

GMDD. List of changes

The references to pages in this document refer to the document diff.pdf, generated by latesdiff.

My answers to the referees comments are in italics

### **Revisions suggested by referee #1**

Specific comments: The abstract should summarize algorithm features.

*a line was added to the abstract describing some specific features of the algorithms*

Many acronyms, although likely familiar to the reader, are not defined – e.g., NORS, NRT, TM-5, MOZART, ECMWF, etc.

*I have run through the paper and clarified abbreviations: e.g. on page5 (added webpages for TM5 and MOZART), page 6. NORS was defined on page2, in the first paragraph.*

Sec. 2.4 p. 8159 (line 14) & Sec. 3.1 p. 8160 (line 4) “To fix thoughts” is not a familiar expression – maybe use “As an example”

*done*

Sec. 3.1 p. 8161 (line 6) – this sentence was confusing when I first read it – it sounded like NaN values would be substituted with something else. Maybe say: “A final step in the re-gridding process is to set profile values to void (or NaN) for layers of the external grid that are not or only partly covered by the source grid.”

*changes have been made.*

Sec. 3.3.3 p. 8162 (line 21) Not sure what is meant by “gluing”.

*I removed the gluing sentence and clarified the next sentence*

Sec. 3.4 p. 8164 (line 5) “brought to the units” should be “converted to the units”

*done*

Sec. 3.4: Figures with representative NDACC AVKs, O3 profiles and comparisons with model profiles (before & after smoothing) would be useful to the reader in visualizing these algorithm steps.

*I inserted 2 plots: a partial column profile plot and an AVK plot*

Sec. 3.4 p. 8165 – Please provide more details about the algorithm for applying a column AVK, i.e., is this applied to a model profile or does it require computing the model total column first. Also, are there many cases where the “measurement grid outranges the model grid”?

*A new formula for column averaging is inserted on page 15. The measurement grid outranges the model grid almost always.*

Sec. 4.1 p 8166 (lines 9,10) “in VMR unit toward partial column units” should be “in VMR units to partial column units” and fix “partia l”

*done*

### **Revisions suggested by referee #2**

Specific comments: Page 8153 line 14: replace ‘partly’ with ‘partially’.

*done*

Page 8153 (lines 10-17): This paragraph is confusing on first read. The distinction between VAL and NORS would be better explained if an example of the type of instrument used for the in situ surface data were provided. Also in line 16: replace ‘almost not the NDACC groundbased’ with ‘little use of the NDACC ground-based’.

*done*

Page 8154 line 3: replace ‘the notation convention’ with ‘notation conventions’.

*done*

Page 8154 line 3: replace ‘scalar,’ ‘scalar;’.

*done*

Page 8154, line 14,15: Most readers will not know what the term consolidated refers to in this context which is simply a statement of whether the data provided in smaller collections (perhaps a file with one month of data) for the RD set and a larger collection (perhaps a file with one year of

data) for the NDACC set. Also, the RD data is not necessarily the final PI reviewed validated NDACC quality controlled product. Since this is rather complex, it is likely better to state here that “the NDACC database rapid delivery directory, if not fully in final form, or to the corresponding NDACC database station directory if in its final form both with respect to data versioning (PI reviewed vs operational) and file temporal coverage”.

*done*

Page 8155, line 16-18: Punctuation should be: The dimension column is only indicative. (These values change when considering other model grids. For notational convenience we fix the dimensions to the grid N128.)

*done*

Page 8156, line 4: should be “model’s pressure levels” or “model pressure levels”.

*done*

Page 8156, line 11: should be “and the”, not “and, the”

*done*

Page 8158, line 11: The phrase “recall that z is the MACC grid height vector is descending” has two verbs “is” and is confusing. Probably should be: “recall that the MACC grid height vector, z, is descending”.

*done*

Page 8159, line 10: should be “and converted to the measurement’s units.”

*done*

Page 8160, line 3: use “conservation of mass in mind”, or with consideration of conservation of mass”.

*done*

Page 8160, line 4-5: As previously “To fix thoughts” is non-standard English. Also the parentheses are misplaced as not all inside of the “( )” relate to the Ozone profile. Lastly “profile” appears as both singular and plural in different parts of the sentence. Possibly correct these items with: “To clarify, assume that an O3 profile (in partial column units for concentration) or an aerosol extinction profile ( in optical depth) is defined. . . .”

*done*

Page 8161, line 20: Should be “; however,”

*done*

Page 8166, line 25: Replace “To fix thoughts” with “For example” or “Specifically”.

*done*

Page 8167, line 14: Use “request to S. Niemeijer”.

*done*

Page 8167, line 19: Use “for many years”.

*done*

Page 8168, line 6: Define QA4EO, and why compliance with this is valuable. Refer to the item in your reference list.

*QA4EO was mentioned in the first section, compliance was mentioned there as well. Added reference.*

Page 8168, line 12: remove the comma between “covered” and “is”.

*done*

Page 8168, Acknowledgements: An appropriate acknowledgment of NDACC should include its web address and a statement that the data are publically available. Perhaps use: The NORS project relies on the data publically available from NDACC. See <http://www.ndacc.org>. The data providers for FTIR, MWR, LIDAR and UVVIS measurements in the NDACC database are acknowledged.

*I added the reference to the ndacc website. This website was mentioned in Section 1.*

Page 8172 Table 2, under description: the phrase ‘latitude instrument’ should either be ‘instrument latitude’ or ‘latitude of the instrument’. Likewise for longitude and altitude.

*done*

Page 8173 Table 3, under description: the phrase ‘array ap is used as translation terms’ is better expressed as ‘array ap is the translation operator’

*done*

Page 8173 Table 3, under description: the phrase 'array bp are used as scaling factors' is better worded as 'array bp contains the scaling factors'. Note this corrects the pluralization of the verb 'are' for the singular noun 'array'.

*done*

Page 8174 Table 4: It is confusing why either all instruments (FTIR for example) or all species (O3 for example) are not listed together. There does not seem to be a logical order to the placement in this table. Grouping would better serve to help the reader recognize the limits of various measurement techniques.

*All tables follow the same ordering: FTIR, MWR, LIDAR, UVVIS ZENITH and UVVIS OFFAXIS*

Page 8175 Table A1: If this table is supposed to be an elucidation of Table 1, then it would make sense for the ordering of the first column to correspond with that of Table 1. Currently Table 1 runs: UVVIS, FTIR, MWR, Lidar; whereas Table A1 runs as: Lidar, FTIR, UVVIS, MWR.

*see above*

Manuscript prepared for Geosci. Model Dev. Discuss.  
with version 2014/09/16 7.15 Copernicus papers of the  $\LaTeX$  class copernicus.cls.  
Date: 6 March 2015

# Description of algorithms for co-locating and comparing gridded model data with remote-sensing observations

**B. Langerock<sup>1</sup>, M. De Mazière<sup>1</sup>, F. Hendrick<sup>1</sup>, C. Vigouroux<sup>1</sup>, F. Desmet<sup>1</sup>, B. Dils<sup>1</sup>, and  
S. Niemeijer<sup>2</sup>**

<sup>1</sup>Belgian institute for Space Aeronomy, Ringlaan 3, 1180 Ukkel, Belgium

<sup>2</sup>S&T, Olof Palmestraat 14, 2616 LR Delft, the Netherlands

Correspondence to: B. Langerock (bavo.langerock@aeronomie.be)

## Abstract

MACC-II,III, Monitoring Atmospheric Composition and Climate, is the current pre-operational Copernicus Atmosphere Monitoring Service (CAMS). It provides data records on atmospheric composition for recent years, present conditions and forecasts for a few days ahead. To support the quality assessment of the CAMS products, the EU FP7 project NORS created a server to validate the gridded MACC-II,III/CAMS model data against remote-sensing observations from the Network for the Detection of Atmospheric Composition Change (NDACC), for a selected set of target species and pilot stations. This paper describes in detail the algorithms used in this validation server. Amongst others, the algorithms take into account the horizontal displacement of the measured profiles from the location of the instrument, the vertical averaging and uncertainty propagation.

## 1 Introduction and notations

MACC-III, Monitoring Atmospheric Composition and Climate (<http://copernicus-atmosphere.eu>), is the current pre-operational Copernicus Atmosphere Monitoring Service (CAMS). It combines state-of-the-art atmospheric modelling with Earth observation data to provide information services covering European air quality, global atmospheric composition, climate forcing, the ozone layer and UV and solar energy, and emissions and surface fluxes. The EU FP7 R&D NORS project (Demonstration Network Of ground-based Remote Sensing Observations in support of the Copernicus Atmosphere Service, <http://nors.aeronomie.be>) was set up to support the development and generation of fit-for-purpose CAMS data products and services by providing quality information based on validation results. In NORS the validation will be carried out using NORS data products which are essentially ground-based remote sensing data from the Network for the Detection of Atmospheric Composition Change (NDACC; see <http://www.ndacc.org>), optimized for the needs of the CAMS validation.

The validation processes carried out in NORS were created to be compliant with best practices as defined by the international community: all validation results include traceability information (see the Global Earth Observation System of Systems (GEOSS), Quality Assurance for Earth Observation, QA4EO, 2010 and follow the validation roadmap for Copernicus atmospheric data and services formalized in the MACC Validation Protocol, Lambert, 2010; Huijnen and Eskes, 2012). The validation support by NORS is delivered as a web-based application that generates default validation reports in an operational and automatic way, but which can also be used for the generation of dedicated user-driven validation reports on demand (<http://nors-server.aeronomie.be>).

NORS is a demonstration project: it focuses on a limited number of target data products from a limited number of pilot NDACC stations representative of four major atmospheric regimes (Table 1).

The validation service is built such that it is easily expandable to a larger number of stations and instruments and to additional CAMS data products for which NDACC can provide independent reference data. The MACC-II,III validation subproject (VAL) has a focus on **the reactive gases and aerosol composition on a global scale, which coincides ~~partly~~ partially with the NORS/NDACC target species. Duplication ~~, however,~~ with existing validation tasks in MACC is not an issue because MACC-II,III VAL is mostly using in situ surface data (Aeronet) or satellite total column data (e.g. MOPITT and IASI CO data), as reference data for the validation of the MACC-II,III products, ~~and almost not the NDACC ground-based remote-sensing data that are the reference data in NORS.~~**

VAL produces 3 monthly validation reports for the near-real time services of MACC-II,III, and 6 monthly validation reports (updates) for the reanalysis services of MACC-II,III ([http://www.copernicus-atmosphere.eu/services/aqac/global\\_verification/validation\\_reports](http://www.copernicus-atmosphere.eu/services/aqac/global_verification/validation_reports)). The results of the NORS validation server have been used in the most recent VAL reports and in the latest reanalysis report.

The paper is organized as follows: in Sect. 2 a brief description of the measurement and model data is presented. For the measurements, the basic GEOMS (Generic Earth Observation Metadata Standard, Retscher et al., 2011) variables and their notations are

introduced. For the model data, the algorithms to set up a vertical height grid and to do unit conversions are outlined. Section 3 contains the essential steps for comparing a single measurement with a model output. It describes the re-gridding, temporal and spatial collocation and smoothing algorithms for model vertical profile data. Section 4 contains further details on how the validation results are presented in their final output: e.g., time averaging of column data, uncertainty propagation, etc.

Regarding ~~the notation convention~~notation conventions, all multiplications are scalar; matrix multiplication is denoted by a central dot  $\cdot$ . Most of the notations used in the algorithms can be found in Tables 2, 3 and A1. For example, the profile array  $\Theta_3^M \cdot O_3^M$  is linked to the MACC parameter go3 and denotes a target species vertical profile. To distinguish between the MACC and NORS vertical profile, a superscript notation is used: e.g.  $\Theta_3^M \cdot O_3^M$  is the MACC model profile and  $\Theta_3^N \cdot O_3^N$  is the NORS measurement profile. If it is clear from the context to what class a data array belongs (model or measurement), the superscript is omitted to ease the notation.

## 2 Description of NORS and MACC data

### 2.1 NORS data

NORS data files are delivered in rapid delivery mode (not later than one month after acquisition) to the NDACC database rapid delivery directory, if not fully consolidated in final form, or to the corresponding NDACC database station directory if fully consolidated in its final form both with respect to data versioning (PI reviewed vs operational) and file temporal coverage. For each measurement technique the NDACC data format has to be compliant with a pre-defined template (see <http://avdc.gsfc.nasa.gov>). The mapping of the template variable names to the mathematical concepts used throughout this paper (e.g. the target profile, the averaging kernel AVK, ...) is depicted in Table A1. Table 2 lists the GEOMS variables that all templates have in common (the ~~given~~ dimensions are exemplary).

All measurement techniques, except LIDAR, report averaging kernels, a priori profiles and uncertainties.

For each measurement technique, site and target species a specific sensitivity range can be determined. For further details on sensitivity ranges and typical AVK's, see the data user document that was developed within the framework of the NORS project (De Mazière et al., 2013).

## 2.2 MACC data

At present, the validation server validates in operational mode the following forecast runs: ~~NRT o-suite~~ near-real-time operation suite (NRT osuite), the experiment ~~with TM-5 chemistry~~ running the TM5 3D atmospheric chemistry-transport model and the experiment with MOZART chemistry (for further details see [https://www.gmes-atmosphere.eu/oper\\_info/nrt\\_info\\_for\\_users](https://www.gmes-atmosphere.eu/oper_info/nrt_info_for_users), <http://tm.knmi.nl>). The model data is generated regularly on specific *output times*  $t_o$  (every 12 h for NRT o-suite and every 24 h for the other two models) and each output contains 3 hourly forecast data valid in the future of the output (see Fig. 1). The experiments are downloaded from the Meteorological Archival and Retrieval System (MARS) archive on their native IFS (Integrated Forecast System) resolutions<sup>1</sup> on regular Gaussian grids: T255N128 for NRT o-suite and T159N80 for the other models.

For each of these experiment versions, the validation server generates a time sequence of MACC model data with a time interval of 3 h between two MACC output times (see Fig. 1) such that the forecast validity times are as close as possible to the corresponding output times.

Table 3 describes the MACC data fields that are used in the validation server. The first column gives the MARS short name or the parameter id. The dimension column is only indicative (these values change when considering other model grids, ~~for~~ For notational

---

<sup>1</sup>An IFS resolution is denoted by  $T_xN_y$ :  $x$  indicates the spectral truncation and  $y$  is the number of latitudes between a pole and the equator (for regular Gaussian grids, the number of longitude bands is  $4y$ ).

convenience we fix the dimensions to the grid N128). Table 3 shows the notations of the MACC variable as they will appear in the algorithms.

The surface height variable  $z_s$  [m] is the geopotential height field corresponding to the surface pressure. It is not available in the MARS archive for the MACC experiments under consideration and is downloaded directly from ECMWF ~~for~~ ([European Centre for Medium-Range Weather Forecasts](#)) ~~for~~ several resolutions available. The server will automatically attach to a given model the geopotential field that matches the native IFS resolution of the model.

### 2.3 Vertical height grid for MACC data

This section contains a detailed description of the calculation of a vertical height grid for MACC data, starting from the vertical pressure coordinate. The algorithm described here calculates directly the height coordinate for the ~~models-model~~ pressure levels (i.e., the middle of a layer) and not for the pressure interfaces (layer boundaries) of the model as it is described in the IFS documentation. Throughout this document  $i_\varphi$  denotes the index for the longitude dimension,  $i_\lambda$  for the latitude and  $i$  for the vertical dimension.

*Algorithm. (Pressure grid)* Following the standard ECMWF procedure, the vertical pressure coordinate  $\mathbf{p}$  is set up in the following way: the surface pressure  $\mathbf{p}_s$  equals

$$\mathbf{p}_s(i_\varphi, i_\lambda) = e^{\text{lnsp}(i_\varphi, i_\lambda)},$$

and  $\bar{p}$  the pressure on the level boundaries is:

$$\mathbf{p}(i, i_\varphi, i_\lambda) = \mathbf{ap}(i) + \mathbf{bp}(i)\mathbf{p}_s(i_\varphi, i_\lambda).$$

The pressure  $\mathbf{p}$  is increasing in the vertical, i.e.  $\mathbf{p}(61, \cdot, \cdot)$  corresponds to the surface pressure  $\mathbf{p}_s$  and  $\mathbf{p}(1, \cdot, \cdot)$  is the pressure at the top of the atmosphere. The vertical pressure

coordinate on model levels then equals

$$\mathbf{p}_m(i, i_\varphi, i_\lambda) = \frac{1}{2} (\mathbf{p}(i+1, i_\varphi, i_\lambda) + \mathbf{p}(i, i_\varphi, i_\lambda)).$$

All profile data (for the MARS variables go3, co, ch4, . . . , except aern03) are defined on model pressure levels  $\mathbf{p}_m$ .

To construct a vertical height vector out of the pressure grid, the server calculates the molar mass of humid air.

*Algorithm. (Relative humidity)* Let  $M_X$  denote the molar mass of species  $X$ . By definition, the value  $\mathbf{q}(i, i_\varphi, i_\lambda)$  is the fraction  $\frac{M_{\text{H}_2\text{O}} n_{\text{H}_2\text{O}}}{M_a n_a}$  where  $n_{\text{H}_2\text{O}}$  represents the number of molecules water vapour in  $n_a$  molecules air ( $n_a$  is the number of molecules in the  $i$ th layer at height  $\mathbf{p}_m(i, i_\varphi, i_\lambda)$ ). Similarly,  $n_{\text{da}}$  denotes the number of particles dry air in  $n_a$  mol molecules air. Using the fact that  $n_a = n_{\text{da}} + n_{\text{H}_2\text{O}}$  and  $M_a n_a = M_{\text{da}} n_{\text{da}} + M_{\text{H}_2\text{O}} n_{\text{H}_2\text{O}}$ , it follows  $M_a n_a = M_{\text{da}} n_a + (M_{\text{H}_2\text{O}} - M_{\text{da}}) n_{\text{H}_2\text{O}}$ , or after division by  $M_a n_a$ :

$$1 = \frac{M_{\text{da}}}{M_a} + \left(1 - \frac{M_{\text{da}}}{M_{\text{H}_2\text{O}}}\right) \mathbf{q}(i, i_\varphi, i_\lambda).$$

In the next algorithm, we use the fraction  $M_{\text{da}}/M_a$  explicitly:

$$\frac{M_{\text{da}}}{M_a} = 1 + \left(\frac{R_{\text{H}_2\text{O}}}{R_{\text{da}}} - 1\right) \mathbf{q}(i, i_\varphi, i_\lambda),$$

with  $M_{\text{da}} = 28.960 \text{ g mol}^{-1}$ ,  $M_{\text{H}_2\text{O}} = 18.015 \text{ g mol}^{-1}$  and  $R_{\text{da}}, R_{\text{H}_2\text{O}}$  the gas constants for dry air and water vapour respectively. For unit conversion algorithms, the molar mass of humid air  $M_a$  ( $\text{g mol}^{-1}$ ) at the gridpoint  $(i, i_\varphi, i_\lambda)$  is used explicitly:

$$M_a(i, i_\varphi, i_\lambda) = \frac{M_{\text{da}} M_{\text{H}_2\text{O}}}{M_{\text{H}_2\text{O}} (1 - \mathbf{q}(i, i_\varphi, i_\lambda)) + \mathbf{q}(i, i_\varphi, i_\lambda) M_{\text{da}}}.$$

*Algorithm. (Height grid)* We use the MACC pressure and temperature to construct a height coordinate in the usual way: we start with  $\mathbf{z}_s$  and construct recursively a height

for the upper levels using the hydrostatic balance equation  $gM_a dz = -R\mathbf{T}d\ln(p)$  (see Andrews, 2010). The latter equation is rewritten using the virtual temperature  $gM_{da} dz = -R\mathbf{T}_\nu d\ln(p)$ , with

$$T_\nu = \frac{M_{da}}{M_a} T = T \left( 1 + \left( \frac{R_{H_2O}}{R_{da}} - 1 \right) q \right).$$

Earth acceleration  $g$  is approximated using WGS-84 (NIMA, 1984)  $g^{84}(\varphi, z)$ :

$$z(60, i_\varphi, i_\lambda) - z_s(i_\varphi, i_\lambda) = \frac{R_{da} \mathbf{T}_\nu(60, i_\varphi, i_\lambda)}{g^{84}(\varphi(i_\varphi), z_s(i_\varphi, i_\lambda))} \ln \frac{\mathbf{p}_s(i_\varphi, i_\lambda)}{\mathbf{p}_m(60, i_\varphi, i_\lambda)}.$$

and recursively for  $i = 59, \dots, 1$ :

$$z(i, i_\varphi, i_\lambda) - z(i+1, i_\varphi, i_\lambda) = \frac{R_{da} \bar{\mathbf{T}}_\nu(i, i_\varphi, i_\lambda)}{g^{84}(\varphi(i_\varphi), z(i+1, i_\varphi, i_\lambda))} \ln \frac{\mathbf{p}_m(i+1, i_\varphi, i_\lambda)}{\mathbf{p}_m(i, i_\varphi, i_\lambda)},$$

with  $\bar{\mathbf{T}}_\nu(i, i_\varphi, i_\lambda) = \frac{1}{2}(\mathbf{T}_\nu(i+1, i_\varphi, i_\lambda) + \mathbf{T}_\nu(i, i_\varphi, i_\lambda))$ . The validation server requires the knowledge of the thickness of the layers for re-gridding purposes. The heights of these boundaries of the layers do not match the MACC pressure grid  $\mathbf{p}$  boundaries. Because these boundaries will only be used for re-gridding purposes, we believe that the differences that this approach may imply, are negligible.

*Algorithm. (MACC boundaries)* The boundary height vector  $z_B$  at a chosen grid point  $(i_\varphi, i_\lambda)$ , consists of a lower boundary  $z_B(1, i)$  and an upper boundary  $z_B(2, i)$  for the  $i$ th layer. The layer boundaries are calculated as the midpoints between layers, i.e. for  $i = 1, \dots, 59$  (recall that  ~~$z$~~   $z$  denotes the decreasing MACC grid height vector  ~~$is$~~ )

descending):

$$z_B(1, i) = \frac{1}{2} (z(i) + z(i + 1))$$

$$z_B(2, i + 1) = z_B(1, i).$$

The outer boundaries are determined from:

$$z_B(2, 1) = z(1) + \frac{1}{2} |z(2) - z(1)|$$

$$z_B(1, 60) = z(60) - \frac{1}{2} |z(60) - z(59)|.$$

The server checks that the lowest boundary does not become negative and that the upper boundary does not exceed the top of atmosphere  $z_{\text{toa}} = 120$  km: if  $z_B(1, 60) < 0$  and  $z(60) \geq 0$ , the lowest boundary is set to  $z_B(1, 60) = 0$  and if  $z_B(2, 1) > z_{\text{toa}}$  and  $z(1) < z_{\text{toa}}$ , the upper boundary is set to  $z_B(2, 1) = z_{\text{toa}}$ .

*Remark.* The above boundary algorithm only requires the knowledge of the layer heights. Therefore, the same algorithm is used for generating boundaries on a NDACC data product, if the product does not provide ALTITUDE.BOUNDARIES.

*Remark.* In the validation server, the above algorithms, although described to be calculated on the full MACC grid, are actually performed on MACC data that is already horizontally interpolated to the site location (see Sect. 3.3). This reduces significantly the required computation time.

## 2.4 Unit conversions

The validation server will align the MACC data with the measurement data. This means that all model profile data is re-gridded to the measurement's vertical grid and **brought converted** to measurement's units. Typically the MACC profile data is given in mass mixing ratio (MMR,  $\text{kg kg}^{-1}$ ), and the measurement data in volume mixing ratio (VMR, ppv) or number density (ND,  $\text{mol m}^{-3}$ ).

*Algorithm. (Unit conversions)* ~~To fix thoughts~~As an example, assume an  $\Theta_3\text{-O}_3$  profile is given in MMR. To convert it to VMR, the profile is multiplied with the factor  $M_a/M_{O_3}$ . To convert VMR to ND:

$$O_3 [\text{mol m}^{-3}] = \frac{P_m}{RT} O_3 [\text{ppv}].$$

The partial column profile is derived from the ND profile, using the layer thickness  $\Delta z = z_B(2, \dots) - z_B(1, \dots)$ :

$$O_3 [\text{mol m}^{-2}] = \Delta z O_3 [\text{mol m}^{-3}].$$

The above formula with the layer thickness is also used to scale a profile of optical depths to an optical thickness profile.

### 3 Essential steps in a validation

#### 3.1 Re-gridding of profile data

Re-gridding of profile data is done with conservation of mass in mind (or total optical depth, in the case of aerosol data). ~~To fix thoughts~~As an example, assume that ~~an  $\Theta_3$~~ a target profile (in partial column units for ~~concentration profiles~~a concentration profile or optical depth for an aerosol extinction profile) is defined on a vertical height grid  $z^S$  with boundaries  $z_B^S$  (the superscript S is used to identify this grid as the *source* grid). To re-grid the profile to a new grid with layer heights  $z^E$  and boundaries  $z_B^E$  (E is *external*), we construct a transformation matrix  $\mathbf{D}$  that contains the fractions of how each external grid layer is covered by a source grid layer. The coefficients of the transformation matrix  $\mathbf{D}$  satisfy  $0 \leq \mathbf{D}(i, j) \leq 1$  where  $i$  runs over the dimension of  $z^E$  and  $j$  runs over the dimension of the source grid  $z^S$ . A situation is depicted in Fig. 2, where the external grid is coarser than the source grid (i.e. external layers overlap multiple source layers).

*Algorithm. (Layer height weighted re-gridding)* Assume the  $i$ th coarse grid layer overlaps with the  $j$ th source grid layer. Then the element in the  $i$ th row of  $\mathbf{D}$  corresponding to the  $j$ th

source layer is an interpolation factor:

$$\mathbf{D}(i, j) = \frac{\min(z_{\mathbf{B}}^{\mathbf{S}}(2, j), z_{\mathbf{B}}^{\mathbf{E}}(2, i)) - \max(z_{\mathbf{B}}^{\mathbf{S}}(1, j), z_{\mathbf{B}}^{\mathbf{E}}(1, i))}{z_{\mathbf{B}}^{\mathbf{S}}(2, j) - z_{\mathbf{B}}^{\mathbf{S}}(1, j)}.$$

If there is no overlap between  $i$ th external layer and  $j$ th source layer, the corresponding element in  $\mathbf{D}(i, j)$  equals 0. For the situation in Fig. 2, the  $i$ th row will take the following form (source layers  $(j+1)$  and  $(j-3)$  contain the lower/upper boundary of the  $i$ th external layer, resp.):

$$\mathbf{D}(i, \dots) = (0 \dots 0 \overset{(j-3)}{0.42} \ 1 \ 1 \ \overset{(j+1)}{0.87} \ 0 \dots 0).$$

The dark grey layers in Fig. 2 have integer coefficients in  $\mathbf{D}$  as they are completely covered by the  $i$ th row of  $z_{\mathbf{B}}^{\mathbf{E}}$ .

The sum of all rows in the transformation matrix is a vector with the dimension of the source grid that contains a coefficient of 1 for every source layer that is completely covered by an external layer. The re-gridded profile is obtained from matrix multiplication:  $\mathbf{O}_3^{\mathbf{R}} = \mathbf{D} \cdot \mathbf{O}_3 \mathbf{O}_3^{\mathbf{R}} = \mathbf{D} \cdot \mathbf{O}_3$ . A final step in the re-gridding process is to ~~substitute in the re-gridded profile~~ set profile values to void (or NaN) ~~values~~ for layers of the external grid that are not or only partly covered by the source grid.

### 3.2 Temporal co-location

Section 2.2 describes the construction of the MACC timely data sequence  $t_1^{\mathbf{M}}, t_2^{\mathbf{M}}, \dots$  The time between two subsequent MACC data validity times is constant and denoted by  $\Delta > 0$  (typically  $\Delta = 3$  h). Around each MACC time  $t^{\mathbf{M}}$  the server puts a pre-defined time window  $\delta$  of length  $0 < \delta \leq \Delta$  as indicated in Fig. 3 with a grey rectangle.

A NORS measurement at time  $t^{\mathbf{N}}$  will be used to validate a MACC data instance valid at time  $t^{\mathbf{M}}$  if  $|t^{\mathbf{N}} - t^{\mathbf{M}}| < \delta/2$ .

This choice implies that a single MACC data instance at time  $t^{\mathbf{M}}$  is validated against several measurements. But a single measurement will not validate multiple MACC data

instances. If a NDACC measurement cannot be related to a MACC time (for instance, in the situation of  $t_4^N$  in Fig. 3), the NDACC measurement will not be used in the validation.

This type of co-location is used for all NDACC products, however depending on the nature of the species and the availability of measurements, the window  $\delta$  may differ from one target species to another (see Table 4). E.g., due to the high diurnal variability of stratospheric  $\text{NO}_2$ , FTIR.NO2 and UVVIS.DOAS.ZENITH.NO2 measurements have  $\delta = \frac{1}{2}$  h.

### 3.3 Spatial co-location and smoothing

Not all measurement techniques measure the state of the atmosphere directly above the instrument's location. E.g., FTIR measurements measure direct sunlight, and the probed column of air varies with the local measurement time. A similar situation occurs for UVVIS.DOAS.ZENITH and UVVIS.DOAS.OFFAXIS measurements. The vertical profile that is extracted from the MACC model at the site's location, should take this off-location of the measured air mass into account.

The UVVIS template includes the latitude and longitude coordinate for the probed air mass at each height grid point (the LATITUDE and LONGITUDE GEOMS variables). For FTIR measurements, the server uses an off-line routine to calculate the LATITUDE and LONGITUDE GEOMS variables when not available in the NDACC file. Depending upon the availability of these variables, the server distinguishes two situations.

#### 3.3.1 LATITUDE and LONGITUDE not available: MWR and LIDAR

In this case a vertical profile is extracted at the site's location by means of a standard bilinear interpolation to get from the MACC latitude–longitude grid to a vertical profile at the instrument's latitude and longitude at all altitudes. The horizontally interpolated profile is re-gridded to the measurement vertical grid (i.e. the external grid is  $z_B^E = z_B^N$  and the source grid is  $z_B^S = z_B^M$ ).

### 3.3.2 LATITUDE and LONGITUDE available: FTIR and UVVIS.DOAS

If the horizontal coordinates of the probed airmass for each measurement layer are available, the co-located model profile is constructed ~~by gluing the MACC profile values interpolated horizontally to the airmass location in each per~~ measurement layer: i.e., the MACC re-gridded profile on-value for the  $i$ th measurement grid layer  $z^N$ , equals the value for layer  $i$  of the horizontally interpolated (towards  $(\varphi^N(i), \lambda^N(i))$ ) and consequently vertically re-gridded (towards  $z^N(i)$ ) MACC profile.

The values for the measurement grid that are not (or only partially) covered by the MACC grid are void.

### 3.3.3 Alignment of re-gridded model data

The re-gridding algorithm described in Sect. 3.1 acts on partial column profiles (or optical depth for aerosol). Further unit conversion is required since the measurement's averaging kernel typically acts on VMR profiles. To do this unit conversion, we use temperature  $\mathbf{T}^N$  and pressure  $\mathbf{p}^N$  profiles provided along with the measurement data, i.e. the partial column of air for each layer in the measurement grid equals

$$\alpha^N = \frac{\mathbf{p}^N(i)}{R\mathbf{T}^N(i)} \Delta z^N.$$

This implies that the validation is based on the model's partial column/optical depth profile and that further manipulation on both model and measurement data is done using equal conversion factors.

Aerosol optical depth profiles require one further manipulation. Typically the measurement's optical depth profile is measured at a specific wavelength (say  $\lambda^N = 477$  nm) which does not coincide with the wavelength of the model optical depth profile at 532 nm.

*Algorithm. (Optical depth wavelength scaling)* To convert the model optical depth profile to the measurement's wavelength, we use the Ångström exponent  $\alpha$  (Ångström, 1929), calculated from the MACC array of optical depths aerng03 (use the pair of total OD's at 532 nm

and at the wavelength closest to the measurement  $\lambda^N = 477$  nm, in this case 469 nm):

$$\alpha = - \left( \ln \frac{OD_{469}}{OD_{532}} \right) / \left( \ln \frac{469}{532} \right).$$

Under the assumption that the Ångström coefficient is height independent and the measurement's wavelength  $\lambda^N$  falls within the validity range of the above estimate for  $\alpha$ , the re-gridded model optical depth profile is then scaled with the factor

$$\left( \frac{\lambda^N}{532} \right)^{-\alpha}.$$

### 3.4 Application of the measurement's averaging kernel

In the following, it is understood that the re-gridded model profile, e.g.  $O_3^{M,R} O_3^{M,R}$ , has been brought converted to the units of the measurement's averaging kernel. Smoothing of model profile data ensures that the model profile and measurement profile coincide on height levels outside the measurement's sensitivity range (see also below), i.e. there can be no biases between model and measurement outside the sensitivity range. Furthermore, according to the GEOMS template files (FTIR, UVVIS, see Retscher et al., 2011), the reported measurement uncertainties do not include the smoothing uncertainty which makes the smoothing of model data mandatory in order to have a complete uncertainty budget for the validation statistics. Examples of typical AVK's for each measurement technique can be found in the NORS documentation (De Mazière et al., 2013; Richter et al., 2013).

*Algorithm. (Smoothing of model data)* The smoothed model profile is obtained by the standard smoothing formula on the re-gridded model profile  $O_3^{M,R}$  (see Rodgers, 2000, matrix multiplication is used):

$$O_3^{M,S} = O_3^{N,ap} + A^N \cdot (O_3^{M,R} - O_3^{N,ap}),$$

where the measurements apriori profile  $O_3^{N,ap}$  is used and where the (possibly) void (NaN) re-gridded model profile values are well taken care of: substitute a zero for these void layers

in the difference profile  $O_3^{M,R} - O_3^{N,ap}$  and as a final step, after applying the above smoothing formula, the values in the resulting smoothed model profile  $O_3^{M,S}$  corresponding to the initial void layers of the re-gridded profile are replaced by NaN. In this way, the AVK is applied to the model target profile, even if the model does *not* provide data on the entire AVK grid. The above method ensures that these outside layers do not contribute in the smoothing operation. [Figures 4 and 5 show an example of the above described regridding and smoothing algorithms applied to a stratospheric O<sub>3</sub> FTIR profile measurement.](#)

In some cases, the measurement data only provides a retrieved total column (e.g. the UVVIS.DOAS.ZENITH measurements). In that case the above formula is adapted such that the column averaging kernel is used. [In this case  \$A^N\$  is a transformation \(with the shape of a vector\) acting on profiles of partial columns, and the result is a scalar total column value:](#)

$$O_3^{M,S,tc} = \sum_{\text{layers}} O_3^{N,ap} + A^N \cdot (O_3^{M,R} - O_3^{N,ap}),$$

The resulting smoothed model total column/optical depth is void if the measurement grid outranges the model grid. [The model grid height typically reaches 65km. For the FTIR, LIDAR and MWR data, the measurements are reported on grids up to 100km.](#)

## 4 Representation of validation results

In order to get statistics on the validation results (see Huijnen and Eskes, 2012), all individual (per measurement) results are brought to a chosen fixed vertical grid. Such a common grid can be a single layer grid to get a partial column or a true vertical grid to calculate a mean (difference) profile. Another possibility implemented by the server, is the re-gridding towards a subgrid of the measurement grid where each partial column overlaps layers whose cumulative sum of the degrees of freedom (the diagonal elements of the AVK) is approximately one.

As a general rule, the representation vertical grid is either a chosen fixed grid from a measurement or from the MACC model. This choice is such that the coarsest grid is chosen for representation purposes: for LIDAR, the vertical grid of the measurement may be finer than the model grid and in this case the representation grid is chosen to be a fixed model grid. Re-gridding towards this fixed representation grid  $z^E$  is done with the algorithm described before (i.e. applied to profile data of partial columns/optical depths).

#### 4.1 Uncertainty propagation

NDACC uncertainties can be reported as standard deviation values  $\sigma$ , or covariances  $\mathbf{S}$ . Uncertainty propagation requires the knowledge of covariance matrices. If either the systematic or random uncertainty matrix contains fill values, the matrix is filled with NaN's. NaN uncertainties are not shown in the reports.

The measurement uncertainty covariance matrix  $\mathbf{S}^N$  (random or systematic) is propagated to this chosen representation grid using the transformation matrix  $\mathbf{D}$  described in Sect. 3.1 with  $z_B^N$  as source grid and the representation grid as external grid (matrix multiplication):

$$\mathbf{S}^E = \mathbf{D} \cdot \mathbf{S}^N \cdot \mathbf{D}^T.$$

The covariance matrices in the formula above are in partial column units (e.g.  $\text{mol m}^{-2}$ ). To convert a covariance matrix in VMR unit towards-to partial column units, the partial+partial column profile of air  $a^N$  is used: the  $(i, j)$ th component of the covariance in VMR units is multiplied with

$$a^N(i)a^N(j).$$

To transform a covariance matrix from optical thickness units towards optical depth, the same formula is used where  $a^N$  is replaced by the vector of layer thickness  $\Delta z^N$ .

## 4.2 Sensitivity and partial columns

Depending on the measurement technique, site and species, the measurement sensitivity may differ. Table 4 contains the lower and upper boundaries used by the server as the boundaries for the partial columns and profile data. The lower boundary is site dependent and adapted with the instrument's height. These height boundaries are typical boundaries in which the measurement has sensitivity and are derived by looking at representative AVK matrices (see De Mazière et al., 2013; Richter et al., 2013, for more information on the different sensitivity ranges).

## 4.3 Time averaging of column data

For long time periods, it is required to average data in time to improve the readability or visibility of the validation statistics. ~~To fix thoughts~~ For example, assume that the monthly mean of a time series of O<sub>3</sub> partial columns is calculated. Due to the nature of systematic and random uncertainties (see Taylor, 1997; JCGM/WG2, 2008), the random uncertainty  $\sigma^r$  of the monthly mean decreases at rate  $1/\sqrt{n}$ , with  $n$  the sample size ( $i$  runs over all measurements in a month):

$$\sigma = \sqrt{\frac{1}{n^2} \sum_{i=1}^n (\sigma_i^r)^2}.$$

This differs from the systematic uncertainty on the monthly mean:

$$\sigma = \frac{1}{n} \sum_{i=1}^n \sigma_i^s.$$

### Code availability

The code used by the validation server is built on the existing GECA Toolset (Generic Environment for Calibration and Validation Activities), developed within the ESA GECA project. It

is foreseen that the GECA Toolset will become publicly available as a separate component in BEAT (Basic Envisat Atmospheric Toolbox) (<http://www.stcorp.nl/beat/>, an ESA funded project). Until then, the GECA code is available upon request [with to](#) S. Niemeijer.

## 5 Conclusions

This paper documents in detail generic tools for comparison between data sets related to atmospheric composition, focusing on ground-based remote-sensing data vs. gridded model data. Although comparisons between data sets from two different sources have been performed [since for](#) many years in the atmospheric scientific community and the basic concepts of co-location and comparisons are known, this is the first time that generic tools have been developed, fully documented and implemented successfully. It is the first time also that the effective location of the remotely-sensed airmasses is taken into account for UVVIS and FTIR measurements, at least in an approximate way. Differences in vertical resolution of the data are also accounted for. The tools comply with the QA4EO guidelines ([see QA4EO, 2010](#)).

The automatic application of the tools requires that the reference data formats comply with the GEOMS generic guidelines and specific templates per data type. During the development of the tools, it appeared that some GEOMS guidelines needed more precise specifications. Also, many inconsistencies in the data files have shown up and were corrected.

The addition of a new data type to the set already covered, is an easy task, because the tools consist of a succession of basic algorithms, of which many are identical for different data types. The tools can also be extended easily to comparisons between the ground-based remote-sensing data and satellite data (instead of gridded model data), by adding different co-location algorithms. Such algorithms have been developed previously in the context of the Generic Environment for Calibration/Validation Analysis (the GECA project, funded by ESA) and will be integrated in the toolset in the near future.

The NORS validation server is operational and example validation reports can be viewed online at <http://nors-server.aeronomie.be>.

*Acknowledgements.* This work was developed under a EU FP7 project, NORS under grant agreement no. 284421. The NORS project relies on the data [publicly](#) available in NDACC (<http://www.ndacc.org>). The data providers for FTIR, MWR, LIDAR and UVVIS measurements in the NDACC database are acknowledged. [We would like to thank the anonymous referees for their constructive comments.](#)

## References

- Andrews, D.: An Introduction to Atmospheric Physics, Cambridge University Press, Cambridge, 2010.
- Ångström, A.: On the atmospheric transmission of sun radiation and on dust in the air, Geogr. Ann. A, 11, 156–166, 1929.
- De Mazière, M., Richter, A., Pastel, M., Hendrick, F., Langerock, B., and Hocke, K.: NORS Data User Guide, available at: [http://nors.aeronomie.be/projectdir/PDF/NORS\\_D4.2\\_DUG.pdf](http://nors.aeronomie.be/projectdir/PDF/NORS_D4.2_DUG.pdf) (last access: 28 September 2014), 2013.
- Huijnen, V. and Eskes, H.: Skill scores and evaluation methodology for the MACC II project. MACC-II VAL D\_85.2, available at: <http://www.gmes-atmosphere.eu/documents/maccii/deliverables/val/> (last access: 28 September 2014), 2012.
- JCGM/WG2: Evaluation of Measurement Data – Guide to the Expression of Uncertainty in Measurement, Working Group 2 of the Joint Committee for Guides in Metrology, [http://www.bipm.org/utls/common/documents/jcgm/JCGM\\_100\\_2008\\_E.pdf](http://www.bipm.org/utls/common/documents/jcgm/JCGM_100_2008_E.pdf) (last access: 28 September 2014), 2008.
- Lambert, J.-C.: Validation protocol, available at: <http://www.gmes-atmosphere.eu/documents/deliverables/man> (last access: 28 September 2014), 2010.
- NIMA: NIMA Technical Report TR8350.2, Department of Defense World Geodetic System 1984, Its Definition and Relationships With Local Geodetic Systems, Geodesy and Geophysics Department, National Imagery and Mapping Agency, 3rd edn., available at: <http://earth-info.nga.mil/GandG/publications/tr8350.2/wgs84fin.pdf> (last access: 28 September 2014), 1984.
- QA4EO: A Quality Assurance Framework for Earth Observation: Principles, available at: [http://qa4eo.org/docs/QA4EO\\_Principles\\_v4.0.pdf](http://qa4eo.org/docs/QA4EO_Principles_v4.0.pdf) (last access: 28 September 2014), 2010.

- Retscher, C., De Mazière, M., Meijer, Y., Vik, A., Boyd, I., Niemeijer, S., Koopman, R., and Bojkov, B.: The Generic Earth Observation Metadata Standard (GEOMS), available at: <http://avdc.gsfc.nasa.gov/index.php?site=1178067684> (last access: 28 September 2014), 2011.
- Richter, A., Godin, S., Gomez, L., Hendrick, F., Hocke, K., Langerock, B., van Roozendaal, M., and Wagner, T.: NORS Spatial Representativeness of NORS observations, available at: [http://nors.aeronomie.be/projectdir/PDF/D4.4\\_NORS\\_SR.pdf](http://nors.aeronomie.be/projectdir/PDF/D4.4_NORS_SR.pdf) (last access: 28 September 2014), 2013.
- Rodgers, C. D.: Inverse Methods for Atmospheric Sounding: Theory and Practice (Series on Atmospheric Oceanic and Planetary Physics), World Scientific Publishing Company, Singapore, 2000.
- Taylor, J.: An Introduction to Error Analysis: The Study of Uncertainties in Physical Measurements, Physics – Chemistry – Engineering, University Science Books, California, 1997.

**Table 1.** List of NORS pilot instruments, stations, and target parameters. UVVIS.DOAS.ZENITH (OFFAXIS) stands for UV-visible (MAX)DOAS spectrometer, FTIR for Fourier transform infrared spectrometer and MWR for microwave radiometer.

Instruments	Locations	Targets
FTIR	Reunion, Izaña, Jungfraujoch, Ny Ålesund	O <sub>3</sub> , CO, CH <sub>4</sub>
MWR	Bern, Ny Ålesund	O <sub>3</sub>
LIDAR	Reunion, OHP	O <sub>3</sub>
UVVIS.DOAS.ZENITH	OHP, Reunion, Jungfraujoch	O <sub>3</sub> , NO <sub>2</sub>
UVVIS.DOAS.OFFAXIS	Xianghe, Jungfraujoch, Izaña, Ny Ålesund	NO <sub>2</sub> , HCHO, Aerosol extinction

**Table 2.** List of GEOMS variable names common for all GEOMS templates (dimensions are indicative for 100 measurements on a grid with 47 layers).

variable	dimension	unit	description	notation
DATETIME	(100)	s	measurement time	$t$
LATITUDE.INSTRUMENT	(1)	rad	latitude of the instrument	$\phi$
LONGITUDE.INSTRUMENT	(1)	rad	longitude of the instrument	$\lambda$
ALTITUDE.INSTRUMENT	(1)	m	altitude of the instrument	$z^{\text{inst}}$
ALTITUDE	(47)	m	altitude grid (always descending)	$z$
TEMPERATURE_INDEPENDENT	(100, 47)	K	temperature profile	$\mathbf{T}$
PRESSURE_INDEPENDENT	(100, 47)	Pa	pressure profile	$\mathbf{P}$
LATITUDE	(100, 47)	rad	latitude of the location of probed air mass at each altitude (optional)	$\lambda$
LONGITUDE	(100, 47)	rad	longitude of the location of probed air mass at each altitude (optional)	$\phi$
ALTITUDE.BOUNDARIES	(2, 47)	m	boundaries of vertical grid layers (optional)	$z_B$

**Table 3.** MACC variables specifications. MMR = mass mixing ratio. The dimensions are indicative for a model with IFS resolution T255N128.

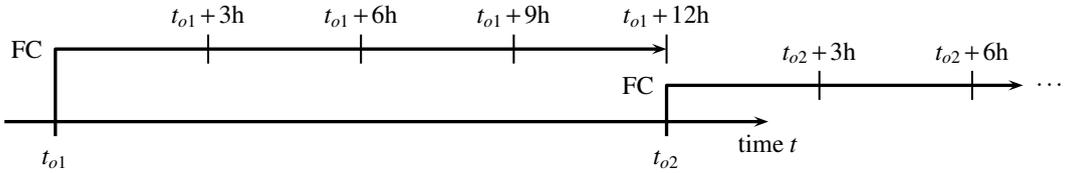
species	dimension	unit	description	notation
Insp	(256, 512)	ln Pa	logarithmic surface pressure (depends on latitude and longitude)	Insp
–	(61)	Pa	array ap contains the translation terms for the construction of the pressure grid	<b>ap</b>
–	(61)	–	array bp contains the scaling factors in the construction of the pressure grid	<b>bp</b>
–	(512)	rad	the array of longitudes ( $i_\lambda$ denotes the index for this array)	$\lambda$
–	(256)	rad	the array of latitudes ( $i_\varphi$ denotes the index)	$\varphi$
$t$	(60, 256, 512)	K	temperature	<b>T</b>
$z$	(256, 512)	m	geopotential height of the surface (corresponds to Insp)	<b>z<sub>s</sub></b>
$q$	(60, 256, 512)	kg kg <sup>-1</sup>	water vapour MMR (w.r.t. moist air)	<b>q</b>
go3	(60, 256, 512)	kg kg <sup>-1</sup>	MACC ozone MMR (w.r.t. moist air)	O <sub>3</sub>
hcho	(60, 256, 512)	kg kg <sup>-1</sup>	Formaldehyde MMR (w.r.t. moist air)	HCHO
co	(60, 256, 512)	kg kg <sup>-1</sup>	CO MMR (w.r.t. moist air)	CO
aern04	(60, 256, 512)	–	Profile of aerosol optical depths at 532 nm	$\tau_{532}$
aern03	(60, 256, 512)	–	Aerosol total optical depth at different wavelengths. Each level contains the total optical depth at different wavelengths: e.g. level 1 at 550 nm, level 2 at 340 nm, level 3 at 355 nm, level 4 at 380 nm, . . . level 7 at 469 nm, level 9 532 nm	<b>OD</b>
ch4	(60, 256, 512)	kg kg <sup>-1</sup>	CH <sub>4</sub> MMR (w.r.t. moist air)	CH <sub>4</sub>
no2	(60, 256, 512)	kg kg <sup>-1</sup>	NO <sub>2</sub> MMR (w.r.t. moist air)	NO <sub>2</sub>

**Table 4.** List of default  $z^{\min}$ ,  $z^{\max}$  values for NORS target species, based upon the measurement's sensitivity range and upper boundaries of MACC models, and of time windows for temporal co-location, based on species temporal variability.  $\Delta$  is the time between two subsequent MACC data validity times – see text (target values may contain wildcards if applicable to multiple targets).

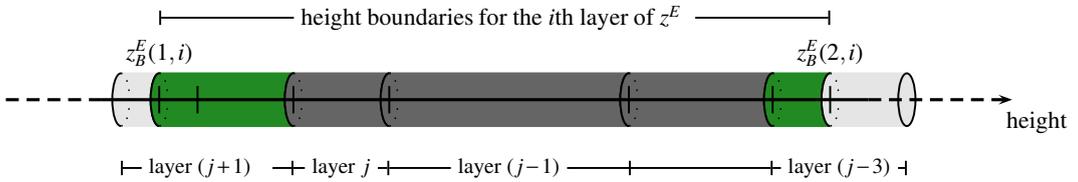
target	$[z^{\min}, z^{\max}]$ , $\delta$	target	$[z^{\min}, z^{\max}]$ , $\delta$	target	$[z^{\min}, z^{\max}]$ , $\delta$
FTIR.O3	[0 km, 60 km], $\Delta$	FTIR.CO	[0 km, 20 km], $\Delta$	UVVIS.ZENITH.NO2	[0 km, 60 km], $\frac{1}{2}$ h
FTIR.NO2	[10 km, 50 km], $\frac{1}{2}$ h	MWR.O3	[25 km, 60 km], 1 h	UVVIS.ZENITH.O3	[0 km, 60 km], $\Delta$
FTIR.CH4	[0 km, 50 km], $\Delta$	LIDAR.O3	[15 km, 45 km], $\Delta$	UVVIS.DOAS.OFFAXIS.*	[0 km, 8 km], 1 h

**Table A1.** List of GEOMS variable names, notations and corresponding GEOMS templates.

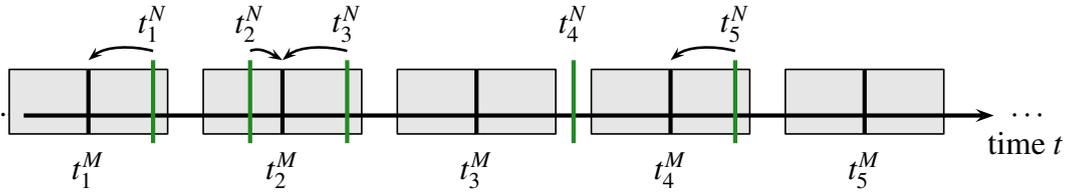
GEOMS template version	GEOMS name	notation
FTIR-002	*.MIXING.RATIO.VOLUME_ABSORPTION.SOLAR	$O_3, CO, CH_4, \dots$
	*.MIXING.RATIO.VOLUME_ABSORPTION.SOLAR_UNCERTAINTY.RANDOM.COVARIANCE	$S^r$
	*.MIXING.RATIO.VOLUME_ABSORPTION.SOLAR_UNCERTAINTY.SYSTEMATIC.COVARIANCE	$S^s$
	*.MIXING.RATIO.VOLUME_ABSORPTION.SOLAR_APRIORI	$O_3^{ap}, \dots$
	*.MIXING.RATIO.VOLUME_ABSORPTION.SOLAR_AVK	$A$
MWR-003	O3.MIXING.RATIO.VOLUME_EMISSION	$O_3$
	O3.MIXING.RATIO.VOLUME_EMISSION_UNCERTAINTY.RANDOM.STANDARD	$\sigma^r$
	O3.MIXING.RATIO.VOLUME_EMISSION_UNCERTAINTY.SYSTEMATIC.STANDARD	$\sigma^s$
	O3.MIXING.RATIO.VOLUME_EMISSION_APRIORI	$O_3^{ap}$
	O3.MIXING.RATIO.VOLUME_EMISSION_AVK	$A$
LIDAR-003	O3.NUMBER.DENSITY_ABSORPTION_DIFFERENTIAL	$O_3$
	O3.NUMBER.DENSITY_ABSORPTION_DIFFERENTIAL_UNCERTAINTY.ORIGINATOR	$\sigma^r$
UVVIS.DOAS.ZENITH-006	*.COLUMN.STRATOSPHERIC_SCATTER.SOLAR.ZENITH	$O_3^c, NO_2^c$
	*.COLUMN.STRATOSPHERIC_SCATTER.SOLAR.ZENITH_UNCERTAINTY.RANDOM.STANDARD	$\sigma_3^r, \sigma_2^r$
	*.COLUMN.STRATOSPHERIC_SCATTER.SOLAR.ZENITH_UNCERTAINTY.SYSTEMATIC.STANDARD	$\sigma_3^s, \sigma_2^s$
	*.MIXING.RATIO.VOLUME_SCATTER.SOLAR.ZENITH_APRIORI	$O_3^{ap}, \dots$
	*.COLUMN.STRATOSPHERIC_SCATTER.SOLAR.ZENITH_AVK	$A$
UVVIS.DOAS.OFFAXIS-006	*.MIXING.RATIO.VOLUME_SCATTER.SOLAR.OFFAXIS	$NO_2, HCHO$
	*.MIXING.RATIO.VOLUME_SCATTER.SOLAR.OFFAXIS_UNCERTAINTY.RANDOM.COVARIANCE	$S^r$
	*.MIXING.RATIO.VOLUME_SCATTER.SOLAR.OFFAXIS_UNCERTAINTY.SYSTEMATIC.COVARIANCE	$S^s$
	*.MIXING.RATIO.VOLUME_SCATTER.SOLAR.OFFAXIS_APRIORI	$NO_2^{ap}, \dots$
	*.MIXING.RATIO.VOLUME_SCATTER.SOLAR.OFFAXIS_AVK	$A$
	AEROSOL.EXTINCTION.COEFFICIENT_SCATTER.SOLAR.OFFAXIS	$\tau$
	WAVELENGTH	$\lambda$
	AEROSOL.EXTINCTION.COEFFICIENT_SCATTER.SOLAR.OFFAXIS_UNCERTAINTY.RANDOM.COVARIANCE	$S^r$
	AEROSOL.EXTINCTION.COEFFICIENT_SCATTER.SOLAR.OFFAXIS_UNCERTAINTY.SYSTEMATIC.COVARIANCE	$S^s$
	AEROSOL.EXTINCTION.COEFFICIENT_SCATTER.SOLAR.OFFAXIS_APRIORI	$\tau^{ap}$
	AEROSOL.EXTINCTION.COEFFICIENT_SCATTER.SOLAR.OFFAXIS_AVK	$A$



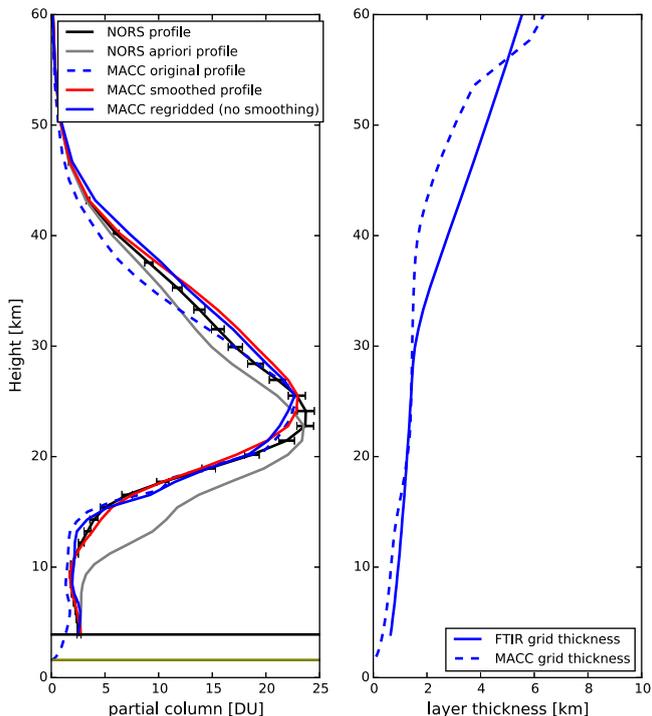
**Figure 1.** Construction of MACC data times for type FC (forecast).



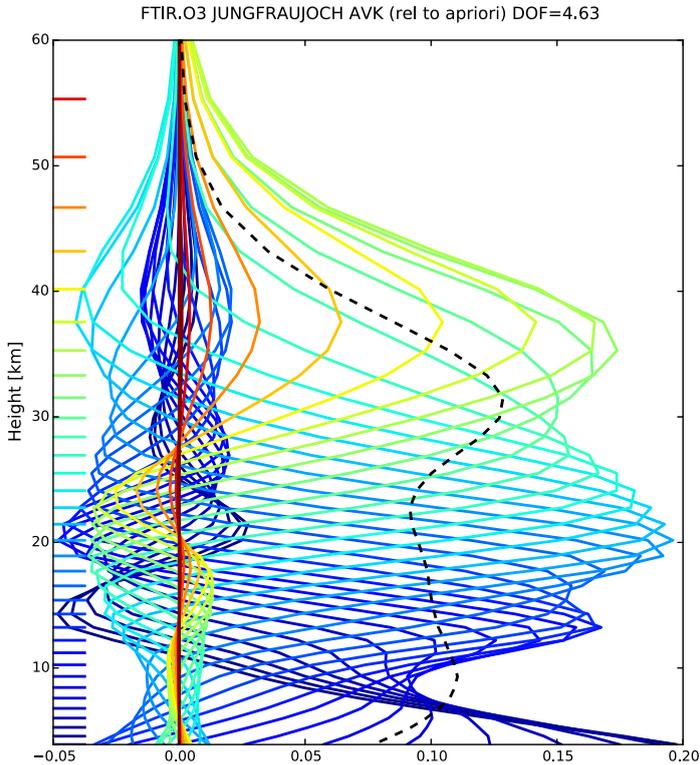
**Figure 2.** Interpolation factors for grid layers: green source grid layers are only partially overlapped by the  $i$ th destination layer.



**Figure 3.** Temporal co-location of MACC (model, black) and NORS (measurement, green) data.



**Figure 4.** Example of a model O<sub>3</sub> partial column profile (dashed blue) and FTIR measured profile (black). The model profile is first regrided to the measurement grid (blue) and smoothed (red) at Jungfraujoch. The lower two horizontal lines show the lowest layer height of the model (lowest) and site (highest). Because a partial column profile depends linearly on the layer thickness of the grid, the right plot shows the layer thickness profile for model and measurement grids. Above 40km the measurement (black), apriori (gray) and smoothed model (red) profiles coincide because the sensitivity of the measurement decreases (see Fig. 5)



**Figure 5.** Example of a  $\text{O}_3$  FTIR measurement averaging kernel. The plot shows the rows of the AVK matrix, color coded according to the height that corresponds to each row. The matrix elements of an AVK are without unit (the depicted AVK acts on  $\text{O}_3$  profiles relative to the a priori profile). The black dashed line is the sensitivity (the sum of each row): for rows with zero sensitivity the smoothing formula returns the a priori.