of available RTM and retrieval algorithms within the FRM4DOAS consortium.

3. The Reference dSCD dataset

The reference dataset consists of dSCDs of HCHO at 343 nm, NO_2 at 343 nm, as well as O_4 at 360 and 477 nm together with according errors at 9 elevation angles for 990 different combinations of trace gas profile, aerosol profile and solar geometry (SZA and RAA). The dSCD errors for each trace gas represent typical values, and are kept constant for all scenarios. The reference dSCD dataset is randomized by assigning a random number to each elevation sequence for a particular scenario. Trace gas and aerosol profile are not specified in the scenario file and thus unknown to the participants.

INDEX	SZA	RAA	Elev	dSCD_HCHO_343	dSCD_HCHO_343_error	dSCD_NO2_460	dSCD_NO2_460_error	dSCD_O4_360	dSCD_O4_360_error	dSCD_O4_477	dSCD_O4_477_error
0	80	180	1	1.41E+17	2.00E+15	1.60E+17	5.00E+14	2.79E+43	2.00E+41	3.12E+43	2.00E+41
0	80	180	2	1.02E+17	2.00E+15	1.09E+17	5.00E+14	2.83E+43	2.00E+41	3.19E+43	2.00E+41
0	80	180	3	7.65E+16	2.00E+15	8.06E+16	5.00E+14	2.87E+43	2.00E+41	3.41E+43	2.00E+41
0	80	180	4	6.06E+16	2.00E+15	6.30E+16	5.00E+14	2.94E+43	2.00E+41	3.69E+43	2.00E+41
0	80	180	5	4.97E+16	2.00E+15	5.11E+16	5.00E+14	3.04E+43	2.00E+41	3.83E+43	2.00E+41
0	80	180	6	4.19E+16	2.00E+15	4.27E+16	5.00E+14	3.13E+43	2.00E+41	3.79E+43	2.00E+41
0	80	180	8	3.13E+16	2.00E+15	3.13E+16	5.00E+14	3.17E+43	2.00E+41	3.44E+43	2.00E+41
0	80	180	15	1.49E+16	2.00E+15	1.47E+16	5.00E+14	2.41E+43	2.00E+41	2.03E+43	2.00E+41
0	80	180	30	5.15E+15	2.00E+15	5.08E+15	5.00E+14	1.03E+43	2.00E+41	7.10E+42	2.00E+41
...	...	...	...	...	...	...	...	...	...	...	...
989	40	0	1	8.05E+15	2.00E+15	8.16E+15	5.00E+14	4.06E+42	2.00E+41	5.02E+42	2.00E+41
989	40	0	2	8.07E+15	2.00E+15	8.16E+15	5.00E+14	4.10E+42	2.00E+41	5.08E+42	2.00E+41
989	40	0	3	8.04E+15	2.00E+15	8.12E+15	5.00E+14	4.15E+42	2.00E+41	5.12E+42	2.00E+41
989	40	0	4	7.98E+15	2.00E+15	8.03E+15	5.00E+14	4.21E+42	2.00E+41	5.14E+42	2.00E+41
989	40	0	5	7.87E+15	2.00E+15	7.91E+15	5.00E+14	4.18E+42	2.00E+41	5.15E+42	2.00E+41
989	40	0	6	7.74E+15	2.00E+15	7.78E+15	5.00E+14	4.18E+42	2.00E+41	5.16E+42	2.00E+41
989	40	0	8	7.42E+15	2.00E+15	7.44E+15	5.00E+14	4.17E+42	2.00E+41	5.13E+42	2.00E+41
989	40	0	15	5.84E+15	2.00E+15	5.70E+15	5.00E+14	3.68E+42	2.00E+41	4.49E+42	2.00E+41
989	40	0	30	2.19E+15	2.00E+15	2.14E+15	5.00E+14	2.52E+41	2.00E+41	1.34E+42	2.00E+41

Table 2: The structure of the reference dataset.

The file *FRM4DOAS_dSCDs_for_inversion_randomized.dat* contains the reference dataset. It is structured as shown in Table 2. The first column (*INDEX*) contains a random index from 0 to 989, which groups the data into blocks of dSCDs at nine elevation angles (1°, 2°, 3°, 4°, 5°, 6°, 8°, 15° and 30°) (column *Elev*) at the same solar zenith angle and relative azimuth angle (columns *SZA* and *RAA*) for a specific combination of trace gas and aerosol profiles. The dSCDs are calculated by subtracting the zenith sky SCD at the same SZA and SAA from the respective SCDs. The columns to the right contain the trace gas concentrations (*dSCD_tg_wl*, with *tg* ∈ [*HCHO, NO2, O4*] and *wl* being the corresponding wavelength) and according errors (*dSCD_tg_wl_error*) in units of molecules/cm^2. dSCDs of each block

with the same index, together with the corresponding errors, serve as input for each individual retrieval run, yielding in total 990 scenarios x 4 trace gases = 3960 retrieval runs.

4. Benchmarking

An important aspect for the selection of the retrieval algorithm as part of the community algorithm is the computational performance. Please keep track of the time required to retrieve the vertical profiles separately for trace gases and aerosols and report these numbers.

5. Retrieval settings

This section describes the settings for the profile retrieval in terms of vertical discretisation, a priori constraints and aerosol optical properties. The retrieval settings are equal to those from the CINDI-2 Profile Retrieval Task Team exercise, but the output format is slightly different.

5.1. *Forward model settings*

Forward model parameters including vertical profiles of temperature and pressure, surface albedo, aerosol optical properties and absorption cross sections are similar to those used for the generation of the synthetic SCDs. For OEM algorithms, the atmosphere is divided into layers of 200 m thickness between the surface and 4km altitude. The settings are listed in Table 3:

Parameter	Value
Temperature vertical profile	Column 'T' in FRMFDOAS_atmospheres.dat [K]
Pressure vertical profile	Column 'p' in FRMFDOAS_atmospheres.dat [hPa]
Surface albedo (Lambertian surface)	0.06
Aerosol single scattering albedo	0.92
Aerosol phase function asymmetry parameter g	0.68
Instrument height	0 km
O4 absorption cross section	o4_thalman_volkamer_293K_inAir.xs
NO2 absorption cross section	no2_298K_vanDaele.xs
HCHO absorption cross section	hcho_297K_Meller.xs
Vertical grid (OEM only)	Layers with 200 m thickness between 0 and 4 km altitude

Table 3: Forward model settings.

5.2. *Retrieval settings*

The settings for the retrieval of trace gas and aerosol vertical profiles are listed in Table 4. The NO_2 retrieval is performed using the aerosol extinction profile retrieved from O_4 at 477 nm. The HCHO retrieval is performed using the aerosol extinction profile retrieved from O_4 at 360 nm. It is assumed that the aerosol extinction has no wavelength dependence (Angström exponent of zero). The specification of the a priori profiles only applies to OEM algorithms.

#	Trace gas	Wavelength [nm]	Aerosol extinction profile	A priori (OEM only)	A priori covariance (OEM only)
1	NO_2	460	From O_4 @ 477 nm	VCD = 9e15 molec/cm^2 1 km scale height	Diagonal elements: 50% of a priori profile Extra-diagonal terms: Gaussian functions with correlation length of 200 m
2	HCHO	343	From O_4 @ 360 nm	VCD = 8e15 molec/cm^2 1 km scale height	
3	O_4	360	n/a	AOD = 0.18 1 km scale height	
4	O_4	477	n/a	AOD = 0.18 1 km scale height	

Table 4: Retrieval settings.

6. Output format

This section describes the output format. For parametrised retrievals, results are stored in a single file (Section 6.4). For OEM algorithms, results are reported in separate files for profiles, profile errors, and averaging kernels (Section 6.3). Furthermore, modelled dSCDs for the retrieved profiles are reported in separate files (**Fehler! Verweisquelle konnte nicht gefunden werden.**). The identifiers for the output file names are as follows:

- [Institute] Acronym of your institute (and/or acronym the RTM)
- [Species_Wavelength] *NO2_460, HCHO_343, O4_360* or *O4_477*
- [vx] Version number, starting with *v1*
- [index] Number of the scenario as listed in the reference dataset file

6.1. *Benchmarks*

The file **[Institute]_benchmark_[vx].dat** contains the (approximate) time required to retrieve the profiles for each individual trace gas/aerosol listed in Table 4. It is structured as follows, with N being the number of retrieved profiles (should be 990) and T the total time in seconds required for the retrieval of all profiles:

Species	*N*	*T*
NO2_460	*990*	*83000*
HCHO_343	*990*	*82500*
O4_360	*990*	*102000*
O4_477	*990*	*105000*

6.2. *Simulated dSCDs*

Modelled dSCDs (for the retrieved aerosol/trace gas profile) are reported in a separate file for each trace gas/aerosol listed in Table 4. The output filename is in the following format:
[Institute]_model_dSCD_[Species_Wavelength]_[vx].dat

The output files consist of a single header line followed by the data and has the following format:

INDEX	*Elev*	*dSCD*
0	*1*	*1.234e16*
0	*2*	*2.345e15*
0	*3*	*1.234e15*
…		

6.3. *Output format for Optimal Estimation algorithms*

6.3.1. Vertical profiles

Trace gas concentrations and aerosol extinction profiles are reported in a separate file for each trace gas/aerosol listed in Table 4. The output filename is in the following format:
[Institute]_profiles_[Species_Wavelength]_[vx].dat

Each of the 990 profiles is reported in a single row, starting with the scenario index as listed in the reference dataset file, followed by *nlayers* columns specifying either the trace gas concentration (molec/cm^2) or the aerosol extinction (km^{-1}) for each layer, sorted from the surface to the top of the retrieval domain. A header line is not mandatory.

INDEX	*Layer1*	*Layer2*	…	*LayerN*
0	*2.345e10*	*1.234e10*	…	*1.234e8*
…	…	…	…	…
989	5.678e10	4.321e10	…	3.456e9

6.3.2. Vertical profile errors

As for the concentrations, profile errors are reported in a separate file for each trace gas/aerosol listed in Table 4. The output filename is in the following format:
[Institute]_error_[Species_Wavelength]_[vx].dat

The format is the same as for vertical profiles. A header line is not mandatory. If possible, noise error (xxx_noiseerror.asc) and smoothing error (xxx_smootherror.asc) should be reported separately.

6.3.3. Averaging Kernels

Averaging kernels are reported in a separate file for each profile and consist of *nlayers x nlayers* numbers. The output filename has the following format:
[Institute]_akernel_[Species_Wavelength]_[vx]_[index].dat

6.4. *Output format for parametrised retrievals*

Groups which use the parameterized approach should submit files with parameters explained in the header, followed by the data, e. g.:

* NofHeaderlines: 10
* NofColumns: 4
* Retrieval code: MPIC parametrized approach

* Version: Aerosol_v1
* X-Axis (Col 1) = Scenario index
* Y1-Axis (Col 2) = AOD
* Y2-Axis (Col 3) = Error on AOD
* Y3-Axis (Col 4) = Aerosol layer height
* Y4-Axis (Col 5) = Aerosol layer height
* INDEX AOD AOD_err h h_err

Retrieval results are reported in a separate file for each trace gas listed in Table 4. The output filename is in the following format:
[Institute]_parametrised_[Species_Wavelength]_[vx].dat