Geosci. Model Dev. Discuss., 7, 8151–8178, 2014
www.geosci-model-dev-discuss.net/7/8151/2014/
doi:10.5194/gmdd-7-8151-2014
© Author(s) 2014. CC Attribution 3.0 License.

Geoscientific Model Development Discussions
Open Access

This discussion paper is/has been under review for the journal Geoscientific Model Development (GMD). Please refer to the corresponding final paper in GMD if available.

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

B. Langerock[1], M. De Mazière[1], F. Hendrick[1], C. Vigouroux[1], F. Desmet[1], B. Dils[1], and S. Niemeijer[2]

[1]Belgian institute for Space Aeronomy, Ringlaan 3, 1180 Ukkel, Belgium
[2]S&T, Olof Palmestraat 14, 2616 LR Delft, the Netherlands

Received: 29 September 2014 – Accepted: 6 October 2014 – Published: 26 November 2014

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

Published by Copernicus Publications on behalf of the European Geosciences Union.

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.

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 Atmospheric 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),

GMDD
7, 8151–8178, 2014

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 with the NORS/NDACC target species. Duplication, however, is not an issue because MACC-II,III VAL is mostly using in situ surface 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). 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 co-location and smoothing algorithms for model vertical profile data. Section 4

GMDD
7, 8151–8178, 2014

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, all multiplications are scalar, matrix multiplication is denoted by a central dot ·. Most of the notations used in the algorithms can be found in Tables 2, 3 and A1. For example, the profile array $\mathbf{O}_3^{\mathrm{M}}$ 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. $\mathbf{O}_3^{\mathrm{M}}$ is the MACC model profile and $\mathbf{O}_3^{\mathrm{N}}$ 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, or to the corresponding NDACC database station directory if fully consolidated. For each measurement technique the NDACC data format has to be compliant with a predefined 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).

GMDD
7, 8151–8178, 2014

2.2 MACC data

At present, the validation server validates in operational mode the following forecast runs: NRT o-suite, the experiment with TM-5 chemistry and the experiment with MOZART chemistry (for further details see https://www.gmes-atmosphere.eu/oper_info/nrt_info_for_users). 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[1] 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 notational 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 $\mathbf{z}_s$ [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 several resolutions available. The server will automatically attach to a given model the geopotential field that matches the native IFS resolution of the model.

[1] An IFS resolution is denoted by TxNy: 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).

GMDD
7, 8151–8178, 2014

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 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_φ denotes the index for the longitude dimension, i_λ for the latitude and i for the vertical dimension.

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

$$\mathbf{p}_\mathrm{s}(i_\varphi, i_\lambda) = e^{\mathbf{lnsp}(i_\varphi, i_\lambda)},$$

and, the pressure on the level boundaries is:

$$\mathbf{p}(i, i_\varphi, i_\lambda) = \boldsymbol{ap}(i) + \boldsymbol{bp}(i)\mathbf{p}_\mathrm{s}(i_\varphi, i_\lambda).$$

The pressure **p** is increasing in the vertical, i.e. $\mathbf{p}(61,\cdot,\cdot)$ corresponds to the surface pressure $\mathbf{p}_\mathrm{s}$ and $\mathbf{p}(1,\cdot,\cdot)$ is the pressure at the top of the atmosphere. The vertical pressure coordinate on model levels then equals

$$\mathbf{p}_\mathrm{m}(i, i_\varphi, i_\lambda) = \frac{1}{2}\big(\mathbf{p}(i+1, i_\varphi, i_\lambda) + \mathbf{p}(i, i_\varphi, i_\lambda)\big).$$

All profile data (for the MARS variables go3, co, ch4, . . . , except aergn03) are defined on model pressure levels $\mathbf{p}_\mathrm{m}$.

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 $\frac{M_{\mathrm{H_2O}} n_{\mathrm{H_2O}}}{M_\mathrm{a} n_\mathrm{a}}$ where $n_{\mathrm{H_2O}}$ represents the number of molecules water vapour in n_a molecules air (n_a is the number of molecules in the

GMDD
7, 8151–8178, 2014

ith layer at height $\mathbf{p}_m(i, i_\varphi, i_\lambda)$). Similarly, n_{da} denotes the number of particles dry air in n_a mol molecules air. Using the fact that $n_a = n_{da} + n_{H_2O}$ and $M_a n_a = M_{da} n_{da} + M_{H_2O} n_{H_2O}$, it follows $M_a n_a = M_{da} n_a + (M_{H_2O} - M_{da}) n_{H_2O}$, or after division by $M_a n_a$:

$$1 = \frac{M_{da}}{M_a} + \left(1 - \frac{M_{da}}{M_{H_2O}}\right) \mathbf{q}(i, i_\varphi, i_\lambda).$$

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

$$\frac{M_{da}}{M_a} = 1 + \left(\frac{R_{H_2O}}{R_{da}} - 1\right) \mathbf{q}(i, i_\varphi, i_\lambda),$$

with $M_{da} = 28.960\,\mathrm{g\,mol^{-1}}$, $M_{H_2O} = 18.015\,\mathrm{g\,mol^{-1}}$ and R_{da}, R_{H_2O} the gas constants for dry air and water vapour respectively. For unit conversion algorithms, the molar mass of humid air $M_a\,(\mathrm{g\,mol^{-1}})$ at the gridpoint $(i, i_\varphi, i_\lambda)$ is used explicitly:

$$M_a(i, i_\varphi, i_\lambda) = \frac{M_{da} M_{H_2O}}{M_{H_2O}(1 - \mathbf{q}(i, i_\varphi, i_\lambda)) + \mathbf{q}(i, i_\varphi, i_\lambda) M_{da}}.$$

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 $g M_a \mathrm{d}z = -R\mathbf{T}\mathrm{d}\ln(p)$ (see Andrews, 2010). The latter equation is rewritten using the virtual temperature $g M_{da} \mathrm{d}z = -R\mathbf{T}_v \mathrm{d}\ln(p)$, with

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

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

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

GMDD
7, 8151–8178, 2014

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

$$z(i, i_\varphi, i_\lambda) - z(i+1, i_\varphi, i_\lambda) = \frac{R_{\mathrm{da}} \overline{\mathbf{T}}_v(i, i_\varphi, i_\lambda)}{\boldsymbol{g}^{84}(\varphi(i_\varphi), z(i+1, i_\varphi, i_\lambda))} \ln \frac{\mathbf{p}_{\mathrm{m}}(i+1, i_\varphi, i_\lambda)}{\mathbf{p}_{\mathrm{m}}(i, i_\varphi, i_\lambda)},$$

with $\overline{\mathbf{T}}_v(i, i_\varphi, i_\lambda) = \frac{1}{2}(\mathbf{T}_v(i+1, i_\varphi, i_\lambda) + \mathbf{T}_v(i, i_\varphi, i_\lambda))$. 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 **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_φ, i_λ), consists of a lower boundary $z_{\mathrm{B}}(1, i)$ and an upper boundary $z_{\mathrm{B}}(2, i)$ for the ith layer. The layer boundaries are calculated as the midpoints between layers, i.e. for $i = 1, \ldots, 59$ (recall that z is the MACC grid height vector is descending):

$$z_{\mathrm{B}}(1, i) = \frac{1}{2}(z(i) + z(i+1))$$
$$z_{\mathrm{B}}(2, i+1) = z_{\mathrm{B}}(1, i).$$

The outer boundaries are determined from:

$$z_{\mathrm{B}}(2, 1) = z(1) + \frac{1}{2}|z(2) - z(1)|$$
$$z_{\mathrm{B}}(1, 60) = z(60) - \frac{1}{2}|z(60) - z(59)|.$$

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

GMDD
7, 8151–8178, 2014

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 to measurement's units. Typically the MACC profile data is given in mass mixing ratio (MMR, $\mathrm{kg\,kg^{-1}}$), and the measurement data in volume mixing ratio (VMR, ppv) or number density (ND, $\mathrm{mol\,m^{-3}}$).

Algorithm. (Unit conversions) To fix thoughts, assume an $\mathbf{O_3}$ 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:

$$\mathbf{O_3}\left[\mathrm{mol\,m^{-3}}\right] = \frac{\mathbf{p}_m}{R\mathbf{T}}\mathbf{O_3}[\mathrm{ppv}].$$

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

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

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

GMDD
7, 8151–8178, 2014

3 Essential steps in a validation

3.1 Re-gridding of profile data

Re-gridding of profile data is done with conservation of mass (or total optical depth, in the case of aerosol data). To fix thoughts, assume that an $\mathbf{O}_3$ profile (in partial column units for concentration profiles or optical depth for aerosol extinction profile) is defined on a vertical height grid z^{S} with boundaries $z_{\mathrm{B}}^{\mathrm{S}}$ (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 $z_{\mathrm{B}}^{\mathrm{E}}$ (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 $0 \le \mathbf{D}(i,j) \le 1$ 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 ith coarse grid layer overlaps with the jth source grid layer. Then the element in the ith row of $\mathbf{D}$ corresponding to the jth source layer is an interpolation factor:

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

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

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

GMDD
7, 8151–8178, 2014

The dark grey layers in Fig. 2 have integer coefficients in **D** as they are completely covered by the ith row of z_B^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: $\mathbf{O}_3^R = \mathbf{D} \cdot \mathbf{O}_3$. A final step in the re-gridding process is to substitute in the re-gridded profile 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 $t_1^M, t_2^M, \ldots$. The time between two subsequent MACC data validity times is constant and denoted by $\Delta > 0$ (typically $\Delta = 3\,\mathrm{h}$). Around each MACC time t^M the server puts a pre-defined time window δ of length $0 < \delta \leq \Delta$ as indicated in Fig. 3 with a grey rectangle.

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

This choice implies that a single MACC data instance at time t^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 δ may differ from one target species to another (see Table 4). E.g., due to the high diurnal variability of stratospheric NO_2, FTIR.NO2 and UVVIS.DOAS.ZENITH.NO2 measurements have $\delta = \frac{1}{2}\,\mathrm{h}$.

GMDD
7, 8151–8178, 2014

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 airmass into account.

The UVVIS template includes the latitude and longitude coordinate for the probed airmass 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 measurement layer: i.e., the MACC re-gridded profile on the ith measurement grid layer z^N, equals the value of the horizontally interpolated (towards $(\varphi^N(i), \lambda^N(i))$) and consequently vertically re-gridded (towards $z^N(i)$) MACC profile.

GMDD
7, 8151–8178, 2014

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 $\mathbf{T}^{\mathrm{N}}$ and pressure $\mathbf{p}^{\mathrm{N}}$ profiles provided along with the measurement data, i.e. the partial column of air for each layer in the measurement grid equals

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

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 $\lambda^{\mathrm{N}} = 477\,\mathrm{nm}$) 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 α (Ångström, 1929), calculated from the MACC array of optical depths aergn03 (use the pair of total OD's at 532 nm and at the wavelength closest to the measurement $\lambda^{\mathrm{N}} = 477\,\mathrm{nm}$, in this case 469 nm):

$$\alpha = -\left(\ln\frac{\mathbf{OD}_{469}}{\mathbf{OD}_{532}}\right) \Big/ \left(\ln\frac{469}{532}\right).$$

Under the assumption that the Ångström coefficient is height independent and the measurement's wavelength λ^{N} falls within the validity range of the above estimate for

GMDD
7, 8151–8178, 2014

α, 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. $\mathbf{O}_3^{M,R}$, has been brought 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 $\mathbf{O}_3^{M,R}$ (see Rodgers, 2000, matrix multiplication is used):

$$\mathbf{O}_3^{M,S} = \mathbf{O}_3^{N,ap} + \boldsymbol{A}^{N} \cdot \left(\mathbf{O}_3^{M,R} - \mathbf{O}_3^{N,ap}\right),$$

where the measurements apriori profile $\mathbf{O_3}^{N,ap}$ 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 $\mathbf{O}_3^{M,R} - \mathbf{O}_3^{N,ap}$ and as a final step, after applying the above smoothing formula, the values in the resulting smoothed model profile $\mathbf{O}_3^{M,S}$ corresponding to the initial void layers of the re-gridded profile are replaced by NaN.

GMDD
7, 8151–8178, 2014

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.

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. The resulting smoothed model total column/optical depth is void if the measurement grid outranges the model grid.

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 σ, or covariances **S**. Uncertainty propagation requires the knowledge of covariance matrices. If either the

GMDD
7, 8151–8178, 2014

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

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

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

$$\boldsymbol{a}^{\mathrm{N}}(i)\boldsymbol{a}^{\mathrm{N}}(j).$$

To transform a covariance matrix from optical thickness units towards optical depth, the same formula is used where $\boldsymbol{a}^{\mathrm{N}}$ is replaced by the vector of layer thickness Δz^{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, assume that the monthly mean of

GMDD
7, 8151–8178, 2014

a time series of O_3 partial columns is calculated. Due to the nature of systematic and random uncertainties (see Taylor, 1997; JCGM/WG2, 2008), the random uncertainty σ^r of the monthly mean decreases at rate $1/\sqrt{n}$, with n the sample size (i runs over all measurements in a month):

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

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 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 many years in the atmospheric scientific community and

GMDD
7, 8151–8178, 2014

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.

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 available in NDACC. The data providers for FTIR, MWR, LIDAR and UVVIS measurements in the NDACC database are acknowledged.

GMDD
7, 8151–8178, 2014

References

Andrews, D.: An Introduction to Atmospheric Physics, Cambridge University Press, Cambridge, 2010. 8157

Ångström, A.: On the atmospheric transmission of sun radiation and on dust in the air, Geogr. Ann. A, 11, 156–166, 1929. 8163

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 (last access: 28 September 2014), 2013. 8154, 8164, 8166

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. 8153, 8165

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/utils/common/documents/jcgm/JCGM_100_2008_E.pdf (last access: 28 September 2014), 2008. 8167

Lambert, J.-C.: Validation protocol, available at: http://www.gmes-atmosphere.eu/documents/deliverables/man (last access: 28 September 2014), 2010. 8153

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. 8157

QA4EO: A Quality Assurance Framework for Earth Observation: Principles, available at: http://qa4eo.org/docs/QA4EO_Principles_v4.0.pdf (last access: 28 September 2014), 2010. 8153

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. 8153, 8164

Richter, A., Godin, S., Gomez, L., Hendrick, F., Hocke, K., Langerock, B., van Roozendael, M., and Wagner, T.: NORS Spatial Representativeness of NORS observations, available at: http://nors.aeronomie.be/projectdir/PDF/D4.4_NORS_SR.pdf (last access: 28 September 2014), 2013. 8164, 8166

GMDD
7, 8151–8178, 2014

Rodgers, C. D.: Inverse Methods for Atmospheric Sounding: Theory and Practice (Series on Atmospheric Oceanic and Planetary Physics), World Scientific Publishing Company, Singapore, 2000. 8164

Taylor, J.: An Introduction to Error Analysis: The Study of Uncertainties in Physical Measurements, Physics – Chemistry – Engineering, University Science Books, California, 1997. 8167

GMDD

7, 8151–8178, 2014

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
UVVIS.DOAS.ZENITH	OHP, Reunion, Jungfraujoch	**O_3**, **NO_2**
UVVIS.DOAS.OFFAXIS	Xianghe, Jungfraujoch, Izaña, Ny Ålesund	**NO_2**, **HCHO**, Aerosol extinction
FTIR	Reunion, Izaña, Jungfraujoch, Ny Ålesund	**O_3**, CO, **CH_4**
MWR	Bern, Ny Ålesund	**O_3**
LIDAR	Reunion, OHP	**O_3**

GMDD
7, 8151–8178, 2014

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 instrument	φ
LONGITUDE.INSTRUMENT	(1)	rad	longitude instrument	λ
ALTITUDE.INSTRUMENT	(1)	m	altitude instrument	z^{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 location of probed air mass at each altitude (optional)	λ
LONGITUDE	(100, 47)	rad	longitude of location of probed air mass at each altitude (optional)	φ
ALTITUDE.BOUNDARIES	(2, 47)	m	boundaries of vertical grid layers (optional)	z_{B}

GMDD
7, 8151–8178, 2014

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
lnsp	(256, 512)	ln Pa	logarithmic surface pressure (depends on latitude and longitude)	**lnsp**
–	(61)	Pa	array ap is used as translation terms for the construction of the pressure grid	***ap***
–	(61)	–	array bp are used as scaling factors in the construction of the pressure grid	***bp***
–	(512)	rad	the array of longitudes (i_λ denotes the index for this array)	λ
–	(256)	rad	the array of latitudes (i_φ denotes the index)	φ
t	(60, 256, 512)	K	temperature	**T**
z	(256, 512)	m	geopotential height of the surface (corresponds to **lnsp**)	$\mathbf{z}_s$
q	(60, 256, 512)	kg kg^{-1}	water vapour MMR (w.r.t. moist air)	**q**
go3	(60, 256, 512)	kg kg^{-1}	MACC ozone MMR (w.r.t. moist air)	$\mathbf{O}_3$
hcho	(60, 256, 512)	kg kg^{-1}	Formaldehyde MMR (w.r.t. moist air)	**HCHO**
co	(60, 256, 512)	kg kg^{-1}	CO MMR (w.r.t. moist air)	**CO**
aergn04	(60, 256, 512)	–	Profile of aerosol optical depths at 532 nm	τ_{532}
aergn03	(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	**OD**
ch4	(60, 256, 512)	kg kg^{-1}	$\mathbf{CH}_4$ MMR (w.r.t. moist air)	$\mathbf{CH}_4$
no2	(60, 256, 512)	kg kg^{-1}	$\mathbf{NO}_2$ MMR (w.r.t. moist air)	$\mathbf{NO}_2$

GMDD
7, 8151–8178, 2014

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. Δ is the time between two subsequent MACC data validity times – see text (target values may contain wildcards if applicable to multiple targets).

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

GMDD
7, 8151–8178, 2014

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

GEOMS template version	GEOMS name	notation
LIDAR-003	O3.NUMBER.DENSITY_ABSORPTION.DIFFERENTIAL	$\mathbf{O_3}$
	O3.NUMBER.DENSITY_ABSORPTION.DIFFERENTIAL_UNCERTAINTY.ORIGINATOR	σ^r
FTIR-002	*.MIXING.RATIO.VOLUME_ABSORPTION.SOLAR	$\mathbf{O_3}$,**CO**, $\mathbf{CH_4}$,...
	*.MIXING.RATIO.VOLUME_ABSORPTION.SOLAR_UNCERTAINTY.RANDOM.COVARIANCE	$\mathbf{S}^r$
	*.MIXING.RATIO.VOLUME_ABSORPTION.SOLAR_UNCERTAINTY.SYSTEMATIC.COVARIANCE	$\mathbf{S}^s$
	*.MIXING.RATIO.VOLUME_ABSORPTION.SOLAR_APRIORI	$\mathbf{O}_3^{ap}$,...
	*.MIXING.RATIO.VOLUME_ABSORPTION.SOLAR_AVK	$\boldsymbol{A}$
UVVIS.DOAS.ZENITH-006	*.COLUMN.STRATOSPHERIC_SCATTER.SOLAR.ZENITH	$\mathbf{O_3}$, $\mathbf{NO_2}$
	*.COLUMN.STRATOSPHERIC_SCATTER.SOLAR.ZENITH_UNCERTAINTY.RANDOM.STANDARD	σ^r
	*.COLUMN.STRATOSPHERIC_SCATTER.SOLAR.ZENITH_UNCERTAINTY.SYSTEMATIC.STANDARD	σ^s
	*.MIXING.RATIO.VOLUME_SCATTER.SOLAR.ZENITH_APRIORI	$\mathbf{O}_3^{ap}$,...
	*.COLUMN.STRATOSPHERIC_SCATTER.SOLAR.ZENITH_AVK	$\boldsymbol{A}$
UVVIS.DOAS.OFFAXIS-006	*.MIXING.RATIO.VOLUME_SCATTER.SOLAR.OFFAXIS	$\mathbf{NO_2}$, **HCHO**
	*.MIXING.RATIO.VOLUME_SCATTER.SOLAR.OFFAXIS_UNCERTAINTY.RANDOM.COVARIANCE	$\mathbf{S}^r$
	*.MIXING.RATIO.VOLUME_SCATTER.SOLAR.OFFAXIS_UNCERTAINTY.SYSTEMATIC.COVARIANCE	$\mathbf{S}^s$
	*.MIXING.RATIO.VOLUME_SCATTER.SOLAR.OFFAXIS_APRIORI	NO_2^{ap},...
	*.MIXING.RATIO.VOLUME_SCATTER.SOLAR.OFFAXIS_AVK	$\boldsymbol{A}$
	AEROSOL.EXTINCTION.COEFFICIENT_SCATTER.SOLAR.OFFAXIS	$\boldsymbol{\tau}$
	WAVELENGTH	λ
	AEROSOL.EXTINCTION.COEFFICIENT_SCATTER.SOLAR.OFFAXIS_UNCERTAINTY.RANDOM.COVARIANCE	$\mathbf{S}^r$
	AEROSOL.EXTINCTION.COEFFICIENT_SCATTER.SOLAR.OFFAXIS_UNCERTAINTY.SYSTEMATIC.COVARIANCE	$\mathbf{S}^s$
	AEROSOL.EXTINCTION.COEFFICIENT_SCATTER.SOLAR.OFFAXIS_APRIORI	$\boldsymbol{\tau}^{ap}$
	AEROSOL.EXTINCTION.COEFFICIENT_SCATTER.SOLAR.OFFAXIS_AVK	$\boldsymbol{A}$
MWR-003	O3.MIXING.RATIO.VOLUME_EMISSION	$\mathbf{O_3}$
	O3.MIXING.RATIO.VOLUME_EMISSION_UNCERTAINTY.RANDOM.STANDARD	σ^r
	O3.MIXING.RATIO.VOLUME_EMISSION_UNCERTAINTY.SYSTEMATIC.STANDARD	σ^r
	O3.MIXING.RATIO.VOLUME_EMISSION_APRIORI	$\mathbf{O}_3^{ap}$
	O3.MIXING.RATIO.VOLUME_EMISSION_AVK	$\boldsymbol{A}$

GMDD
7, 8151–8178, 2014

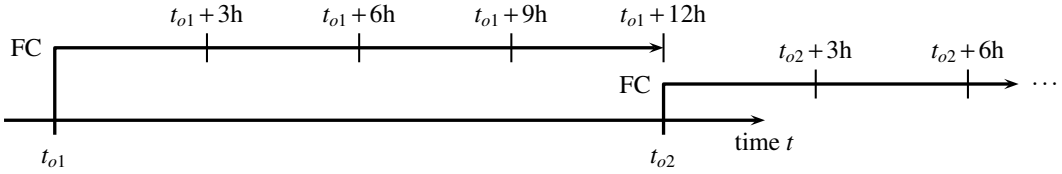

Figure 1. Construction of MACC data times for type FC (forecast).

GMDD
7, 8151–8178, 2014

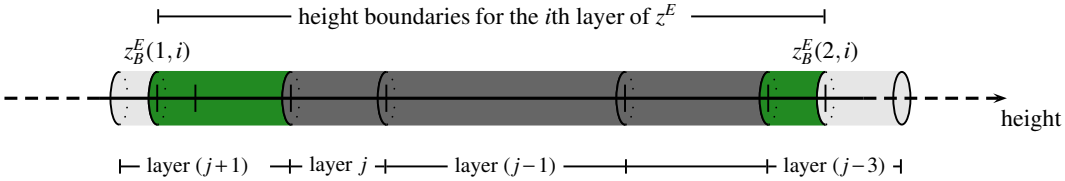

Figure 2. Interpolation factors for grid layers: green source grid layers are only partially overlapped by the ith destination layer.

GMDD
7, 8151–8178, 2014

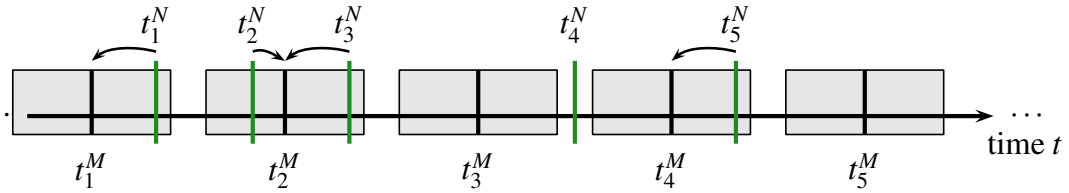

Figure 3. Temporal co-location of MACC (model, black) and NORS (measurement, green) data.

GMDD
7, 8151–8178, 2014