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

FRM4DOAS 2.0

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

<u>Deliverable D4.5</u>:
Technical implementation plan (TIP)

Date: 2026-05-12

Version: 1.1

Contributing authors:

Andreas Richter, Kai Krause (IUP-UB)

Martina M. Friedrich (IASB-BIRA)

Content

Acronyms and abbreviations

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

ATBD	Algorithm Theoretical Basis Document
CPS	Central Processing System
FRM_4DOAS	Fiducial Reference Measurements for Ground-Based DOAS Air-Quality Observations
VCD	Vertical Column Density
WP	Work Package
MMF	Mexican MAX-DOAS Fit
MAPA	Mainz Profile Algorithm

1 Introduction

This document presents the essential technical steps needed for the implementation of the advanced urban tropospheric NO_2 product developed in WP4 and described in D-4.2 and D-4.3. The goal in this WP was to improve the characterization of the horizontal NO_2 distribution around MAX-DOAS instruments in urban environments and ultimately provide information about horizontal inhomogeneities to the user.

2 New variables

The horizontal variability of the NO_2 field around a MAX-DOAS station is evaluated using six different inhomogeneity indicators:

1. Inhomogeneity in azimuth: absolute and relative
2. Inhomogeneity in time: absolute and relative
3. Inhomogeneity in time and azimuth: absolute and relative

which are described in detail in deliverable D-4.2. The inhomogeneity indicators are continuous indicators and not a simple flagging. For each indicator there is also a variable that keeps track of the number of individual VCDs that were used to calculate the respective indicator. Based on these indicators and predefined thresholds, representative columns are calculated and saved in conjunction with auxiliary data into individual subgroups in the data file.

1. Representative mean column density filtered for inhomogeneity in azimuth
2. Representative mean column density filtered for inhomogeneity in time
3. Representative mean column density filtered for inhomogeneity in time and azimuth

Each subgroup has its own independent time grid. However, for every representative VCD, the individual scan indices of the contributing VCDs are saved as ancillary data within the subgroup.

In total, this will add 9 new variables and 3 new subgroups with 21 variables in each subgroup to every MMF or MAPA result. This will increase the file size by around 7% compared to the input files.

3 Requirements, modifications, and extensions in the CPS

The inhomogeneity indicators and the representative columns are calculated in a stand-alone python script with about 650 lines of commented code.

The program first calculates the inhomogeneity indicators and, in a second step, provides representative VCDs for the station based on these inhomogeneity indicators and filter settings presented in D-4.3. The program is structured as follows:

1. Read LV2 FRM4DOAS input files and automatic identification of used variable paths
2. Calculation of inhomogeneity indicators for each retrieval algorithm
3. Calculation of representative columns based on inhomogeneity indicators and initially suggested values filter values from D-4.3 for each retrieval algorithm
4. Appending the results to the level 2 files and saving the updated file

Currently, the following Python packages are used to implement the processor:

1. python=3.12.7
2. numpy=2.3.5
3. xarray=2025.10.1
4. pandas=2.3.3
5. netcdf4=1.7.2

while the CPS uses:

1. python=3.12.3
2. numpy=1.26.4
3. xarray=2025.1.2
4. pandas=2.2.2
5. netdf=1.6.5

Since the current CPS and the demonstrator software use different package versions, testing is required to verify whether the demonstrator software is compatible with the package versions available on the CPS.

At the moment, the inhomogeneity indicators are processed on yearly files, but daily processing as it is done in the CPS should work without any changes to the code base. Since the inhomogeneity indicators do not interact with existing algorithms, the processing can be run in a separate step after all other existing processors, so no changes to the existing code base are to be expected.

However, there are some changes that need to be made to the configuration files. First of all, the L2_configure.yml needs to be updated with the new variables, their datatypes, descriptions, units, and dimensions for each inhomogeneity indicator and its representative columns. New station specific config files need to be created to specify where the needed variables can be found and which inhomogeneity thresholds should be applied for each station.

List of affected components:

- code base:

  ProfilePreprocessBira: write module not affected, however config file for writing needs to be adapted

- config files

  Change of L2_configure.yml to include proper fillvalues, datatypes, descriptions, long names, units and dimensions for the new output variables

  Introduction of new config files (station specific), to specify the variable paths in the level 2 files and to specify constants for the representative columns (medians, thresholds)

4 Workload and computational costs

Computational time for the algorithm on the yearly files differs substantially, depending on the amount of data and number of azimuth directions. However, the exact impact of the number of azimuth directions could not be determined and seems to be less important than the overall number of observations. For stations such as Athens, the algorithm takes about 1.5 hours to run, while for other stations it is below 10 minutes.

As the inhomogeneity indicators can be added to the processor as another add-on and interference with existing program parts is not expected, most of the workload is expected to be used for testing and verification of the results of the daily files against the already existing yearly results. A workload of 2 person-months is to be expected.