Assessment of CAS products

Deliverable title	Assessment of GAS products
Deliverable number	D9.2
Revision	06
Status	Final
Planned delivery date	30/11/2014
Date of issue	01/12/2014
Nature of deliverable	Report
Lead partner	University of Liège
Dissemination level	Public

This work has received research funding from the European Community's Seventh Framework Programme ([FP7/2007-2013]) under grant agreement n°284421.

DOCUMENT PROPERTIES

	FUNCTION	NAME	DATE	SIGNATURE
LEAD AUTHOR	WP9 lead	Mahieu, Emmanuel	01/12/2014	
CONTRIBUTING AUTHORS	Scientist	Langerock, Bavo		
	Scientist	Franco, Bruno		
	Scientist	Bader, Whitney		
	Scientist	Hendrick, François		
	Scientist	Henne, Stephan		
	Scientist	Hocke, Klemens		
	Scientist	Palm, Mathias		
	Software architect	Breebaart, Leo		
Reviewed by	Coordinator	Martine De Mazière	07/12/2014	

Table of Contents

Executive summary

An important objective of the NORS project is to assess the quality of the Copernicus Atmospheric Service products characterizing the chemical and physical states of the atmosphere. To this end, a subset of techniques, sites and species available in the Network for the Detection of Atmospheric Composition Change (NDACC) has been carefully selected. Ground-based FTIR, LIDAR, microwave radiometers and UVVIS instruments contribute to the project by providing data in a rapid delivery mode (WP3) as well as consolidated measurements (WP7), using in all cases GEOMS templates which have been defined or updated within the framework of NORS (WP4). A suite of sites were included in the project from the beginning (Ny Alesund, the Alpine stations, Izana and Reunion Island) while others were added through the capacity building effort (WP10). The six targets are ozone, nitrogen dioxide, carbon monoxide, methane, formaldehyde and aerosol extinction. The corresponding products have been extensively characterized (WP4), they have undergone numerous comparison and validation exercises (between techniques, with satellite measurements), giving confidences in the related NDACC data. The representativeness of a subset of them has been assessed in WP5 for Jungfraujoch and Izana, after comparison with in-situ measurements of CH_4, CO, O_3 and NO_2.

Appropriate procedures have then been defined to account for the NDACC product performance, time sampling, horizontal extent such that the validations of the CAS products are performed in an optimum and automatic way. These procedures were implemented in the NORS validation server (NVS; WP8), which is a key element of the project. Thanks to this web-based tool, validation reports are automatically produced, allowing characterizing the performance of the MACC model simulations, for a broad range of atmospheric conditions, encompassing latitudes from the Arctic down to the southern sub-tropical region, from highly polluted sites to remote locations.

In the present work package, which builds on all efforts undertaken in the other WPs and benefits from the NORS achievements, the performance of the MACC simulations have been evaluated for all NORS products, techniques and sites. We present here an overview of the validation results, pointing towards some specific shortcoming of the MACC forecast or of the NORS products.

For O_3, we generally noted good to very good agreement between the NORS measurements and the MACC simulations in a broad range of altitudes, latitudes and for all seasons, in particular of the fnyp runs. Relative biases are often limited to a few percent and the atmospheric variability is well captured by the models. However, in specific circumstances, we observed significant discrepancies, e.g. for Ny Alesund where winter low partial columns are not matched by the models (by up to 80 %), or for Reunion Island where the ozone variability can sometimes be poorly represented by the models.

For CO, we noted reasonable to good agreement between the models and the observations. The seasonal variability seems systematically underestimated during springtime and we observed contrasted performances of the models in the representation of the CO abundance, depending on the season and site.

Comparisons involving FTIR measurements of CH_4 and fsd7 model simulations indicated satisfactory agreement for the total columns, with good statistics in terms of correlation and

representation of the (somewhat limited) atmospheric variability of methane. Relative mean biases are generally in the 3–5 % range. However, the retrieved FTIR profiles often suggest a slope in the troposphere which is likely unrealistic given the lifetime of CH_4. And indeed, the model indicates constant mixing ratios from the surface up to the tropopause, as expected. Hence, the comparison points here towards deficiencies of the FTIR to produce sensible tropospheric profile shapes for methane, and this is probably the result of remaining inconsistencies in the spectroscopic line parameters available to date. Efforts are ongoing in the laboratory and it can be anticipated that the situation will improve in the years to come.

For NO_2, FTIR and fsd7 data were compared for Izana and Reunion-Maido. The main conclusion is that in both cases the model has difficulties to represent the NO_2 seasonal cycle, with significant overestimation of the NO_2 column in winter and an underestimation during summertime.

For formaldehyde and aerosol, NORS near surface data are available from the Chinese site of Xianghe, near Beijing, where a MAX-DOAS is in operation. This is a polluted area, and the comparisons lead to similar conclusions for these two targets, i.e. that if the models appropriately reproduce background conditions, they are far from able to forecast the frequent polluted episodes, with simulations significantly biased low. Possible explanations include missing emissions in the model(s), and the impact of an insufficient horizontal resolution of the simulation. Nevertheless, the impact of the clouds on the quality of the UVVIS data will have to be investigated before drawing definitive conclusions.

Finally, we have identified HCHO FTIR and SO_2 MAX-DOAS measurements as obvious candidates to complete the list of NDACC products relevant for the validation of MACC model simulations. The first addition would help to complement the comparisons of formaldehyde up to the tropopause level while increasing the number of sites. As an important indicator of air quality, SO_2 surface concentrations would help challenging the models in polluted areas.

Applicable and reference documents

NORS Description of Work
Validation server User Requirements Documents (D8.1), May 14, 2012
Validation server Design Document (D8.2), July 9, 2012
Validation server in test-phase (D8.3), August 1, 2013
Feedback report regarding the NORS Validation Server (D9.1), November 8, 2013

Acronyms and abbreviations

CAS	Copernicus Atmospheric Service
DOAS	Differential Optical Absorption Spectroscopy
FTIR	Fourier Transform InfraRed spectroscopy
GEOMS	Generic Earth Observation Metadata Standard
GMES	Global Monitoring for Environment and Security
IFS	Integrated Forecasting System – MOZART model
LIDAR	Light detection and ranging
MACC	Monitoring Atmospheric Composition & Climate
MAX-DOAS	Multi Axis Differential Optical Absorption Spectroscopy
MWR	Microwave radiometer
NDACC	Network for the Detection of Atmospheric Composition Change
NORS	Network of Remote Sensing Ground-Based Observations for the GMES Atmospheric Service
NVS	NORS Validation Server
OHP	Observatoire de Haute Provence
RDD	Rapid Data Delivery
SAOZ	Système d'Analyse par Observation Zénitale
UVVIS	Ultraviolet visible spectroscopy
VMR	Volume mixing ratio

1. Introduction

The NORS validation server (aka NVS) is a key component of the project which has been specifically designed and developed to perform automatic validation of the Copernicus Atmospheric Service (CAS) model products using rapid delivery or consolidated data available from the NDACC data base. So-called validation reports including NORS and CAS products are automatically generated, providing figures and associated statistics allowing meaningful comparisons. In the present document, we will use standard NVS validation reports as well as comparison results from specific investigations to evaluate the overall quality of the CAS products. Indeed, the NVS tool can also be used to perform additional analyses. This capability has been exploited to perform validation exercises for inclusion in the MACC reports. Specific investigations involving tropospheric columns have been conducted, using the NVS architecture in offline mode.

2. Products used in the comparisons

Here we identify the products which are the subject of the various comparisons. Table 1 lists the NORS products and techniques. Four sites are primarily involved, namely Ny Alesund, the Alpine stations, Izana and Reunion Island. It should be noted however that additional sites are also included in the comparisons, as a result of the capacity building effort pursued in the project (NORS WP10) or because other (non-NORS) NDACC data can be ingested by the NVS. This is the case if the following holds true: (i) the data correspond to a NORS target and technique, (ii) the data are available from the RDD or consolidated directory, in HDF (GEOMS 1.0), using the most recent template, (iii) the data can be successfully processed by the NVS backend tool chain. This significantly broadens the number of intercomparisons by expanding the latitude coverage but also the parameters. For example, aerosols products can be compared at the Chinese station of Xianghe where an UVVIS.DOAS instrument is operated by BIRA-IASB, contributing to the RDD chain.

	O_3	**NO_2**	**CO**	**CH_4**	**HCHO**	**Aerosol**
DOAS	SC	SC, SP, TC				
MAX-DOAS	SC	SC, TC			TC	E
FTIR	TC, SC	SC	TC, SC	TC, SC	(C)	
LIDAR	SP					
MWR	SP					

Table 1. NORS products and techniques involved in the intercomparisons. Abbreviations as follows: Stratospheric Column (SC); Stratospheric Profile (SP); Tropospheric Column (TC), Total column (C); Extinction (E).

Table 2 identifies the MACC forecast models used in NORS and provides information on the time periods currently covered. Given these, the comparisons will often be limited to 2013 and after. More information about the models is available from the MACC website, in the Operational Info section. It should be noted that in contrast to fkya and fsd7, fnyp includes data assimilation, using essentially various satellite products, and this can have implications as to the relative performances of the models.

	Description	Forecast	Availability*
fnyp	Pre-operational MACC DA/FC run (MACC_osuite)	O_3, CO, HCHO and aerosol	>20120704
fkya	Pre-operational MACC DA/FC run (MACC_fcnrt_MOZ)	O_3, CO, HCHO	>201109
fsd7	NRT run based on C-IFS, without DA using CB05 trop. chemistry (no aerosol) (MACC_CIFS_TM5)	O_3, CO, CH_4, NO_2, HCHO	>20121102

Table 2. MACC forecast models used in the NORS validation server and available products. The colour key (red for fnyp, orange for fkya and blue for fsd7) is consistently used throughout the standard and custom validation reports. (*) Note that all "f* models" were replaced (by the g4e2 MACC_osuite) at the end of September 2014 and that the NVS already handles the new model data. However, for statistical reasons and data availability, we will stick here to comparisons involving the "f* forecasts".

3. Taylor diagrams

By simply combining the entries of Table 1 and 2, it quickly appears that a large number of comparisons is available from the NVS. Table 3 lists the NVS reports available at the end of October 2014, per year and per technique, per year and per target. Altogether, more than 2100 monthly reports have been generated by the validation server.

Per year / per technique	**2013**	**2014**
DOAS.ZENITH	66	51
DOAS.OFFAXIS	74	13
FTIR	866	551
LIDAR	44	40
MWR	355	122
Per year / per target	**2013**	**2014**
O3	571	358
NO2	130	68
CO	323	223
CH4	167	103
HCHO	44	0
Aerosol	10	5
TOTAL	**> 1400**	**> 770**

Table 3. Inventory of NVS monthly reports (31/10/2014).

To ease or even allow the assessment of the overall agreement between the forecasts and the observations, Taylor diagrams or plots (Taylor, 2001) have been implemented in the NVS, allowing evaluating the relative skill of the MACC forecast models to match the NORS observations, in terms of correlation and variability (the statistics are established after correction of a possible mean bias between observed and modelled data). A sample Taylor plot is shown in Figure 1, for the comparison of ozone total columns above Ny Alesund, as derived from FTIR observations performed in May 2013. This plot provides the following information: (i) the number of coincidences used to evaluate the statistics (here 12); (ii) the correlation coefficient between the observed and synthetic data; e.g., for fnyp, a very good correlation of 0.99 is observed; (ii) the radial position of the data points informs on the standard deviations of the model relative to the observations (normalised to 1), here again the fnyp model performs really well with an amplitude in the variation of ozone only slightly larger than for the observations (1.1 of standard deviation). Conversely, fkya underestimates the ozone variability at high northern latitude. The black symbol would represent an ideal simulation, with a perfect match with the observations. Note that Taylor diagrams include a second quadrant in the case where anti-correlations are found between modelled and observed quantities. Access to these diagrams is restricted to the VIP users of the server.

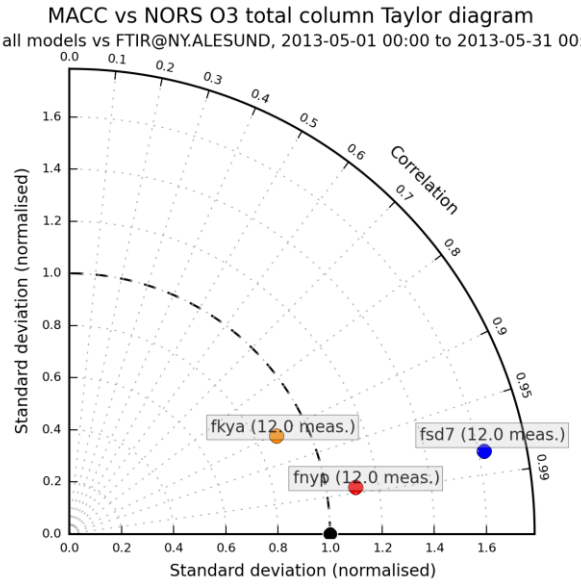

Figure 1. Taylor plot showing the statistical comparison results between ozone total columns measured by the FTIR instrument at Ny Alesund and as simulated by the fkya, fsd7 and fnyp MACC models.

4. Validation of the CAS products

In this section, we successively present statistical comparisons for the NORS products. Some comparisons are restricted to the overall performance of the MACC models to match the observations as a whole (i.e. the models vs all sites). When relevant, specific examples are provided to illustrate some shortcomings of the models to represent specific seasonal and/or latitudinal situations.

4.1. Ozone (O_3)

Table 1 and Table 2 indicate that ozone is a target of all NORS techniques, and that it is available from all MACC models.

4.1.1. Comparison with FTIR total columns

Figure 2 shows Taylor plots for the comparison between the MACC models and all FTIR observations available in 2014 from the NORS sites, from January to August. It can be seen that the agreement is generally good, with correlations often larger than 0.9. The fnyp model (red dots) performs really well, with correlation factors often greater than 0.95 and ozone variability close the “truth” (dashed curve). The situation is more variable for fkya (in orange) which shows some poor comparisons (e.g. in May) and fsd7 (in blue) which tends to overestimate the ozone variability for the April-August months.
If we look into some specific site-by-site comparisons, the overall performance of the models is confirmed and similar at all sites, except at Reunion. In some instances, the correlation factors are really low for that site (e.g. below 0.4 for fsd7 and fnyp in August 2014) or the ozone variability is poorly represented by the models (e.g. with a normalised standard deviation of 2.5 for fkya in August 2013, or at another extreme of 0.45 for fnyp in June 2014).

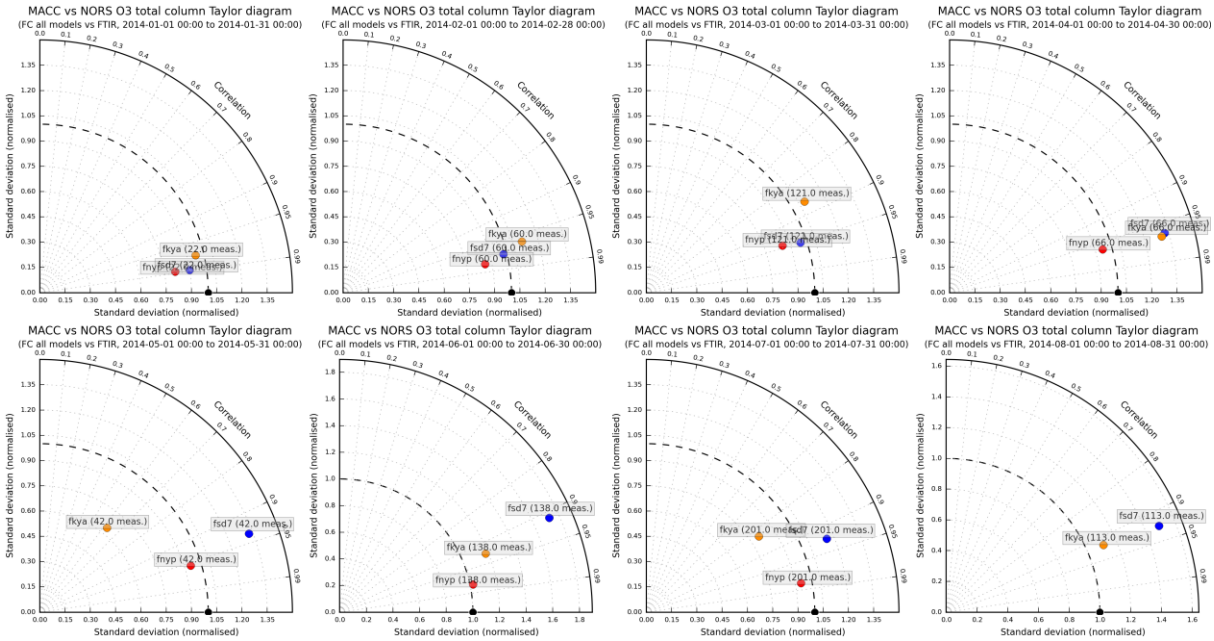

Figure 2. Relative performance of the MACC models in reproducing the FTIR ozone total columns for January to August 2014 (from left to right and top to bottom).

4.1.2. Comparison with LIDAR measurements (10–50 km)

Regular LIDAR measurements are available for OHP (2013/02-2014/07) and Reunion-Maido (2014/06-2014/09) and all the MACC models. Figure 3 shows comparisons for the Haute Provence LIDAR, for 10 consecutive months, in the 15 to 45 km altitude range. Overall, the ozone variability is well captured by the models; mean biases close to 4% are derived for fnyp and fkya while fsd7 predicts ozone columns larger by 8.7% on average. When looking into profile comparisons (not shown), it appears that fsd7 overestimates ozone between 22 and 30 km by about 10-15%, without any clear seasonal pattern in the differences.

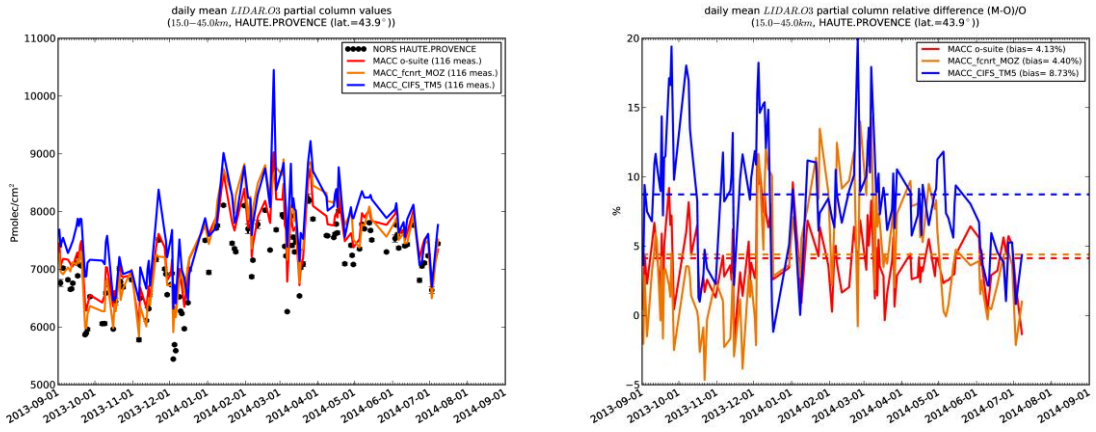

Figure 3. Comparisons between LIDAR and MACC model partial ozone columns above OHP.

4.1.3. Comparison with MWR measurements (25–60 km)

MWR ozone measurements are currently available from Bern (2013/02-2014/09), Ny Alesund (2013/02-2014/09), Lauder (2013/02-2014/05), Mauna Loa (2013/02-2014/05) and all the MACC models. These measurements are sensitive above about 25 km, and comparisons have

been performed for the NORS stations in the 25 to 60 km altitude range, over a time period of one year, starting in September 2013. Figure 4 presents the observed and modelled time series for Bern and Ny Alesund as well as the relative differences on the partial columns. It appears that the fsd7 model tends to overestimate the ozone columns, in particular for September to January and high latitudes, when the minimum columns are observed. fnyp and fkya are closer (NYA) or in agreement (Bern) with NORS, with fkya generally providing the lowest partial columns. The mean biases are reasonable for the mid-latitude site of Bern, ranging from about -3 to 6%. For Ny Alesund, they are much larger (from ~13 to 43%), as a result of the poor agreement noted in autumn.

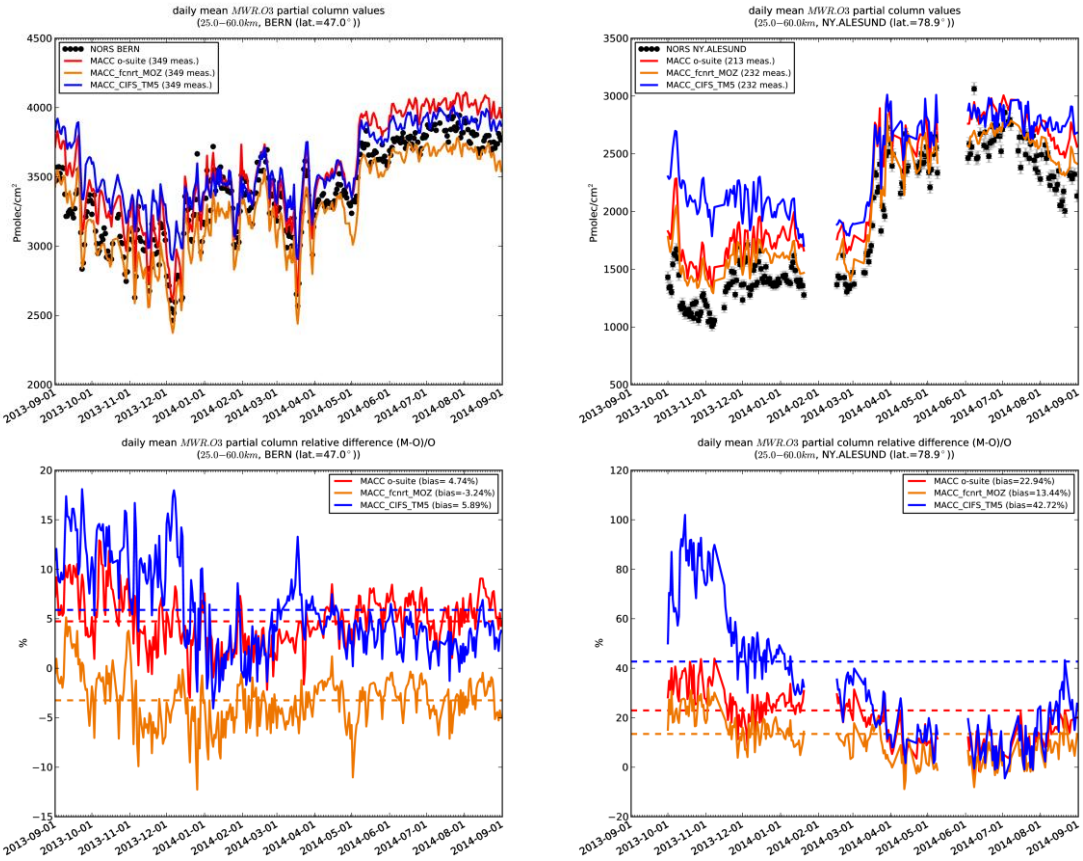

Figure 4. Comparisons between MWR and MACC model partial ozone columns above Bern (left) and Ny Alesund (right). The upper and lower panels show the time series and the relative differences, respectively.

A recent study investigated the ozone diurnal variations in the Polar Regions using MWR observations at Ny Alesund. These variations are strong (from 15% in summer to 45% in winter) and can affect ozone trend evaluations. The authors conclude that the MACC reanalyses provide a realistic representation of the ozone diurnal variations; the models could then be used to correct for the diurnal sampling effects, allowing improving the ozone trend determinations. More information is available in Schanz et al. (2014).

4.1.4. Comparison with UVVIS total columns

Intercomparisons are available for Jungfraujoch (2013/02-2014/09), Izana (2013/02-2014/09) and Harestua (2013/02-2014/08), for all the MACC models. The latter site is situated at 60ºN, often on the vortex edge at the end of winter/beginning of spring. Mean biases of 6 % (fnyp),

7 % (fkya) and 14 % (fsd7) have been determined for ozone partial columns (0.6 – 50 km), for comparison involving over a year of observations by the DOAS instrument (346 measurements). It is interesting to look at the March 2014 month during which significant variability has been observed, with ozone columns varying by about a factor 2 between March-12 and March-23. The figure 5 shows model and observed ozone columns, pair wise. Even for these challenging conditions, the models do a good job in reproducing the day-to-day changes in ozone. More specifically, fnyp shows a mean positive bias of 2.5% and overestimate the high ozone columns by about 10%; fkya is also biased high (7.3%) and provide ozone columns too large for the end of the month (by about 20%); fsd7 provides a mean bias of 8.4% and in contrast with fkya and fnyp, this bias is more consistent throughout the whole month.

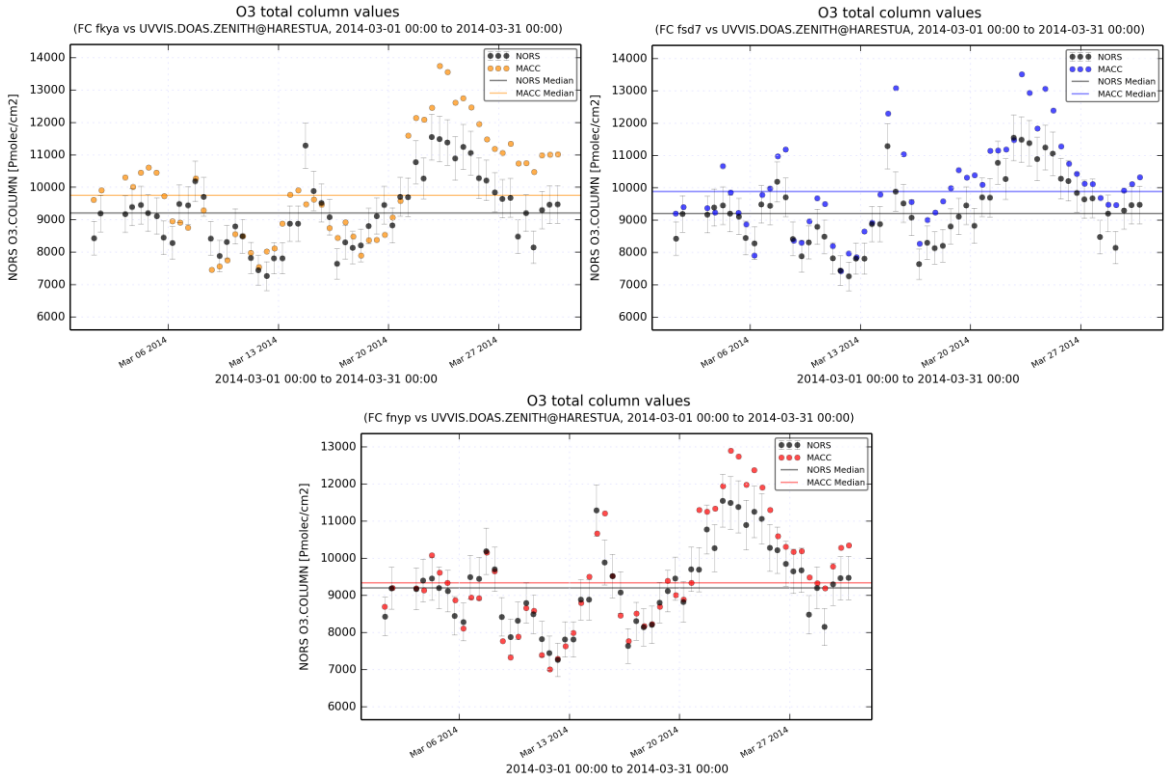

Figure 5. Ozone total column comparison for Harestua in March-2014. The DOAS measurements are successively compared to the fkya (upper left), fsd7 (upper right) and fnyp MACC models.

4.2. Carbon monoxide (CO)

Carbon monoxide is available from the fnyp, fsd7 and fkya MACC models and from FTIR measurements at up to 7 NDACC sites. This species presents significant seasonal variation and intraday/spatial variability. Taylor diagrams involving the three models and all available sites are reproduced in Figure 6, for the period September 2013 to August 2014, i.e. a full seasonal cycle. Overall, the fnyp model provides the best correlations with the observations with a good representation of the CO variability (close to the dashed curved for most instances). However, the models seem to systematically underestimate the atmospheric variability of CO during springtime. fkya also performs well, also a little less satisfactorily than fnyp in fall and summer. fsd7 is the only model suggesting CO variability larger than the

one observed by the FTIR instruments, in September, December and to a lesser extent in January.

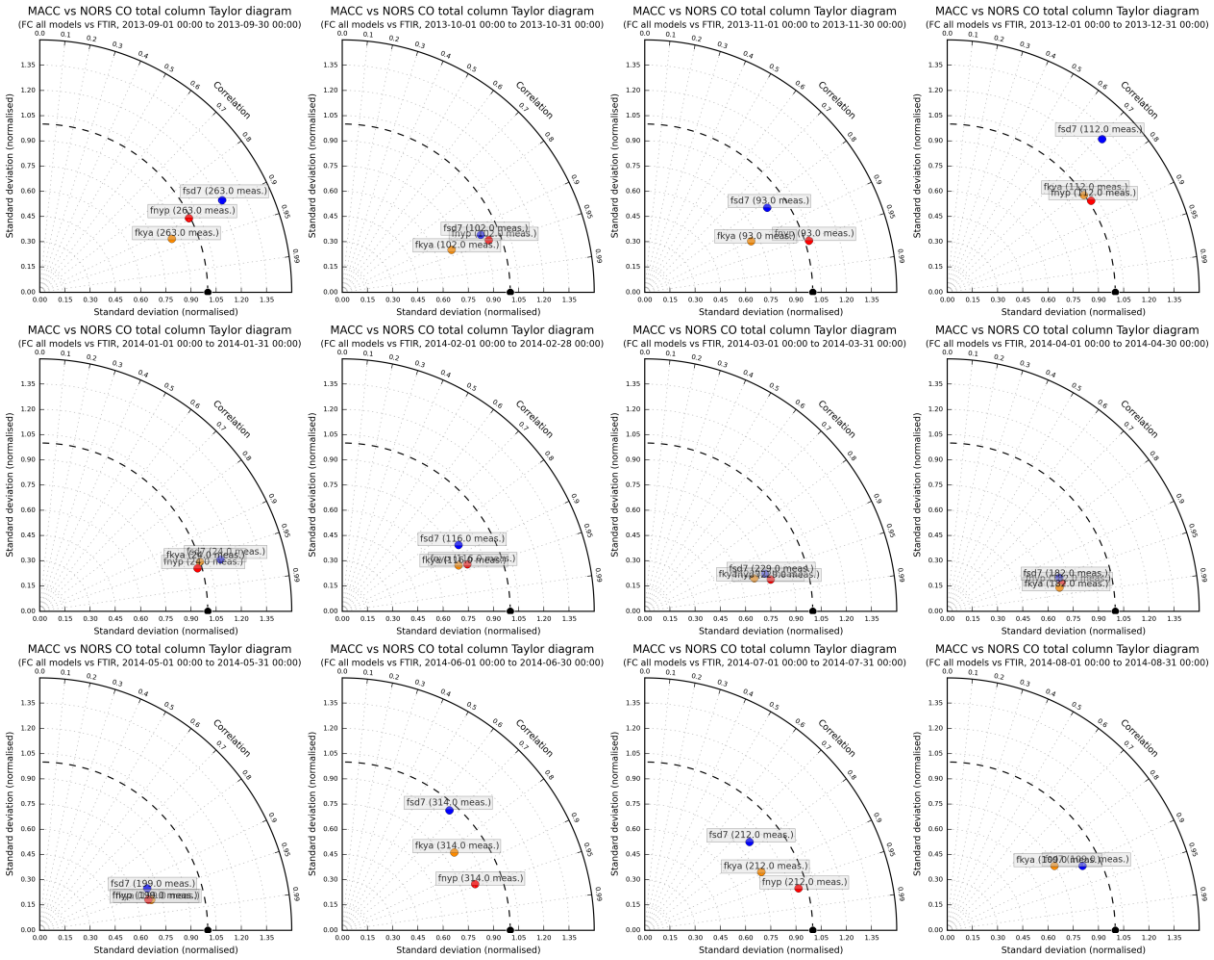

Figure 6. Relative performance of the MACC models in reproducing the FTIR CO total columns for a complete seasonal cycle, from September 2013 to August 2014 (from left to right and top to bottom), for the available sites.

4.2.1. Comparison with FTIR observations at Jungfraujoch and Reunion (Maido)

Here below, we show comparisons between the CO profiles of the MACC models and the FTIR measurements at Jungfraujoch (47°N, 8°E, Alpine station, altitude 3.58 km) and La Reunion Maido (21°S, 55°E, i.e. southern tropics, altitude 2.2 km). These ground-based, remote-sensing instruments are sensitive to the CO abundance in the troposphere and lower stratosphere, i.e. between the surface and up to 20 km altitude. Tropospheric CO profiles and columns are validated (up to 10km).

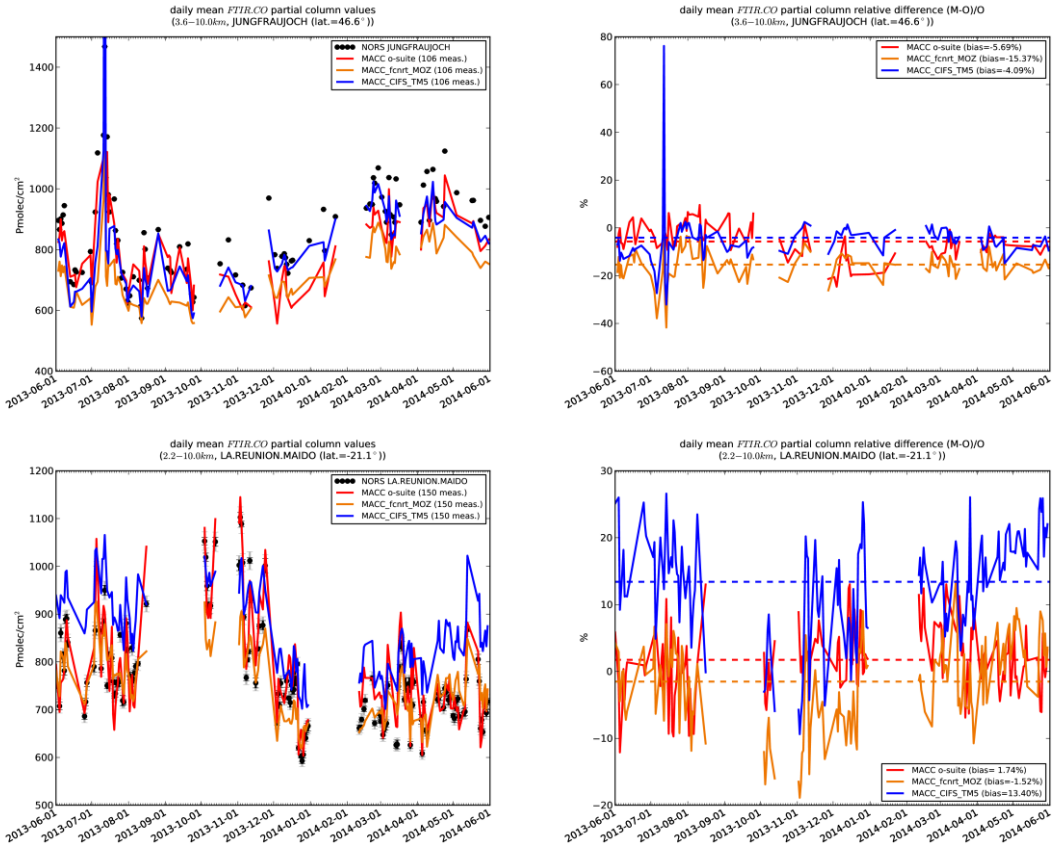

Figure 7. Daily mean tropospheric CO columns (up to 10 km) (left) and relative differences (right) by MACC_osuite (or fnyp, red), MACC_fcnrt_MOZ (or fkya, orange), MACC_CIFS_TM5 (or fsd7, blue, full line) compared to NDACC FTIR data at Jungfraujoch (47°N, 8°E) (top) and La Reunion Maido (21°S, 55°E) (bottom) for the period June 2013-May 2014. The number of measurement days and yearly bias are indicated in the legend.

4.2.1.1. MACC_osuite (fnyp)

Table 4 shows that the tropospheric columns of CO agree well during the summer season (JJA), with a seasonal bias less than 1%. There is a large negative bias (up to -10% during winter) in the following seasons at Jungfraujoch. At Maido the o-suite overestimates the measured CO columns with 2% during spring and autumn and 5% during the southern summer months.

4.2.1.2. MACC_fcnrt_MOZ (fkya)

fkya underestimates the tropospheric CO abundance with an annual bias of 15 % at Jungfraujoch. The situation seems better at Maido except during spring where the bias raises to 10%.

4.2.1.3. MACC_CIFS_TM5 (fsd7)

fsd7performs better at the Alpine station, where the model underestimates the measurements (6%). At Maido high overestimations are seen (16%).

4.2.1.4. Conclusions

fnyp performs well during the whole year at Maido. For Jungfraujoch a large underestimation is seen starting from autumn 2013. fkya and fsd7 do not perform as good as the o-suite. fkya underestimates at both measurement sites. fsd7 is different in that it overestimates the Maido measurements.

		JJA			SON			DJF			MAM		
		MB	stddev	nobs	MB	stddev	nobs	MB	stddev	nobs	MB	stddev	nobs
fnyp	Jung	-0.77	6.05	143	-4.40	8.41	45	-11.6	5.54	112	-6.30	3.88	163
	Maido	-0.57	5.86	184	1.90	4.37	96	4.97	5.00	96	2.10	4.61	272
fkya	Jung	-15.0	10.09	143	-15.3	5.82	45	-15.3	3.09	112	-15.6	3.20	163
	Maido	-1.14	4.78	184	-10.7	6.92	96	-2.88	4.94	96	1.94	5.34	272
fsd7	Jung	-3.47	20.18	143	-7.32	4.74	45	-2.34	2.90	112	-4.96	3.79	163
	Maido	15.25	6.03	184	2.17	9.10	96	12.91	5.90	96	16.30	6.14	272

Table 4. Seasonal relative mean bias (MB, %), standard deviation (STD, %) for the considered period and number of observations used (NOBS), compared to NDACC FTIR observations at Jungfraujoch and Maido (mean bias and stddev in %). "fnyp" corresponds to MACC_osuite, "fkya" to MACC_fcnrt_MOZ and "fsd7" to MACC_CIFS_TM5.

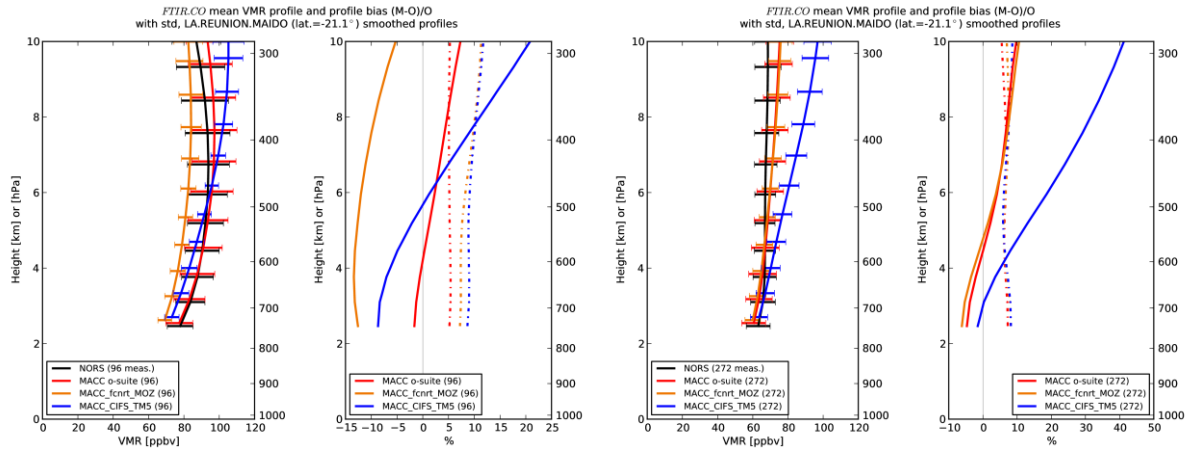

Figure 8. Seasonally mean tropospheric CO profiles (left) and relative profile differences (right) by MACC_osuite (red), MACC_fcnrt_MOZ (orange), MACC_CIFS_TM5 (blue, full line) compared to NDACC FTIR data at La Reunion Maido (21°S, 55°E) for the periods SON 2013 (left) and MAM 2014 (right). MACC_CIFS_TM5 overestimates during MAM (corresponding to the 16% bias in tropospheric column data).

4.3. Methane (CH_4)

The number of possible combinations for intercomparison is limited by the availability of MACC forecasts to the sole fsd7 model while methane is only measured by the FTIR instruments in NORS. Nevertheless, more than 260 individual reports are available at the time of writing for 6 NDACC sites. The atmospheric variability of CH_4 is somewhat limited at most sites, hence a validation of all total column measurements versus fsd7 at once provide very good statistics, both in terms of correlation (larger than 0.95 in all cases, i.e. from 02/2013 to 08/2014) and of representativeness of the atmospheric variability. However, when looking into individual intercomparisons, we observe some systematic differences between the fsd7 and the FTIR data. The columns derived from the FTIR data are often systematically

higher (e.g. by ~3% for Ny Alesund, Jungfraujoch and Zugspitze, ~5% for Izana and Reunion-Maido). The modelled and retrieved profiles also show some systematic differences: fsd7 simulating constant CH_4 vmr in the troposphere, while the FTIR retrieved profiles often suggest a slope in the lower troposphere and larger concentration at the surface level. This is particularly true for Jungfraujoch, Zugspitze and Izana. This is illustrated in Figure 9, for Zugspitze in March 2014.

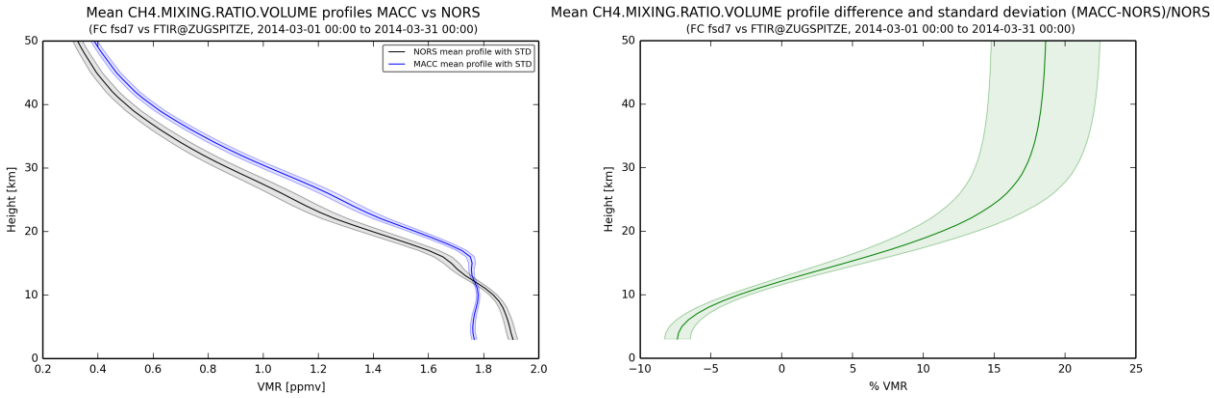

Figure 9. Intercomparison between fsd7 and FTIR mean profiles for March 2013 above the Zugspitze mountain site. The shape of the relative vmr difference is typical of most of the comparisons.

4.4. Nitrogen dioxide (NO_2)

Nitrogen dioxide simulations are produced by the fsd7 model, comparisons with FTIR total column at 3 sites (Izana, Jungfraujoch and Reunion) and MAX-DOAS boundary layer measurements at Xianghe are available in NVS. This species presents strong diurnal and seasonal cycles: it is important to consider observations and model data with close coincidence in time (time collocation interval for NO_2: below 30 min for FTIR and DOAS.ZENITH, below 1h for DOAS.OFFAXIS) and to check whether the seasonal signal is well reproduced by fsd7. Note that UVVIS (SAOZ) and FTIR column measurements of NO_2 above Jungfraujoch have been compared by Hendrick et al. (2012), showing a good agreement throughout the years and seasons, with FTIR measurements lower than SAOZ by 7.8±8.2 % on average.

4.4.1. Comparison with FTIR total columns

The number of coincidences per month for Jungfraujoch is quite limited at the time of writing; hence we will concentrate on the comparisons available for Izana and Reunion (Maido). We looked at the Taylor diagrams for the NO_2 total column comparisons above Izana, for the period 2013/09 to 2014/08 (not shown). The correlations are good, with most factors larger than 0.9. Also, the modelled NO_2 variability is generally close to the observed one, with the two most notable underestimations in July and August (normalised standard deviations of ~0.6 instead of 0.9–1.1 in the other cases). However, Figure 10 shows the seasonal distribution of the mean biases for Izana (in dark blue) and Reunion (Maido, in red, shifted by 6 months), with the error bars representing one standard deviation around the mean. We clearly identify a consistent seasonal signal in the mean biases derived from both sites, with a significant overestimation of the NO_2 total columns by fsd7 of up to 50% in winter, an underestimation by 5–10% of the NO_2 columns during summertime, at the limit of statistical significance (1-σ), and smoothed transitions between those two extremes in the intermediate months. It will be interesting to see if this is confirmed at mid- and high- latitudes when more data and reports will be available in the NVS.

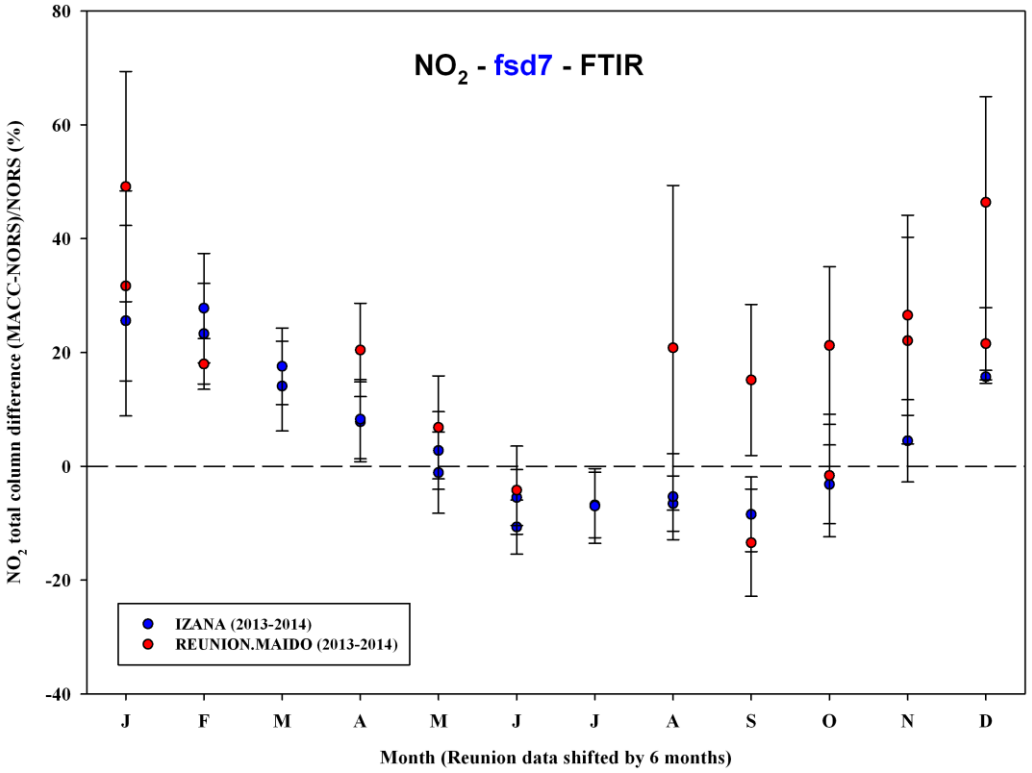

Figure 10. Mean relative bias ((MACC-NORS)/NORS)), in per cent, as derived from comparisons between FTIR and fsd7 NO_2 total columns for the sub-tropical high altitude sites of Izana and Reunion-Maido.

4.4.2. Comparison with MAX-DOAS

At the moment, such comparisons are only available for the Chinese site of Xianghe (39.8°N, 117°E, station near Beijing, altitude 92 m). The MAX-DOAS instrument operated there by BIRA-IASB is sensitive to near-surface tropospheric NO_2, and the statistics available in the NVS concern the first few kilometers above the surface. Observations and fsd7 model data have been compared for February to December 2013 and May to September 2014. Inspection of the Taylor diagrams reveals correlation factors in the 0.1–0.7 range, with an atmospheric variability of NO_2 systematically underestimated by the model, probably because of missing emissions. Despite the large number of available coincidences, covering more than a complete seasonal cycle, no clear picture emerges when looking at the statistics of the comparison. Mean biases are much scattered and characterized by very large standard deviations. When considering all available data at once, a mean bias of about 1 ± 40 % (1-σ) is derived.

4.5. Formaldehyde (HCHO)

Since FTIR products are not yet available in the NDACC database, the comparisons involve MAX-DOAS measurements and MACC model simulations for the sole Chinese site of Xianghe. The ground-based, remote-sensing instrument is sensitive to the HCHO abundance in the lower troposphere, up to 1 km altitude. Tropospheric HCHO profiles and columns are validated (up to 3.5 km). It is important to mention here that the model partial column values between the surface and 3.5 km are calculated for the smoothed model profiles (see Fig. 11, left). This guaranties that the model levels where the measurement is not sensitive do not contribute to the observed bias. Due to an instrumental failure, no observations are available for the first half of 2014.

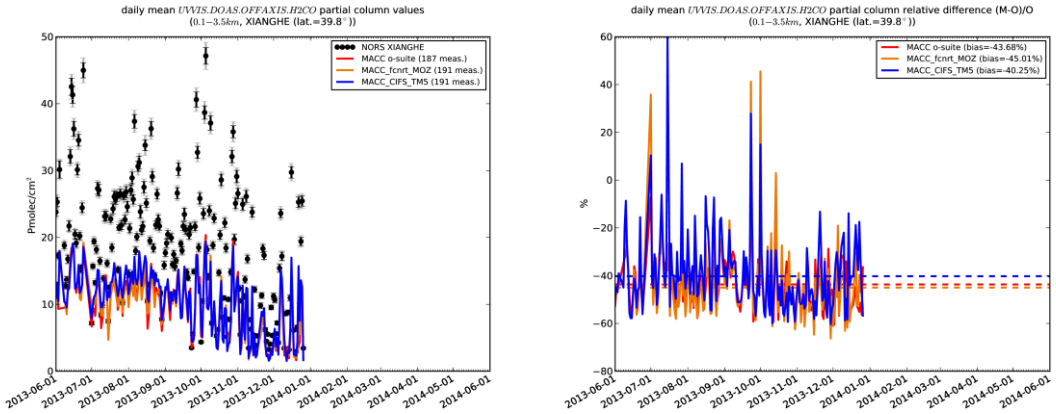

Figure 11. Daily mean relative differences of tropospheric HCHO columns (till 3.5km) by MACC_osuite (fnyp; red), MACC_fcnrt_MOZ (fkya; orange), MACC_CIFS_TM5 (fsd7; blue, full line) compared to NDACC UVVIS DOAS data at Xianghe (39.8°N, 117°E) for the period June 2013-June 2014. The number of measurement days and yearly bias are indicated in the legend. Data are missing after December 2013 due to an instrumental failure.

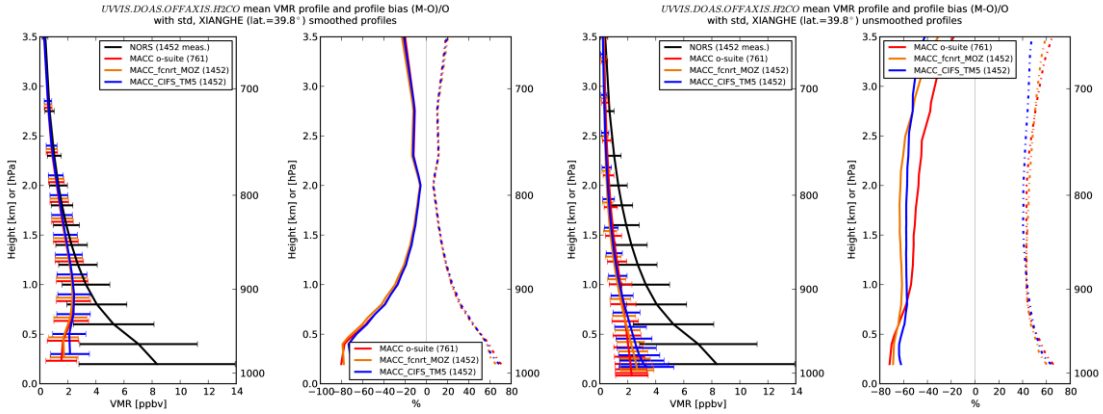

Figure 12. Annual mean tropospheric HCHO profiles by MACC_osuite (fnyp; red), MACC_fcnrt_MOZ (fkya; orange), MACC_CIFS_TM5 (fsd7; blue, full line) compared to NDACC UVVIS DOAS data at Xianghe (39.8°N, 117°E) for the period June 2013-June 2014, smoothed (left) and unsmoothed (right) profiles.

From Figs. 11 and 12 we see that the models underestimate the observation below 1 km. Although the background column values are well captured by the models, the high emission events are not. This might be due to the model's horizontal resolution which is too low to capture the local emission events in this highly polluted area as well. The impact of the clouds on the quality of the UVVIS data and therefore on the agreement with MACC will be also investigated in the future based on the cloud screening method developed in Gielen et al. (2014).

4.6. Aerosol

As for HCHO, the only site contributing measurements is Xianghe while only the fnyp MACC model simulations are available for comparison. Inspection of the averaging kernels indicates that the MAX-DOAS instrument is sensitive to the aerosol optical depth in the lowest layers, with a sensitivity decreasing rapidly above 2 – 2.5 km. Figure 13 displays the available observed and simulated time series for aerosol extinction over the September-2013

to August-2014 time period. Here again, the model demonstrates a good capability to reproduce aerosol background levels, but fails to capture pollution episodes, with in some instances relative differences larger than 200%. Conclusions drawn for HCHO are probably valid for the forecast of aerosol optical depth over a polluted area. As for HCHO and NO_2, the impact of the clouds on the quality of the UVVIS data needs to be investigated.

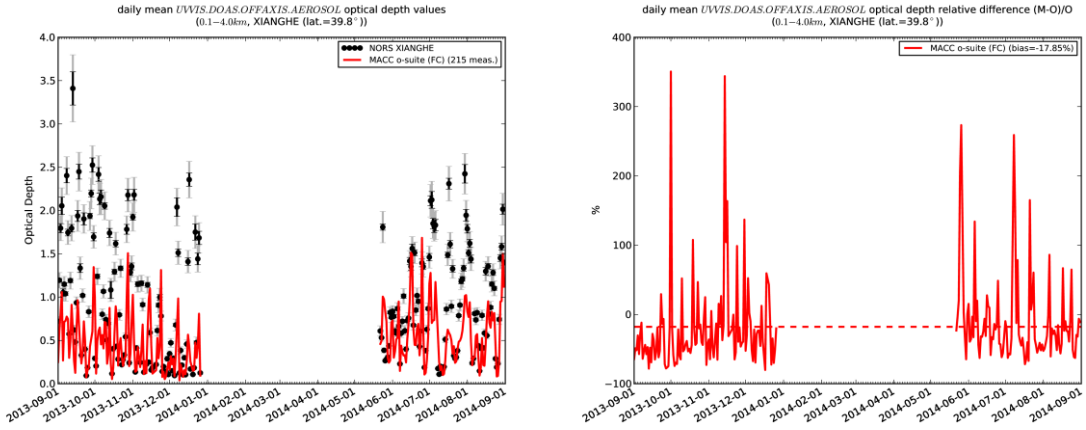

Figure 13. Comparison between MAX-DOAS measurements of aerosol optical depth and fnyp model simulations for the Xianghe site. Data are missing between January and May 2014 due to an instrumental failure.

5. Suitable NDACC products for additional comparisons

In this section we identify two additional NDACC products which might valuably complement the current set of targets for validation of the MACC model simulations.

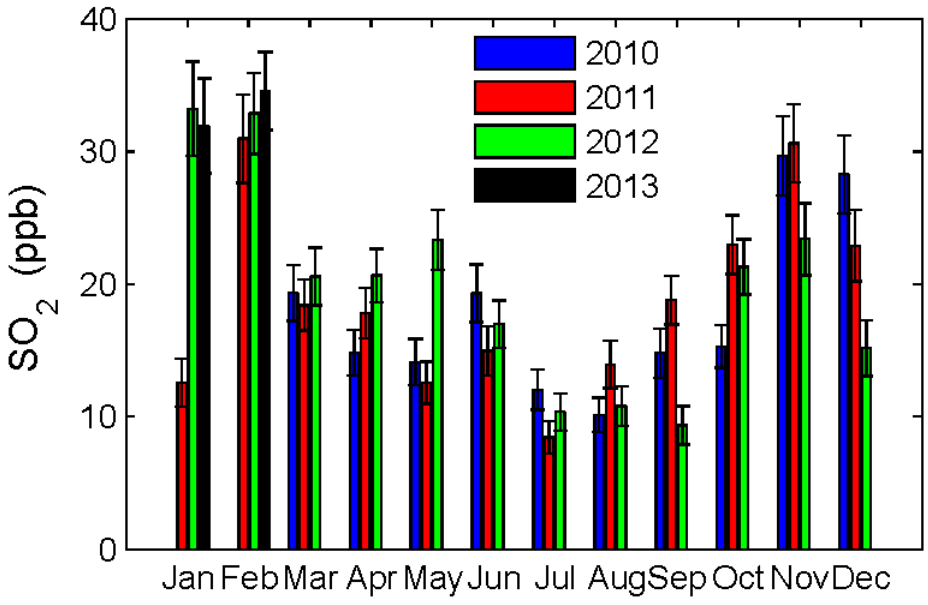

Figure 14. SO_2 concentrations derived from MAX-DOAS observations at Xianghe.

5.1. HCHO from FTIR instruments

Two studies have compared FTIR and MAX-DOAS products for two very different sites, namely Reunion Island (Vigouroux et al., 2009) and the dry unpolluted and high altitude site of the Jungfraujoch (Franco et al., 2014). Both studies conclude that the HCHO products derived from the UVVIS and FTIR techniques are very complementary, with MAX-DOAS providing high sensitivity in the first layers near the Earth surface while the FTIR instruments are sensitive to the whole troposphere. Hence including both techniques, i.e. adding FTIR data to the NDACC targets, would help validating the MACC model forecasts in a broader altitude region than at present. This would also contribute to increase data availability at many more sites, spanning a wide range of atmospheric conditions.

5.2. SO_2 from MAX-DOAS

Sulphur dioxide (SO_2) is one of the main indicators for air quality assessment. Recent developments have allowed the determination of vertical profiles from MAX-DOAS observations. Surface concentrations and vertical columns can be derived. The potential of the method has been demonstrated in a recent study by Wang et al. (2014) in which observations performed at Xianghe since March 2010 are discussed. Figure 14 shows the time series of SO_2 surface concentrations over 2010-2013, with a clear seasonal signal characterized by maximum concentrations in winter, mainly due to domestic heating. It should be noted that the MAX-DOAS method is not applicable in background conditions because of a too low signal-to-noise ratio.

6. Possible improvements of the NORS validation server

Following intensive use of the NORS validation server, we have identified some features that could further expand the server usefulness:

- The output files produced by the NVS and available for download as zip archives have clear filenames, allowing to easily identify the products which are compared. This is however not the case for the files inside the zip archive. After extraction, it is difficult or impossible to identify the file content or to perform automatic access to them with a script. It would be good then to set up a strict naming convention and to apply it systematically, also for the figures and statistics ascii files.
- At the moment, there are problems with comparisons for stations located in valleys (when the MACC model pixel has a bottom altitude significantly higher than the site altitude): this is the case at Xianghe (Beijing) and Bujumbura; an orography with higher resolution would be needed

7. References

Franco, B., F. Hendrick, M. Van Roozendael, J.-F. Müller, T. Stavrakou, E. A. Marais, B. Bovy, W. Bader, C. Fayt, C. Hermans, B. Lejeune, G. Pinardi, C. Servais, and E. Mahieu, Retrievals of formaldehyde from ground-based FTIR and MAX-DOAS observations at the Jungfraujoch station and comparisons with GEOS-Chem and IMAGES model simulations, *Atmos. Meas. Tech. Discuss.*, *7*, 10715-10770, 2014.

Gielen, C., M. Van Roozendael, F. Hendrick, G. Pinardi, T. Vlemmix, V. De Bock, H. De Backer, C. Fayt, C. Hermans, D. Gillotay, and P. Wang, A simple and versatile cloud-screening method for MAX-DOAS retrievals, *Atmos. Meas. Tech.*, 7, 3509-3527, 2014.

Hendrick, F., E. Mahieu, G. Bodeker, K.F. Boersma, M.P. Chipperfield, M. De Mazière, I. De Smedt, P. Demoulin, C. Fayt, C. Hermans, K. Kreher, B. Lejeune, G. Pinardi, C. Servais, R. van der A, J.-P. Vernier, and M. Van Roozendael, Analysis of stratospheric NO_2 trends above Jungfraujoch using ground-based UV-visible, FTIR, and satellite nadir observations, *Atmos. Chem. Phys., 12*, 8851-8864, 2012.

Schanz, A., K. Hocke, N. Kämpfer, S. Chabrillat, A. Inness, M. Palm, J. Notholt, I. Boyd, A. Parrish, and Y. Kasai, The diurnal variation in stratospheric ozone from the MACC reanalysis, the ERA-Interim reanalysis, WACCM and Earth observation data: characteristics and intercomparison, submitted to *Atmos. Chem Phys.*, 2014.

Taylor, K. E., Summarizing multiple aspects of model performance in a single diagram, *J. Geophys. Res., 106*, 7183-7192, 2001.

Vigouroux, C., F. Hendrick, T. Stavrakou, B. Dils, I. De Smedt, C. Hermans, A. Merlaud, F. Scolas, C. Senten, G. Vanhaelewyn, S. Fally, M. Carleer, J.-M. Metzger, J.-F. Müller, M. Van Roozendael, and M. De Mazière, Ground-based FTIR and MAX-DOAS observations of formaldehyde at Réunion Island and comparisons with satellite and model data, *Atmos. Chem. Phys.*, *9*, 15891-15957, 2009.

Wang, T., F. Hendrick, P. Wang, G. Tang, K. Clémer, H. Yu, C. Fayt, C. Hermans, C. Gielen, G. Pinardi, N. Theys, H. Brenot, and M. Van Roozendael, Evaluation of tropospheric SO2 retrieved from MAX-DOAS measurements in Xianghe, China, *Atmos. Chem. Phys.*, *14*, 6501-6536, 2014.