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

ESA Contract No. 4000135355/21/I-DT-lr

Deliverable D2.5:

Technical implementation plan (TIP)

Date: 2024-09-30

Version: 1d

Contributing authors:

Steffen Beirle, Lucas Reischmann, Steffen Ziegler, Thomas Wagner (MPIC)
Martina Friedrich (IASB-BIRA)

Content

Acronyms and abbreviations

Acronyms and abbreviated terms that are specific for this document can be found below.

AMF	Air Mass Factor
ATBD	Algorithm Theoretical Basis Document
CCA	Cloud Classification Algorithm
CI	Color Index
CPS	Central Processing System
dAMF	Differential Air Mass Factor
dSCD	Differential Slant Column Density
EA	Elevation Angle
FRM_4DOAS	Fiducial Reference Measurements for Ground-Based DOAS Air-Quality Observations
FRS	Fraunhofer Reference Spectrum
LOC	Lines Of Code
MPIC	Max-Planck Institute for Chemistry
SCD	Slant Column Density
SZA	Solar Zenith Angle
VCD	Vertical Column Density
WP	Work Package

This document presents the current status of the implementation of the cloud classification algorithm at MPIC (section 1) and the essential technical steps needed for its implementation in the central processing system (section 2).

1 The Cloud Classification Algorithm (CCA)

The CCA developed at MPIC Mainz, as described in the ATBD [RD-1], is implemented at MPIC as a standalone Python script with currently about 700 LOC. Central part of the CCA is a CloudClassification class, providing methods for data input, calibrations, cloud classification, and output.

The program structure is as follows:

- Read input files:
  The code supports different file formats, including the standard FRM4DOAS input file format. The data is internally re-organized in lists of "sequences" of color indizes (CI) and O4 dSCD, where each sequence of elevation angles (EAs) is framed by the two zenith sky measurements before and after the sequence.
- Perform calibrations:
  - The CI is calibrated based on the accumulation point of the CI for clouded scenes in zenith.
  - The O4 AMF of the Fraunhofer reference spectrum (FRS) is calibrated based on the lower accumulation point of cloud-free O4 measurements with solar zenith angle (SZA) between 30° and 50°. As the O4 VCD is provided, this is equivalent to the calibration of the O4 SCD of the FRS.
- Perform cloud classification:
  Cloud classification is based on the (calibrated) CI and O4 dAMF and their temporal variation. For details see [RD-1] and Wagner et al., 2016.
- Provide classification results:
  Currently, the classification results are returned encoded as binary, where each cloud category takes one bit.

Note that

- Calibration and cloud classification are currently not strictly separated. Reason for this was that part of the cloud classification result (primary classification based on CI) is currently used in order to mask out clouded scenes *before* the O4 calibration is done.
- During calibration, figures are created documenting the frequency distribution of values and the identified accumulation points. These figures need to be manually inspected and checked for plausibility. Thus, the calibration that has to be performed for each station cannot run fully automated without supervision.
- Some modifications/extensions made during the activities within FRM4DOAS 2.0 WP 2, as described in detail in D2.2, still need to be implemented in the current CCA:
  - The new sub-categories "constantly clear" and "constantly cloudy" based on more strict thresholds.
  - The change flag based on the change of the cloud classification from sequence to sequence.

  These extensions also require to modify the output/return format of the cloud classification, as more (sub)categories and flags require additional bits. Details of the revised output format still need to be discussed at MPIC.

After implementation of the new features, the source code will be refactorized and polished in the coming months before delivering it as D2.4 in March 2025.

The settings for calibration will be controlled by a configuration file, i.e. are de-coupled from the source code.

2 Requirements, modifications and extensions for the implementation of the CCA in the Central Processing System (CPS)

The CCA exists as standalone program developed at MPIC Mainz. The cloud classification algorithm works very fast: 8 months of MAX-DOAS data were processed in less than 1 second. Thus, computational costs of the CCA are completely negligible within the CPS.

For the implementation of the CCA within the CPS, the python code developed at MPIC could be used. Before this, however, some modifications and extensions of the CCA as well as the CPS are required (see also Table 1). These are detailed in the following sections.

2.1 Separation of calibration and cloud classification

In the current implementation of the CCA, the calibration is performed each time it is called for the provided timeseries. Within the CPS, however, cloud classification results should be provided for daily updates of the MAX-DOAS input data, without re-processing the historical time series. Therefore, the code needs to be re-structured such that calibration and cloud classification can be run separately. For this purpose, MPIC will provide a new calibration module for the calibration of the CI and the O4 AMF of the FRS. This module has to be called for each station for different time periods. For the calibration, the following steps need to be performed:

2.1.1 Calibration period

A period used for calibration has to be defined. This period has to be long enough to cover various cloud conditions, in particular a sufficient amount of cloud free as well as clouded observations. On the other hand, the time period should not be too long in order to avoid poor performance or erroneous results in case of instrument degradation. In mid-latitudes, typically a few months of input data are expected to yield reasonable results. We thus recommend to perform the calibration for a period of 6 months.

Calibrations should be updated every 3 months. I.e., two succeeding calibrations have an overlap of 3 months in order to avoid jumps in the calibration results (which would be more likely to occur if the calibration would be updated every 6 months).

For the application of the CCA, the results from the latest (w.r.t. date of measurement) available period should be used.

2.1.2 The FRS

For each calibration period, a FRS from within the calibration period has to be selected. The FRS should be chosen as a zenith measurement around local noon. Saturated spectra or spectra of low intensities (e.g. under thick cloud cover) should be avoided. In order to select a suitable FRS, the following steps should be applied:

- A pool of FRS candidates is compiled from a time interval of +/- 10 days around the center of the calibration period. All zenith measurements within 1 hour of local noon should be kept.

- If the number of candidates is below 30, the intervals should be doubled (+/- 20 days, +/- 2 hours around local noon).
- From these candidates, the FRS with median radiance (counts per second) at 380 nm (+/- 0.5 nm) should be selected.

This FRS has to be used for the DOAS retrieval of O4 dSCDs (see section 2.2).

2.1.3 Manual and automated checks

During calibration, figures are created documenting the frequency distribution of values and the identified accumulation points. For each station, especially the calibration results of the first calibration period need to be carefully inspected and checked for plausibility.

For the subsequent calibration periods, calibration results can be compared against the preceding period, and the results will provide information on the stability of instrument and algorithm. These comparisons can – at least partly - be automated:

- Significant changes in the CI calibration can be due to
    - Instrument degradation
    - Failure of the accumulation point determination

  The concrete reason requires manual inspection of the calibration results. But the change of the CI calibration can be used to create an automated alert in order to make the user aware of the inconsistency. In addition, warning/error flags can be added automatically to the CCA results.
- As different FRS are used for the different calibration periods, the O4 AMF of the FRS is expected to differ as well. I.e. a change of the calibrated O4 AMF is not indicating any issues. Still, the O4 calibration might be affected by a change of instrumental performance. In order to check for instrument stability, we propose to check the results of the DOAS analysis such as the RMS or spectral shift and stretch. Again, large increases of the RMS or large changes of the shift and stretch during the selected calibration period should raise an automated alert, and add warning/error flags to the CCA results.

So far we do not have experience with cloud classification results for long time series. Thus, calibration results should be checked from time to time for the regular calibrations based on the overview plots created by the CCA calibration module.

2.1.4 Configuration

The calibration settings will be controlled by configuration files. I.e., it will be possible to modify the calibration periods (individually for each station) without modifying the source code in order to be able to deal with degradation or other instrumental effects.

2.2 Input data for CCA and CCA calibration

The CCA needs input data as lists of elevation sequences, organized in a similar way as the data passed to MAPA within the CPS. Each sequence is required to include the preceding and succeeding zenith measurements. Which input data is needed is described in Sections 2.2.1 -- 2.2.3.

The CCA is going to use the sequence definition of the CPS. For comparisons of the standalone CCA with e.g. validation data sets, it might be helpful to also allow the different contributing groups to define the elevation sequences in the level1 spectra file. For consistency, a sequence ID has to be assigned for each off-axis measurement, as well as a measurement number for the zenith before and after. This would also allow to interpret the default Pandora "v-scan" as two separate sequences. This change affects the Level1 check module (netCDF_L1.py) of the CPS. For each sequence, apart from information about the SZA and the telescope EA, a number of quantities are required for each measurement in a sequence. These quantities are described in the following sub-sections. Note that the CCA calibration would only require the zenith sky measurements. For simplicity, however, we propose to have the same input data for CCA calibration and CCA.

Extracting the data and passing it to the CCA in the right format, affects the Tropospheric Preprocessing module (ProfilePreprocessBira.py) of the CPS. Including the returned information from the CCA in the FRM4DOAS level2 files also affects the same module in the CPS.

2.2.1 Color index: Radiances at 330 nm and 390 nm

The color index, i.e. the ratio of radiances at two wavelengths, is a key quantity for cloud detection in MAX-DOAS measurements. In the CCA, the color index is calculated from radiances at 330 nm and 390 nm. These radiances (averaged within +/- 0.5 nm) thus have to be provided. This change affects the QDOAS processing module and settings in the CPS.

2.2.2 O4 dSCD

The O4 dSCD for each EA is required. In contrast to the O4 dSCD used within the aerosol profile retrieval, which is already part of the CPS, the CCA requires the O4 dSCD analysed with the fixed FRS for the time interval of the respective calibration, see above, which thus needs to be added to the CPS. This change, as the previous one, affects the QDOAS processing module and settings in the CPS.

2.2.3 O4 VCD

For the conversion of the O4 dSCD into dAMFs, the VCD of O4 is required. Currently, this quantity is calculated in the CPS within MAPA, based on vertical profiles of temperature and pressure. The MAPA output contains the O4 VCD, which might be used as input for the CCA in the CPS.

However, instead we propose a more modularized approach, in which the O4 VCD is calculated in a first step in the CPS and then passed to both, CCA and MAPA:

- MPIC will modify the MAPA code such that the O4 VCD is calculated in a separate function from input data containing vertical profiles of pressure and temperature, organized analogue to the current data input to MAPA.
- The modified MAPA routine then requires the O4 VCD as additional input field. Pressure and temperature profiles are not needed any more.
- The O4 VCD matching the FRS has to be calculated and passed to the CCA calibration.
- The O4 VCD matching the FRS as well as the O4 VCD for each sequence need to be passed to CCA.
- For the CPS, the workflow and passed parameters need to be adopted accordingly.

2.2.4 CCA calibration

For cloud classification, the results of the CCA calibration have to be provided for the matching calibration period for each instrument.

2.3 Empirical correction for missing 390 nm

Some instruments do not cover 390 nm, or apply filters that reduce radiances at 390 nm. For these instruments, the current CCA implementation is expected to yield suboptimal results or might not work at all.

For these instruments, an empirical adjustment might be applied based on measurements at 380 nm. However, this approach requires further investigations:

As a first step the consistency of the radiances at 380 and 390 nm for different sky conditions and SZA should be investigated. The underlying idea is that radiances at 380 nm might be used instead of those at 390 nm. To investigate the consistency of the radiances at 380 and 390 nm, the $CI_{380/390}$ should be calculated for an instrument which covers the radiance at 390 nm with good signal to noise ratio. The mean and standard deviation of the $CI_{380/390}$ should be calculated as function of the SZA.

As next step, the derived $CI_{330/380}$ will be multiplied by the average $CI_{380/390}$ to approximate the $CI_{330/390}$. Possibly the SZA dependence could be considered. The scaled CI is then used as input for the CCA. The results are compared to the CCA results for the true $CI_{330/390}$.

If the deviations between both CCA results are small, for instruments without (useful) radiance measurements at 390 nm, the $CI_{330/390}$ might be simply replaced by the $CI_{330/380}$.

If the deviations between both CCA results are large, new threshold values for the $CI_{330/380}$ need to be calculated.

In any case, to allow for investigations of such a potential empirical adjustment in the future, we recommend to include the radiance @ 380 nm in the CPS already now.

2.4 Application of the cloud classification results for the VIS channel

The CCA yields cloud classification results based on radiances in the UV. Several instruments, however, (also) provide a VIS channel. In general, the CCA results might be applied here as well, as clouds are expected to affect measurements in the UV and VIS similarly.

In case of synchronous measurements for UV and VIS, a direct match of elevation sequence is provided, and the CCA results could directly be passed to the VIS results.

In case of unsynchronous measurements, such a direct link is not given, and the useability of the CCA results for VIS is probably limited. We expect that CCA results can be interpolated in time, in particular if the change flag is not raised. However, so far, this has not been tested, and a possible application of the CCA results for VIS measurements requires further investigations.

Technically, the usage of CCA results from the UV channel within the VIS channel would also require fundamental modifications in the CPS processing chain, where the concrete effort is hard to estimate. We thus propose to postpone this step as a potential future upgrade.

Table 1. Identified tasks needed for implementing the CCA into the CPS and estimates of the respective workload.

Task	Affected module(s) in the CPS	Estimated Workload (person months)
Coding		
De-couple calibration and cloud classification in the CCA	New module CCA calibration New module CCA New set of configuration files	1.5 (MPIC)
Allow for pre-defined sequence id in submitted spectra (optional)	netCDF_L1.py	0.5 (BIRA)
Include new wavelength for radiances, adjust wavelength delta for radiances, include specific O4 fitting window with a fixed FRS and add logic for fixed FRS adjustment	QDOAS config files and wrapper	0.5 (BIRA)
Write wrappers for the new module CCA calibration, to perform calibration and adjust the config files for QDOAS and the CCA	New wrapper	1.0 (BIRA)
Create module for O4 VCD calculation detached from MAPA	New module O4 VCD Change to MAPA	0.5 (MPIC)
Adjust CPS interfaces for the interaction with the new CCA module and the new O4 VCD module, as well as the change in data stream to and from MAPA	ProfilePreprocessBira.py	1.5 (BIRA)
Determine suitable FRS for O4 analysis	New function/wrapper	1.0 (BIRA)
Quality control and further investigations		
Perform initial calibration for each instrument, check results (human interaction needed)	CCA calibration (new)	0.25 per 15 instruments
Perform calibration in regular time periods for each instrument, check results (human interaction needed)	CCA calibration (new)	0.25 per 30 instruments per year
Investigate $CCI_{330/380}$ as potential replacement for $CCI_{330/390}$		1.0 (MPIC)

3 References

[RD-1] FRM4DOAS-2.0 deliverable D-2.2 ‘ATBD for the cloud classification algorithm’

Wagner, T., Apituley, A., Beirle, S., Dörner, S., Friess, U., Remmers, J., and Shaiganfar, R.: Cloud detection and classification based on MAX-DOAS observations, Atmos. Meas. Tech., 7, 1289–1320, https://doi.org/10.5194/amt-7-1289-2014, 2014.

Wagner, T., Beirle, S., Remmers, J., Shaiganfar, R., and Wang, Y.: Absolute calibration of the colour index and O4 absorption derived from Multi AXis (MAX-DOAS measurements and their application to a standardised cloud classification algorithm, Atmos. Meas. Tech., 9, 4803–4823, https://doi.org/10.5194/amt-9-4803-2016, 2016.