GEMS GRG Comprehensive Validation Report

With contributions from :

BIRA-IASB, Brussels, Belgium

CNRM, Météo-France, Toulouse, France

CNRS-LA, Toulouse, France

DWD, Hohenpeissenberg, Germany

ECMWF, Reading, UK

FZJ, Jülich, Germany

IUP, Bremen, Germany

KNMI, De Bilt, Netherlands

NKUA, Thessaloniki, Greece

SA-UPMC, Paris, France

13/08/2009

cite as :
Cammas, J.-P., A. Gilles, S. Chabrillat, F. Daerden, N. Elguindi, J. Flemming, H. Flentje, C. Granier, V. Huijnen, A. Inness, L. Jones, E. Katragkou, F. Khokhar, L. Kins, K. Law, K. Lefever, J. Leitao, D. Melas, P. Moinat, C. Ordonez, V.-H. Peuch, G. Reich, M. Schultz, O. Stein, V. Thouret, T. Werner, C. Zerefos, GEMS GRG Comprehensive Validation Report, Toulouse, 2009. Available as project report at http://gems.ecmwf.int.

1. Introduction

The GEMS project has created a new European operational system for operational global monitoring of atmospheric chemistry and dynamics and an operational system to produce improved medium-range & short-range air-chemistry forecasts, through much improved exploitation of satellite data. The GEMS consortium comprises 28 partners among which regional centres and leading European labs. Labs have research capabilities & models on all aspects of atmospheric chemistry, capabilities for global operational weather capabilities (ECMWF), and capabilities for global diagnostic capabilities (EU's JRC). Under the support from the 6th EU framework programme (2006-2009), GEMS has developped sub projects on greenhouse gases (GHG), global reactive gases (GRG), aerosols (AER) and regional air quality (RAQ) in Europe, together with system integration and overall validation.

The GRG sub project has developed a coupled model approach with data assimilation and transport in the ECMWF Integrated Forecasting System (IFS) and chemistry (and emissions) in one of three chemistry transport models. The three CTMs are MOCAGE by Meteo France (Josse et al., 2004 ; Bousserez et al., 2007), MOZART-3 by NCAR, MPI Hamburg, FZ Juelich (Horowitz et al., 2003 ; Kinnisson et al., 2007) and TM5 by KNMI (version KNMI-cy3-GEMS, Krol et al., 2005). Three CTMs were selected because previous model intercomparison studies showed considerable spread of results, and a 3-model ensemble can provide some guidance with respect to the robustness of the simulation results. A summary of the different CTM specifications can be found in Annex 1, Table 1. For technical reasons, MOZART was selected as primary CTM in the coupled set-up for the routine forecasts and analyses. GRG covers stratospheric and tropospheric ozone and other trace gases, notably CO, NO_x, CH_2O and SO_2.

Within the GRG subproject, the system set-up has initially involved uncoupled ("offline") simulations. These were used as benchmark for coupled system and for case studies. First coupled simulations led to first reanalysis. A second reanalysis was started in **** and ran jointly with GHG and AER reanalyses covering the period **** to ****. Routine forecasts since May 2007 with some upgrades occurring in between.

Validation of the model system is an essential part of an integrated forecasting system. GRG has taken the approach of "distributed and independent" validation, i.e. all model results were evaluated by the providers of observational data, who are more knowledgeable about the strengths and weaknesses of their data sets and should show less preference of one model over another one.
In contrast to other, more scientifically oriented projects, GEMS had to place a strong emphasis on "readiness for operations". With respect to the validation activities this implied on the one hand that some efforts were directed towards the development of automatic software applications which allow fast turnaround times, and on the other hand towards the development of quantitative and objective validation metrics ("skill scores") which shall allow for a fast assessment of model performance and rapid detection of sudden performance drops due to model version changes or other factors. It should be noted that such scores present a novel approach in the field of global atmospheric chemistry modelling and the definitions laid out in this document are not yet finalized for operational purposes.
Another implication of the quest for future operational forecasting is the need for realtime (or near-realtime) observational data. This aspect has also been covered in the GEMS GRG sub project, but is covered elsewhere (see report *****).

This report describes the validation activities performed within the GEMS GRG sub project. It is structured according to the data sets used. Each chapter contains an independent evaluation for one data set. A general discussion is given in chapter 13 and final conclusions are presented in chapter 14. For complete information regarding the Chemistry Transport Models (CTMs) used in the GEMS-GRG work, one should consult Annex 1 for the general CTM specifications, Annex 2 for MOZART, Annex 3 for TM5, Annex 4 for MOCAGE, and Annex 5 for the coupled system of the IFS model with one of the three previous CTMs. Annex 6 describes each of the model simulations used in this evaluation. Annex 7 describes the common part of the scoring methods used for evaluation.

2. Evaluation with satellite datasets

Contributors: Leitao Joana, Andreas Richter (IUP, Bremen, Germany), Vincent Huijnen (KNMI), Fahim Khokhar (SA-UPMC), and the modeller teams

In order to evaluate the outputs of the different Chemical Transport Models (CTMs) used within the Global Reactive Gases subgroup (GRG) of the GEMS project – MOCAGE, MOZART and TM5 –, data calculated by these models were compared with measurements of NO_2 and CO retrieved from several satellite instruments: SCIAMACHY, OMI and MOPITT. Here, we present the results for both tropospheric and stratospheric columns. The respective methods of retrieval of satellite data and analysis of model results are described in the sections below.

In this report, we also compare the tropospheric NO_2 columns as determined by the models from the GRG and RAQ-subgroups (see section 2.5), to the measured OMI-NO_2 column data. Tropospheric NO_2 is one of the key trace gases both in RAQ and GRG, but the focuses in the two subgroups are different. Whereas a RAQ model aims at modeling the tropospheric trace gases over Europe on a very high resolution (temporal/spatial), in order to predict boundary layer concentrations, the GRG models are more interested in the evaluation of good background values and correct global budgets, so that they can deliver appropriate boundary conditions to RAQ models. The temporal and spatial resolution in the GRG models is naturally coarser as compared to the RAQ models. Still it is interesting to compare the GRG models to the RAQ models.

2.1 Datasets and methodologies

2.1.1 Datasets

Global MOCAGE model:

For the year 2003, data of the runs v1 (*60LEV01)* and v2 (*60LEV02)* were compared with satellite results. From this model (as for TM5), only datasets of standalone runs were available for an evaluation.

MOCAGE V2 is the name for standalone reference run for the year 2003 and 2004. It is available since June 2008. 8days GFEDv2 for CO fire emissions inventories are used.

Global MOZART model:

From this model several datasets were analyzed and in this report we present data for:

- standalone runs: *V1, V7* and *V10* for the year 2003, and *V7, V8, V9* and *V10* for the year 2004;
- reanalysis run: *expid f026*, from May 2003 to September 2006;
- forecast run: *expid ez2m* for the period of February-December 2008.

As mentioned in Annex 5, the MOZART-IFS forecast run (experiment *ez2m*) has run continuously since the beginning of 2008, delivering global forecasts of trace gases up to three days ahead. This experiment is based on a free-running coupled system, i.e. without data assimilation.

The chemical solver is based on the MOZART-tracer model. It applies a gaussian grid with a resolution of about 1.9 deg lat/lon. MOZART applies a distribution of 60 layers, with the top layer at 10 hPa. The emission inventory is based on the RETRO-2003 inventory created for the GEMS-GRG project. For the biomass burning inventory a ten-year averaged (1997-2006) database is used for the forecast-run.

V10 is the latest stand alone reference run for the year 2003. GRG anthropogenic emission inventories and 8days GFEDv2 for fire emissions inventories are used and confined to the lower MOZART level. The simulation results are available since December 2008. The new JPL06 evaluation of the reaction CO + OH, leading to a reaction rate which is about 10-20% lower than before in the troposphere, is introduced in the *V10* simulations.

F026 second reanalysis run includes fixes (i-IFS convection and diffusion for CO, CTM convection and diffusion for GO_3, NOx, Coupled to MOZART. ii-Two-way coupling for CO and GO_3. iii- NOx and HCHO initialized from CTM in every forecast. iv- Bugfix for CTM fluxes) from **eyih** (1st reanalysis run – with CO (MOPITT) and GEMS ozone assimilation (SCIAMACHY, MIPAS, GOME, SBU) and it is first 1st GEMS reanalysis (GRG together with GHG and AER). This run is available for 2003 -2007. Monthly GFEDv2 for fire emissions inventories are incorporated in MOZART- *F026* simulations and confined to the lower MOZART level.

Global TM5 model:

From this model, mainly standalone runs were considered for the evaluation study: data of version *V7* for the year 2003 and data of versions *V9* and *V10* for summer 2004 were analysed. In addition, for the comparison with the RAQ model results, a special reanalysis run (Y08) of the TM5 model (Krol *et al.*, 2005) has been setup, which uses the operational meteorological fields from ECMWF. The horizontal resolution is still 2° x 3° lat/lon but, for this particular evaluation, the model applies 34 model layers with the top layer at 10 hPa. The chemistry is based on a modified CBM-IV scheme, which is updated to the latest recommendations from JPL. The model version that is used is mainly identical to the TM5-V10 version that is used for the 2004 standalone evaluations presented in section 2.4.2 in this chapter. It also applies the same emission inventory over Europe as the coupled MOZART experiment, but has enhanced emissions over China and South-east Asia (the REAS inventory). A five-year average biomass burning inventory (2001-2006) from GFED-v2 is used.

SCIAMACHY – IUP-Bremen

SCIAMACHY is a 8 channel UV/vis/NIR spectrometer launched on ENVISAT in March 2002. It observes the atmosphere in nadir, limb and solar and lunar occultation geometry. For this work, only the nadir measurements are used. In nadir mode, SCIAMACHY achieves global coverage within about 6 days with a spatial resolution of 60 x 30 km^2. The local overpass time of SCIAMACHY is at about 10:00 LT for low and midlatitudes.

The IUP Bremen SCIAMACHY product consists of two parts – a tropospheric column product and a stratospheric column product which are compared to the respective sub-columns of the model fields.

NO_2 slant columns are retrieved with the DOAS technique in the wavelength region 425 – 450 nm. For the stratospheric product, they are then converted to a vertical column applying an airmass factor based on the US-standard atmosphere with the tropospheric part removed. No screening for clouds or pollution is applied, and as a result, the values over polluted regions are too large. Details on the retrieval can be found in Richter and Burrows (2002).

For the tropospheric columns, the average of the NO_2 slant columns measured on the same day at the same latitude over the Pacific sector (180° – 220° longitude) are subtracted to derive the tropospheric slant columns. Vertical tropospheric columns are then computed using airmass factors based on a MOZART run for 1997 and averaged over months. Tropospheric columns are filtered for cloud fractions smaller than approx. 30% but no further cloud correction is applied. As a result of the simplified correction of the stratospheric contribution, measurements in winter and spring at high latitudes have larger uncertainties, in particular if the polar vortex is displaced from the Pole. Also, as the difference relative to the Pacific is

taken, the background NO_2 column also present over remote regions is missing and negative columns can occur over clean areas. Details on the tropospheric retrieval can be found in Richter *et al.* (2005).

The HCHO product is derived from SCIAMACHY spectra in a similar procedure than that for NO_2. First, a DOAS algorithm to derive HCHO slant columns from earth irradiance spectra measured by SCIAMACHY has been developed. A spectral fitting window of 334.3 - 348.5 nm was selected. The retrieved slant columns were normalized by assuming a mean value of $3.5x10^{15}$ molec/cm^2 in the region between 180° - 200° E. This is necessary to compensate for offsets introduced by the solar reference measurements. The next step is the conversion of the slant columns to tropospheric vertical columns using appropriate airmass factors (AMF) based on calculations with the full spherical radiative transfer model SCIATRAN (Rosanov et al., 2005). Since the AMF for satellite observations depends strongly on several meteorological boundary conditions it is necessary to account for all the parameters in the radiative transfer calculations. In order to obtain a realistic global picture of HCHO, the most practical way is to create pre-calculated tables of AMF for a set of SZAs which depend on time and location and are interpolated to the conditions of an individual measurement. In order to be able to investigate the satellite measurements also on a regional scale, the spatial resolution for the AMF tables was selected to be 0.5° x 0.5° in latitude and longitude.

Some information on parameterisation and conditions of the data sets for the AMF calculation is given below:

1. Use of the surface albedo climatology of Tanskanen *et al.* (2003).
2. Use of a variable tropospheric aerosol loading
3. Correction for topography effects (TerrainBase, Row *et al.*, 1994)
4. Use of four different HCHO profiles depending on origin (anthropogenic, background, biogenic, biomass burning)
5. Use of a 20% cloud cover threshold based on intensity criteria

More details are given in Wittrock (2006) and Wittrock *et al.* (2006).

OMI – KNMI

The retrieval algorithm for tropospheric NO_2 columns from the Ozone Monitoring Instrument (OMI) is described in Boersma *et al.* (2007). This method is based on the combined retrieval-assimilation-modeling approach developed at KNMI for off-line tropospheric NO_2 from the GOME and SCIAMACHY satellite instruments and uses an off-line system that provides a priori information on the profile shapes and stratospheric background NO_2. Slant columns for NO_2 are retrieved using differential optical absorption spectroscopy (DOAS) in the 405– 465 nm range. Cloud fraction and cloud pressure are provided by a new cloud retrieval algorithm that uses the absorption of the O_2-O_2 collision complex near 477 nm. Online availability of stratospheric slant columns and NO_2 profiles is achieved by running the TM4 chemistry transport model (CTM) forward in time based on forecast ECMWF meteo and assimilated NO_2 information from all previously observed orbits. OMI NO_2 slant columns, after correction for spurious across-track variability, show a random error for individual pixels of approximately 0.7×10^{15} molec/cm^2. Cloud parameters from OMI's O_2-O_2 algorithm have similar frequency distributions as retrieved from SCIAMACHY's Fast Retrieval Scheme for Cloud Observables (FRESCO) for August 2006. On average, OMI cloud fractions are higher by 0.011 and OMI cloud pressures exceed FRESCO cloud pressures by 60 hPa. OMI overpass time is at 13:30 h local time. For the evaluation of tropospheric NO_2 only pixels with less than 50% cloud cover are used.

For the second half of 2008 the OMI instrument suffered from some failures in retrievals for a number of tracks. The erroneous tracks have been filtered out in the evaluation procedure. In case of multiple retrievals at the same day for the same location a weighting of the retrieval pixel has been applied based on the squared cosine of the viewing angle. In this way more weight is given to retrievals with small angles, and high resolution.

MOPITT

Measurements Of Pollution In The Troposphere (MOPITT) is an instrument onboard NASA's EOS Terra spacecraft launched in December 1999. It is primarily designed for measuring the global distributions of carbon monoxide (CO) and methane (CH_4) in the troposphere (Deeter *et al.*, 2003). MOPITT operates by sensing infrared radiation from either the thermal emission/absorption at 4.7 μm for CO profiles, or reflected sunlight at about 2.2-2.4 μm for CO and CH_4 column measurements in daylight. The use of solar channels enhances the instrument sensitivity to the atmospheric boundary layer. The satellite is deployed in a polar synchronous orbit with 10:30 am local equator crossover time. Spatial resolution is 22 km x 22 km, and cross-track scanning achieves approximate global coverage in 3 days.

MOPITT retrievals are reported on 7 vertical levels (surface, 850, 700, 500, 350, 250, and 150 hPa), and as a total column, for all cloud-free scenes. The measurement technique relies on thermal contrast between the surface and the atmosphere, leading to a retrieval dependence on surface temperature, and little sensitivity to CO in the boundary layer.

2.1.2 Methodologies

As mentioned above, for both MOZART and TM5 data were available for the years 2003 and 2004, while for MOCAGE only the year 2003 could be evaluated. The complete analysis is not included here, as we want to limit ourselves in this final report to provide a comprehensive evaluation and present only those graphs that are most important for the conclusions achieved. More detailed graphs can be found in Annex 8 with the NO2 monthly scores calculated for each of the model versions evaluated when compared to SCIAMACHY data, for both stratosphere and troposphere.

To compare the model results with the satellite measurements of stratospheric and tropospheric amount of a trace gas (either NO_2 or CO), different types of analysis were considered. To be more precise, in the case of the satellite measurements, the stratospheric columns are in fact total columns, i.e., a weighted part of the tropospheric NO_2 is also included in this amount. As a result, a comparison of the stratospheric fields cannot be done directly but mostly qualitatively. However, the impact of tropospheric NO_2 on the compared averages is small in comparison to the differences between model and measurement as discussed later.

In this chapter, we present 2D global maps for 3-month averages that allow for a quick look comparison of the overall results. In addition, the seasonality was studied for specific pre-defined regions. In this step, it is possible to identify if the model is capturing the correct annual/monthly cycle of the considered compound. In the case of the stratosphere, areas were divided into:

- North and South Pole: 90°-60° latitude, North and South respectively;
- Midlatitudes North and South: 60°-30° latitude, North and South respectively;
- Tropics: 30°N-30°S latitude.

For the troposphere; regions representative of polluted areas and biomass burning locations were selected (see figure 2.1 below).

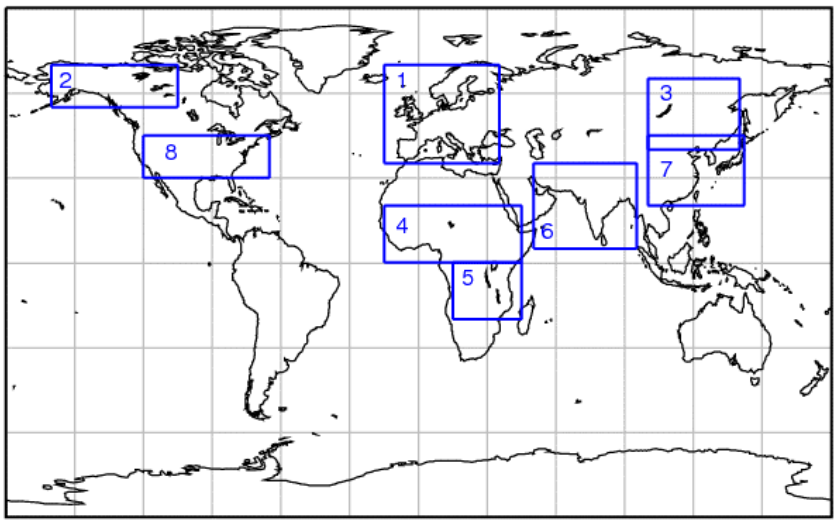


Figure 2.1: Map with pre-defined regions selected for the evaluation of tropospheric columns of model results: 1) Europe, 2) Alaska, 3) Siberia, 4) Northern Africa, 5) Southern Africa, 6) South-Asia, 7) East-Asia and finally 8) US. Regions 1, 6-8 (green outlines) correspond to normally polluted places and regions 2-5 (red outlines) refer to typical biomass burning areas.

It is worth mentioning that, for the evaluation of the tropospheric columns, when comparing model data with SCIAMACHY measurements, only data over continents were considered and, therefore, mainly the NO_2 over emission sources is being compared, excluding the oceanic regions with very low columns and occasional effects of long-range transport. Scatter plots are also presented to see how good the model tropospheric columns (in each of the regions) correlate with the satellite measurements. Finally, scores based on the bias between model and measurement, according to the method described in the following section, are presented in different forms: maps of regions, average per region for different months or averaged in certain time periods or for each of the regions defined above. This score is constructed such that it reaches unity for low bias between modeled columns and observations.

In the context of the GEMS-project, the document 'Scoring approaches for the GEMS project' was published. In this document, the application of uniform scoring approaches within the work developed in the different sub-groups was suggested (see Annex 7). The aim of the scores is to offer the user a sense of the quality of the model results. Therefore, one can say that a score is mainly "user-oriented", i.e. it is not primarily used to express the quality of physical processes, such as variables for the model bias, temporal/spatial correlation, but merely the quality of model behavior for a specific (combination of) parameter(s) over a specific period and region.

It is important to define a score based on underlying physical parameters that may reflect the user needs as good as possible. Further on, scores should be robust, i.e. not sensitive to the quality of measurement data, and they should be bounded between 0 and unity. In this way it is possible to combine scores in order to obtain the most relevant final score. The application of the same score over a longer period allows for documentation of model improvement.

The score is typically based on the model bias compared to independent observations. Additionally, as suggested by Kinne (2008), the score may contain information on the spatial correlation as well as on its ability to reflect the seasonal change. For any of these sub-scores a treatment should be provided, while finally the scores can be combined to yield a single number.

In this report a score for modeled trace gas columns using satellite retrievals is described, which is based on the average local model bias B_j at location j:

$$B_j = \left(\frac{1}{T} \sum_T \left(M_i - O_i \right) \right)_j \tag{1}$$

where:

- M_j is the modeled column at time t (T elements),
- O_j the corresponding satellite retrieval.

An averaged bias can be obtained by averaging over a predefined region and time frame. To create a score that is bounded between 0 and unity, normalization of this bias is required.

Standard scoring approaches

In the GEMS-document on scoring (see Annex 7), one can find several general scores that are based on a bias measures, namely scores based on the modified normalized mean bias (MNMB) and the fractional gross error (FGE):

$$MNMB = \frac{2}{N} \sum_j \frac{M_j - O_j}{M_j + O_j} \tag{2}$$

$$FGE = \frac{2}{N} \sum_j \left| \frac{M_j - O_j}{M_j + O_j} \right| \tag{3}$$

Note that the MNMB is limited between -2 and 2, whereas the FGE is bounded between 0 and 2. A score can then be constructed for instance as $S = 1 - FGE/2$. However, one of the main shortcomings of these scores is that they assume negligible uncertainty in the measurement error. This might be problematic when the reference measurements are not perfect themselves, which is the case when using satellite retrievals of tropospheric trace gases.

From analysis performed, it was revealed that for small measurement informations the FGE and MNMB do not reach zero, but rather their (absolute) maxima. This is because retrieved column data contains scatter that is of the same order of magnitude (and larger) as the observation. The observations can even yield negative values, which cannot simply be neglected. This makes the MNMB and FGE ill-posed for this purpose. When the denominator gets very small the FGE can become large, which in turn leads to a very low local contribution to the area-averaged score. Therefore, based on the points enumerated above, the use of the FGE/MNMB is not practical for the analysis intended with this evaluation work.

We also choose to normalize the bias using the observations only, such that normalization for the different models is treated with the same values. So the score is based on a formulation that can be written as:

$$S = 1 - \frac{1}{N} \sum_j \left| \frac{B_j^*}{O_j^*} \right| \tag{4}$$

where:
- B_j^* is a measure that represents the local *average* model bias,
- O_j^* represents the local *average* observation.

The summation expresses the procedure for the area-averaging of the score.

When working with satellite data as a reference for evaluation of model results the situation is slightly more complex and, therefore, the above Eq. (4) cannot be used directly. Some adaptations are required and in the next subsections a new score will be introduced based on this equation, which fulfills the criteria as described above: for small biases the score should go to zero, it should be robust and it should reflect what the user is requesting.

The use of medians

In order to create a score that is robust, for a certain class of retrieval products it makes sense to use statistics based on medians for B and O, instead of means. In this way, we hope to remove any outliers from the evaluation of statistics. The use of medians is particularly relevant in measurement observations that cannot be described by a Gaussian distribution, but contain outliers. However, this approach does not fully deal with the regular (Gaussian) scatter that is present in most satellite retrievals. Together with the limited number of measurement data (often, in some months and on some locations, less than 5 observations are available) any score based on the FGE would still be spoiled by very small and/or negative average observations, even when using medians.

Accounting for measurement errors

Following the attempt to obtain a robust score, one needs to account for the measurement error ε. This can be achieved by limiting the minimum magnitude of the local average observations to ε:

$$O_j^* = \max(O_j, \varepsilon) \tag{5}$$

Based on the idea that the measurements of total columns can only be known with a limited accuracy, it is necessary to take this uncertainty into account not only in the evaluation of the denominator O_j^* but also in the (median) bias B_j^*. Hence, we assume that the local average model bias can only be evaluated with a limited accuracy. A simple way to limit the model bias is by assuming that any bias cannot get below the accuracy-level of the measurement, written as:

$$B_j^* = \max(\left|B_j\right|, \varepsilon) \tag{6}$$

Note that taking the absolute value implies that we are considering an (absolute) error scale, where positive and negative biases will not cancel out, similar to a score based on the FGE. However, in the above description of the bias, the score can never reach unity. To solve this issue, we need to ensure that the bias can become zero. The simplest way to do so is by subtracting the measurement error again:

$$B_j^* = \max(\left|B_j\right|, \varepsilon) - \varepsilon \tag{7}$$

Note: by decreasing the bias, proportional to the measurement error, the score artificially increases. It is found that when accounting for the measurement error it is no longer necessary to adopt the use of median biases rather than normal biases.

Representativity of the score

To ensure a good range for the model bias compared to the observation magnitude, a proportionality scale is introduced. In this way, the score can be tuned to the user's needs. Our judgment is that a model bias B_j^* of the size of the observation O_j^* should lead to a score of $S=0.5$. Therefore, the proportionality of the bias to the observation should be a factor 2. The proportionality factor may also be adapted, in case of less (or better) constraint model behavior. Based on that, the score is finally written as:

$$S = 1 - \frac{1}{N} \sum_j \left| \frac{B_j^*}{2O_j^*} \right| \tag{8}$$

where:

- B_j^* and O_j^* are defined as in Eqs. (7) and (5), respectively.

In principle, this score can still become negative, when averaging over a small region with a very large model bias. Strictly considered, the score should therefore always be limited to positive values. However, in practice, negative scores will never occur when taking averages over a larger region, and when a proper choice for the error scale and for the proportionality factor is made.

2.2 Offline simulations

2.2.1 CTMs vs SCIAMACHY

a) MOCAGE

We present the 3 month average global maps of NO_2 only for half a year since the model performance is quite similar in the rest of 2003. Comparing the NO_2 stratospheric fields of the model and the satellite measurements presented in figure 2.2 below, one can quickly see that they differ significantly. Actually, the MOCAGE V02 stratospheric results for NO_2 columns are much lower than those obtained from MOZART and TM5 (see Figs. 2.4, 2.5, 2.8 and 2.9).

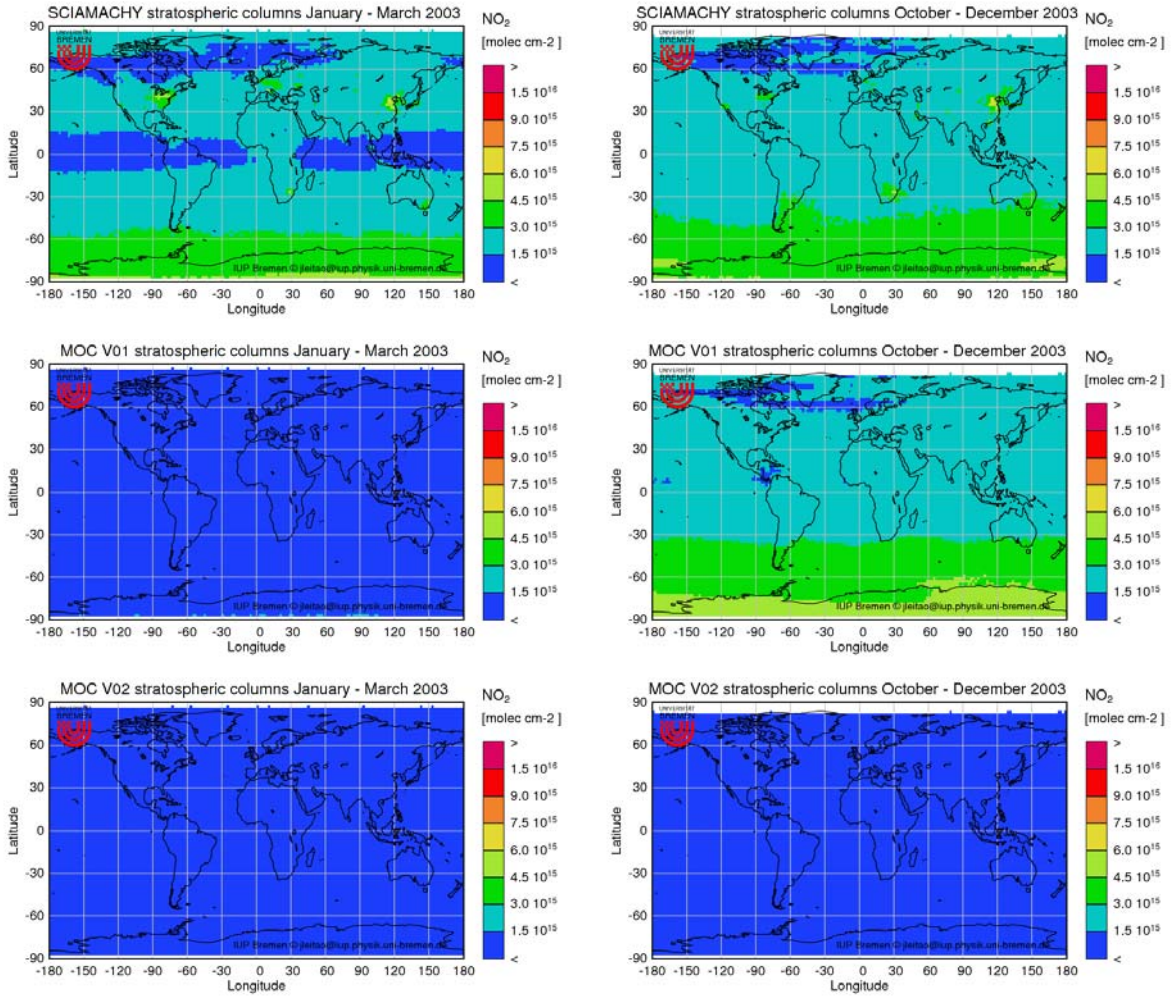


Figure 2.2: Three month averages of global stratospheric NO_2 determined by MOCAGE V01 (middle), MOCAGE V02 (bottom) and total column of NO_2 measured by SCIAMACHY (top) for January – March (left) and April – June (right) of the year 2003.

The main conclusion drawn from this analysis is that MOCAGE V01 performs better that the succeeding MOCAGE V02 run. Some of the stratospheric NO_2 fields from MOCAGE are so low that all values are in the lowest color bin when using the appropriate color scale for comparisons of the measured annual cycles. When analyzing the data from V02 with a different color scale it is possible to see that only a small part of the overall pattern of NO_2 concentration is well captured. In the beginning of the year both versions failed to model the stratospheric layer, however, in the later months V01 begins to achieve better results with a very good comparison for the last 3 months.

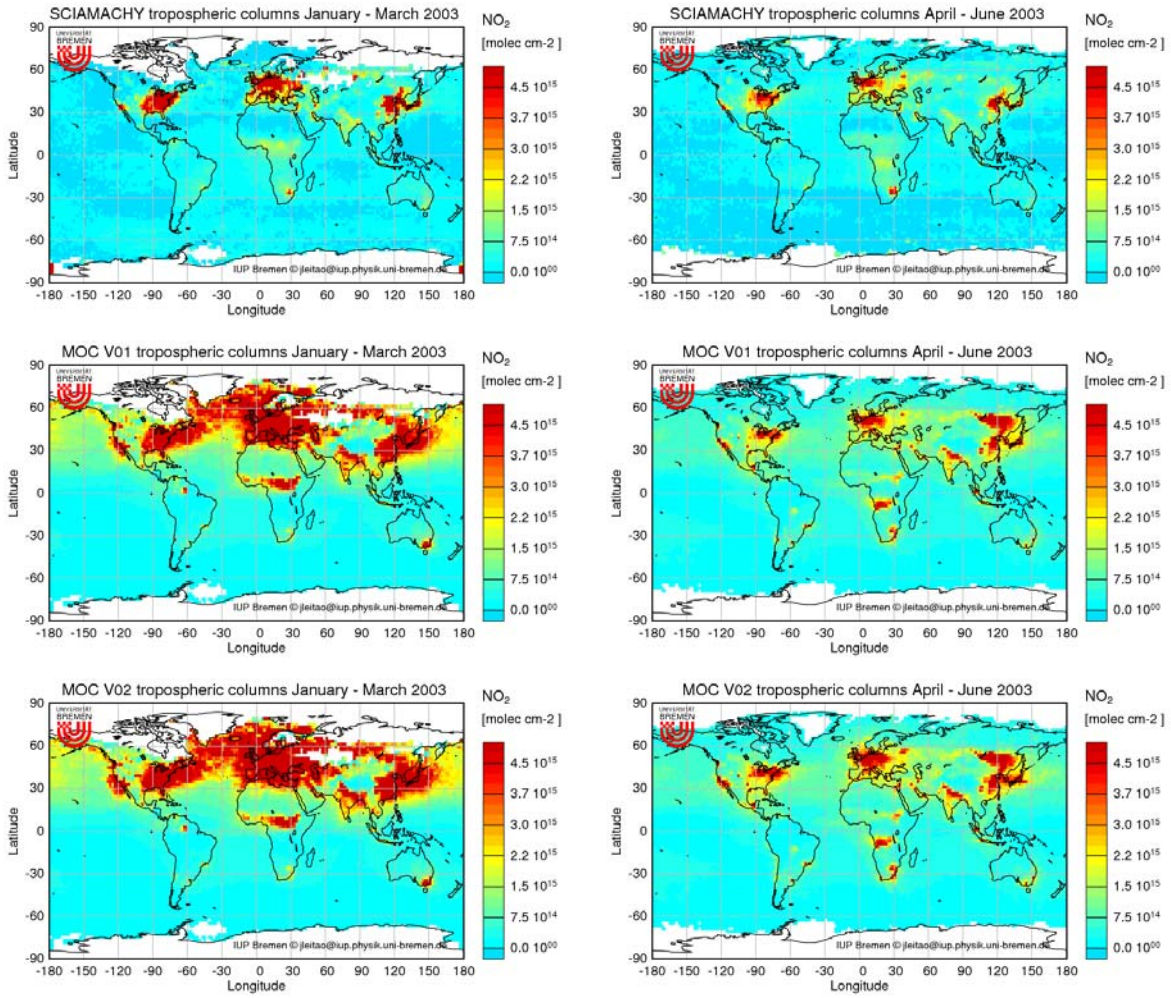


Figure 2.3: Three month averages of global tropospheric NO_2 determined by MOCAGE V01 (middle), MOCAGE v02 (bottom) and measured by SCIAMACHY (top) for January – March 2003 (left) and April – June 2003 (right).

Regarding the tropospheric values, in general MOCAGE overestimates the winter values in the Northern hemisphere, but in the summer months the comparison looks better. However, over fire regions in Siberia and in Africa the NO_2 values estimated by MOCAGE are much higher than those retrieved from the satellite data. For the troposphere, there are no significant differences between the two versions of MOCAGE used.

b) MOZART

Fig. 2.4 represents 3 month averages of stratospheric NO_2 calculated by MOZART for the first and last trimester of 2003, and for the second half of 2004 in Fig. 2.5. As mentioned before, the SCIAMACHY data correspond to the total columns of NO_2. Only part of the year is presented because the results and conclusions are the same for the rest of the year. The tropospheric maps are shown in Figs 2.6 and 2.7.

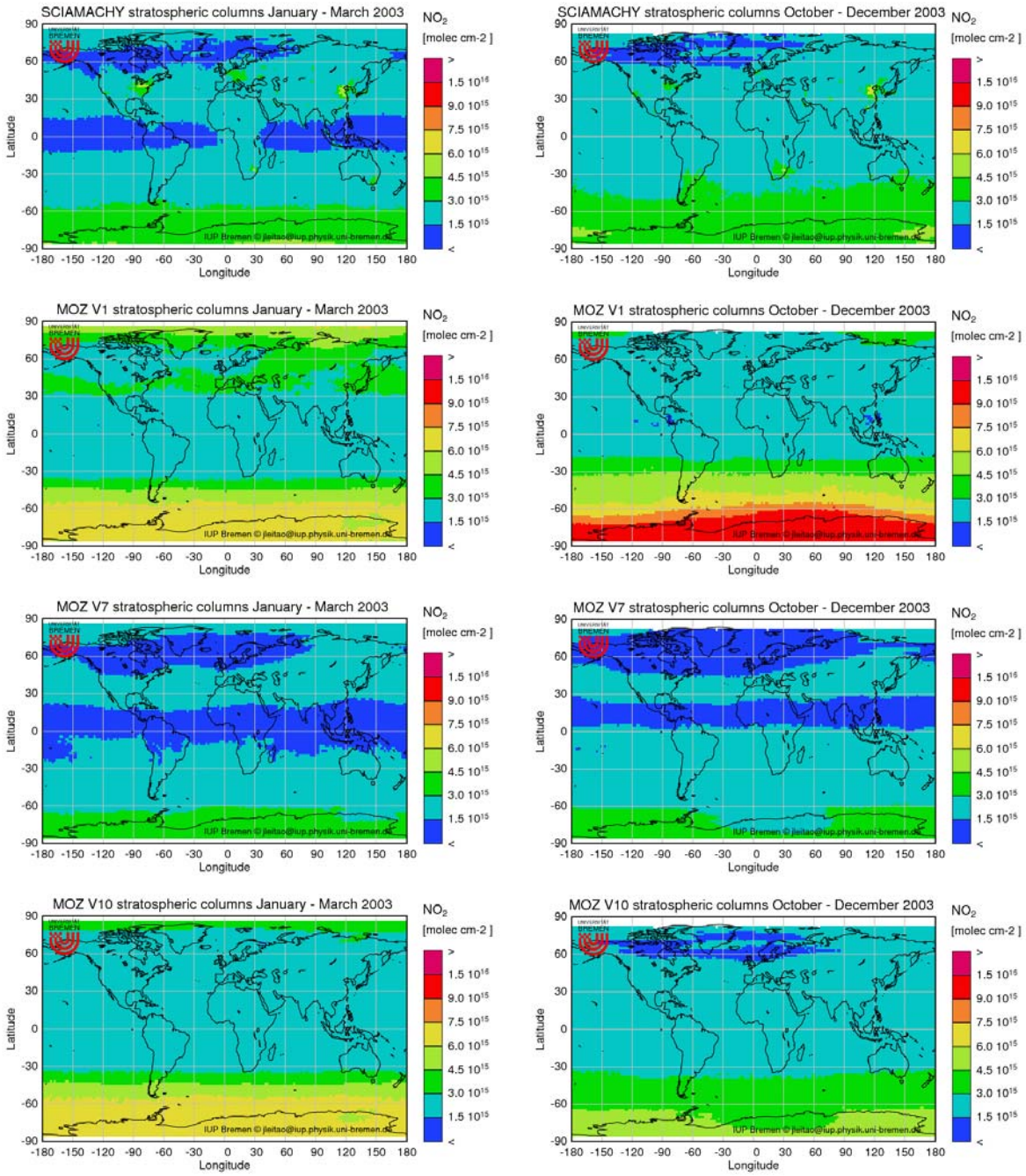


Figure 2.4: Three month averages of global stratospheric NO_2 determined by (from second of the top to bottom) MOZART V1, V7 and V10 and total column of NO_2 measured by SCIAMACHY (top) for January – March 2003 (left) and October – December 2003 (right).

From the results presented above one can see that the initial version of MOZART was not performing well regarding the stratospheric NO_2. The standalone V7 shows NO_2 fields that agree nicely with the satellite data, both in terms of spatial distribution and amounts calculated. However, the latest version V10 again overestimates the NO_2 columns in the stratosphere, at least during the spring.

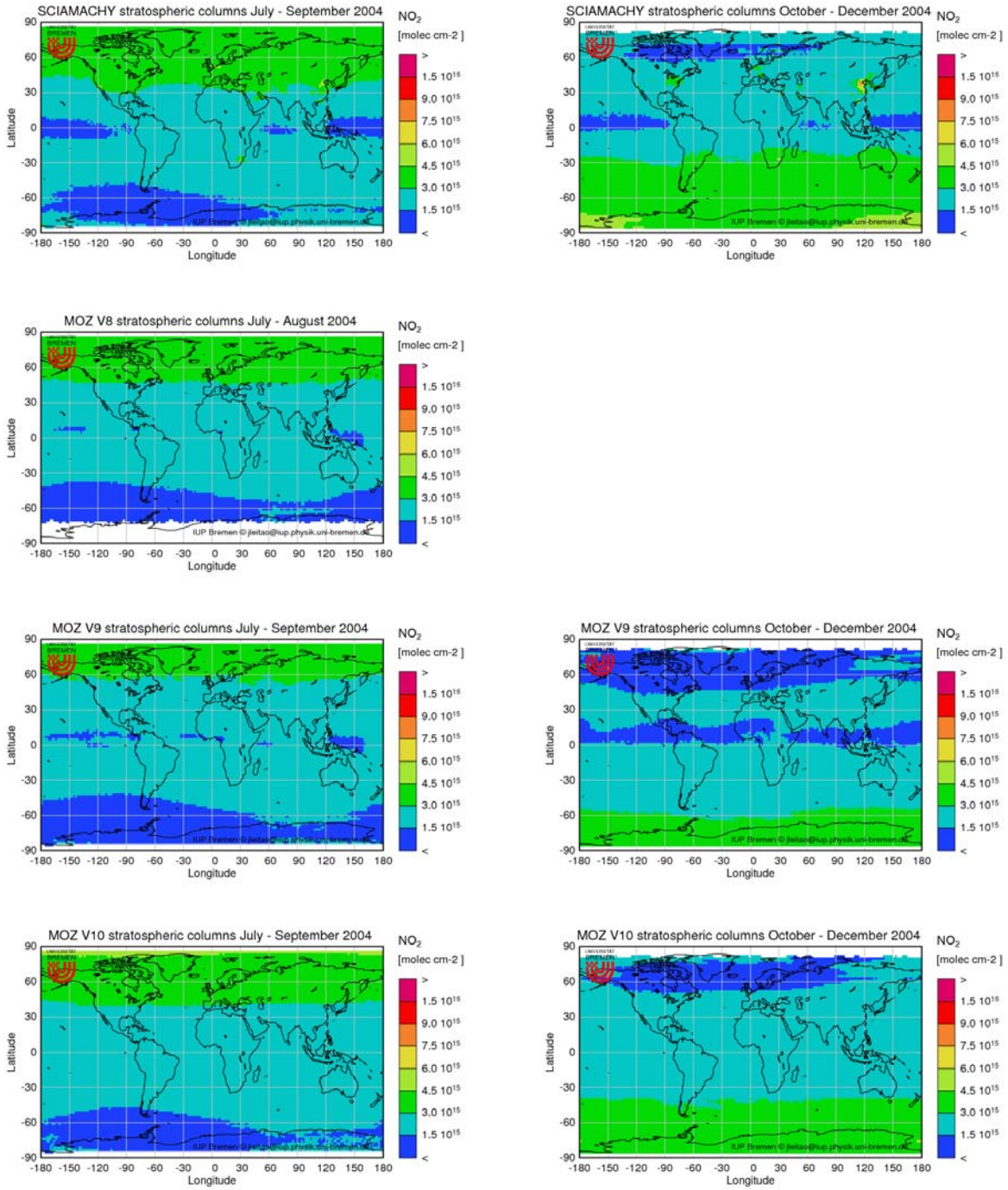


Figure 2.5: Three month averages of global stratospheric NO_2 determined by (from second of top to bottom) MOZART V8 (data only for July and August), V9 and V10 and NO_2 measured by SCIAMACHY (top) for July – September 2004 (left) and October – December 2004 (right).

In the above Fig. 2.5, only data for MOZART V8, V9 and V10 are presented, but also the results of V7 for this same year were analyzed. However, this evaluation was limited to 2 months in the summer period where the model performance was comparable to its performance in the year 2003, which was also very similar to version V8. We can see that for the year 2004 all the model versions have very good results modeling nicely the stratospheric NO_2. Still, in winter, V10 has a slight overestimation of the values in the tropics.

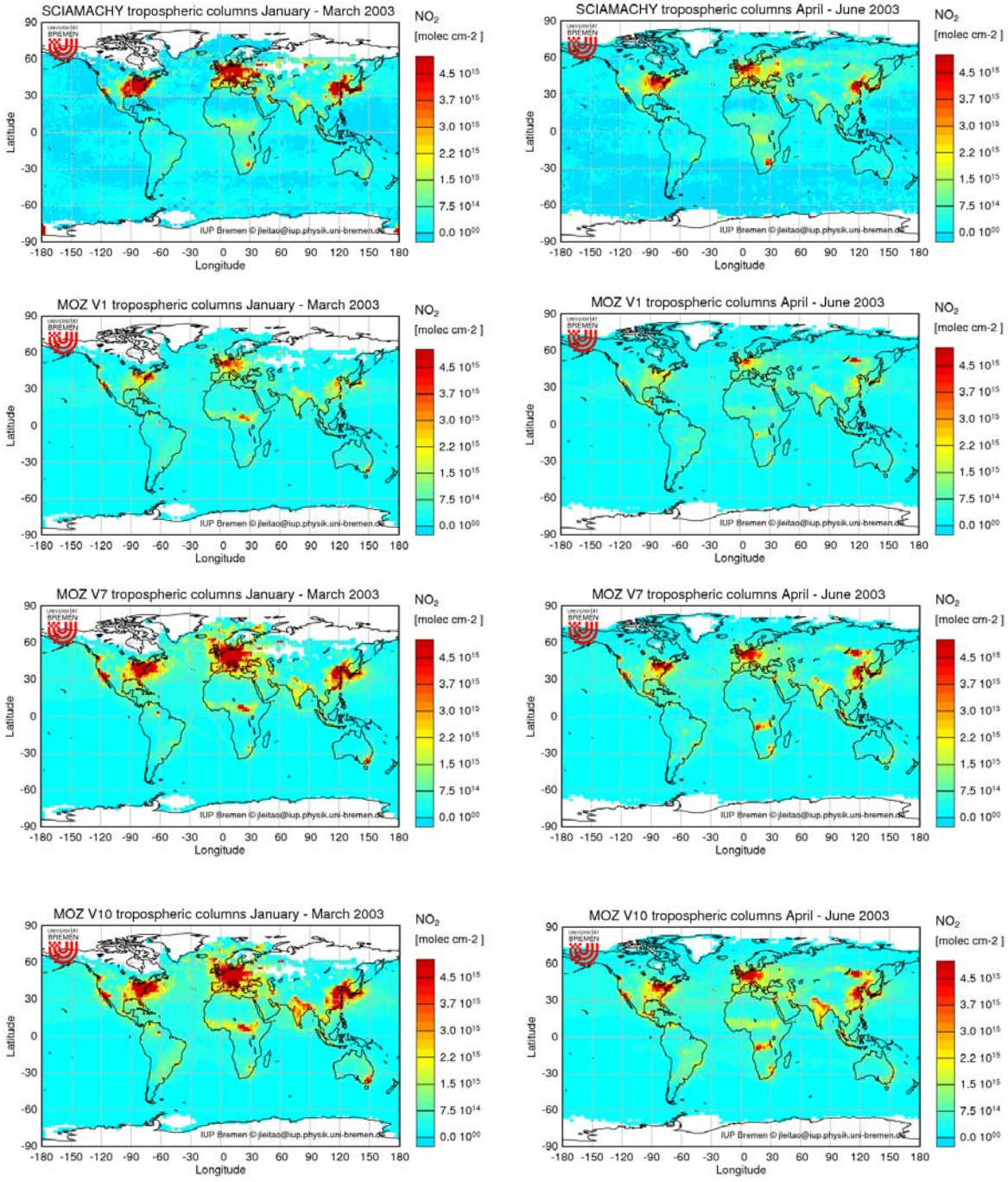


Figure 2.6: Three month averages of global tropospheric NO_2 determined by (from second of top to bottom) MOZART V1, V7 and V10 and total column of NO_2 measured by SCIAMACHY (top) for January – March 2003 (left) and April – June 2003 (right).

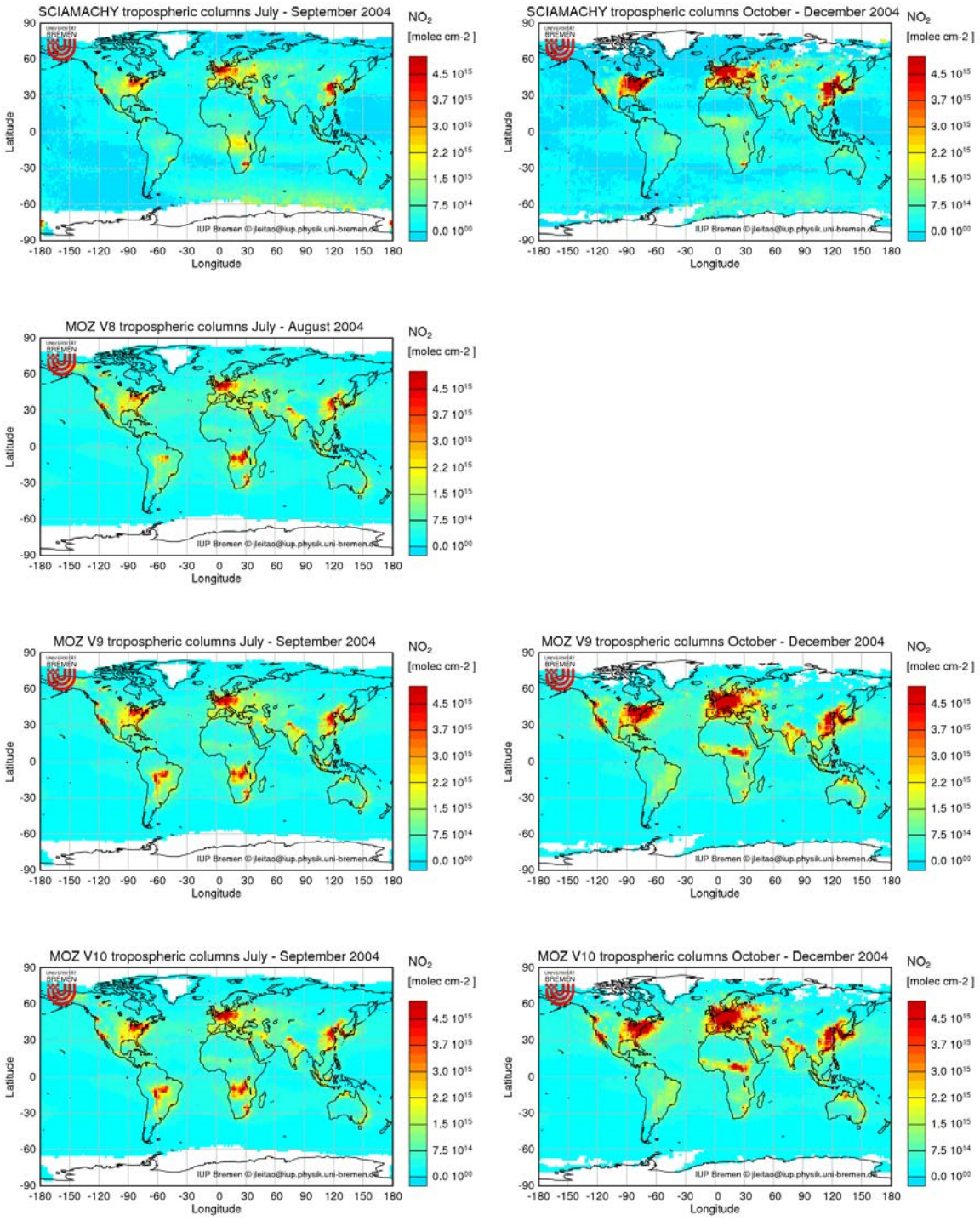


Figure 2.7: Three month averages of global stratospheric NO_2 determined by (from second of top to bottom) MOZART V8 (data only for July and August), V9 and V10 and NO_2 measured by SCIAMACHY (top) for July – September 2004 (left) and October – December 2004 (right).

Looking at the overall results presented above, the main conclusions regarding the tropospheric NO_2 are:

- Above biomass burning regions the model tends to overestimate the NO_2 emissions both in 2003 and 2004. This is the case for all model versions except for V1 which is able to simulate the right order of magnitude in Africa in 2003. We see that in Siberia in the summer of 2003, in Alaska and South America in the summer of 2004, and in central Africa both in 2003 and 2004, the

discrepancies between model emissions and satellite measurements are quite high, i.e., the satellite data shows no strong NO_2 enhancement while the models predict a clear signature of the fires.

- Above polluted regions, the values are in general closer to the satellite measurements, except for MOZART V1 that underestimated largely this amount. Still, in the winter of 2004, the NO_2 transported away from the eastern coast of USA is much higher in the model results.

The results of version V7 are not presented here but the main differences between this version and the next V8 are observed over India where V8 displays very high values and over central Africa where V7 does not reproduce the emissions from wild fire events.

c) TM5

In the following we present the evaluation of the TM5 standalone results, i.e. the V7 run for 2003, and the V9 (summer period) and V10 runs for 2004. The stratospheric model results are compared to the satellite measurements in the next figures.

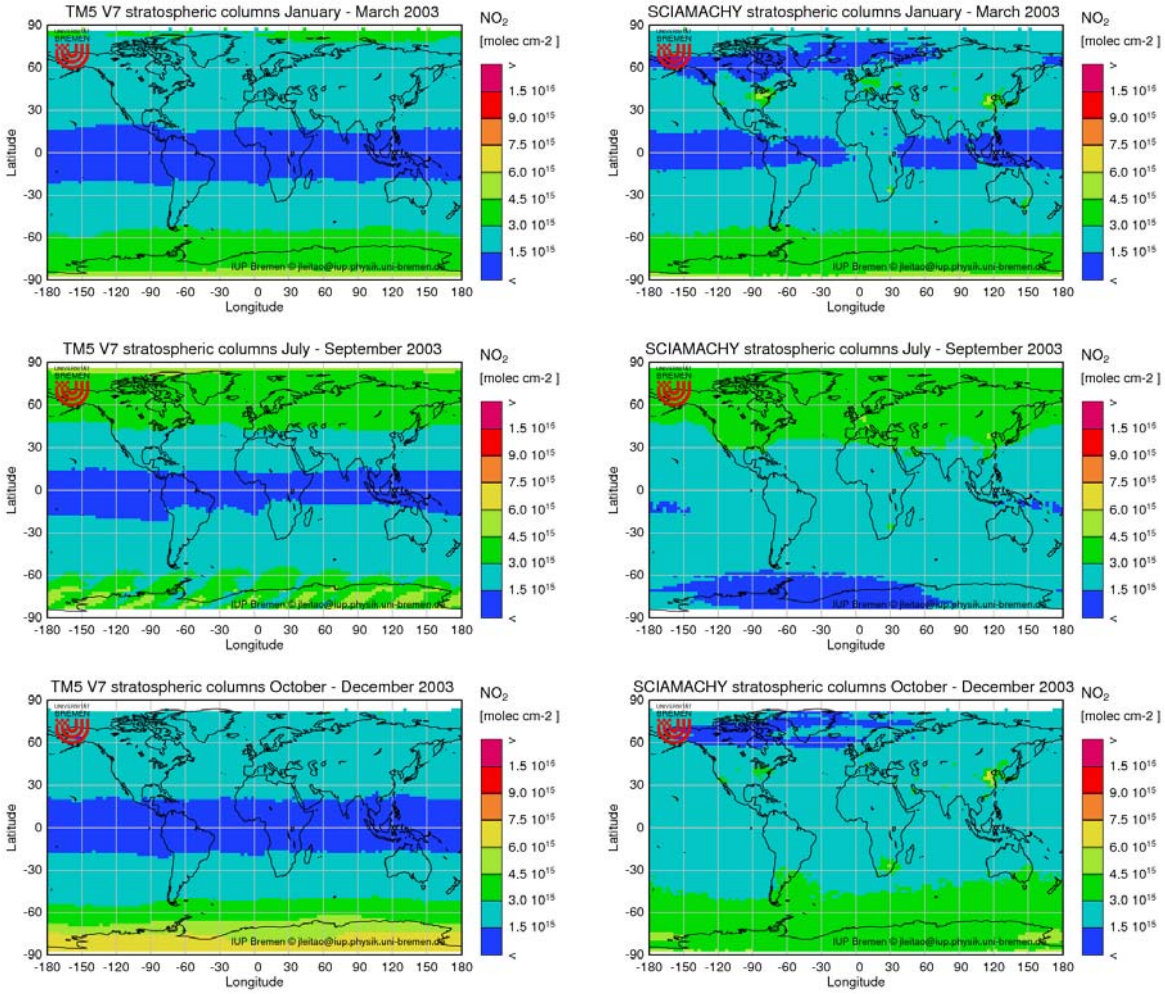


Figure 2.8: Three month averages of global stratospheric NO_2 determined by TM5 V7 (left) and total column of NO_2 measured by SCIAMACHY (right) for part of the year 2003.

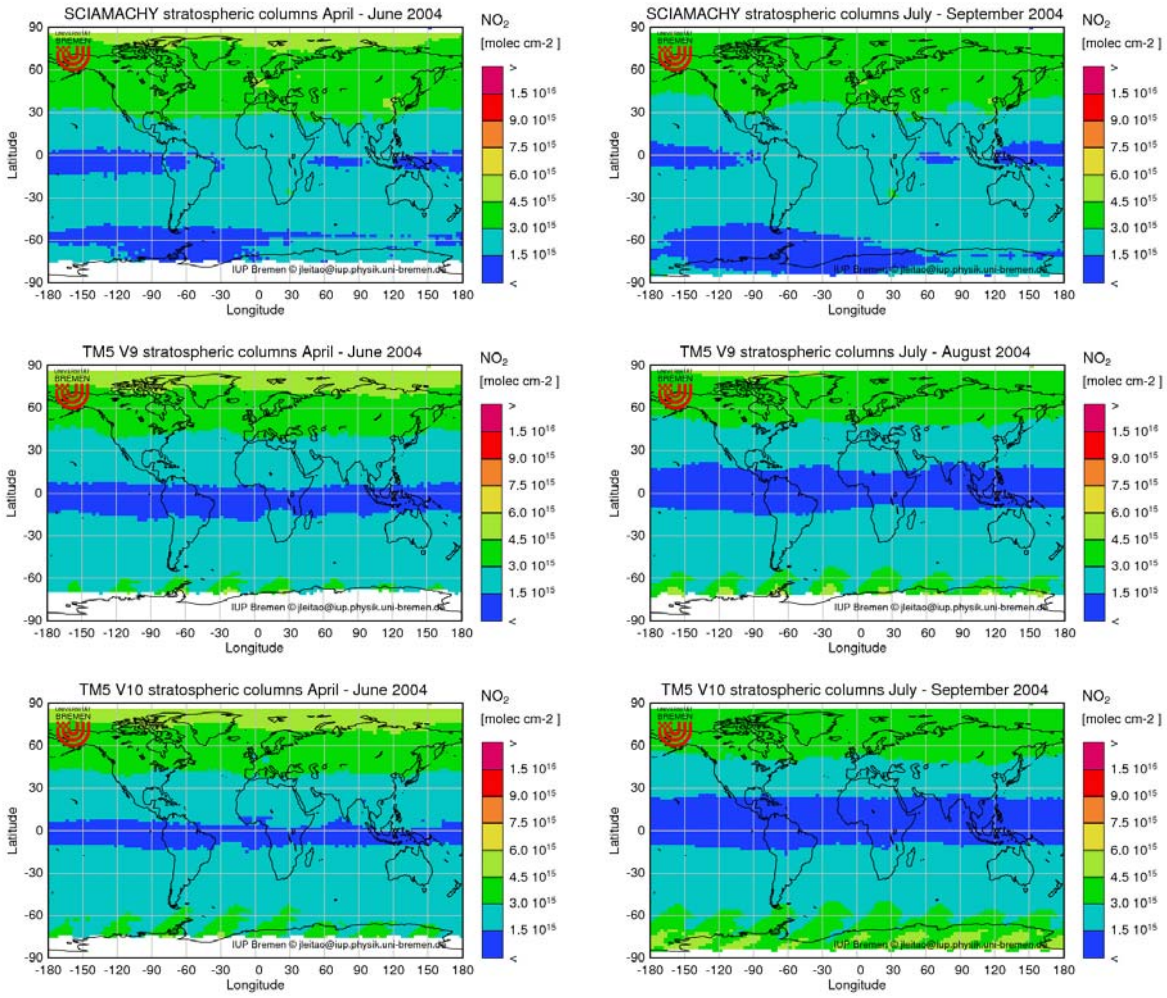


Figure 2.9: Three month averages of global stratospheric NO_2 determined by TM5 V9 (middle) and TM5 V10 (bottom) and total column of NO_2 measured by SCIAMACHY (top) for April – June (left) and July – September (right) of the year 2004.

From the evaluation of the TM5 stratospheric results it can be observed that in some months of the year 2003 (from April to June and October to December, the latter not presented here for V10) the NO_2 values from satellite measurements are lower than the ones of the model at high latitudes. However, as can be verified from the seasonal plots presented in figure 2.18, the versions that provide data for the year 2004 were able to give slightly better results over the high latitude regions. Still, the latest version V10 overestimates the stratospheric NO_2 over the South Pole region, especially in the second half of 2004.

At high southern latitudes, stripes appear in the TM5 maps. This is the result of a variation in stratospheric NO_2 chemistry with time which is not fully compensated by the linear interpolation applied to the TM5 fields which are saved in three hour steps. Similar patterns were observed in MOZART data for those data which were not yet saved on an hourly basis.

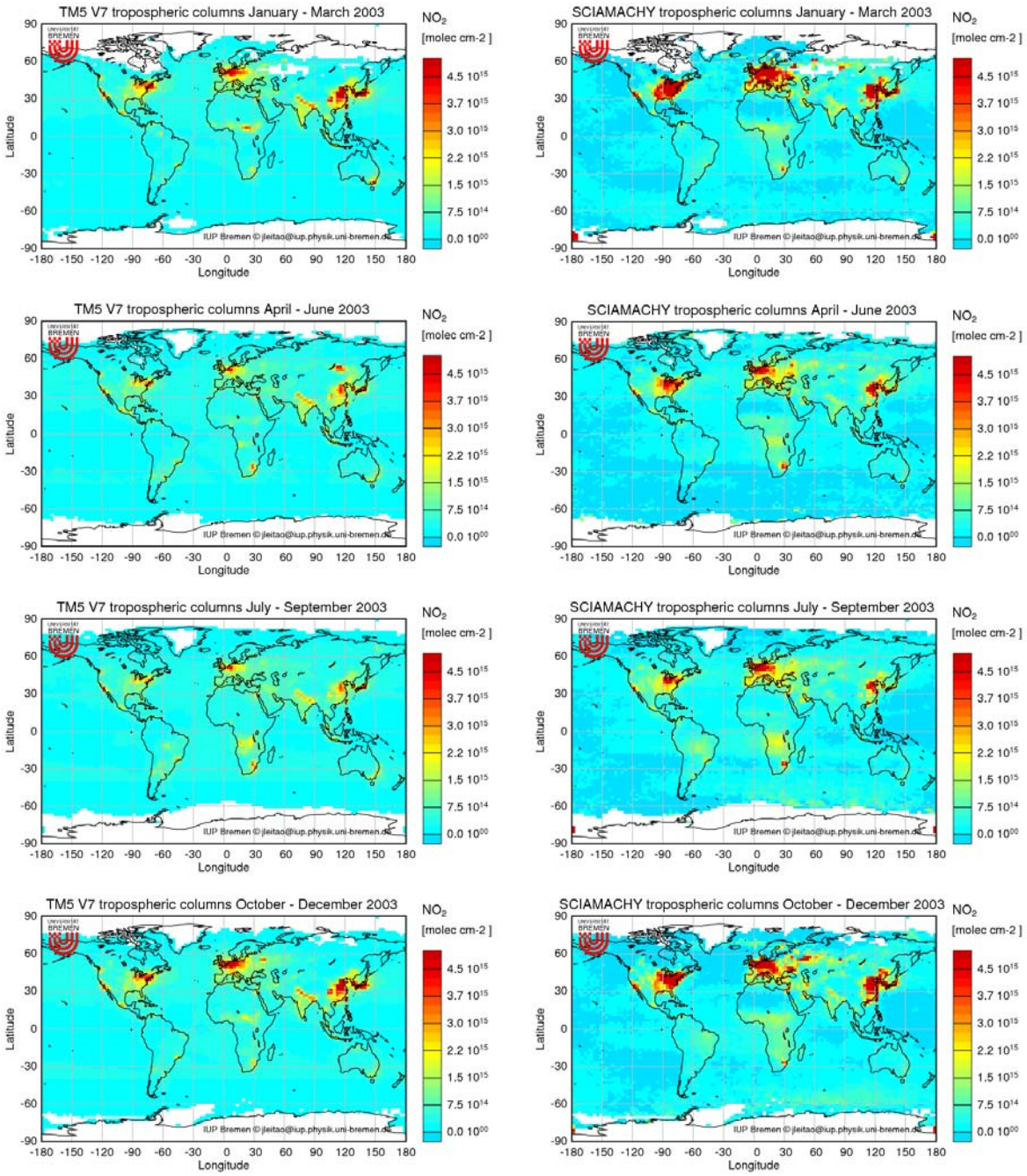


Figure 2.10: Three month averages of global tropospheric NO_2 determined by TM5 V7 (left) and measured by SCIAMACHY (right) for the year 2003.

From Fig. 2.10 above we can see that tropospheric NO_2 determined by the TM5 V7 is mostly too low over polluted areas, such as Europe, China or the US. On the other hand, when looking at regions characterized with emissions mostly from fire events, there is no systematic over- or underestimation of the NO_2 amounts. For example, in the period of July-September, the NO_2 columns are too small over South America and central Africa when compared to the SCIAMACHY measurements. On the contrary, in the period of April-June the model emissions for the boreal fires in Siberia or those in Australia are higher than the NO_2 detected by the satellite.

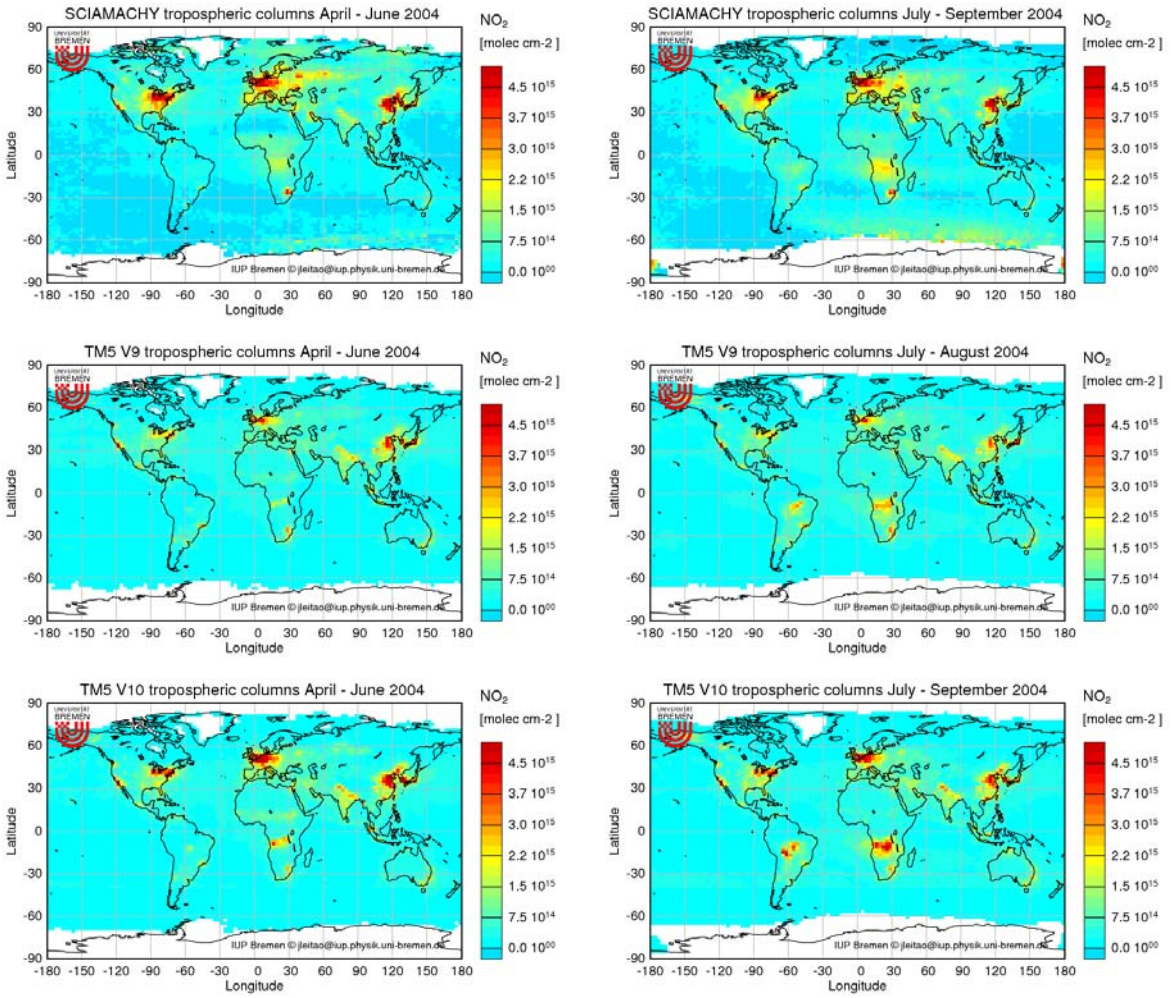


Figure 2.11: Three month averages of global tropospheric NO_2 determined by TM5 V9 (middle) and TM5 V10 (bottom) and measured by SCIAMACHY (top) for April – June (left) and July – September (right).

In general it can be concluded that compared to the previous versions, the TM5 V10 has improved on the emissions from urban regions (the previous version V9 was predicting too low values), but the biomass burning emissions are still much higher than the satellite measurements. While V7 had often too low values, this latest version has too high columns in regions like southern Africa (see for example the last 3 months of the year 2003 in figure 2.10 and for the period April-September 2004 in figure 2.11 for both V9 and V10).

Just like for MOZART, the Alaska fire period (June to August 2004) was also taken as a specific case-study for evaluation of tropospheric NO_2. From that analysis it was concluded that, likewise the MOZART versions, the TM5 V9 also overestimated the NO_2 in this region. However, this discrepancy was smaller for this model.

Seasonality graphs

In the next section we present the seasonality graphs for each model version for both the stratosphere and troposphere. Results are presented for the regions defined in section 2.1.

a) MOCAGE

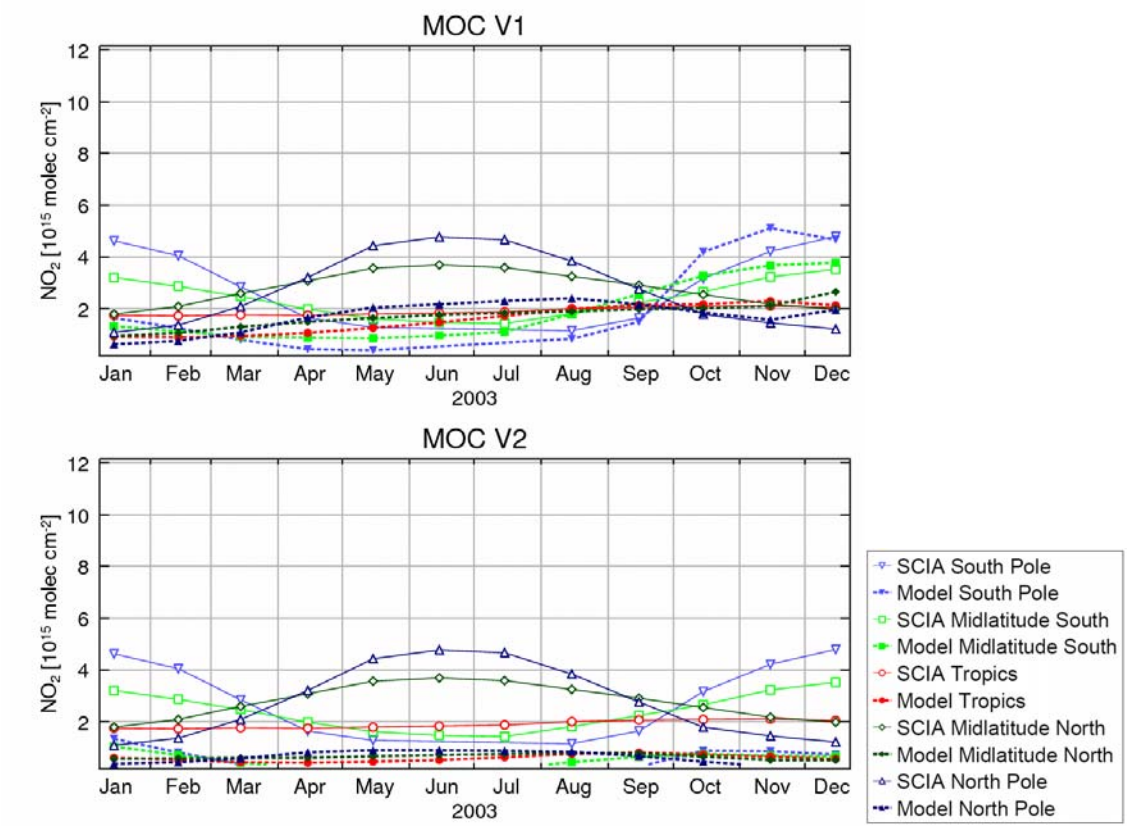


Figure 2.12: Seasonality curves for 2003 of stratospheric NO_2, in pre-defined regions, determined by the standalone run of MOCAGE V01 (top) and V02 (bottom) and total NO_2 columns measured by SCIAMACHY (open symbols).

As already shown in the previous section, the stratospheric NO_2 values modeled by MOCAGE are too low with the exception of the Polar Regions in NH winter. In the figure above it is possible to see that the first version of the model is closer to the satellite measurements, but the NO_2 annual variation in V02 is basically inexistent.

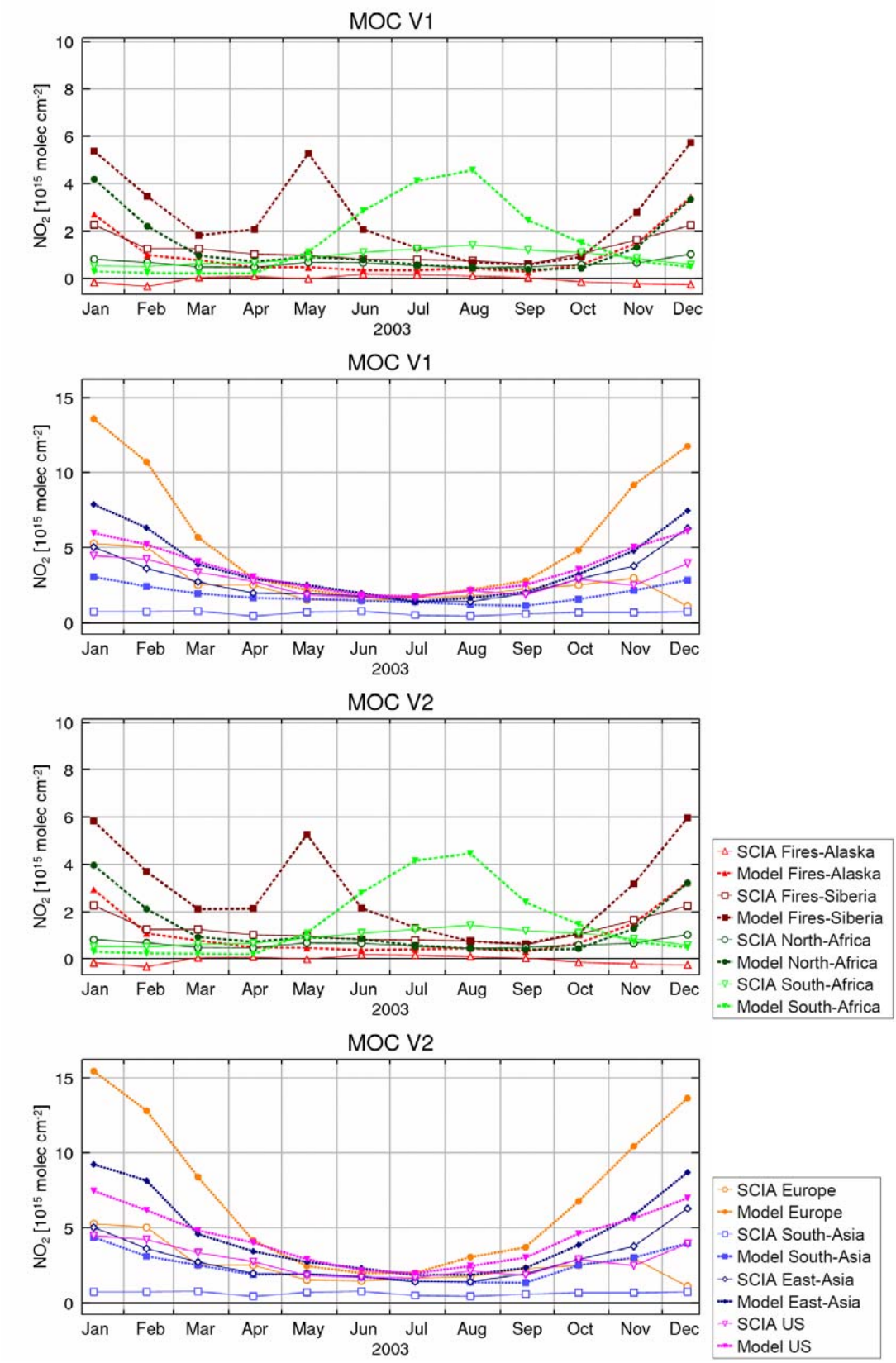


Figure 2.13: Seasonality curves for 2003 of tropospheric NO_2, in pre-defined regions divided into the typical biomass burning and polluted areas, determined by the standalone run of MOCAGE V01 (top two) and V02 (bottom two) and measured by SCIAMACHY (open symbols).

Regarding the tropospheric values, the model performs slightly better than in the stratosphere. The annual variation is in general captured. However, the values over Europe are extremely high and in the remaining polluted regions the values are also systematically overestimated. The differences in the tropospheric NO_2 between the two versions are quite small when looking at the seasonal variation (and also from the previous maps shown in figure 2.3) so no conclusions can be drawn from analyzing these figures alone.

b) MOZART

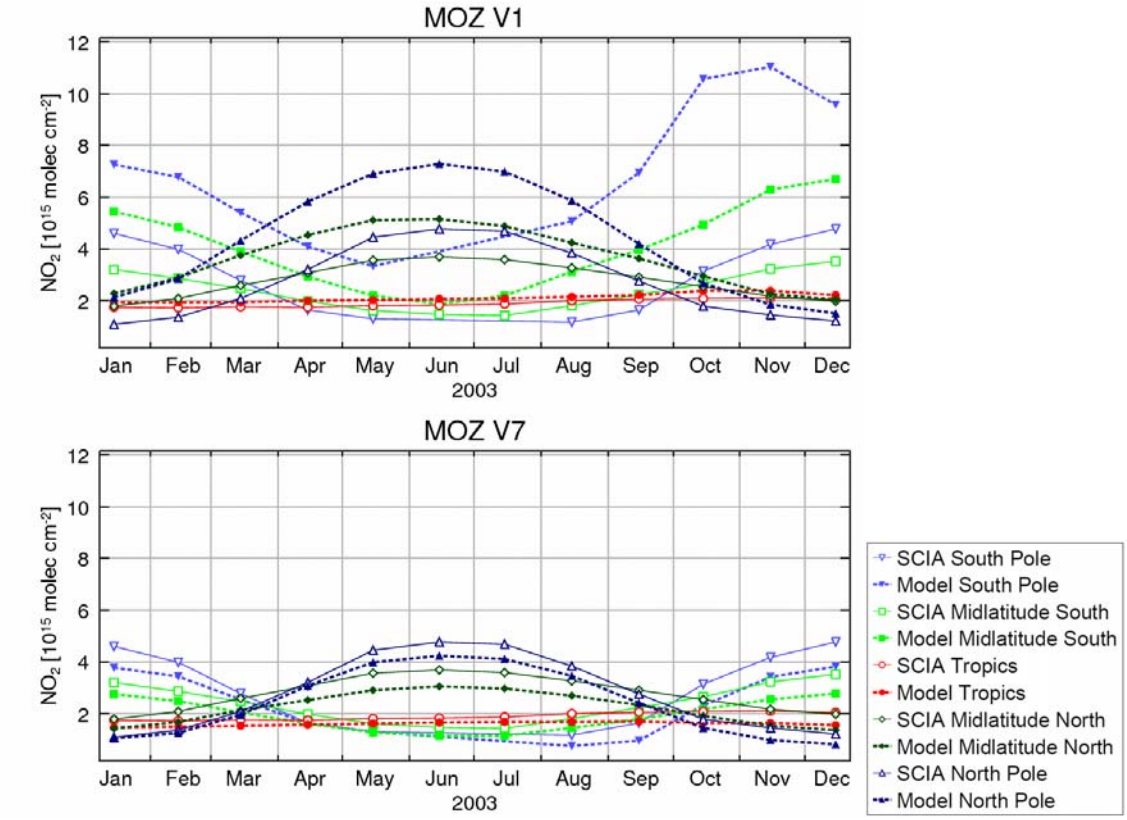


Figure 2.14: Seasonality curves for 2003 of stratospheric NO_2, in pre-defined regions, determined by the standalone run of MOZART V1 (top) and V7 (bottom) and total NO_2 columns measured by SCIAMACHY (open symbols).

From the seasonality curves in figure 2.14, it can be concluded that the stratospheric fields of NO_2 modeled by MOZART V7 agree quite well with the satellite data, being just a bit lower. This shows a great improvement compared to version V1 that had extremely high values in the winter periods of the Polar Regions. In the "tropics"-region, the model and the satellite present very similar values.

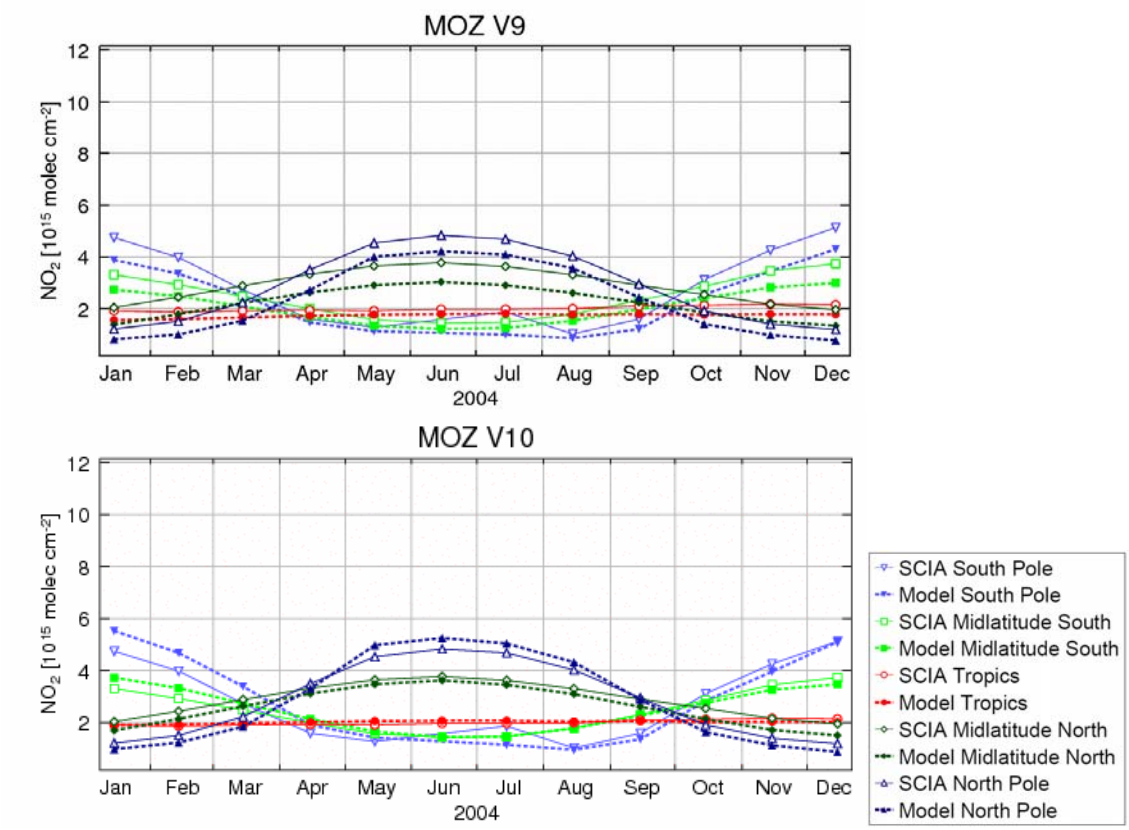


Figure 2.15: Seasonality curves for 2004 of stratospheric NO_2, in pre-defined regions, determined by the standalone run of MOZART V9 (top) and V10 (bottom) and total NO_2 columns measured by SCIAMACHY (open symbols).

As shown in figure 2.15, shifting to version V9 and V10 does not change the overall picture very much. Both model versions agree quite well with the satellite data, showing the correct annual variation and very close values. However, looking closer to the curves one can see that while MOZART V9 has lower values than the satellite measurements, version V10 now slightly overestimates the satellite providing overall the best agreement.

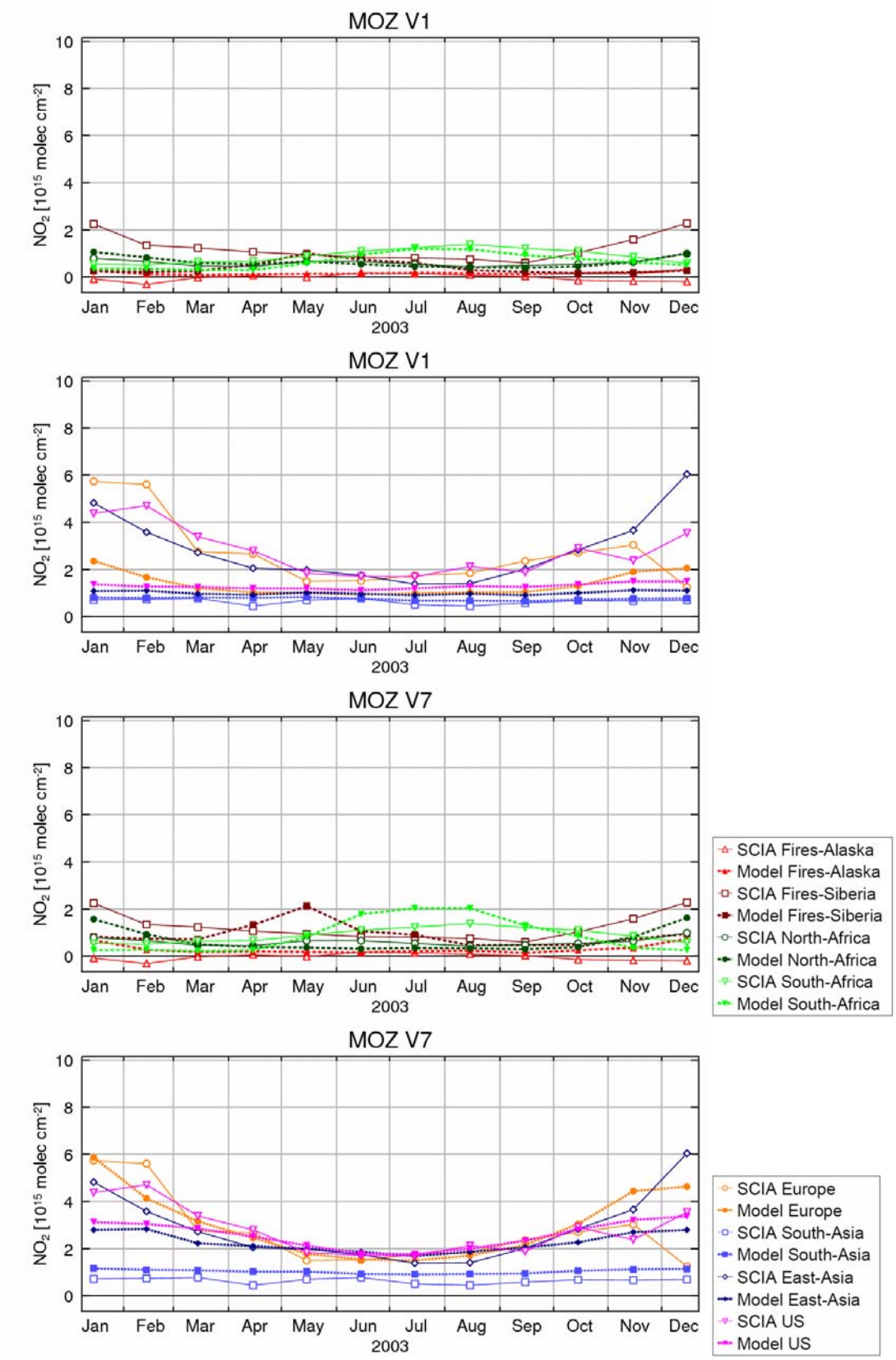


Figure 2.16: Seasonality curves for 2003 of tropospheric NO_2, in pre-defined regions divided into the typical biomass burning and polluted areas, determined by the standalone run of MOZART V1 (top two) and V7 (bottom two) and measured by SCIAMACHY (open symbols).

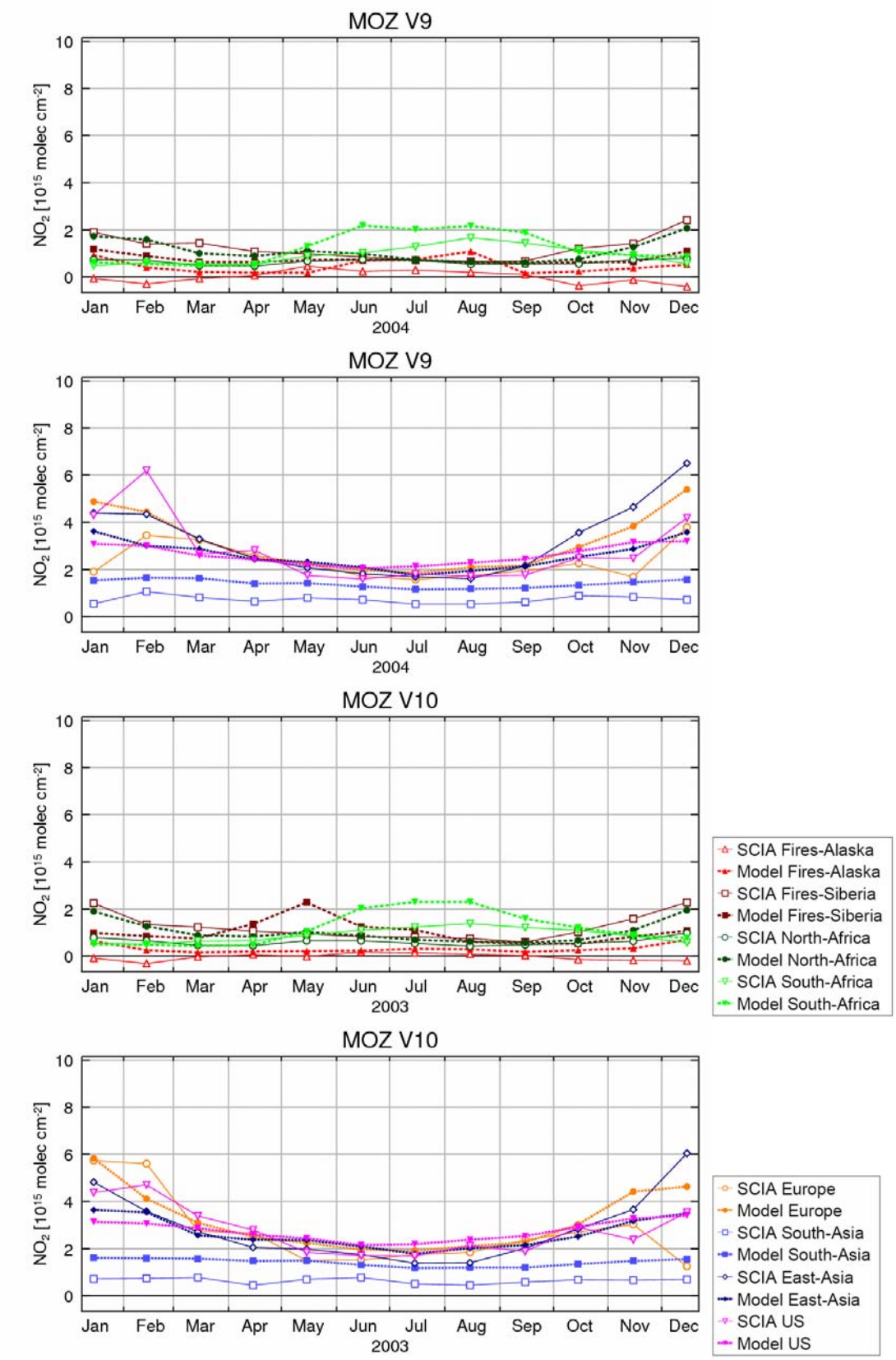


Figure 2.17: Seasonality curves for 2004 of tropospheric NO_2, in pre-defined regions divided into the typical biomass burning and polluted areas, determined by the standalone run of MOZART V9 (top two) and V10 (bottom two) and measured by SCIAMACHY (open symbols).

The results for 2003 (figure 2.16) show that MOZART V7 tends to overestimate biomass burning emissions. See for example the peak in May corresponding to NO_2 from fires in Siberia that was not measured by the satellite. For this particular region, the results of the previous version V1 are quite the opposite with underestimation of the values almost all year. Nevertheless, V1 has better results, for example for South-Africa region where the model results are very close to the satellite data. Version V7 is doing a good job on simulating the seasonality of the tropospheric NO_2 showing also good agreement over Europe and the US. On the other hand, this is not the case for V1 where the model NO_2 emissions in these regions are too low and the variation throughout the year is not reproduced. East-Asia is the most problematic polluted region where winter columns are strongly underestimated probably due to too low emission values taken from inventories which do not yet reflect recent increases in NOx emissions in China.

Once again we can see that the differences between V9 and V10 are rather small (as expected for the tropospheric results), showing the same similarities and differences with the satellite data. Nonetheless, from these results we see that in general the MOZART model performs quite well when modeling tropospheric NO_2. There is agreement between measurements and model results in the biomass burning regions (except for South-Africa) and over the urban areas. Only in winter we see too low NO_2 determined by the model.

c) TM5

The seasonal trends of TM5 runs for the years 2003 and 2004 are presented in the figures 2.18 to 2.20 below. We can clearly see that all the model versions are very similar. The South Pole values are overestimated by V7 and V10 (no data was analyzed for V9 in these months) in the winter period (i.e., local summer). This maximum seems to be shifted by 1-2 months of when it should take place according to the measurements. The remaining stratospheric model values are in good agreement with the satellite data. However, one should note that little seasonality occurs in mid-latitudes of both hemispheres.

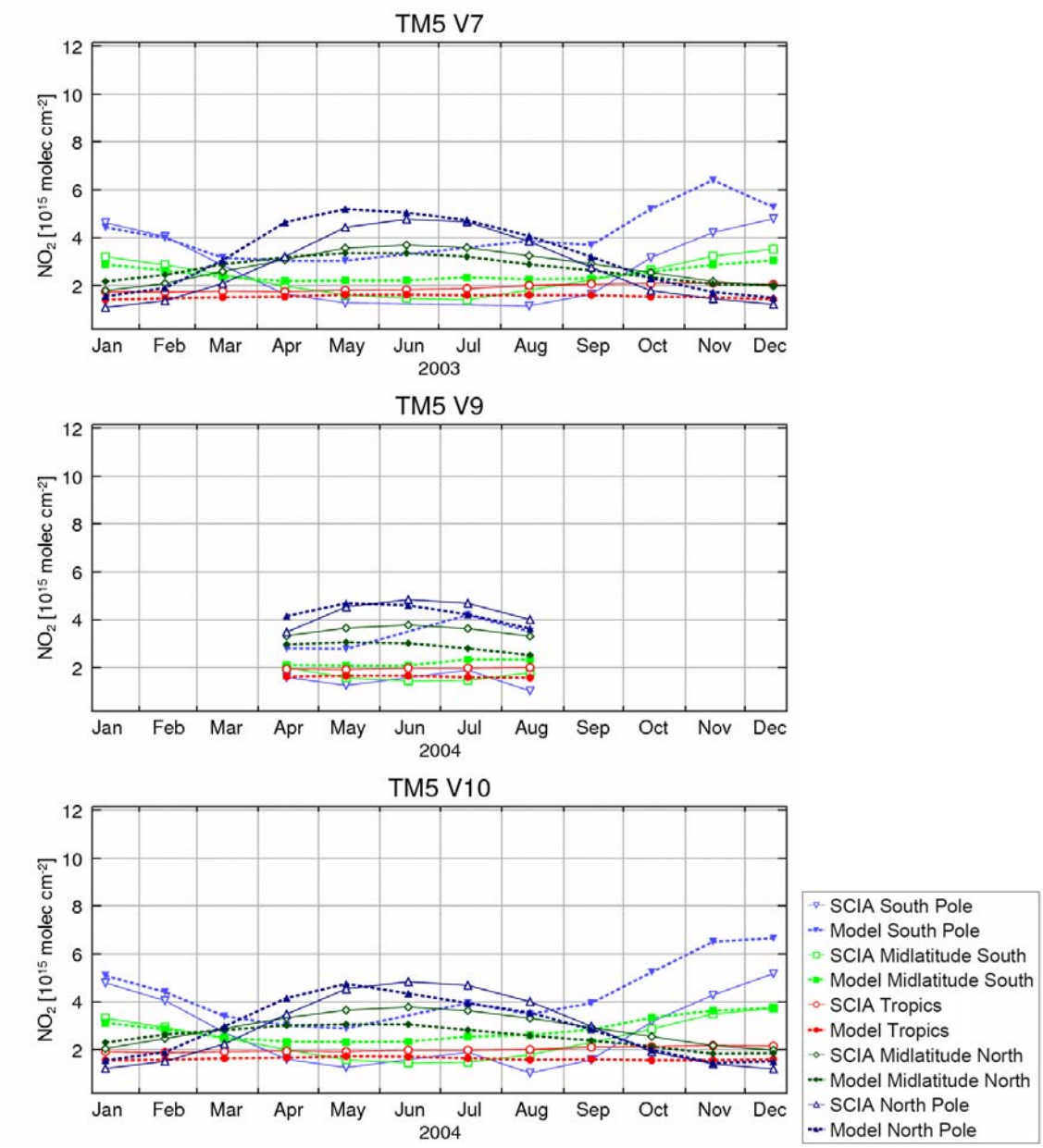


Figure 2.18: Seasonality curves for 2003 (top) and 2004 (bottom two) of stratospheric NO_2, in pre-defined regions, determined by the standalone run of TM5 V7 (top), V9 (middle, with data only from April till August 2004) and V10 (bottom) and total NO_2 columns measured by SCIAMACHY (open symbols).

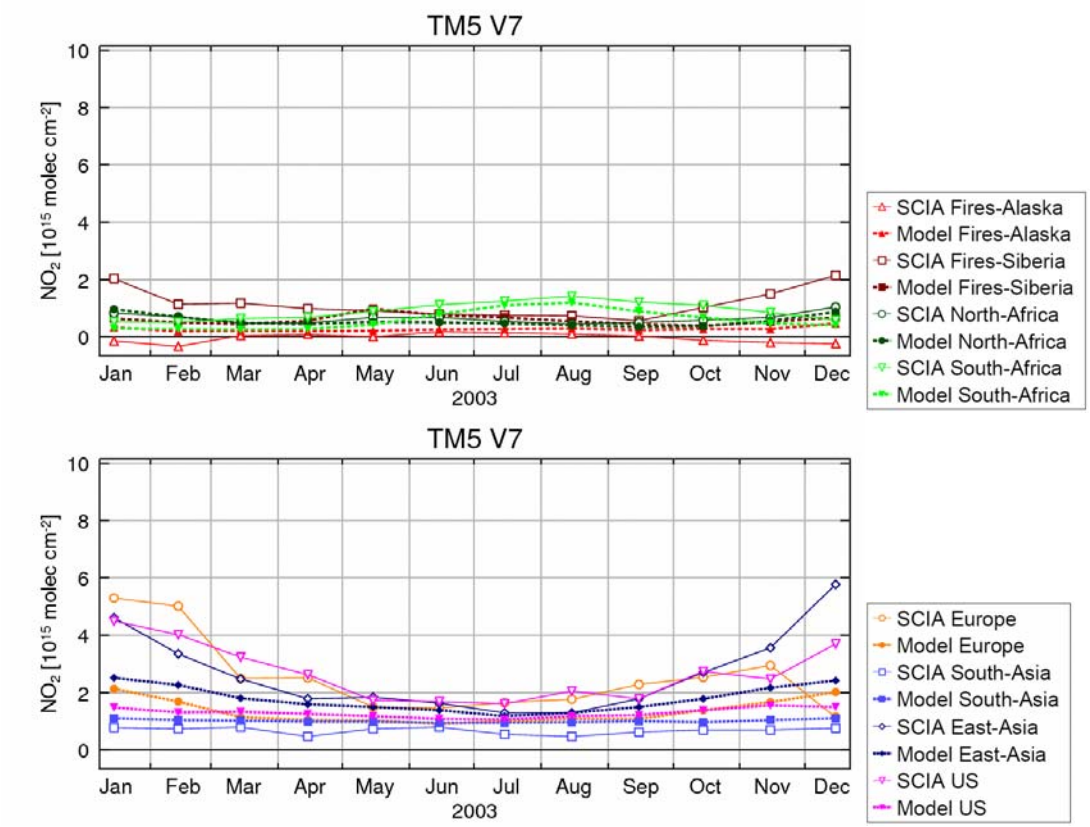


Figure 2.19: Seasonality curves for 2003 of tropospheric NO_2, in pre-defined regions divided into the typically biomass burning and polluted areas, determined by the standalone run of TM5 V7 and measured by SCIAMACHY (open symbols).

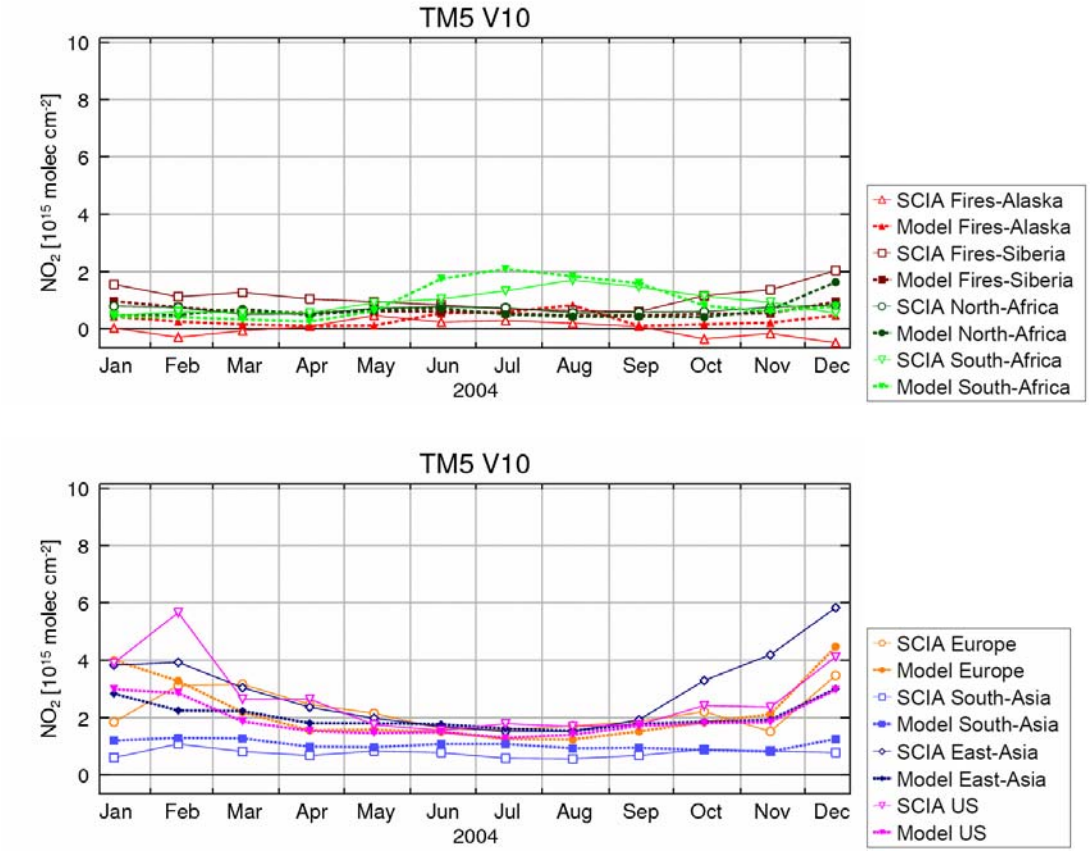


Figure 2.20: Seasonality curves for 2004 of tropospheric NO_2, in pre-defined regions divided into the typically biomass burning and polluted areas, determined by the standalone run of TM5 V10 and measured by SCIAMACHY (open symbols).

For TM5, the emissions over biomass burning regions are slightly underestimated in the year 2003 but overestimated in the summer of 2004. Nevertheless, the seasonality is well reproduced in these regions. In anthropogenic source regions, TM5 underestimates both the seasonality and the absolute columns of tropospheric NO_2, but it looks that the 2004 run has performed better in this matter than the previous V7. As for MOZART, the largest discrepancies are observed over East-Asia in winter.

Monthly scatter plots

Scatter plots allow for a detailed comparison between the modeled tropospheric columns (in each of the regions) and the satellite measurements. These tell us how good the satellite and model data correlate for individual regions and how well the absolute values agree.

a) MOCAGE

In Fig. 2.21 below we show the scatter plots for both MOCAGE V01 and V02 for a selection of months that are representative of the overall conclusions drawn from the analysis of these plots. From the 'January 2003'-plots we can see, once more, that MOCAGE results are much higher than the satellite measurements in the winter months. And, in fact, if we look at the scatter plots of the others models (figures 2.22 to 2.25) we notice that, for these months, the tropospheric NO_2 values are on average much lower than from MOCAGE. In the plots for May, we see an example of the overestimation of emissions in biomass burning regions (fires in Siberia in this case), but the values look much better for the remaining regions. September 2003 is a month where the model correlates nicely with the satellite measurements.

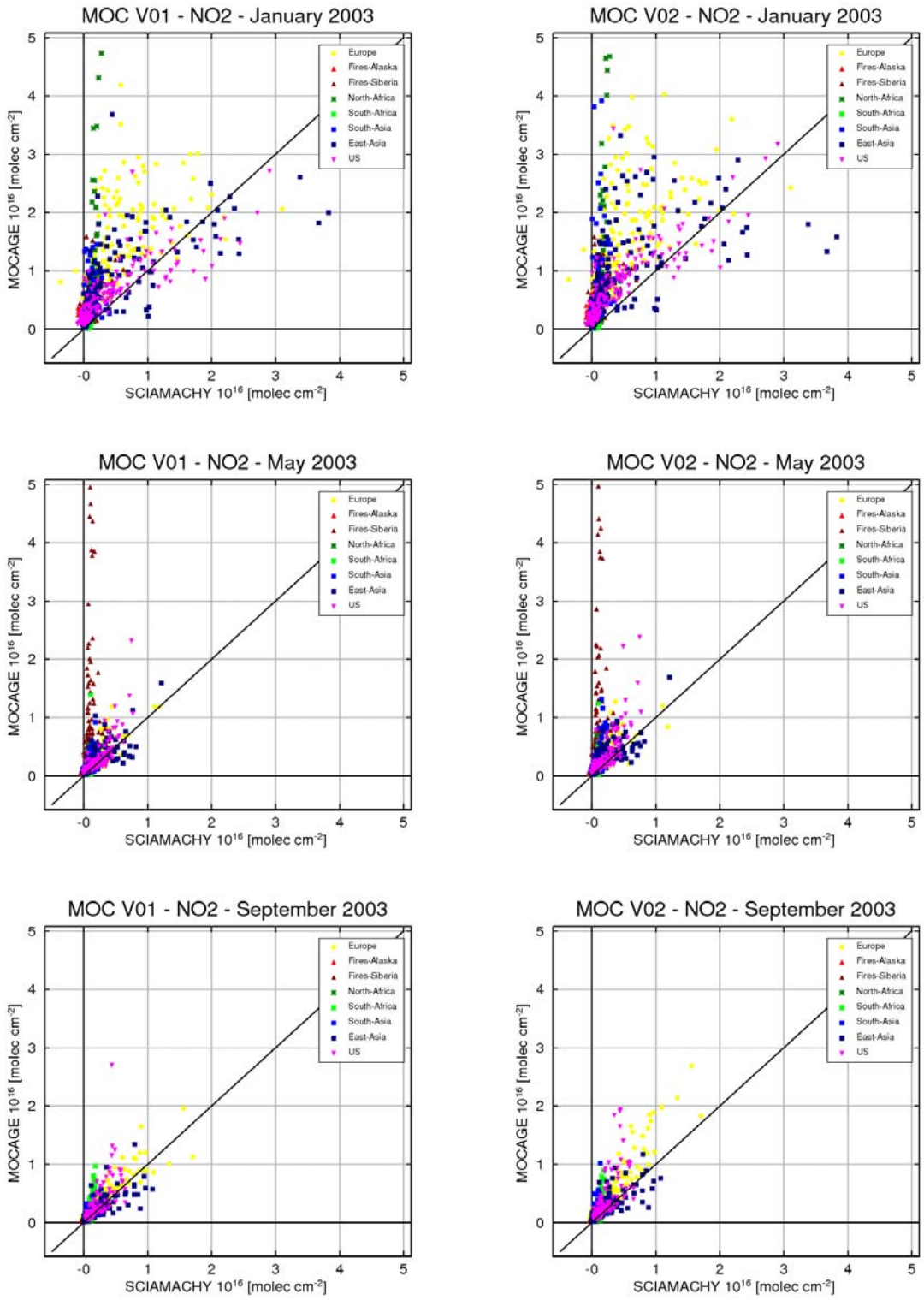


Figure 2.21: Scatter plots of monthly averages of tropospheric NO_2 for January, March, May and September of 2003, determined by standalone runs of MOCAGE V01 (left) and V02 (right) and measured by SCIAMACHY.

b) MOZART

In the following we present a selection of scatter plots for results of different MOZART versions for the years of 2003 and 2004.

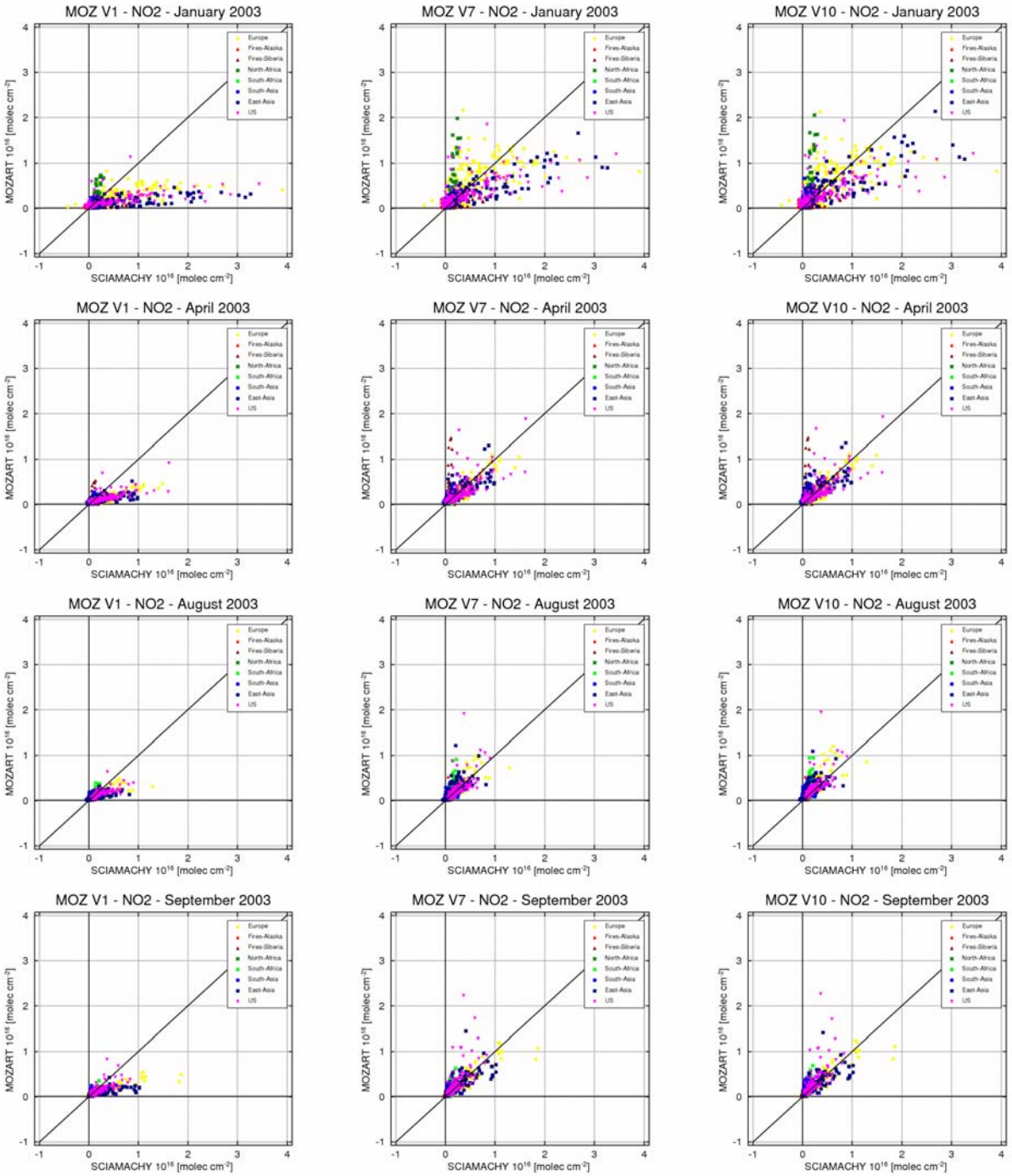


Figure 2.22: Scatter plots of monthly averages of tropospheric NO_2 for January, April, August and October of 2003 determined by standalone runs of MOZART V1 (left), V7 (middle) and V10 (right) and measured by SCIAMACHY.

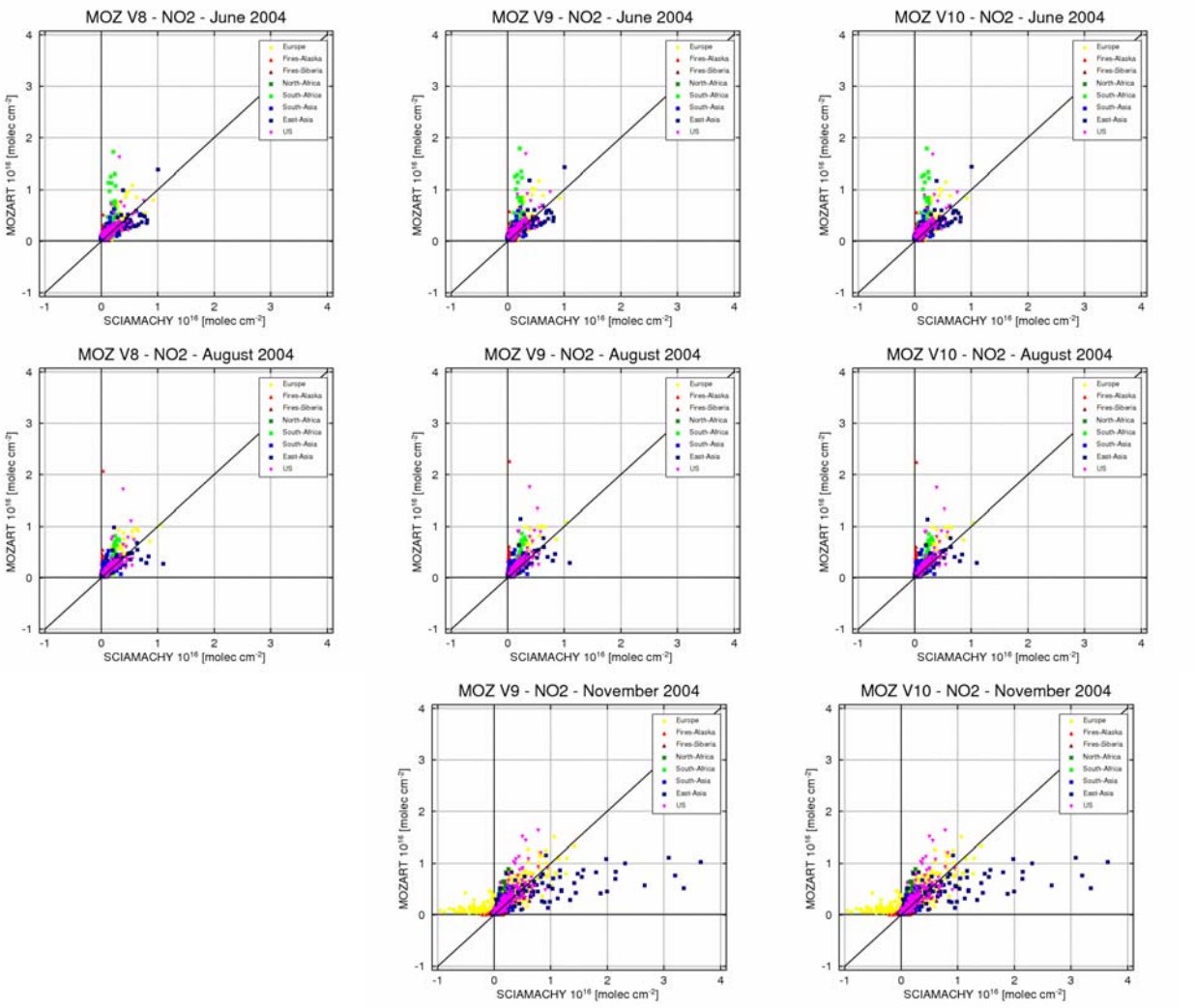


Figure 2.23: Scatter plots of monthly averages of tropospheric NO_2 for June, August and November of 2004 determined by standalone runs of MOZART V8 (left), V9 (middle) and V10 (right) and measured by SCIAMACHY.

In the correlations presented above in Figs. 2.22 and 2.23, we see that the NO_2 data calculated by standalone MOZART runs fit better to the satellite measurements in the summer months. For the other months, a very high dispersion can be found especially in the winter period (see for example January 2003 or November 2004). The different model versions behave mostly in the same way and the overall scenario is quite good. Yet, we see that, in the winter of 2003, data from MOZART V1 show much less scatter compared to SCIAMACHY measurements than the other 2 versions, albeit at overall much too low values.

The negative values over Europe in November 2004 are a result of the current retrieval method of the satellite data and not an indication of model problems.

TM5

For TM5 there were 2 standalone versions available to be evaluated in terms of NO_2. In figure 2.24, we present some scatter plots for version V7 (year 2003) and in Fig. 2.25 for versions V9 and V10 (year 2004).

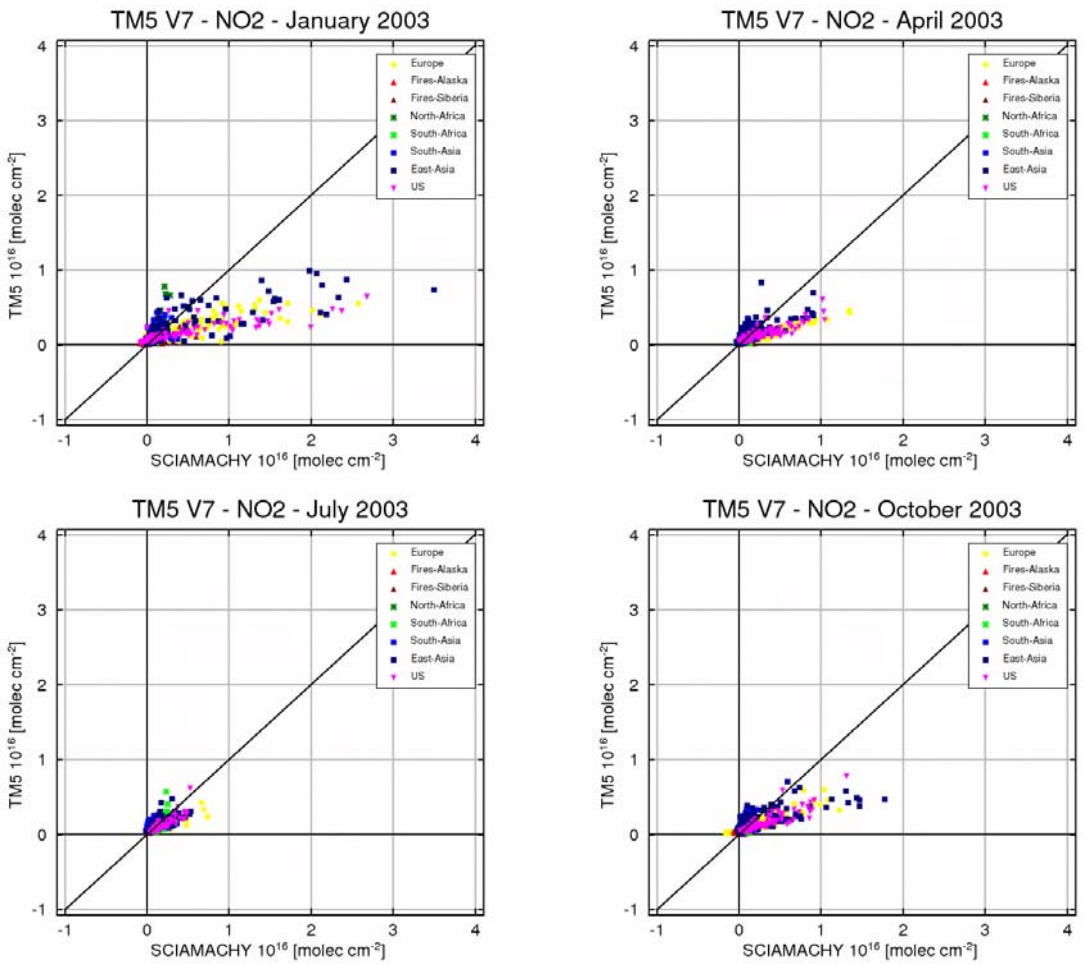


Figure 2.24: Scatter plots of monthly averages of tropospheric NO_2 for January, April, July and October 2003 determined by standalone runs of TM5 V7 and measured by SCIAMACHY

From these plots presented above we see an outstanding difference between the data calculated for the winter and the summer months. While in January and October 2003 the data are somehow scattered and the model is underestimating the tropospheric NO_2 in almost every region, in the summer months (see July for example) the correlation of data is very good. Overall, compared to the other models, this version was performing quite well. This might be in part explained by the coarser grid which leads to more smoothing and a smaller amount of values considered in the correlation.

However, some further improvements were necessary and, for 2004, there were also data available for V9 and, later on also for V10. In the figure below we present the scatter plots for the summer months (the period in which V9 was evaluated). As expected (from the previous version available), in this time of the year the model performs quite well and the data nicely correlate with the satellite measurements. Still, in the winter months the same dispersion and systematic underestimation of NO_2 columns was found for V10 as it was for V7.

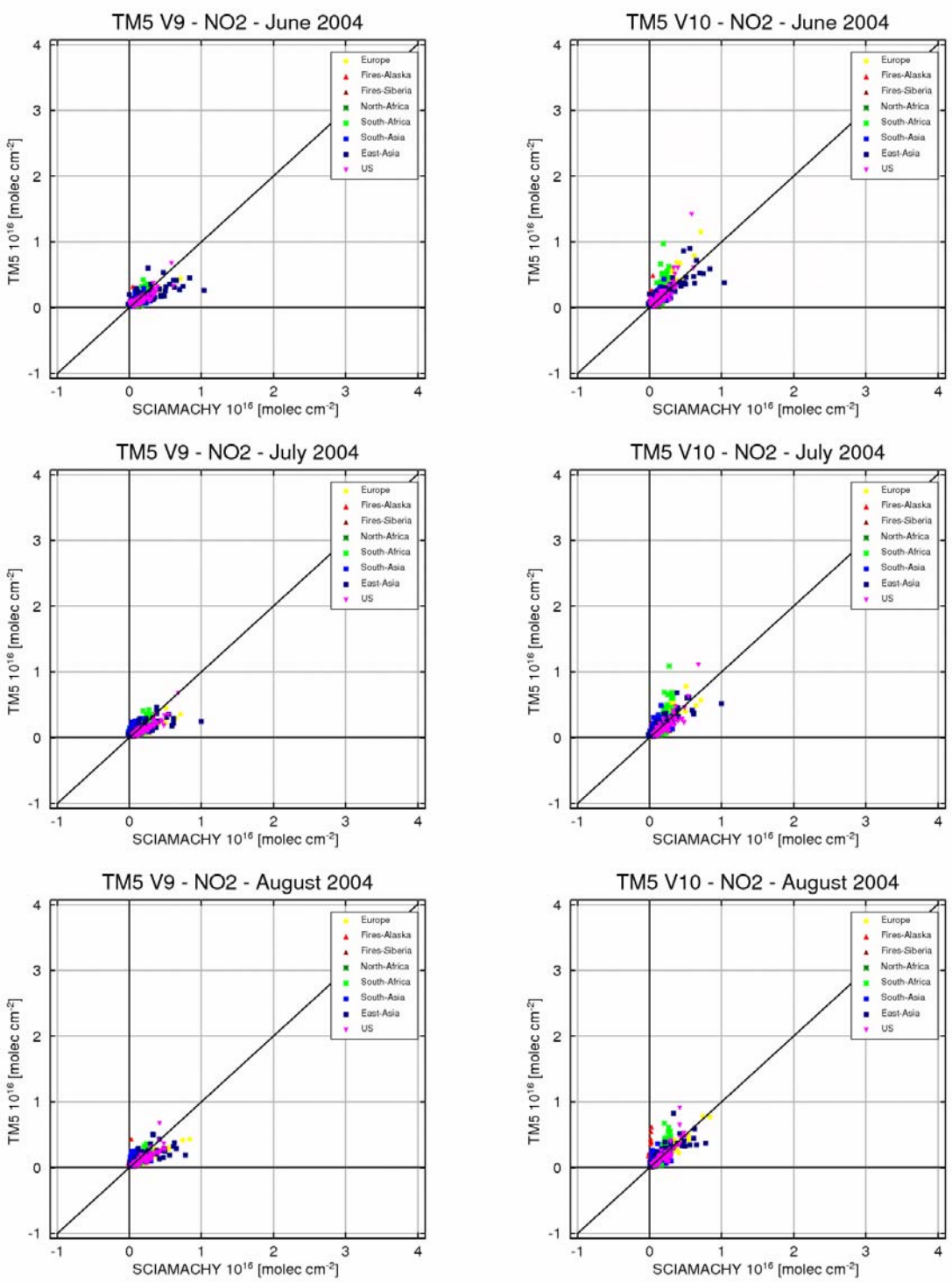


Figure 2.25: Scatter plots of monthly averages of tropospheric NO_2 for the summer months (June to August) from the year 2004 determined by standalone runs of TM5 V9 (left) and TM5 V10 (right) and measured by SCIAMACHY

Scores plots

In the following figures we present the time averaged scores for the regions defined in section 2.1, for all the model runs of the three CTMs considered in this evaluation. The averages are calculated over the months for which model data were analyzed. Using the scores, the models can be compared side by side in terms of their performance in the years 2003 and 2004, and also in the stratosphere and troposphere. These graphs are a compilation of an extensive record of scores determined. The graphs with scores calculate per region, per month, per model run can be found in Annex 8. In those, one can better identify and understand

the reason behind when and where the model is performing well and where it still requires further improvement.

It is important to make clear that the tropospheric data considered here are only using measurements performed over land and, for this reason, accounting only those regions where NO_2 is normally emitted. This provides a more truthful score as the usually good results of models over the ocean are not included. The satellite measurements over ocean are close to the detection limit and therefore result in large scores by definition.

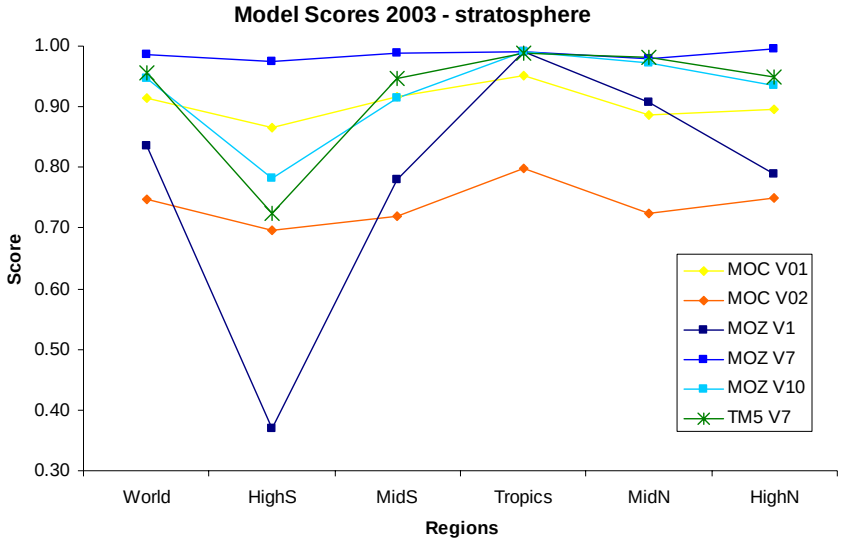


Figure 2.26: NO_2 annual averages (for 2003) of scores for the stratospheric regions for the different standalone runs of the CTMs.

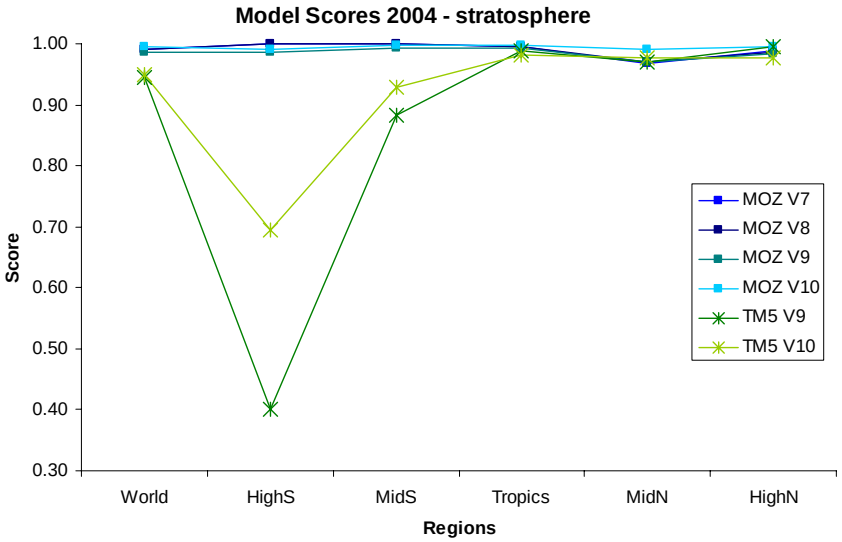


Figure 2.27: NO_2 annual averages (for 2004) of scores for the stratospheric regions for the different standalone runs of the CTMs.

Looking at the scores obtained for the stratosphere regions we can conclude that the South Pole (high southern latitudes – highS) is the most problematic region for almost every model run in 2003, except MOZART V7 that obtains quite high scores for all the regions. From this graph we can also notice how much better MOCAGE V01 is than V02. In addition, we also see how much MOZART improved from version V1 to version V7. As expected, the differences of scores obtained for V7, V8 and V9 are minimal and, in 2004, V10 has improved for the northern hemisphere results. Still, it is important to notice that the scores for MOZART V10 in the year 2004 are higher than those determined for 2003, with the larger difference over the South Pole where the score is 0.78 in 2003 and then 0.99 for 2004. Looking at the evolution of TM5 we clearly see an improvement from version V9 to V10, where the score over the South Pole has increased from 0.4 to 0.69.

In terms of the tropospheric scores we see, in the figures below, a large difference for 2003 between MOCAGE and the other models. It is also possible to conclude that the polluted regions are those for which lower scores are calculated. Looking at the different MOZART runs we notice that, for both 2003 and 2004, version V10 of this model has the lowest scores, while V1 is the best performing version for 2003 and V8 for 2004. Previous evaluation steps (see, e.g., the seasonal plots in Figs. 2.16 and 2.17) revealed that the latest version of MOZART was better when simulating the tropospheric NO_2. The better scores for V8 are a consequence from analyzing only data in summer months (April to August) which are months where the model usually performs better (i.e., the lower scores obtained for winter months are lowering the V9 and V10 annual score average). In 2003, TM5 V7 is on average the model with the best scores and the versions with data for 2004 (i.e. V9 and V10) have scores in the same range as the MOZART runs (see Fig. 2.29).

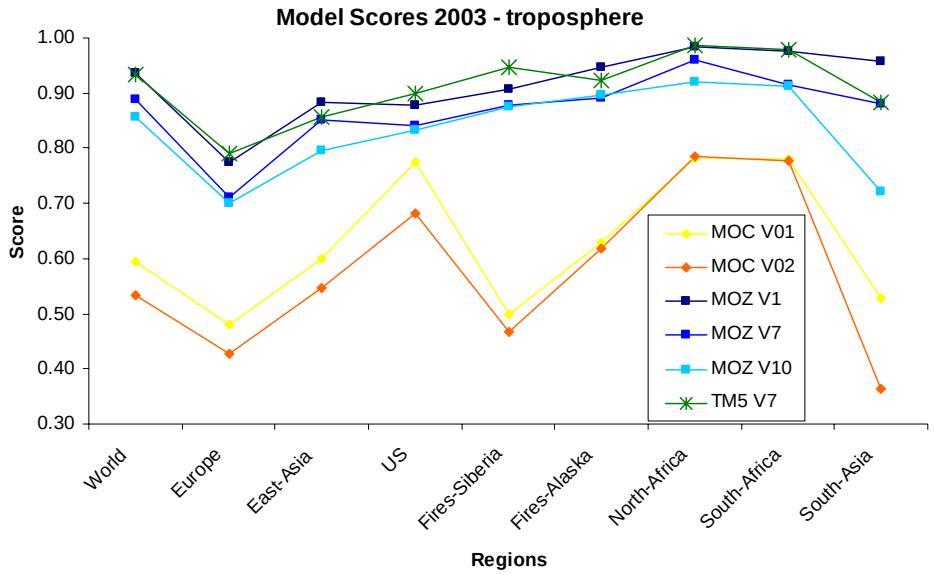


Figure 2.28: NO_2 annual average (for 2003) of scores for the tropospheric regions for the different standalone runs of the CTMs.

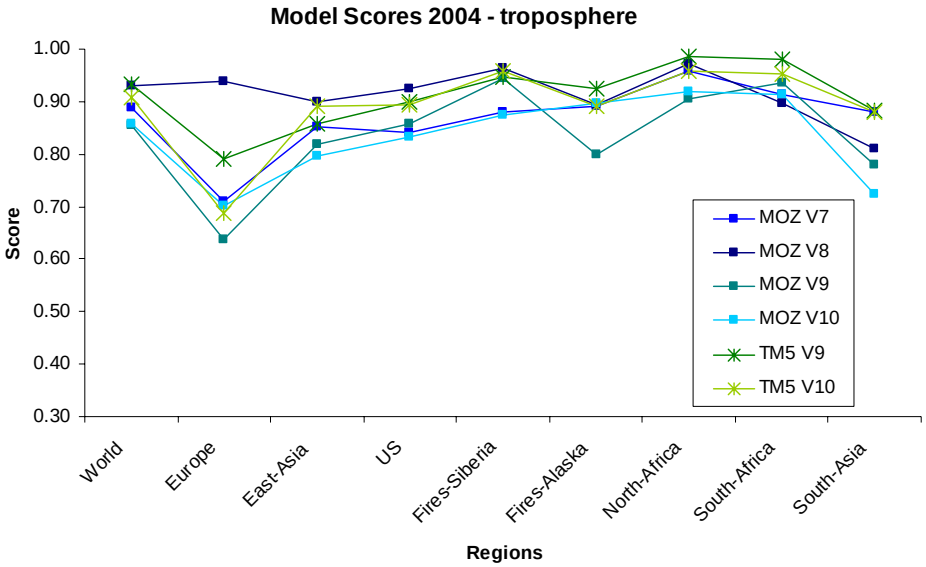


Figure 2.29: NO_2 annual averages (for 2004) of scores for the tropospheric regions for the different standalone runs of the CTMs.

Looking at the values from the graphs above one would believe that, in general, the models perform quite well in the boreal fire regions of Siberia and Alaska. This is in part true, but in fact, during the actual fire events (May 2003 for Siberia and August 2004 for Alaska), most of the model results and satellite data differ, with models overestimating the NO_2 emissions. This can be better identified when analyzing the monthly scores determined for a region and not simply considering an annual average. These values can be seen below.

Table 2.1: Scores obtained for NO_2 from the standalone model versions for the region "Fires-Siberia" in the period of April – July, 2003. May is highlighted as the month of the boreal fires in Siberia.

	MOC V01	***MOC V02***	***MOZ V1***	***MOZ V7***	***MOZ V10***	***TM5 V7***
April	0.57	0.53	0.91	0.82	0.82	0.97
May	0.00	0.00	0.82	0.45	0.44	0.88
June	0.47	0.45	0.92	0.92	0.90	0.95
July	0.79	0.74	0.98	0.91	0.90	0.98

From these tables we see that for almost every model run the scores on the month of the fire events are much lower than the rest of the year. During the month of the boreal fires in Siberia (May 2003), we see that both versions of MOCAGE have score 0, and both versions V7 and V10 of MOZART obtain scores of approximately 0.4 where they normally 0.8 or even 0.9. On the other hand, MOZART V1 and TM5 V5 still obtain reasonably good scores for this region: lower, but still close to the values in the rest of the year.

Table 2.2: Scores obtained for NO_2 from the standalone model versions for the region "Fires-Alaska" in the period of June – September, 2004. August is highlighted as the month of the boreal fires in Alaska.

	MOZ V8	***MOZ V9***	***MOZ V10***	***TM5 V9***	***TM5 V10***
June	0.88	0.85	0.85	1.00	0.95
July	0.92	0.87	0.87	0.98	0.97
August	0.71	0.63	0.63	0.95	0.88
September	-	1.00	1.00	-	1.00

From Table 2.2, we see that, in August 2004 (the month of the boreal fires in Alaska), all MOZART versions have lower scores than the previous and following months. TM5 however is able to simulate well the NO_2 columns over Alaska and therefore the score values in August for both versions are still quite high, i.e. 0.95 for TM5 V9 and 0.88 for TM5 V10.

This detailed analysis reveals once again the importance of looking further in detail at the origin of averaged scores (either by region or over time). One can consider an average for a quick initial analysis but to really understand the model performance the monthly scores should not be disregarded. For that reason, we strongly recommend further consultation of Annex 8 where the graphs with monthly scores for each of the model versions are presented.

2.3 GEMS-GRG reanalysis

2.3.1 MOZART f026 vs SCIAMACHY

In this section we present the results obtained from the analysis of the GRG-reanalysis model. The same evaluation steps that were performed for the standalone model runs were carried out for the NO_2 and HCHO modeled with the MOZART *f026* run (see more details in Annex 2, 5, and 6). The interesting point of this run is the long time period for which data is available (approximately 4 years) and that one can analyze the performance of the model in different years and whether or not the model is able to capture certain episodes.

a) NO_2

The figures below show a comparison of the global NO_2 maps for stratospheric and tropospheric columns averaged over a 3 month period. Again, due to the large amount of data analyzed, only a selection of some months is presented.

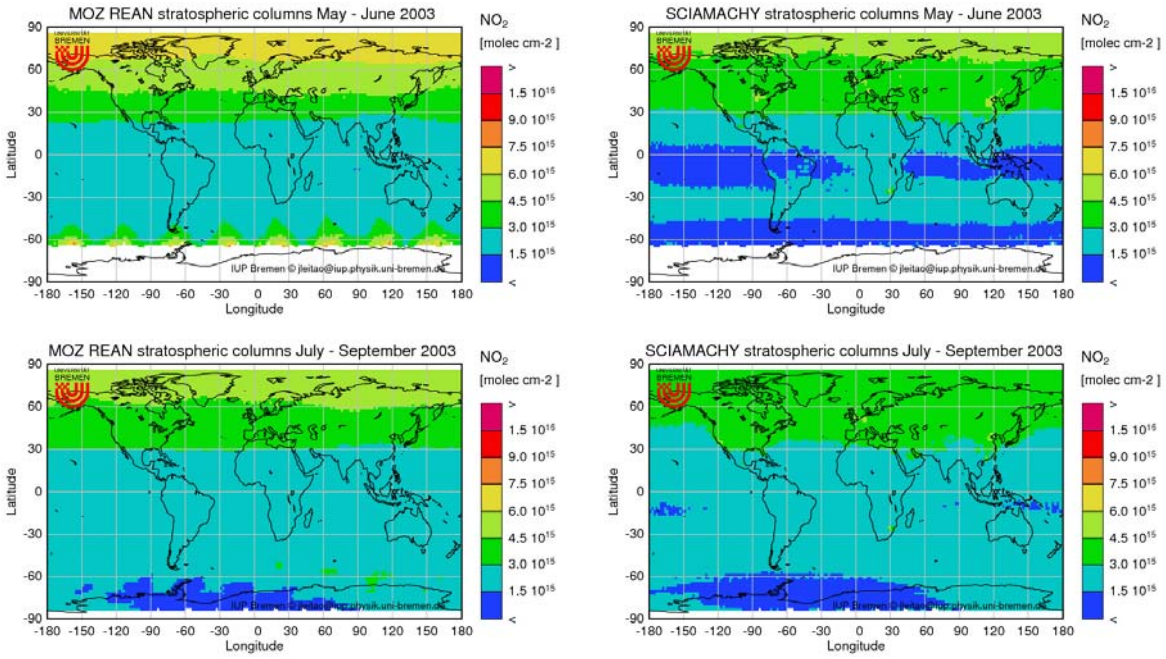


Figure 2.30: Three month averages (May-June and July-September 2003) of global stratospheric NO_2 determined by MOZART *f026* (left) and total column of NO_2 measured by SCIAMACHY (right).

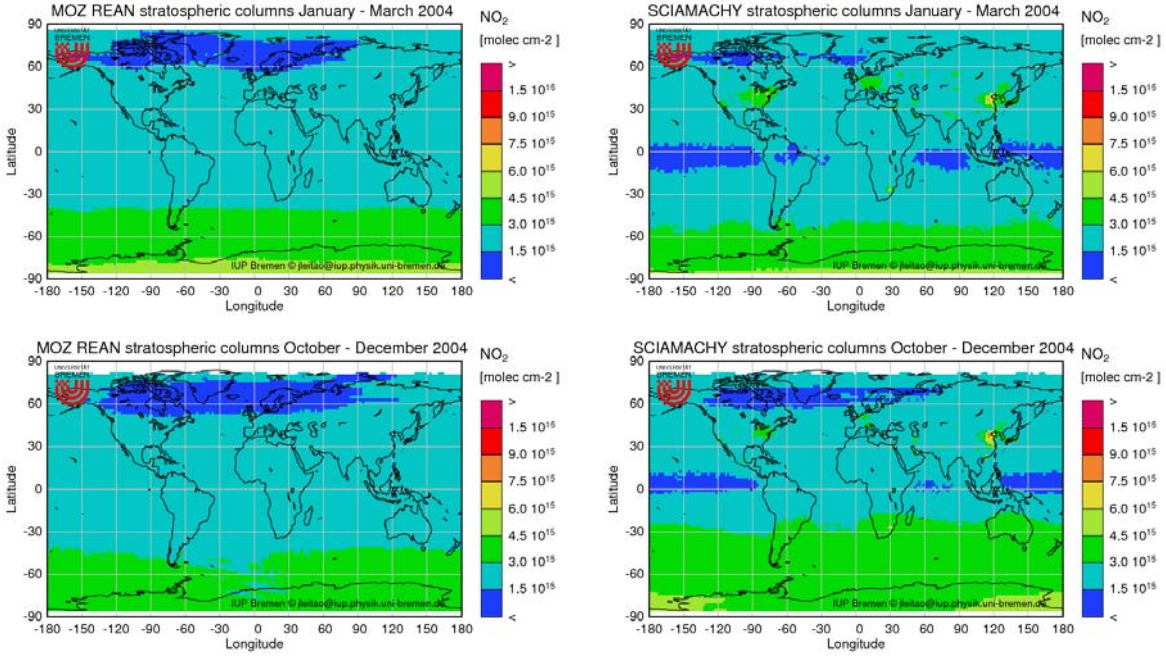


Figure 2.31: Three month averages (January-March and October–December 2004) of global stratospheric NO_2 determined by MOZART *f026* (left) and total column of NO_2 measured by SCIAMACHY (right).

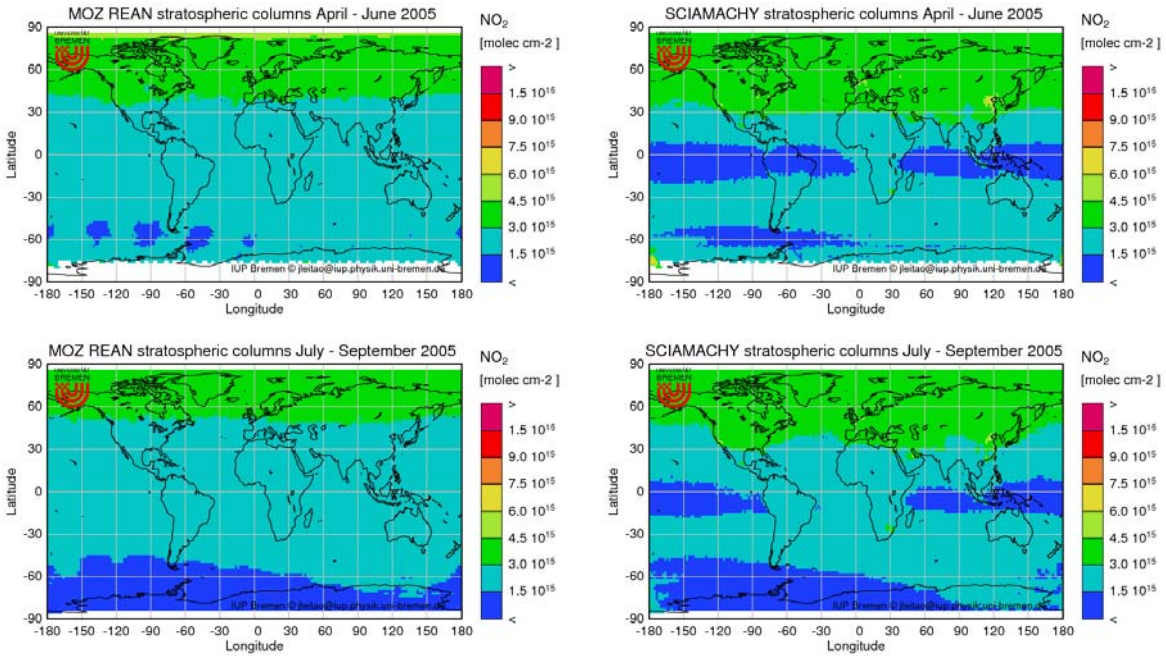


Figure 2.32: Three month averages (April-June and July-September 2005) of global stratospheric NO_2 determined by MOZART *f026* (left) and total column of NO_2 measured by SCIAMACHY (right).

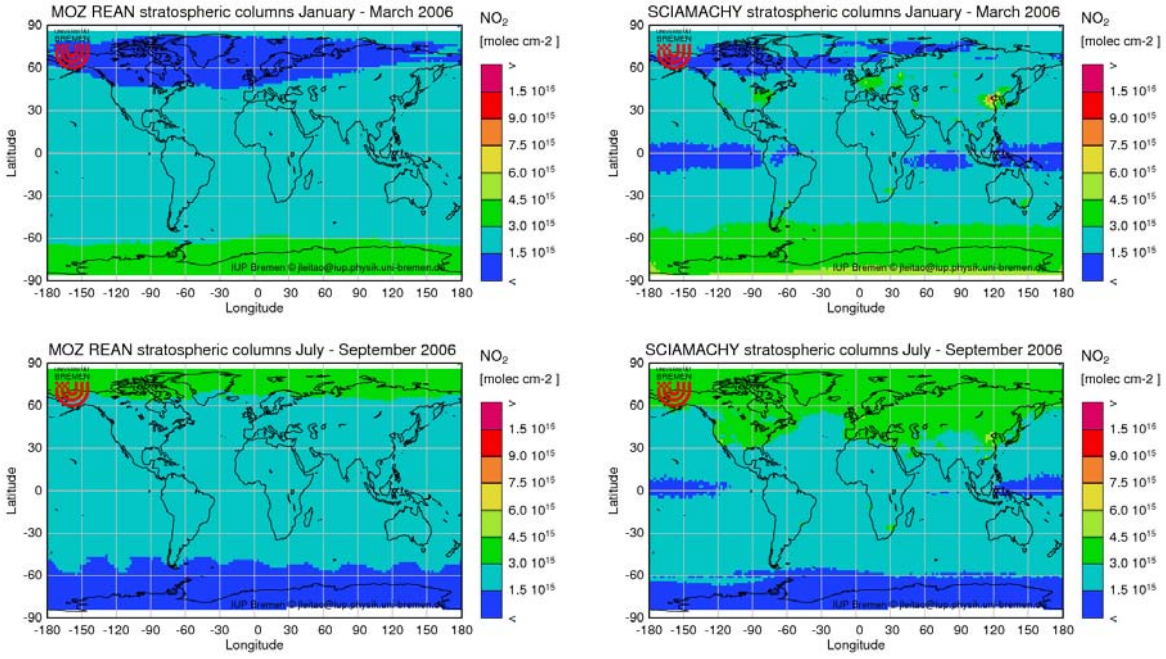


Figure 2.33: Three month averages (January-March and July-September 2006) of global stratospheric NO_2 determined by MOZART *f026* (left) and total column of NO_2 measured by SCIAMACHY (right).

Looking at the different maps presented above the first detail that it is important to notice is the very good agreement between model and satellite measurements. In the beginning of the time series (first months available of 2003) a problem in the stratosphere can be seen. This aspect was also identified in MOZART standalone V1 (the basis for this reanalysis run) and was then corrected in the *f026* run. Still, we see a systematic overestimation of stratospheric NO_2 in the tropics.

In some months, especially the period January-March (see Fig. 2.31 for 2004 and 2.33 for 2006), we can see the some hot-spots for highly polluted areas (particularly East-Asia) in the satellite data. As it was clarified before, this is simply due to the fact that the SCIAMACHY data presented here are not only the stratospheric signal of the measurements but total columns instead, i.e., tropospheric NO_2 also contributes to this amount. Therefore, some differences between satellite and model are expected in these areas.

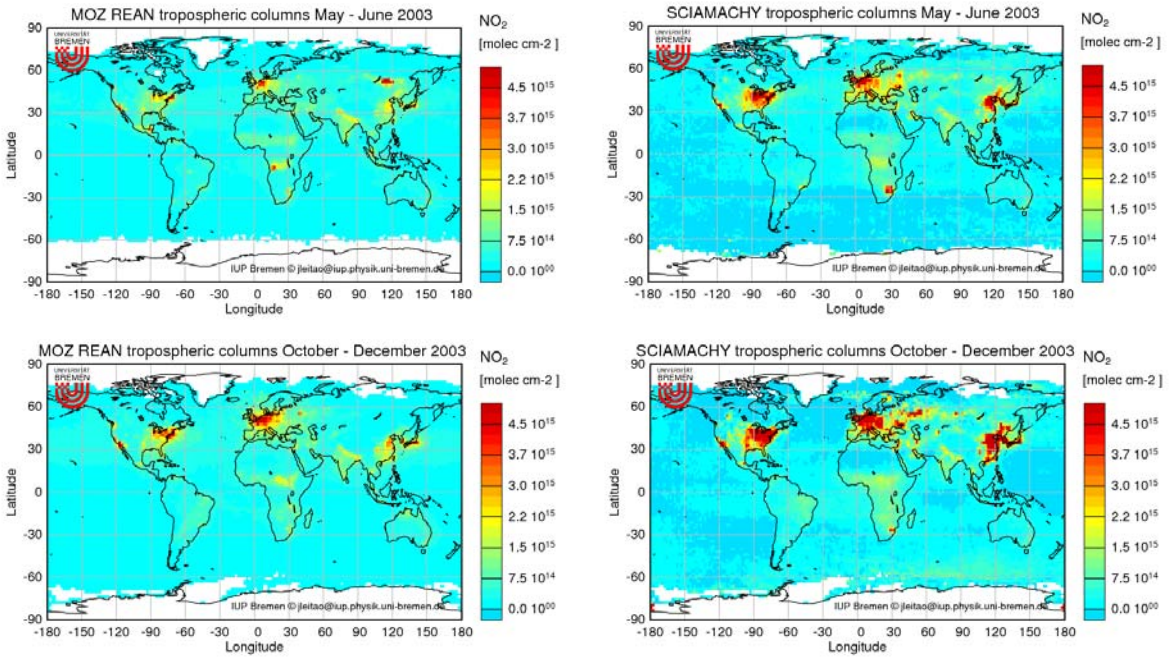


Figure 2.34: Three month averages (May-June and October-December 2003) of global tropospheric NO_2 determined by MOZART *f026* (left) and measured by SCIAMACHY (right).

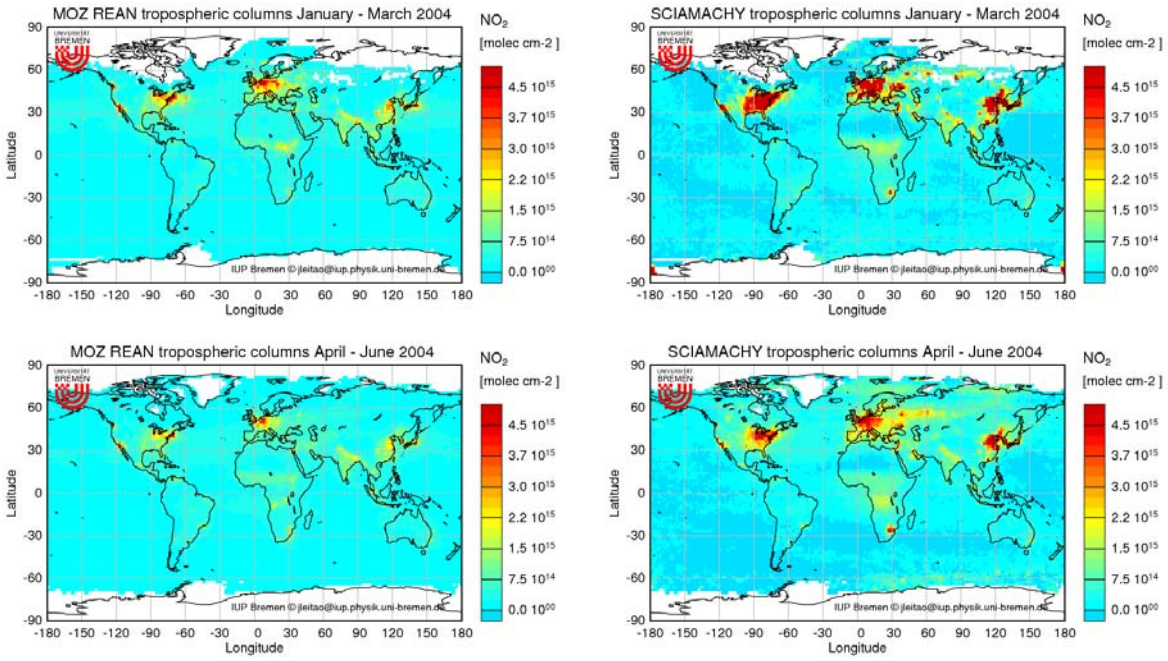


Figure 2.35: Three month averages (January-March and April-June 2004) of global tropospheric NO_2 determined by MOZART *f026* (left) and measured by SCIAMACHY (right).

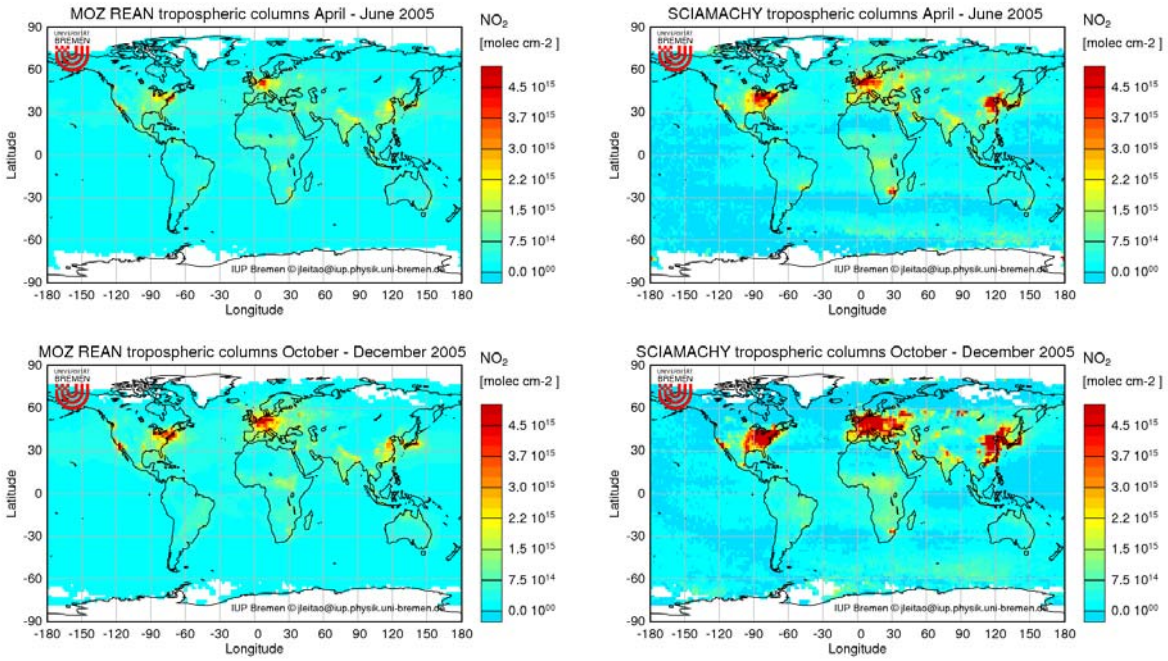


Figure 2.36: Three month averages (April-June and October-December 2005) of global tropospheric NO_2 determined by MOZART *f026* (left) and measured by SCIAMACHY (right).

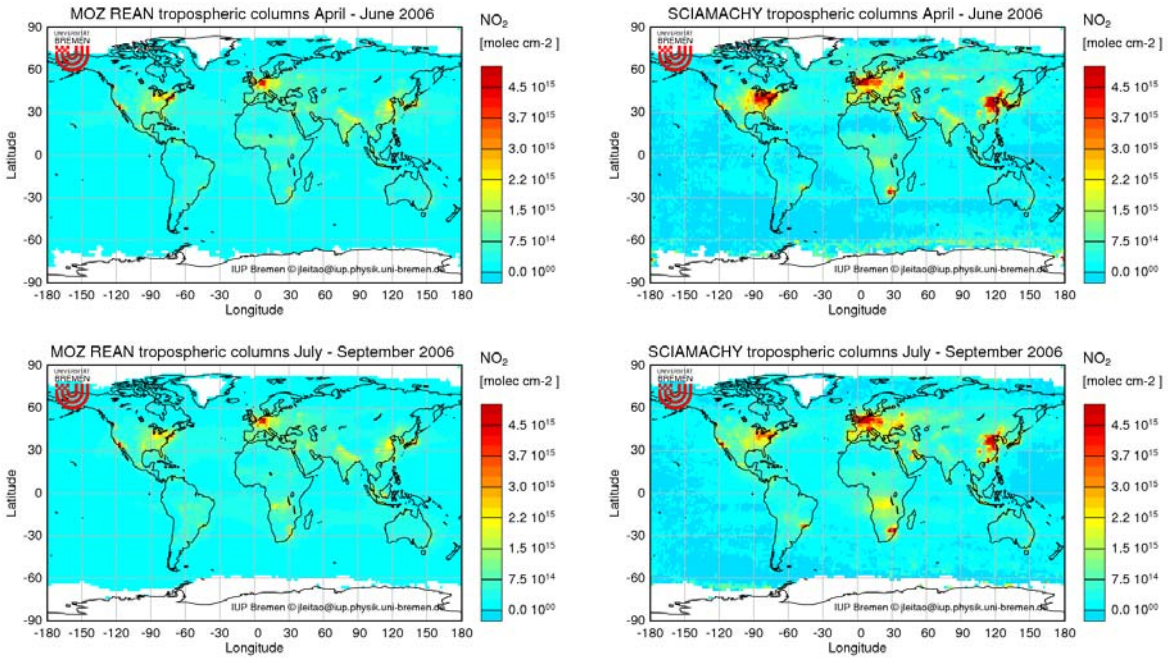


Figure 2.37: Three month averages (April-June and July-September 2006) of global tropospheric NO_2 determined by MOZART *f026* (left) and measured by SCIAMACHY (right).

For tropospheric NO_2, the main conclusion is that the models display a better match to the satellite data in summer than in winter. However, the overall performance is not satisfactory. Over the regions typically characterized by high anthropogenic emissions the model tends to underestimate the NO_2 column. Also, the biomass burning regions in central and southern Africa are not fully captured by the model. On the other hand, in the period of May-June 2003, we can see once more that the high emissions in the Siberia region (originating from large fires) are not confirmed by the satellite data. All these observations are in line with what was already expected, because, as we mentioned above, the reanalysis run is based on the MOZART V1 version that presented the same problems for the troposphere.

Seasonal Plots

In the following figures, the seasonal trends for the years 2004 and 2005 are compared between the models and the satellite measurements. The seasonality is presented for the selected latitude bands of the stratosphere and pre-defined regions for the tropospheric data (see Section 2.1 for a more thorough description of these regions).

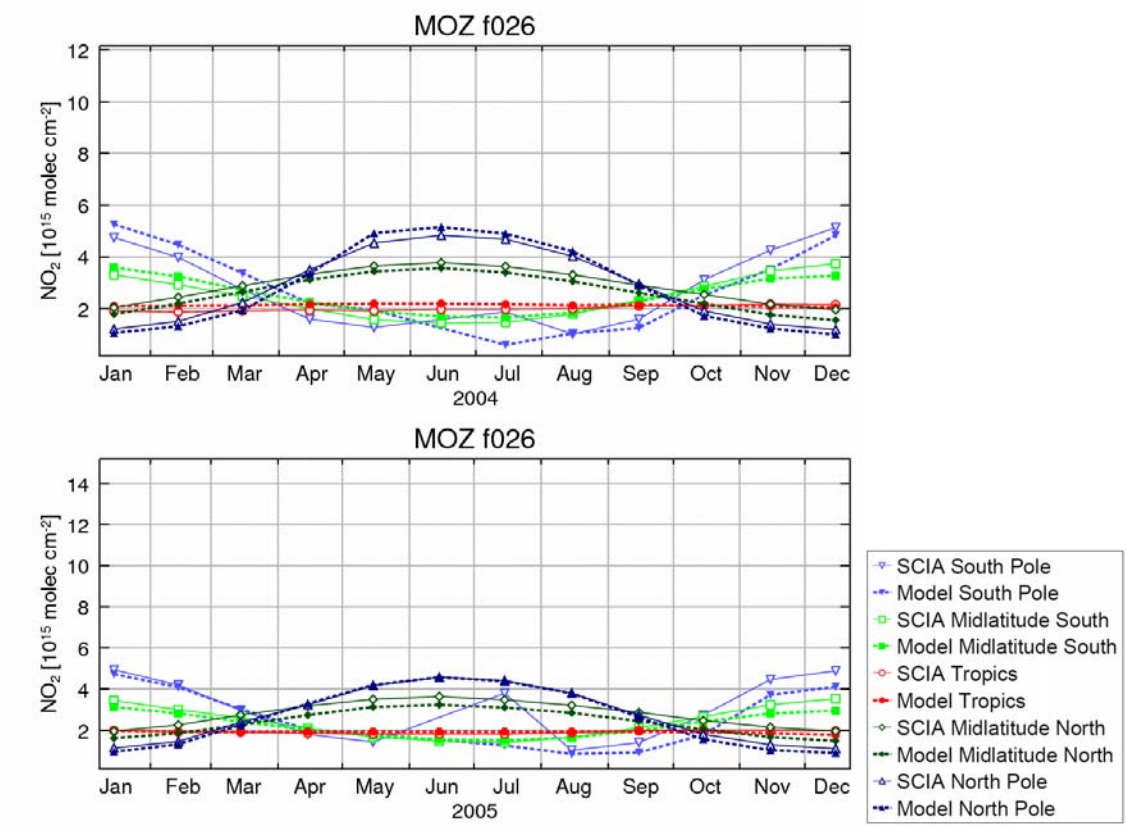


Figure 2.38: Seasonality curves for 2004 (top) and 2005 (bottom) of stratospheric NO_2, in pre-defined regions, determined by the reanalysis run of MOZART *f026* and total NO_2 columns measured by SCIAMACHY (open symbols).

From figure 2.38 above, we confirm that MOZART *f026* generally performs quite well when simulating stratospheric NO_2. Apart from small differences over the South Pole region in the summer months (local winter) we see almost a perfect match between measurements and model results.

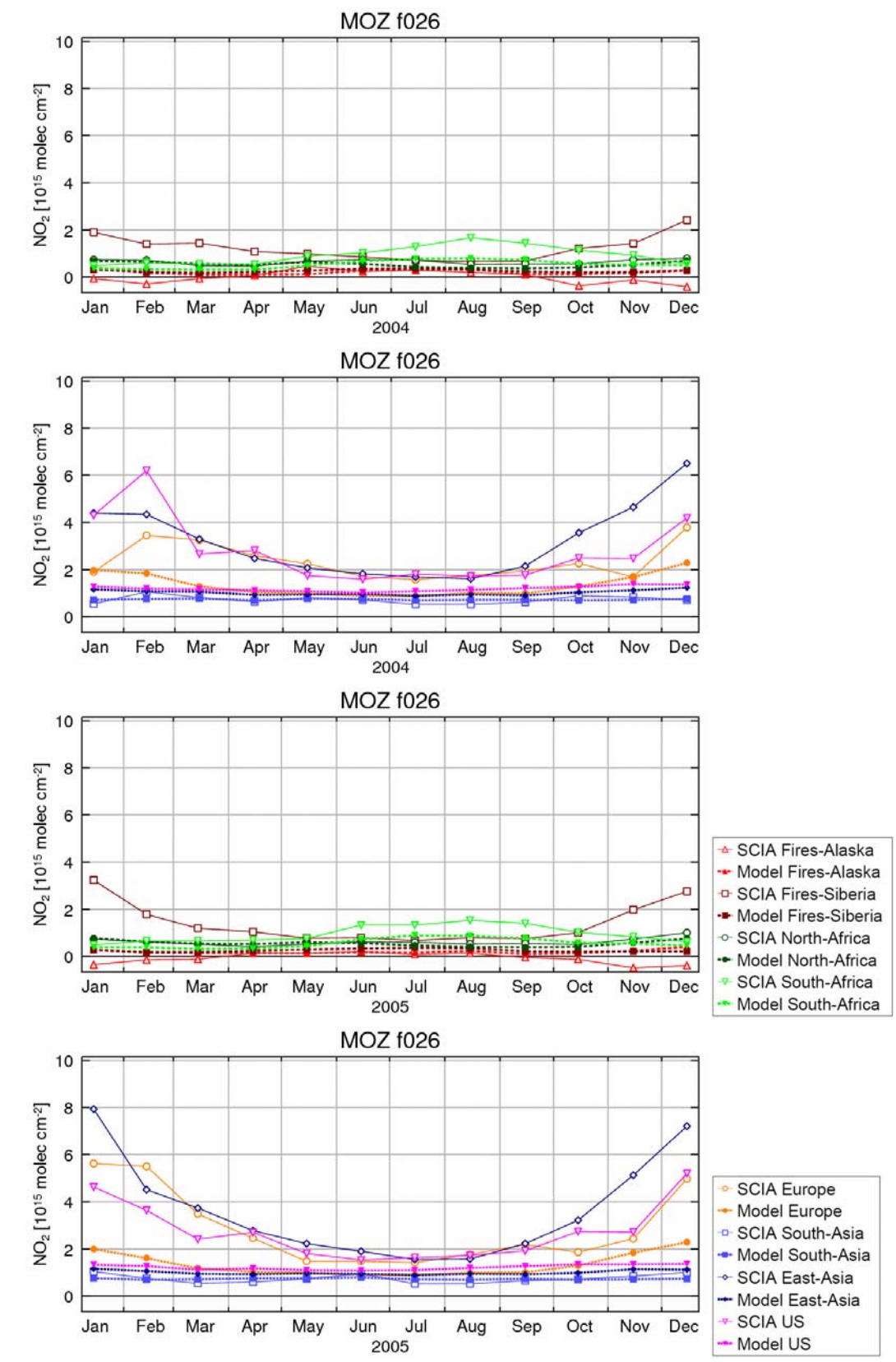


Figure 2.39: Seasonality curves for 2004 (top 2) and 2005 (bottom 2) of tropospheric NO_2, in pre-defined regions divided into the typical biomass burning and polluted areas, determined by the MOZART *f026* and measured by SCIAMACHY (open symbols).

However, as it was already mentioned before, the tropospheric results are not so satisfying. Over biomass burning regions the differences are not so large (the more obvious is over South-Africa), but the seasonal cycle over polluted areas is not well reproduced in the simulations. At this point, one should note that in the more recent years, in the winter, the satellite presents high NO_2 values in the "Fires-Siberia" region. This, however, does not originate from boreal fires but mostly from anthropogenic emissions in China (this is also observed in the 2006 seasonal curve not presented here).

Monthly scatter Plots

From the scatter plots presented in Fig. 2.40, we can see how well the model data correlate with the satellite measurements in the different regions analyzed and in the different months.

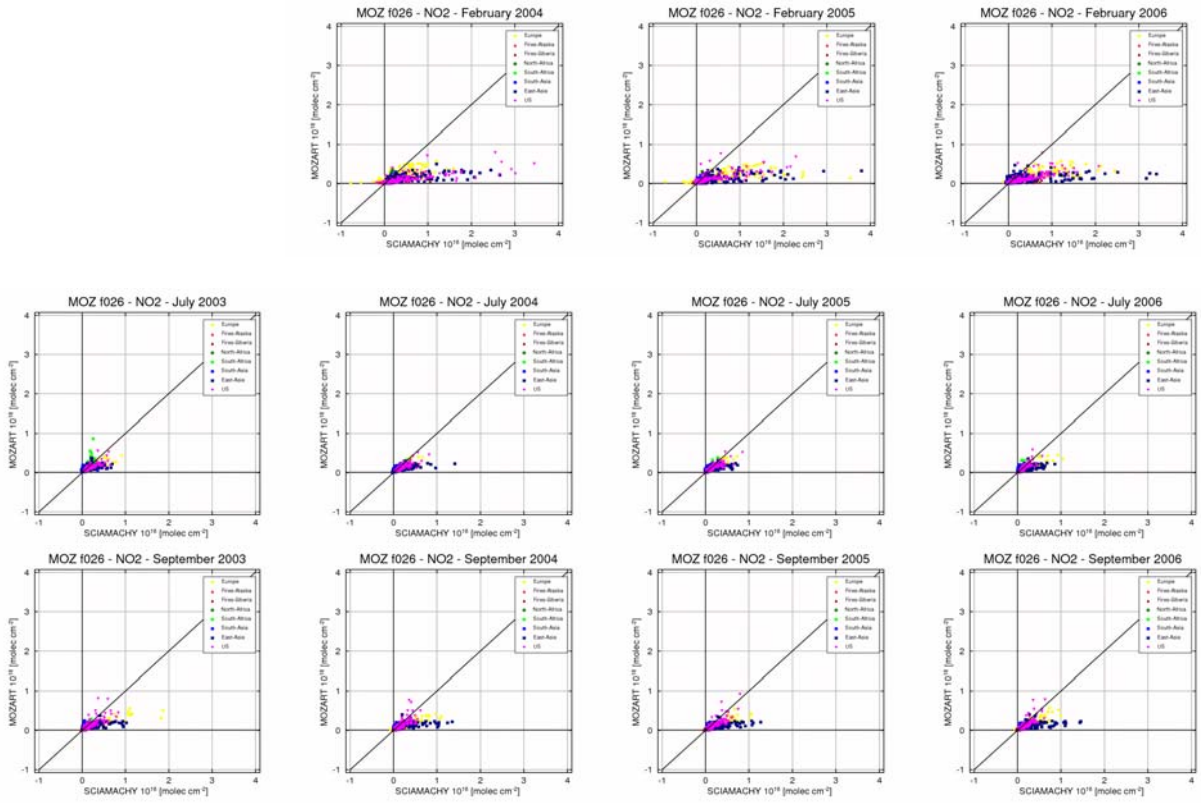


Figure 2.40: Scatter plots of monthly average of tropospheric NO_2 for February, July and September for each for the years from 2003-2006 (left to right) determined by MOZART *f026*s and measured by SCIAMACHY.

February is presented here as an example of the winter months. For all the years evaluated and in the period from November till February the data dispersion is very similar, leading us to the conclusion that over Europe, East-Asia and the US, the model keeps underestimating the NO_2 values. This scenario is quite different for the rest of the months. As can be seen, in July there is better agreement between model data and satellite measurements. This is always the case for the period from May to August.

b) HCHO

In Section 2.1, the retrieval method of tropospheric HCHO measured by SCIAMACHY was briefly described. Compared to NO_2, the HCHO signal in the satellite measurements is weak resulting in much larger scatter. Also, at low sun the data cannot be analyzed with sufficient accuracy resulting in data gaps over the winter hemisphere at mid and high latitudes. For HCHO there is no need to correct for the stratospheric component which, in the case of NO_2, leads to some uncertainty and occasionally to (unphysical) negative

columns. Nevertheless, systematic biases exist in the HCHO retrieval over some regions which can also result in negative columns. Another systematic feature in the satellite derived fields is the significant amounts of HCHO observed over some oceanic regions. The origin of this HCHO is not yet fully understood but both biologic sources in the ocean and secondary production from long-lived precursors in the gas or aerosol phase have been proposed. As these processes are not well understood, they are often not included in current models which can therefore not reproduce the enhanced HCHO (and glyoxal) columns over water. In summary, the comparison of modeled and measured HCHO fields is more difficult than that for NO_2 and is in parts more qualitative than quantitative.

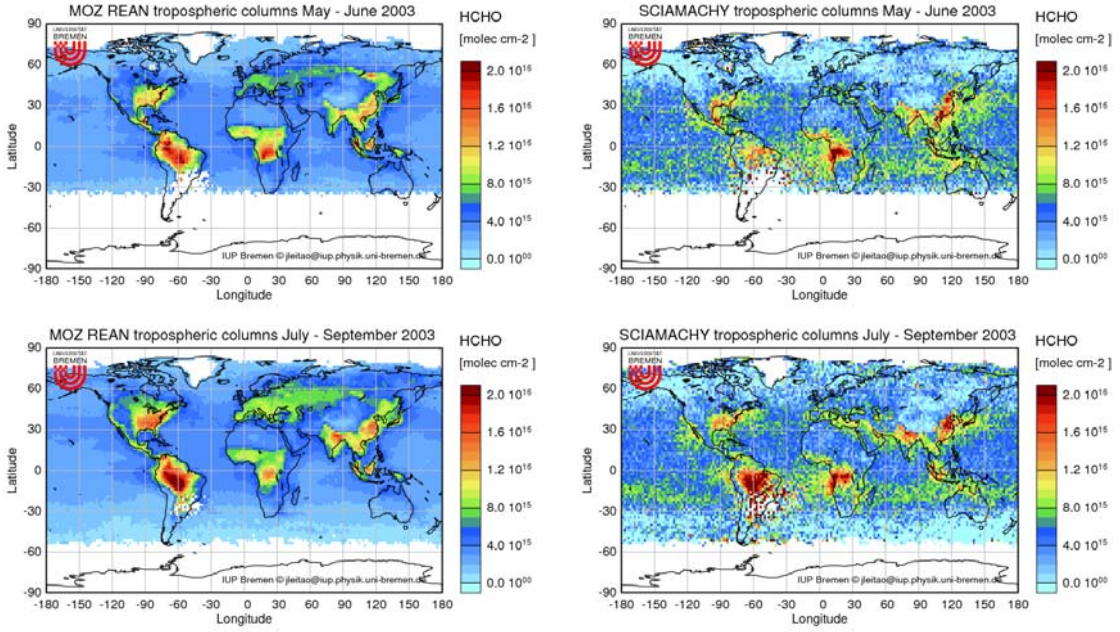


Figure 2.41: Three month averages (May-June and July-September 2003) of global tropospheric HCHO determined by MOZART *f026* (left) and measured by SCIAMACHY (right).

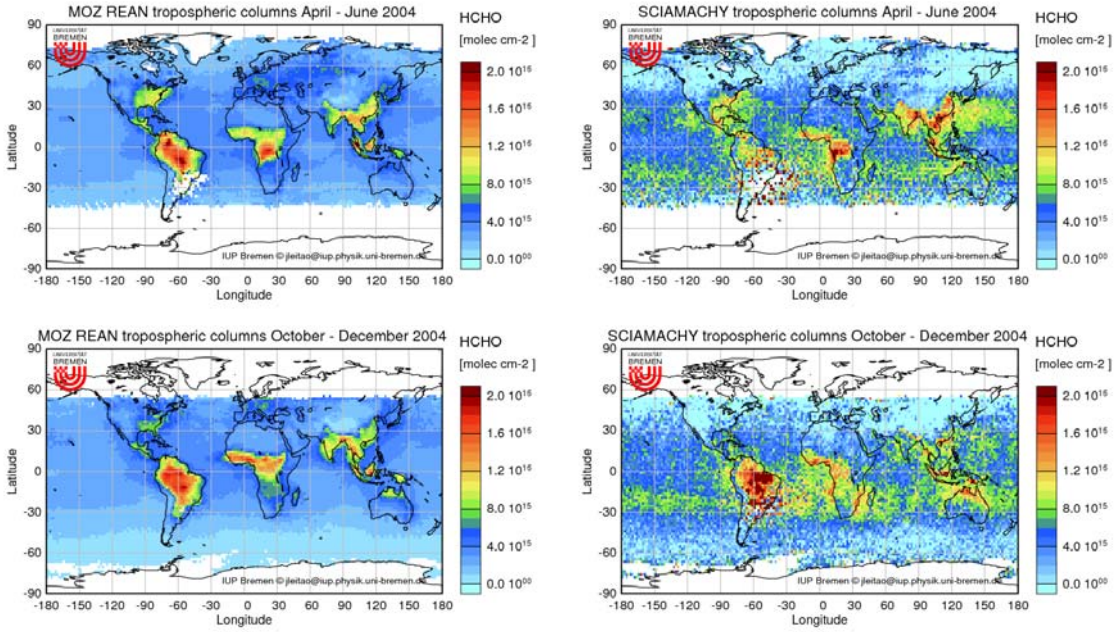


Figure 2.42: Three month averages (April-June and October-December 2004) of global tropospheric HCHO determined by MOZART *f026* (left) and measured by SCIAMACHY (right).

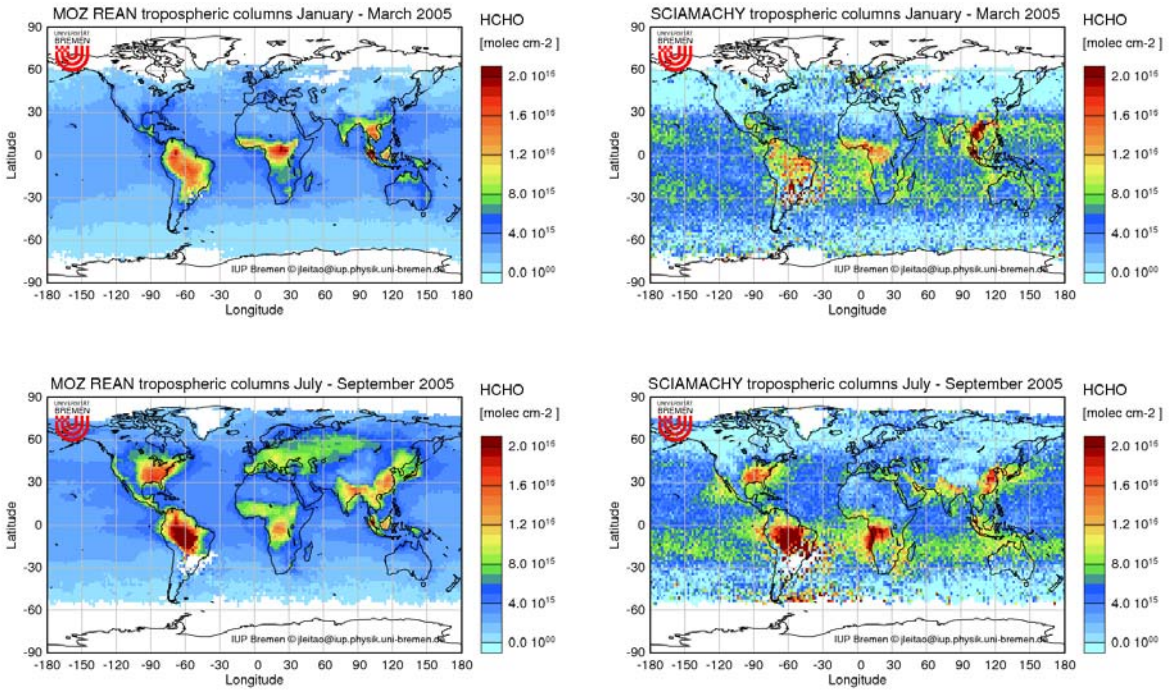


Figure 2.43: Three month averages (January-March and July-September 2005) of global tropospheric HCHO determined by MOZART *f026* (left) and measured by SCIAMACHY (right).

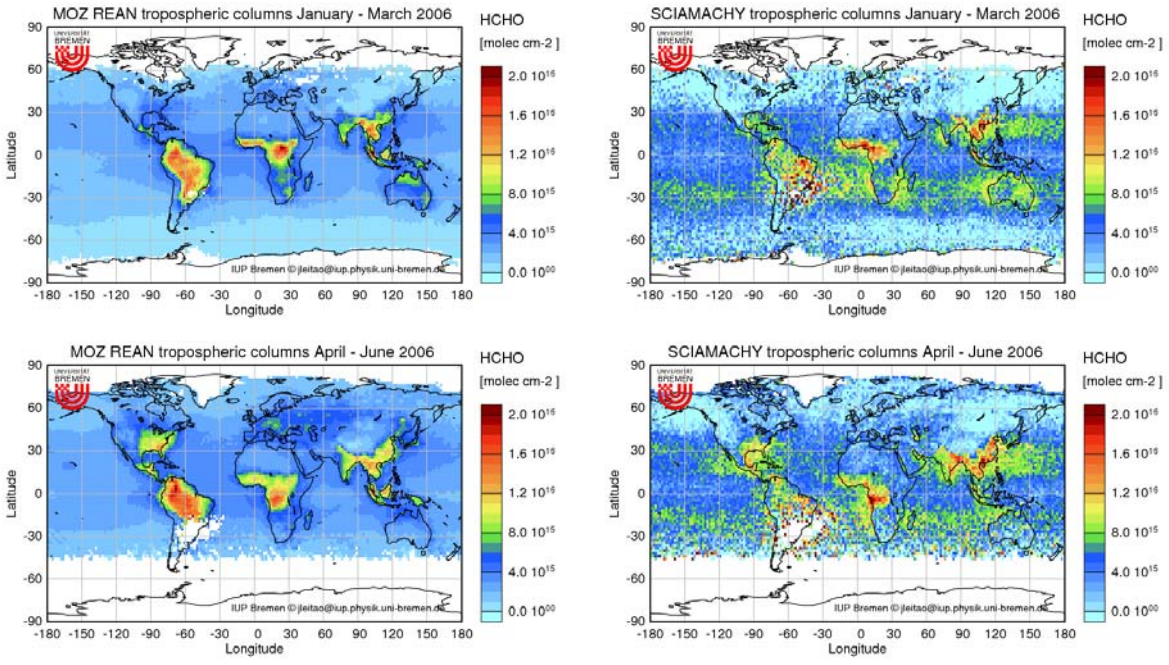


Figure 2.44: Three month averages (January-March and April-June 2006) of global tropospheric HCHO determined by MOZART *f026* (left) and measured by SCIAMACHY (right).

The figures presented above show that, in general, the model performs quite well in reproducing the overall pattern and absolute columns. The high HCHO values measured over the South American continent are nicely modeled, as well as the hot-spot over central Africa. In the South-Asia region, the HCHO values measured by the satellite are systematically higher than those provided by MOZART. In addition, the model presents almost no transport of HCHO or its precursors. How much (exactly) HCHO is transported away from sources is yet to be proven but it is certain that the satellite measures HCHO over the Atlantic and also further away from the Chinese coast. Many of these patterns are not captured by the model simulations and where there are similarities, the absolute values are much larger in the satellite measurements.

Over the urban areas the HCHO concentrations are not very high and therefore the signal of the satellite measurements is mostly low. In Europe, for example, the model tends to overestimate the HCHO tropospheric columns (see also for example Fig. 2.41, for the year 2003). In certain regions of the US (south-east), high values of HCHO are found, which are of biogenic origin, and the model results are very good here (see Fig. 2.47 with seasonal trends for the regions selected specifically for HCHO).

Seasonal Plots

In the seasonal plots shown below we can analyze more in detail the HCHO calculated by the MOZART reanalysis run. We present figures only for 2004 and 2005 because the other 2 years analyzed (2003 and 2006) were not complete.

The pre-defined regions (presented in section 2.1) selected for the evaluation process are not the most suitable for HCHO analysis as they were not selected to include the locations where higher HCHO columns are normally registered (e.g. South America) but rather include many urban locations which are not so relevant for HCHO. In addition, for regions like “Fires-Siberia” and “Fires-Alaska” there are no data available in some months and in the rest of the year many of the monthly averages are affected by systematic biases in the satellite data making comparison to model data difficult. For the sake of consistency in the evaluation we show here the trends for all these regions, and some conclusions can be drawn from there. But in addition, below, we defined 5 additional regions that are more suitable for an analysis of HCHO tropospheric columns.

We see that, both in 2004 and 2005, the model is able to reproduce the seasonal cycle of HCHO over most of the selected regions, both those affected mainly by biogenic emissions and biomass burning and those classified as polluted. As already mentioned above, over certain regions, like Europe, the model results are systematically higher than the satellite measurements which could in part be related to biases in the measurements as is evident from the negative columns in winter. There is also an interesting deviation of the model from the observations over the “South-Africa” region where the values are underestimated in both years.

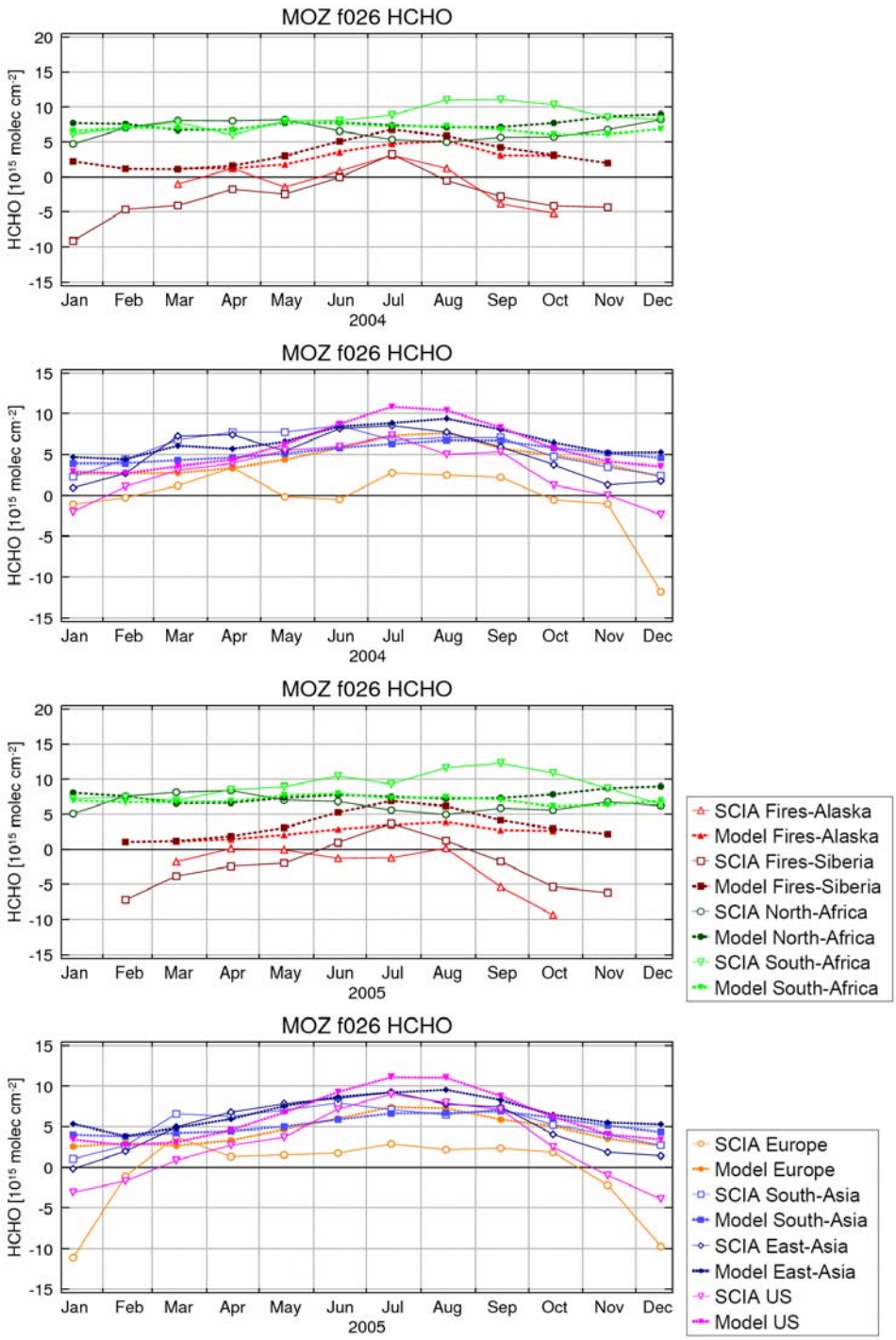


Figure 2.45: Seasonality curves for 2004 (top 2) and 2005 (bottom 2) of tropospheric HCHO, in pre-defined regions divided into the typical biomass burning and polluted areas, determined by MOZART *f026* and measured by SCIAMACHY (open symbols).

Adapted regions for HCHO

As mentioned above, for the case of HCHO, the selected regions were not the most appropriated ones for the evaluation process, as this selection disregards some areas where usually high emissions of HCHO are registered or, in other cases (like US and East-Asia) the selected area was too wide. Therefore, an additional analysis was performed for what we consider to be typical regions with high HCHO values. The figure below shows these 5 regions: North-Africa, South-Africa, China, USA and Indonesia.

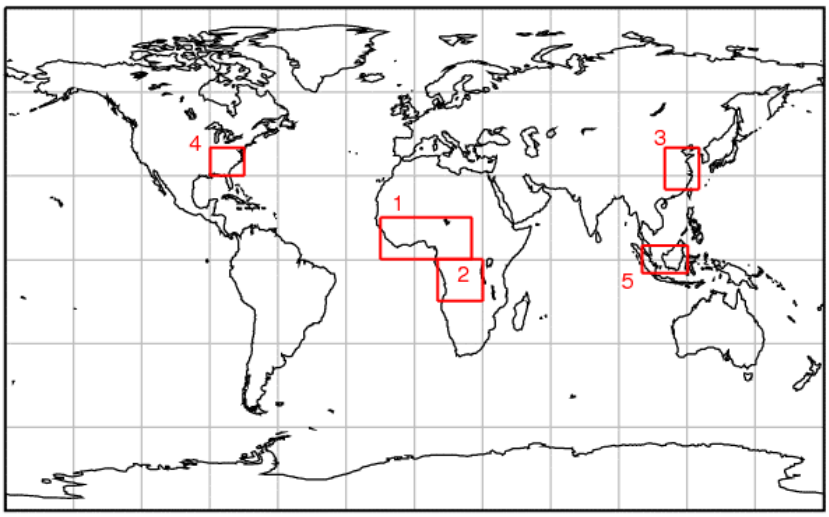


Figure 2.46: Map with the new regions selected (and adapted) especially for the evaluation of tropospheric columns of HCHO model results: 1) North-Africa (HCHO), 2) South-Africa (HCHO), 3) China (HCHO), 4) USA (HCHO), and 5) Indonesia (HCHO). Regions 1, 2 and 5 correspond to typical biomass burning areas, and 3 and 4 are regions with mainly biogenic and some anthropogenic emissions.

For these regions, we decided to present the graphs with the seasonal variation trends (see Figure 2.47 below). We can see that both in the year 2004 and 2005 the model and the measurements are quite close, except for South-Africa, where larger differences are observed, and China, where some larger discrepancies are found in the summer of 2005. In general, it is possible to conclude that the model is able to capture the seasonal variation of HCHO emissions over these regions.

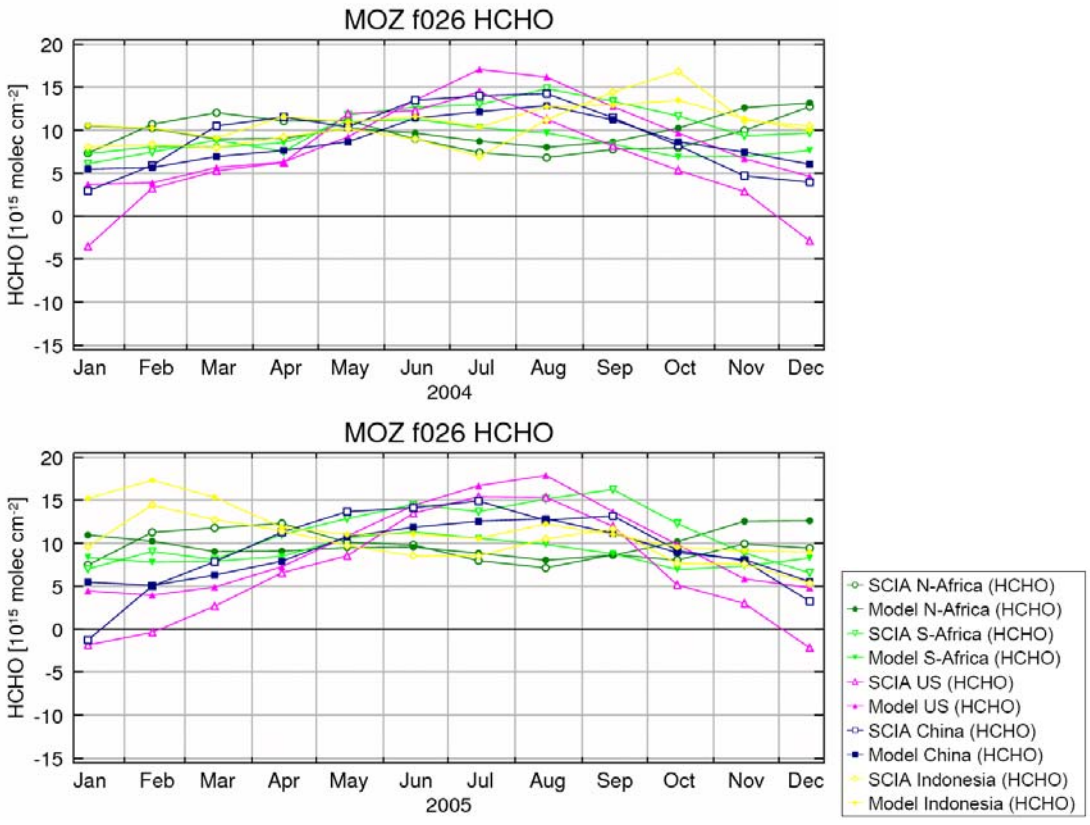


Figure 2.47: Seasonality curves for 2004 (top) and 2005 (bottom) of tropospheric HCHO, in the special regions defined for HCHO, determined by the MOZART *f026* and measured by SCIAMACHY (open symbols).

Monthly scatter Plots

The scatter plots below focus mainly on the summer months of the dataset available for HCHO. The reason for this choice is that, in winter, the satellite data have too many negative values which are not useful for the model evaluation. The plots reveal that the model behaves different in each of the regions considered but that the overall scenario is quite good. We can see that, for each year, the data from South-Africa (light green) has more dispersed data points in June. For this same region, one can see that, from August onwards (rest of the months not presented here), and especially in the latest years, the values are more underestimated than in the other months. The US data (magenta), for example, nicely fit the 1 to 1 line, with shifts from over to underestimation from year to year.

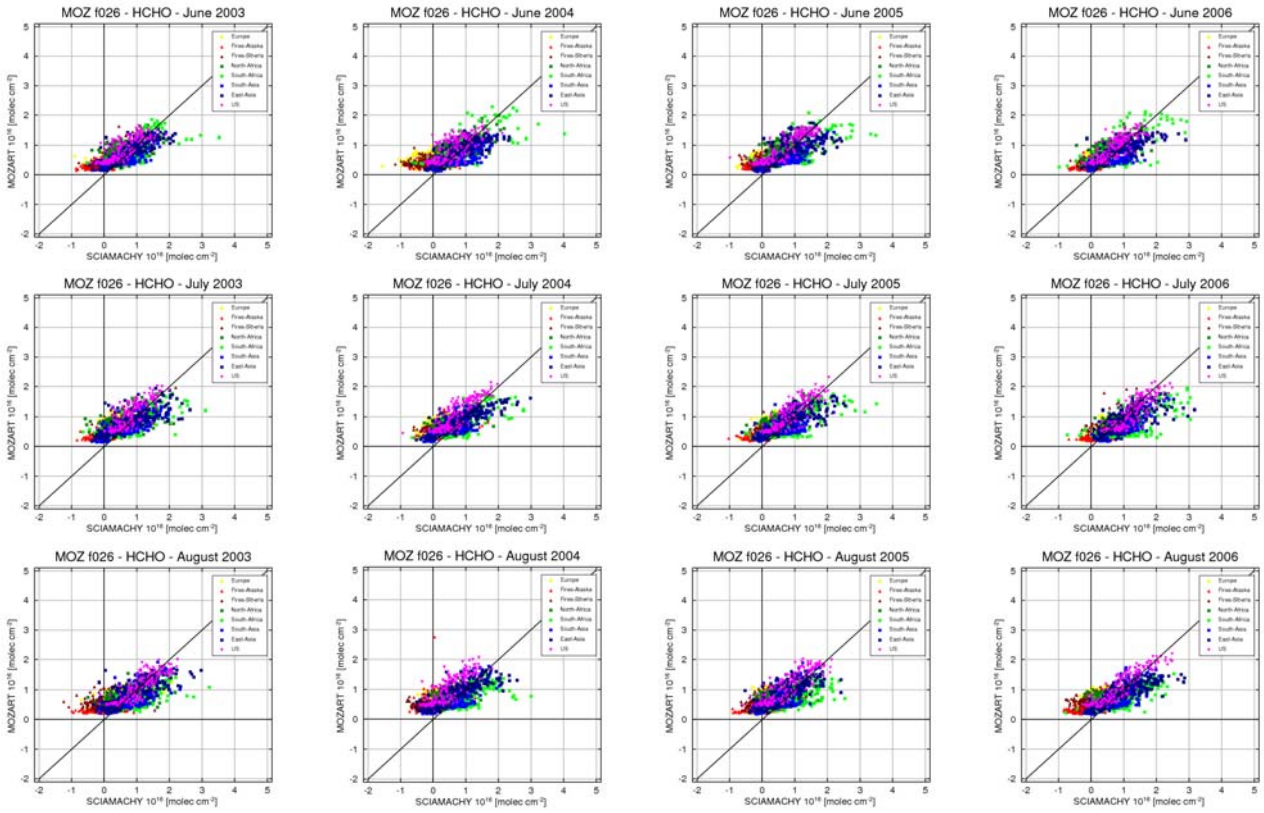


Figure 2.48: Scatter plots of monthly average of tropospheric HCHO for June, July and August for the years from 2003-2006 (left to right) determined by MOZART *f026* and measured by SCIAMACHY. Different colors represent the different models.

Scores

As a final step in the evaluation of the MOZART reanalysis run (*expid f026*), the scores for both NO_2 and HCHO were calculated. We show those in the following graphs (with HCHO scores only in the troposphere).

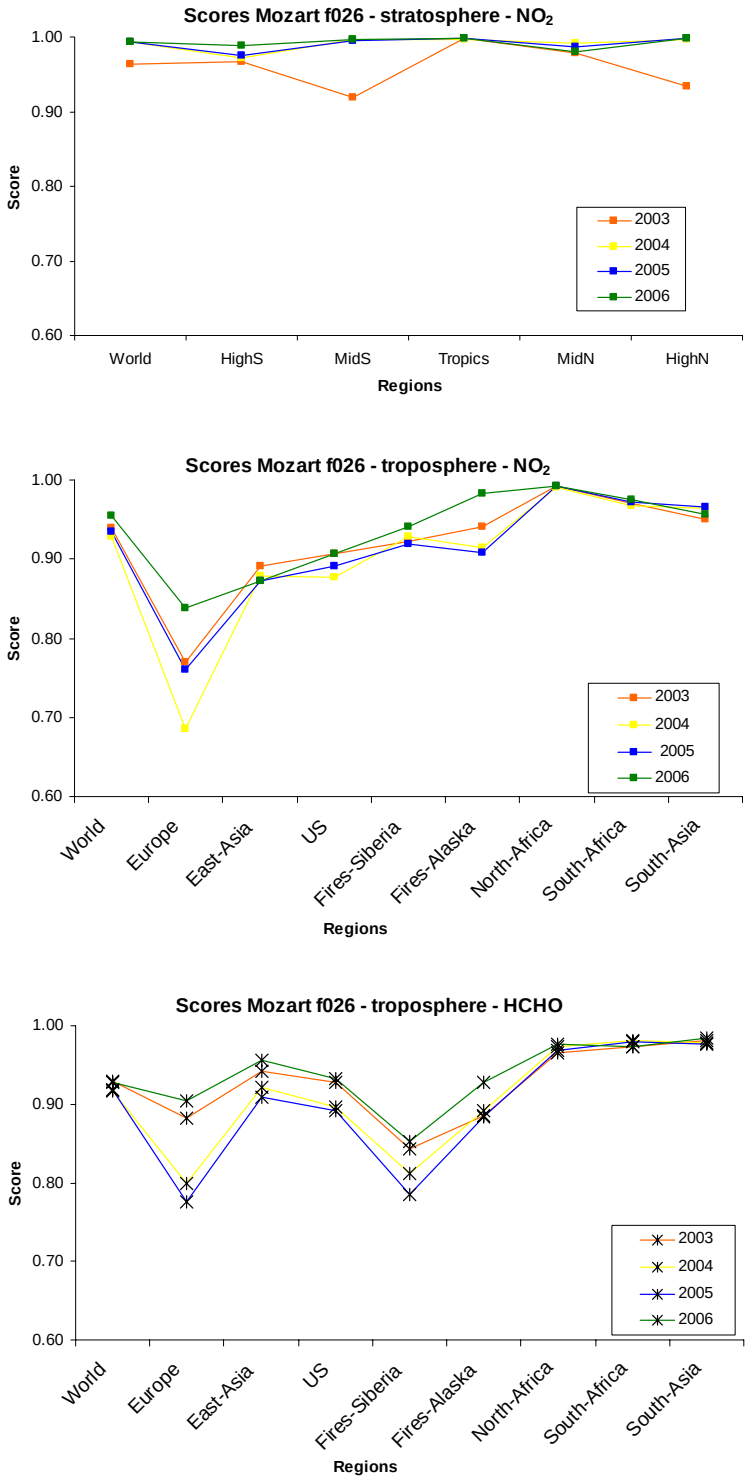


Figure 2.49: NO_2 and HCHO (only troposphere) annual average scores (for the years 2003-2006) obtained for the MOZART reanalysis run *f026*, for the different regions of the stratosphere (top) and troposphere (middle for NO_2 and bottom for HCHO).

From these graphs, it is clear that the model performance is quite good for stratospheric NO_2, with an average score per region in all years (except 2003) always close to 1. The tropospheric case is quite different. As we have seen before, the emissions of NO_2 and HCHO vary strongly between regions and the model performance for the two compounds cannot be compared directly. Still, we see that the lower scores of each year for both compounds are found over Europe, with the minimum in the year 2004: NO_2 with 0.69 and HCHO with 0.80. Actually these lower values over Europe were found for almost all models and versions analyzed. The best scores are achieved in the North-Africa region for both the NO_2 and HCHO columns.

Comparison of these scores with those obtained for the standalone MOZART versions illustrates once more the inheritance of V1 in the reanalysis values as the scores for the tropospheric NO_2 determined by both V1 and *f026* are very similar.

2.4 GEMS-GRG forecast

In this section, we present the MOZART data from *expid ez2m* for the evaluation period February–December 2008. Details about this run are presented in Annex 6 and were also repeated in Section 2.1.

2.4.1 MOZART vs SCIAMACHY

For the comparison of the NO_2 measurements performed by SCIAMACHY and the data from the MOZART forecast run we consider the period February – November, 2008. From the global stratospheric fields presented in the figure below we conclude that for the majority of the year 2008 the NO_2 data calculated by the MOZART-IFS system capture the seasonality of the stratospheric values. An overestimation of the columns was identified for the first months. This was the same problem identified in version V1 that was corrected in the forecast runs and, therefore, from July on, the model results are very good. There are still some minor differences with too high values in the tropics and high latitudes, but the main features are correct.

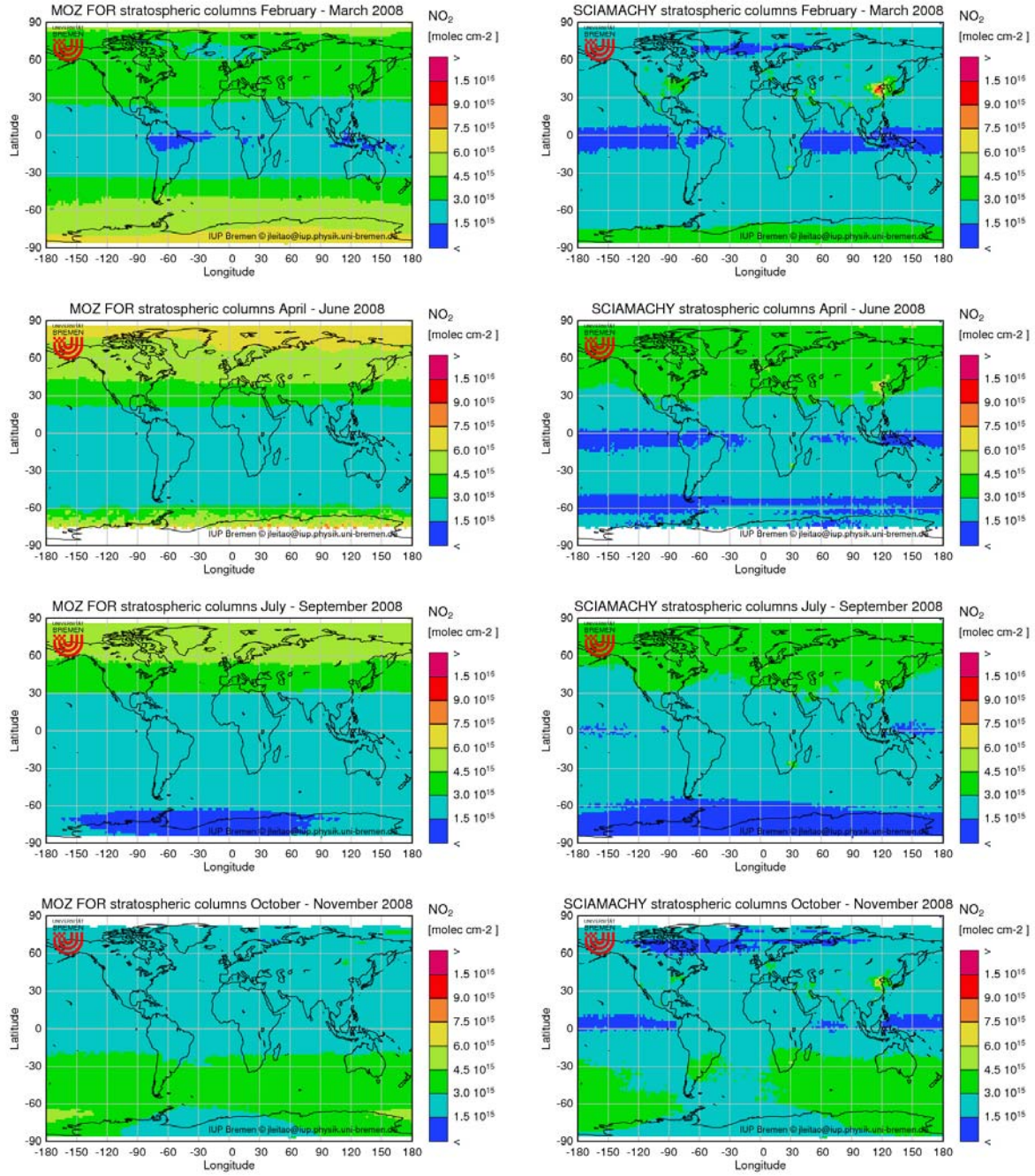


Figure 2.50: Three month averages of global stratospheric NO_2 columns determined by MOZART-IFS run *ez2m* (left) and total column of NO_2 measured by SCIAMACHY (right), for the year 2008 (excluding January and December).

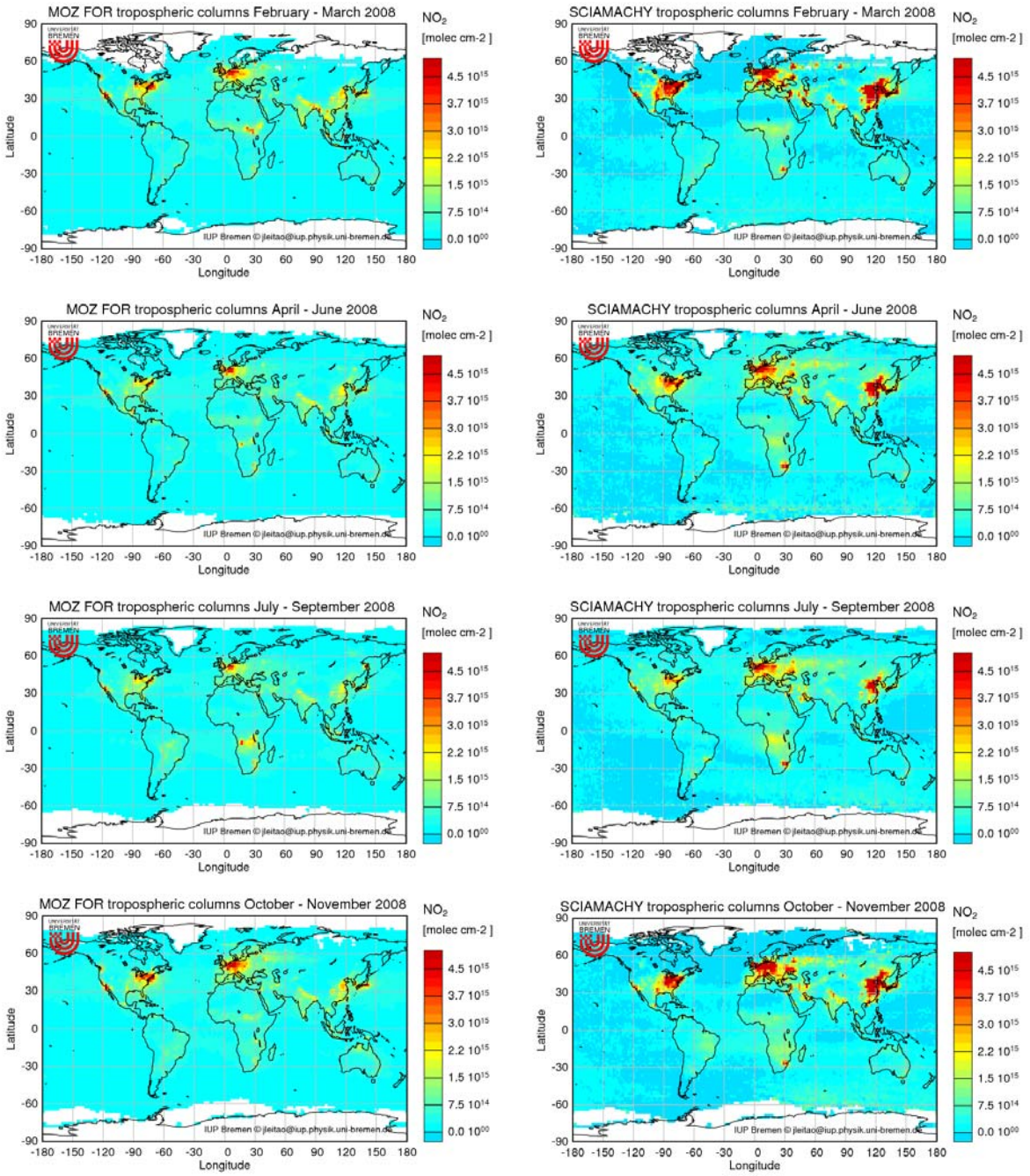


Figure 2.51: Three month averages of global tropospheric NO_2 determined by MOZART-IFS run *ez2m* (left) and measured by SCIAMACHY (right), for the year 2008 (excluding January and December).

In Fig. 2.51, the tropospheric fields from SCIAMACHY and MOZART-IFS forecast run are compared. Two conclusions can be drawn from these figures: while the overall pattern is similar, the forecast run of MOZART mostly underestimates the NO_2 emitted over polluted regions like Europe and China and it presents too high emissions over biomass burning regions in particular in Africa. This is in line with all other versions of MOZART that were analyzed before.

Seasonal Plots

In Figs. 2.52 and 2.53, the seasonality of stratospheric and tropospheric NO_2 (over regions of active biomass burning and polluted areas) is shown for SCIAMACHY and the forecast run of MOZART (*expid ez2m*). The seasonality of the stratospheric NO_2 columns is nicely modeled by MOZART-IFS. However, in the beginning of the year the differences between measurements and model were much higher than in the last months, the model overestimating the satellite columns.

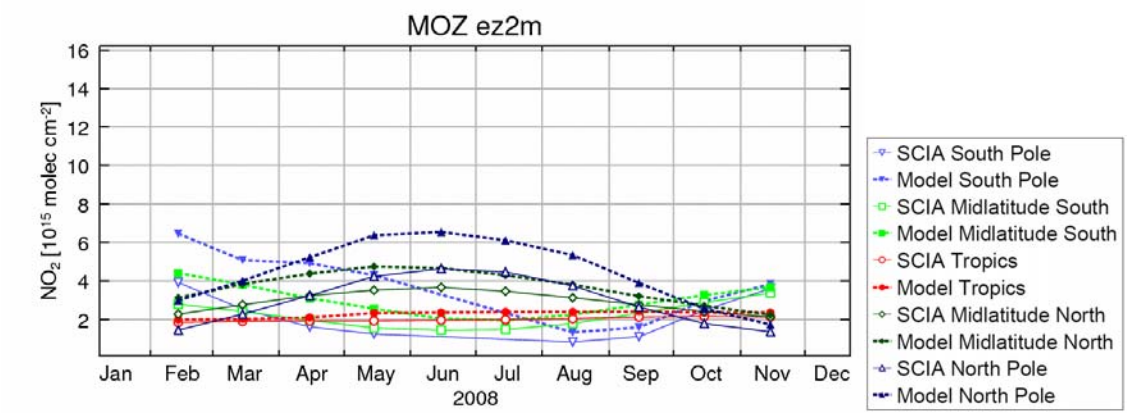


Figure 2.52: Seasonality curves for 2008 of stratospheric NO_2, in pre-defined regions, determined by the MOZART *ez2m* (forecast run) and total NO_2 columns measured by SCIAMACHY (open symbols).

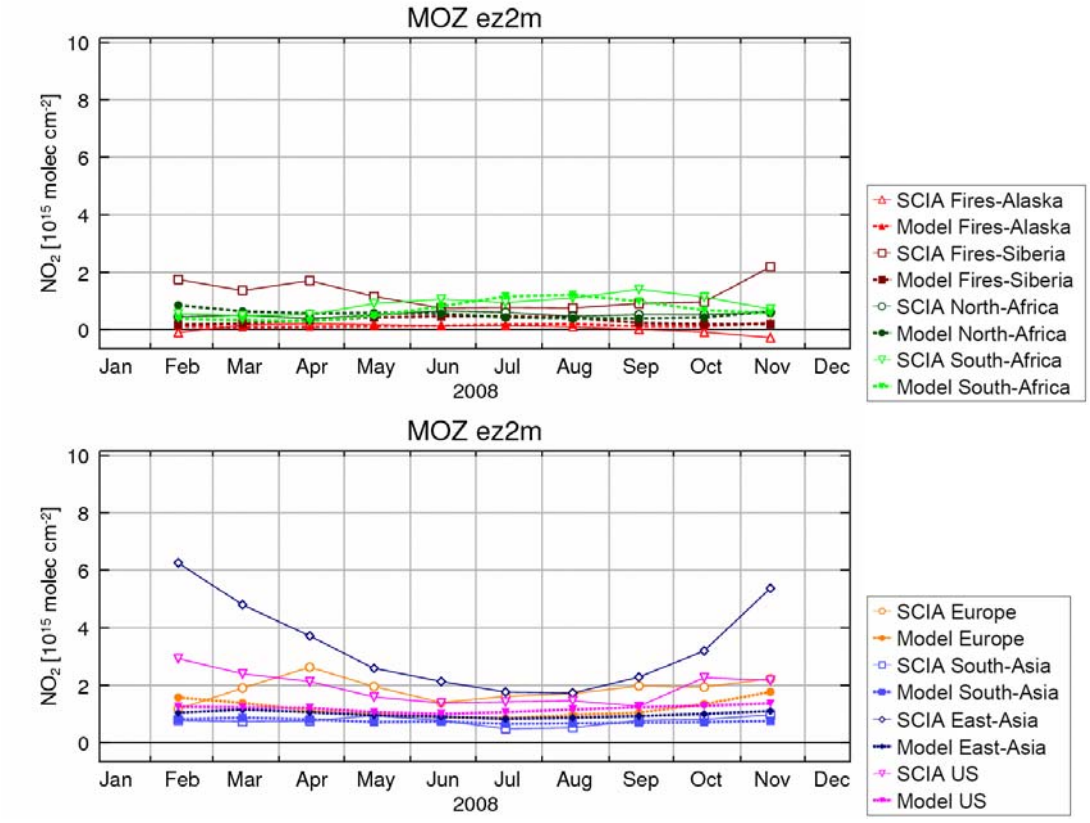


Figure 2.53: Seasonality curves for 2008 of tropospheric NO_2, in pre-defined regions divided into the typically biomass burning and polluted areas, determined by the MOZART *ez2m* (forecast run) and measured by SCIAMACHY (open symbols).

In general, this model version is able to forecast the right magnitude of values of tropospheric NO_2 in the selected regions and also captures the seasonality of biomass burning emissions. In contrast to the good agreement for biomass burning areas, over regions with high anthropogenic emissions of NO_2, the model underestimates the values and does not reproduce a seasonal cycle strong enough to capture the differences between winter and summer periods. The largest differences are observed over the East-Asia region, probably as a result of an underestimation of the rapid increase in anthropogenic emissions over the last decade in China.

Monthly Scatter plots

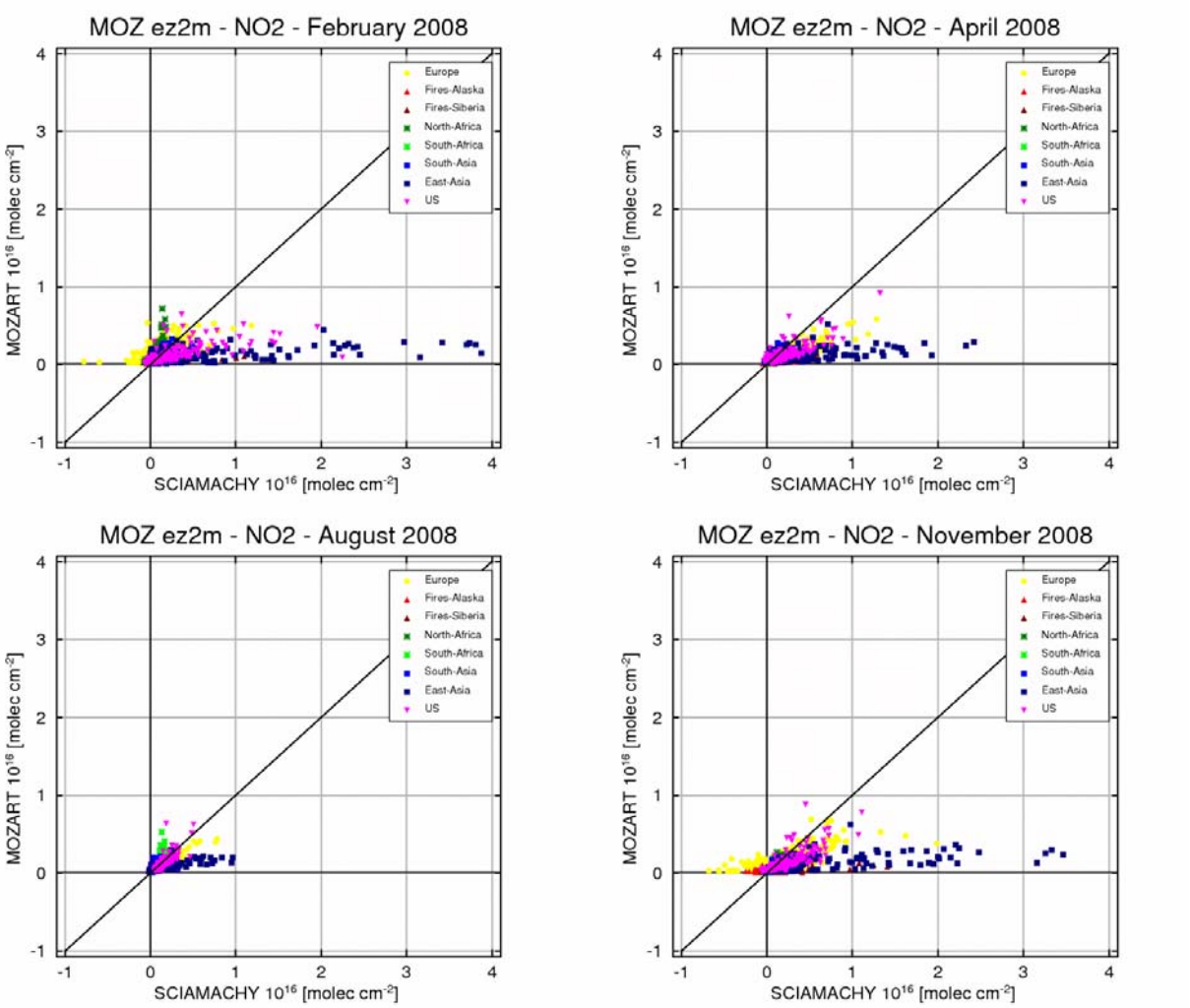


Figure 2.54: Scatter plots of tropospheric NO_2 for several months (February to August) in the year 2008 determined by MOZART-IFS run *ez2m* and measured by SCIAMACHY.

The NO_2 data calculated by MOZART-IFS for the year 2008 fits the satellite measurements better in the summer months (see for example the scatter plot for August in Fig. 2.54). In winter, the model underestimates the satellite data (see scatter plots for February and November above), in particular for large observed NO_2 concentrations. Overall, the scatter of the comparison is large, indicating problems of the model to reproduce the observed NO_2 fields.

Scores

In analogy to the other model versions, scores were calculated for the forecast data so that the evaluation process could be summarized into a few numbers telling us how good the model performs. Here, we present the averages over the selected regions, but the monthly scores for each of the regions can be found in figures presented in the Annex 8.

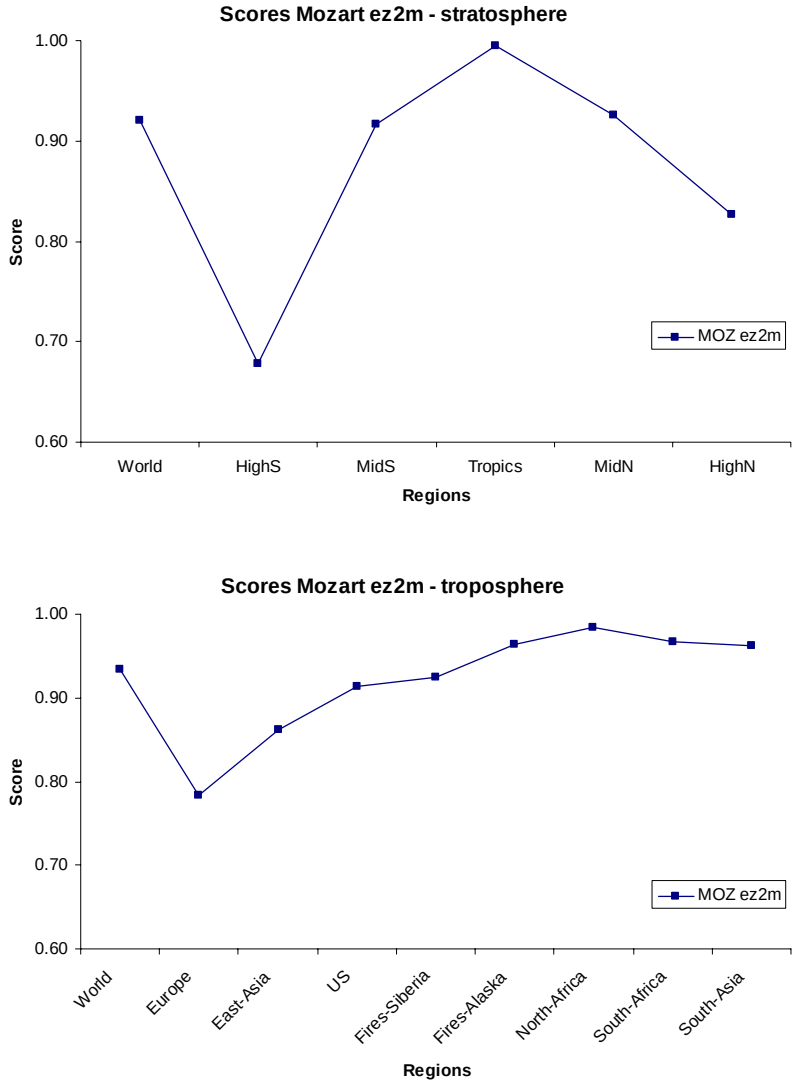


Figure 2.55: Annual average (period February to November 2008) scores obtained for the MOZART run *ez2m* for the different regions of the stratosphere (top) and troposphere (bottom).

From the calculated scores, we see that stratospheric NO_2 forecasted by MOZART for the year 2008 is better at the Tropics (0.99) and worst in the South Pole region ("HighS" – 0.68). This very low score in the South Pole is caused by the wrong NO_2 values in the first half of the year. As soon as the upper boundary conditions were improved, the stratospheric results looked much better and the scores actually range from 0.96 to 0.99 in the second half of the year. This was also verified for the remaining regions but the "HighS" is the one with more sticking increase of values. In the case of the tropospheric columns, the model is doing a fine job predicting the NO_2 emissions in "North-Africa", but needs further improvement over Europe and other polluted areas. Further details can be found in the next section where the model data are compared to measurements from another satellite instrument, OMI.

2.4.2 MOZART vs OMI

In this section, the tropospheric columns of MOZART-IFS run *ez2m* and TM5Y08 are evaluated based on a definition for the tropopause where O_3 exceeds 150 ppb, and tropospheric columns correspond to the total amount of NO_2 below this level. The evaluation period is from July to the end of 2008 and in this section we present only figures for a few months of data, given that these can be taken as an example for the conclusions achieved.

For this evaluation the model data considered corresponds to the overpass time of the satellite, 13.30h local time and, for the analysis, all (OMI/model) data has been regridded onto a common grid of approximately: 0.1 x 0.1 deg for the European evaluation, and 0.5 x 0.5 deg for the global evaluation and for the statistical evaluation over other regions at the globe. Based on the daily samples statistical evaluation has been performed. The monthly averaged tropospheric columns are evaluated, as well as their biases. In addition, the correlations in time over the complete evaluation series are analyzed.

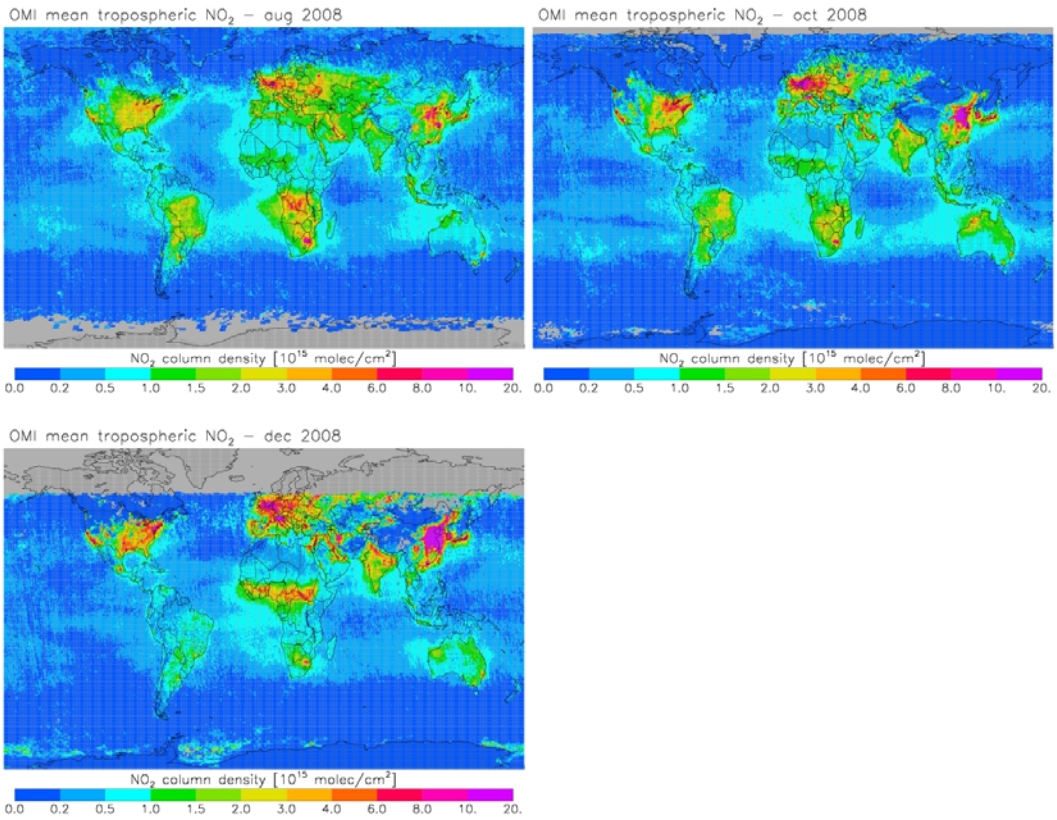


Figure 2.56: Global NO_2 tropospheric column from OMI, for the months of August, October and December 2008.

On average both TM5 and MOZART-IFS properly represent the main features in tropospheric NO_2 (see Figs. 2.56 and 2.57): high concentrations over the urban regions and the biomass-burning regions. Over the continents in July to October both models show a negative bias of tropospheric NO_2, except at higher latitudes on the Northern hemisphere. In those regions, both TM5 and MOZART-IFS exhibit zero / positive biases. In both models red spots in Siberia can be found. This is probably due to the applied inventory for biomass burning, which is a five-yearly average (2002-2007). This database contains events of bursts of biomass burning in Siberia, which took place in 2003. A similar error is causing the over-prediction for TM5 in August in Alaska. In December (Fig. 2.58), the MOZART-IFS and TM5 models actually over-predict tropospheric NO_2 in a larger part of the Northern hemisphere: the northern part of USA/Canada and Europe. Both models show a negative bias over the Chinese east-coast, for all months. The model(s) properly represent the large rise of tropospheric NO_2 in the winter months over the Northern-hemisphere urban regions. This rise can be attributed partly to lower photolysis rates and reduced losses of NO_2 through reaction with OH.

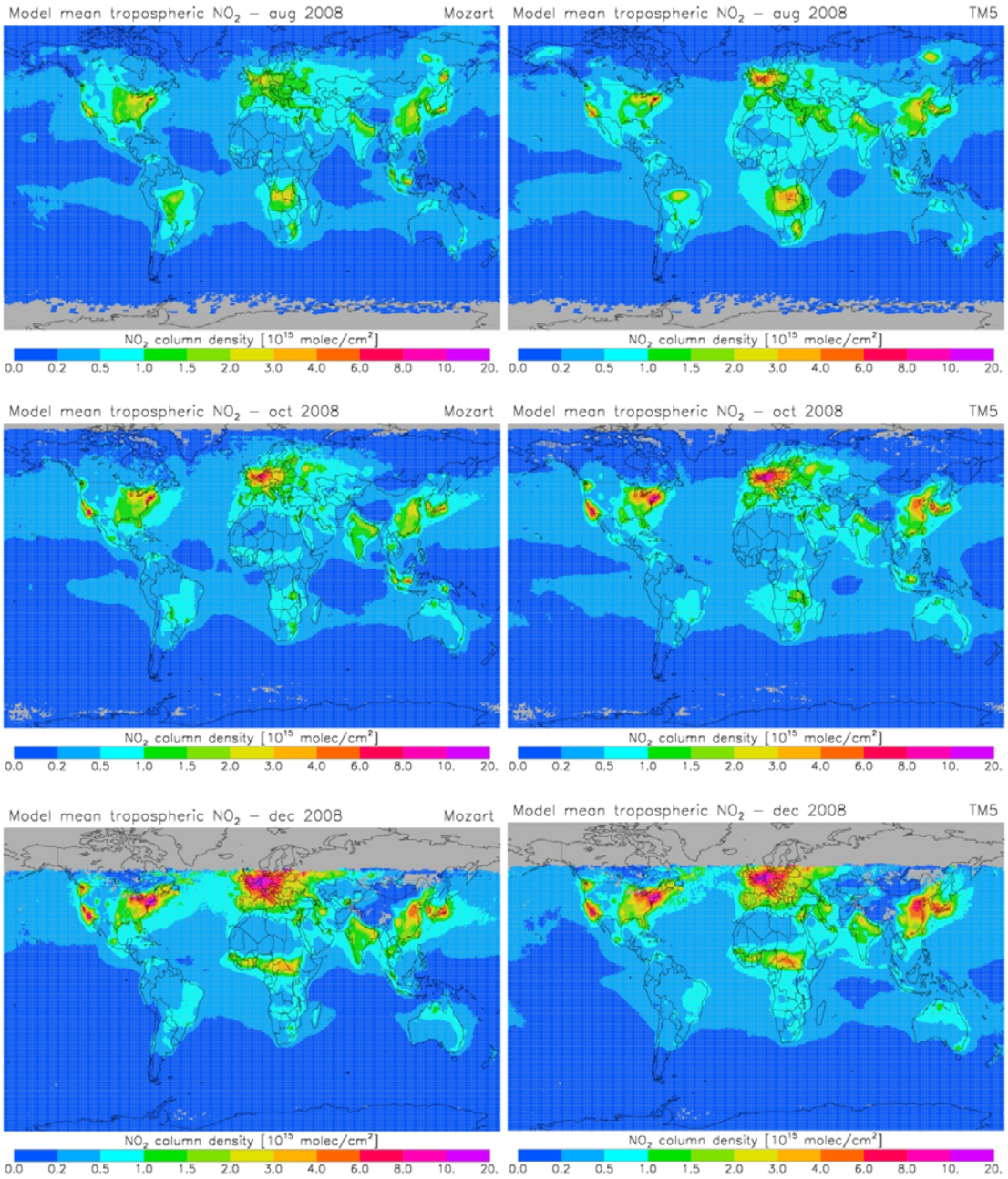


Figure 2.57: Global NO_2 tropospheric column for MOZART-IFS and TM5 for the months August (top), October (middle) and December (bottom) 2008.

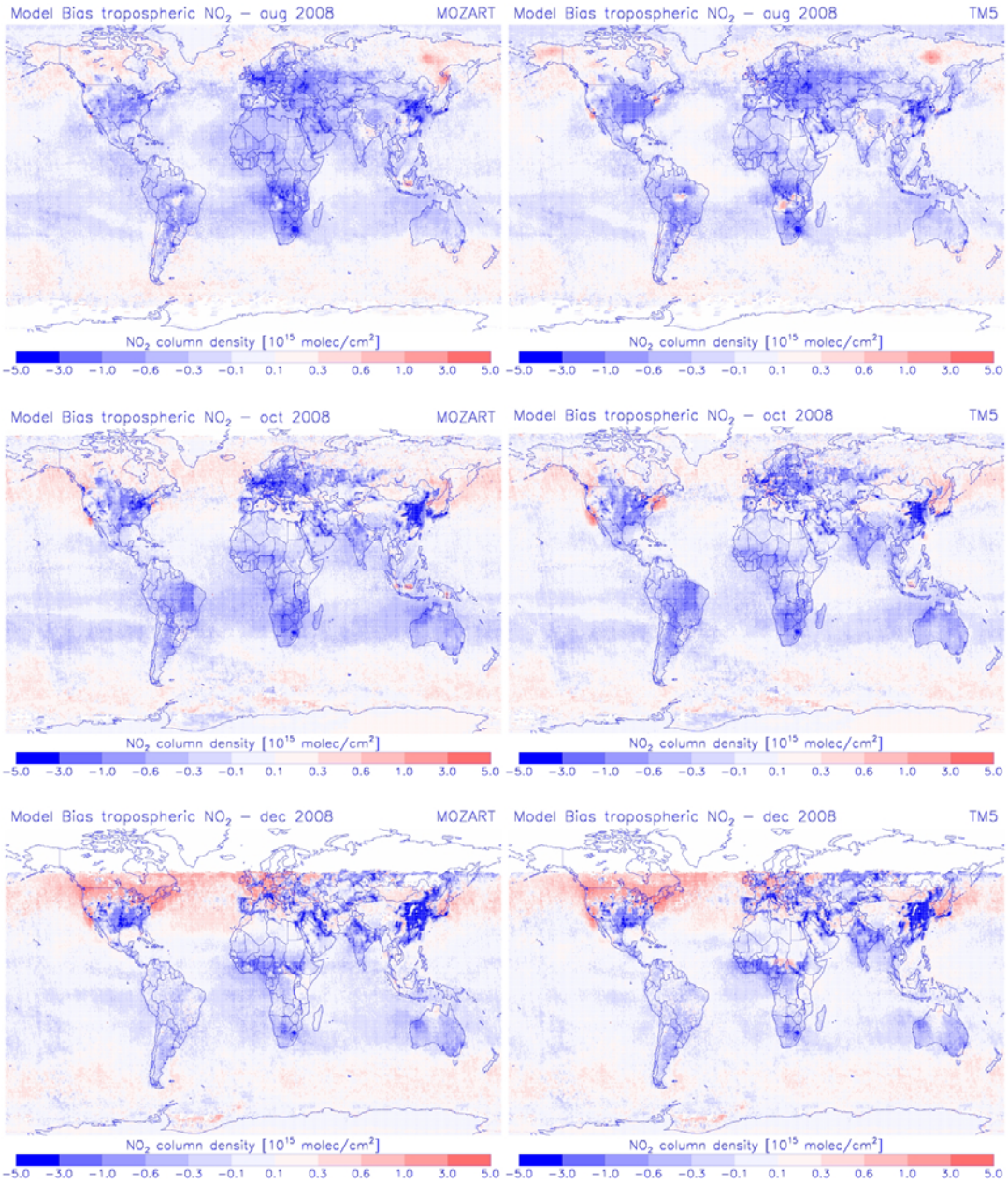


Figure 2.58: Bias in global NO_2 tropospheric column between MOZART-IFS / TM5 and OMI, August (top), October (middle) and December (bottom) 2008.

Scores

The resulting scores for the MOZART-IFS and TM5 model are given in Fig. 2.59. These values cannot be compared directly with those obtained from the comparison between the model and the SCIAMACHY observations because the comparison with OMI data was limited to land measurements as in the previous analysis.

Similar features as those presented in previous figures are found again, with lower scores over the European and USA regions and high scores away from the sources. The erroneous emissions in Siberia can be observed again. MOZART-IFS has a problem with the outflow from the African region in August (top panels in Fig. 2.59).

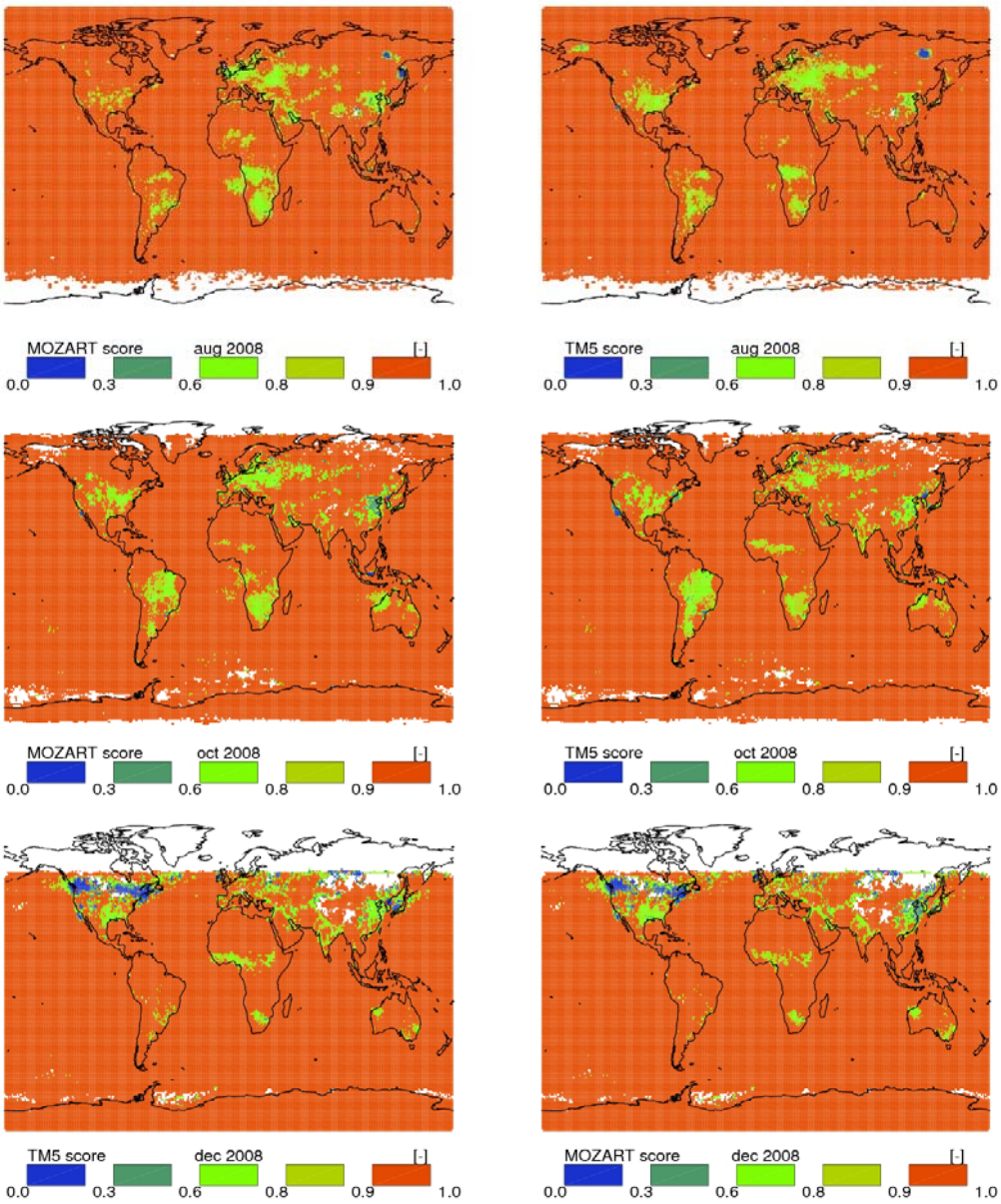


Figure 2.59: Scores for global NO_2 tropospheric columns for MOZART-IFS (left) and TM5 (right), for the months August (top), October (middle) and December (bottom) 2008.

The average scores over different regions on the globe are given in Fig. 2.60. The average score for the full globe for both TM5/MOZART-IFS is very high, as the background score is always quite high by definition. When looking at different regions, the score over the European RAQ domain is high in the summer/autumn, but decreases a bit towards the winter time. This is (partly) due to the fact that the region with available measurements is reduced towards the winter time, as no measurements are possible during winter months at high latitudes. Therefore less contribution from the (high-score) Northern region are not accounted for in the area-averaged score. The low score over the USA towards the winter months can be attributed to a high bias compared to OMI over the northern part of the USA and Canada. The Siberian fire region suffers from the same problems as the EU-RAQ region. The scores over East Asia, South Asia and Africa are relatively stable and similar for TM5 and MOZART-IFS. The scores over North Pole and South Pole regions are not very relevant here.

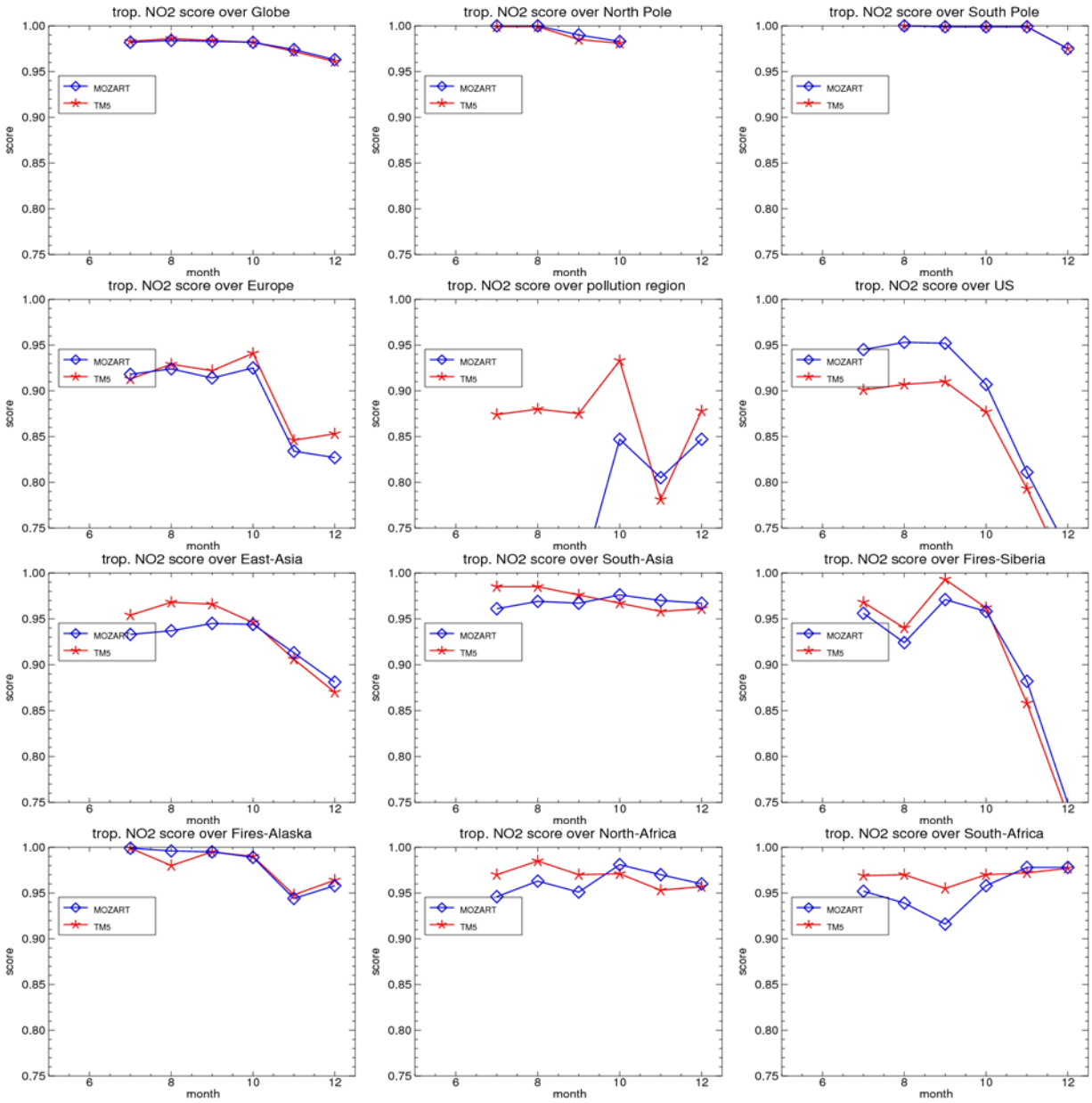


Figure 2.60: Evolution of scores over pre-defined regions for the period of July-December 2008.

2.5 Comparison between RAQ and GRG simulations

In this study, the tropospheric NO_2 columns for the second part of 2008 (July – December) derived from different RAQ models and two global models (the MOZART-IFS forecast experiment *ez2m* and TM5 reanalysis) are evaluated over the European RAQ domain, by comparing them to OMI NO_2 data. Both datasets are sampled at 13:30h local time (the overpass time of the satellite) according to the cloud criteria described in Section 2.1. For the models in forecast mode we use data from the first forecast day. Since we have half a year of data, the summer-autumn-winter model results can be evaluated. This analysis provides insight in the quality of the models in the different seasons, as well as their ability to model the seasonal variation.

RAQ models

The RAQ model domain ranges from 35/70 deg latitude and -15/35 deg longitude. The contributing GEMS-RAQ models for this evaluation are listed in Table 2.3. More detailed information can be found in the GEMS-RAQ reports. Here, only the most relevant information is provided. Note that the RAQ models themselves differ substantially in resolution, applied chemical scheme, underlying meteorology and boundary conditions. Also the height of the RAQ domains differs from model to model. In the current set of evaluations (as given in this report) the results of some of the models are based on earlier model versions. In the meantime model upgrades have been performed, as reported in the documents from the RAQ subproject. For instance, some of the models applied inappropriate choice of fixed boundary conditions and in the later model versions these have been replaced using the boundary conditions from the MOZART-IFS system. The emission inventory for the RAQ models is in principle equal for all models, and is based on the TNO 2000 inventory (Visschedijk and Denier van der Gon, 2005; Visschedijk *et al.*, 2007).

Table 2.3: Specifications of the contributing RAQ models.

Model name	**Resolution (deg)** **Levels, top level**	**Meteorology**	**Chemistry: based on:**
BOLCHEM	0.4 x 0.4, L-, - hPa	3hr ECMWF	CB-IV
CAC	0.2 x 0.2, L25, 250 hPa	HIRLAM /ECMWF	CB-IV
CAMx	0.3 x 0.3, L15, 300 hPa	MM5/ ECWMF	CB-IV
CHIMERE	0.5 x 0.5 L8, 500 hPa	3hr ECMWF	EMEP
EMEP	.25 x .25 L20, 100 hPa	3hr ECMWF	MSC-W
EURAD	.4 x .4 L23, 100 hPa	3hr ECMWF	RACM
MATCH	.2 x .2 L30, 400 hPa	6hr ECMWF	EMEP
SILAM	.2 x .2 L-, - hPa	3hr ECMWF	-

Within this analysis, also scores (as defined before and in Annex 7) are calculated and they were averaged over the full RAQ domain, as well as over a "pollution region" at 48/54 deg lat and -3/10 deg long, which is the area over the London region, the Benelux, the north of France and the west of Germany (see Fig. 2.61).

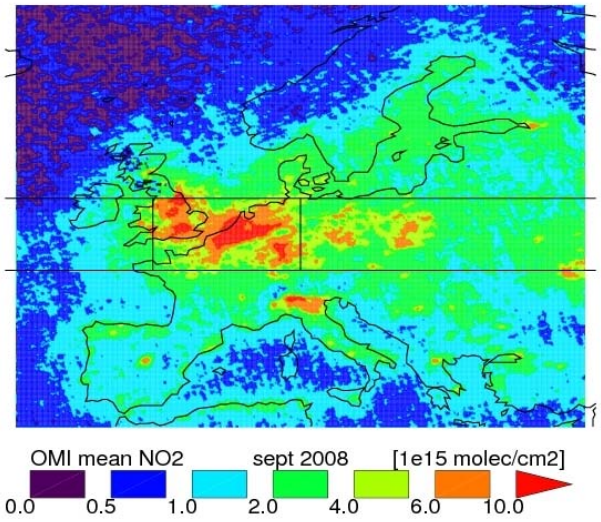


Figure 2.61: Tropospheric NO_2 from the OMI instrument in September, showing the 'pollution region'.

As given in Table 2.3, the height of the RAQ models changes from one model to the other. Therefore the tropospheric column from these models is also changing, having a different height. However, this contribution from the upper troposphere to the total column is marginal, as shown in Figs. 2.62 and 2.63 below. Figure 2.62 shows a monthly-averaged profile over the RAQ domain, as well as over the "pollution region" as defined in Fig. 2.61. This shows that the largest concentrations are found close to the surface, especially in the "pollution region". The partial contribution to the total column is shown in Fig. 2.63. On the average, for the RAQ region, about 90% of the total columns originates from the levels below 500 hPa. Over the "pollution region" this number is 95%.

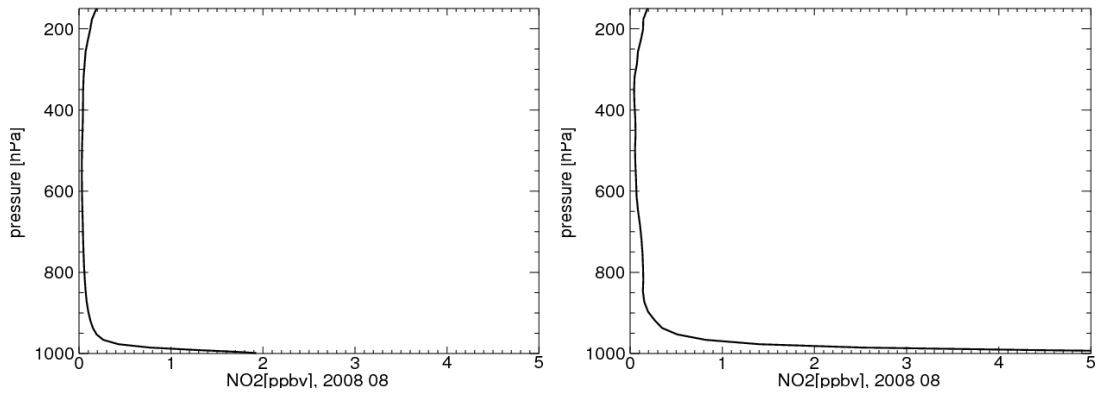


Figure 2.62: Area-averaged profile of NO_2 over the full European region (left) and region with high pollution (right), derived from MOZART-IFS data for August 2008.

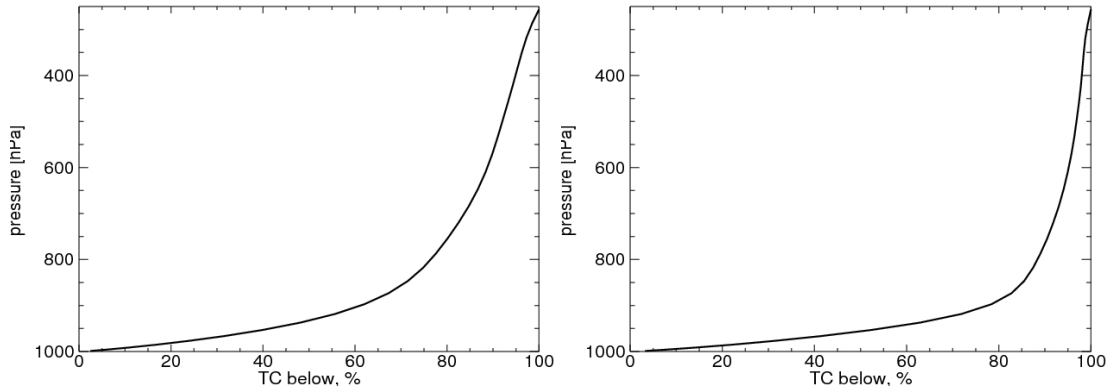


Figure 2.63: Fraction of NO_2 contribution from bottom to specific height, scaled to total column below 250 hPa, for August 2008, from MOZART-IFS for the full European region (left) and region with high pollution (right).

Figure 2.64 shows the average total columns over half a year sampling time for all models and OMI data. As compared to OMI, MOZART-IFS has a negative bias. Here the TM5 global model performs better. The RAQ models show a more fine-scaled spatial distribution in their tropospheric column data. In most of the models the urban regions can clearly be identified.

In Fig. 2.65 standard deviations in the observations are compared to the corresponding model results. The standard deviation has been evaluated for every single month, and filtered at the available retrievals. Afterwards, the monthly values for the standard deviation are averaged for all available months. In this way, we corrected for the seasonal variation. Moreover, only locations with more than four samples available in a month time are included. Therefore, the white regions in the figure correspond to locations that do not fulfill these requirements, and that have been removed from the analysis for this reason. This figure illustrates the dynamics of the models as compared to the observed data. The OMI observations show a background level of about 0.5×10^{15} molec/cm^2, which can be (partly) attributed to the measurement uncertainty. Over the 'pollution region' and northern Italy a higher variability is observed. As for the models, they capture a higher variability of about 1×10^{15} molec/cm^2 over the continent, but the highest variability over the Benelux is underestimated for most of the models. The large fluctuations in the north of Italy are not captured by all models.

In Fig. 2.66 shows the seasonal change of the average tropospheric NO_2 columns, for the southern part of the European domain (35-57 deg latitude, to prevent a changing evaluation domain in winter months) as well as for the pollution region. There is an increase in NO_2 in both regions that is captured by all models. However, there are substantial differences, where BOLCHEM has an overprediction and CHIMERE an underprediction of the seasonality. The diurnal cycle in tropospheric NO_2 columns over the pollution region and the full RAQ domain is illustrate in Figure 2.67. These data are scaled to the daily mean column. With respect to the average column over the full domain the models are well in line with each another. More deviation in the diurnal cycle can be observed for the pollution region. In this analysis the BOLCHEM model has a relatively high peak in the afternoon, which may be due to a coding-error in the emissions in this model version. MOZART-IFS and TM5 have a relatively flat profile over the day, while TM5 has relatively high column data at night, compared to the other models.

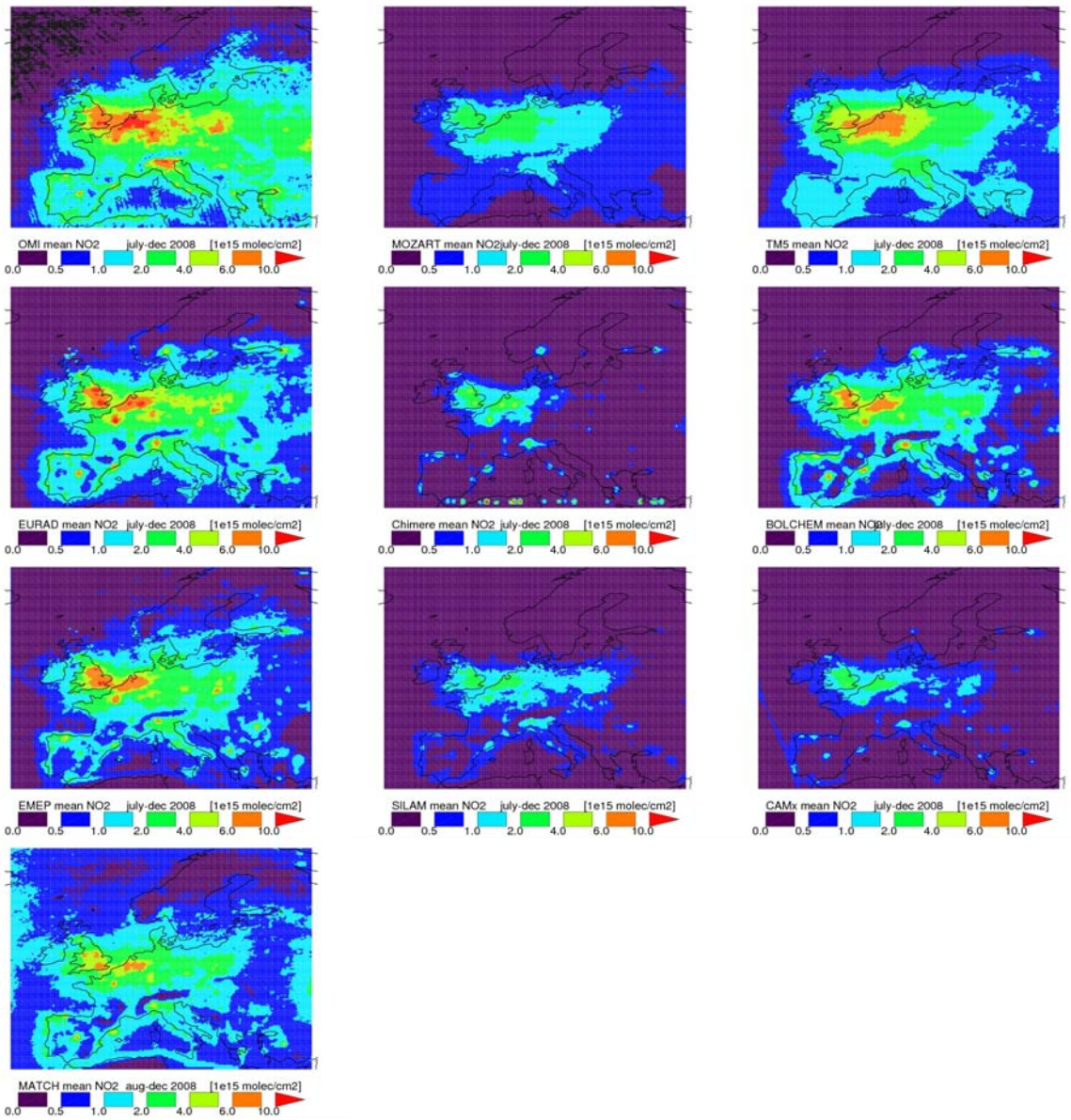


Figure 2.64: Half-year average NO_2 columns from Global and RAQ models over Europe and OMI average (top left corner).

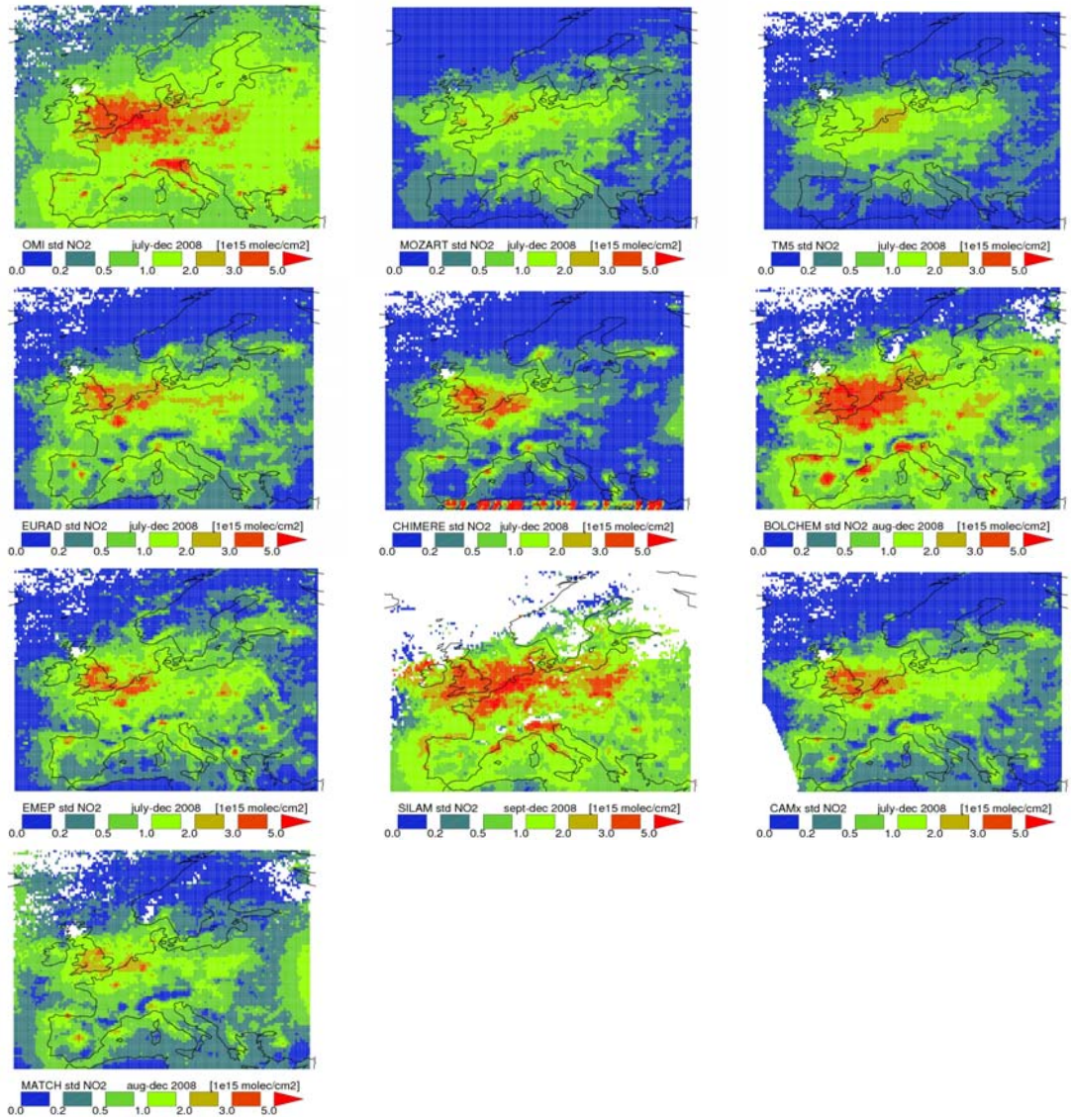


Figure 2.65: Standard deviation illustrating the dynamics in OMI NO_2 retrievals, and corresponding dynamics of global and RAQ models, July-December 2008.

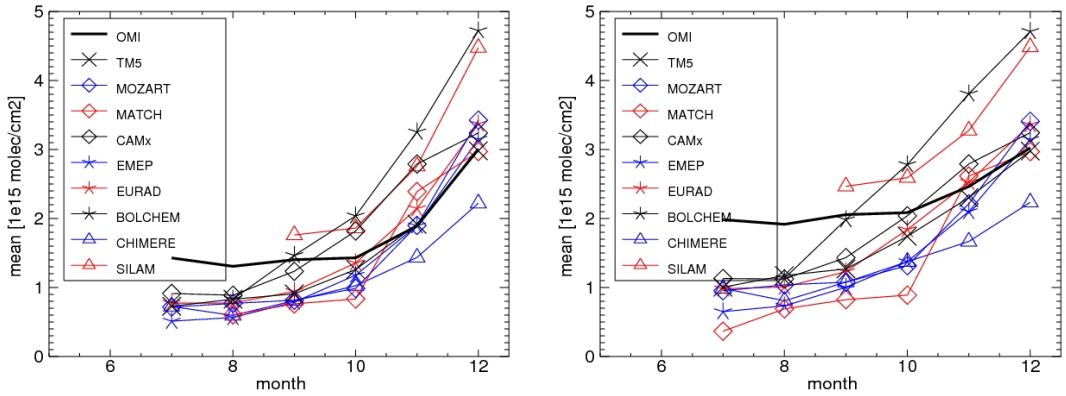


Figure 2.66: Mean column over (left) southern RAQ domain and (right) 'pollution region', as a function of month.

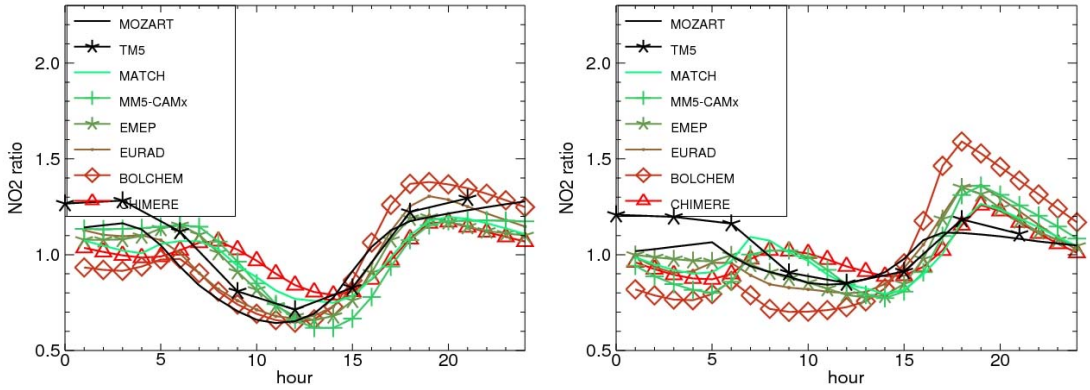


Figure 2.67: Diurnal cycle in September, full RAQ domain and 'pollution region'.

Scores

Fig. 2.68 shows the average score for the contributing models. One observation is that the global models MOZART-IFS and TM5 are reasonably well in line with other models. Generally a good score at the background regions (the Atlantic Ocean, Mediterranean region, and at higher latitudes) is obtained, whereas the score decreases over the regions with higher emission levels. The EURAD, CAMx, EMEP and SILAM models are still giving good scores. CHIMERE and MATCH show a relatively low score over the Eastern part of Europe. This can be attributed to an error in the emission treatment in the current MATCH version. The CAC model only started to deliver column data by November 2008, so there is not a sufficient amount of data available for a full evaluation. Fig. 2.69 shows the evolution of the scores as a function of month. In general, most models perform well and are stable over the entire period. The decrease in November-December for the average score over the full RAQ domain can be attributed to the changing evaluation region.

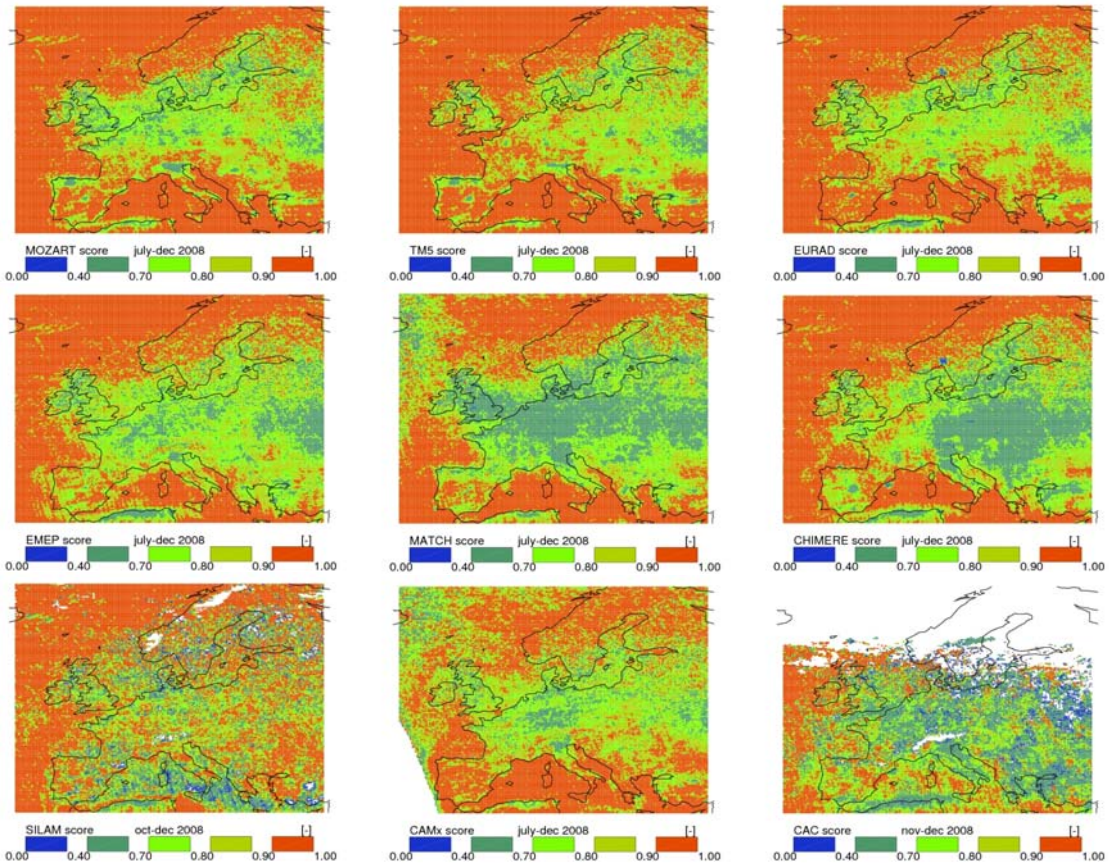


Figure 2.68: Average score for RAQ models, July-December (in some cases the averaging of the scores starts later).

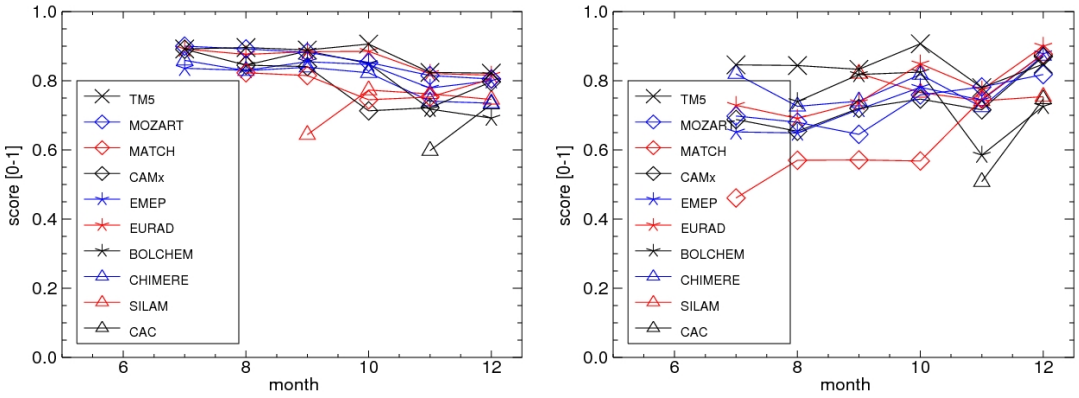


Figure 2.69: Mean score over full RAQ domain (left) and 'pollution region' (right), as function of the month.

2.6 CO evaluation

In this section, we present the evaluation of the CO simulations by comparing the model results (both standalone - MOCAGE V2, TM5 V7 and MOZART V10 - and reanalysis - MOZART *f026*) to the MOPITT CO measurements.

The MOPITT retrieval uses a fixed "global" a priori profile. This a priori profile is generated from a master set of 525 in-situ profiles measured from aircraft during eight atmospheric chemistry field campaigns and at two fixed sites over the past few years. In order to compare model outputs and satellite retrievals, all data are re-gridded to the same spatial resolution. In our case, MOCAGE, MOZART and MOPITT data (with 60 vertical levels) are interpolated to the TM5 spatial resolution (2° latitude x 3° longitude). MOPITT's averaging kernels are calculated for a given time frame over each area of interest. After this step, the model output fields are convoluted with average kernels of MOPITT and the results (new model CO profiles resolved at 7 MOPITT levels) are compared with MOPITT CO profiles. Table 2.4 gives the injection heights of different fractions of CO in each of the global CTMs model considered within the GRG sub-group.

Table 2.4: Fraction of CO fire emissions at injected heights by GRG-CTMs.

Injection height (m)	**MOZART**	**TM5**	**MOCAGE* 2003**	
	2003	**2003**	**forest**	**shrubs**
0-100	100%	20%	3%	40%
100-500	-	20%	3%	40%
500-1000	-	20%		
1000-2000	-	40%	6%	20%
2000-3000	-	-	13%	-
3000-4000	-	-	16%	-
4000-5000	-	-	19%	-
5000-6000	-	-	19%	-
6000-7000	-	-	13%	-
7000-8000	-	-	5%	-
8000-9000	-	-	2%	-
9000-10000	-	-	1%	-
10000-11000	-	-	-	-
Resolution	1.9°x 1.9°	3° x 2°	2° x 2°	

**: Information about injection height is only valid for the Alaska fire region for year 2003. For other regions, it may vary in numbers. For details: see Rast et al., submitted to JGR, November 2008.*

As it was already mentioned in Section 2.1, several regions were defined for the evaluation of the model results. In the tropospheric case, there are 8 regions (see Fig. 2.1): Europe, US and East Asia, corresponding to highly polluted regions; South Asia, which is relatively clean/less polluted; and Fires-Siberia, Fires-Alaska, North Africa, South Africa which have experienced extensive fire activities. In this chapter, we will present the analysis made over only six of these regions, as the results for evaluation of CO over Alaska and Siberian fire regions are presented in more details in Chapter 8. Furthermore, the evaluation presented here focuses only on the results for the year 2003. For a more direct and easy comparison, both standalone and reanalysis (MOZART *f026*) runs are shown in the same figures.

The following section deals with the comparison of modeled CO with MOPITT measurements at different levels. These results (monthly averages over a region) are presented for the following levels: surface, 850,

700, 500, 350, 250, and 150 hPa. The vertical bars give the standard deviation that, in fact, for this analysis, is calculated as spatial variance of the CO concentrations within the region being analyzed.

Europe

An intercomparison of MOPITT CO concentrations over Europe and those retrieved from GRG-CTMs after application of MOPITT averaging kernels is presented in Fig. 2.70. All CTMs show a consistent temporal pattern over Europe. Especially, MOCAGE results show a good agreement with MOPITT data for pressure levels lower than 500 hPa. TM5 V7 and MOZART V10 show an underestimation for most of the different height levels. MOZART reanalysis run *f026* slightly underestimates CO concentrations during the period February – April, 2003, and shows a fair agreement with MOPITT CO concentrations during the remaining months of 2003.

USA

For the USA region, MOCAGE V2 CO concentrations compared to MOPITT CO data (see Fig. 2.71), exhibit an overestimation during the whole year with larger differences in the lower troposphere. TM5 V7 and MOZART V10 show a consistent seasonal pattern with a slight underestimation throughout the year. Additionally, MOZART V10 exhibits larger variation in CO mixing ratios during the January, June and September months, particularly at 150, 250, 350 and 500 hPa levels. A similar behavior is reflected in monthly mean scores over the USA region displayed in Fig. 2.71 presented below. The MOCAGE V2 performed least good when compared to the other CTMs particularly during December 2003 (with a 40% deviation).

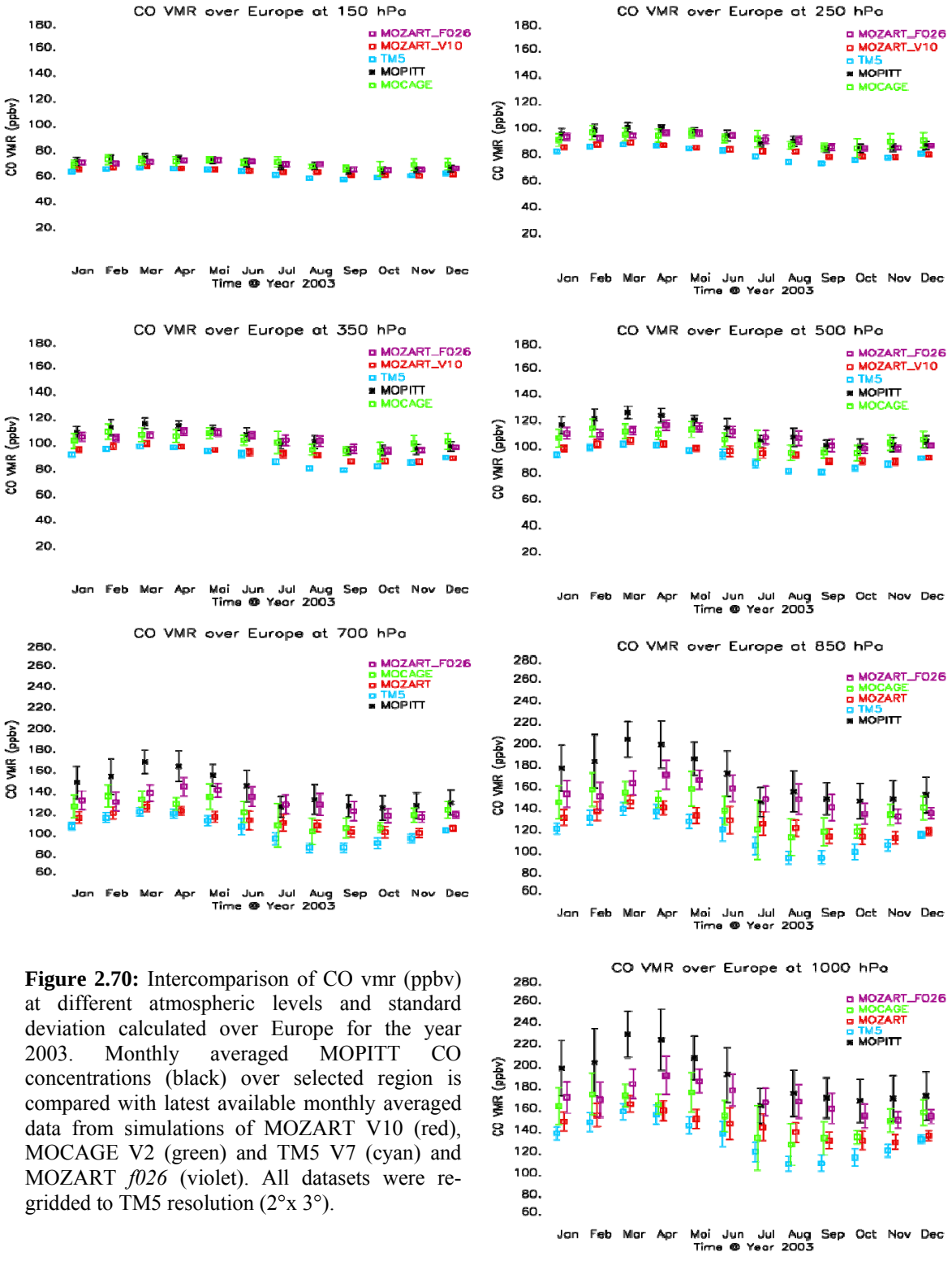


Figure 2.70: Intercomparison of CO vmr (ppbv) at different atmospheric levels and standard deviation calculated over Europe for the year 2003. Monthly averaged MOPITT CO concentrations (black) over selected region is compared with latest available monthly averaged data from simulations of MOZART V10 (red), MOCAGE V2 (green) and TM5 V7 (cyan) and MOZART *f026* (violet). All datasets were re-gridded to TM5 resolution (2°x 3°).

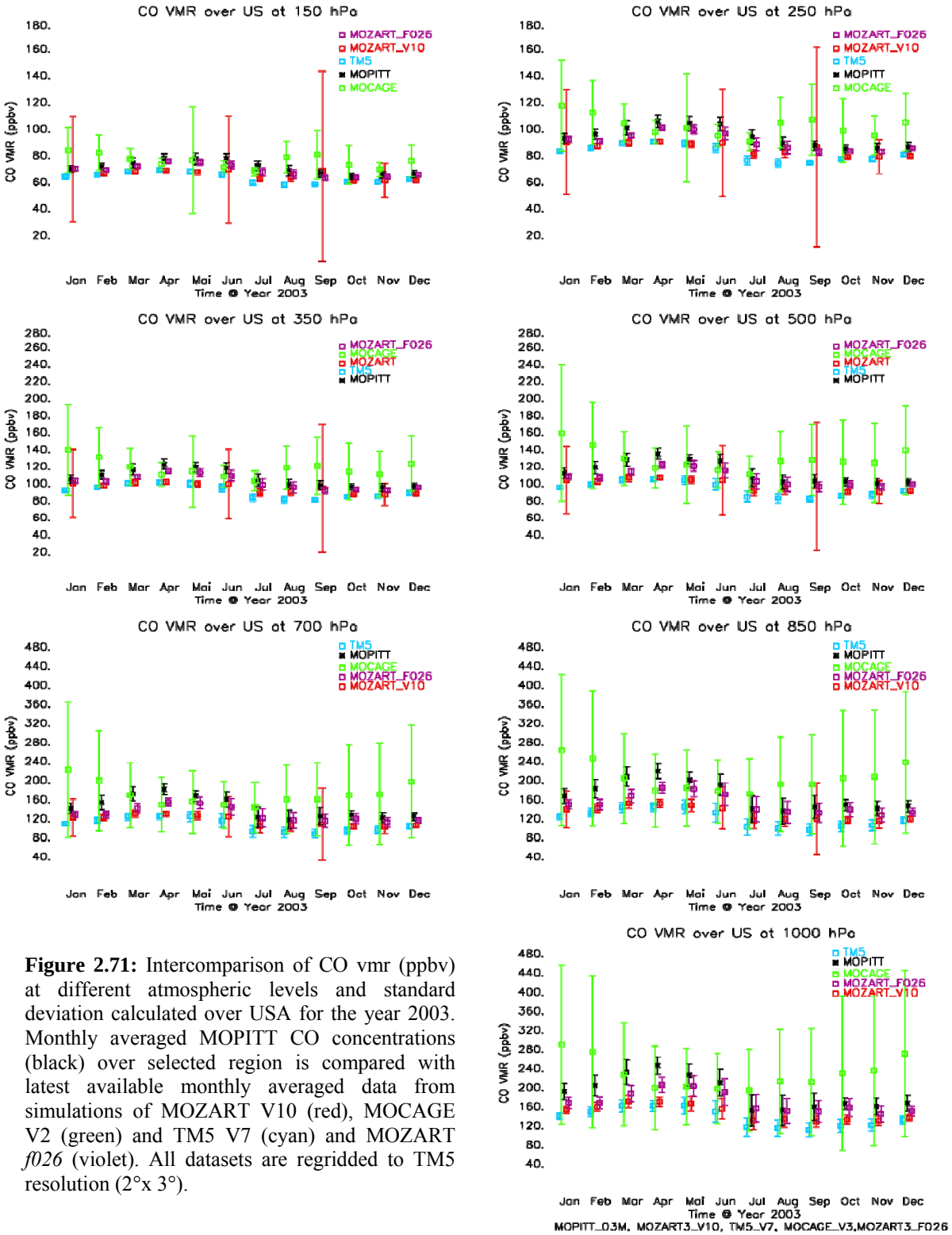


Figure 2.71: Intercomparison of CO vmr (ppbv) at different atmospheric levels and standard deviation calculated over USA for the year 2003. Monthly averaged MOPITT CO concentrations (black) over selected region is compared with latest available monthly averaged data from simulations of MOZART V10 (red), MOCAGE V2 (green) and TM5 V7 (cyan) and MOZART *f026* (violet). All datasets are regridded to TM5 resolution (2°x 3°).

East Asia

Over East Asia, all CTMs reasonably catch the seasonality in CO emissions. These results are presented in Fig. 2.72 below. In general, MOZART standalone V10 and reanalysis *f026* show a fair agreement with MOPITT CO data. MOCAGE and TM5 data are lower than MOPITT CO observations, with a relatively larger dispersion in the MOCAGE data during January at upper tropospheric levels.

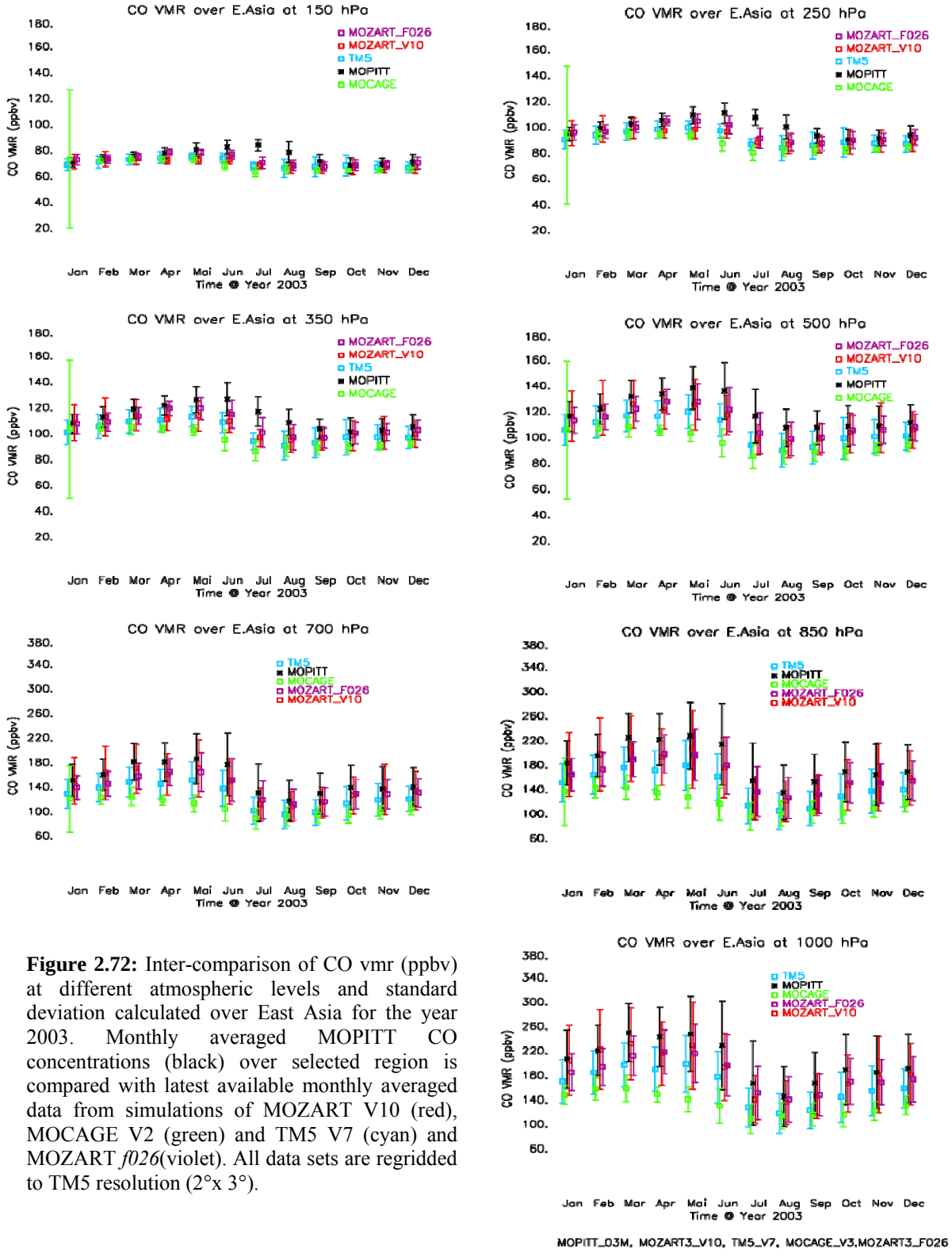


Figure 2.72: Inter-comparison of CO vmr (ppbv) at different atmospheric levels and standard deviation calculated over East Asia for the year 2003. Monthly averaged MOPITT CO concentrations (black) over selected region is compared with latest available monthly averaged data from simulations of MOZART V10 (red), MOCAGE V2 (green) and TM5 V7 (cyan) and MOZART *f026*(violet). All data sets are regridded to TM5 resolution (2°x 3°).

South Asia

When looking at the results from all GRG-CTM data compared to MOPITT, CO observations in Fig. 2.73 show a consistent temporal pattern over South Asia. MOCAGE V2 and TM5 V7 slightly underestimate CO concentrations over south Asia. The latest standalone run for MOZART, version V10, and reanalysis run *f026* are in a good agreement with MOPITT data, for the year 2003, with slightly overestimated concentrations at lower levels during winter months.

North Africa

The CO analysis presented in Fig. 2.74 corresponds to the evaluation over North Africa region. All GRG-CTM CO simulations, except MOCAGE V2, show consistent temporal patterns in this area, when compared to MOPITT CO observations. MOCAGE V2 underestimates CO concentrations over the year. This could be due to the fact that MOCAGE V2 did not use the same anthropogenic emission inventories as used by the other GRG-CTM and/or different injection heights for fire CO emissions, as it happened in the comparison over Alaska fire region during summer 2004 (see Chapter 8 for further details). MOZART V10 simulation results are in a good agreement with MOPITT data. However, both data sets show slightly larger variability at lower levels during winter months.

South Africa

The South Africa region is a region where distributions are mainly controlled by fire activities and results of the evaluation are presented in Fig. 2.75. All GRG-CTMs, except MOCAGE V2, show a consistent temporal pattern over South Africa, when compared to MOPITT observations. MOZART *f026* results are in good agreement with MOPITT data. However, MOZART V10 and TM5 V7 results show slightly underestimated CO concentration during three months (September to November) of the year 2003. MOCAGE V2 data exhibits lower CO concentrations for all tropospheric levels considered and does not catch the temporal patterns in the CO concentrations over South Africa.

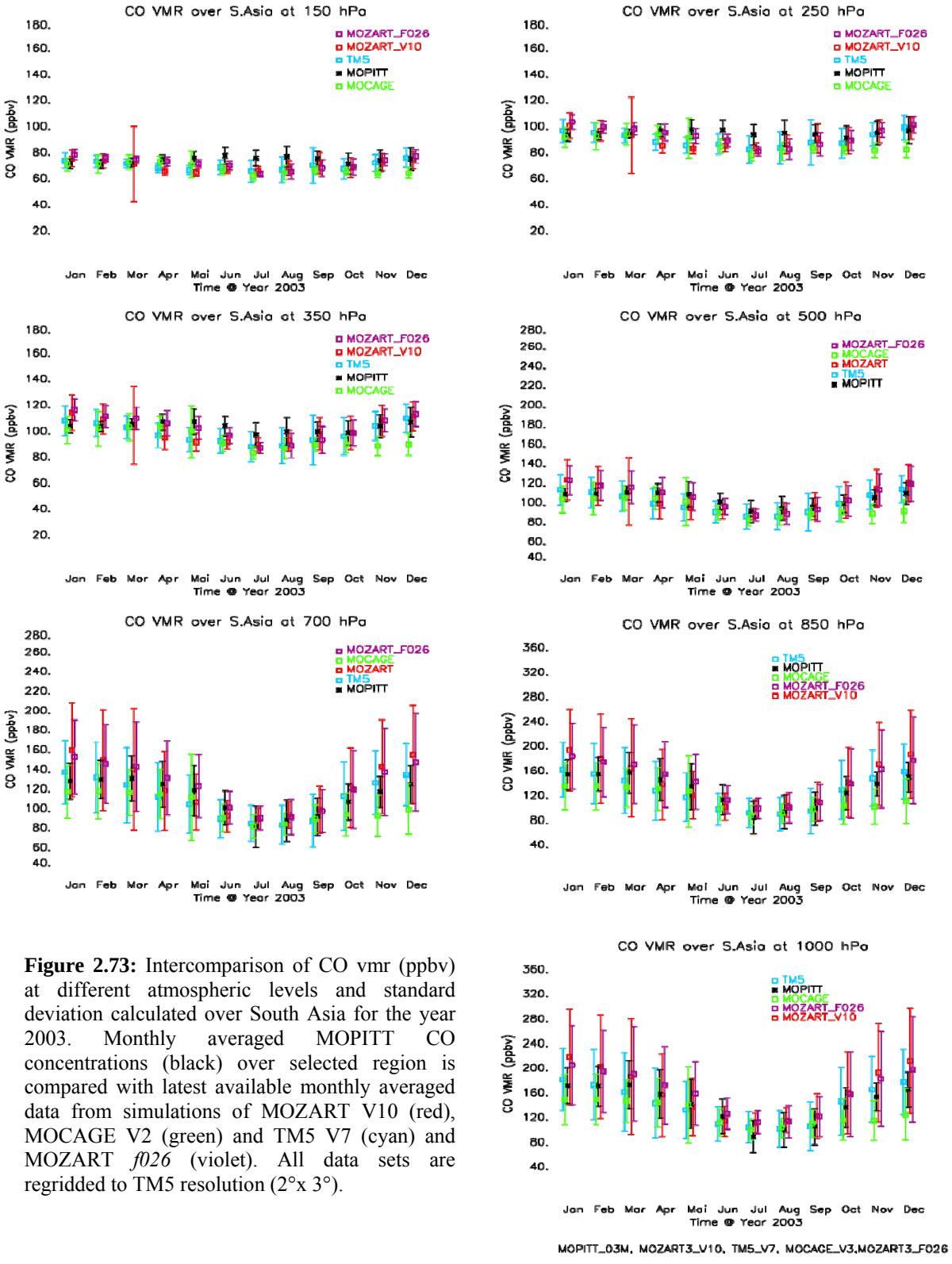


Figure 2.73: Intercomparison of CO vmr (ppbv) at different atmospheric levels and standard deviation calculated over South Asia for the year 2003. Monthly averaged MOPITT CO concentrations (black) over selected region is compared with latest available monthly averaged data from simulations of MOZART V10 (red), MOCAGE V2 (green) and TM5 V7 (cyan) and MOZART *f026* (violet). All data sets are regridded to TM5 resolution (2°x 3°).

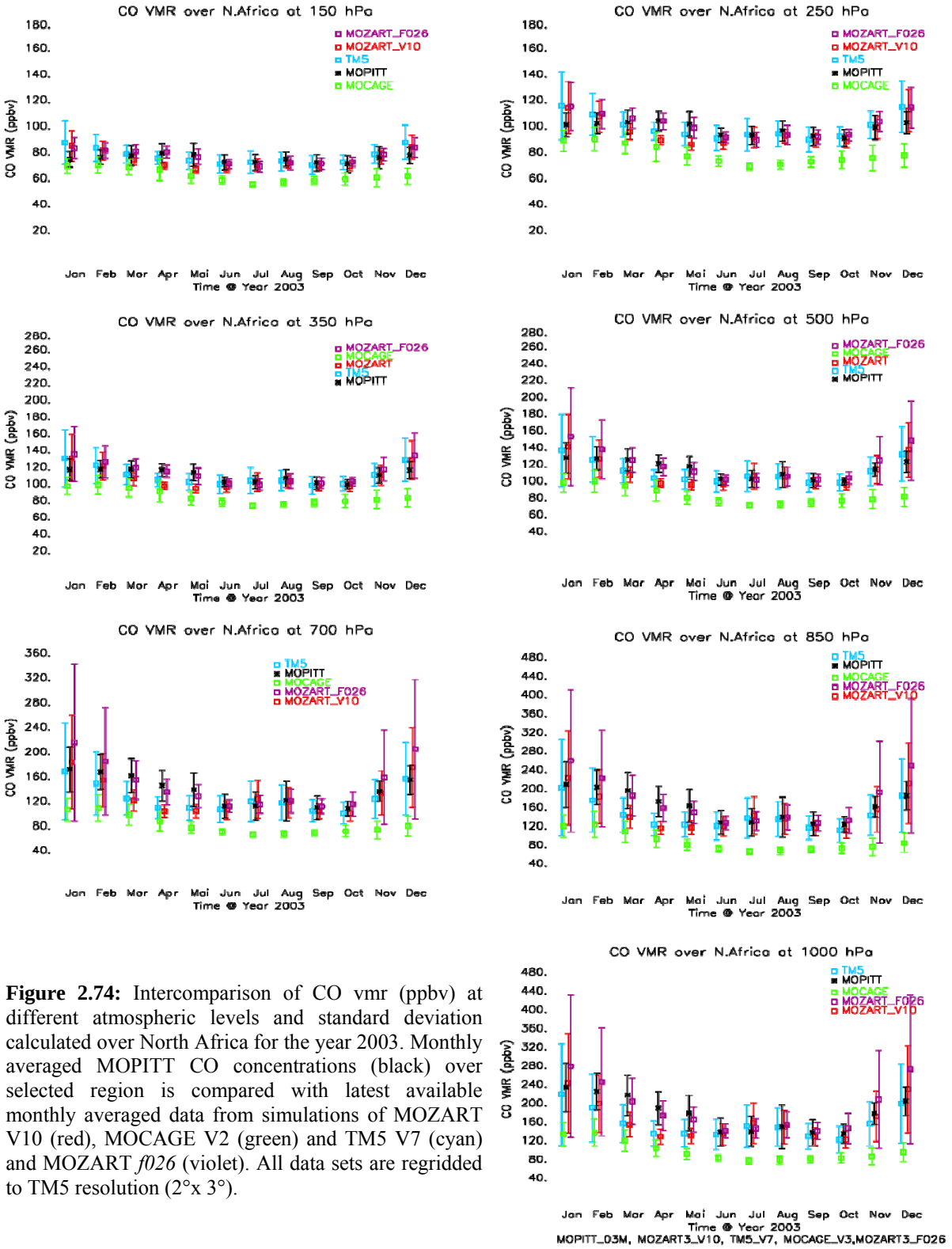


Figure 2.74: Intercomparison of CO vmr (ppbv) at different atmospheric levels and standard deviation calculated over North Africa for the year 2003. Monthly averaged MOPITT CO concentrations (black) over selected region is compared with latest available monthly averaged data from simulations of MOZART V10 (red), MOCAGE V2 (green) and TM5 V7 (cyan) and MOZART *f026* (violet). All data sets are regridded to TM5 resolution (2°x 3°).

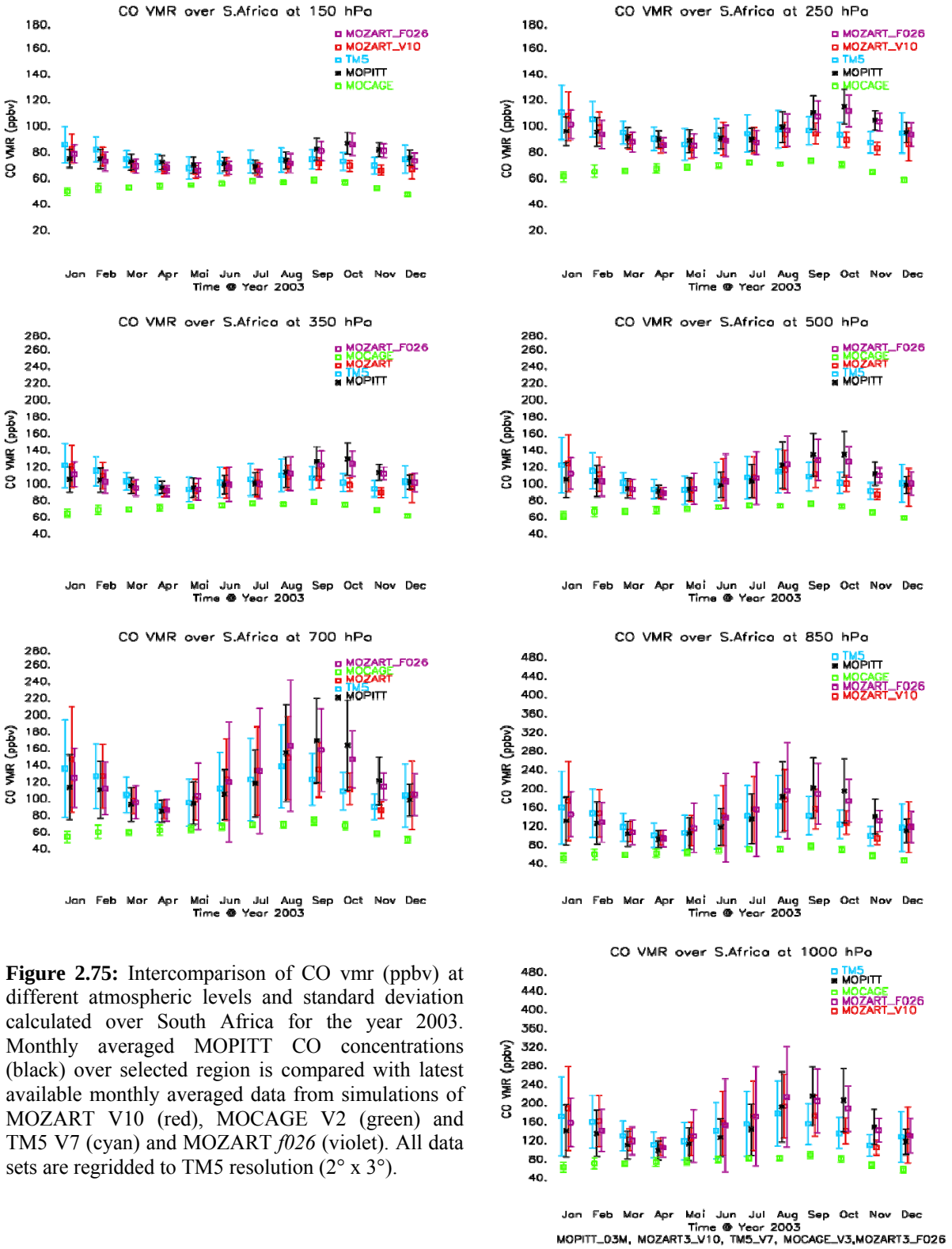


Figure 2.75: Intercomparison of CO vmr (ppbv) at different atmospheric levels and standard deviation calculated over South Africa for the year 2003. Monthly averaged MOPITT CO concentrations (black) over selected region is compared with latest available monthly averaged data from simulations of MOZART V10 (red), MOCAGE V2 (green) and TM5 V7 (cyan) and MOZART *f026* (violet). All data sets are regridded to TM5 resolution (2° x 3°).

Scores

Fig. 2.76 displays the GRG-CTM performances based on the monthly scores calculated for the different regions considered for the evaluation of CO simulations.

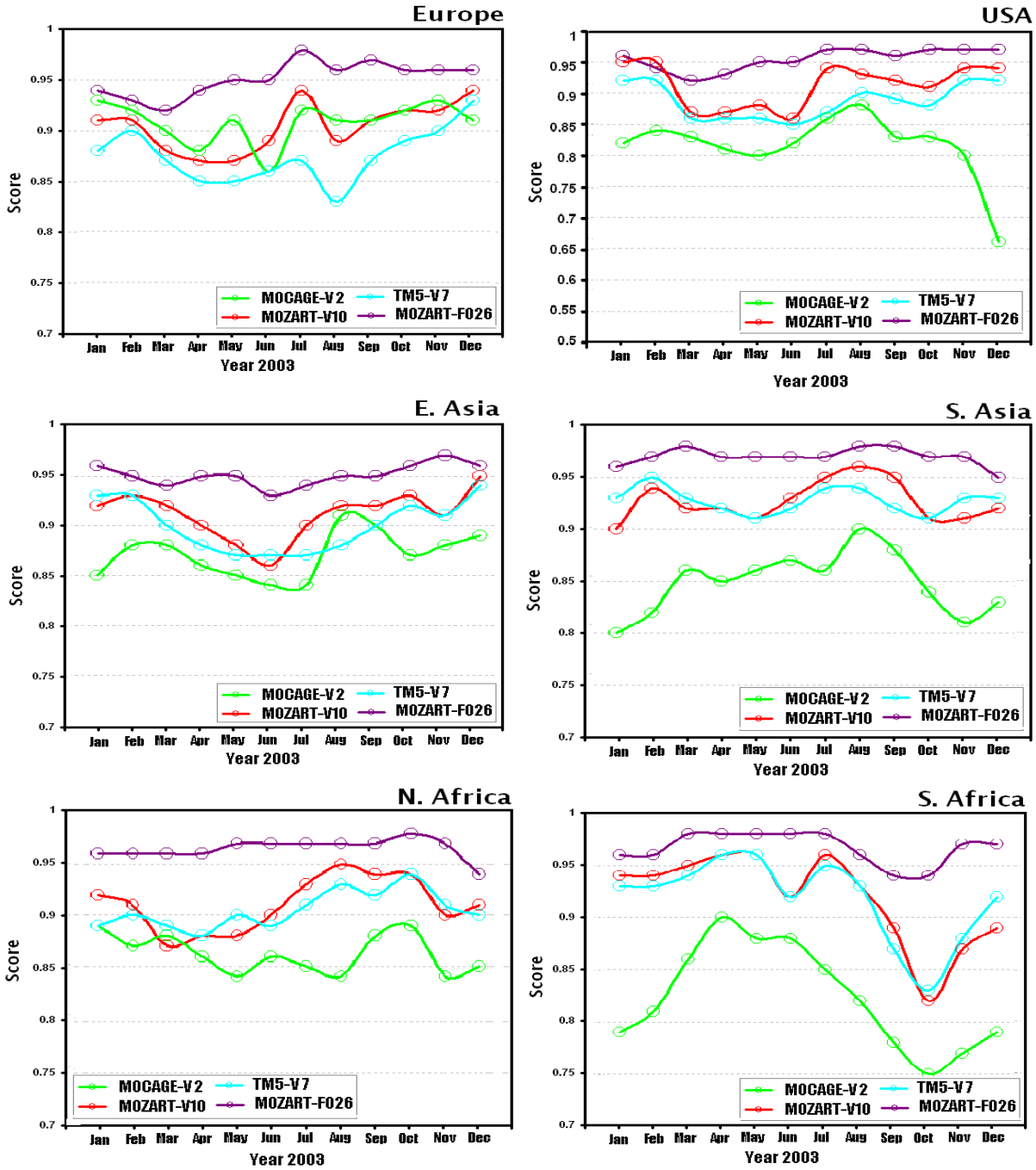


Figure 2.76: Monthly mean score over the different regions considered for the evaluation of tropospheric model results. Different colors represent the GRG-CTM.

Scores over Europe show that MOZART *f026* performs very well when compared to the other CTMs. The CTM performance is at its best during July 2003, except for TM5 V7 which shows a higher value during December 2003.

Regarding the USA, MOZART reanalysis run *f026* has the highest scores, while TM5 V7 and MOZART V10 show an almost similar performance with minimum values from March to June 2003. The poor performance of MOCAGE V2 over the USA region could be due to the use of different emission inventories as compared to MOZART and TM5 (GRG emissions inventory).

In general, the scores of all CTM for the East Asian region are more than 80%, with even higher values for MOZART reanalysis run *f026*, i.e. more than 92% performance throughout the year 2003. However, the performances of all CTM are lower during June 2003.

The GRG-CTM performances over the South Asian region exhibit almost similar patterns with maximum performance during August 2003. However, MOCAGE V2 scores are lower, TM5 V7 and MOZART V10 give moderate levels and MOZART *f026* displays the largest scores.

MOCAGE V2 shows lower scores with respect to performance over North African region.

For South Africa region the monthly scores presented in Fig. 2.76 reveals a similar behavior as those calculated over the South Asian region. MOCAGE V2 shows a lower sensitivity to CO emissions resulting from savannah fires. MOZART *f026* displays the highest GRG-CTM performances over the South African region.

The performance of all GRG-CTM based on the yearly mean scoring index is given in Table 2.5 below. As mentioned before, these values are calculated from the medians of daily CO total columns from the GRG-CTM simulations and MOPITT observations over each region.

Table 2.5: Yearly mean score over different regions for the year 2003

GRG-CTMs	**Europe**	**E. Asia**	**S. Asia**	**USA**	**N. Africa**	**S. Africa**
MOZART V10	0.904	0.911	0.926	0.913	0.910	0.919
MOCAGE V2	0.908	0.870	0.848	0.815	0.862	0.823
TM5 V7	0.875	0.90	0.927	0.887	0.905	0.918
MOZART *f026*	0.951	0.950	0.97	0.955	0.965	0.966

On the basis of Table 2.5, MOZART second reanalysis run (*f026*) obtains the best scoring for almost all regions, followed by MOZART V10, TM5 V7 and MOCAGE V2, respectively.

2.7 Summary

NO_2 standalone

For the standalone runs for the year 2003 and 2004 different conclusions can be drawn for the three models analyzed: MOCAGE, MOZART and TM5.

From the evaluation of MOCAGE data some problems were identified. The seasonality of the NO_2 concentrations in both the stratosphere and the troposphere is well reproduced in qualitative terms. However, in general, the stratospheric columns are too low. Surprisingly, MOCAGE V01 was able to simulate the NO_2 values more correctly than the succeeding version V02 – V01 has the time averaged scores between 0.87 (South Pole) to 0.95 (Tropics) and V02 from 0.70 to 0.80. This is only the case for the high levels in the atmosphere. When looking at the results for tropospheric NO_2, both model versions perform quite badly with severe overestimation of values in the winter months. The highest time averaged score obtained in the troposphere is 0.78 for V01 and 0.79 for V02, in both African regions. But if one looks into the monthly scores zero values are found for regions such as Europe in the winter months.

Stratospheric NO_2 for both years is well modeled by the latest version of MOZART V10 that obtains scores close to 1 for the stratospheric fields. The initial version V1 had clearly wrong results and as a result of the frequent evaluation improvements that were carried through very good results were obtained already for version V7. In terms of the troposphere, the NO_2 values above polluted regions are quite well estimated with some low bias in the winter months. However, over some biomass burning regions, like Siberia and Alaska, tropospheric NO_2 is still too high compared to satellite measurements. The NO_2 data calculated by all MOZART runs fit better to the satellite measurements in the summer months. Some high dispersion can be found especially in the winter months, but the overall scenario is quite good. Looking at the scores for the tropospheric regions one would conclude that the model version that has achieved the best results in modeling the NO_2 in this layer was actually the early V1 for 2003 and V8 for the year of 2004. However, this is not completely true as, in fact, we saw from the global maps and seasonality plots that, in general, the tropospheric NO_2 simulated by the initial version was very low and V10 shows better agreement with the satellite data. The high values of V8 are a result of the analysis of this version being limited only to the summer months when the model normally performs better and therefore we can say that for the year 2004 the scores are very similar to all the versions analysed.

For TM5 V7, in the year 2003, some stratospheric results, mostly in the high latitudes during winter, are higher than the NO_2 measured by the satellite. However, in the year 2004, with version V9 and the latest V10, the results of the model and satellite present a better match, with just some slight overestimation in the South Pole region where the lowest score values are achieved (time averaged score is 0.72 in 2003 and 0.4 for version V9 in 2004). For the tropospheric values, the case is quite different. In general, TM5 underestimates the NO_2 fields in the lower atmosphere, but over biomass burning sources the model results are closer to the measurements (in V7 the values were slightly lower and in V10 NO_2 columns are to some extent higher). Nevertheless, the calculated scores vary between 0.79 and 0.99 in 2003 and from 0.91 to 0.99 in 2004 with V9. But it is important to notice that, with version V10, the NO_2 in the European region obtains a score of 0.69.

In any case, the seasonality of NO_2 is in general well captured by both MOZART V10 and TM5 V10, excluding the area over China (in the East-Asia region) where, for example, winter columns are, just like for 2008, strongly underestimated. Also over Europe we can see that MOZART V10 overestimates the NO2 values in the winter months.

Comparing the three CTMs, the TM5 data are less scattered relative to the satellite measurements, but this is partly related to the coarser grid of the model (2° x 3° instead of the 2° x 2° of MOCAGE and 1.8° x 1.8° of MOZART).

NO_2 MOZART reanalysis – *f026*

The dataset analyzed for this model version consisted of 4 years, more precisely from May 2003 to September 2006. In the overall scenario one can see a very good agreement between model and satellite measurements in terms of stratospheric columns. For this layer, it is only relevant to notice the minor overestimation of NO_2 in the tropics.

However, for the tropospheric columns, the situation is quite different. The model results present a better match to the satellite data in the summer but the overall performance is not good over the regions characterized by high anthropogenic emissions. Here the model tends to underestimate the NO_2 column. The seasonal cycle, as it was observed for the standalone versions, is not strong enough and therefore does not capture the high NO_2 values over East-Asia, for example. Also, the biomass burning regions in central and southern Africa present too low values. On the other hand, the high emissions modeled for the fires over the Siberia region in 2003 are not measured by SCIAMACHY. From year to year the model results are very similar and this is explained by the constant anthropogenic emission field. Therefore, the evolution in anthropogenic emissions, e.g. the slight increase of NO2 over polluted areas as Europe and East-Asia, are not captured by the MOZART-IFS system in the period of 2003 to 2006. The better agreement for the biomass burning emissions is a result of the yearly update of these emissions (monthly values) in the model runs.

The time averaged scores obtained for the tropospheric regions vary from 0.69 (Europe in 2004) to 0.99 (North Africa for every year), with the lower values normally calculated for the polluted regions. We can see that, for the year 2003, the values for the run *f026* are similar to those of version V1 which is the basis for the run of MOZART-IFS system, with very small improvement only in the US and fire regions.

HCHO MOZART reanalysis – *f026*

The HCHO fields modeled by the reanalysis run of MOZART *f026* are in general very good results with good simulation of both the seasonal patterns and also the absolute tropospheric columns. We have seen that the model was able to reproduce the seasonal cycle of HCHO over most of the regions selected, both those used for the NO_2 evaluation and those selected especially for this trace gas. Europe is once again a problematic region with model results constantly higher than the satellite measurements which, as explained above, could in part be related to biases in the measurements, evident from the negative columns in winter. Discrepancies between simulated and measured columns are also found above South-Africa, normally in the second half of the years analyzed. The annual averaged scores for tropospheric HCHO calculated by MOZART *f026* run range from 0.88 (2003 and 2005) to 0.93 (2006). The best monthly values are achieved in the North-Africa region and the minimum scores in Europe.

NO_2 MOZART-IFS forecasts – *ez2m*

The data of the coupled system MOZART-IFS that forecasts NO_2 in the year 2008 was compared with both SCIAMACHY and OMI measurements. For these runs the model mostly underestimates the NO_2 emitted over polluted regions like Europe and China and also over biomass burning regions, in particular in Africa. Also, from the global model intercomparison for this period, i.e., the evaluation of MOZART-IFS and TM5 Y08, it was possible to observe that the models are well in line with one another. Both models overestimate tropospheric NO_2 in the USA. This might be related to outdated emissions inventory - emissions reductions have taken place in the USA since 2000.

The global stratospheric NO_2 fields were initially overestimated by the MOZART-IFS forecasts but in the last months of the year 2008 the results look much better. Still, the variation throughout the year was well forecasted. While the seasonal cycle of biomass burning emissions is well reproduced, over China no differences between winter and summer months are modeled in contrast to observations. This might be explained, once again, by a lack of update in the emission fields that, in this case, do not reflect the rapid increase in anthropogenic emissions over the last decade and the change in emission patterns. Just as for the run *f026*, the scores for this version are very similar to the standalone MOZART V1, revealing that the coupled system has not improved much the initial NO_2 simulated by MOZART.

Concerning the annually averaged scores we can see that in the troposphere the lower value is obtained for the South Pole region where in general all models obtain worst results, with exception of the MOZART V8 and the reanalysis run. For the tropospheric regions analyzed, the main conclusion is that Europe is the most problematic area and in Fig. 2.60, one can see that this is most probably related to the emissions over the "pollution region" that includes Benelux. From this figure, we also see that normally the TM5 Y08 results for the year 2008 are in close agreement with the MOZART results except for regions like US where MOZART performs slightly better or East-Asia, and South Africa, EU-RAQ domain where TM5 achieves higher scores.

In addition, these results for the summer of 2008 were also compared with data from some Regional Air Quality (RAQ) models included within the GEMS framework. This intercomparison exercise showed that the global models are well in line with the RAQ models, with respect to the average tropospheric column, its change from summer to winter time, and the area-averaged diurnal cycle. This gives confidence in the use of the boundary conditions from the global models for the regional model. The global MOZART-IFS forecast run gives a slightly larger negative bias over the "pollution region" with high concentrations of NO_2 compared to TM5 and most of the RAQ models. The regional models perform in most cases better than the global models over the regions with high pollution levels and with larger spatial differences.

CO evaluation – standalone and reanalysis

The overall conclusion achieved after comparing the CO model results with MOPITT measurements is that the CTMs are in general able to capture the right seasonal trend of CO values. The results are better over polluted areas, where TM5 V7 and MOZART V10 slightly underestimate CO, and over the fire regions considered here (South and North Africa) mostly MOCAGE tends to simulate lower CO emissions. The reanalysis run of MOZART *f026* is comparable to the MOZART standalone version evaluated. However, when looking at the scores calculated (Fig. 2.76), this version always obtains the higher values. On the other hand, for MOCAGE we often find the lower values, performing better mostly over Europe. The scores of standalone TM5 V7 and MOZART V10 are very similar.

2.8 Conclusions and recommendations

In general, from the evaluation performed by comparing the model results with the satellite data, a few common aspects were found for all the models (considering the latest versions of the standalone runs):

- the stratospheric NO_2 is well modeled with exception of MOCAGE that presents too low values for this layer;
- the NO_2 emissions are not correctly simulated over the polluted regions mostly during the winter months with on average an underestimation of values;
- tropospheric NO_2 values over biomass burning regions are often too high, but the seasonality is well simulated.

The lack of seasonality and trend for the anthropogenic emissions might be one of the most problematic issues in the simulations analyzed. This can especially be noticed in the results for the most recent years (for the reanalysis and forecast runs of MOZART-IFS system) where the differences between model results and satellite measurements are very high over polluted regions, in particular over China. The model results have not followed the general change of NOx emissions in the last years in the major urban centers.

The HCHO fields modeled by the reanalysis run of MOZART-IFS *f026* in general agree very well with the measurements representing both the seasonal patterns and also the absolute tropospheric columns within the accuracy of the satellite retrievals.

The global models are well in line with the RAQ models, with respect to the average tropospheric columns, their seasonality and the area-averaged diurnal cycle. This supports the use of boundary conditions from the global models for the regional models. Still, it is important to keep in mind that the performance over Europe is not the best for any of the models analyzed.

The CO calculated by all models (standalone and reanalysis version) is very similar to the MOPITT measurements, with better results over polluted areas.

2.9 References

1. Boersma, K.F., D.J. Jacob, H.J. Eskes, R.W. Pinder, J. Wang and R.J. van der A (2008), Intercomparison of SCIAMACHY and OMI tropospheric NO_2 columns: Observing the diurnal evolution of chemistry and emissions from space, *J. Geoph. Res*. 113, D16S26
2. Boersma, K.F., H.J. Eskes, J.P. Veefkind, E.J. Brinksma, R.J. van der A, M. Sneep, G.H.J. van den Oord, P.F. Levelt, P. Stammes, J.F. Gleason, and E.J. Bucsela (2007), Near-real time retrieval of tropospheric NO_2 from OMI, *Atmos. Chem. Phys.*, 7, 2103-2118.
3. Deeter, M.N., L.K. Emmons, G.L. Francis, D.P. Edwards, J.C. Gille, J.X. Warner, B. Khattatov, D. Ziskin, J.-F. Lamarque, S.-P. Ho, V. Yudin, J.-L. Attié, D. Packman, J. Chen, D. Mao, and J.R. Drummond (2003), Operational carbon monoxide retrieval algorithm and selected results for the MOPITT instrument, *J. Geophys. Res.*, 108, D14, 4399, doi:10.1029/2002JD003186.
4. Kinne, S (2008), Scoring in AER, GEMS report.
5. Krol, M.C., S. Houweling, B. Bregman, M. van den Broek, A.J. Segers, P. van Velthoven, W. Peters, F. Dentener, P. Bergamaschi (2005), The two-way nested global chemistry-transport zoom model TM5: algorithm and applications, *Atmos. Chem. Phys.*, 5, 417-432.
6. van Noije, T.P.C. *et al*. (2006): Multi-model ensemble simulations of tropospheric NO_2 compared with GOME retrievals for the year 2000, *Atmos. Chem. Phys.*, 6, 2943-2979.
7. Richter, A., and J.P. Burrows (2002), Tropospheric NO_2 from GOME measurements, *Advances in Space Research*, 29(11), 1673-1683.
8. Rast et al., submitted to *J. Geophys. Res.*, Nov. 2008.
9. Richter, A., J.P. Burrows, H. Nüß, C. Granier and U. Niemeier (2005), Increase in tropospheric nitrogen dioxide over China observed from space, *Nature*, 437(7055), 129-132
10. Rozanov, A., V. Rozanov, M. Buchwitz, A. Kokhanovsky, and J.P. Burrows; SCIATRAN 2.0 - A new radiative transfer model for geophysical applications in the 175 - 2400 nm spectral region; Adv. Space Res.; 36(5); 1015-1019; doi:10.1016/j.asr.2005.03.012, 2005
11. Row, L.W., Hastings, D.A. and Dunbar, P.K. (1994): Terrainbase worldwide digital terrain data. Release 1.0 National Geophysical Data Center, Boulder
12. Tanskanen, A., A. Arola and J. Kujanpää (2003): Use of the moving time-window technique to determine surface albedo from the TOMS reflectivity data, *Proc. SPIE 4896*, pp. 239-250.
13. Visschedijk A.J.H., P.Y.J. Zandveld, H.A.C. Denier van der Gon (2007), A High Resolution Gridded European Emission Database for the EU Integrated Project GEMS, TNO-report 2007-A-R0233/B, TNO, Apeldoorn.
14. Visschedijk, A.J.H. and H.A.C. Denier van der Gon (2005), Gridded European anthropogenic emission data for NOx, SO_2, NMVOC, NH_3, CO, PM_{10}, $PM_{2.5}$ and CH_4 for the year 2000, TNO B&O-A Repport 2005/106, 2nd version November, TNO, Apeldoorn.
15. Wittrock, F. (2006), The retrieval of oxygenated volatile organic compounds by remote sensing techniques, 192 pp, University of Bremen, Bremen.
16. Wittrock, F, A. Richter, H. Oetjen, J.P. Burrows, M. Kanakidou, S. Myriokefalitakis, R. Volkamer, S. Beirle, U. Platt, T. Wagner (2006), Simultaneous global observations of glyoxal and formaldehyde from space, *Geophys. Res. Lett.*, 33, L16804, doi:10.1029/2006GL026310.

3. Evaluation with ozone soundings from global networks

Contributors Lucia Kins (DWD, Hohenpeissenberg, Germany), and the modeller teams

3.1 Datasets and methodologies

3.1.1 Datasets

The vertical ozone distribution of the MOZART_V1 and TM5_V3 stand-alone runs, the coupled IFS+MOZART runs exoz and ez2m, the IFS+TM5 coupled forecast runs eybl and ez3h and the IFS+MOZART coupled reanalyses runs eyih,eyq6 and f026 was evaluated by comparisons with ozone profile measurements from balloon sondes.

Applicable ozone sonde data can be found in various data bases: NDACC (Network for the Detection of Atmospheric Composition Change), WOUDC (World Ozone and Ultraviolet Radiation Data Centre), NILU (Norwegian Institute for air research) and SHADOZ (Southern Hemisphere Additional Ozonsondes). However, input into these data bases is quite irregular which largely reduces the availability of ozone sonde profiles for NRT comparisons. For that reason the stations used for the evaluation of the GEMS NRT forecast (ez2m) is somewhat different from the stations used for the other model runs. Table 3.1 gives an overview of the stations used for this NRT evaluation, their geographical location, the measurement interval and near real time availability. As can be seen, only a few stations in the northern hemisphere deliver their sonde data actually in near real time i.e. within one or two days after measurement. Table 3.2 presents the ozone sonde stations used for the stand-alone, the IFS+TM5, the reanalyses and the IFS+MOZART exoz comparisons. The global distribution of the ozone sonde stations is uneven, most of them are located in the mid latitude and polar regions of the northern hemisphere and only a few stations are found in the tropics and the southern hemisphere.

With exception of the Japanese stations and Hohenpeissenberg, electrochemical concentration cell (ECC) ozonesondes are launched. These sondes have a precision of better than 3-5% and an accuracy of ± 5-10% depending on the ECC manufacturer type and the sensing solutions used in the sondes. At Hohenpeissenberg, Brewer-Mast sondes are used. These sondes have a precision of 3-5% in the stratosphere and of ~10% in the troposphere. In comparison studies, the Japanese KC96 sondes underestimated the ozone values by about 5-10% at air pressures higher than 30hPa, at lower pressures their agreement was within 5% (T. Deshler et al., 2008, H.G.J. Smit et al, 2007).

A sonde profile ranges from the surface up to approx. 10 hPa, but due to instrumental reasons the lowermost measurements are not reliable. The presentation of our comparison results includes all levels, however the results near the ground are neglected in the discussion.

Table 3.1: Ozone sonde stations used for IFS+MOZART ez2m evaluation, their measurement interval, and NRT availability (d=day, w=week, m=month).

Station	***Lon***	***Lat***	***measurement interval***	***note***	***average time delay for data retrieval***
Alert (Canada)	-62.3	82.4	1/w	Jan-March only	2 m
Eureka (Canada)	-86.5	80	2/w	Jan-March only	2 w
Ny Alesund (Svalbard)	12	79	2/w		1-2 d

Thule (Greenland)	-68.8	76.5	1/w	Jan-March only	2 w
Scoresbysund (Greenland)	-22	70.5	1/w		1-2 d
Sodankyla (Finland)	27	67	1/w		1 w
Edmonton (Canada)	-114	53.5	1/w	Jan-March only	2 w
Goose Bay (Canada)	-60.5	53.2	1/w	Jan-March only	2 w
Legionowo (Poland)	20.9	52.4	1/w		1-2 d
Valentia (Ireland)	-10.3	51.9	1/w		4 m
Uccle (Belgium)	4	51	3/w		1-2 d
Prag (Czech Republic)	14.4	50	3/w		1-2 d
Hohenpeissenberg (Germany)	11	48	2/w summer 3/week winter		1 d
Yarmouth (Canada)	-66.1	43.8	1/w	Jan-March only	2 w
Sapporo (Japan)	141	43	1/w		3 m
Madrid (Spain)	-3.8	40.5	1/w		1m
Wallops Isl. (USA)	-75	38	1/w		2 m
Naha (Japan)	127.7	26.2	1/w		3 m
Nairobi (Kenia)	37	-1	1/w		4 m
Java	113	-8	2/m		3 m
Ascension Isl. (St. Helena)	-14	-8	1/w		4 m
American Samoa	-171	-14	1/w		6 m
La Reunion	55.45	-21.1	1/w		1 m
Lauder (New Zealand)	170	-45	1-2/w		6 m
Macquarie Island (Australia)	158.9	-54.5	1/w		3 m
Neumayer (Antarctica)	-8	-71	1/w 3-4/w Sep-Nov		1 m

Table3.2: Ozone sonde stations and number of ozone profiles used for evaluation of model runs: MOZART_V1 and TM5_V3stand-alone, coupled IFS+MOZART exoz, IFS+MOZART eyih, eyq6, f026 and IFS+TM5 eybl and ez3h.

Station	***Lon***	***Lat***	***Alt [m]***	***no. of O3 profiles Jul-Aug 2003***	***no. of O3 profiles May-Dec 2003***	***no. of O3 profiles Jan-Dec 2003***
Ny Alesund (Svalbard)	12	79	12	10	35	86

Sodankyla (Finland)	27	67	179	8	32	77
Uccle (Belgium)	4	51	100	21	92	135
Hohenpeissenberg (Germany)	11	48	976	18	77	129
Wallops Isl. (USA)	-75	38	13	12	38	60
Izana (Spain)	-16	28	51	9	36	55
Hilo (USA)	-155	20	11	9	33	47
Paramaribo (Suriname)	-55	6	25	10	30	47
Nairobi (Kenia)	37	-1	1795	9	59	May-Dec only
Java	113	-8	50	9	43	59
Ascension Isl. (St. Helena)	-14	-8	91	7	29	42
American Samoa	-171	-14	10	9	25	40
Lauder (New Zealand)	170	-45	370	9	45	66
Neumayer (Antarctic)	-8	-71	42	25	64	78
South Pole (Antarctic)	-25	-90	2835	16	60	77

3.1.2 Methodologies

Model values at the station's locations in the horizontal are interpolated linearly from the model gridded data. In the vertical the Geopotential at model-levels and model-halflevels is calculated. The methods of the horizontal interpolation and the vertical level calculation are described in detail in chapter 5 (Evaluation with ground-based data from global networks).

To make the sonde measurement results comparable to model data, the sonde ozone units have to be converted from partial pressure to volume mixing ratio in case of the MOZART and TM5 stand-alone runs and to mass mixing ratio for all other model runs, furthermore the sonde profile has to be fitted to the model levels.

Conversion of O_3 units:

$ppmv = p_{O3} * 10/p$ *p_{O3}=ozone partial pressure*

$ppmm = p_{O3} * 10/p * M$ *M=molecular ratio O_3/air = 1.657*

Fitting to model levels:

In a first step, the sonde values are linearly interpolated to the model levels, in a second step they are averaged between the model levels.

No temporal interpolation is done, instead the model results are compared to the sonde profiles closest in time. From these individual comparisons, monthly mean differences are calculated provided that three single comparison results per month are available. These monthly mean results are used for the evaluation of time series. For a more general overview, mean differences of all comparison results are calculated for each model run. The deviation of the model from the observations is either expressed as relative difference (100*(model-sonde)/sonde) or modified bias normalized with the median differences (for details see Annex 7).

3.2 Offline simulations

The evaluation results of MOZART_V1 and TM5_V3 stand-alone runs are discussed by means of five different sonde locations (Ny Alesund, Hohenpeissenberg, Java, Lauder and Neumayer) each representing different geographical regions: Arctic, northern hemispheric mid latitudes, tropics, mid latitudes of the southern hemisphere and Antarctic. Table 3.3 summarizes the normalized bias for the entire evaluation period (Jan 2003 - Dec 2003) grouped by three different vertical sections, the free troposphere, the UT/LS (upper troposphere/lower stratosphere) and the free stratosphere. The extension of these vertical sections is not fixed, but differs according to the altitude of the tropopause.

Table3.3: Mean normalized bias of MOZART_V1 and TM5_V3 stand-alone runs compared to ozone sonde measurements averaged for 3 different atmospheric altitude layers (negative values=model lower than measurement).

	TM5_V3			MOZART_V1			Approx. pressure levels of vertical layers [hPa]		
	Trop	**UT/LS**	**Strat**	**Trop**	**UT/LS**	**Strat**	**Trop**	**UT/LS**	**Strat**
Ny Alesund	-0.2±0.2	-0.4±0.5	-0.1±0.2	-0.3±0.2	-0.3 - +0.2 ±0.5	+0.2±0.2	800-400	300-100	95-10
Hohenpeissenberg	-0.1±0.3	-0.7±0.9	-0.4±0.3	-0.1±0.2	-0.7±0.8	-0.5±0.2	800-300	250-100	95-10
Java	-0.1±0.4	-0.4±0.4	-0.1±0.2	-0.2±0.3	+0.4±0.5	±0±0.1	800-150	100-60	50-10
Lauder	-0.1±0.2	+0.3±0.6	-0.1±0.2	-0.3±0.2	-0.6 - +0.1 ±0.5	-0.3±0.2	800-300	250-100	95-10
Neumayer	-0.2±0.1	-0.3 - +0.2 ±0.3	+0.1±0.2	-0.3±0.1	-0.3 - +0.3 ±0.5	-0.2±0.2	800-400	300-100	95-10

Both models have consistently negative biases in the troposphere, which are lower for the TM5 model. The bias in the stratosphere is mostly negative and again higher in the MOZART model output. Both models show largest deviations from the measurements in the UT/LS both in positive and negative direction. Quite often this layer is characterized by a steep gradient from negative to positive bias and back (see fig.3.1), especially in the MOZART model. Figure3.1 is an example for the variation of the individual comparison results. It shows the model output (black crosses) and the mean model and sonde ozone profiles on the left

and centre plots, the right plot shows the differences (model-sonde) in black and the mean normalized bias. To enhance the tropospheric ozone values the centre plot shows ozone mixing ratios on a logarithmic scale and the altitude on a linear scale. The normalized bias has its largest variation in the UT/LS where also the ozone mixing ratios show the largest variation.

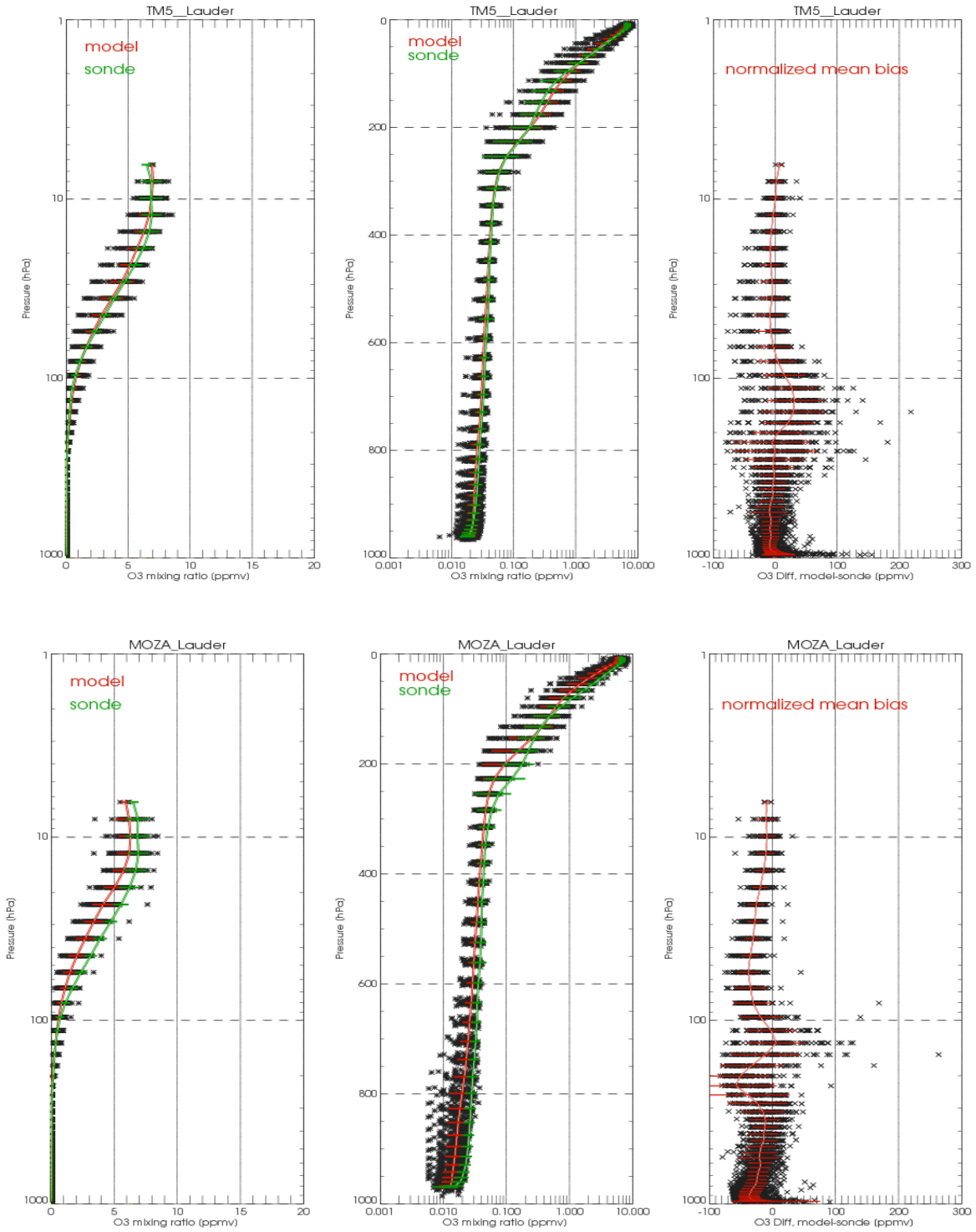


Fig.3.1. Variation of model ozone mixing ratio (TM5_v3 on top, MOZART_v1 below) and ozone differences (model-sonde) in the time period Jan 2003-Dec 2003 (blach crosses) and mean model O3 profiles(red) and ozone sonde profiles (green) (normalized bias is multiplied by factor 100).

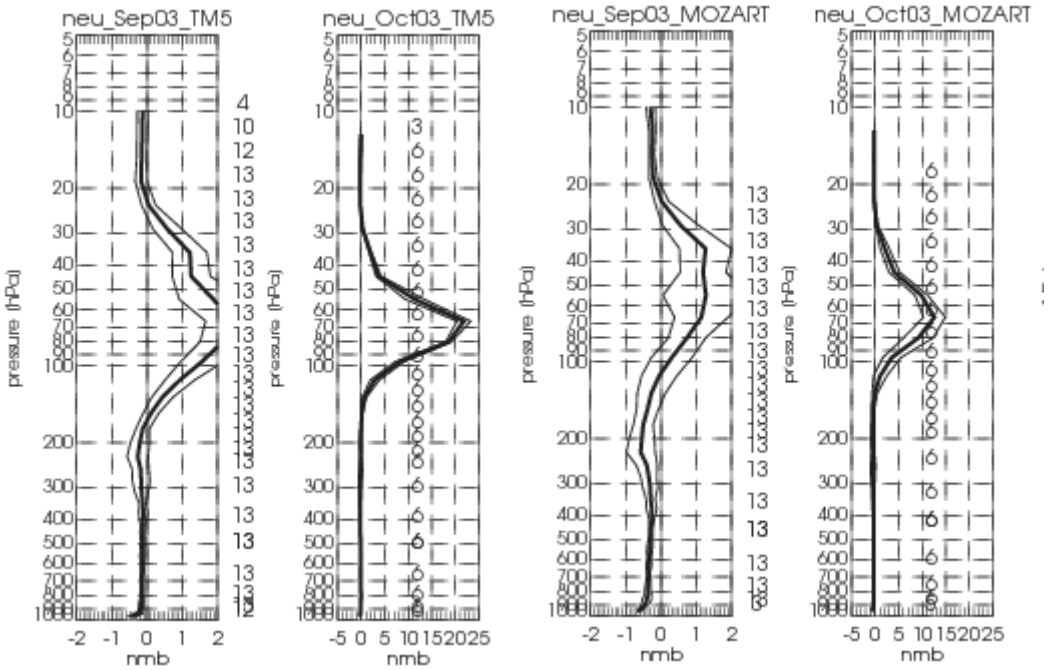


Fig.3.2 Normalized bias (nmb) or TM5_V3 left and MOZART_V1 (right) at location Neumayer (Antarctica) under ozone hole conditions. Numbers on the right hand side of the figures are numbers of ozone profiles available for comparisons. Please note the different nmb scales in September and October. Monthly means are calculated with a minimum of 3 ozone profile comparisons.

The behaviour of the TM5 and MOZART stand-alone model runs under ozone hole conditions is shown in figure 3.2 for the location of Neumayer. The ozone hole evolved at the end of September 2003. In both models the reduced ozone concentrations are not reflected, but the bias of the MOZART model is about half that of the TM5.

3.3 GEMS-GRG reanalysis

Figures 3.3-3.7 show the outcome of the comparisons of model runs exoz, eydy, eyih, eyq6, f026, eybl and ez3h with sonde profiles from Ny Alesund, Hohenpeissenberg, Paramaribo, Lauder and Neumayer averaged over the time period July to August 2003.

All model runs have in common an underestimation of the ozone mixing ratios in the troposphere and an overestimation in the stratosphere. Runs exoz and eydy have negative ozone mixing ratios in the troposphere at the last forecast time. For that reason, mean differences at Lauder and Neumayer, which are stations with a high number of late balloon launches, are out of range in the troposphere.

With the introduction of ozone assimilation the Mozart model improved in the stratosphere, runs eyih and f026 have only small or insignificant biases when compared to the ozone sondes. In the troposphere however, model runs with ozone assimilation show only a small decrease of differences between sonde and model, if any.

TM5 runs are comparable to Mozart runs with ozone assimilation in the stratosphere but have lower negative bias in the troposphere, except at tropical and antarctic stations. In the tropical stratosphere the TM5 runs eybl and ez3h underestimate the ozone mixing ratios by as much as -30%, whereas in the

antarctic stratosphere the TM5 runs overestimate the ozone mixing ratio. Especially run ez3h which is up to 58%±38% higher than the measurements in the stratosphere.

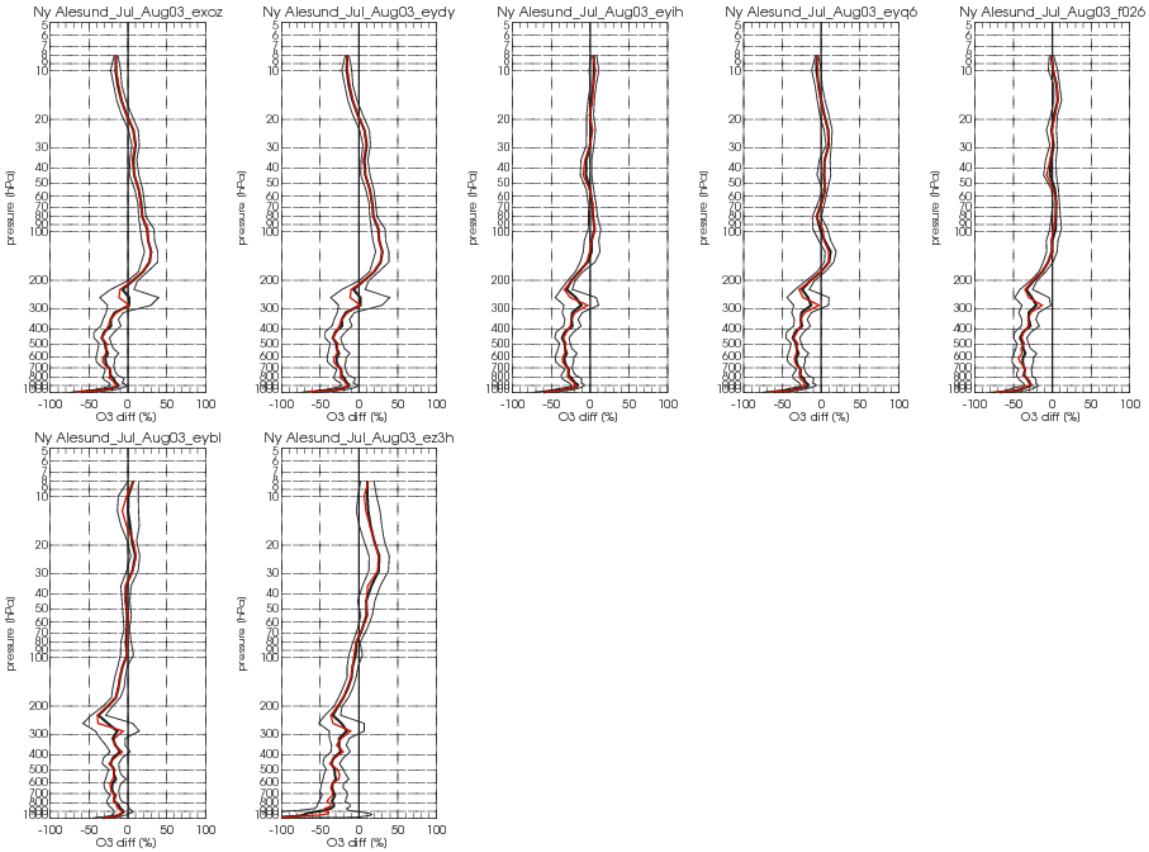


Fig.3.3: Relative differences between ozone sonde measurements at Ny Alesund and Mozart model runs exoz, eydy, eyih, eyq6 and f026 and TM5 model runs eybl and ez3h (model-sonde)/sonde. Mean differences in black, median in red. Thin black lines: 1standard deviation. Monthly means are calculated with a minimum of 3 ozone profile comparisons.

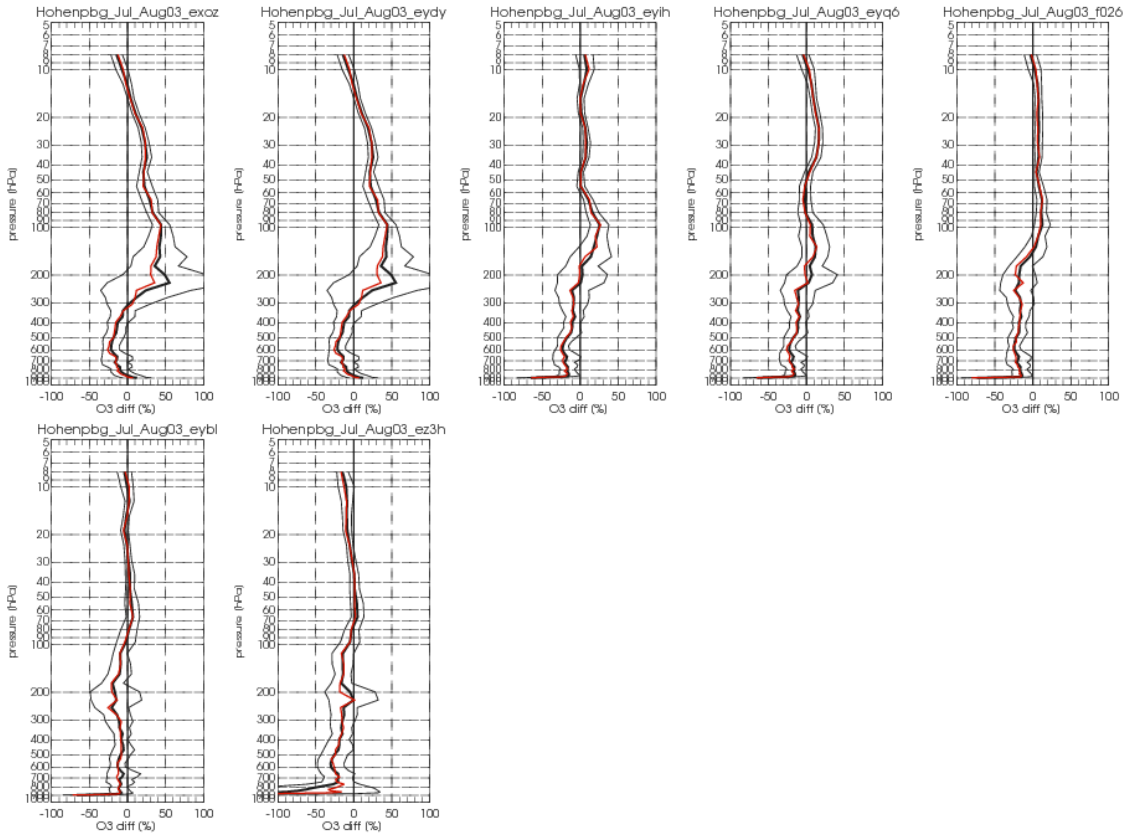


Fig.3.4: Same as fig.3.3 but for Hohenpeissenberg.

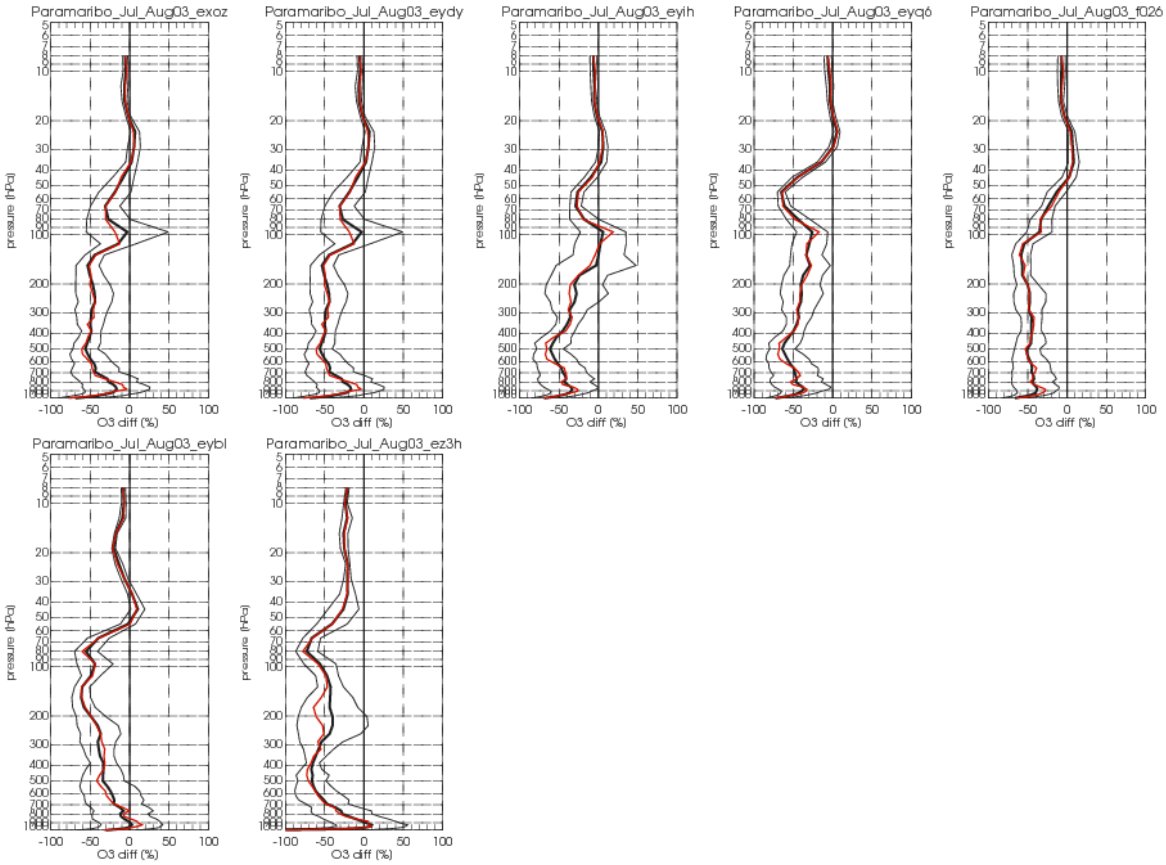


Fig3.5: Same as fig.3.3 but for Paramaribo.

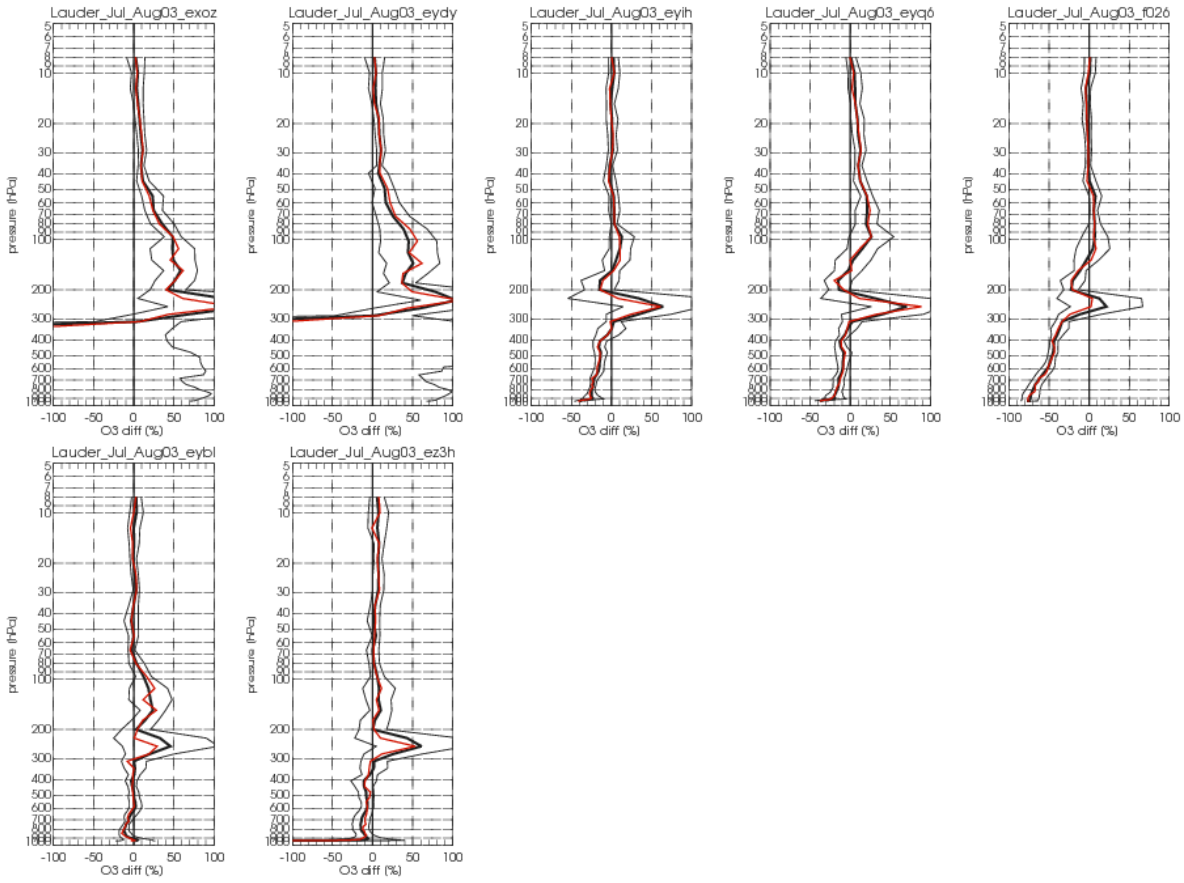


Fig3.6: Same as fig.3.3 but for Lauder.

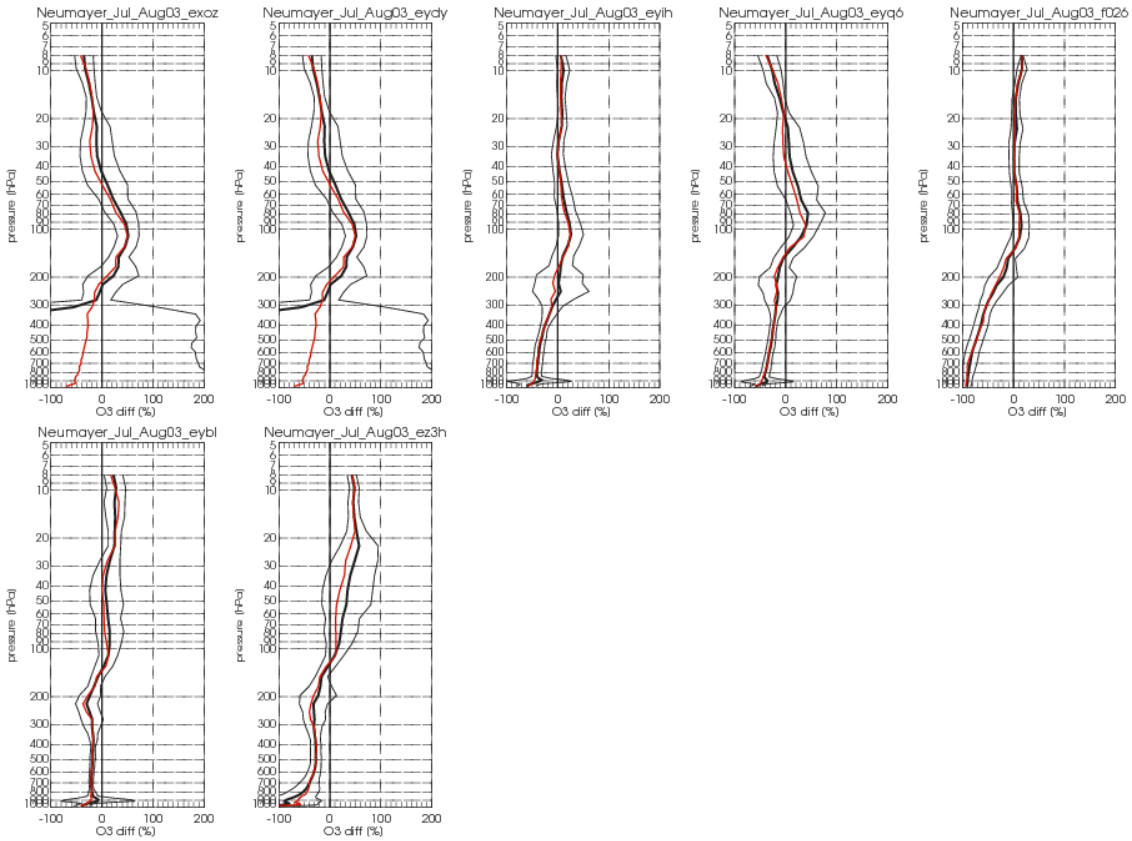


Fig3.7: Same as fig.3.3 but for Neumayer.

A closer look at the differences between the various model runs using Hohenpeissenberg as an example, reveals an improvement of the Mozart model in the UTLS and the stratosphere when ozone assimilation from satellite data is introduced (eyih, eyq6,f026), although tropospheric ozone mixing ratios are still too low and stratospheric mostly somewhat too high. In the UTLS the mean differences as well as the standard deviations are reduced at the 223 hPa level from +55%±82% (exoz, eydy) to +1%±31% (eyih), +3%±31% (eyq6) and -18%±25% (f026) and at the 113 hPa level from +43%±16% to +22%±13% (eyih), +8%±19% (eyq6) and +7%±9% (f026). In the stratosphere the eyih and f026 runs show a small positive bias of less than 10%, whereas the eyq6 run has a maximum difference of +16%±5% between 20 and 30 hPa. In the troposphere however the negative bias remains unchanged with ozone assimilation and ranges from -10% to -20%, in all model runs, except ez3h where it is slightly higher with up to -30%. TM5 model runs are comparable to Mozart runs with ozone assimilation, in the stratosphere TM5 run eybl is slightly better with insignificant differences between measurements and model, ez3h has a small negative bias above 30hPa of -8%.

Comparisons with model runs eyq6 and eyih cover the time period May 2003 to Dec 2003 and the entire year 2003 for model run f026. Figures 3.8a-c show monthly mean relative differences for Hohenpeissenberg of these three model runs. The largest variations in the course of the year are observed in the UTLS region, where the relative differences and the standard deviations between all three model runs and the sonde measurements are largest. The levels of highest variability roughly cover 300 hPa to 100 hPa. In late winter and early spring the relative differences are largest at these levels, which is most probably due to the high variability of the ozone layer caused by atmospheric dynamical processes at this time of the year.

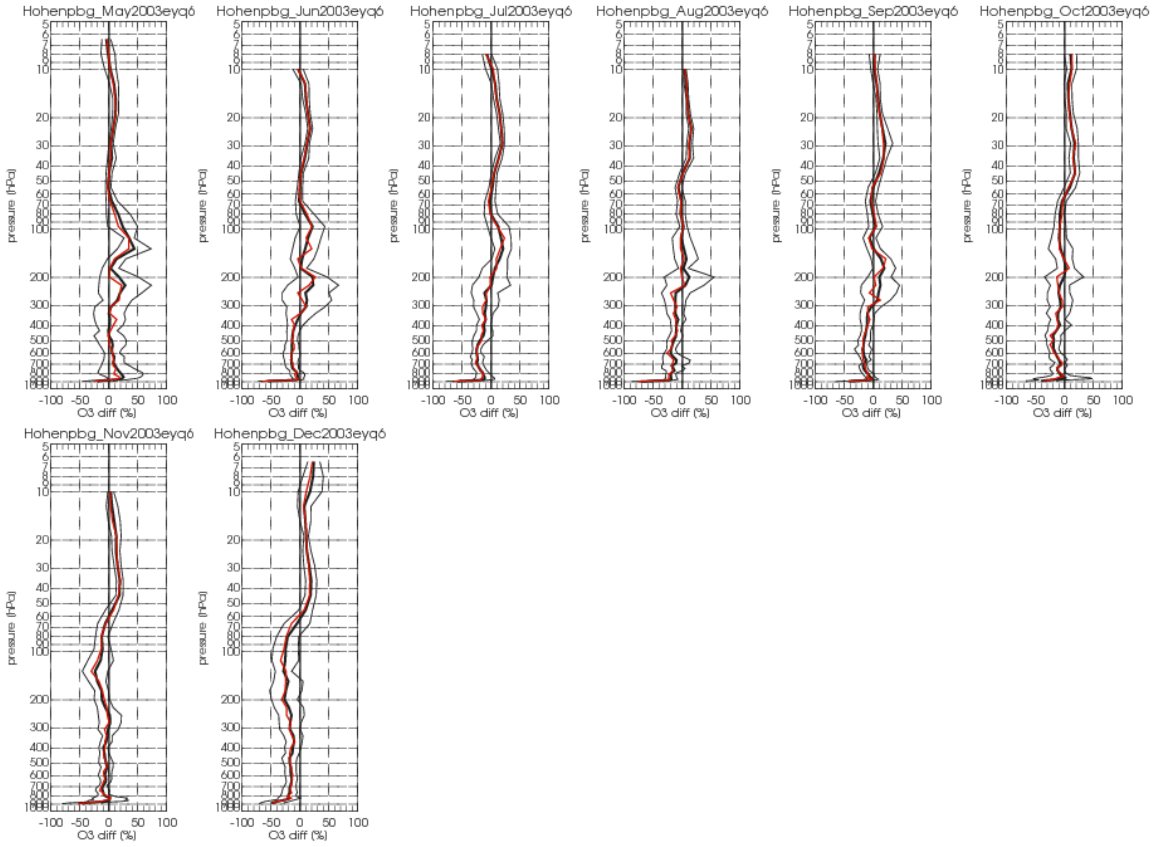


Fig.3.8a: Relative differences between ozone sonde measurements at Hohenpeissenberg and Mozart run eyq6 (model-sonde)/sonde (monthly mean differences in black, median in red). Thin black lines: 1standard deviation. Monthly means are calculated with a minimum of 3 ozone profile comparisons.

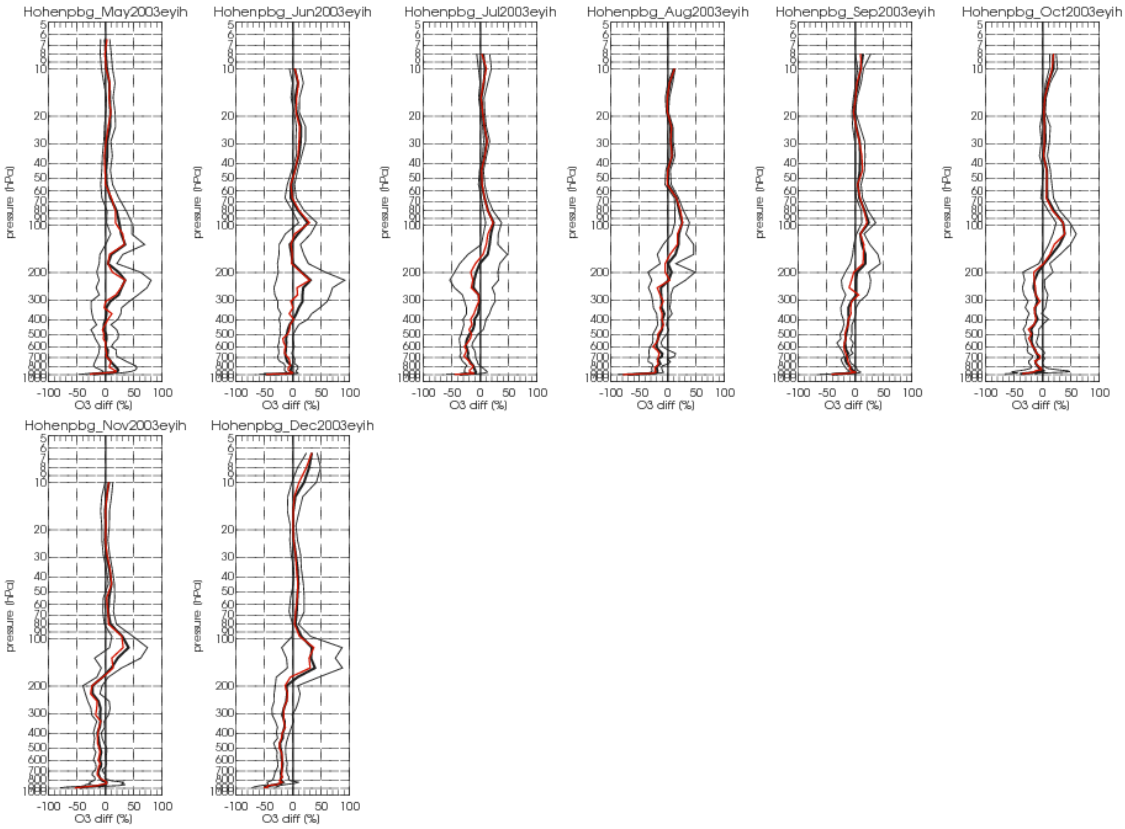


Fig 3.8b: Same as fig 3.8a but for model run eyih.

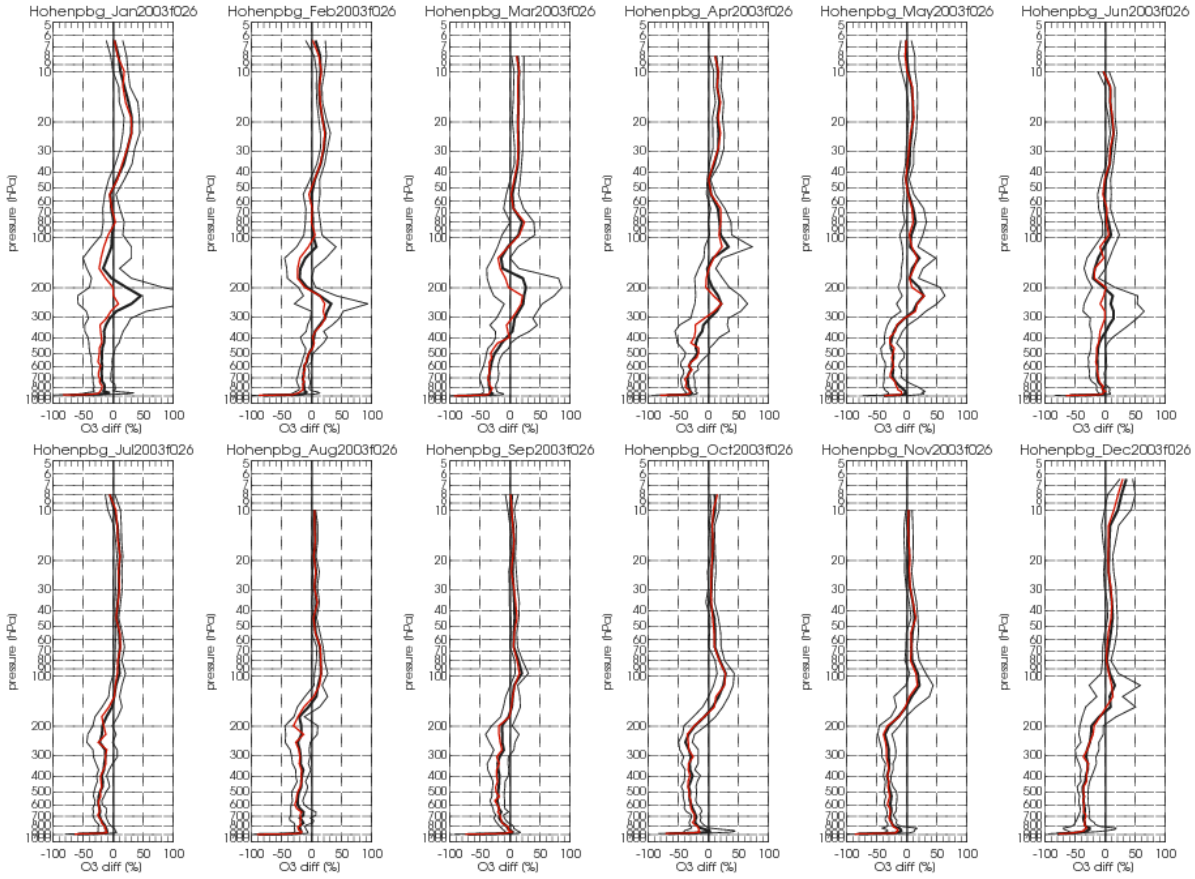


Fig 3.8c: Same as fig 3.8a but for model run f026.

Ozone hole conditions

Figure 3.9 presents the monthly mean relative differences between model runs eyih, f026, eyq6 and the sonde measurements under ozone hole conditions in the Antarctic. The ozone hole starts to develop at the end of August. While runs eyih and f026 are still in good agreement with the measurements in August, run eyq6 already shows too high ozone mixing ratios compared to the measurements. In September and October when the ozone hole is fully developed, run eyq6 is more than twentyfold higher than the measurements, run eyih is up to 200% higher and model run f026 has the smallest differences compared to the measurements but is still up to 135% too high.

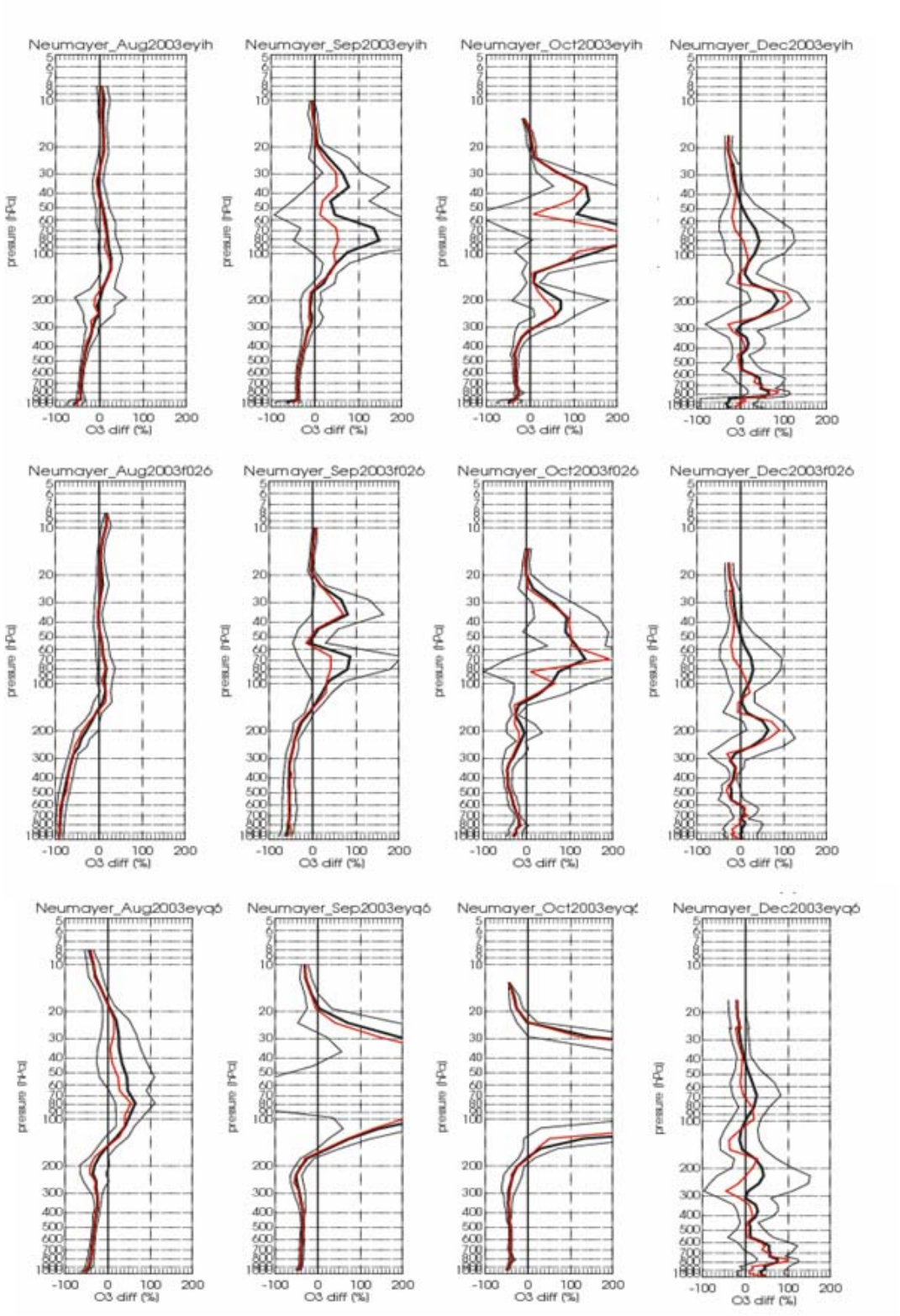


Fig. 3.9: Relative differences between ozone sonde measurements at Neumayer and Mozart runs eyih, f026 and eyq6 [(model-sonde)/sonde] under ozone hole conditions. Monthly mean differences in black, median in red. Thin black lines: 1standard deviation. Monthly means are calculated with a minimum of 3 ozone profile comparisons.

3.4 GEMS-GRG forecast

In the regions north of 40deg latitude, the comparison results show largely the same variations over the course of the year, although the magnitude and the altitude of these variations change from the mid latitude to the polar regions. In the free troposphere the model ozone mixing ratios decrease compared to the ozone sonde values from winter to summer. Around 700hPa for example, the monthly mean relative difference is between +7% at Sodankyla and +30% at Hohenpeissenberg in January and between -35% at Sodankyla and -12% at Hohenpeissenberg in July. The monthly mean differnce of all other mid latitude stations is found within these limits and shows a comparable variation from winter to summer. In the stratosphere the monthly mean differences change in the opposite direction. At the 23hPa level the monthly mean differences in January vary from -29% at Sodankyla to +1% at Legionowo. In July, monthly mean relative differences between +11% at Ny Alesund and +19% at Lerwick are observed. Largest differences at all stations north of 40deg latitude are observed in the upper troposphere/lower stratosphere region from December to March, where the monthly mean differences are higher than +50% and may even exceed +100%.The reason for these high deviations is most probably a wrong position of the tropopause height in the model.

In the tropics, between 18deg north and 14deg south, no significant changes of monthly mean differences with season are observed. The tropical regions do not have the large variability in ozone concentrations as the more northern regions and less variation of the tropopause altitude, both factors lead to the insignificant temporal change of monthly mean relative differences. The pattern of the monthly mean differences at the different tropical stations, especially in the troposphere, is not as consistent as further north. One reason for that is certainly the quite low number of balloon soundings available for comparisons in the tropics. A mostly negative bias is found in the troposphere ranging from ±0% to -45%. In the UTLS region these mostly negative bias shifts to positive values which at some stations exceed +50%.

In the following section a more detailed discussion of the comparison results is presented by means of three stations Ny Alesund, Hohenpeissenberg and Neumayer, representing arctic, mid latitude and Antarctic regions.

Figure 3.10a shows the monthly mean normalized bias at Ny Alesund. In the winter months ozone mixing ratios of the NRT model start to increase at lower levels than measured by the sonde, which allows the conclusion that the large differences in the UTLS are caused by insufficient representation of the tropopause altitude by the model. At the same time the ez2m model is lower than the measurements in the mid stratosphere, which might also be caused by the low tropopause height in the model. In the free troposphere a small negative bias (~0.1) is observed from January to March which increases to ~0.3 for the rest of the year. Highest biases are observed at the UT/LS with exception of September and October 2008. Standard deviation has its maximum also at the UT/LS. Both facts might be assigned to incorrect tropopause height in the model. Figures 3.11a and b from February 2008 may highlight this assumption.

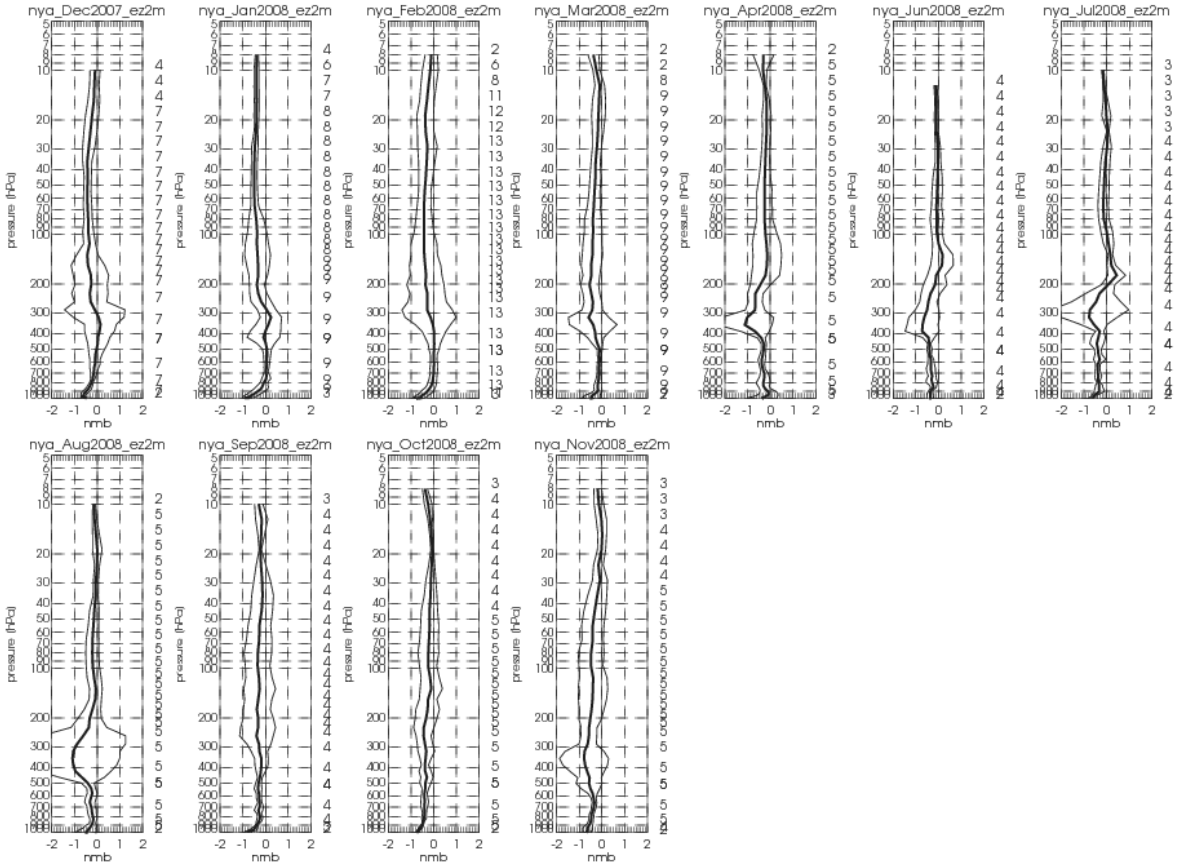


Fig. 3.10: Ny Alesund monthly mean normalized bias (nmb) profiles of the GEMS NRT forecast (ez2m) model , thin black lines: 1 standard deviation. Monthly means are calculated with a minimum of 3 ozone profile comparisons only. Figures on the right hand side of plots are numbers of ozone profiles comparisons.

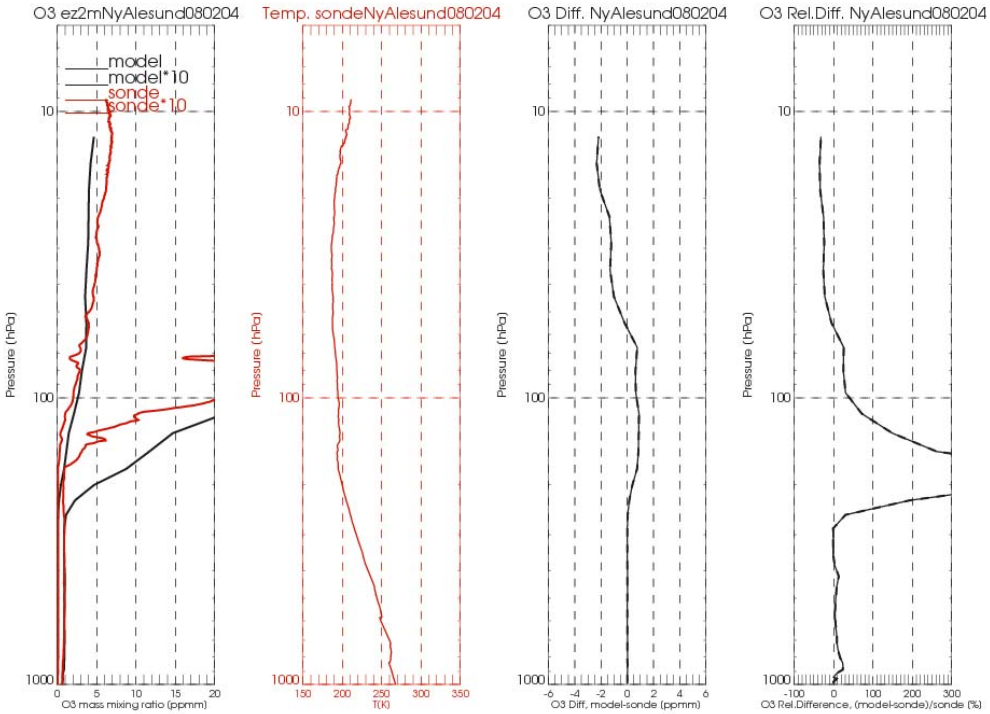


Fig.3.11a: Ozone profiles, temperature profile, ozone difference and ozone relative difference profiles on 4th April2008 at Ny Alesund.

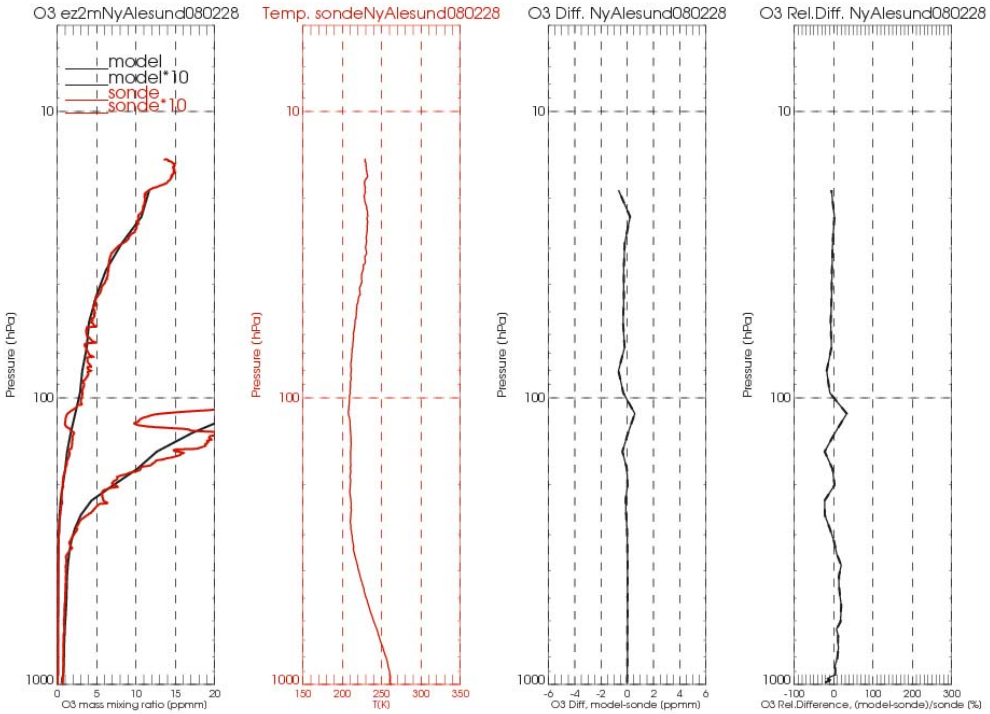


Fig.3.11b: Same as fig.2a but on 28th April 2008.

In February a number of comparisons e.g. 04, 15, and 29 show very large differences in the altitude of the tropopause between model and measurements, whereas on other days e.g. 10, 16, 28 of this month the tropopause altitude of model and measurement are in good correspondence. It is well known that in February the ozone layer is highly variable due to dynamic processes in the atmosphere. A more detailed analysis of the atmospheric situation on the different days might be helpful to detected under which circumstances the model yields the best and worst agreement with the sonde measurements.

To gain more insight into the differences between model output and sonde measurements, scatter plots with model ozone as function of sonde ozone for different model levels are plotted (fig.3.12). It is surprising that correlations are poor at almost all levels. From fig.3.10 one might expect better correlations especially in the free troposphere where standard deviations are very small. This result leads to the conclusion that although the mean bias and its variation is quite small, the model results might differ largely when compared on an individual bases, as already demonstrated in figure 3.11

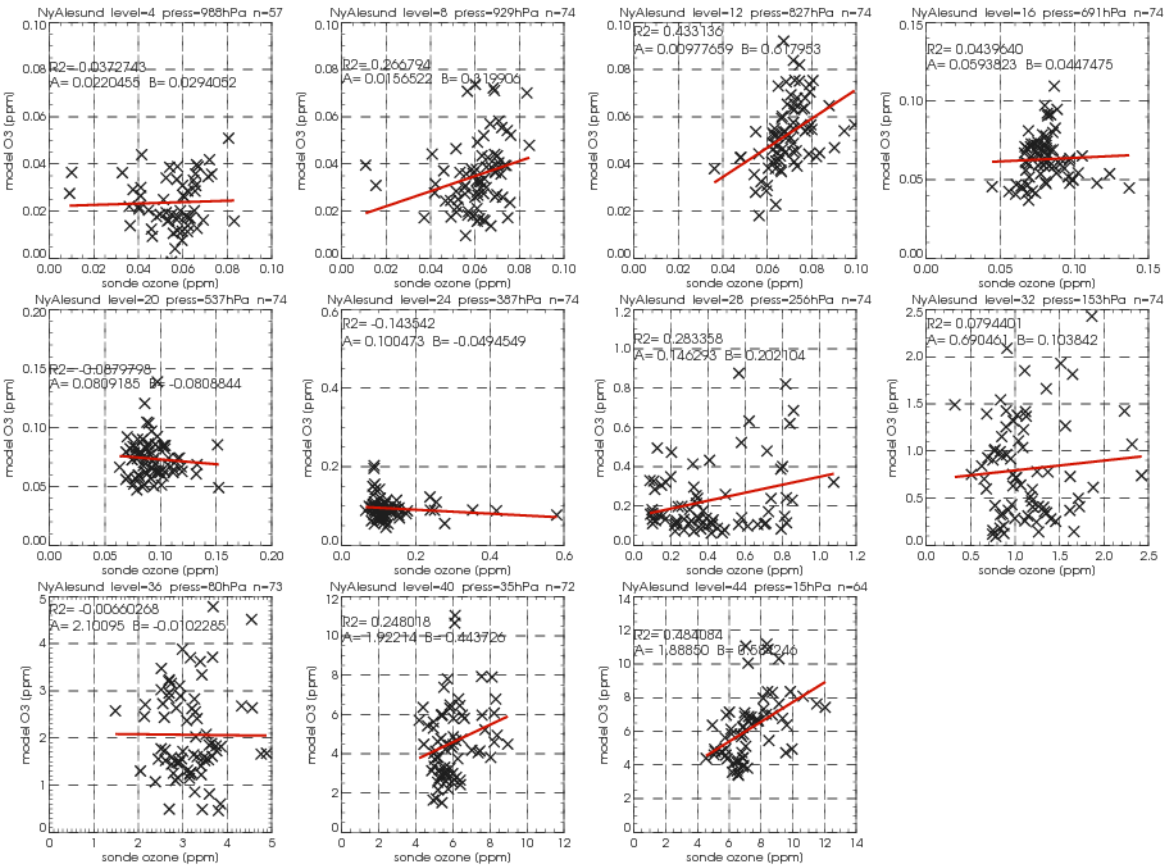


Fig. 3.12 Scatter plot with GEMS-GRG Forecasts model ozone as function of measured sonde ozone.

Monthly mean normalized biases for the location of Hohenpeissenberg are shown in figure 3.13. Seasonal variation is most striking in the UT/LS. In winter the bias increases to 1± 0.8 at these levels and decreases within the rest of the year to approx. 0.1± 0.2 with exception of July and September when a negative bias of up to -0.5± 0.7 is observed. The reasons for the larger bias and standard deviation in winter are probably the same as for Ny Alesund, i.e. higher natural variability of the ozone layer itself in this time of the year and variability in the tropopause altitude which is not always correctly represented in the model. In contrast to Ny Alesund however, the standard deviation of the bias is significantly smaller at all altitudes for the location of Hohenpeissenberg. Scatter plots of model ozone as function of sonde ozone (fig.3.13) show much better correlations for Hohenpeissenberg. Especially in the stratosphere above the 200hPa level the correlations are good. However, in the free troposphere between 750hPa and 500hPa the correlations are poor. Only further analyses, probably of individual days, may reveal the reasons for the poor correlations in these levels.

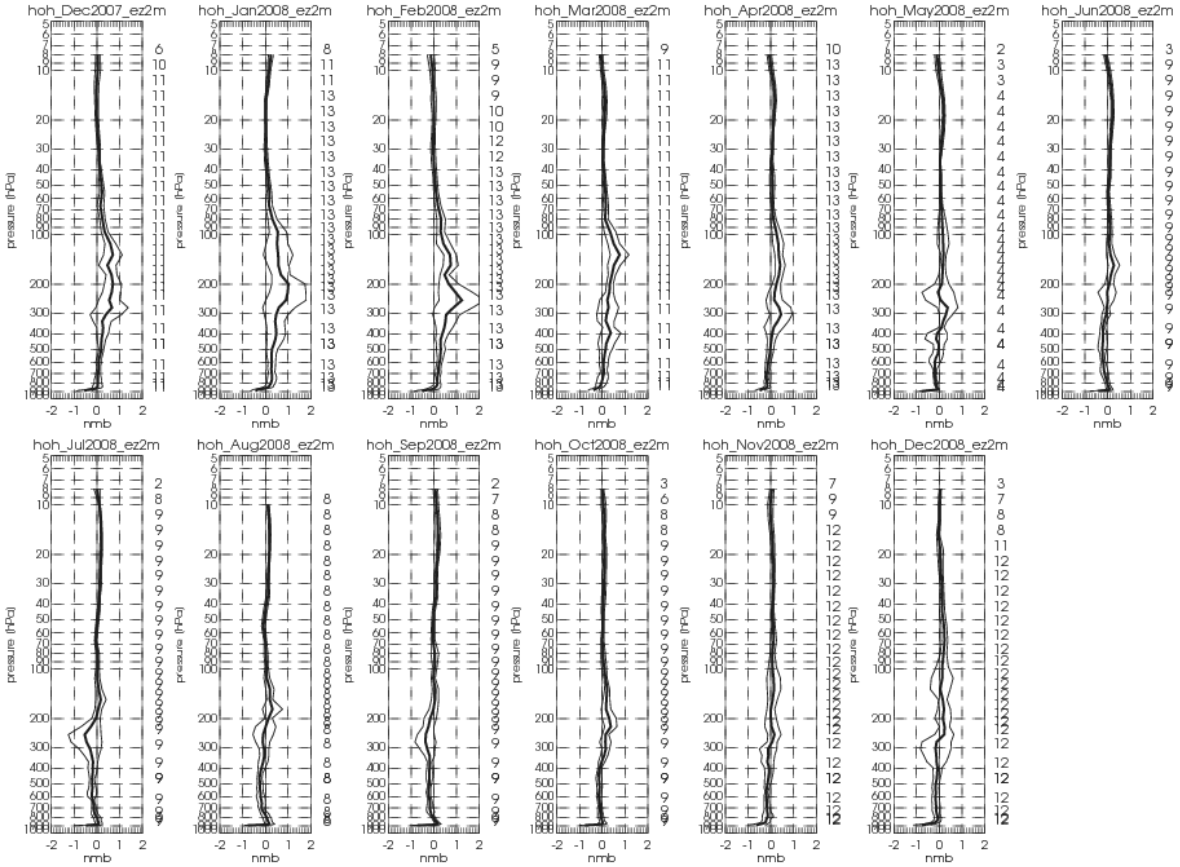


Fig. 3.13: Same as fig.3.10, but for Hohenpeissenberg.

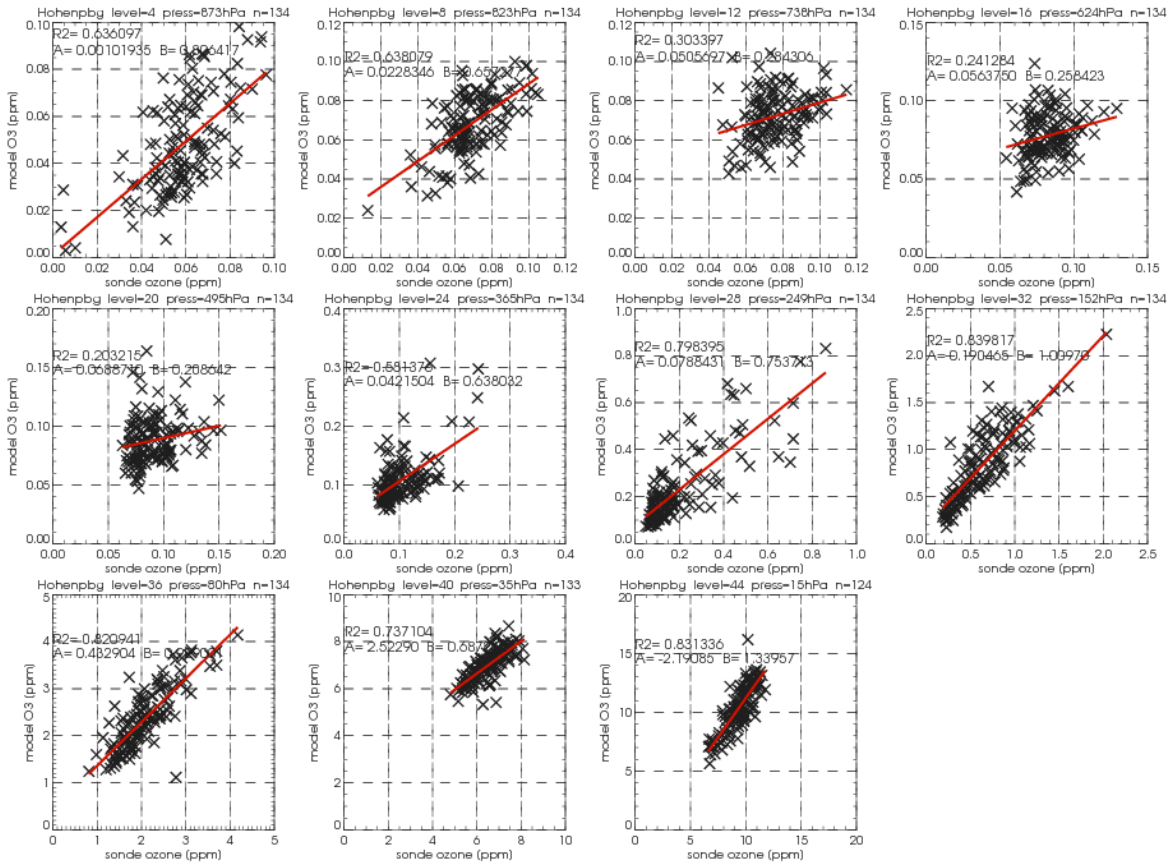


Fig. 3.14: Same as fig. 3.12 but for Hohenpeissenberg.

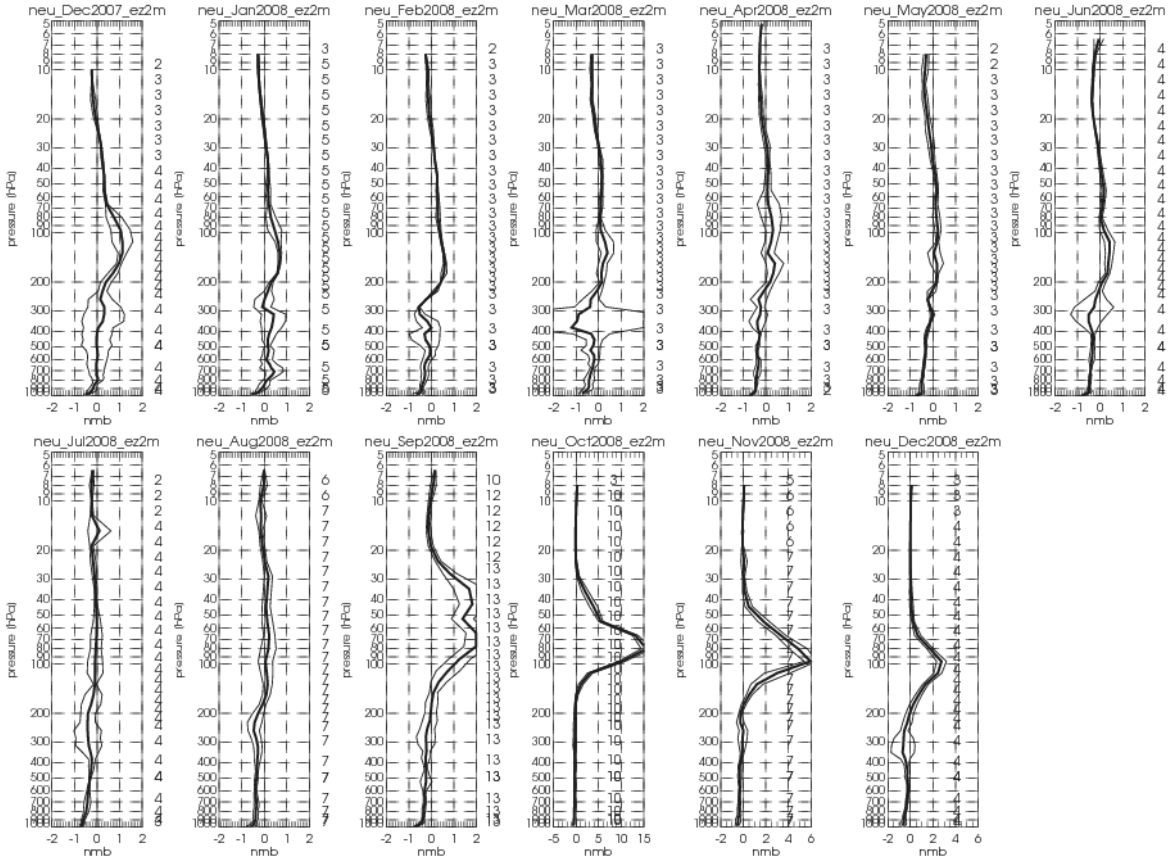


Fig. 3.15: Same as fig.3.10, but for Neumayer. Pleas note the different nmb scale from October to December 2008

Figure 3.15 brings the focus to the model output under ozone hole conditions in the Antarctic. It shows the monthly mean normalized biases for the location of Neumayer. It is obvious that the model does not reflect the ozone hole correctly and largely overestimates the ozone values.

3.5 Summary

The vertical ozone distribution of different MOZART and TM5 model simulations was evaluated by comparisons with ozone profile measurements from balloon sondes. Model versions under investigation were stand-alone runs, coupled runs with and without data assimilation and the GEMS-NRT forecast.

MOZART and TM5 stand-alone model simulations both have consistently negative biases in the troposphere, which are smaller for the TM5 model than for MOZART. The bias in the stratosphere is mostly negative and again smaller for the TM5 model output. Both models show largest deviations from the sonde measurements in the UT/LS in positive as well as negative direction. Quite often this layer is characterized by a steep gradient from negative to positive bias and back, especially in the MOZART model.

MOZART re-analyses runs all have in common an underestimation of the ozone mixing ratios in the troposphere and an overestimation in the stratosphere. With the introduction of ozone assimilation the coupled MOZART-IFS models improved in the stratosphere, leading to only small or insignificant biases

when compared to the ozone sondes. In the troposphere however, model runs with ozone assimilation show only a small decrease of differences between sonde and model, if any.

Coupled TM5-IFS model simulation results are comparable to MOZART runs with ozone assimilation in the stratosphere but have a lower negative bias in the troposphere, except at tropical and antarctic stations. In the tropical stratosphere the TM5 runs eybl and ez3h underestimate the ozone mixing ratios whereas in the antarctic stratosphere the TM5 runs largely overestimate the ozone mixing ratio.

GEMS-NRT forecasts for regions north of 40°N, show typical variations over the course of the year. In the free troposphere the model ozone mixing ratios decrease compared to the ozone sonde values from winter to summer. In the stratosphere the differences change in the opposite direction. Largest differences and also largest standard deviations are observed in the upper troposphere/lower stratosphere region from December to March. In the tropics, no significant changes of monthly mean differences with season are observed. The pattern of the monthly mean differences at the different tropical stations, especially in the troposphere, is not as consistent as further north. A mostly negative bias is found in the troposphere which shifts in the UTLS region to positive values

All model simulations largely underestimate the antarctic ozone hole, i.e. the models produce too much ozone. Although the re-analyses runs eyih and f026 with data assimilation had a significantly lower bias than the offline runs, the bias of the GEMS-NRT forecast in winter 2008/2009 was as high as in the offline model simulations.

3.6 Conclusions and recommendations

The evaluation of the vertical ozone distribution by comparisons of the model results with ozone sonde profiles identified several aspects common to all model simulations:

- UTLS bias is largest with highest standard deviations
- negative bias in the troposphere
- data assimilation has a negligible effect on tropospheric model results
- ozone hole in the Antarctic is not reflected by the models.

A large bias in the UTLS has already been identified by several model evaluation studies (Pan 2006, Pan 2007) and has been primarily assigned to two factors. First, the meteorological fields used by the model system including the way the dynamics is coupled to the chemistry and second, the coarse model resolution in the vertical as well as in the horizontal direction. A false transport through the tropopause might also be the cause for biases in the troposphere and stratosphere, when either too much ozone is transported in one or the other direction through the tropopause. Whether the negligible improvements in the troposphere by the use of data assimilation are also connected to the transport processes at the UTLS is not yet proved but conceivable.

It is therefore recommended to test models with reduced vertical and horizontal resolution mainly in the UTLS.

To get a better reflection of the Antarctic ozone hole, the already started model developments should be continued.

3.7 References

1. Deshler,T., J.L.Mercer, H.G.J.Smit, R.Stubi, G.Levrat, B.J.Johnson, S.J.Oltmans, R.Kivi, A.M.Thompson, J.Witte, J.Davies, F.J.Schmidlin, G.Brothers, T. Sasaki (2008) Atmospheric

comparison of electrochemical cell ozonesondes from different maufacturers, and with different cathode solution strengths: The Balloon Experiment on Standards for Ozonsondes. J. Geophys. Res.113, D04307, doi:10.1029/2007JD008975

2. Pan,L.L., P.Konopka, E.V.Browell (2006) Observations and model simulations of mixing near the extratropical tropopause. J. Geophys. Res.111, D05106, doi:10.1029/2005JD006480
3. Pan,L.L., J.C.Wei, D.E.Kinnison, R.R.Garcia, D.J.Wuebbles (2007) A set of diagnostics for evaluating chemistry-climate models in the extratropical tropopause region. J. Geophys. Res.112, D09316, doi:10.1029/2006JD007792
4. Smit;H.G.J., W.Straeter, B.J.Johnson, S.J.Oltmans, J.Davies, D.W.Tarasick, B.Hoegger, R.Stubi, F.J.Schmidlin, T.Northam, A.M.Thompson, J.C.Witte, I.Boyd (2007), Assessment of the performance of ECC-ozonesondes under quasi-flight conditions in the environmental simulation chamber: Insights from the Juelich Ozone Sonde Intercomparison Experiment (JOSIE), J. Geophys. Res. 112, D19306, doi:10.1029/2006JD007308

4. Evaluation with routine aircraft soundings

Contributors: Elguindi Nellie, Ordonez Carlos, Thouret Valérie, Athier Gilles, Cammas Jean-Pierre (CNRS-LA, Toulouse, France), and the modeller teams

4.1 Datasets and methodologies

4.1.1 Datasets

The MOZAIC program provides ozone (O3) and carbon monoxide (CO) automatically recorded on board commercial aircraft since August 1994 for O3 and December 2001 for CO (*Marenco et al.*, 1998). Details on the program itself including scientific achievements can be found on the following web site : http://mozaic.obs-mip.fr/web.

The MOZAIC ozone analysers are dual-beam UV absorption Model 49-103 from Thermo Environment Instruments, USA. Their performance and installation are completely described in Thouret et al. (1998). The instruments are laboratory calibrated before and after the flight periods, whose duration is generally 12 to 18 months. The laboratory calibration is performed with a reference analyzer periodically cross-checked with a National Institute of Standards and Technology in France. Additionally and during the flight operation period, each instrument is checked every flight for the zero and for the calibration factor, using a built-in ozone generator. Finally, intercomparisons are made between aircraft when they fly close in location and time, which happens several times a month and guarantees the permanence of the analyzers specifications: ± 2 ppbv, ± 2%.

The MOZAIC CO analyzer is an improved version of a commercial Model 48CTL from Thermo Environmental Instruments, based on the Gas Filter Correlation principle of infrared absorption by the 4.67μm fundamental vibration rotation band of CO. Radiation from an infrared source is chopped and passes through a gas filter which alternates between CO and N2 via the rotation of the filter wheel. The radiation then passes through a narrow band pass filter and a multiple optical pass sample cell where absorption by the sample gas occurs. The IR radiation exits the sample cell and falls on a PbSe solid state IR detector. Other gases do not cause modulation of the detector signal since they absorb the reference and measure beams equally. Thus, the Gas Filter Correlation System responds specifically to CO. The Model 48CTL is qualified by U.S. EPA designated Method (EQSA-0486-060). The accuracy specification of the commercial instrument is 10 ppbv CO for 300 s integration time. Several major modifications have been made by CNRS-LA in order to improve the Model 48CTL instrument characteristics in order to achieve performance suitable for routine aircraft measurements : ± 5 ppbv, ± 5% precision for a 30 s response time (Nédélec et al., 2003).

Quality and usefulness of this MOZAIC data set for model evaluation has been recognized since the beginning of the program. For example, Law et al., (1998) presented the first validation study of the TOMCAT model with the MOZAIC ozone profiles. A following study by Law et al., (2000) has compared several state-of the-art CTMs at this time to the MOZAIC data. On the other hand, Thouret et al., (1998) have shown that the MOZAIC tropospheric ozone climatology was in reasonable agreement with the one based on the ozone soundings network which used to be the reference in terms of data set for models evaluation. Both data sets are now widely used for models and satellites validation on a complementary way. Nonetheless, it is still worth mentioning that as with any dataset there are associated uncertainties. Namely, the representativeness of aircraft measurements within highly frequented take-off/landing flight corridors near large airports and the extension of the profiles over a 50-100 km horizontal distance during the aircraft's ascent/descent.

In the framework of GEMS-GRG, the years 2003 and 2004 have been chosen for the exercise. For technical reasons, only the period between May and December 2003 will be evaluated for all the models. In 2004, MOCAGE has been run only for the period June to September, while MOZART and TM5 were run for the full year. A comparison will be made for tropospheric vertical profiles of O3 and CO. During this period, the MOZAIC aircraft visited 35 different airports. Table 1 below gives the number of profiles available for the comparison for the 2 years over the 35 airports. For the discussion and summary, we have selected 19 as being the most documented airports and sufficiently representing the different regions of the world for this period (in bold in Table 4-1). During this period, only European airlines (Air France, Lufthansa and Austrian Airlines) were involved in the program. Thus, it is not surprising to see that Paris, Frankfurt, Munich and Vienna present the better statistics with at least one profile a day.

	2003								2004											
Months	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Paris	8	28	29	21	22	26	23	23	26	15	19			17	30	28	22	29		17
Frankfurt	23	25	28	30	21	25	21	30	29	27	30	29	31	30	30	30	29	30	20	12
Munchen	19	25	4	11	12	6	25	28	27	8		17		26	30	28	27	31	28	30
Vienna	22	20	25	23	20	26	26	7			24	30	29	28	24	28	24	11	18	28
New York	7	6		7	8	2	14	11	17	7	8	5	3	9	10	5	2	5		9
Boston		7	3	11	3	3	3	3												
Washingt	1			4	4	7	2													
Chicago	11	16	17	10	9	6	9	6	16								2	12		
Detroit	2	2	3	2		2		4												
Toronto		7	3	5	7	14	2	0												
Montreal		8	5	4	2		2	5												
Dallas	4	2	1	4	1	5	1		1		3	3	6	3	5	3	3	2	0	0
Atlanta				4	5	6	5	7	2	10	11	27	26	28	28	19	10	20	0	0
Houston	2	3	3				2	4												
Beijing		1	2	4	6	4	4				2	2	4	4	6	12	4	8	6	4
Tokyo	14	22	10	4		4	10	8	2		8	12	8	13	15	13	4	8	10	14
Osaka	6	3	14	14	11	14	10	1		2	6	14	12	6	4	5	9	4	4	10
Cairo		5	5	8	3	4	4		2	7	6		2					3		
AbuDhabi		2	4	1	2			2			2		3	11	7	8	8	2	4	15
Koweit	2	3	2		2			2			3		2				3			
Riyadh		2	2				2					2		2		4	2	2		
Dubai	2		4	2			2	8												
Tel Aviv			6	8	7	5	4	8	10	7	7									
Teheran			2					2												
Caracas	9	13	13	12	4	3	1	2	8		5		2	2	3		6	5		7
Abidjan			3			2														
Lagos	4		3	3	1	1			2		3									
Accra	2			2	1	1														
Douala				2		4	6													
Abuja			3	2																
Novakchott	1					2		2												
Conakry			2		2	3														

Adisabeba				3	1		1												
Delhi		6		6	3	3	4	3	2	1	7	3	2	7	7	21	12	6	5	6
Khartoum				4	1	2														

Table 4-1: Number of MOZAIC profiles available for the period of the evaluation (May-December 2003 and January-December 2004) over the 19 selected airports.

For each airport, monthly mean profiles of MOZAIC data are compared with model outputs by evaluating either the Modified Normalized Mean Bias (MNMB) or the Fractional Gross Error (FGE) as defined in the introduction, for 4 different altitude ranges in the troposphere. The surface layer (SL) is defined as the layer below 950 hPa, the boundary layer (BL) lies just above between 950 and 850 hPa, the free troposphere (FT) extends between the boundary layer up to 1 km below the tropopause (based on the usual temperature gradient), and the upper troposphere (UT) is defined as the uppermost 1 km below the tropopause. Concerning the interpolation issues, we extract the models' outputs over the grid box which is nearest to the selected MOZAIC airport, except for TM5 where bilinear interpolation between 4 neighboring points is used. The model output is then linearly interpolated between two time steps (6 hours difference) to the time of the MOZAIC flight. The MOZAIC profile data are interpolated to the model's vertical grid. Then, a daily profile average is calculated and finally the monthly means. In this method, the time difference between the MOZAIC profile and the model output can be as large as 3 hours. This may be considered quite high regarding the ozone diurnal cycle in the boundary layer, however, we did not notice any significant correlation between the time difference and biases.

4.1.2 Methodologies

The scoring method used is described in Annex 7.

4.2 Offline simulations

We ended up having 6 standalone runs to evaluate. They are :

MOZART_v1 for 2003 (May to Dec)
MOZART_v10 for 2004 (Jan to Dec)
TM5_v7 for 2003 (May to Dec)
TM5_v10 for 2004 (Jan to Dec)
MOCAGE-V2 for 2003 (May to Dec)
MOCAGE_HTAP_ES1 for 2004 (June to September)

Figures 4-1 to 4-8 present the vertical profiles of CO and O3 for 2003 (left panel) and 2004 (right panel) of MOZAIC observations (black dots with standard deviation as shaded area) and the coupled runs (color lines). The main characteristics by regions are described below.

The general behavior regarding O3 and CO profiles are the following:

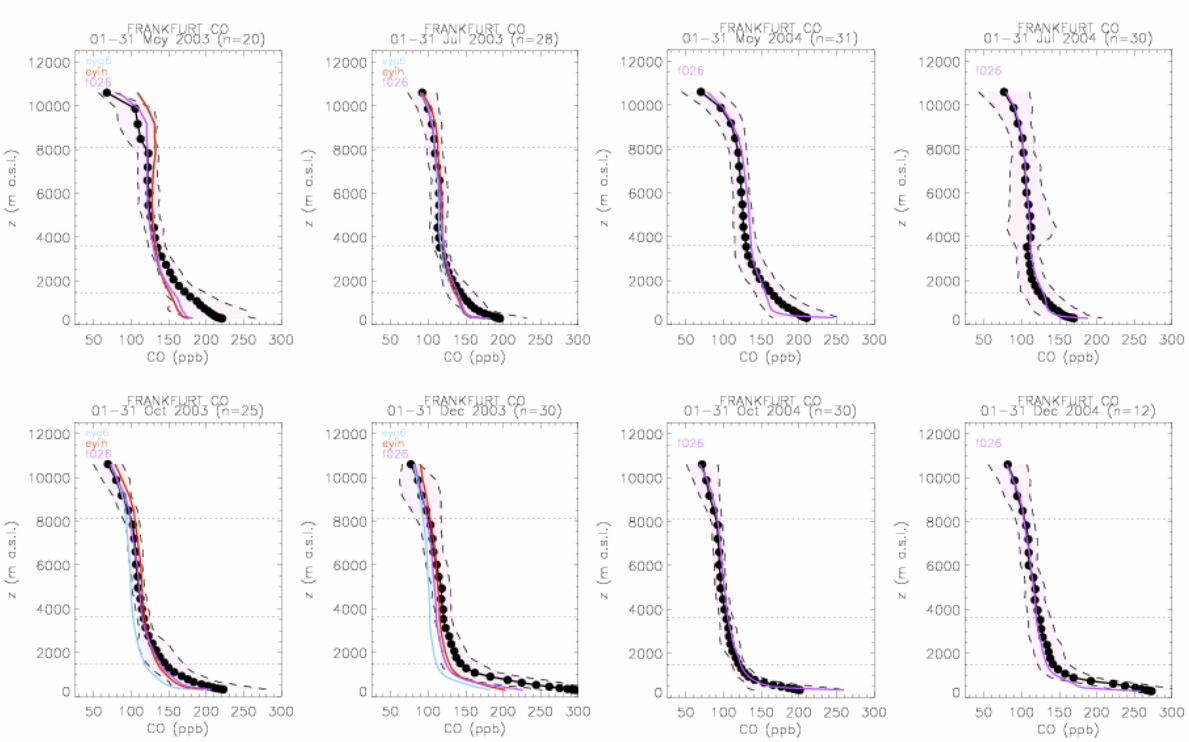


Figure 4-1: Vertical profiles of CO over Frankfurt from the coupled models compared to MOZAIC observations (2003 on the left and 2004 on the right).

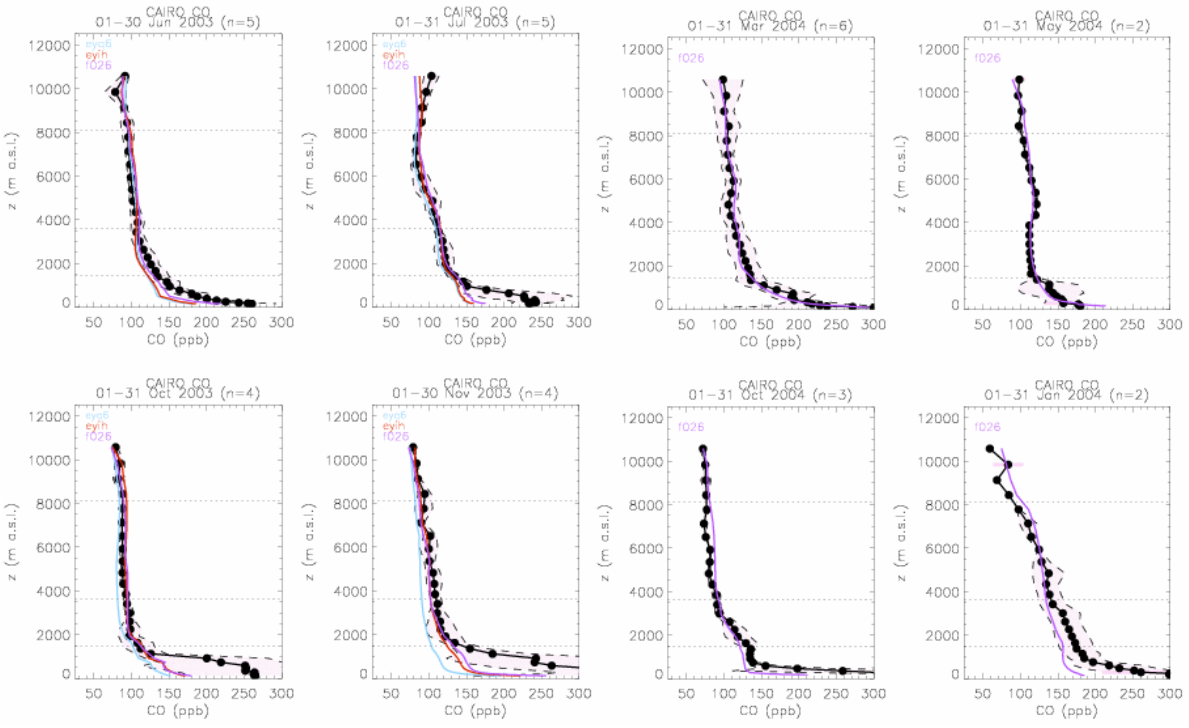


Figure 4-2: Vertical profiles of CO over Cairo from the coupled models compared to MOZAIC observations (2003 on the left and 2004 on the right).

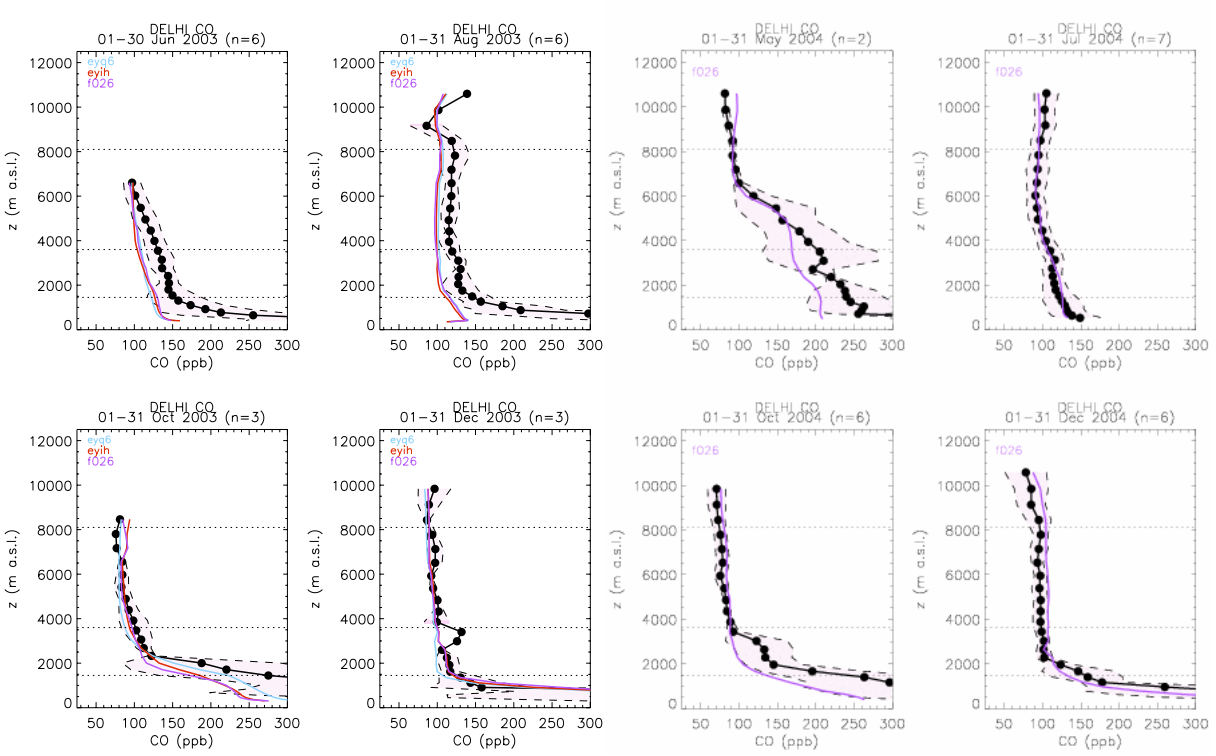


Figure 4-3: Vertical profiles of CO over Dehli from the coupled models compared to MOZAIC observations (2003 on the left and 2004 on the right).

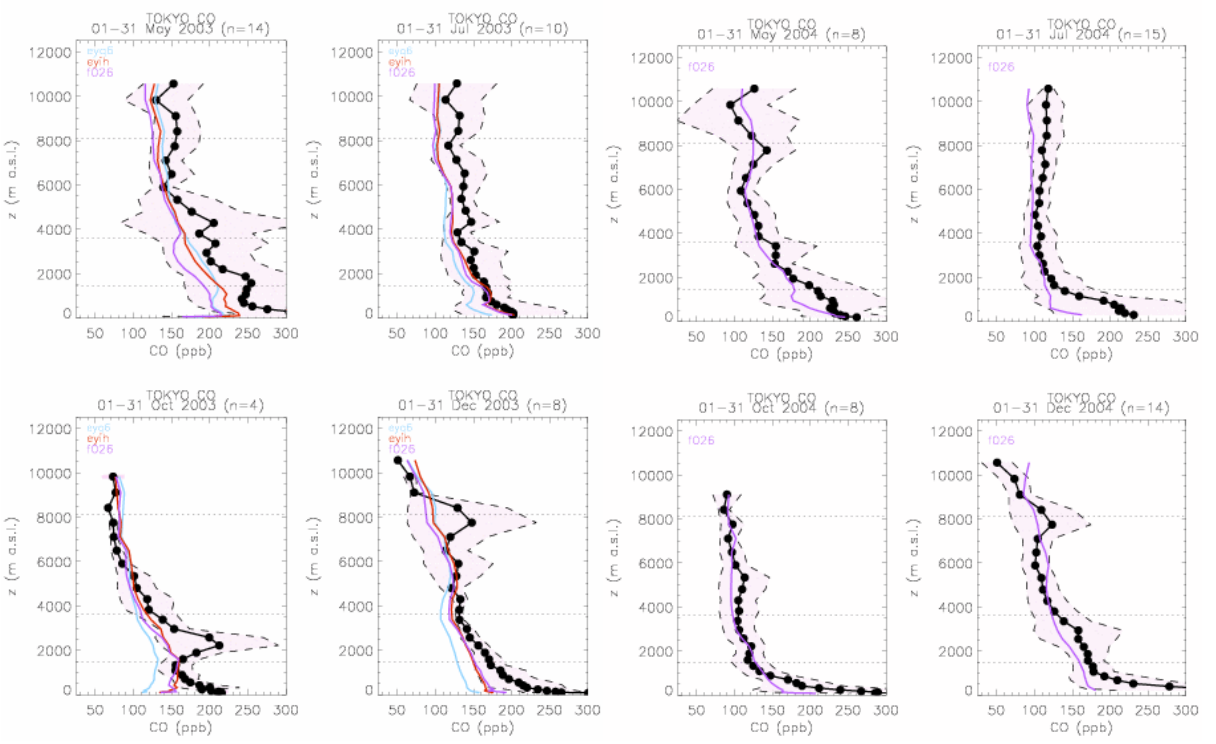


Figure 4-4: Vertical profiles of CO over Tokyo from the coupled models compared to MOZAIC observations (2003 on the left and 2004 on the right).

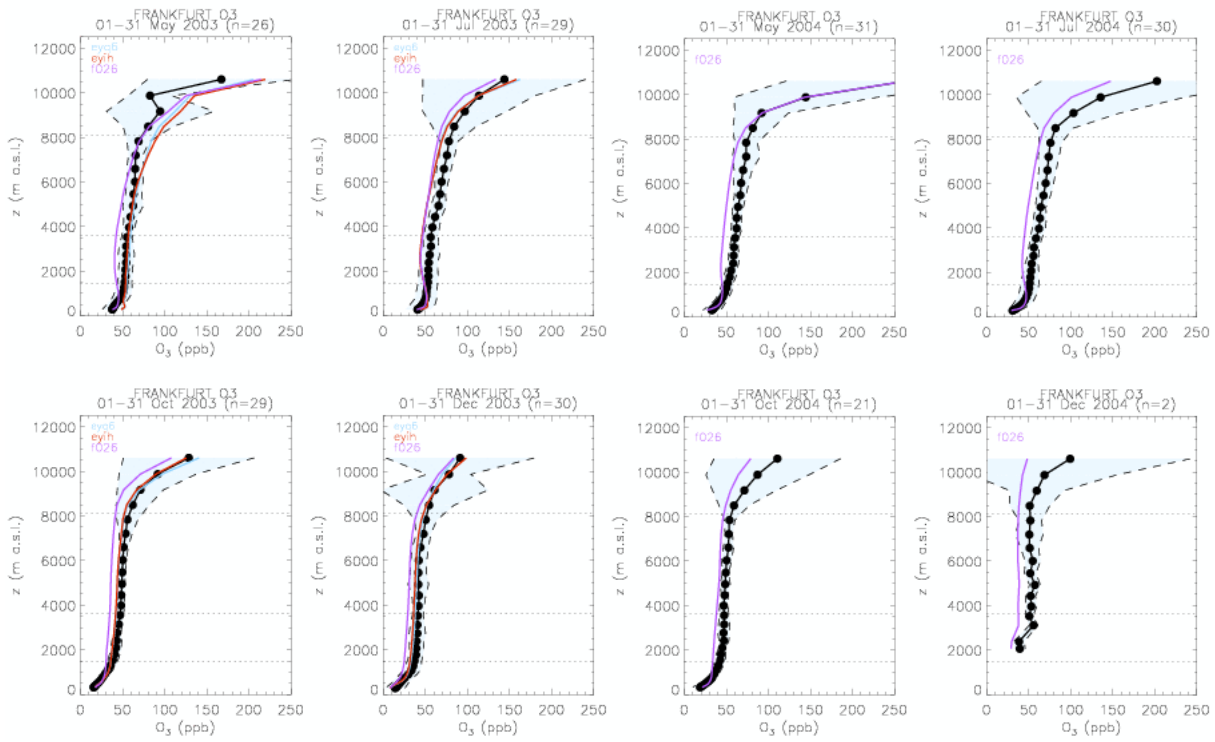


Figure 4-5: Vertical profiles of O3 over Frankfurt from the coupled models compared to MOZAIC observations (2003 on the left and 2004 on the right).

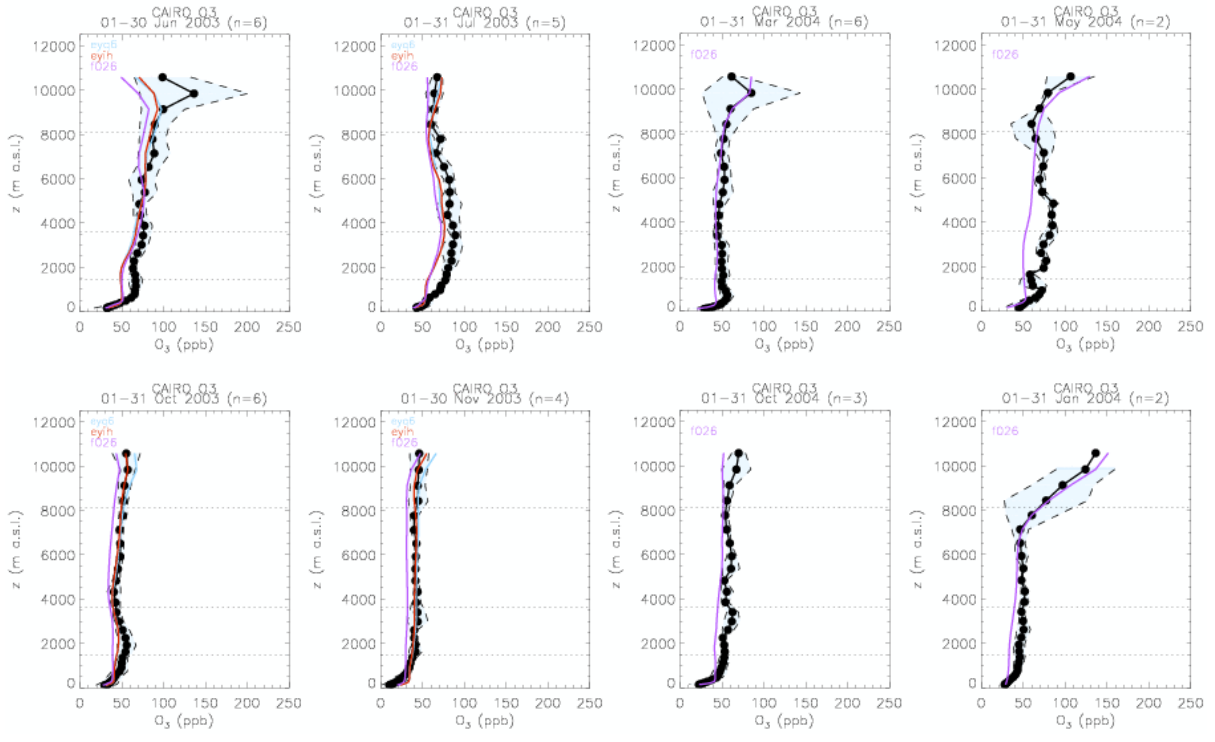


Figure 4-6: Vertical profiles of O3 over Cairo from the coupled models compared to MOZAIC observations (2003 on the left and 2004 on the right).

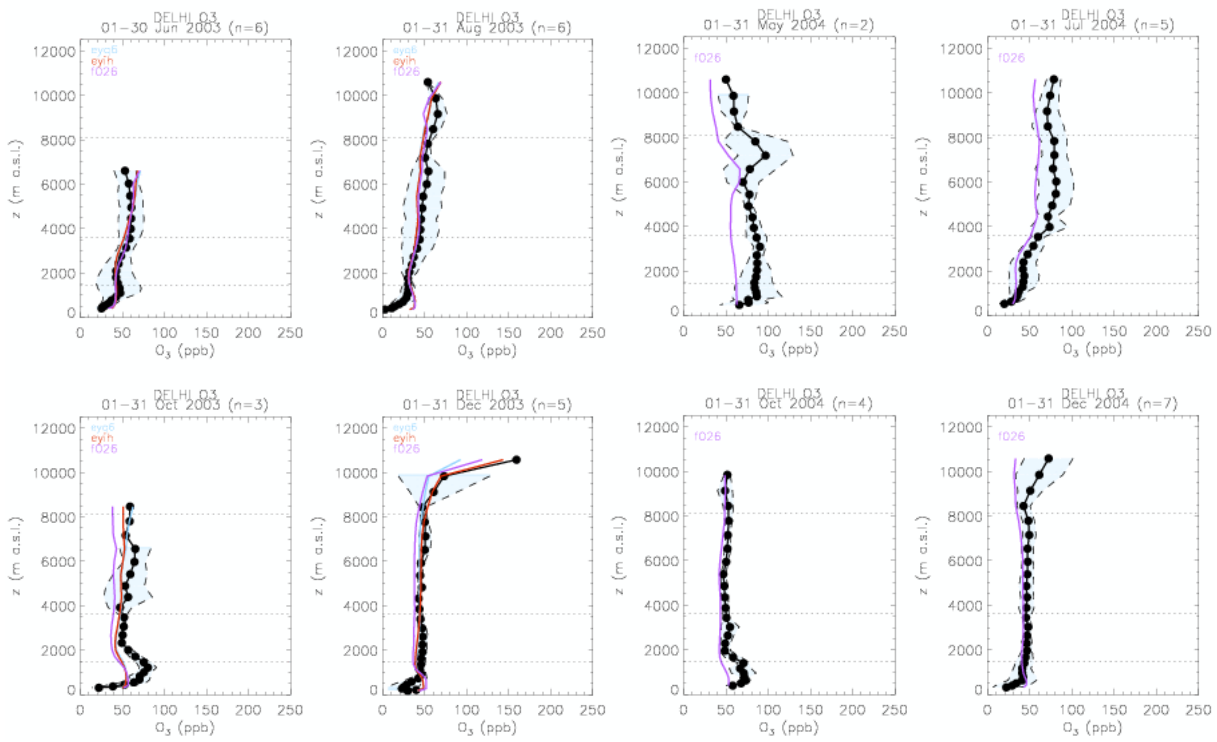


Figure 4-7: Vertical profiles of O3 over Dehli from the coupled models compared to MOZAIC observations (2003 on the left and 2004 on the right).

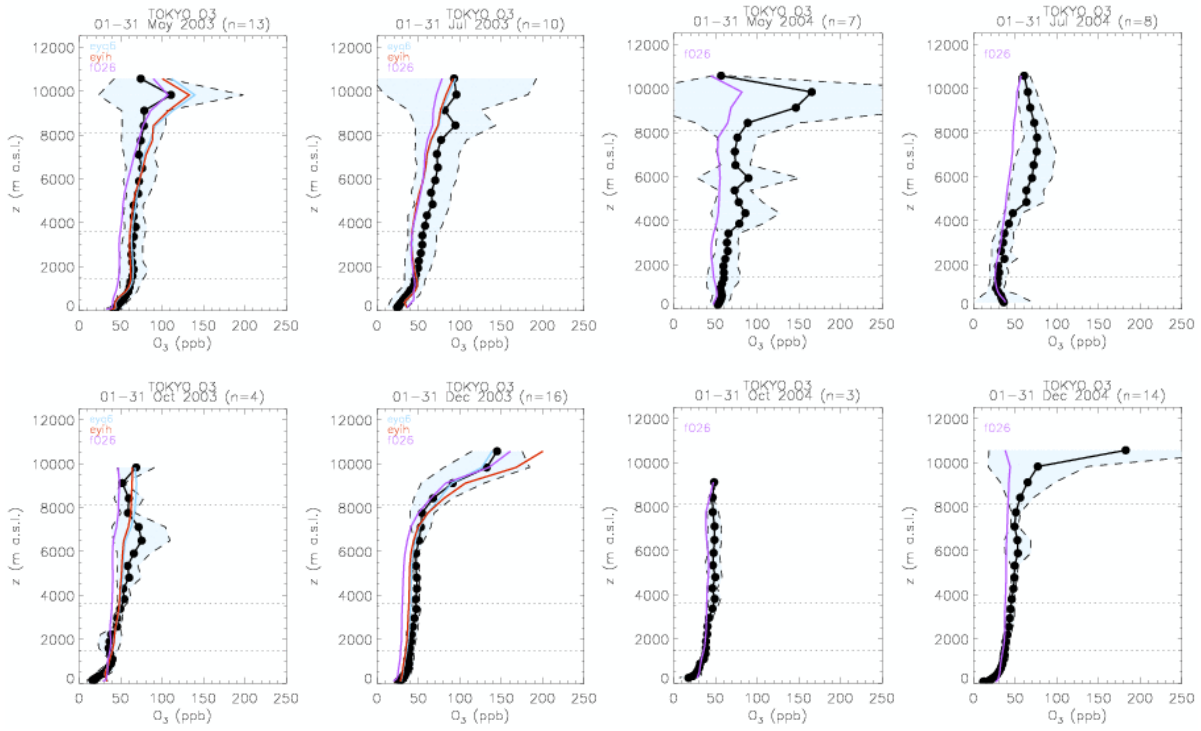


Figure 4-8: Vertical profiles of O3 over Tokyo from the coupled models compared to MOZAIC observations (2003 on the left and 2004 on the right).

- Over Europe and US (Paris, Frankfurt, Vienna, Munchen, NewYork, Chicago, Dallas, Atlanta):

In 2003, Europe experienced a summer heat wave leading to extreme O3 and CO concentrations in the lower and mid troposphere. This particular episode will be further discussed in chapter 7 and in *Ordonez et al.*, in preparation. In summer 2004, the intense boreal fires from Alaska and Canada were also responsible for enhanced O3 and CO layers in a large number of the profiles over the east coast of the US and over Europe. This is the subject of chapter 8 and *Elguindi et al.*, in preparation. Figure 4-1 above presents the vertical profiles of O3 and CO (MOZAIC and the 3 stand alone runs for each year) over Frankfurt for the months of May, July, October and December for the 2 years. Over this region, the 3 CTMs underestimate CO by around 5-40% throughout the troposphere. The highest biases are found at the surface and in the boundary layer, which is probably indicative of a problem in the emissions inventory. TM5_v7 underestimates CO the most, especially during the summer. It is worth noting that MOCAGE-V2 is quite close to the CO measurements over Europe in October and December. O3 biases are mostly in the +/- 20% range in the troposphere, with larger biases existing near the surface. Both runs from MOCAGE are the exception in that they systematically overestimate O3, especially in the lower troposphere. Both runs from MOZART and TM5 are very similar.

- Over East Asia (Beijing, Tokyo, Osaka):

Tokyo and Osaka in Japan, downwind of Northern China, characteristically have high CO concentrations up to 6 km altitude on average. Contrarily, the summer monsoon there is responsible for the observed ozone minimum in summer. Figure 4-3 presents vertical profiles similar to Fig 4-1 but for Tokyo. GEMS Models underestimate CO observations by 20-30% at the surface and in the boundary layer. CO is generally underestimated by the GEMS CTMs throughout the troposphere except by MOCAGE-V2 and MOZART_v3 in October above 4 km. MOCAGE-v2 does an excellent job in reproducing the long range transport event of CO in December 2003. TM5_v10 also performs very well in July 2004 above 2 km. For O3, we find that the CTMs overestimate O3 in the lower troposphere, especially during summer, and underestimate O3 in the free troposphere. Over this region, TM5_v10 is doing the best job. MOCAGE-v2 is overestimating O3, particularly in the free troposphere.
Beijing is probably the most polluted city at mid latitudes (MOZAIC aircraft measured CO concentrations up to 2 ppm close to the surface). Note there are no results for MOZART_2003 over Beijing. In this region (not shown), all the CTMs largely underestimate CO. However, it is worth noting that MOCAGE-v2 is doing very good in December 2003 throughout the troposphere.

- Over the Middle East (Cairo, AbuDhabi, Koweit, Riyadh, Tel Aviv):

This region is characterized by rather high CO concentrations in the boundary layer throughout the year, which is due to the high amount of local emissions. Cairo and Tel Aviv have quite high ozone concentrations in the lower and mid-troposphere during summer, probably due to transport from further north in Europe and an impact of the Asian Monsoon Anticyclone (*Kalabokas et al.*, 2008, *Lelieveld et al.*, 2002, 2009). Figure 4-5 presents vertical profiles similar to Fig 4-1 but for Cairo. The high levels of CO (250 ppb) up to 1 km are not reproduced by the 3 CTMs. However, both runs from MOZART are the closest to the observations. In the free troposphere, the three CTMs show reasonable agreement with the observations, except in May 2004 when MOZART_v10 and TM5_v10 underestimate by 20% from 2 to 10 km. For O3, the summer maximum in the mid-troposphere is well reproduced by MOZART_v3 and TM5_v7 in 2003, while MOCAGE-v2 overestimates O3 in the lower part of the troposphere.

- Over Low latitudes (Lagos, Caracas, Delhi) :

This region merges Lagos in West Africa, Caracas in tropical South America and Delhi in India,

and therefore covers diverse regions. The characteristics of Lagos are enhanced layers of O3 and CO in the lower troposphere during JJA due to the transport of biomass burning from southern central Africa (*Sauvage et al.*, 2005, 2007, *Mari et al.*, 2008). Caracas is characterized by very thick and rich CO layers throughout the year in the lower to mid-troposphere, while Delhi has high CO concentrations near the surface. Figure 4-7 presents vertical profiles similar to Fig 4.1 but for Delhi, the most representative tropical site within this MOZAIC data set. Contrary to the other regions, there is no general behavior with regard to the CO profiles from the CTMs. MOCAGE-v2 is underestimating in June, however in August, October and December it largely overestimates the CO. In 2004, MOZART_v10 and TM5_v10 underestimate CO in May but tend to overestimate it in October and December, especially in the free troposphere. The overall best agreement is given by MOZART_v3. For O3, all of the models are able to reproduce the observed low concentration (around 20 ppb or less) in the lower troposphere. In the free troposphere we find an overall good agreement by all models except MOCAGE-v2 which generally overestimates O3 throughout the troposphere. However, MOCAGE is much better in July 2004 with very good agreement above 4 km.

4.3 GEMS-GRG reanalysis

In this time frame, we ended up having 4 coupled simulations to evaluate. They are :

- IFS_eyih_2003 for IFS+MOZART for 2003 (May-Dec) (with assimilation)
- IFS_eyq6_2003 for IFS+MOZART for 2003 (May-Dec) (control run, = eyih without assimilation)
- IFS_f026_2003 for IFS+MOZART for 2003 (May-Dec) (with assimilation, no control run available)
- IFS_f026_2004 for IFS+MOZART for 2004 (Jan-Dec) (with assimilation, no control run available)

The data assimilation utilizes CO from MOPITT and O3 from SCIAMACHY, MIPAS, GOME, and SBUV. The control run is performed in order to evaluate the effect of the coupling with the IFS dynamical model compared to the additional effects of the assimilation. Unfortunately due to technical issues, nothing has been feasible in this time frame for the coupling of IFS with MOCAGE and TM5. One of the main issues we wish to address is to what extent the dynamical coupling and assimilation improve the simulations in comparison to the MOZAIC observations? Figures 4-9 to 4-16 show the vertical profiles of CO and O3 for 2003 (left panel) and 2004 (right panel) of MOZAIC observations (black dots with standard deviation as shaded area) and the coupled runs (color lines). The main characteristics by regions are described below.

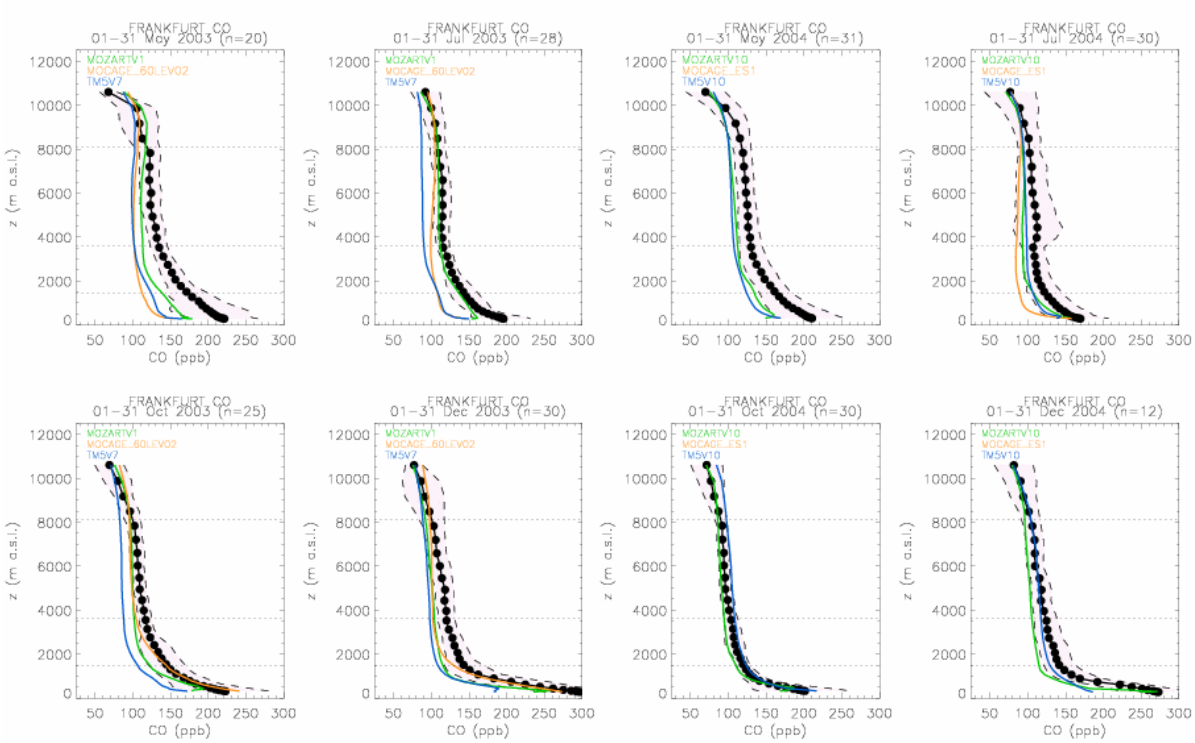


Figure 4-9: Vertical profiles of CO over Frankfurt from the CTMs compared to MOZAIC observations (2003 on the left and 2004 on the right).

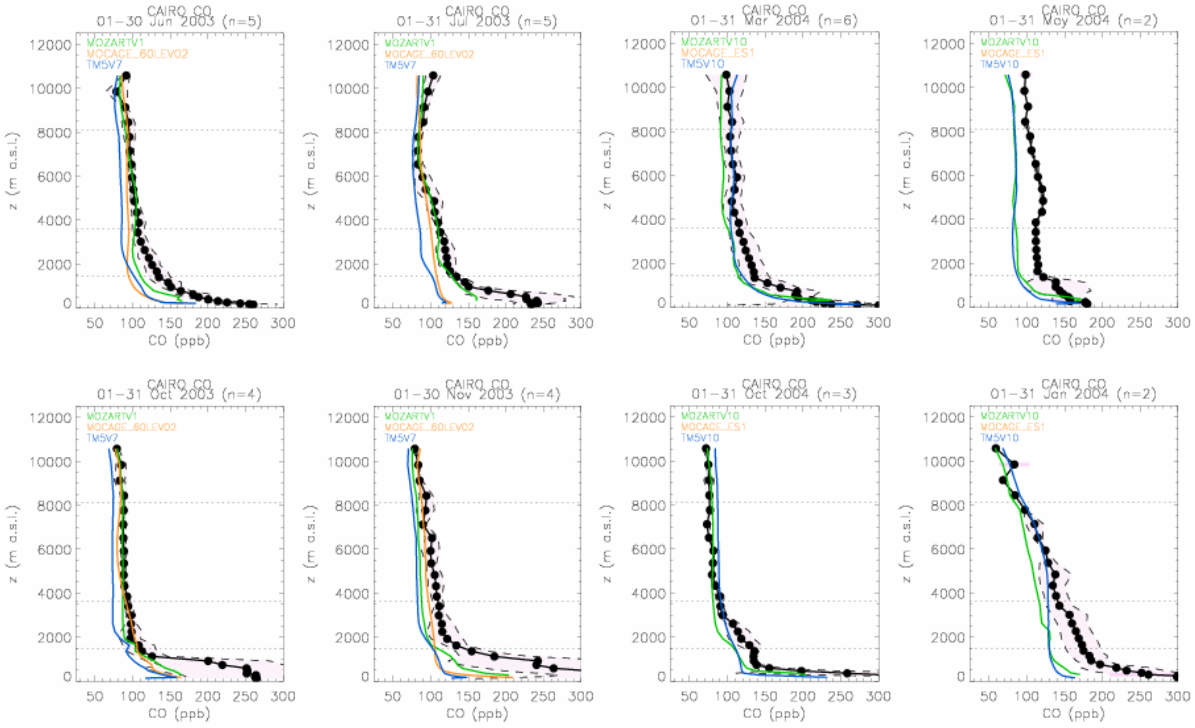


Figure 4-10: Vertical profiles of CO over Cairo from the CTMs compared to MOZAIC observations (2003 on the left and 2004 on the right).

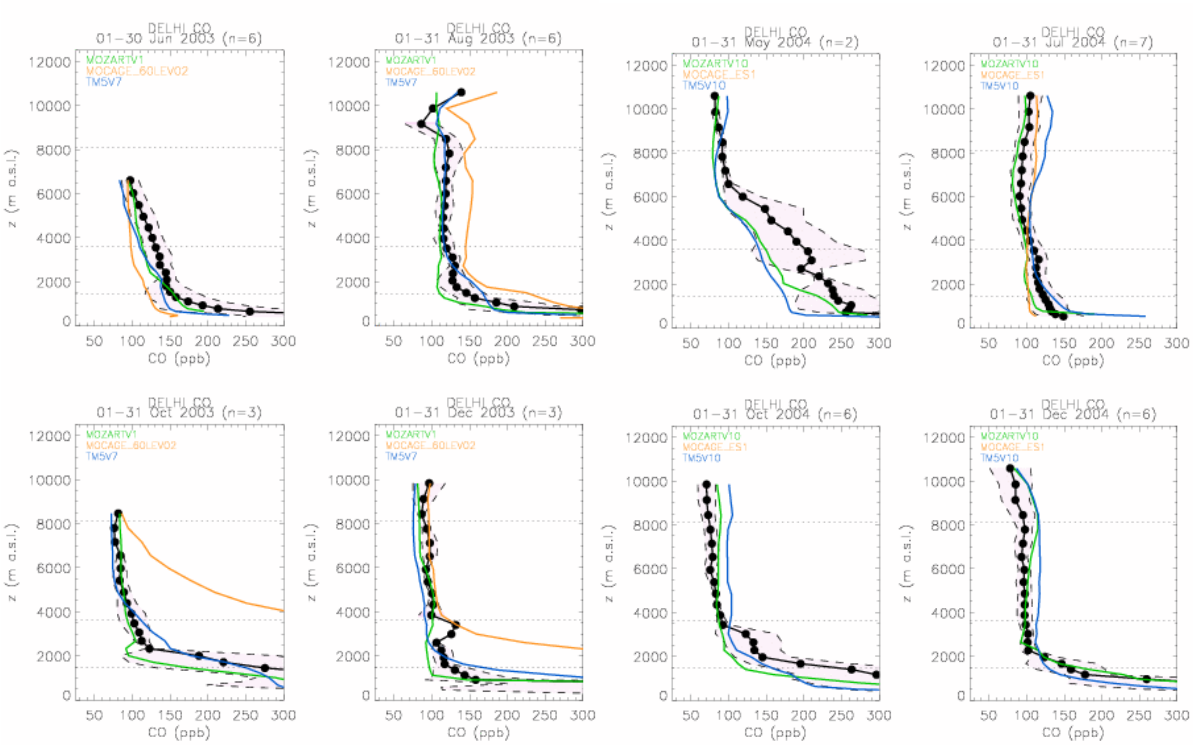


Figure 4-11: Vertical profiles of CO over Dehli from the CTMs compared to MOZAIC observations (2003 on the left and 2004 on the right).

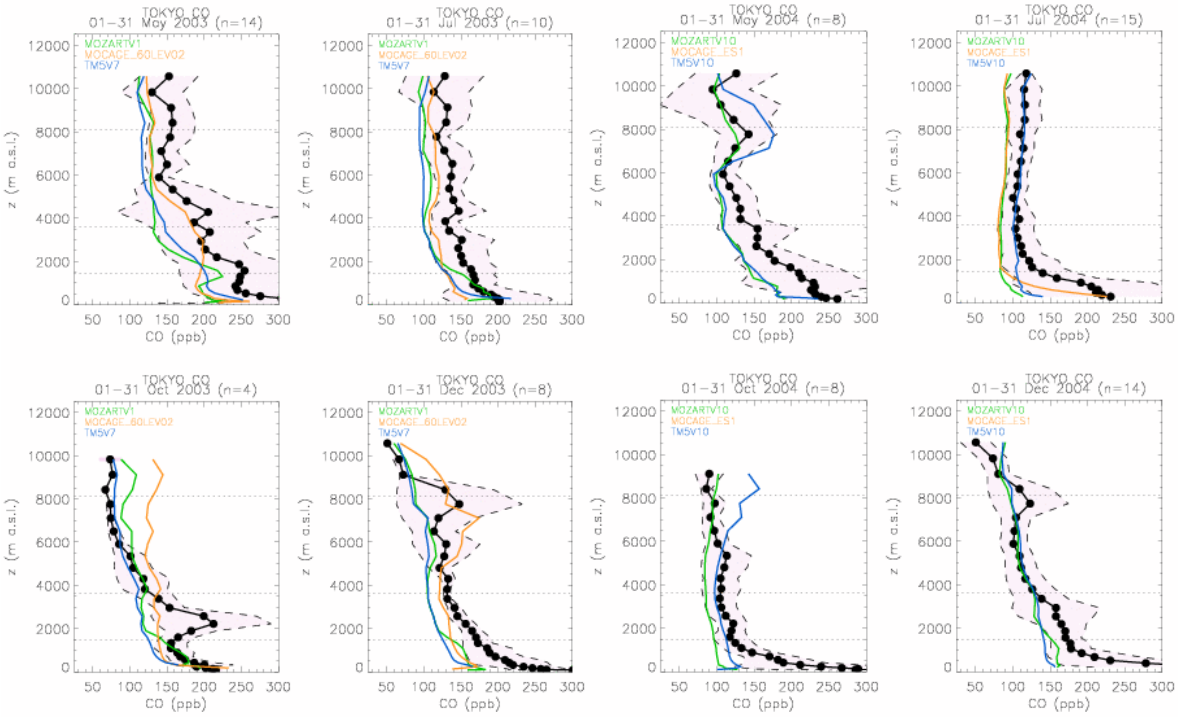


Figure 4-12: Vertical profiles of CO over Tokyo from the CTMs compared to MOZAIC observations (2003 on the left and 2004 on the right).

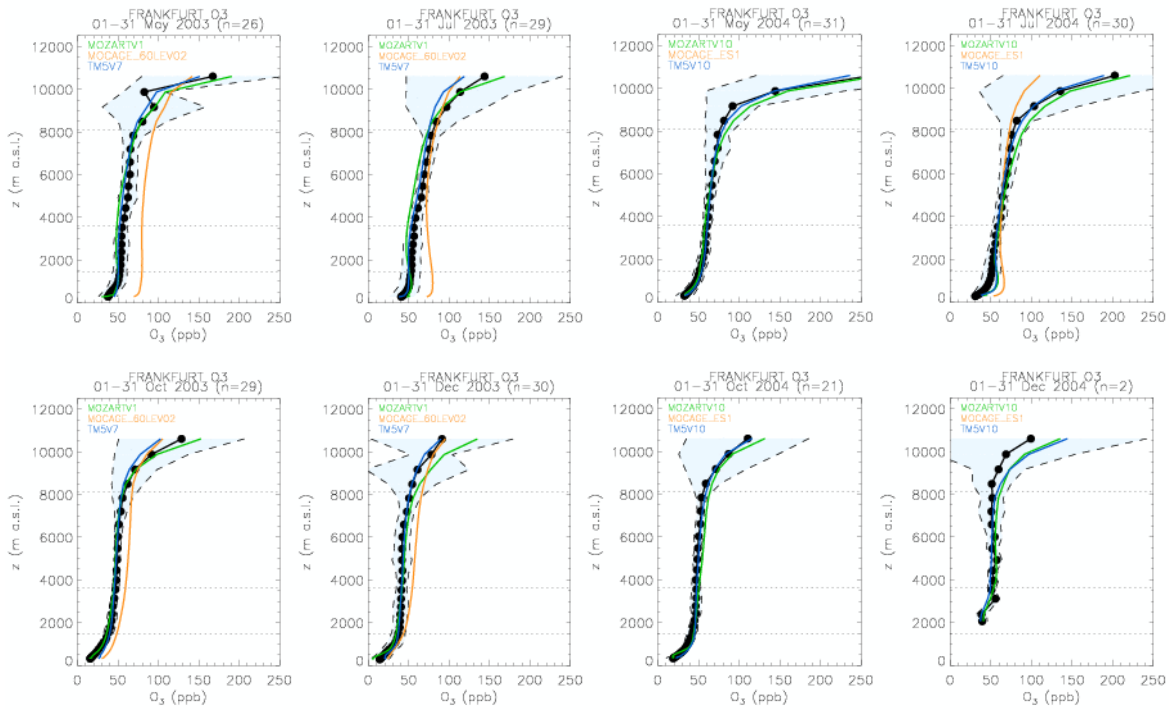


Figure 4-13: Vertical profiles of O3 over Frankfurt from the CTMs compared to MOZAIC observations (2003 on the left and 2004 on the right).

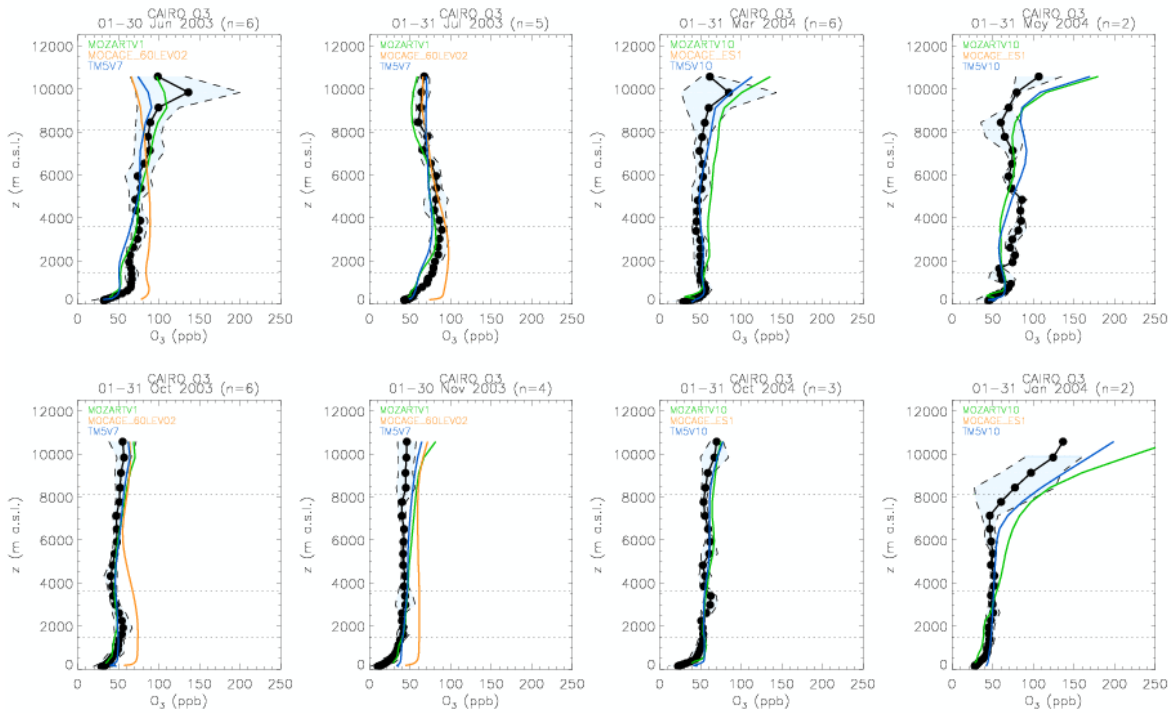


Figure 4-14: Vertical profiles of O3 over Cairo from the CTMs compared to MOZAIC observations (2003 on the left and 2004 on the right).

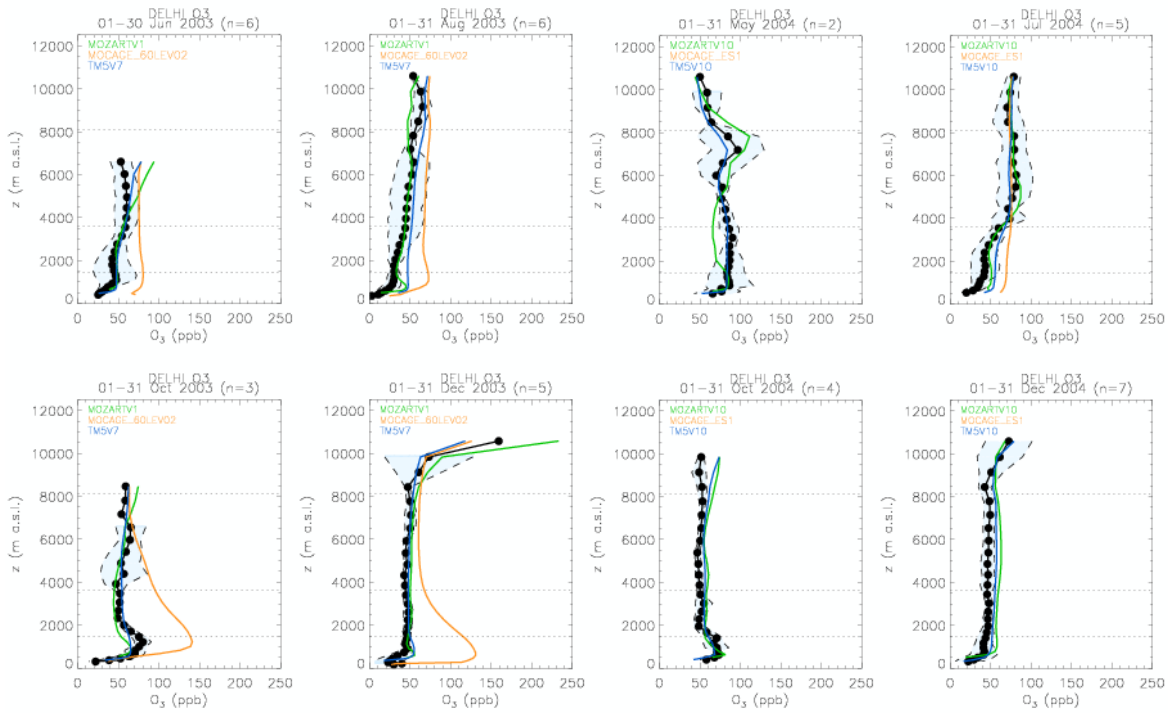


Figure 4-15: Vertical profiles of O3 over Dehli from the CTMs compared to MOZAIC observations (2003 on the left and 2004 on the right).

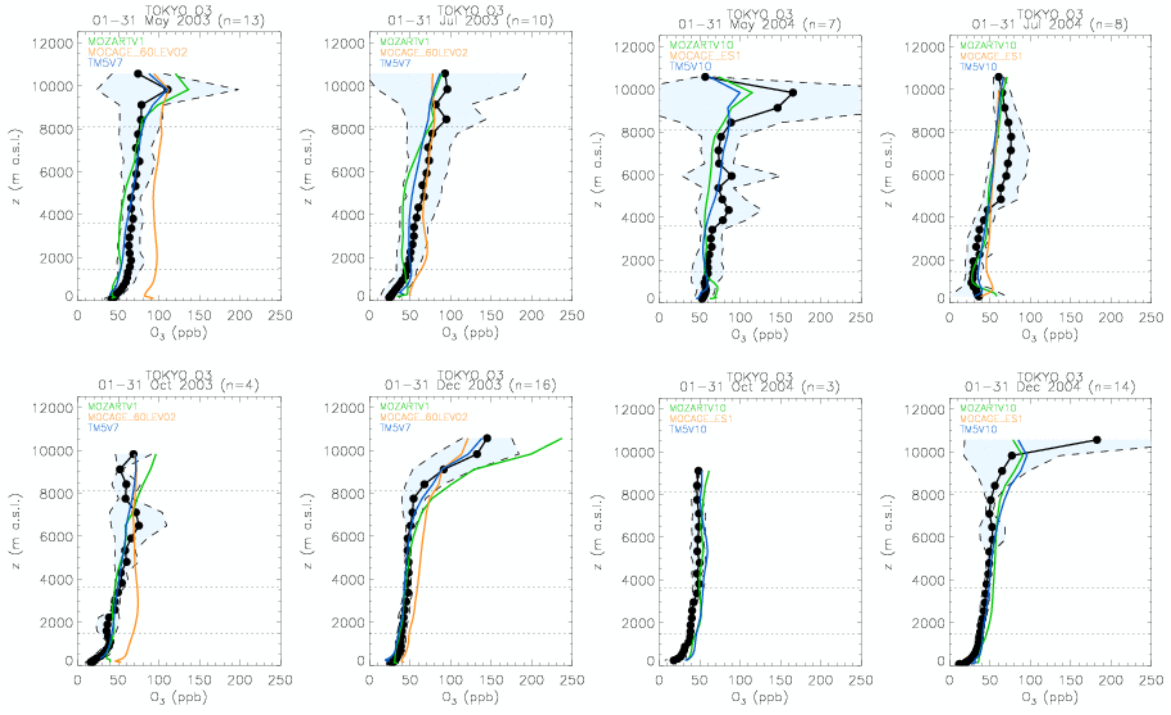


Figure 4-16: Vertical profiles of O3 over Tokyo from the CTMs compared to MOZAIC observations (2003 on the left and 2004 on the right).

- Over Europe and US:

Figure 4-9 shows a very good agreement between CO from MOZAIC and the four coupled runs. Except at the surface during the winter, model distributions are within the variability of the measurements. CO in-situ data are particularly well reproduced in 2004 by IFS_f026_2004. In 2003, higher biases are obtained with the control run eyq6 showing thus the improvement reached with assimilation. For O3 shown in Figure 4-10, the agreement between the models and observations is not as good. The second reanalysis IFS_f026_2003 performs worse than the first one in 2003 (eyih or eyq6) or than the standalone runs in 2003. O3 is underestimated by IFS_f026_2004 while it was in the range of the variability with previous standalone runs in 2004. However, we do notice an improvement with IFS_f026_2004 during summer in the surface and boundary layers in 2004 compared to the standalone runs.

- Over East Asia:

As for Frankfurt, CO profiles over Tokyo from the coupled runs better fit the MOZAIC data than the standalone runs. The better agreement is especially noticeable in 2004 and in the lower troposphere. The vertical gradients are better represented and the long-range transport events better reproduced by IFS_f026_2004. In 2003, the coupled runs with assimilation are better than eyq6 without assimilation. However, IFS_eyih_2003 is slightly better than IFS_f026_2003. The overall good agreement is most evident in the free and upper troposphere. This probably reveals the positive impact of the assimilation from MOPITT (higher sensitivity at this altitude). This fact might indicate that the assimilation process is compensating for the deficiencies found in the surface emissions inventory. For O3, the overall good agreement observed with the standalone runs (Fig 4-4) has not really been improved by the coupled runs. Indeed, IFS_f026_2004 shows highest (negative) biases than the (positive) biases from CTMs runs.

- Over Middle East:

As for the other regions, the coupled runs with assimilation give good results with CO profiles fitting the MOZAIC distribution within the range of variability. O3 however is not as encouraging. IFS_f026_2004 is systematically low biased in the free troposphere where CTMs were closer to the measurements. However, it is worth noting the good agreement between IFS_f026_2004 and MOZAIC O3 in the UT especially during March, May and January. The O3 vertical gradient is very well reproduced.

- Over Low latitudes:

In 2004, the CO profiles are quite well reproduced by IFS_f026_2004. The agreement is much better than with the standalone runs in 2004. The strong vertical gradient in the lower troposphere in October and December, as well as the CO enhancement between 3 and 6 km in May, are well captured by IFS_f026_2004. Contrarily, in 2003, the coupled runs do not provide better profiles than the MOZART_v1 standalone run, and are even worse as in August. For O3, as for the other regions, IFS_f026_2004 is actually worse than MOZART_v10. In 2003, the coupled runs provide an improvement only in August in the lower troposphere.

In conclusion, Figures 4-9 to 4-16 have highlighted the overall very good agreement between the second reanalysis (IFS_f026_2004) and CO data from MOZAIC in 2004. In 2003, surprisingly, this coupled run (IFS_f026_2003) is not as good as the first reanalysis (eyih). In many cases, MOZART_2003 standalone does as good (or better) as the IFS coupled systems for

the year 2003, especially near the surface (effect of MOPITT assimilation ?). For O3, improvement in latest IFS coupled run with assimilation (IFS_f026_2004) is not systematically reached. Often O3 is decreased (for the worse) in IFS_f026_2003 compared to eyih and in IFS_f026_2004 compared to MOZART_v10 or TM5_v10, especially in the free troposphere.

4.4 Scoring approach

In the following, we present the so-called "carpet plots" and several tables to synthesize the results of the evaluation and to propose a seasonal and annual score for each of the models. Figures 4-17a to 4-21b show seasonal matrices for CO (Figs *a) and O3 (Figs *b) of a combined score for the 10 different runs over the 19 selected airports. This finally gives an overview of the biases observed at locations not presented above as vertical profiles. The coding for the models runs is the following:

1: IFS_eyq6_2003, IFS+MOZART for 2003 (May-Dec) control run with no assimilation

2: IFS_eyih_2003, coupled run IFS+MOZART for 2003 (May-Dec) (with assimilation)

3: IFS_f026_2003 for 2003, new coupled run IFS+MOZART for 2003 (May-Dec) (with assimilation)

4: MOZART_v1, standalone run for 2003,

5: MOCAGE_v02, standalone run for 2003,

6: TM5_v7, standalone run for 2003,

7: IFS_f026_2004 for 2004, new coupled run IFS+MOZART for 2004 (Jan-Dec) (with assimilation)

8: MOZART_v10, standalone run for 2004,

9: MOCAGE_HTAP_ES1, standalone run for 2004 (JJAS only) and

10: TM5_v10, standalone run for 2004.

The seasonal score is defined as follows:

$$seasS = sign(MNMB) * (1 - sqrt[abs(mnmb)/2* (fge/2)]).$$

The score ranges from -1 to 1 where scores closer to +/-1 are better. The scores are plotted in Figures 4.17-4.22 to allow for a quick and visual comparison between the runs and between the regions, for the 4 seasons and the 4 altitude layers. Notice that the DJF season for 2003 is actually only represented by the month of December and the MAM season is only May 2003. As the MOCAGE_HTAP_ES1 run for 2004 is only between June and September, the 9th column is empty for DJF and MAM.

In DJF (Fig. 4-17a), it is worth noting how similar the different runs are over the different regions in the free troposphere. All models have low biases but this seems acceptable on seasonal average compared to the large underestimate observed in the surface layer. Surprisingly, the models tend to

overestimate CO in the UT especially over EU/US and Japan. Concerning O3 in the free troposphere, Fig 4-17b highlights very well the different behaviors of the coupled runs compared to the standalone ones for the 2 years. As noticed previously, coupled runs underestimate O3 while standalone runs tend to overestimate it. Among the coupled runs, it is also more evident that IFS_f026 (2003 and 2004) is doing "worse" than the first reanalysis. This is the general behavior over all the regions.

In MAM, CO in the free troposphere is slightly overestimated by the coupled runs, especially over EU/US and Middle East, while the standalone runs still underestimate it everywhere (Fig. 4-18a). In the surface layer, we notice an improvement over EU/US compared to DJF, especially for the coupled runs. For O3 (Fig. 4-18b) in the free troposphere, it is also evident that the second reanalysis (IFS_f026) has not improved the ozone distributions compared to the standalone runs (MOZART and TM5). Fig 4-18b highlights the large overestimation made by MOCAGE-v2 in the boundary layer and in the free troposphere, although it does perform much better in the UT.

In JJA (Fig. 4-19a), it is worth noting that a reasonable agreement in "surface and boundary layer" CO is found between the coupled runs (and MOZART) and the observations, especially over EU and northern US. This is actually the season when the biases are the lowest over this region. In the free troposphere, the second reanalysis is remarkable as it slightly overestimates CO almost everywhere in 2004 and especially in the Middle East in 2003. For O3 (Fig. 4-19b), MOCAGE is also noticeable for its general high overestimation of O3 from the surface to the free troposphere. However, we may see an improvement in the boundary layer and the free troposphere from 2003 to 2004. As previously mentioned, the second reanalysis for 2004 (IFS_f026_2004) underestimates O3 everywhere in the free troposphere while the standalone runs have slightly positive biases. However, the expected improvement from this IFS_f026_2004 run is reached at the surface and in the boundary layer. This is also valid during SON. It is worth noting here that several years have been run using the f026 model with MSL (Microwave Limb Sounder) O3 assimilation. Although there was not enough time to include a full analysis of these simulations in this document, preliminary analysis indicates that in many cases tropospheric O3 has been improved, particularly during the warmer months.

In SON (Fig. 4-20a), CO at the surface is largely underestimated everywhere by the 10 runs, with the lowest biases being observed over EU. Figure 4-20a clearly shows that TM5_v7 is particularly high biased during this season. However, notice the great improvement between this run and the following TM5_v10. The latter is slightly overestimating CO in the free troposphere. As already mentioned, MOCAGE-v2 is systematically overestimating O3 except in the UT, but we notice a real improvement from 2003 to 2004 (MOCAGE_HTAP_ES1).

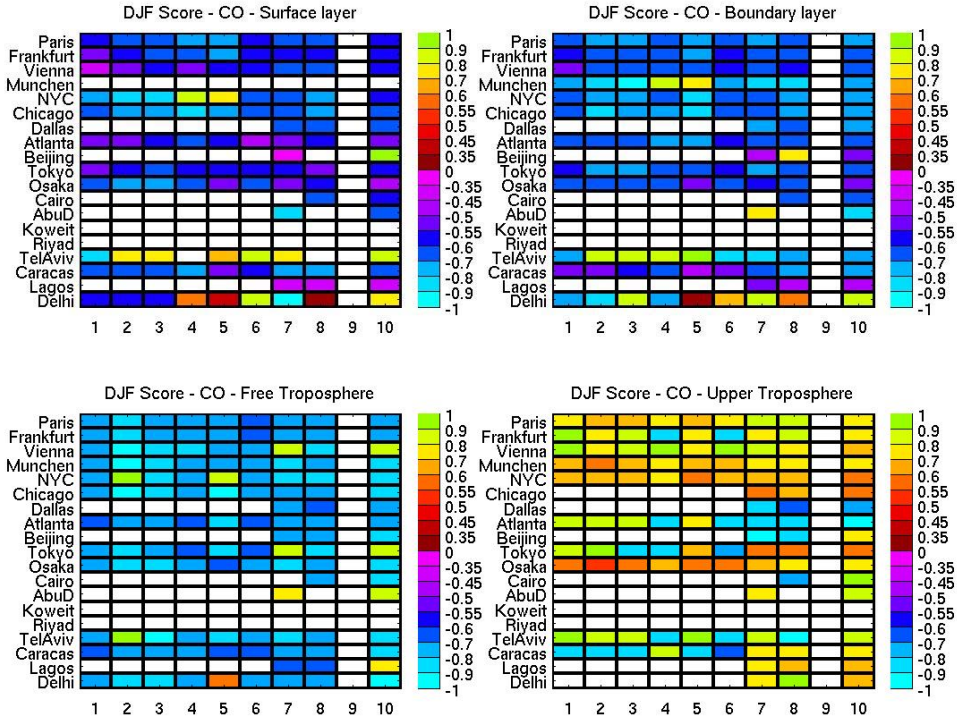


Figure 4-17a: Seasonal (DJF) score (=sign(MNMB)*(1-sr[(abs(mnmb)/2) * (fge/2)])) for CO, for the 10 model runs and the 19 selected airports, for 4 layers in the troposphere.

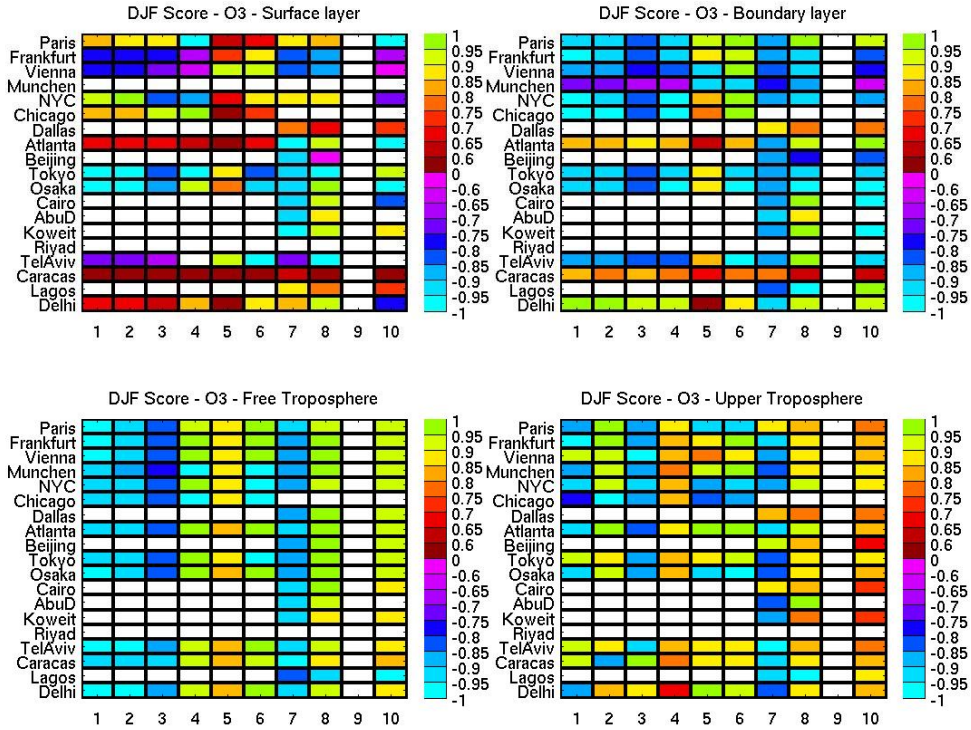


Figure 4-17b: Seasonal (DJF) score (=sign(MNMB)*(1-sqrt[(abs(mnmb)/2) * (fge/2)])) for O3, for the 10 model runs and the 19 selected airports, for 4 layers in the troposphere.

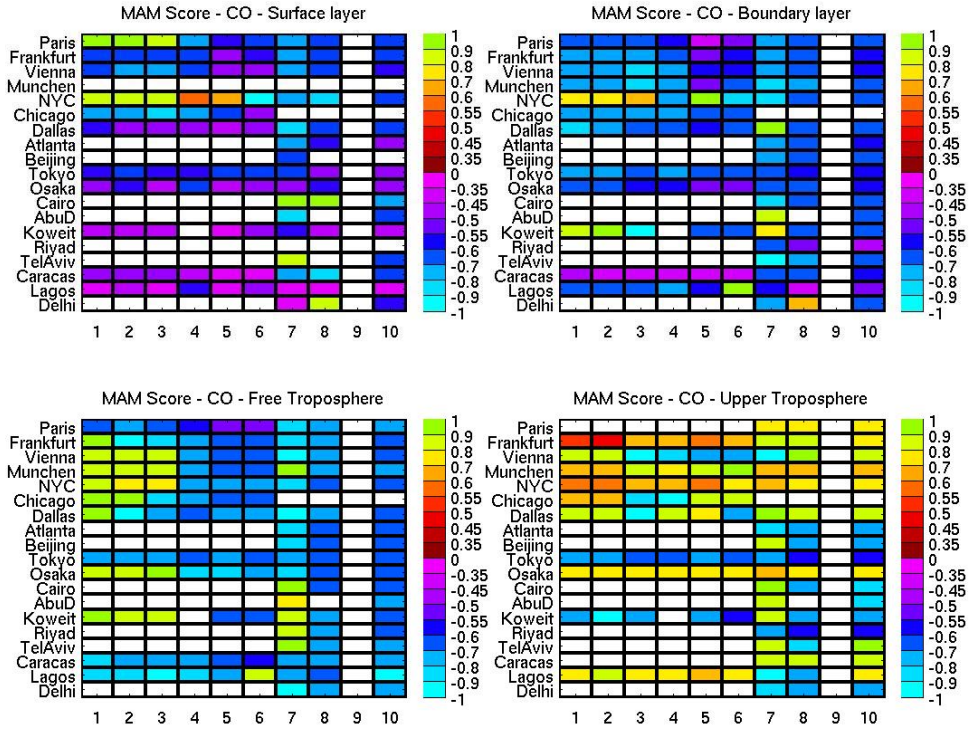


Figure 4-18a: Seasonal (MAM) score (=sign(MNMB)*(1-sqrt[(abs(mnmb)/2) * (fge/2)])) for CO, for the 10 model runs and the 19 selected airports, for 4 layers in the troposphere.

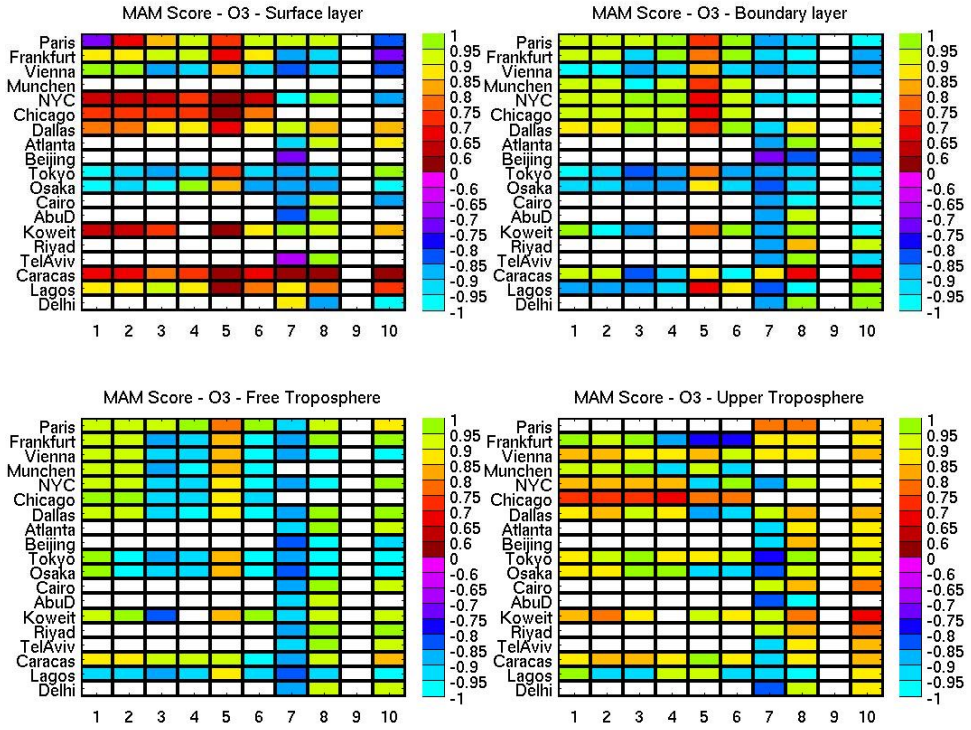


Figure 4-18b: Seasonal (MAM) score (=sign(MNMB)*(1-sqrt[(abs(mnmb)/2) * (fge/2)])) for O3, for the 10 model runs and the 19 selected airports, for 4 layers in the troposphere.

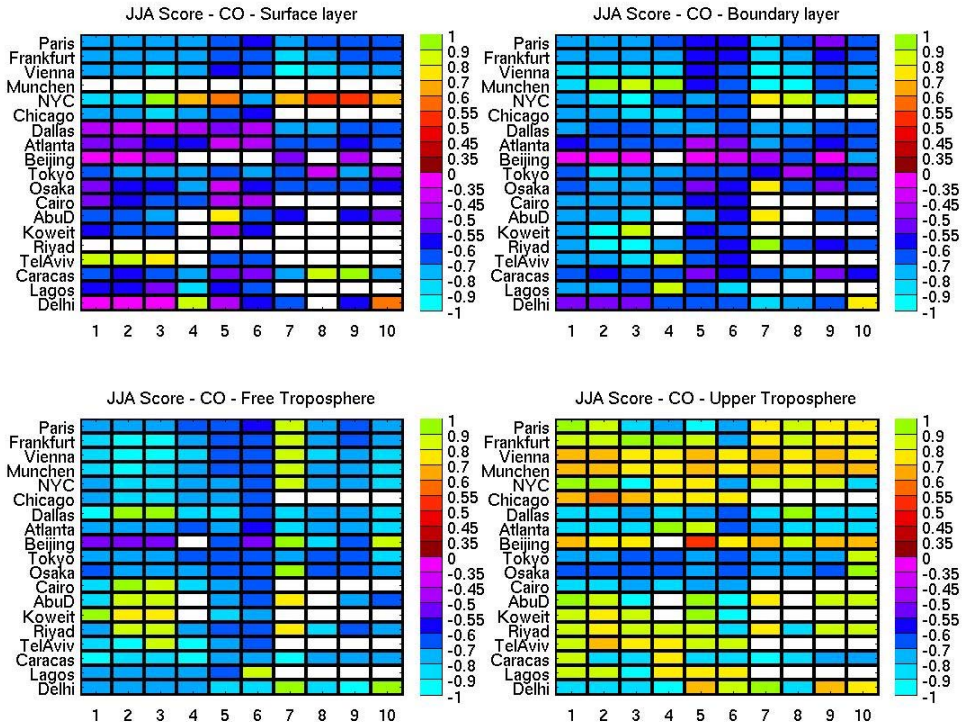


Figure 4-19a: Seasonal (JJA) score (=sign(MNMB)*(1 - sqrt[(abs(mnmb)/2) * (fge/2)])) for CO, for the 10 model runs and the 19 selected airports, for 4 layers in the troposphere.

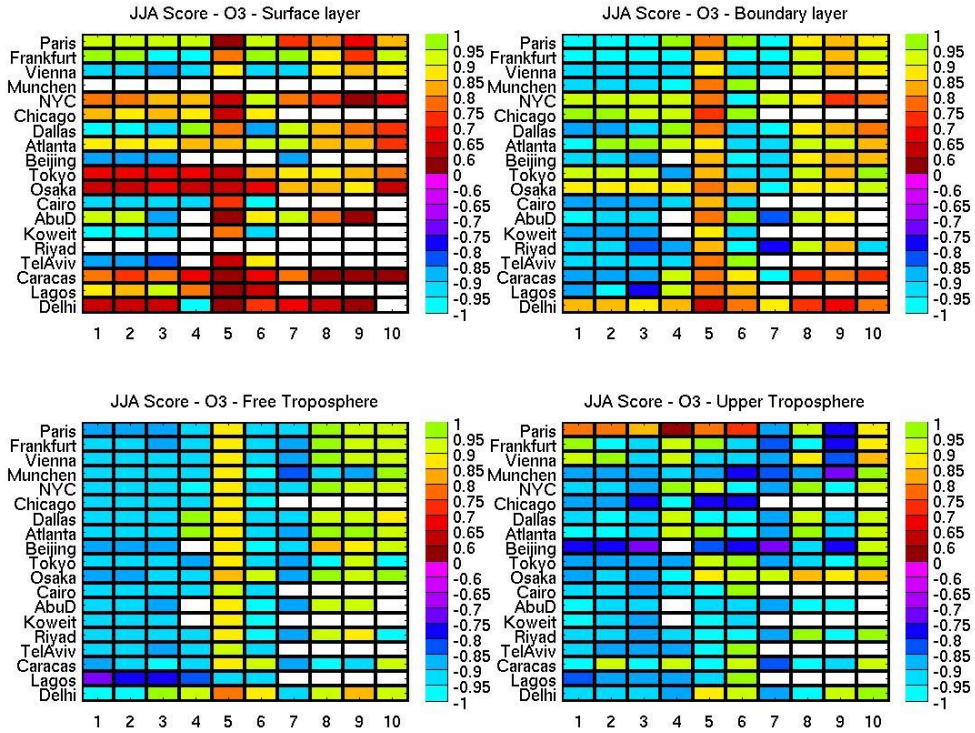


Figure 4-19b: Seasonal (JJA) score (=sign(MNMB)*(1-sqrt[(abs(mnmb)/2) * (fge/2)])) for O3, for the 10 model runs and the 19 selected airports, for 4 layers in the troposphere.

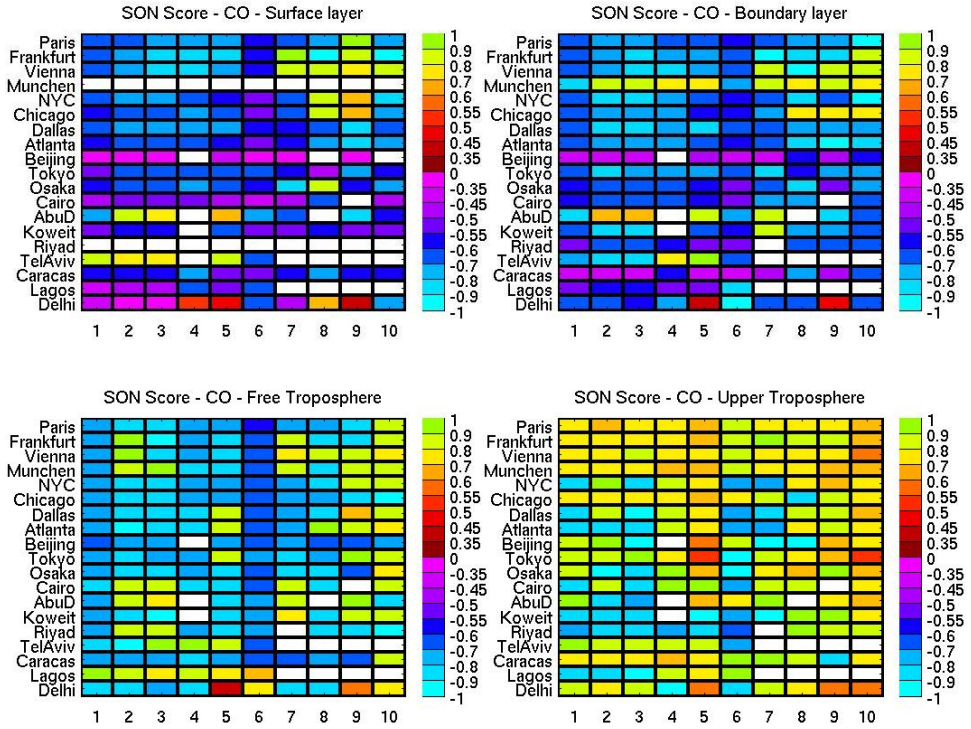


Figure 4-20a: Seasonal (SON) score (=sign(MNMB)*(1-sqrt[(abs(mnmb)/2) * (fge/2)])) for CO, for the 10 model runs and the 19 selected airports, for 4 layers in the troposphere.

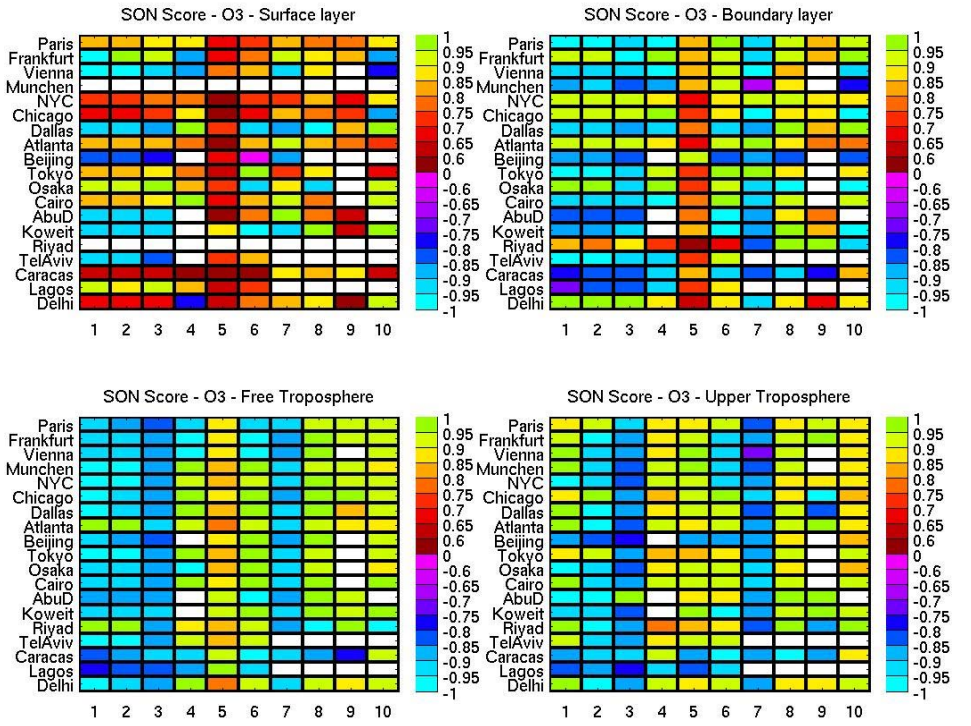


Figure 4-20b : Seasonal (SON) score (=sign(MNMB)*(1-sqrt[(abs(mnmb)/2) * (fge/2)])) for O3, for the 10 model runs and the 19 selected airports, for 4 layers in the troposphere.

To go further in the scoring approach, we have calculated what we call the annual score. This latter is a combination of the previous seasonal score (actually calculated for every month) with the correlation factor (r) between the monthly means from models and from MOZAIC :

$$S = \text{sign}(MNMB) * (1 - [\text{sqrt}[\text{abs}(mnmb/2)*(fge/2)]) * r$$

This allows us to take into account in the final score the ability of the models to reproduce the correct tropospheric seasonal cycle which is characteristic of particular physical and chemical processes responsible of the O3 and CO distributions. This would give additional information on the ability of the models to reproduce the correct or acceptable distributions for the "good" reasons. For example, are any of the model able to reproduce the O3 and CO maximum in the lower troposphere over West Africa in July and August? The answer is none, even though some of them present averaged concentrations close to the observations. Is this latter quality sufficient to give the model the better score? We are inclined to say no. A model with high biases but with a good seasonal cycle should be given a good score too. This is the reason why we set up this annual score presented in figures 4-21a and 4-21b for CO and O3 and for the 4 altitude layers. As previously, the score ranges between -1 and 1 in which scores closest to either +1 or -1 are best.

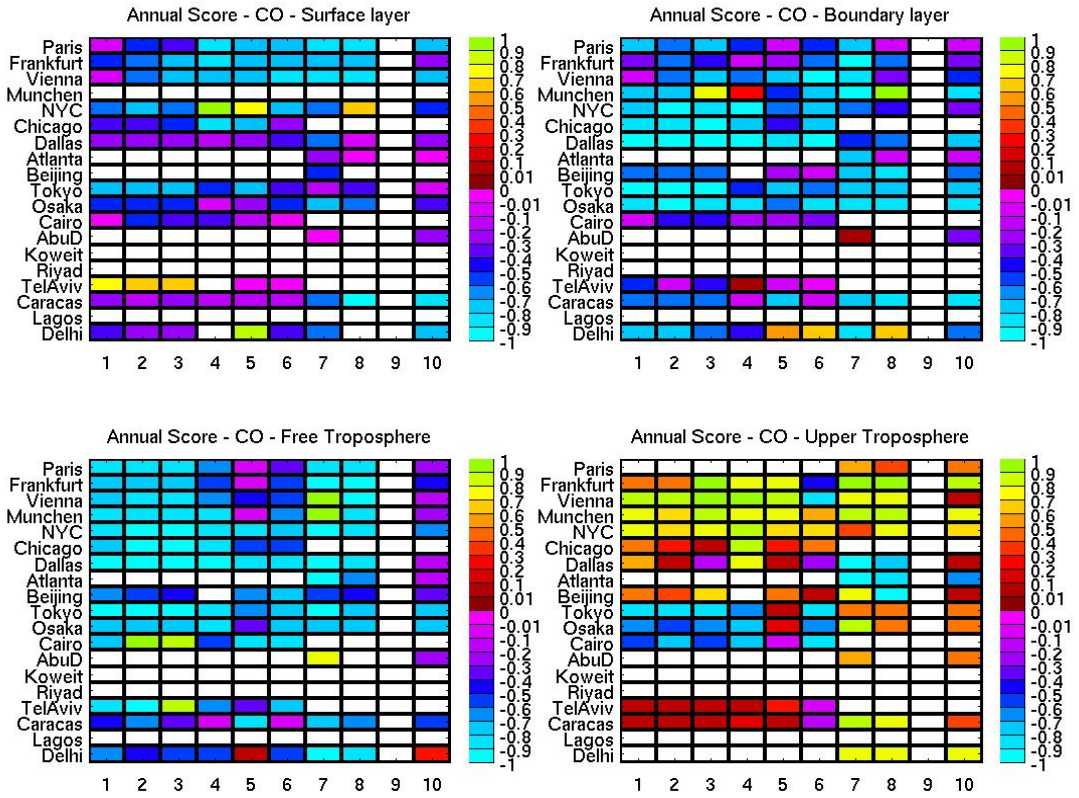


Figure 4-21a : Annual score (=sign(MNMB)*(1- sqrt[(abs(mnmb)/2) * (fge/2)] * r) for CO, for the 10 model runs and the 19 selected airports, for 4 layers in the troposphere.

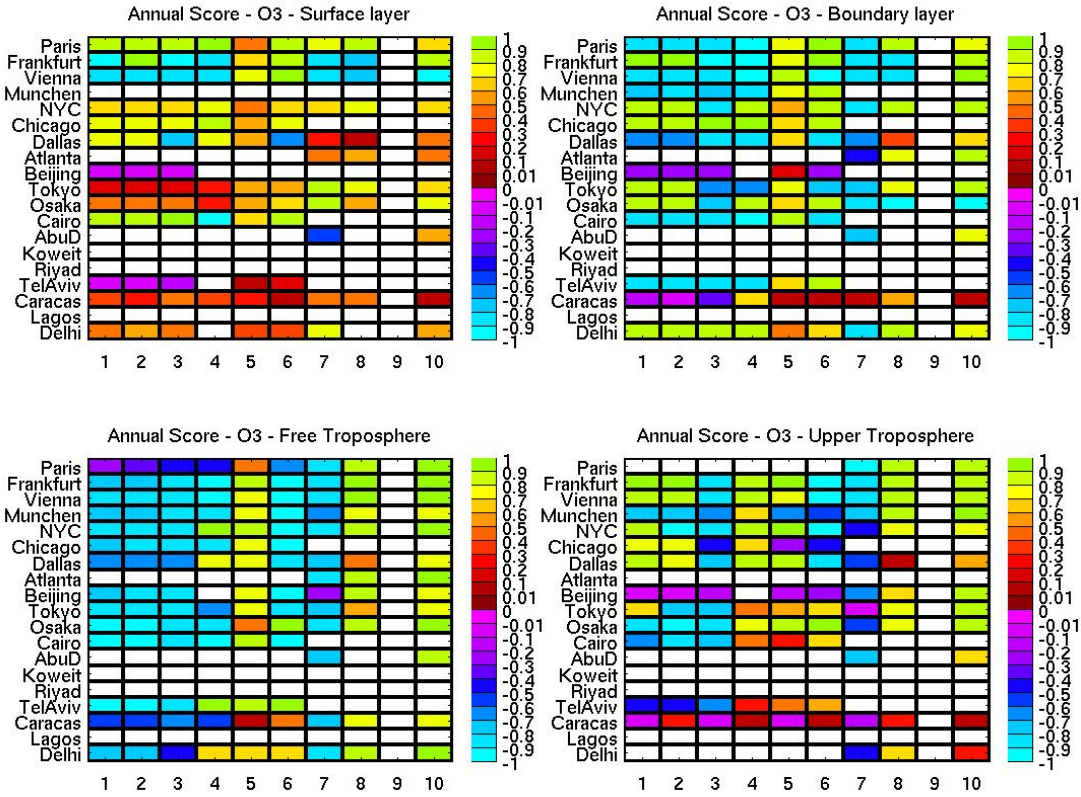


Figure 4-21b:.Annual score (=sign(MNMB)*(1- sqrt[(abs(mnmb)/2) * (fge/2)]) * r) for O3, for the 10 model runs and the 19 selected airports, for 4 layers in the troposphere.

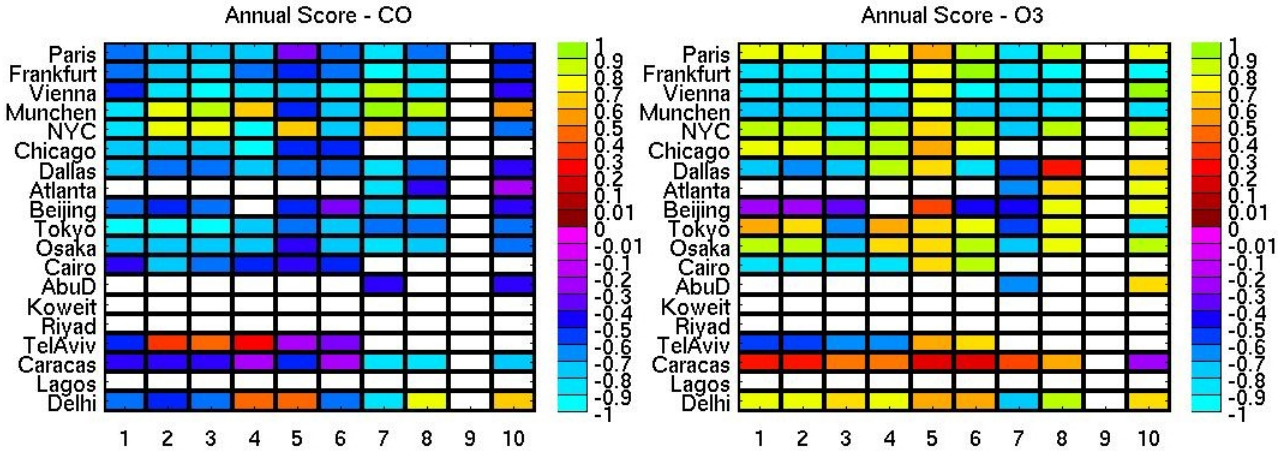


Figure 4-22 : Annual score (=sign(MNMB)*(1- sqrt[(abs(mnmb)/2) * (fge/2)] * r) for CO and O3, for the 10 model runs and the 19 selected airports, averaged throughout the troposphere.

Figure 4-21a summarized all the seasonal informations for the CO distributions. Generally, all model simulations underestimate CO from the surface layer to the free troposphere (purple to blueish colors), while a slight CO overestimation can be seen in the upper troposphere. Lowest scores, due to highest biases can be seen in cities such as Caracas, Delhi, Atlanta and Cairo illustrating a problem in local

emissions. Agreement in the free troposphere is very reasonable for all model runs except for TM5_v10 and MOCAGE_v02 (5th and 10th column). They both perform worse than the other models. It is worth noting how better the scores are for the second reanalysis in 2003 and 2004 (4th and 7th columns) compared to stand alone runs (4th,5th,6th, 8th and 10th columns). This is especially noticeable in the free troposphere highlighting thus the improvement gained through assimilation of MOPITT data.

Figure 4-21b summarized all the seasonal information for the O3 distributions. Surface O3 concentrations are generally overestimated by all the models (reddish to green colors), especially by MOCAGE_v02 (5th column is relatively more red than the other ones). However, an exception is seen over Beijing and Tel Aviv where the coupled runs for 2003 largely underestimate surface ozone. In the free troposphere, the overall agreement models/MOZAIC observations is much better than at the surface. From the boundary layer to the UT it is particularly interesting to see that in many cases, IFS_f026_2004 (7th column) shows negative biases while the standalone runs show positive one (8th and 10th columns). The improvement outside the surface layer is not straightforward then. The second reanalysis for 2003 (3rd column) is underestimating O3 in the UT. Globally, scores in absolute values are better for MOZART, TM5_v10 similar to the second reanalysis in 2003.

Finally, Figure 4-22 is an average of Figure 4-21 over the 4 layers. Therefore, for each model, we average this score over the regions by taking the absolute values (not to compensate between regions/layers high or low biased). Results are presented in table 4-2, to 4-7 below. The global line is always the average between the O3 and the CO lines. Notice there is no annual score for MOCAGE_HTAP_ES1 as the run was only for 4 months, so no significant correlation could be derived.

	1	2	3	4	5	6	7	8	9	10
CO	.58	.62	.63	.58	.45	.52	**.72**	.66	-	.45
O3	.69	.70	.71	**.74**	.59	.72	.69	**.74**	-	**.74**
global	.64	.66	.67	.66	.52	.62	**.70**	**.70**	-	.60

Table 4-2: Averaged annual scores over the 4 layers of the troposphere, and over the 19 selected airports. The 9th column is empty as it represents MOCAGE_HTAP_ES1 but too few months (JJAS only) have been run to compute the annual score.

Notice also that the evaluation for the year 2004 has been done for the full year while the year 2003 is for only the last 7 months (May to December). Besides, some airports have not been scored for all the runs. This is the case for example for Beijing. As the biases are very large over there, we provide the final score without this airport. Then, if we omit Beijing and if we consider the same 7 months for 2004, table 4-2 becomes table 4-3 below:

	1	2	3	4	5	6	7	8	9	10
CO	.58	.63	.64	.58	.45	.53	**.72**	.66	-	.47
O3	.72	.73	**.74**	**.74**	.61	**.74**	.72	.72	-	**.74**
global	.65	.68	.69	.66	.53	.64	**.72**	.69	-	.60

Table 4-3 : Same as Table 4-2 but without Beijing and for the same 7 months (May to December) for the 2 years.

Behind this global and averaged scores, we may find different specific behaviors. Tables below show the scores for the different regions, seasons and layers to better answer specific questions like "*Who is doing the best job in the surface layer, in the boundary layer, in the free troposphere or in the UT ? Who is doing the best job over Europe, US, Asia, tropical latitudes, middle east ? Who is doing the best job in summer, winter, spring or fall ?*".

	CO										O3										Global									
	1	2	3	4	5	6	7	8	9	10	1	2	3	4	5	6	7	8	9	10	1	2	3	4	5	6	7	8	9	10
EU/US	.62	.67	.71	.71	.50	.62	**.78**	.64	-	.41	.80	.80	.81	**.84**	.69	**.84**	.75	.75	-	.81	.71	.73	.76	**.78**	.60	.73	.77	.69	-	.61
E Asia	.66	.68	.69	.62	.45	.53	**.67**	.65	-	.48	.56	.56	.56	.58	.55	.67	.61	.76	-	**.80**	.61	.62	.62	.60	.50	.60	.64	**.71**	-	.64
Low lat	.48	.44	.43	.29	.46	.32	**.75**	**.75**	-	.62	.50	.52	.53	.58	.35	.35	.58	**.66**	-	.44	.49	.48	.48	.43	.40	.33	.66	**.71**	-	.53
Mid East	.41	**.53**	.50	.37	.25	.34	.35	-	-	.34	.71	.73	.74	.74	.58	**.75**	.68	-	-	.67	.56	**.63**	.62	.55	.42	.54	.51	-	-	.51

Table 4-4: Annual scores for the 4 different regions, averaged over the 4 different layers in the troposphere.

Globally, the absolute **highest scores are obtained over EU/US**. **For CO**, the second reanalysis (**IFS_f026_2004) is definitely the best.** For O3, MOZART and TM5 standalone runs perform really well. Lowest scores are obtained over the low latitudes and the middle East probably traducing the weak correlation factor due a poor representation of the seasonal cycle (as for example, all model runs miss the lower tropospheric summer maximum of ozone over Lagos).

	CO										O3										Global									
	1	2	3	4	5	6	7	8	9	10	1	2	3	4	5	6	7	8	9	10	1	2	3	4	5	6	7	8	9	10
Surface	.42	.53	.54	.58	.53	.47	**.59**	.56	-	.91	.56	.55	.56	.67	.51	.60	**.66**	.60	-	.60	.90	.49	.54	.62	.52	.53	**.62**	.58	-	.54
B Layer	.55	.62	.62	.43	.41	.50	**.64**	.52	-	.45	.76	.76	.76	**.84**	.61	.74	.71	.77	-	.75	.65	.69	.69	.63	.51	.62	**.67**	.64	-	.60
Free Tro	.78	.81	.78	.69	.47	.62	**.86**	.78	-	.37	.75	.77	.76	.79	.67	**.86**	.77	.79	-	**.86**	.76	.79	.77	.74	.57	.74	**.81**	.79	-	.62
Upp Tro	.52	.47	.54	.66	.36	.47	**.73**	.72	-	.46	.65	.67	.63	.61	.55	.65	.59	.66	-	**.69**	.59	.57	.59	.63	.46	.56	**.66**	.69	-	.57

Table 4-5: Annual score for the 4 different layers in the troposphere, averaged over the 19 selected airports.

For CO, IFS_f026_2004 is giving the best scores throughout the troposphere. Regarding **O3, IFS_f026_2004 gives the best results in the surface layer, while MOZART_V1 gives the best results in the boundary layer, TM5_V7 and V10 are the best in the free troposphere and TM5_V10 gives the best results in the upper troposphere.** This reinforces the finding that improvements due to the dynamical coupling and assimilation are more conclusive for CO than for O3. Globally, the best score is still from **IFS_f026_2004 throughout the troposphere.** Biases are globally higher in the surface layer, in the boundary and in the upper troposphere. This probably reveals the default of emission inventory linked to

the horizontal resolution of the models and the problem of representing the stratospheric influence. Overall, **highest scores are obtained in the free troposphere.**

EU/US only	CO										O3										Global									
	1	2	3	4	5	6	7	8	9	10	1	2	3	4	5	6	7	8	9	10	1	2	3	4	5	6	7	8	9	10
Winter	.88	.91	**.92**	**.92**	.91	.88	**.92**	.89	-	.90	.89	.89	.83	.87	.82	**.91**	.87	.89	-	.84	.89	**.90**	.88	.89	.86	.89	**.90**	.89	-	.87
Spring	.92	.92	.93	.90	.85	.87	**.94**	.87	-	.89	.87	.87	.89	.90	.77	.90	.89	**.92**	-	.89	.90	.89	.91	.91	.81	.89	**.92**	.90	-	.89
Summer	.91	.92	.93	.91	.87	.85	**.95**	.93	.88	.93	**.92**	**.92**	.90	**.92**	.80	**.92**	.90	.91	.84	.87	**.92**	**.92**	**.92**	**.92**	.84	.89	**.92**	**.92**	.86	.90
Fall	.91	.94	.95	.93	.91	.87	**.96**	.94	.93	.95	**.92**	**.92**	.88	.91	.80	.90	.88	.90	.86	.88	.91	**.93**	.91	.92	.85	.89	.92	.92	.89	.92

Table 4-6: Seasonal score for the 4 layers in the troposphere, averaged over the 6 airports in EU and US only. *EU/US only because these seasons refer to this part of the world with a marked seasonal cycle of O3 and CO (summer max for O3, spring max for CO).*

For CO during the winter, **IFS_f026_2003**, **MOZART_2003 and IFS_f026_2004** give the best results. The **IFS_f026_2004 also performs the best during the spring, summer and fall**. For O3, **TM5_V7** is doing the best in winter while MOZART V10 performs best in the spring. During the summer and fall**, the coupled models IFS_eyih_2003 and IFS_eyq6_2003 perform best.**

4.5 Summary

Based on table 4-3 annual scores, we are finally able to provide the following table ranking the models for O3, CO and both of them.

	CO	O3	global
1	**IFS_f026_2004 (0.72)**	**IFS_f026_2003 (0.74)**	**IFS_f026_2004 (0.72)**
2	MOZART_2004 (0.66)	**MOZART_2003 (0.74)**	MOZART_2004 (0.69)
3	IFS_f026_2003 (0.64)	**TM5_2003 (0.74)**	IFS_f026_2003 (0.69)
4	IFS_eyih_2003 (0.63)	**TM5_2004 (0.74)**	IFS_eyih_2003 (0.68)
5	MOZART_2003 (0.58)	IFS_eyih_2003 (0.73)	MOZART_2003 (0.66)
6	IFS_eyq6_2003 (0.58)	IFS_f026_2004 (0.72)	IFS_eyq6_2003 (0.65)
7	TM5_2003 (0.53)	MOZART_2004 (0.72)	TM5_2003 (0.64)
8	TM5_2004 (0.47)	IFS_eyq6_2003 (0.72)	TM5_2004 (0.60)
9	MOCAGE_2003 (0.45)	MOCAGE_2003 (0.61)	MOCAGE_2003 (0.53)

Table 4-7 : Ranking of the model runs following their annual score (numbers in brackets) for CO, O3 and average of both (global column).

Globally, for all the models, **scores are better for O3 than for CO**. It may reveal that the problem of having a good emission inventory is essential and is still not compensated by the dynamical coupling and assimilation to reproduce the correct CO distribution. On contrary, it is worth noting that the CTMs are able to correctly reproduce and integrate all the non linear processes leading the tropospheric ozone distribution (photochemical production/destruction, import from the stratosphere as main contributors).

O**verall, the best scores are obtained with IFS_f026_2004, over Europe&US** (the CTMs were designed for that) **in fall** (when photochemistry is not very active and CO is not at its seasonal maximum)**, in the free troposphere** (where the surface and stratosphere influences are lower)**.**

4.6 Conclusions and recommendations

The observations of CO and O_3 from the MOZAIC commercial aircraft measurements program were used to evaluate the simulations for the years 2003 and 2004. For each airport with sufficient data coverage, the simulation results were compared to the observed vertical profiles obtained during take off and landing. A MOZAIC vertical profile score was defined using the following formula:

$$S = \text{sign}(MNMB)*(1- [\text{sqrt}(MNMB/2 * FGE/2)]) * r$$

where MNMB and FGE are the modified normalized bias and the fractional gross error and r is the correlation factor between monthly means from models and observations. The score ranges between -1 an +1 in which scores closest to +/1 are the best. Combining the three informations as MNMB, FGE and r allows us to rank the ability of the models to reproduce the observations for the "good reasons". Keeping the sign of the MNMB is also interesting in the sense we clearly see that models generally tend to systematically underestimate or overestimate. This way, the developed scoring approach seems well appropriate to rank the different model runs. It has allowed a synthetic summary well representing our "by eye" experience based from thousand of profiles.

Model simulations generally underestimate CO especially in the surface layer, while 03 concentrations are generally overestimated, especially by MOCAGE_v02. In the free and upper troposphere, the agreement between models and MOZAIC observations is much better, especially with coupled runs including assimilation. Finally, it is worth noting how the scores have been improved by the method of dynamical coupling and assimilation set up in the frame of GEMS:

- The second reanalysis, IFS_f026_2004 has lead to a real and significant improvement in the CO distributions, especially noticeable in the boundary layer and the free troposphere.
- Improvement for O3 is not as clear. MOZART and TM5 standalone were already good. Both reanalysis (IFS_f026_2003 and IFS_f026_2004) tend to underestimate O3 in the free troposphere.

As perspectives of this work, we can expect some ways of further improvements through assimilation procedures. As already mentioned, preliminary analysis indicates that the f026 model has already been improved through MSL O3 assimilation. In the future, we also plan to further focus the evaluation with the MOZAIC data in the UTLS region.

4.7 References

1. Kalabokas P. D., A. volz-Thomas, J. Brioude, V. Thouret, J-P. Cammas and C. C. Repapis, Vertical ozone measurements in the troposphere over the Eastern Mediterranean and comparison with Central Europe, Atmos. Chem. Phys., 7, 3783-3790, 2007.
2. Law, K.S., Plantevin, P.-H., Shallcross, D.E., Pyle, J.A., Grouhel, C., Thouret, V., et Marenco, A., Evaluation of modelled O3 using MOZAIC data, J. Geophys. Res., 103, D19, 25,721- 25,737, 1998.
3. Law K., P-H. Plantevin, V. Thouret, A. Marenco, W.A.H. Asman, M. Lawrence, P. Crutzen, J-F. Muller, D. Hauglustaine, and M. Kanakidou, Comparison between Global Chemistry Transport Model Results and Measurement of Ozone and Water Vapour by Airbus In-Service Aircraft (MOZAIC) Data, J. Geophys. Res., 105, 1503-1525, 2000.
4. Lelieveld J. et al., Global air pollution crossroads over the Mediterranean, Science, 298, 794-799, 2002.
5. Lelieveld J. P. Hoor, P. Jockel, A. Pozzer, P. Hadjinicolaou, J.-P. Cammas, and S. Beirle,Severe ozone air pollution in the Persian Gulf region,Atmos. Chem. Phys., 9, 1393–1406, 2009
6. Marenco A. V. Thouret, P. Nédélec, H.G. Smit, M. Helten, D. Kley, F. Karcher, P. Simon, K. Law, J. Pyle, G. Poschmann, R. Von Wrede, C. Hume, and T. Cook, Measurement of ozone and water vapour by Airbus in-service aircraft : The MOZAIC airborne programme, an overview, J. Geophys. Res., 103, D19, 25,631-25,642, 1998.
7. Mari C.H., H., G. Cailley, L. Corre, M. Saunois, J. L. Attié, V. Thouret, and A. Stohl, Tracing biomass burning plumes from the Southern Hemisphere during the AMMA 2006 wet season experiment, Atmos. Chem. Phys., 8, 3951-3961, 2008.
8. Nedelec P., J.-P. Cammas, V. Thouret, G. Athier, J.-M. Cousin, C. Legrand, C. Abonnel, F. Lecoeur, G. Cayez, and C. Marizy An improved infrared carbon monoxide analyser for routine measurements aboard commercial airbus aircraft: Technical validation and first scientific results of the MOZAIC III programme , Atmos. Chem. Phys., 3, 1551-1564, 2003.
9. Sauvage B., V. Thouret, J- P. Cammas, F. Gheusi, G. Athier and P. Nédélec, Tropospheric ozone over Equatorial Africa: regional aspects from the MOZAIC data. Atmos. Chem. Phys., 5, 311-335, 2005.
10. Sauvage B., F. Gheusi , V. Thouret, Cammas J-P., Duron J., Escobar J., Mari C., Mascart P., and Pont V., Medium-range mid-tropospheric transport of ozone and precursors over Africa: two numerical case-studies in dry and wet seasons, Atmos. Chem. Phys., 7, 5357-5370, 2007.
11. Thouret, V., Marenco, A., Logan, J., Nédélec, P., et Grouhel, C., Comparisons of ozone measurements from the MOZAIC airborne programme and the ozone sounding network at eight locations, J. Geophys. Res., 103, D19, 25,695-25,720, 1998.

5. Evaluation with CMDL ground-based data

Contributors: Khokhar Fahim, Law Kathy, Granier Claire (SA-UPMC, Paris, France), and the modeller teams

5.1 Datasets and methodologies

5.1.1 Datasets

NOAA Climate Monitoring Diagnostics Laboratory Stations (CMDL) listed in the table below, have been selected for model evaluation in the framework of GRG's work packages. These stations are selected by adapting a criterion based on their strategic location with respect to forest and agricultural fires (FF), high latitude regions (PR), Industrial activities (IA) and continental outflows (CO), as indicated in Table 1.

Table 1: Details about CMDL stations and selection criterion

S. No	CMDL Station [name, (geo co-ordinates), altitude (m)]	Category	Details
1	ALT Alert, Canada (82 27 0 N - 62 31 12 W), 210	PR	- close to north pole
2	ICE Storhofdi, Iceland, (63 20 24 N 20 17 24 W), 127	PR	- close to north pole, - inter-continental transport
3	MHD Mace Head, Ireland (53 19 48 N 9 54 0 W), 25	CO	- continental outflow - centre of activities for different field measurements
4	ZEP Zeppeline Ny-Alesund, (78 54 0 N 11 52 48 E), 475	PR	- close to north pole
5	SPO South Pole, (89 58 48 S 24 48 0 W), 2810	PR	- close to south pole
6	HBA Halley Station,(75 34 48 S 26 30 0 W), 33	PR	- close to south pole
7	HUN Hegyhatsal, Hungary, (46 57 0 N 16 39 0 E), 344	FF/IA	- European Agricultural fires - industrial activities

8	BSC Black Sea, Romania, (44 10 12 N 28 40 48 E), 3	FF	- European Agricultural fires
9	BAL Baltic Sea, Poland, (55 25 12 N 17 4 12 E), 28	FF/CO	- European Agricultural fires - continental outflow
10	UUM Ulaan Uul, Mongolia, (44 27 0 N 111 6 0 E), 914	FF/ CO	- Siberian fires - Chinese transport
11	BRW Barrow, Alaska, USA, (71 19 12 N 156 36 0 W), 11	FF	- Siberian fires
12	TAP Tae-ahn Peninsula, Korea, (36 43 48 N 126 7 48 E), 20	CO/IA	- Chinese transport - industrial activities
13	CGO Cape Grim, Tasmania, Australia (40 40 48 S 144 40 48 E), 94		- continental outflow
14	KEY Key Biscayne, Florida USA, (25 40 12 N 80 12 0 W), 3	IA/CO	- continental outflow - industrial activities
15	NWR Niwot Ridge, Colorado, USA (40 3 0 N 105 34 48 W), 3475	IA/FF	- industrial activities - Forest Fires
16	CBA Cold Bay Alaska, USA (55.20°N, 162.72°W)	CO/PR	- continental outflow - close to north pole

CMDL data for different green house and trace gases is collected by using three different platforms i.e fixed sites (observatories, towers), ships and onboard aircrafts). Here we only present data collected from fixed sites enlisted in Table 1. Flasks filled with polluted air samples (mainly containing CO and H_2) from NOAA/CMDL stations are injected into a Gas chromatograph (GC) containing ambient air. The air samples are collected by two general methods: flushing and then pressurizing glass flasks with a pump, or opening a stopcock on an evacuated glass flask. During each sampling event, a pair of flasks is filled.

CO and H_2 are reacted with hot HgO bed to produce mercury (Hg). Hg is then determined photometrically. Measurements are reported in units of parts per billion (ppb) (Novelli et al., 1991, Novelli et al., 1998, Novelli et al., 2003).

Sampling frequencies are four times (approximately weekly) a month for the fixed sites on random days, thus, valid measurements on the sampling days are referred as CMDL "event "data for respective species.

Monthly means are produced for each site by first averaging all valid measurement results in the event file with a unique sample date and time. Values are then extracted at weekly intervals from a smooth curve (Thoning et al., 1989) fitted to the averaged data and these weekly values (referred as "global view" data) are averaged for each month to give the monthly means (referred as "monthly mean" data) recorded in the files. Three quality control flags (Rejection, Selection and Comments – for details see website http://www.esrl.noaa.gov/gmd/) are used. Flagged data are excluded from the curve fitting process.

Three types (event – sampling event, monthly mean and global view) of CMDL CO data sets are analysed and compared with the latest versions of CO simulation results from the GRG- chemistry transport models (GRG-CTMs: MOZART-V10, MOCAGE-V3 and TM5-V10) for these given stations.

5.1.2 Methodologies

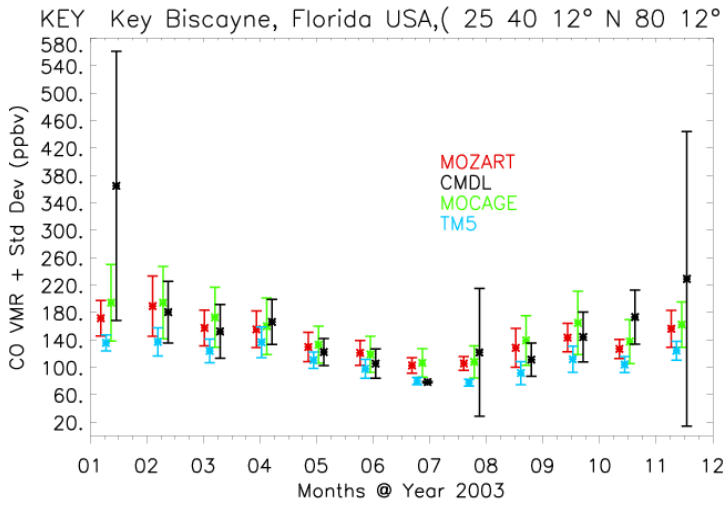
The scoring method used for the final evaluation of the performances of GRG-CTMs using satellite observation was agreed upon during GRG-subgroup meetings and afterwards on-line discussions, as described in Annex 7. In order to account for relatively large measurement uncertainties (scatter) in satellite observations, the use of the medians was introduced. So for surface observations (CMDL) at a single point, scatter is relatively less. However, due to limitation of CMDL data availability (only four measurements per month), the scoring index is based on mean of observation instead of median values. The scoring index based on mean values, is defined as:

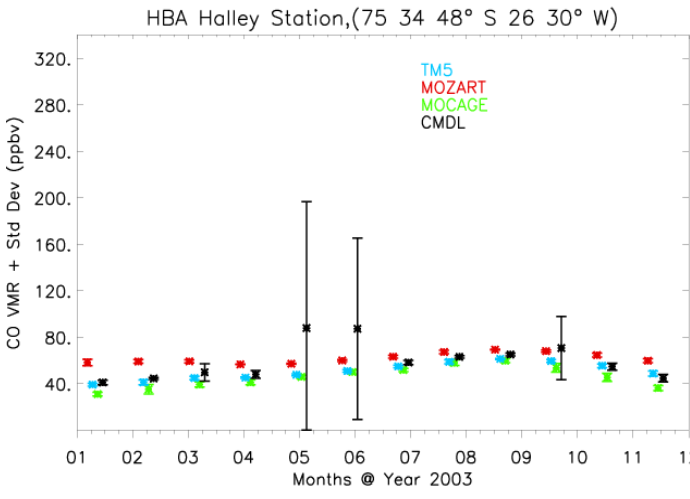
S=1- (abs (Mean (Mi-Oi) / 2* Mean(Oi))) Eq (1)

Where *mean (Mi − Oi)* is mean of difference of model and observation outputs for *i* values (i stands for Observations/model outputs per week in this case)

5.2 Offline simulations

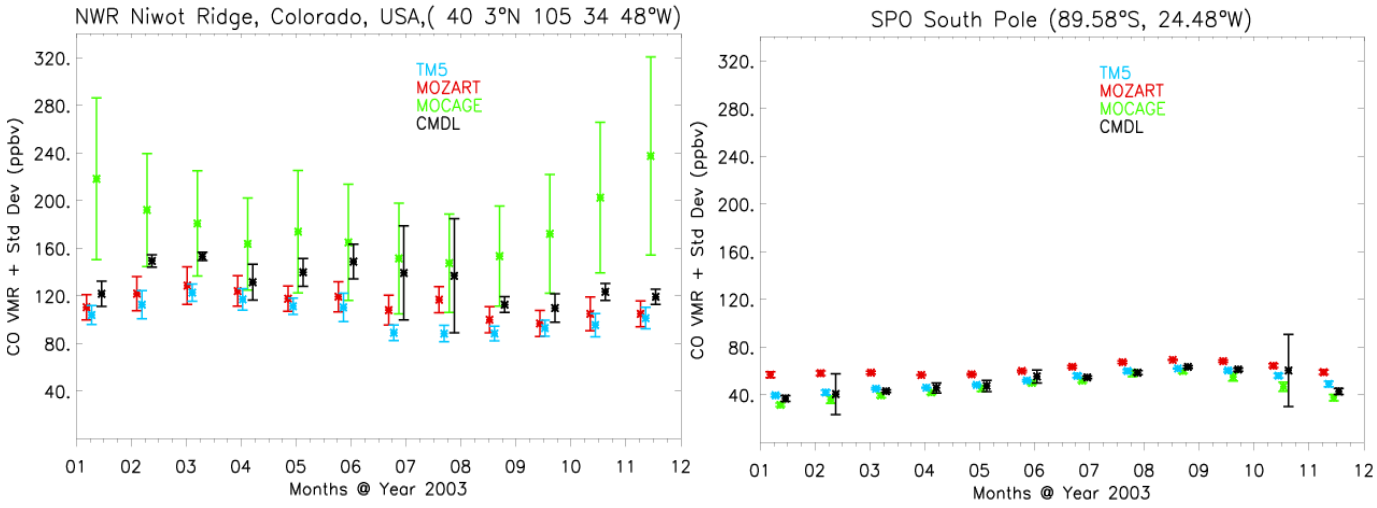
Sampling frequency of CMDL (event - sampling event) data is approximately four measurements per month. The figures shown below give a few examples of CMDL CO (event – sampling event) data compared with all of three GRG-CTMs simulation results. The model outputs are extracted for specific geo co-ordinates of a CMDL station (see Table 1). Standard deviation over the time period of one month is calculated and plotted as vertical bars in the figures for each GRG-CTMs.





KEY (Key Biscane) Station: Except for the months of January and December 2003, MOCAGE-V3 and MOZART-V10 show a fairly good agreement with CMDL data, while TM5-V10 is slightly underestimated. CMDL data exhibit larger variability during January and December months.

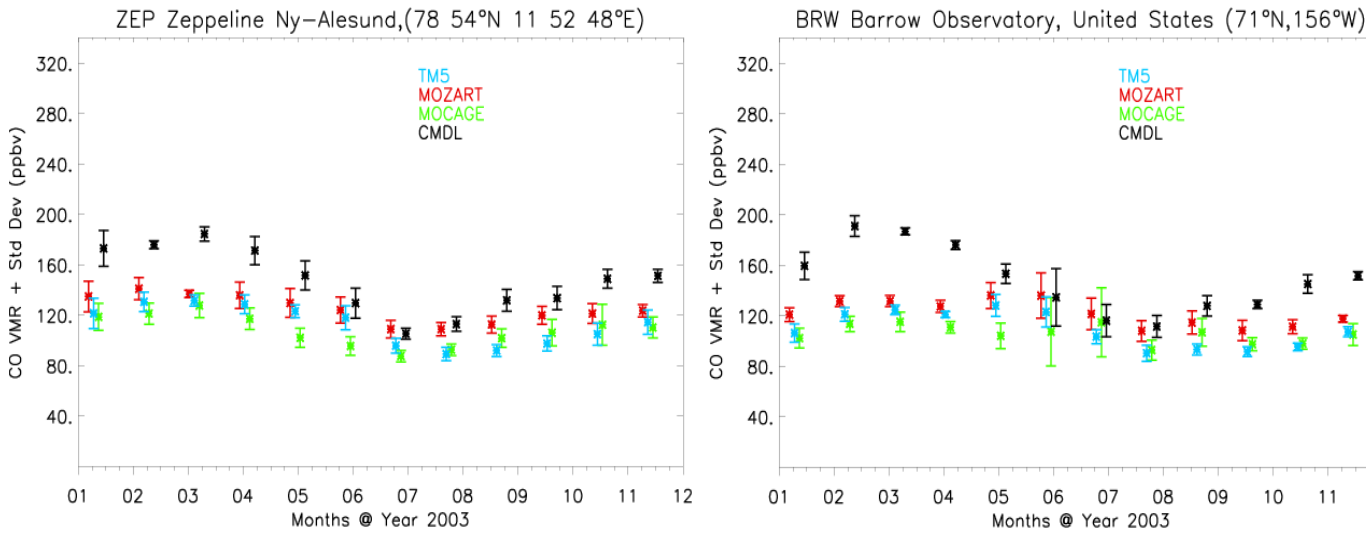
HBA Station: Except for the months of May and June 2003, MOCAGE-V3 and TM5-V10 show a fairly good agreement with CMDL data, while MOZART-V10 simulations overestimate the concentrations throughout the year.



NWR Station: MOCAGE-V3 data exhibit larger dispersions with overestimated CO concentrations. The reasons for the significant differences in MOCAGE-V3 output are not clear yet. However, it could be due to some induced artifact and/or higher emission factors used over North American region as compared to other GRG-CTMs. TM5-V10 and MOZART-V10 data is slightly underestimated as compared to CO CMDL observations.

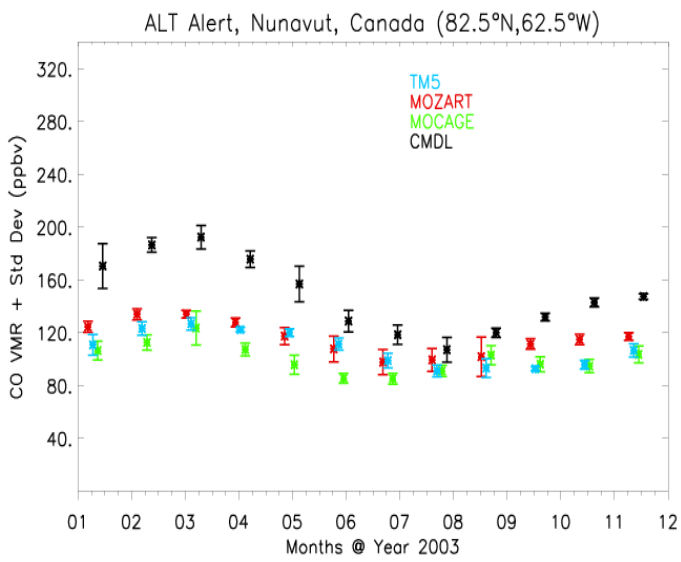
TM5-V10 and MOZART-V10 model results are slightly underestimated when compared to CO CMDL observations.

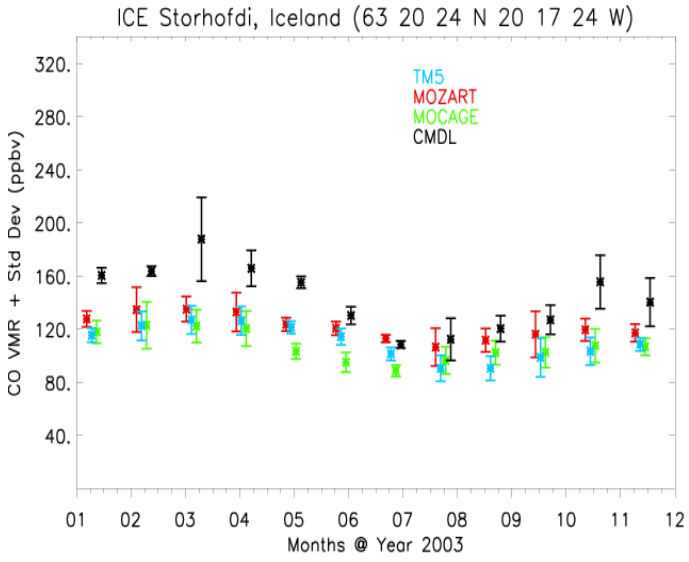
SPO Station: MOCAGE-V3 and TM5-V10 show a good agreement with CMDL observations over South Pole region, while MOZART-V10 CO concentrations are slightly overestimated



ZEP Station: all GRG-CTMs exhibit less CO concentrations, particularly during winter and spring months as compared to CMDL data

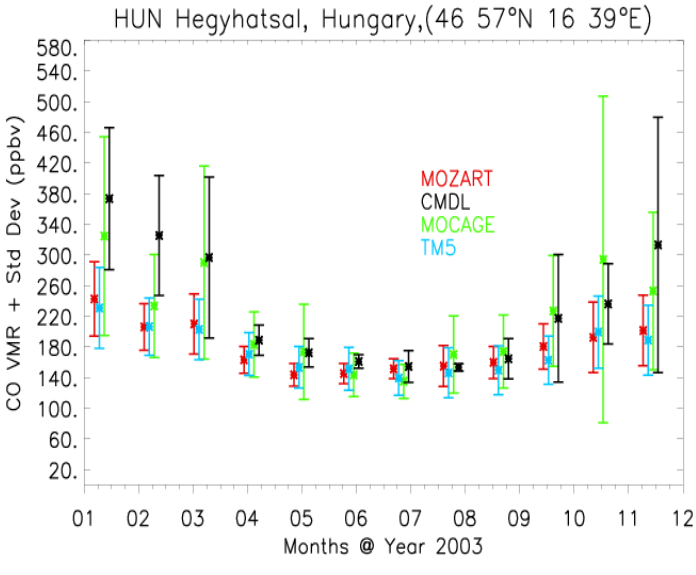
BRW Station: Similar behavior are observed over BRW station from GRG-CTMs as well

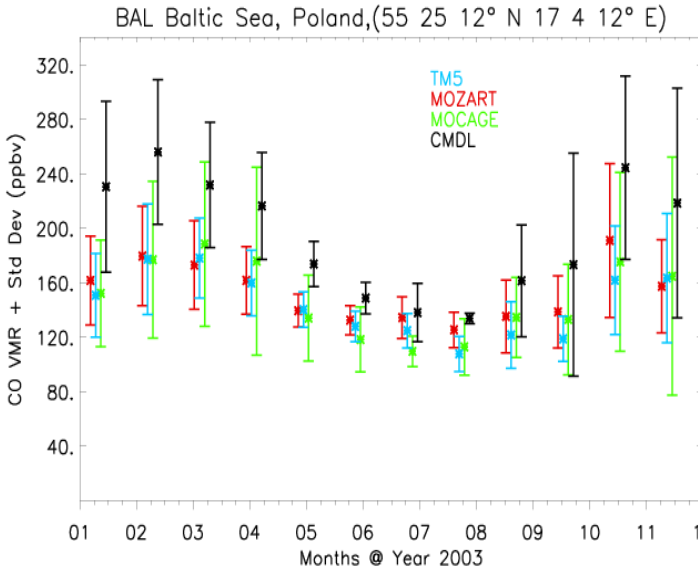




ALT Station: GRG-CTMs underestimate CO with relatively less underestimation during July – September 2003 period.

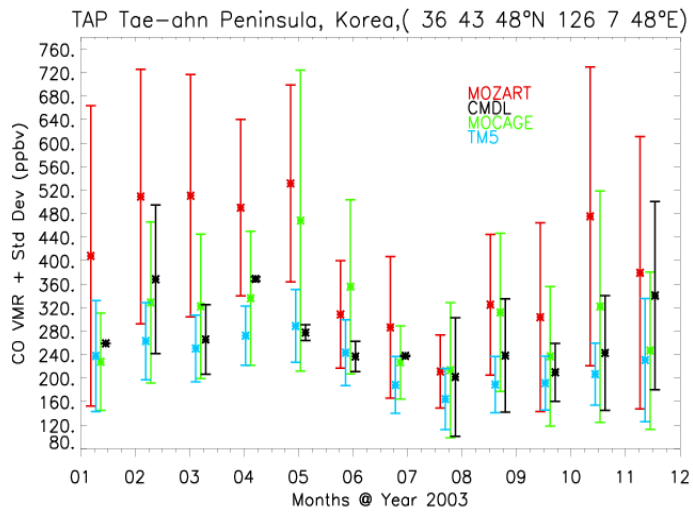
ICE Station: All GRG-CTMs exhibit similar seasonal patterns with slightly underestimated CO concentrations. However, from June to October 2003, MOZART-V10 model results are fairly well in agreement with CMDL observations.

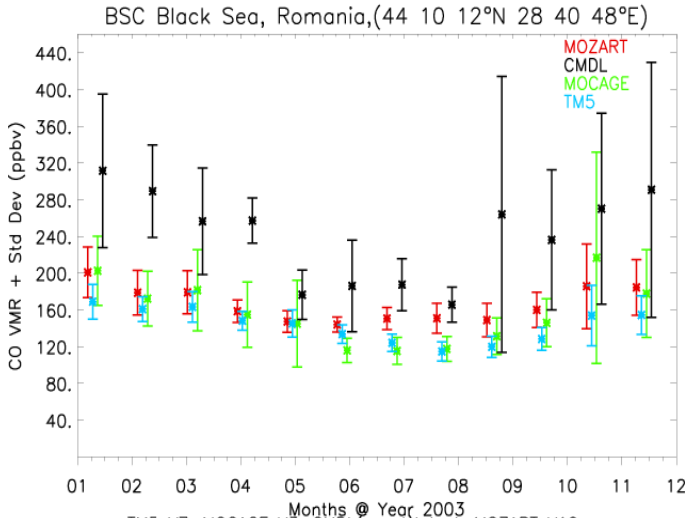




HUN Station: GRG-CTMs exhibit similar seasonal pattern with slightly underestimated CO concentrations, however, from April to September 2003 are fairly well matched with CMDL data.

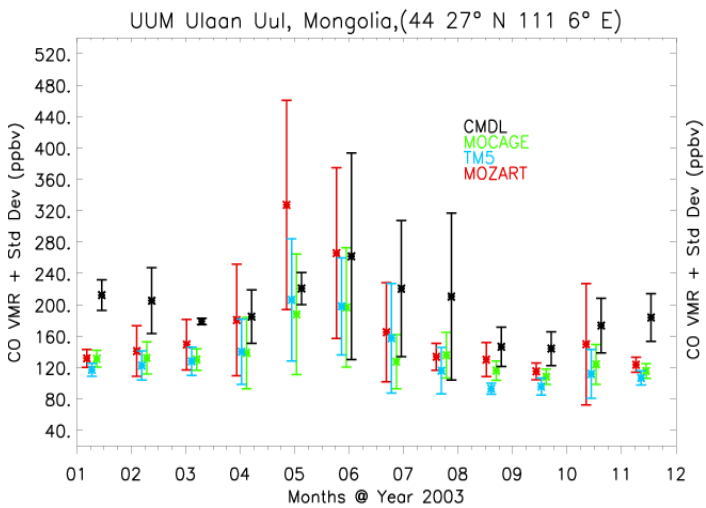
BAL Station: All CTMs exhibit similar seasonal pattern over BAL station with slightly underestimated CO concentrations as compared to CMDL data during winter months

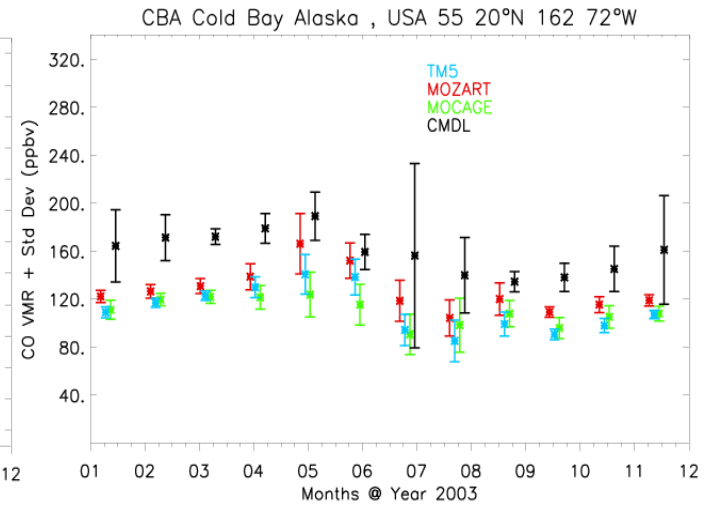




TAP Station: Besides June 2003, MOZART-V10 overestimates CO concentrations with a large standard deviation. However, MOCAGE-V3 and TM5-V10 data exhibit multiple behaviors of fair agreement, under- and overestimated CO concentrations.

BSC Station: All CTMs data exhibit similar seasonal patterns with relatively higher underestimated CO concentrations during winter months

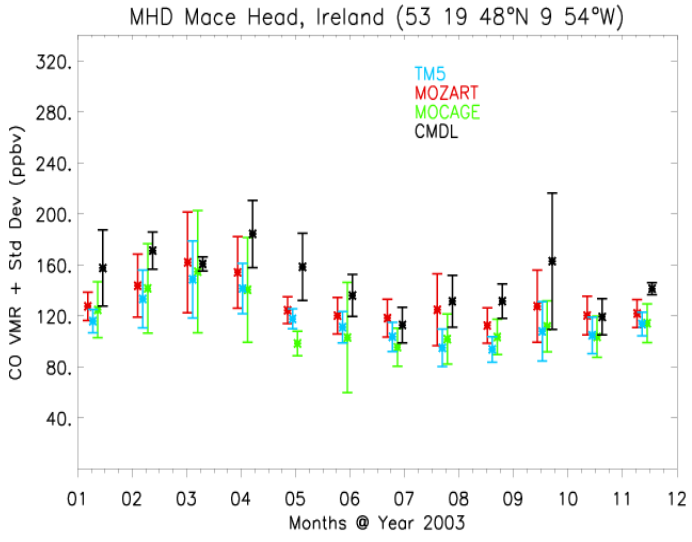


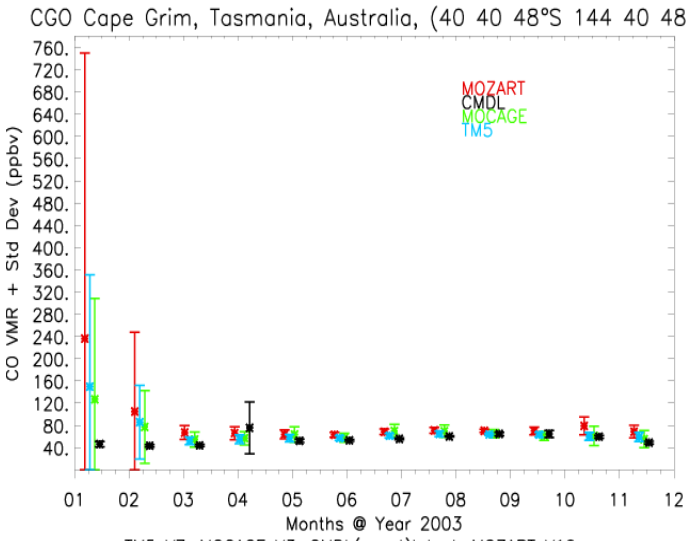


UUM station: GRG-CTMs exhibit similar seasonal variations, which are fairly consistent with observation data. However, CTMs model outputs are lower than observations with

exception of MOZART-V10 CO concentrations, which overestimate CO concentrations for May.

CBA Station: All CTMs data exhibit similar seasonal patterns with relatively higher underestimated CO concentrations during the first four months, when compared to the rest of year with slightly underestimated CO concentrations.





MHD Station: All GEMS- CTMs show slightly underestimated CO concentrations during the whole year except for MOZART-V10, which exhibits a fair agreement during March, July, August and November months.

CGO Station: except for first two months of year 2003, MOCAGE-V3 and TM5-V10 show a fairly good agreement with CMDL data, while MOZART-V10 CO concentrations are much higher. A large dispersion observed from GRG-CTMs for the first two months indicates some induced artifact/and or very special events which are not observed by the CMDL station. The calculated standard deviation for the month of January is 181, 201 and 525 ppbv from MOCAGE-V3, TM5-V10 and MOZART-V10 respectively. The source of this discrepancy in CO concentration has been further investigated

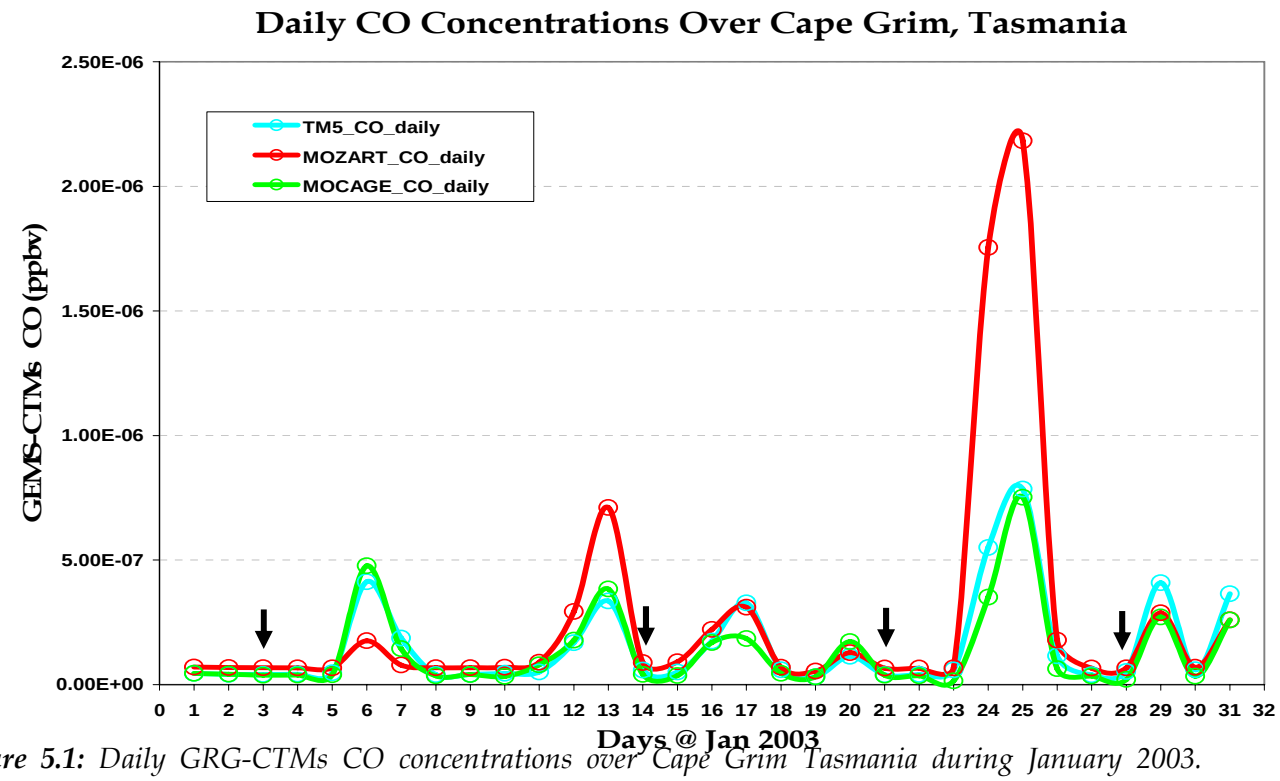


Figure 5.1: *Daily GRG-CTMs CO concentrations over Cape Grim Tasmania during January 2003. Different colors represent MOZART-V10 (red), TM5-V10 (cyan) and MOCAGE-V10 (green); Black arrows indicate sampling time of CMDL event data*

GRG-CTMs show inter-model consistency by exhibiting higher daily CO concentrations over Cape Grim Tasmania during the month of January 2003 presented in Figure (5.1). It seems to be resulting from fire events, transport and/or some artifact induced in meteorological data (wind fields) used by all GRG-CTMs. Since the sampling dates do not correspond to the high CO events, the high CO values are not indicated in the CMDL record.

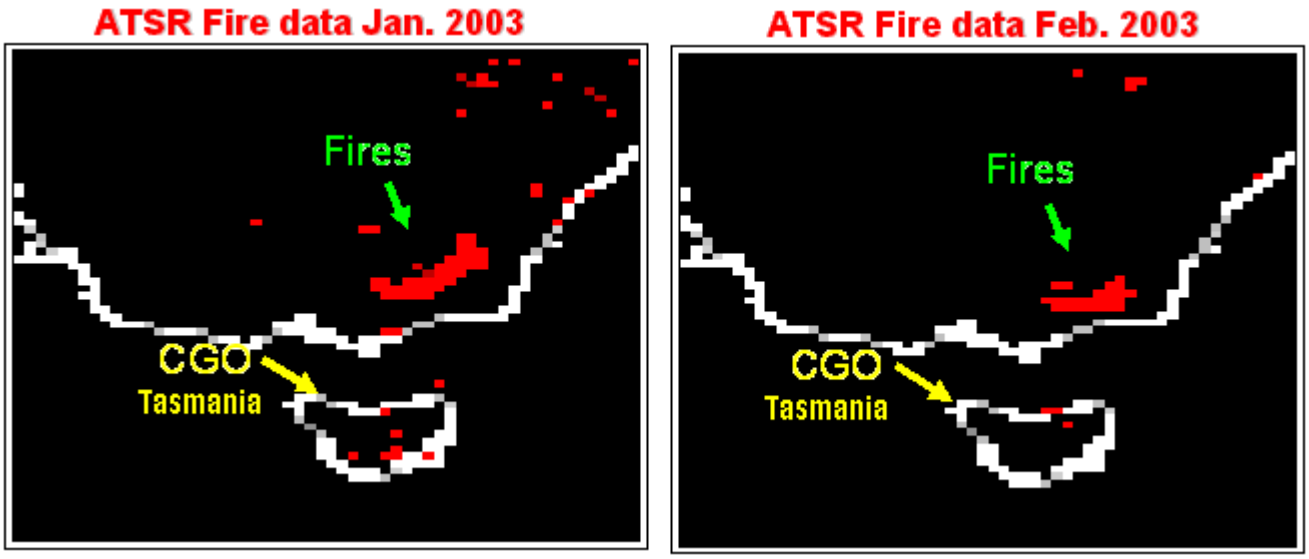


Figure 5.2: *Monthly ATSR Fire data (red spots) over Southern Australia and Tasmania during January and February 2003. Yellow arrow indicates the location of CGO station.*

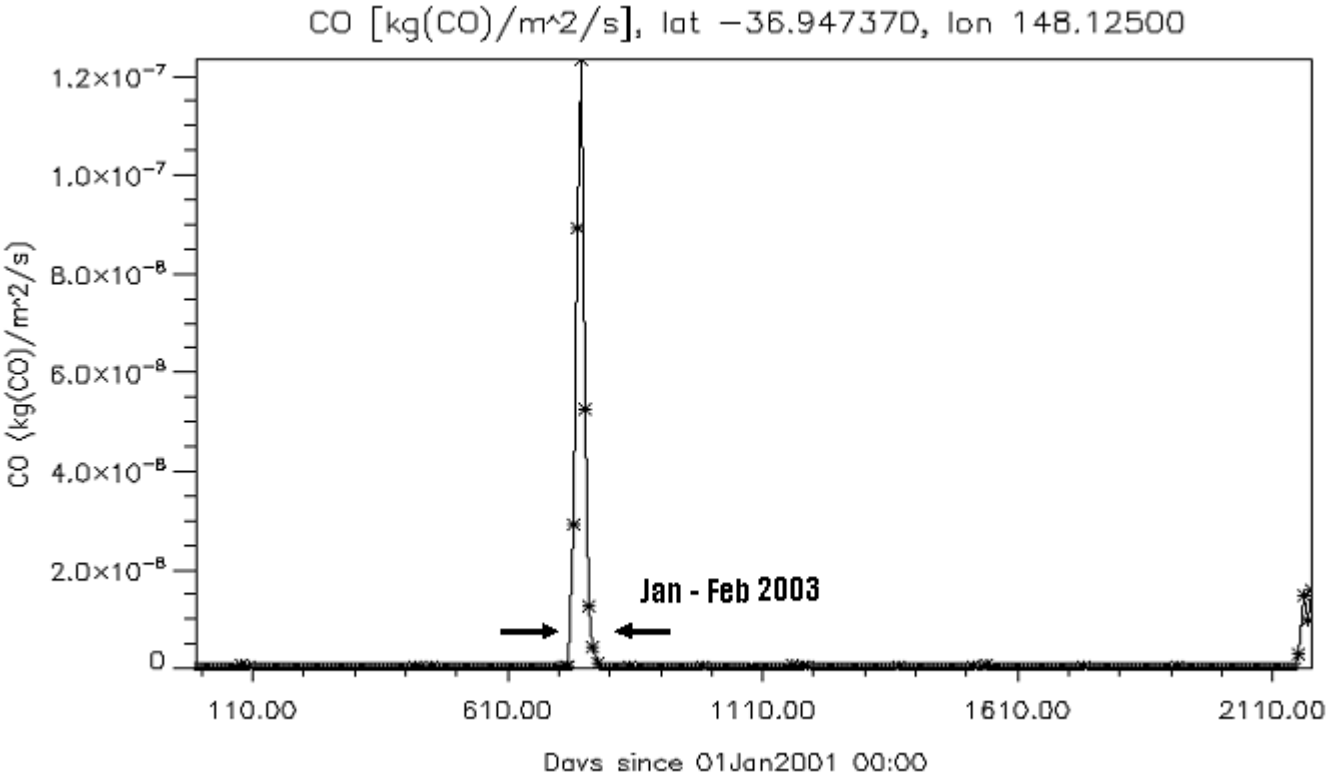


Figure 5.3: *Daily CO emissions (Kg/m²/s) resulting from fire emissions extracted from GFEDv2 8days emission inventories used by all GRG-CTMs over south of Australia since January 2001. Black arrows indicate the time period of Jan – Feb. 2003.*

ATSR fire data were investigated during this time period. Monthly mean distributions of ATSR (Along Track Scanning Radiometer - onboard ESR-2) active fire data (red spots – indicated by green arrows), for January and February 2003 are presented in Figure 5.2. Yellow arrow in both figures indicate the location of Cape Grim station. ATSR shows a large amount of active fires in January, the number of active fires is still significant, but decreasing in February All GRG-CTMs simulations show large CO accents in January, decreasing in February. This shows a good consistency between the ATSR fire data and GRG-CTMs.

CO emissions (in kg/m²/s) over the fire region (figure 5.2) from south of Australia (36.94°S and 148.125°E) as extracted from the GFEDv2 8days fire emissions during the time period of 2001-2006 are presented in Figure 5.3. Emission time series exhibit a very large peak in fire CO emissions during the Jan -Feb. 2003 period. The model simulations reflect the transport of CO to the Cape Grim station.

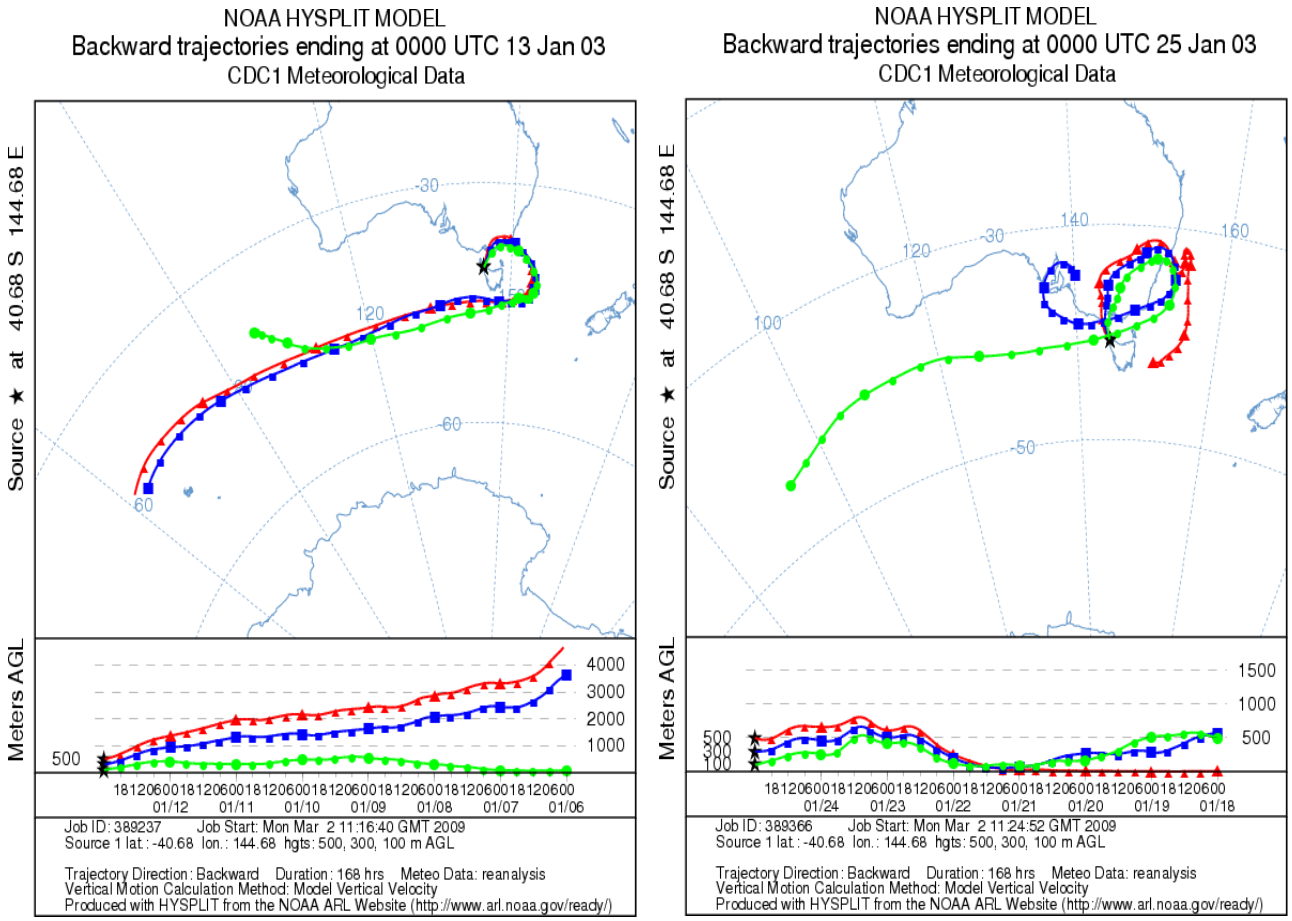


Figure 5.4: *Backward trajectories calculated at Cape Grim station, Tasmania (Black Star) on 13 January (left) and 25 the January (right). Trajectories are calculated from on-line version of HYSPLIT trajectories model at three different altitudes and with 6 hours temporal resolution for last 7 days.*

This transport is further investigated by calculating the HYSPLIT backward trajectories ending at the Cape Grim station (Black Star in Figure 5.4) during the two larger peaks in GRG-CTMs model results. Three backward trajectories at different altitudes of 100, 300 and 500 meter above ground levels are presented in Figure 5.4. These trajectories are calculated from an on-line version of the NOAA HYSPLIT trajectory model available from the Air Research laboratory's (ARL) website (www.arl.noaa.gov/ready/). Further

details about HYSPLIT are available from *Draxler et al., [2003]*. Seven days old backward trajectories on 13 and 25 January 2003 presented in Figure 5.4 reveal that air masses reaching Cape Grim are originating from the Pacific Ocean, making a circle over the fire region from South of Australia and intercepted with air masses enriched with fire CO and finally arriving at the Cape Grim station with in few days. Therefore, on the basis of our analysis we can conclude that the peaks in CO concentration exhibited by the GRG-CTMs are not due to local emissions rather related to fires from south of Australia. NOAA/ CMDL CO observation stations Cape grim could not catch this event of transport of fire emissions apparently because of sampling on irrelevant dates (black arrows in Figure 5.1.

Minor inter-model differences, as daily CO concentrations peaks from MOZART-V10 during 13, 17 and 24 January 2003 are larger than TM5-V7 and MOCAGE-V3 over Cape Grim, could be due to reasons that GRG-CTMs use different GEFEDv2 CO fire emission inventories for year 2003. As MOZART-V10 used CO fire emissions with 8 days temporal resolution while MOCAGE-V3 uses RETRO wild fire inventory and TM5-V7 used monthly GFEDv2 fire emissions inventory. Furthermore, all GRG-CTMs differ in fire CO emissions injection heights. MOZART-V10 used fire CO injected at lowest model level, while TM5-V7 and MOCAGE-V2 use vertical profiles of fire CO emissions (see Annex 1 to 3 for further details). Additionally, MOZART-V10 switched to new JPL-06 reaction rate for CO+OH and evaluation showed that the new (personal communication with Olaf Stein, FZJ, Germany) JPL-06 reaction rate is about 10-20% lower than reaction rate in previous simulations (V1-V9) and consequently CO concentrations from MOZART-V10 are about 10-20% higher in the troposphere.

The performance of all GEMS-CTMs based on monthly scoring index is compared in Table2. The score is obtained from the monthly mean scores obtained using Eq. 1. They are calculated from the means of weekly CO concentrations from GRG-CTMs simulations and CMDL CO observations from respective stations. The yearly mean scores are calculated from monthly scoring index from respective CMDl station.

On the basis of yearly scoring index presented in Table 2, all GRG-CTMs performance for all selected CMDL stations except CGO is more than 75%. MOZART-V10 performed less than 70 percent over CGO station. This is because of least performance of MOZART-V10 during first two months of year 2003. In fact, CMDL observations could not catch the fire CO transported from south of Australia (as discussed in previous section) to CGO station in Tasmania and thereby, affecting the scoring index of all GRG-CTMs during January and February 2003. However, in general performances of all CTMs is more than 85% (with some exceptions) based on monthly scoring indices and some times reaches 100%.

Monthly mean CO concentrations from GRG-CTMs are compared with *monthly* and *global view* CMDL data from each station. The analyses exhibit similar behaviour for each station as discussed in the previous section 3.1. Therefore, we decided to not discuss these intercomparisons explicitly, however, the intercomparison plots are presented in Annex A and B for CMDL *monthly* and *global view* data sets respectively.

Table 2: Monthly and yearly scoring index of GRG-CTMs for different CMDL surface stations

Month @ 2003	GRG-CTMs	ALT	ICE	MHD	ZEP	SPO	HBA	BAL	UUM	BRW	NWR	KEY	CGO	HUN	BSC	TAP	CBA
Jan	MOZ	0.86	0.9	0.9	0.89	0.73	0.79	0.85	0.81	0.88	0.95	0.73	-1.0	0.82	0.82	0.71	0.87
	TM5	0.82	0.86	0.87	0.85	0.96	0.98	0.83	0.78	0.83	0.93	0.69	-0.1	0.81	0.77	0.96	0.83
	MOC	0.81	0.87	0.9	0.84	0.93	0.88	0.83	0.81	0.82	0.6	0.77	0.13	0.93	0.83	0.94	0.84

Feb	MOZ	0.86	0.91	0.92	0.9	0.78	0.83	0.85	0.84	0.84	0.91	0.98	0.29	0.82	0.81	0.81	0.87
	TM5	0.83	0.87	0.89	0.87	0.98	0.96	0.85	0.8	0.82	0.88	0.88	0.51	0.82	0.78	0.86	0.84
	MOC	0.8	0.88	0.91	0.84	0.94	0.9	0.85	0.82	0.8	0.86	0.96	0.61	0.86	0.8	0.95	0.85
Mar	MOZ	0.85	0.86	1	0.87	0.82	0.91	0.87	0.92	0.85	0.92	0.98	0.74	0.85	0.85	0.54	0.88
	TM5	0.83	0.84	0.96	0.86	0.98	0.95	0.88	0.86	0.83	0.9	0.91	0.9	0.84	0.82	0.97	0.86
	MOC	0.82	0.83	0.98	0.85	0.96	0.89	0.91	0.86	0.81	0.91	0.93	0.88	0.99	0.85	0.89	0.85
Apr	MOZ	0.86	0.9	0.92	0.9	0.88	0.91	0.87	0.99	0.86	0.97	0.97	0.94	0.93	0.81	0.84	0.89
	TM5	0.85	0.88	0.88	0.88	1	0.97	0.87	0.88	0.84	0.95	0.91	0.86	0.95	0.79	0.87	0.86
	MOC	0.8	0.86	0.88	0.84	0.96	0.93	0.91	0.87	0.81	0.88	0.98	0.88	0.99	0.8	0.96	0.84
May	MOZ	0.87	0.9	0.89	0.93	0.9	0.83	0.9	0.76	0.94	0.92	0.97	0.89	0.92	0.92	0.54	0.94
	TM5	0.88	0.89	0.87	0.91	0.99	0.77	0.9	0.97	0.92	0.9	0.95	0.95	0.94	0.91	0.98	0.87
	MOC	0.8	0.83	0.81	0.84	0.98	0.76	0.89	0.93	0.84	0.88	0.96	0.89	1	0.91	0.66	0.83
Jun	MOZ	0.92	0.96	0.94	0.98	0.96	0.84	0.95	0.99	0.99	0.9	0.93	0.91	0.95	0.89	0.85	0.98
	TM5	0.93	0.94	0.91	0.96	0.97	0.79	0.93	0.88	0.96	0.87	0.96	0.96	0.97	0.86	0.99	0.93
	MOC	0.83	0.86	0.88	0.87	0.95	0.79	0.9	0.88	0.9	0.95	0.94	0.96	0.95	0.81	0.75	0.86
Jul	MOZ	0.91	0.98	0.98	0.98	0.92	0.96	0.99	0.87	0.98	0.89	0.84	0.89	0.99	0.9	0.9	0.88
	TM5	0.92	0.97	0.96	0.95	0.99	0.97	0.95	0.86	0.95	0.82	0.99	0.95	0.95	0.83	0.9	0.8
	MOC	0.86	0.91	0.92	0.92	0.98	0.94	0.9	0.79	0.99	0.96	0.82	0.87	0.94	0.81	0.98	0.79
Aug	MOZ	0.96	0.97	0.97	0.98	0.92	0.97	0.97	0.82	0.98	0.93	0.93	0.91	0.99	0.96	0.98	0.87
	TM5	0.92	0.9	0.86	0.89	0.99	0.97	0.9	0.78	0.91	0.82	0.82	0.96	0.98	0.85	0.91	0.8
	MOC	0.92	0.93	0.89	0.91	1	0.96	0.92	0.82	0.92	0.96	0.94	0.92	0.95	0.85	0.97	0.85
Sep	MOZ	0.92	0.96	0.93	0.93	0.95	0.97	0.92	0.94	0.95	0.94	0.92	0.96	0.99	0.78	0.82	0.95
	TM5	0.89	0.88	0.86	0.85	0.99	0.97	0.88	0.82	0.87	0.89	0.91	0.99	0.95	0.73	0.9	0.87
	MOC	0.93	0.92	0.89	0.89	0.97	0.96	0.92	0.9	0.92	0.82	0.87	1	0.97	0.75	0.85	0.9
Oct	MOZ	0.92	0.96	0.89	0.95	0.94	0.98	0.9	0.9	0.92	0.94	1	0.96	0.92	0.84	0.78	0.9
	TM5	0.85	0.89	0.83	0.86	0.99	0.92	0.84	0.83	0.85	0.92	0.89	0.99	0.87	0.77	0.96	0.83
	MOC	0.86	0.9	0.84	0.9	0.95	0.88	0.88	0.88	0.88	0.72	0.93	0.97	0.98	0.81	0.94	0.85
	MOZ	0.9	0.88	1	0.91	0.97	0.91	0.89	0.93	0.88	0.92	0.87	0.84	0.91	0.84	0.52	0.9

Nov	TM5	0.83	0.83	0.94	0.85	0.96	0.99	0.83	0.82	0.83	0.89	0.8	0.99	0.92	0.78	0.93	0.84
	MOC	0.83	0.85	0.93	0.88	0.89	0.92	0.86	0.86	0.84	0.68	0.9	0.99	0.88	0.9	0.84	0.86
Dec	MOZ	0.9	0.92	0.93	0.91	0.81	0.83	0.86	0.84	0.89	0.94	0.84	0.8	0.82	0.82	0.94	0.87
	TM5	0.86	0.89	0.9	0.88	0.93	0.96	0.87	0.79	0.85	0.93	0.77	0.9	0.8	0.77	0.84	0.83
	MOC	0.85	0.88	0.9	0.86	0.94	0.91	0.88	0.81	0.85	0.5	0.85	0.94	0.9	0.81	0.86	0.83
Yearly Mean	MOZ	**0.89**	**0.93**	**0.94**	**0.93**	**0.88**	**0.89**	**0.9**	**0.88**	**0.91**	**0.93**	**0.91**	**0.67**	**0.91**	**0.85**	**0.77**	**0.9**
	TM5	**0.87**	**0.89**	**0.89**	**0.88**	**0.98**	**0.93**	**0.88**	**0.84**	**0.87**	**0.89**	**0.87**	**0.82**	**0.9**	**0.81**	**0.92**	**0.85**
	MOC	**0.84**	**0.88**	**0.89**	**0.87**	**0.95**	**0.89**	**0.89**	**0.85**	**0.87**	**0.81**	**0.9**	**0.84**	**0.95**	**0.83**	**0.88**	**0.85**

5.3 Conclusions and recommendations

Sixteen NOAA/CMDL stations were selected on the basis of their strategic location in both hemispheres. The selected stations were categorized into four groups with respect to their location surrounded by and/or with potential of being influenced by some natural and anthropogenic activities such as agricultural and forest fires (FF), industrial activities (IA), high latitude/Polar regions (PR) and continental outflows (CO).

The intercomparison of GRG-CTMs model output (stand-alone run) with CMDL ground-based CO observations was performed over selected NOAA/CMDL surface stations (see Table 1) around the globe during 2003. In general, GRG-CTMs were able to capture the correct seasonal behaviour in CO concentrations; however, they were unable to simulate the exact amount of CO concentrations compared with CMDL observations.

Especially, GRG-CTMs exhibited a hemispheric dependent behaviour in the CO concentrations as compared to CMDL ground-based observations. The comparison over stations from Northern hemisphere (ALT, BRW, ICE, ZEP), GRG-CTMs underestimated CO concentrations during the whole year except for the months May to September 2003. Exceptionally, MOCAGE and MOZART overestimated CO concentrations over NOAA/CMDL stations NWR and TAP, respectively.

For surface stations (SPO, HBA, CBA) in the southern hemisphere, MOCAGE and TM5 showed an overall good agreement with CMDL CO observations. MOZART exhibited slightly overestimated CO concentrations.

In case of CGO station our analysis (section 5.2) showed that the peaks in CO concentration exhibited by the GRG-CTMs are not due to local emissions rather related to fires from south of Australia. NOAA/ CMDL CO observation stations CGO could not catch this event of transport of fire CO apparently because of sampling on irrelevant dates (black arrows in Figure (x1).

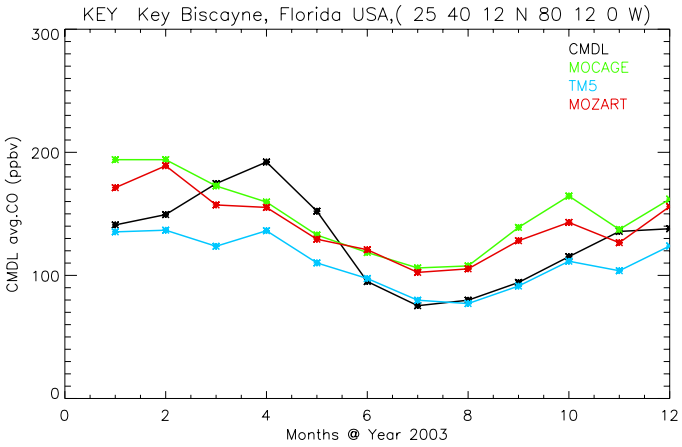
The performances of GRG-CTMs (stand-alone runs) over selected regions using CMDL observation data (using a scoring method given by Eq. 1) were calculated on monthly and yearly basis. Scoring index presented in Table 2 shows GRG-CTMs performance is more than 75% for all selected CMDL stations except for CGO station. In fact, CMDL observations could not catch the fire CO transported from south of Australia and thereby, affecting the scoring index of all GRG-CTMs over CGO station during first two months of year 2003.

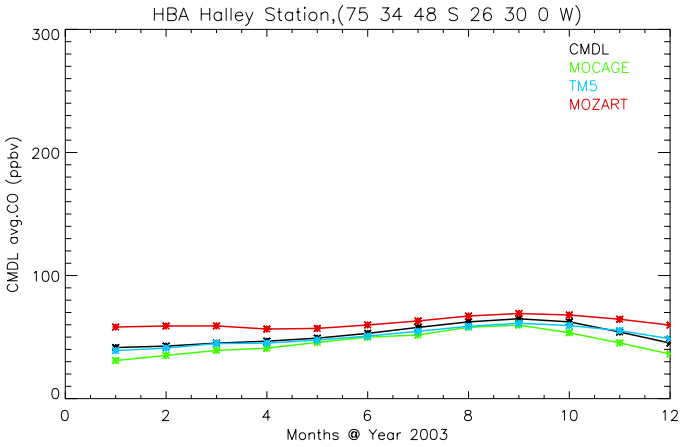
The inter-model differences can be due to the reason that GRG-CTMs differ from each other in various aspects such as, reaction rates, anthropogenic and fire emission inventories, horizontal girds and fire CO injection heights that were used in the different simulations.

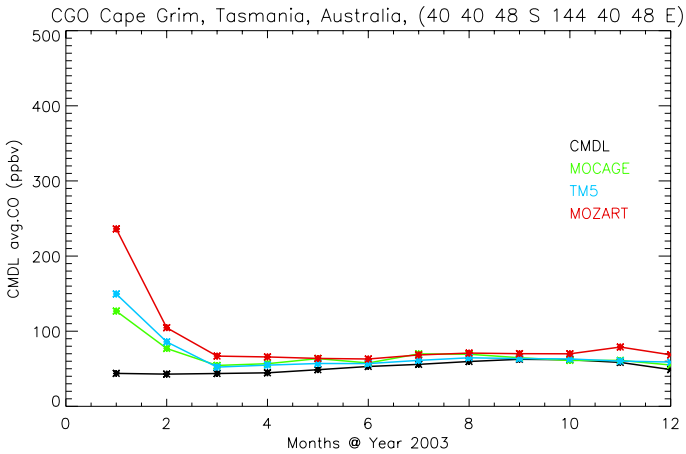
5.4 References

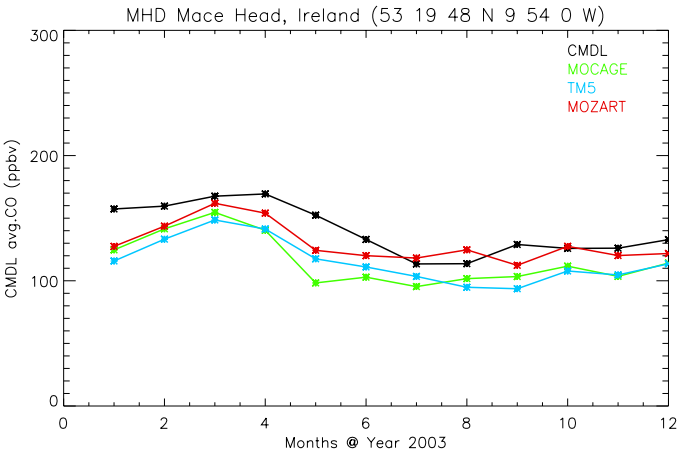
1. *Draxler, R.R. and Rolph, G.D., 2003. HYSPLIT (HYbrid Single-Particle Lagrangian Integrated Trajectory) Model access via NOAA ARL READY Website (http://www.arl.noaa.gov/ready/hysplit4.html). NOAA Air Resources Laboratory, Silver Spring, MD*
2. Novelli, P.C., J.E. Elkins, and L.P. Steele, The development and evaluation of a gravimetric reference scale for measurements of atmospheric carbon monoxide, J. Geophys. Res., 96, 13,109-13,121, 1991.
3. Novelli, P.C., K.A. Masarie, and P.M. Lang, Distributions and recent changes in carbon monoxide in the lower troposphere, J. Geophys. Res., 103, 19,1015- 19,033, 1998.
4. Novelli, P.C., K.A. Masarie, P.M. Lang, B.D. Hall, R.C. Myers, and J.W. Elkins, Re-analysis of tropospheric CO trends: Effects of the 1997-1998 wild fires, J. Geophys. Res., 108, D15 : 4464, doi:10.1029/2002JD003031, 2003.
5. Thoning, K.W., P.P. Tans, and W.D. Komhyr, Atmospheric carbon dioxide at Mauna Loa Observatory, 2: Analysis of the NOAA/GMCC data, 1974-1985, J. Geophys. Res., 94, 8549-8565, 1989.

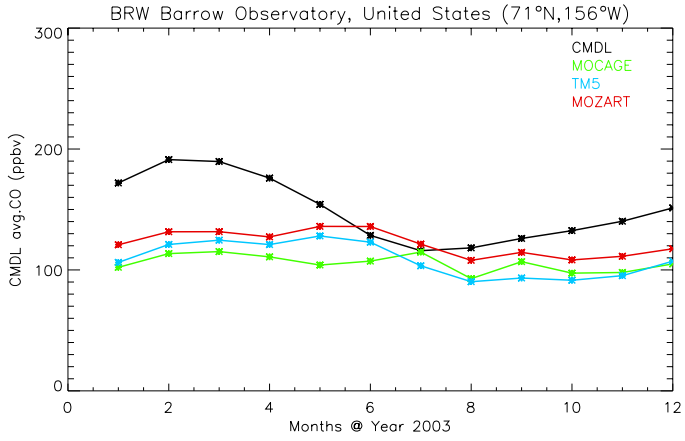
5.5 Annex A: CMDL CO monthly data and GRG-CTMs

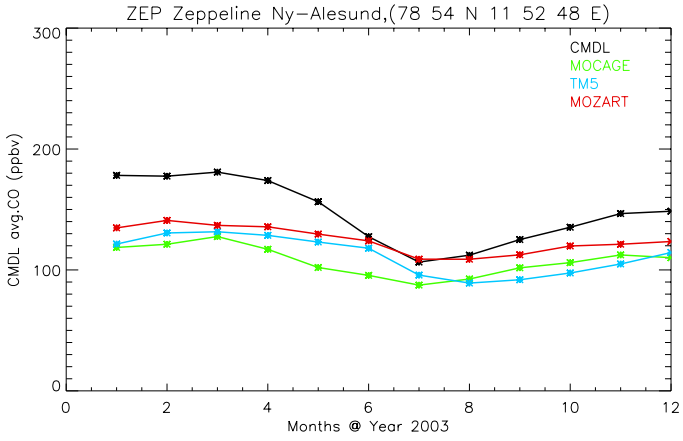


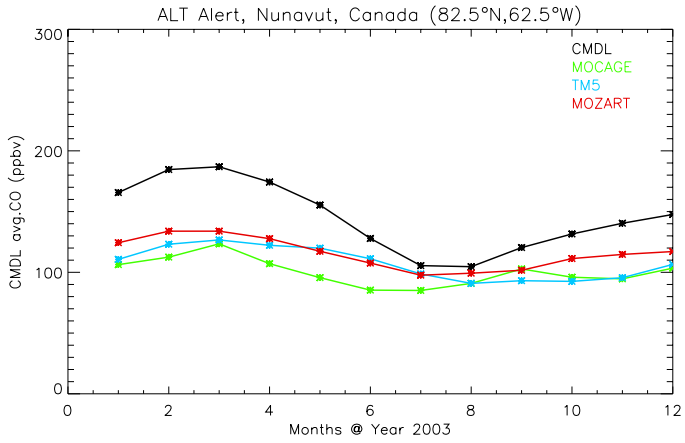


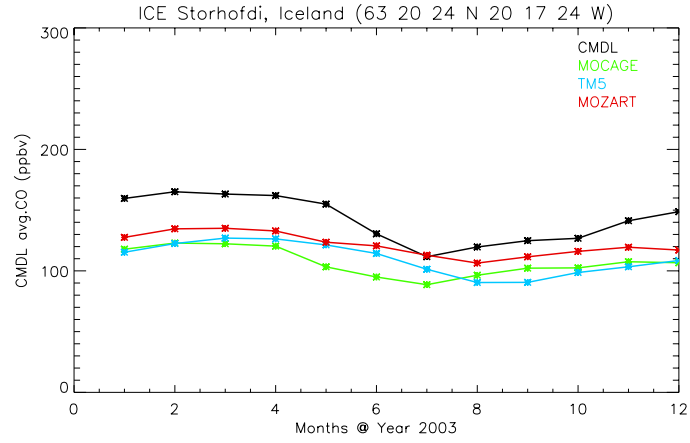


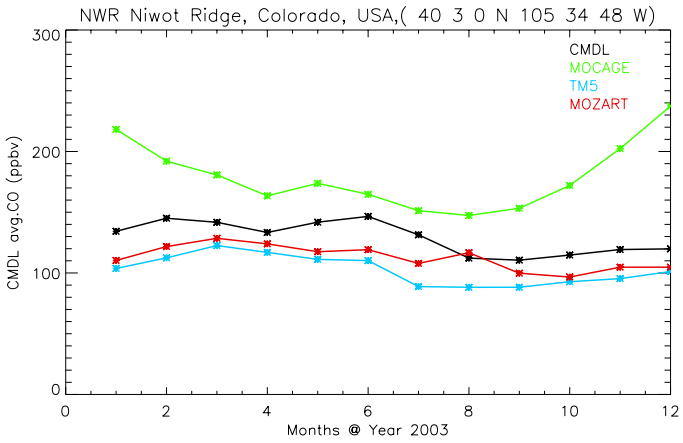


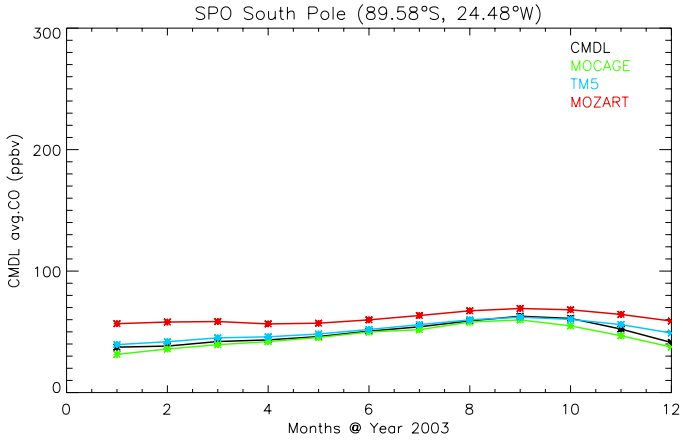


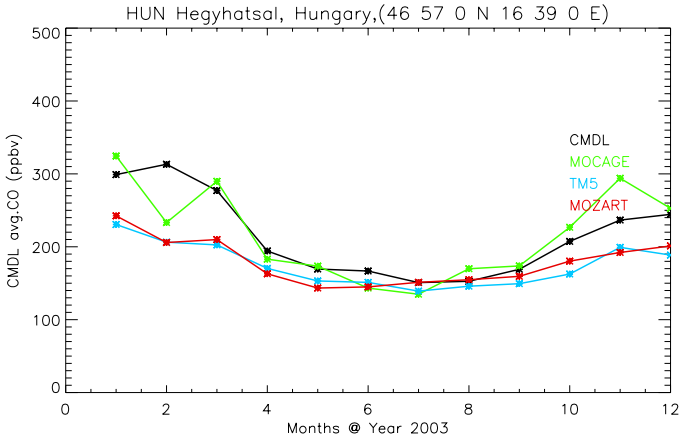


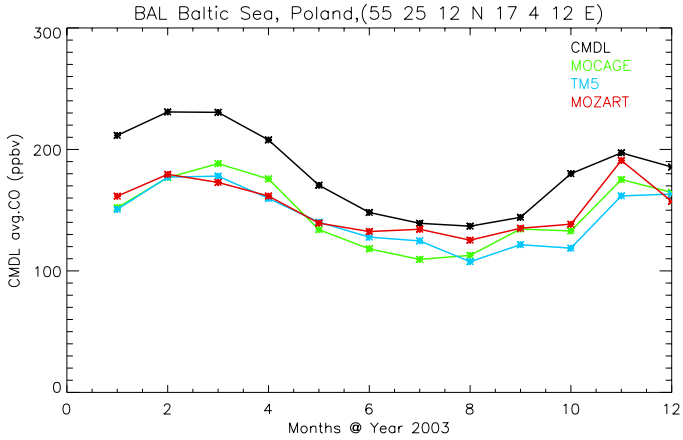


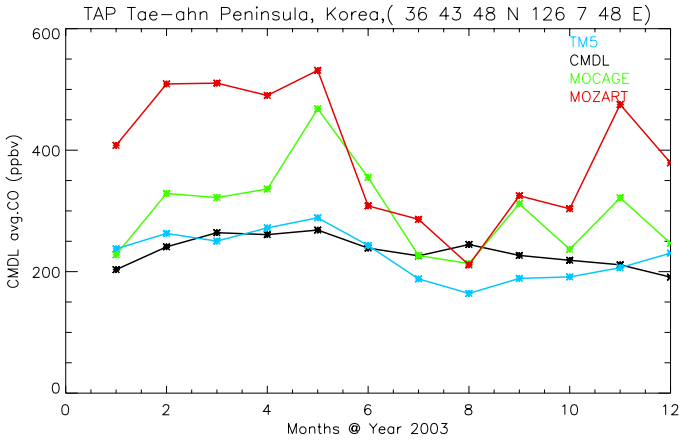


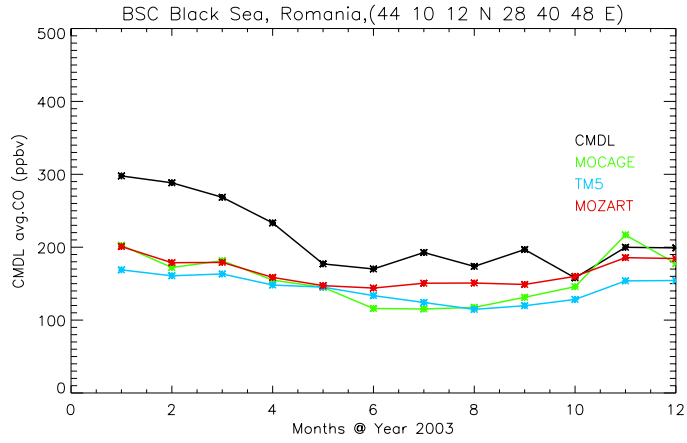

NWR Niwot Ridge, Colorado, USA,(40 3 0 N 105 34 48 W)
CMDL
MOCAGE
TM5
MOZART
CMDL avg.CO (ppbv)
Months @ Year 2003

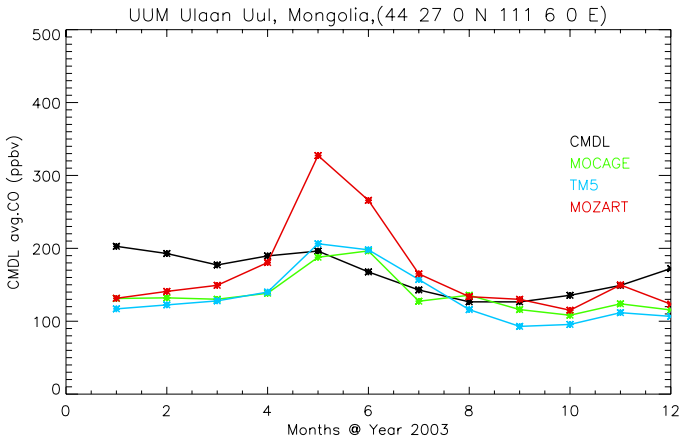

SPO South Pole (89.58°S, 24.48°W)
CMDL
MOCAGE
TM5
MOZART
CMDL avg.CO (ppbv)
Months @ Year 2003

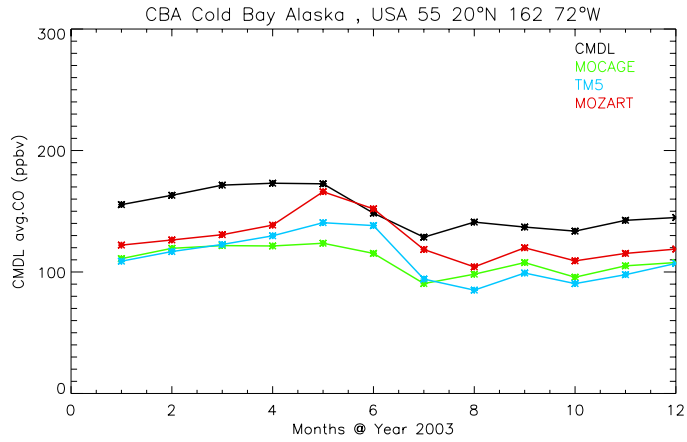

HUN Hegyhatsal, Hungary,(46 57 0 N 16 39 0 E)
CMDL
MOCAGE
TM5
MOZART
CMDL avg.CO (ppbv)
Months @ Year 2003


BAL Baltic Sea, Poland,(55 25 12 N 17 4 12 E)
CMDL
MOCAGE
TM5
MOZART
CMDL avg.CO (ppbv)
Months @ Year 2003

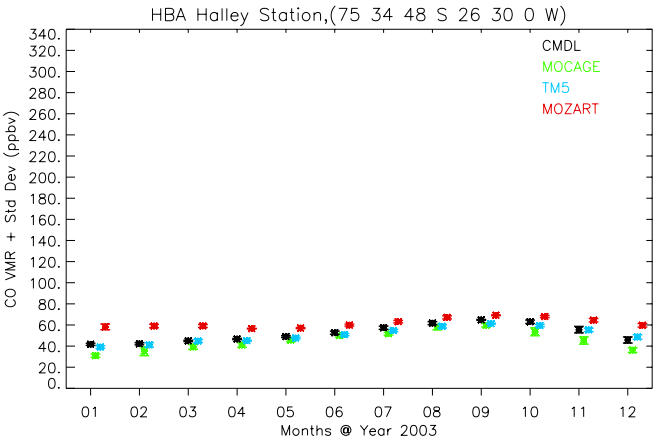

TAP Tae-ahn Peninsula, Korea,(36 43 48 N 126 7 48 E)
TM5
CMDL
MOCAGE
MOZART
CMDL avg.CO (ppbv)
Months @ Year 2003

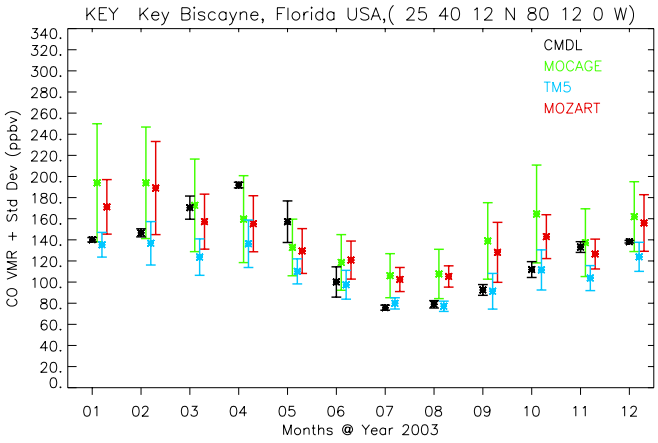

BSC Black Sea, Romania,(44 10 12 N 28 40 48 E)
CMDL
MOCAGE
TM5
MOZART
CMDL avg.CO (ppbv)
Months @ Year 2003

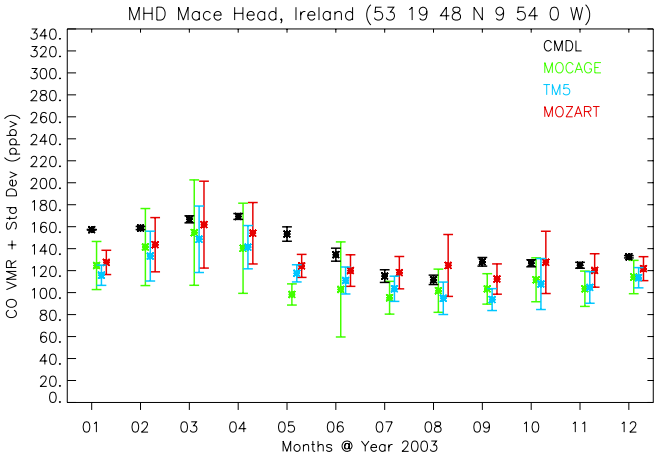

UUM Ulaan Uul, Mongolia,(44 27 0 N 111 6 0 E)
CMDL
MOCAGE
TM5
MOZART
CMDL avg.CO (ppbv)
Months @ Year 2003

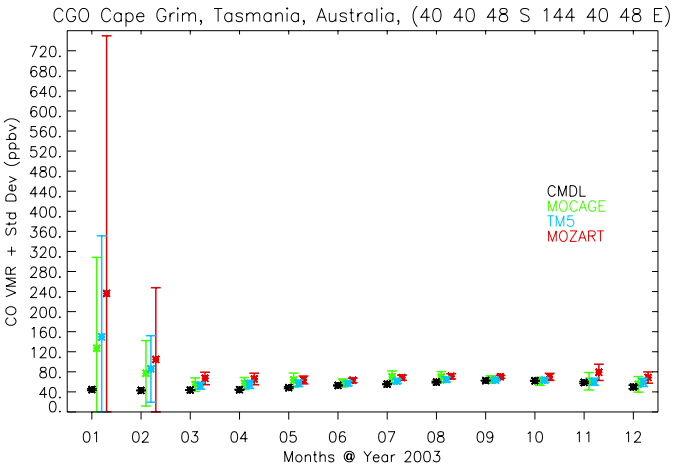

CBA Cold Bay Alaska , USA 55 20°N 162 72°W
CMDL
MOCAGE
TM5
MOZART
CMDL avg.CO (ppbv)
Months @ Year 2003

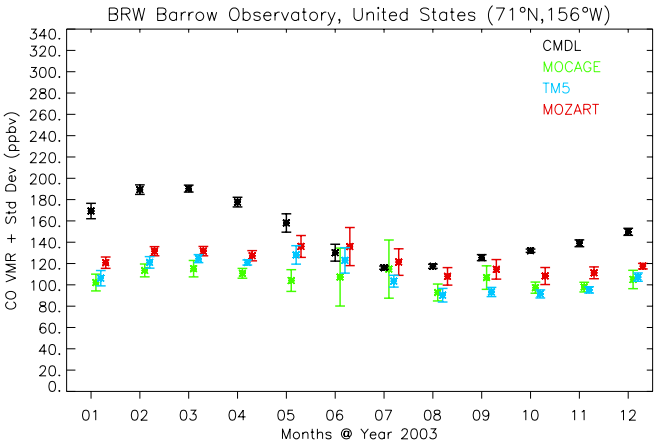
5.6 Annex B: CMDL CO global view data and GRG-CTMs

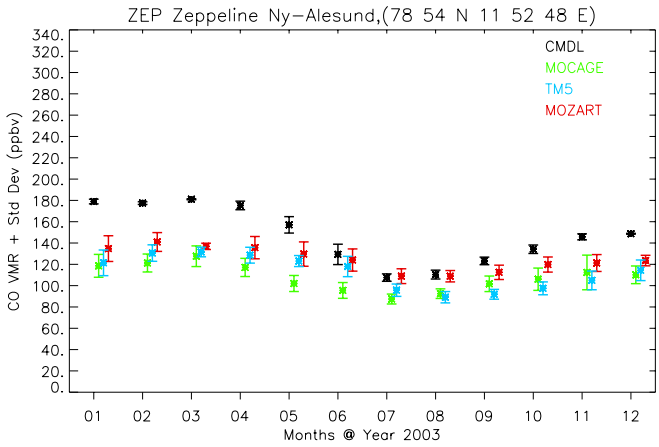


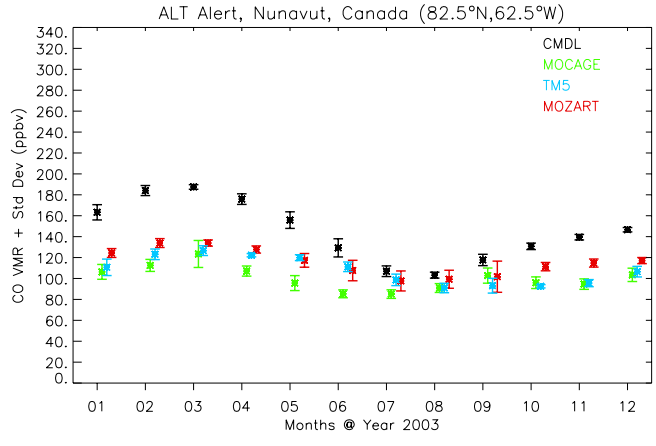


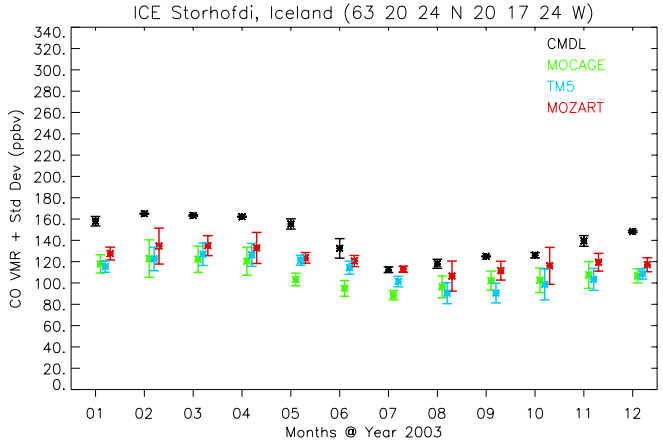


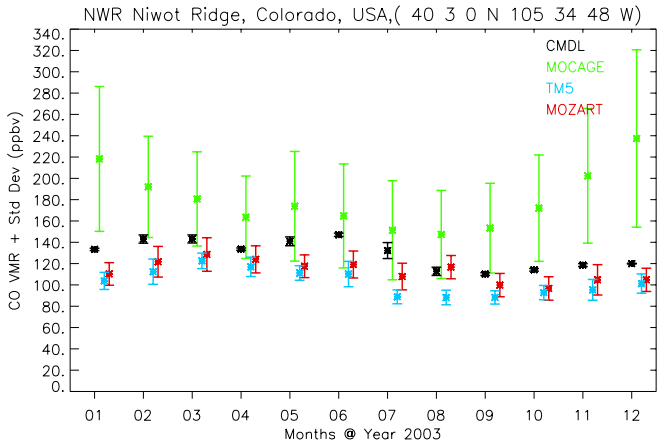


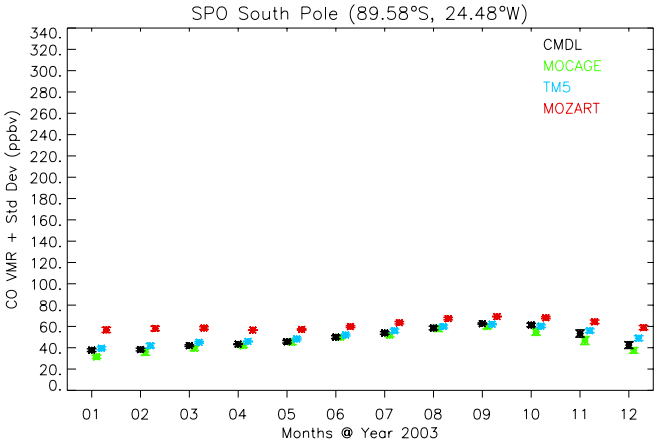


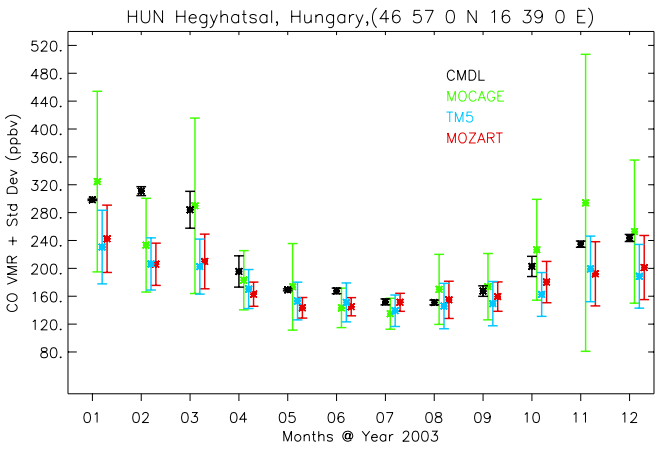


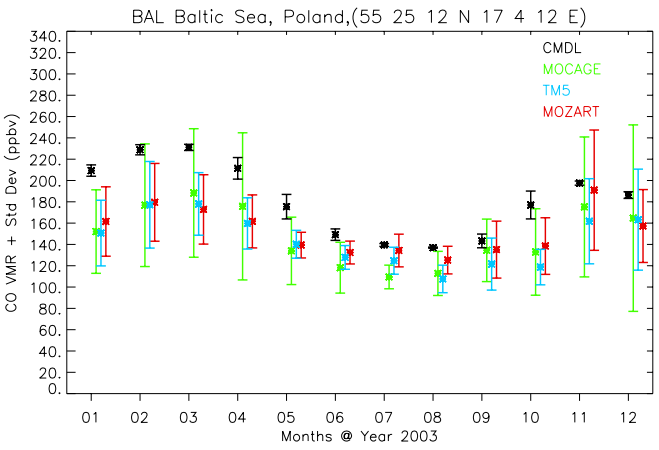


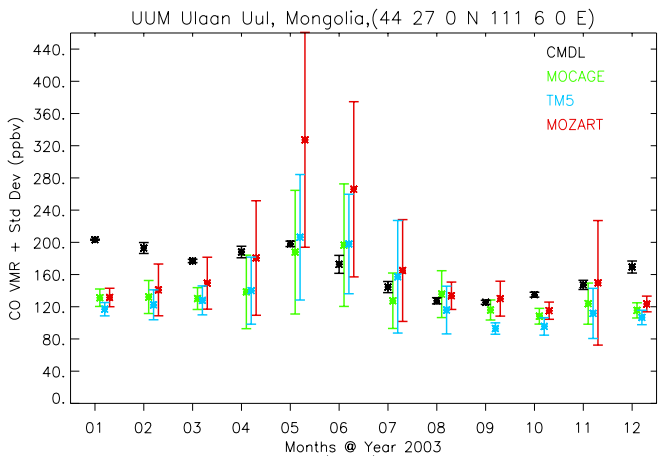


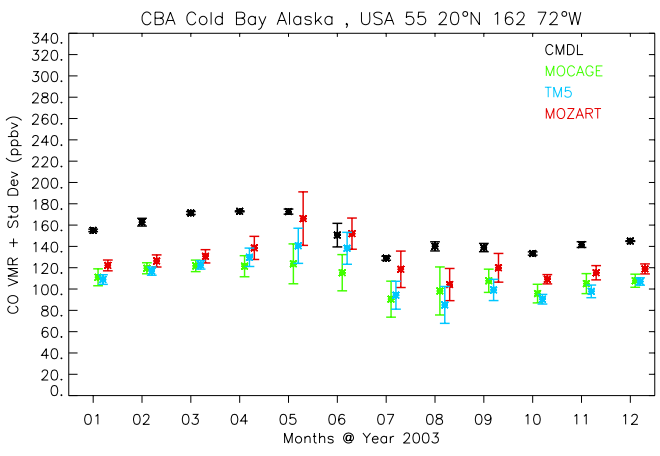

NWR Niwot Ridge, Colorado, USA,(40 3 0 N 105 34 48 W)
CMDL
MOCAGE
TM5
MOZART
CO VMR + Std Dev (ppbv)
Months @ Year 2003

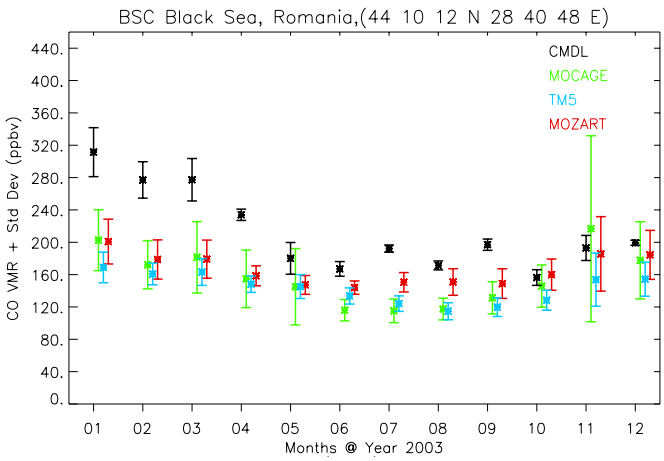

SPO South Pole (89.58°S, 24.48°W)
CMDL
MOCAGE
TM5
MOZART
CO VMR + Std Dev (ppbv)
Months @ Year 2003

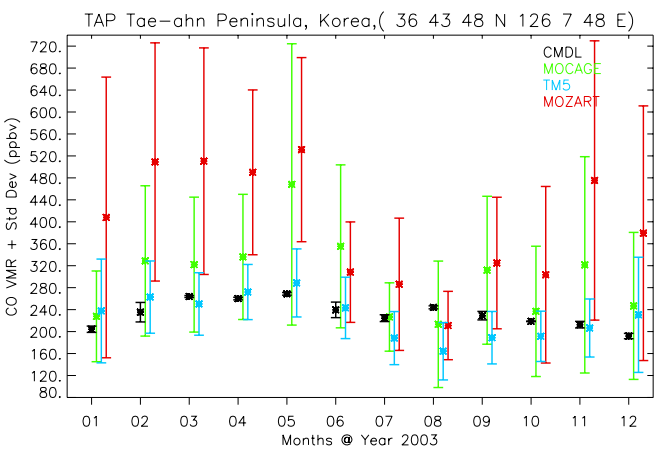

HUN Hegyhatsal, Hungary,(46 57 0 N 16 39 0 E)
CMDL
MOCAGE
TM5
MOZART
CO VMR + Std Dev (ppbv)
Months @ Year 2003


BAL Baltic Sea, Poland,(55 25 12 N 17 4 12 E)
CMDL
MOCAGE
TM5
MOZART
CO VMR + Std Dev (ppbv)
Months @ Year 2003


UUM Ulaan Uul, Mongolia,(44 27 0 N 111 6 0 E)
CMDL
MOCAGE
TM5
MOZART
CO VMR + Std Dev (ppbv)
Months @ Year 2003


CBA Cold Bay Alaska , USA 55 20°N 162 72°W
CMDL
MOCAGE
TM5
MOZART
CO VMR + Std Dev (ppbv)
Months @ Year 2003


BSC Black Sea, Romania,(44 10 12 N 28 40 48 E)
CMDL
MOCAGE
TM5
MOZART
CO VMR + Std Dev (ppbv)
Months @ Year 2003


TAP Tae-ahn Peninsula, Korea,(36 43 48 N 126 7 48 E)
CMDL
MOCAGE
TM5
MOZART
CO VMR + Std Dev (ppbv)
Months @ Year 2003

6. Evaluation with GAW ground-based data

Contributors: Flentje Harald (DWD, Hohenpeissenberg, Germany), and the modeller teams

6.1 Datasets and methodologies

6.1.1 Datasets

Observations from the Global Atmosphere Watch (GAW) network are used to evaluate global model simulations (1.125°-2° lat x 1.125°-3° lon, 60 vertical levels). A complete list of all relevant model runs can be found in Annex 6. While many runs were evaluated during the development phase of GEMS, only the latest relevant offline and coupled runs (see table below) are discussed in this report.

Table 6.1: Latest relevant offline and coupled model runs used in this section

Run ID / Info	**Start- End**	**Notes**	**Purpose**
Mozart offline – V1	2003-2004		Reference stand-alone run
TM 5 offline – V7	2003-2004		Reference stand-alone run
Mocage offline – V2	2003-2004		Reference stand-alone run
Mozart exoz	0701-0831	CO and O3 feedback mode, IFS vertical transport	Fluxes corrected
Mozart eydy	0701-0831	CTM constrained mode, CTM vertical transport	Fluxes corrected + improved cloud restart
Mozart-eyih	20030501- 20031231	IFS convection and diffusion for CO. CTM convection and diffusion for GO3, NOx. 2-way coupling to MOZART for CO and GO3.NOx and HCHO initialized from CTM in every forecast. Bugfix for CTM fluxes	1st reanalysis.CO (Mopitt) and GEMS ozone assimi- lation (SCIAMACHY, MIPAS, GOME, SBUV).
Mozart-eyq6	20030501 20031231	CO data and GEMS ozone data passive.	Control run for eyih.
Mozart-f026	20030101 20041231	All fixes from eyih SBUV ozone MIPAS ozone (from20040501) CO-Mopitt	Second -reanalysis (together with GHG and AER)
TM5-eybl	20030701	CTM constrained mode, CTM vertical transport	- first trial

	20030830	24 h Forecast	
TM5-ez3h	20030701 20030830	CO feedback IFS vertical transport 24 h Forecast	-first trial.Updated version is under construction.

The Global Atmosphere Watch (GAW) network consists of 24 'global' stations, covering (almost) the complete set of GAW parameters, more than 300 ‘regional’ stations, which participate in one or more GAW programs, and a number of associated or contributing networks that are either regionally based (e.g. EMEP), or thematically based (e.g. BSRN). As GAW is a nationally funded network, most stations do only measure a subset of parameters, often for limited periods. Up to date information on contributing stations and their measurements are available from GAWSIS via 5 world data centres operated by WMO members. Trace gas data are available from the World Data Centre for Greenhouse Gases (WDCGG) (Japan Meteorological Agency JMA, Tokyo, Japan). These data are freely available to the scientific community. The observation sites for surface O_3 and CO are shown in Figures 6.2, whereby not all of them delivered data for year 2003.

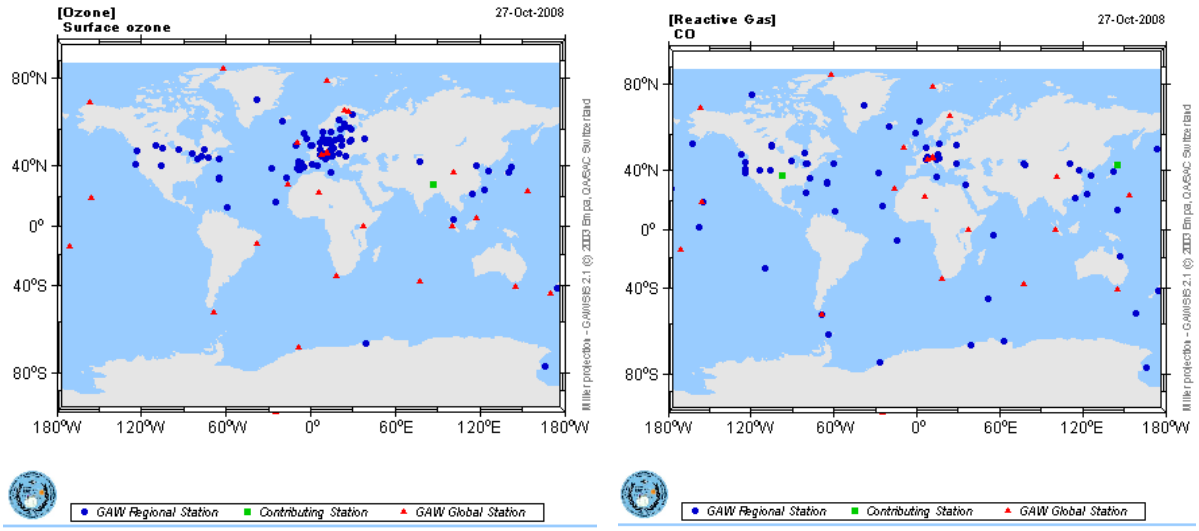


Figure 6.2: GAW surface observation sites for O_3 and CO.

Table 6.2: Number of stations by parameter, country and region

	Region I	Region II	Region III	Region IV	Region V	Region VI	Antarktika	Mobile	Total
Station	**12**	**30**	**8**	**62**	**33**	**117**	**21**	**29**	**312**
Country	**8**	**10**	**5**	**6**	**7**	**31**	**8**	**3**	**66**
O_3	4	6	2	17	5	47	4		**85**
CO	6	9	3	16	14	21	8	3	**80**
NO_2				1	2	42			**45**
NO						11			**11**

NOx						6			**6**
NOy						3			**3**
SO_2					2	46			**48**

Region I: Africa, II: Asia, III: South America, IV: North/Central America, V: South-West Pacific, VI: Europe

6.1.2 Methodologies

Evaluation Procedure:

Though GAW sites are established at regionally (or even globally) representative sites, individual stations are excluded since the applicability for evaluation of global models is questionable. Model values at the station's locations in the horizontal are interpolated linearly from the model gridded data. In the vertical they are extracted at the model level which matches the GAW stations' real altitude (geopotential) which is equivalent to matching the mean pressure of model level and station. As upper levels lack ground effects, this approach is not really satisfactory for mountain stations - thus the sensitivity to the chosen model level is used to estimate the uncertainty.

The Geopotential at model-halflevels is calculated by:

$$\Phi_{k+1/2} = \Phi_{surf} + \sum_{j=k+1}^{NLEV} R_{dry}\,(T_v) \ln \frac{p_{j+1/2}}{p_{j-1/2}}$$

whereby $p_{k+1/2} = A_{k+1/2} + B_{k+1/2} \cdot p_{surf}$

and $A_{k+1/2}$, $B_{k+1/2}$

are functions of the model level describing the transition from surface-following model levels at ground to pure pressure levels at the top of the model domain. p_{surf} being surface pressure, R the gas constant, T virtual temperature. The deviation of the model from the observations is either expressed as

modified normalized mean bias $$B_{mnm} = \frac{2}{N} \sum_i^{region} \frac{f_i - o_i}{f_i + o_i} \quad ,$$

station-normalized median bias $$B_{mnm} = Median \sum_i^{region} \frac{f_i - o_i}{Median(o_i)}$$

or as fractional gross error $$E_f = \frac{2}{N} \sum_i^{region} \left| \frac{f_i - o_i}{f_i + o_i} \right|$$

where f_i and o_i are the forecasted and the observed values at site i.

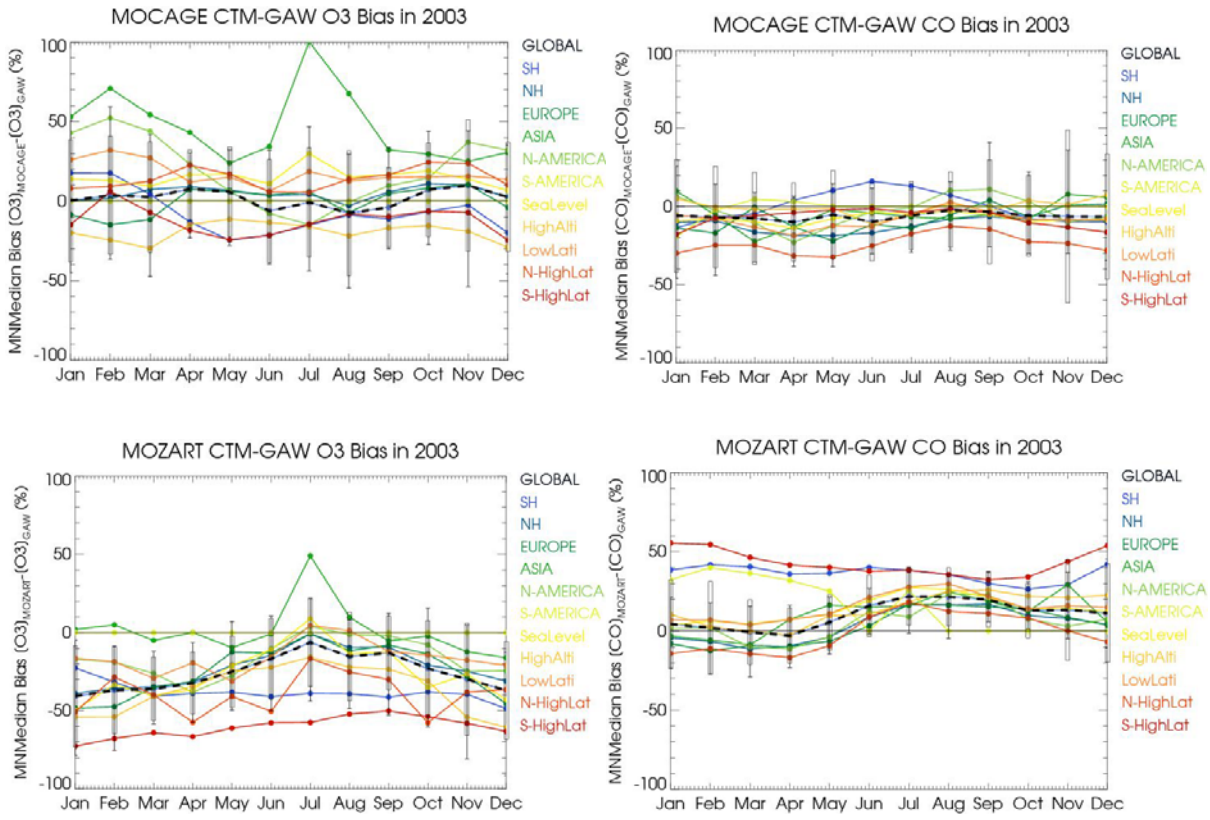
Limitations:

Comparison of surface in-situ data with global model output comprises several issues of representativity: In the horizontal, this is relatively uncritical in the sense that errors can be estimated. The verical matching remains critical. The relevant model level must be selected for each station individually, here depending on

the deviation of it's real altitude/pressure from that of the smoothed model surface layer. Generally stations in flat terrain are less critical. In the frame of GEMS global evaluation, we cannot investigate local sources or local meteorological effects in detail – obviously 'problematic' sites are rather ignored.

6.2 Offline simulations

The three CTM initially have been run in their original configuration without coupling to the IFS in order to have a reference and to allow intercomparison independent from their compatibility to ECMWF's integrated forecasting system (IFS). Later on they were coupled to the ECMWF-IFS and produced re-analyses for the years 2003 and 2004. In correspondence to the agreed evaluation strategy and metrics (described in Annex 7), the modified normalized mean/median bias (MNMB) between simulations and measurements was studied on a monthly mean basis and separately for different regions. The bias' annual and diurnal cycle were derived as well as correlation and covariance of time series, providing Taylor plots based on averaging over different periods and regions. Specific events, such as forest fires, summer smog and stratospheric ozone depletion, are examined in detailed case studies (see Chapters 8 to 12) and are not discussed here. Note that median and mean bias emphasize different features since the former disregards extreme events (e.g. summer 2003) as 'outliers' while the latter may be less representative for 'typical' conditions. This section summarizes the evaluations only for the latest runs of each CTM all previous iterations are documented on the GEMS website.



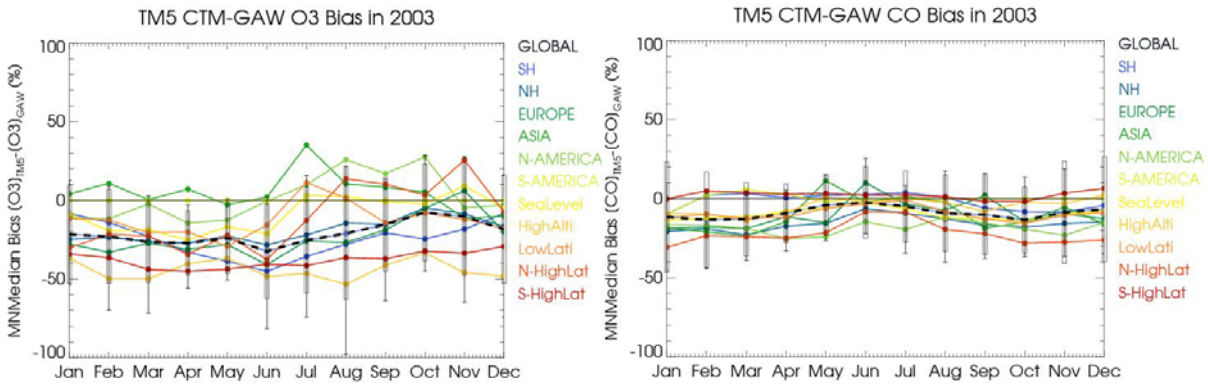


Figure 6.3: Annual variation of the modified normalized median bias in 2003 for the three GEMS-CTM running in offline mode. Left panels: ozone, right panels: CO. The monthly boxes and whiskers indicate the standard deviation and the 10/90%-iles over the regions' stations collective.

Figure 6.3 summarises the monthly median bias separately for different regions according to the evaluation template available at the GEMS website. The boxes and the whiskers indicate the standard deviation and 10/90%-iles for the global mean line (dashed black). The large intra-station variability is obvious also from the example in Figure 6.4 where the individual stations' biases are shown for CO measuring GAW stations all over the globe. Using the median bias takes into account the possibility of measurement errors, instrument artefacts or local effects.

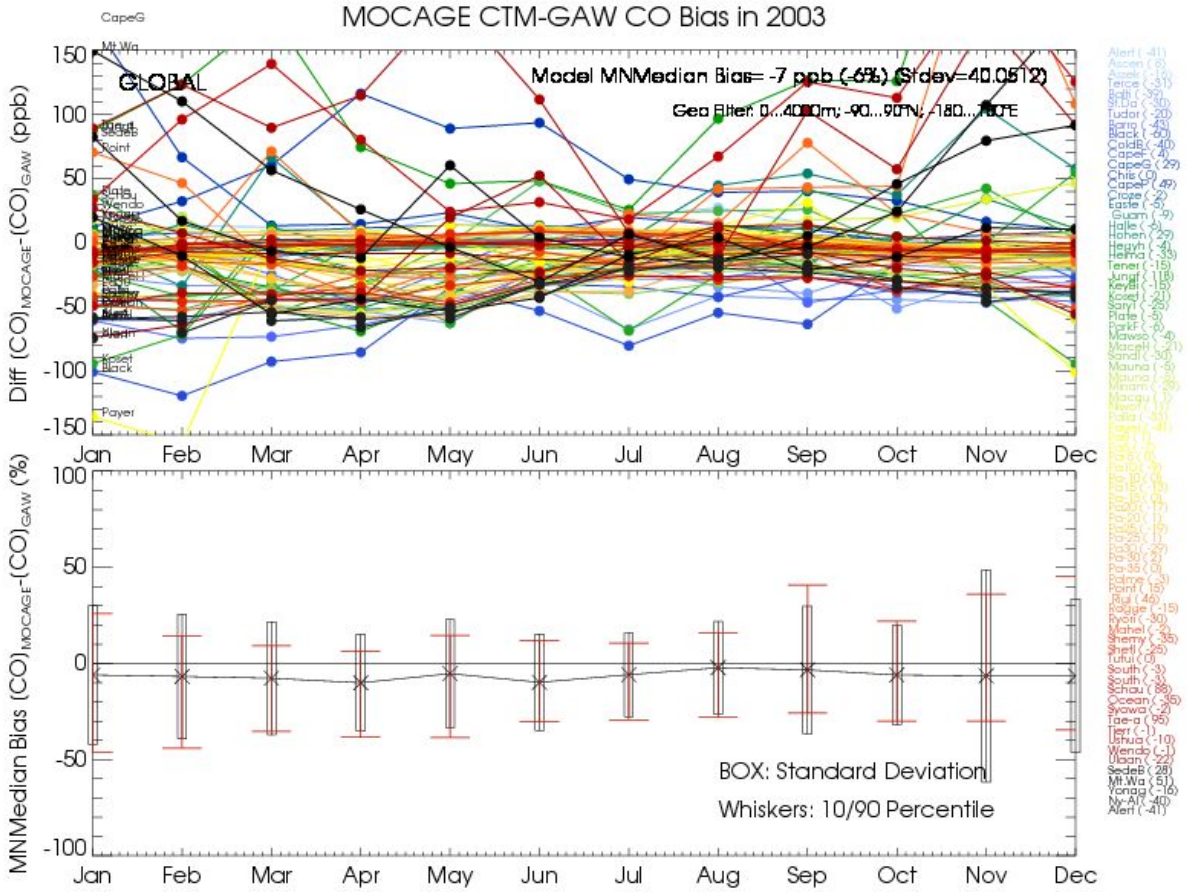


Figure 6.4: Modified normalized median CO-bias for MOCAGE running offline for 2003. Upper panel: individual stations, lower panel: monthly boxes for standard deviation and whiskers for 10/90%-iles calculated for all stations.

The MOCAGE CTM shows lowest absolute values and smallest annual amplitude of the bias, whereby part of the annual cycle in one region is balanced by an inverse bias in another. The MOZART CTM exhibits the largest annual cycles: a lower negative O_3-bias in summer than in winter and a higher positive CO-bias in summer than in winter. The most relevant region Europe (bluish green) is in all cases relatively close to the global average. Observations in the southern hemisphere, tends to be worse reproduced in all CTM, particularly at high latitudes. The globally averaged MNMedian biases are summarized in table 6.4. Evaluations of NO, NO_2 and SO_2 have been performed for few GAW stations and the results show that NO_2 is reproduced much better than NO and SO_2. An example is given for GAW station Hohenpeissenberg in Fig. 6.5. While NO_2 is captured reasonably well, the NO mixing ratios and their variability are much too high. As ozone is calculated via photochemical equilibrium, deficiencies in one of the involved parameters inevitably transfers to the others as well, particularly to ozone. These shortcomings also reflect in the diurnal variation of the bias as shown for selected stations in Figure 6.6. The simulated amplitude of the diurnal cycle is too large at stations in regions with elevated precursor levels. Particularly the nocturnal minimum is too low in the model. While simulated summer noon values approach the observations, winter values remain too low all day, indicating overestimated nocturnal sink processes.

Table 6.3: Relative Bias (%) of the simulated mixing ratios of CO und O_3 with respect to the GAW surface stations in year 2003.

Model run	*O_3*	*CO*
MOZART v1	**-26**	**12**
TM5 v7	**-21**	**-8**
MOCAGE v2	**2**	**-6**

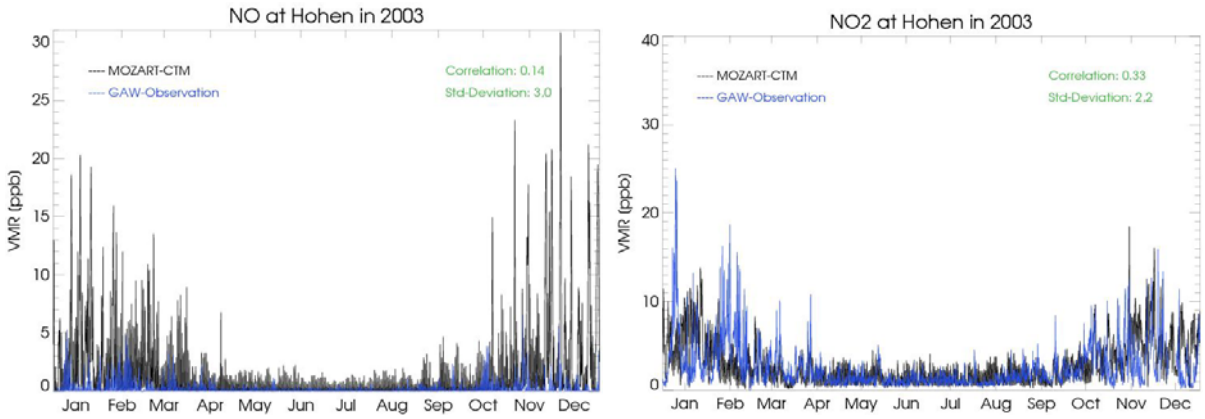


Figure 6.5: NO (left) and NO_2 (right) mixing ratios in year 2003 by MOZART_v1 and GAW observation at Hohenpeissenberg.

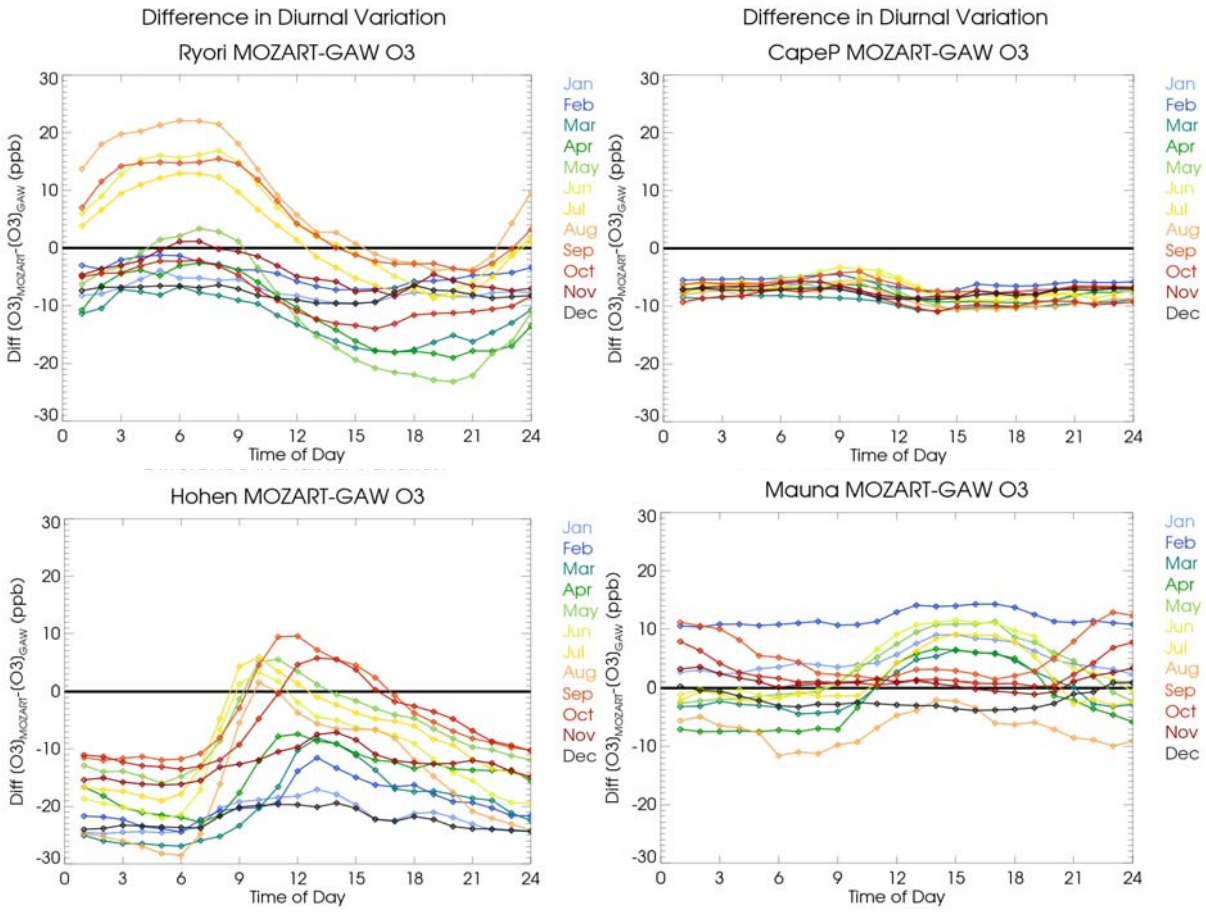


Figure 6.6: Monthly averages of the O_3 bias' diurnal cycle for MOZART_v1 minus GAW observation at Ryori (Japan), Cape Point (South Africa), Hohenpeissenberg (Germany) and Mauna Loa (Hawaii, USA). Times of day are given in Universal time.

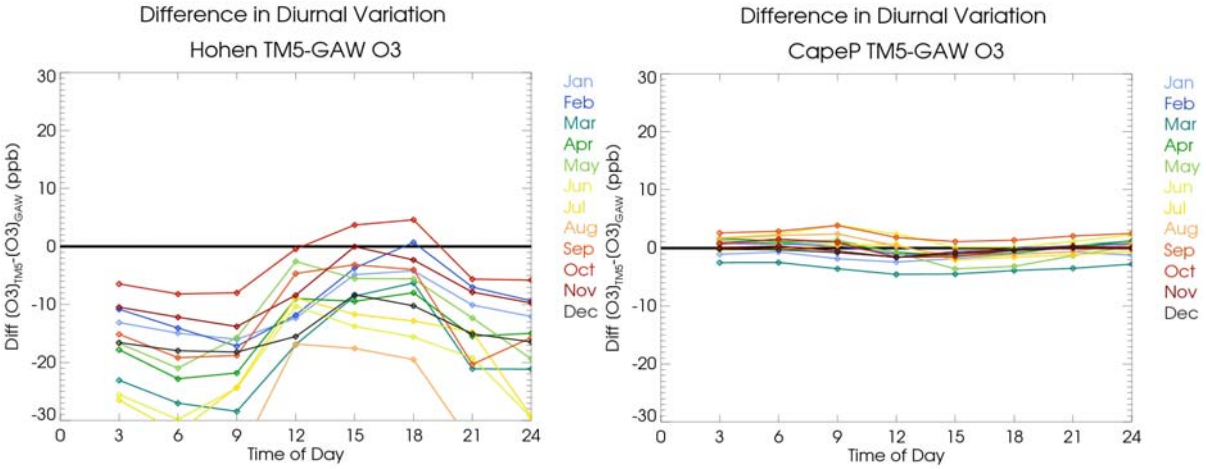


Figure 6.7: As Fig 6.6 but for TM5_v7 minus GAW observation at Cape Point (South Africa) and Hohenpeissenberg (Germany). Times of day are given in Universal time.

Looking to the TM5 offline simulation for Hohenpeissenberg, the phase is slightly shifted and the nocturnal minimum values even more underestimated as by MOZART_v1. MOCAGE data has not yet been processed in this regard. The CO-bias shows a distinct diurnal cycle, too but this must be more related to inadequate source strengths which are assigned to each model grid box but are not necessarily representative for the individual stations. This particularly refers to potential errors introduced by the mean transport time to the GAW site which may well be few hours and indicates the limitations in applying global model data to in-situ stations.

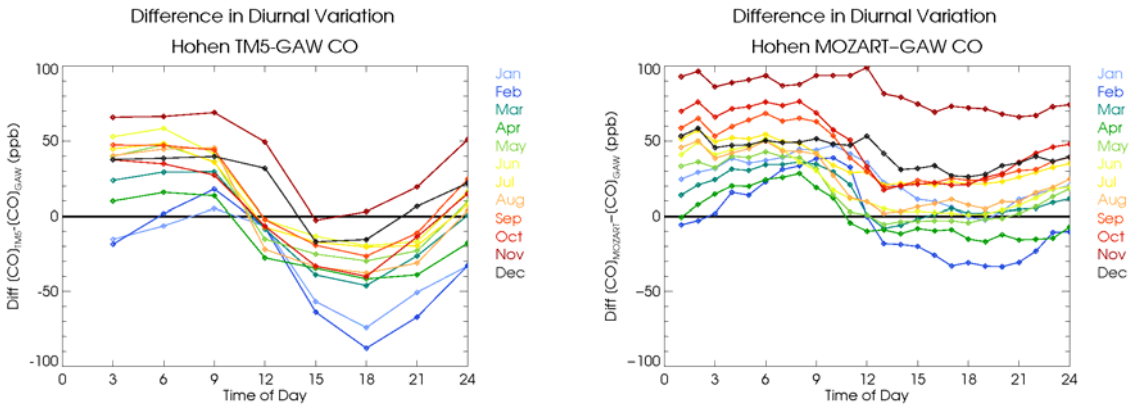
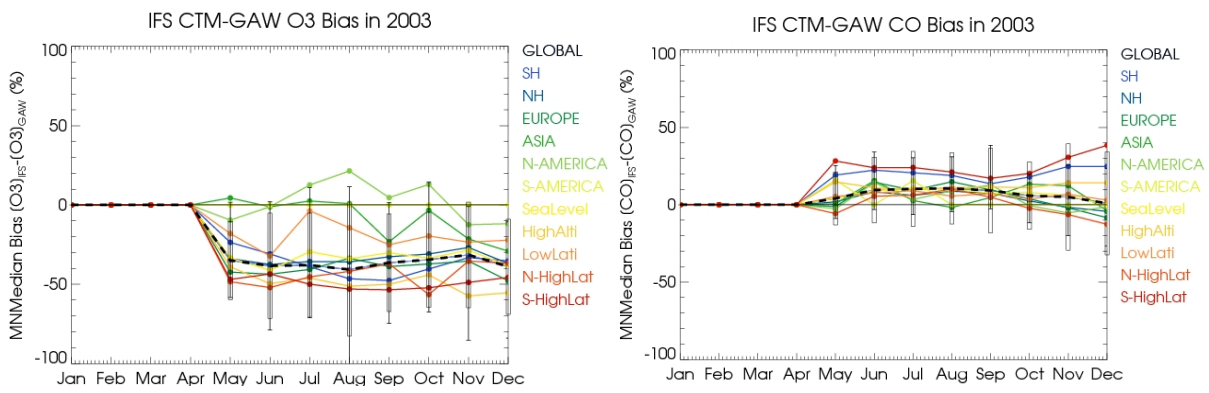


Figure 6.8: As Fig 6.6 but for CO: TM5_v7 (left) and MOZART_v1 (right) minus GAW observation at Hohenpeissenberg (Germany). Times of day are given in Universal time.

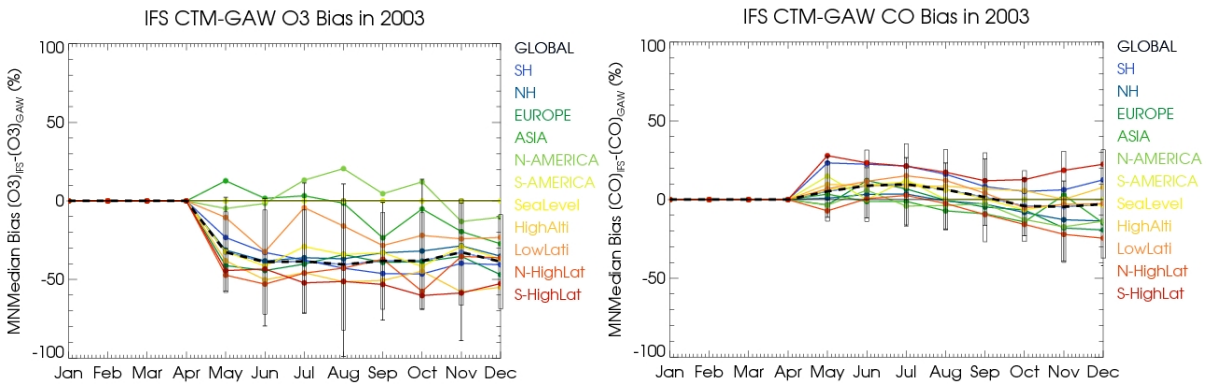
6.3 GEMS-GRG reanalysis

In this section we present the evaluation of the latest runs resulting from the interactive model development-evaluation chain performed in GEMS-GRG. In order to compare the performance of the relevant coupled reanalyses with assimilation, the evaluation results are condensed stepwise: the station ensembles in pre-defined regions are averaged to annual variations of monthly MNMedian bias and further into a global average. The first reanalysis run (eyih, control eyq6) starting only in May 2003, and the second run covering year 2003 and 2004 are shown in Fig. 6.9.

Run eyih (1st GEMS reanalysis)



Run eyq6 (control run for eyih)



Run f026 (2nd GEMS reanalysis)

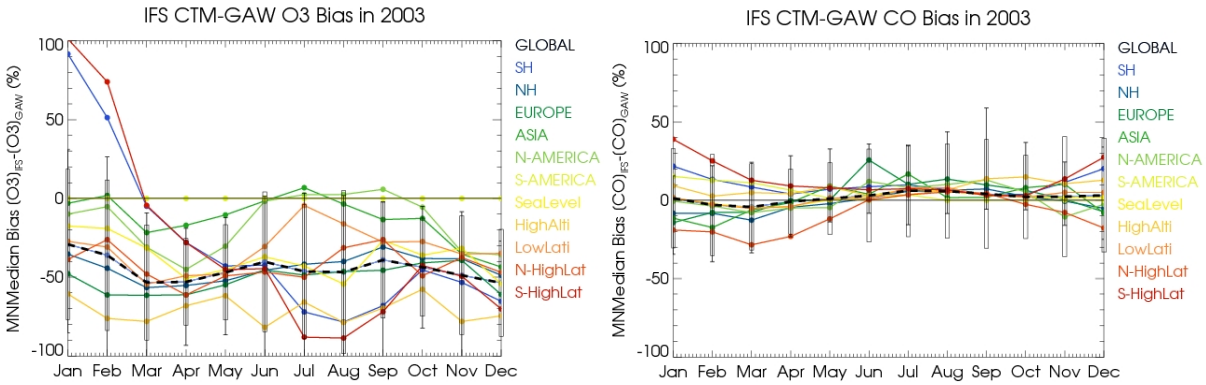
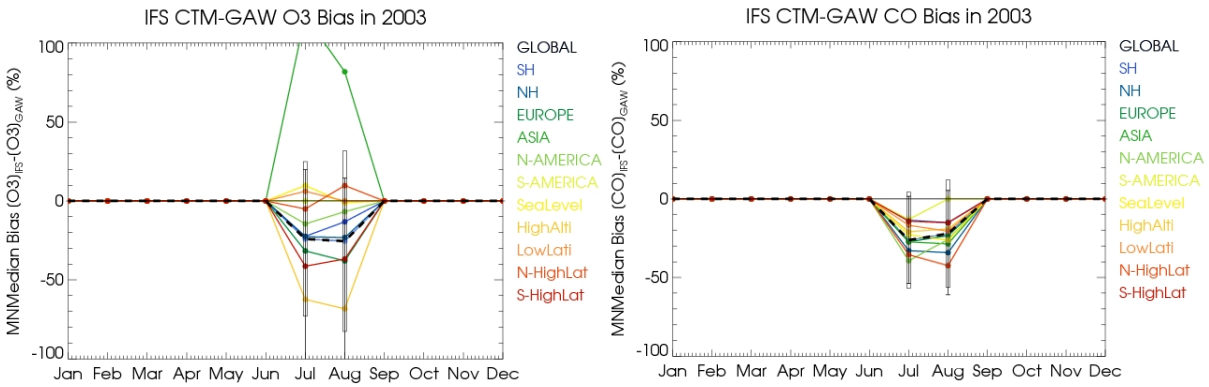


Figure 6.9: Modified normalized median bias in year 2003 for O_3 (left panels) and CO (right panels) from the GEMS reanalysis runs with the MOZART CTM coupled to the IFS. The monthly boxes and whiskers indicate the standard deviation and the 10/90%-iles over the regions' stations collective.

While the first reanalysis run eyih showed little improvement compared to the control run eyq6, the second reanalysis f026 improved significantly for CO. For O_3 hardly any improvement is found, particularly the performance for both ozone and CO degraded in the southern hemisphere. Ozone mixing ratios in the model are generally too low, but CO bias is only few %. With respect to the coupled TM5, we so far only had a look at the trial run eybl (control ez3h) for Jul+Aug 2003 shown in Fig. 6.10. We find that the MNMedian bias for O_3 is less and that for CO significantly larger than for the coupled MOZART runs (1st, 2nd reanalysis). Both O_3 and CO are simulated too low.

Run eybl (coupled TM5)



Run ez3h (control run for coupled TM5)

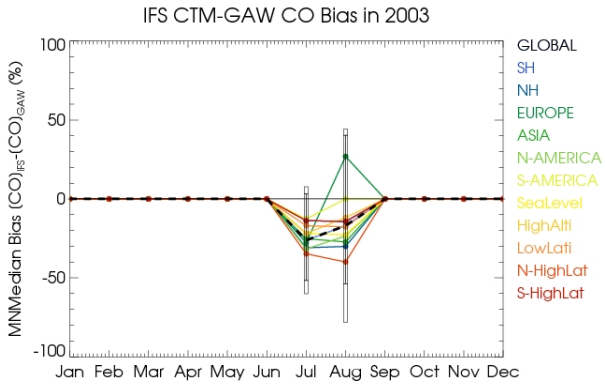


Figure 6.10: As in Fig 6.9 but for TM5 CTM coupled to the IFS and only for the period July-August 2003.

The most condensed overview of the GRG model runs' performance in terms of the Modified Normalized Mean/Median Bias for the year 2003 is shown in Fig 6.11 for all relevant runs. The Figures contain both the reference offline runs (MOZART_v1, TM5_v7 and MOCAGE-v2) as well as the coupled GEMS-GRG reanalysis runs with assimilation performed for year 2003. The comparisons refer to averages over all stations and over the European GAW stations, respectively. Both the mean and median bias are shown as they highlight different aspects of model performance which are more relevant to events or climatological studies, respectively. For both Gases, O_3 and CO the most relevant 2nd re-analysis run f026 is not the best one, particularly it exhibits larger biases than the 1st re-analysis eyih. The CO MNMean bias shows a seasonal variation with relatively higher model CO in summer and lower CO in winter, the MNMean bias for CO is lower than for O_3. The reproduction of short-term pattern in the time series is discussed for the coupled forecasts in the next section. The sophisticated skill scores have not yet been applied to the global models but are discussed with respect to the regional air quality model evaluation.

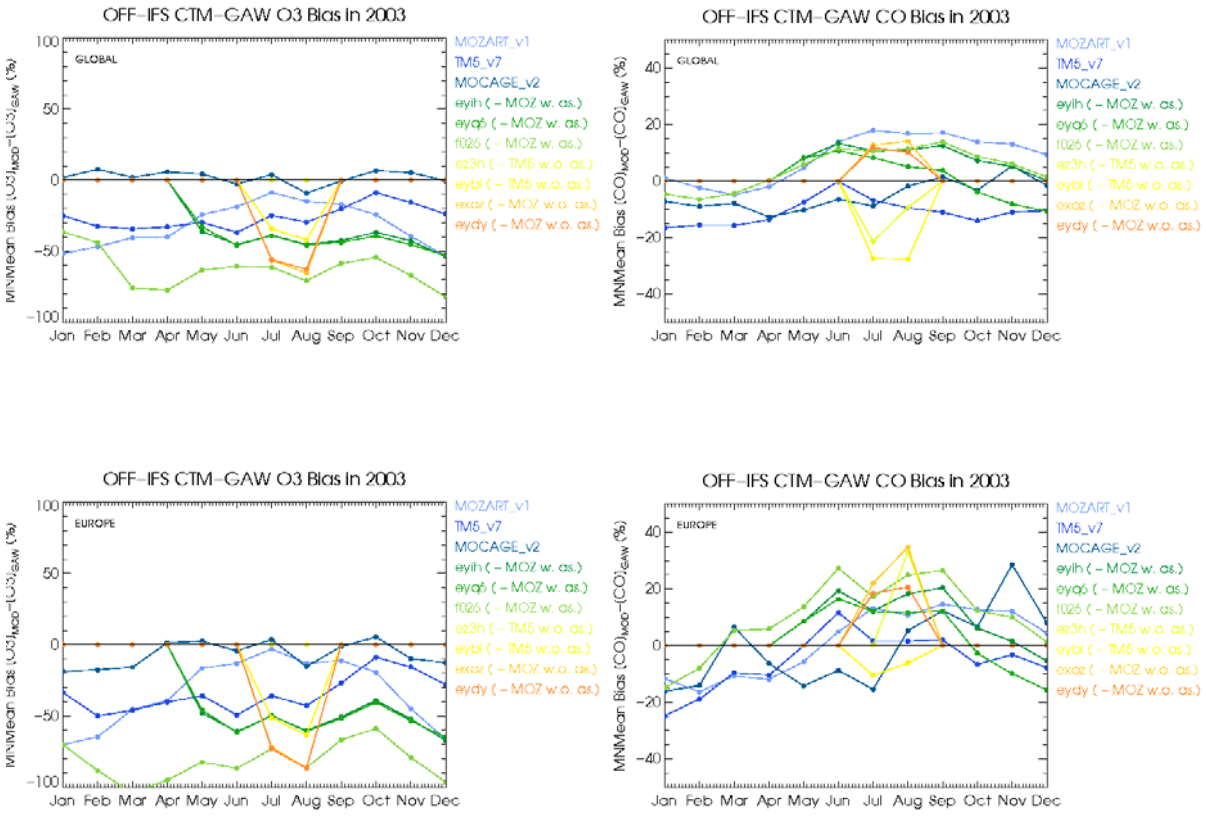
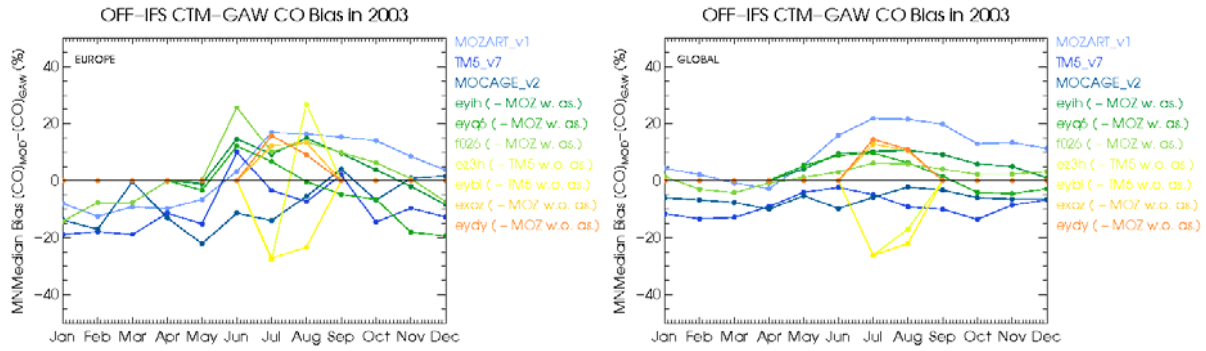


Figure 6.11: Modified normalized mean bias in year 2003 for O_3 (left) and CO (right) for all relevant GEMS-GRG offline and reanalysis runs. Top: global average, bottom: European average.



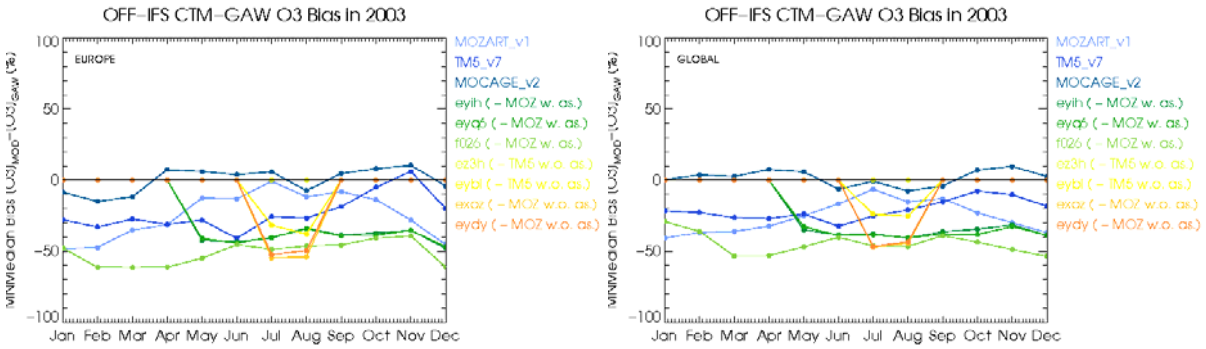


Figure 6.12: As Fig 6.11 but expressed as modified normalized median bias

6.4 GEMS-GRG forecast

Eight GAW stations contributed to near-real-time (NRT) evaluation of the GRG-forecasts. The data from most of these stations are regionally representative after removal of obvious local effects and are thus suitable for evaluation of time-averaged global model output. Even day-to-day variability can be verified in the absence of local sources. The stations Monte Cimone (MCI), Mousalla (BEO), Izana (IZO) and to some extent also Hohenpeissenberg (HPB) are mountain sites with steep relief for which model verification is sensitive to the selected model level. Against this the stations Cape Point (CPT), Neumayer-Antarctica (NEU) and Tamanrasset (TAM) are in relatively flat terrain. Santa Cruz-Tenerife (STC) station is influenced by sub-grid-scale regional pollution. Mousalla (BEO) may be affected by local CO emissions or be subject to intermittent convective orographic transport. Both BEO and Monte Cimone exhibit background conditions but their altitudes deviate roughly 2000m from the model orography. More stations are required to fill geographical gaps. Regarding local characteristics will become more important when town-scale regional air quality models are evaluated in the GEMS successor project MACC (Monitoring of Atmospheric Composition and climate). Yet, only carbon monoxide (CO) and ozone (O_3) have been received in near-real-time from the GAW stations, SO_2, NO_x are to follow in MACC. In table 6.4 the GAW stations are given with their geographic location including altitudes and the corresponding model level (ML) which is selected for evaluation.

Table 6.4: GAW station contributing to evaluation of IFS forecasts.

TaID	Station	Latitude	Longitu	Altitude	M-Alti	ML
HPB	Hohenpeissenberg (Germany)	47.8	11.0	980	968	60
CPT	Cape Point (South Africa)	-34.3	18.5	230	112	56
TAM	Tamanrasset (Algeria)	22.8	5.5	1377	1082	54
MCI	Monte Cimone (Italy)	44.2	10.7	2165	259	46
IZO	Izana (Tenerife)	28.3	-16.5	2367	320	46
BEO	Mousalla (Bulgaria)	42.2	25.4	2925	390	44

STC	Santa Cruz (Tenerife)	28.5	-16.3	50	333	60
NEU	Neumayer (Antarctica)	-70.6	-8.25	42	159	60

GEMS GRG emission data are based on the recently completed emission data sets for the year 2000 from the RETRO project (http://retro.enes.org). Wildfire emission are taken from the Global Fire Emission Database GFED version 2 (containing also aerosols), which spans the period 1997-2004, based on analysis of MODIS fire data and fuel load modelling with the CASA model. Biogenic emissions are from two sources: a decadal mean data set of Lathiere et al., 2004, and emissions from GEIA 1985. All data sets are provided as netcdf files with monthly temporal resolution and spatial resolution of 0.5° x 0.5° (some fields were interpolated from an original resolution of 1° x 1° or 5° x 5°). A mass (flux) conserving interpolation routine for regridding these data onto the required model grid is provided as an IDL routine (ncregrid.pro) or described on the tools page of the RETRO emissions web site (http://retro.enes.org/emissions/tools.html - e.g. the cdo tools). The data sets are available at ftp://ftp.retro.enes.org/pub/emissions/gems/grg. Thus emissions at present enter the IFS model as monthly mean mass per area and time ($kg\ m^{-2}\ s^{-1}$). For CO large scale transport is relevant whereas assimilation from MOPITT satellite probably has a minor effect only. South Africa: Emissions in the Cape region are dominated by traffic and wildfires, both peaking near Capetown (wildfires are speculative since based on 2003).

As shown in Figs. 6.13, the relative bias of the IFS-forecasts with reference to the near real time GAW observations is different at the available stations. Simulated CO is always too high (positive bias) over the SH maritime mid-latitudes (Cape Point - CPT), probably partly due to offset background levels by wildfires, and at the Central European mountain stations Hohenpeissenberg HPB and Monte Cimone MCI, probably due to underestimated emissions. At the Saharan station Tamanrasset TAM, the bias changes from small negative to small positive values throughout the year 2008. The high regional CO levels at Santa Cruz (STC - Teneriffe) are not captured by the model but at the globally more representative station Izana (IZO - Teneriffe) only a small positive bias is found in all seasons. Thus, CO generally tends to be high and better reproduced at lower latitudes.

Against this, ozone levels are always too low (negative bias) except for the beginning of year 2008. The O_3-bias is comparable at all stations except for the Antarctic, where it is most severely underestimated. This may be due to little information from the ozone column assimilation or shortcomings in the (precursor) emissions. It must be noted however, that small scale transport in the steep terrain of the mountain stations may ventilate air-masses to the station which are not representative for the model level chosen on the basis of the station altitude. The negative biases at STC and HPB during the winter periods may be enhanced by regional traffic emissions. At CPT wildfires occurred which cannot be captured as the emissions are based on 2003. Generally low ozone bias is observed at Tamanrasset.

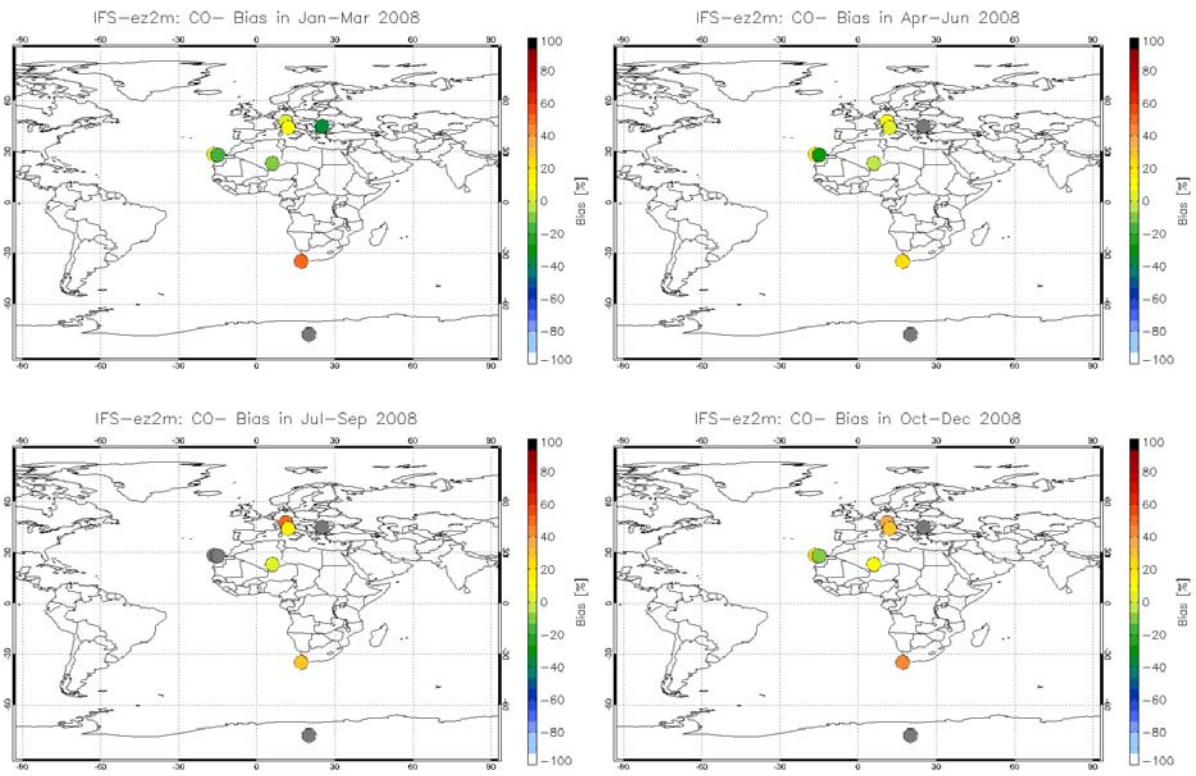


Figure 6.13: Relative bias of CO simulations from run ez2m with respect to NRT GAW surface station observations for the four quarters of 2008. Grey colour indicates no data.

In the time series at Hohenpeissenberg and Santa Cruz, the simulated diurnal cycle is much stronger than at the other stations. It's amplitude is neither confirmed by the observations of ozone nor of CO. In the first case this may indicate shortcomings in the photochemical scheme, precursor emissions or surface sink strength, in the latter the diurnal cycle of emissions seems inadequate for rural environment.

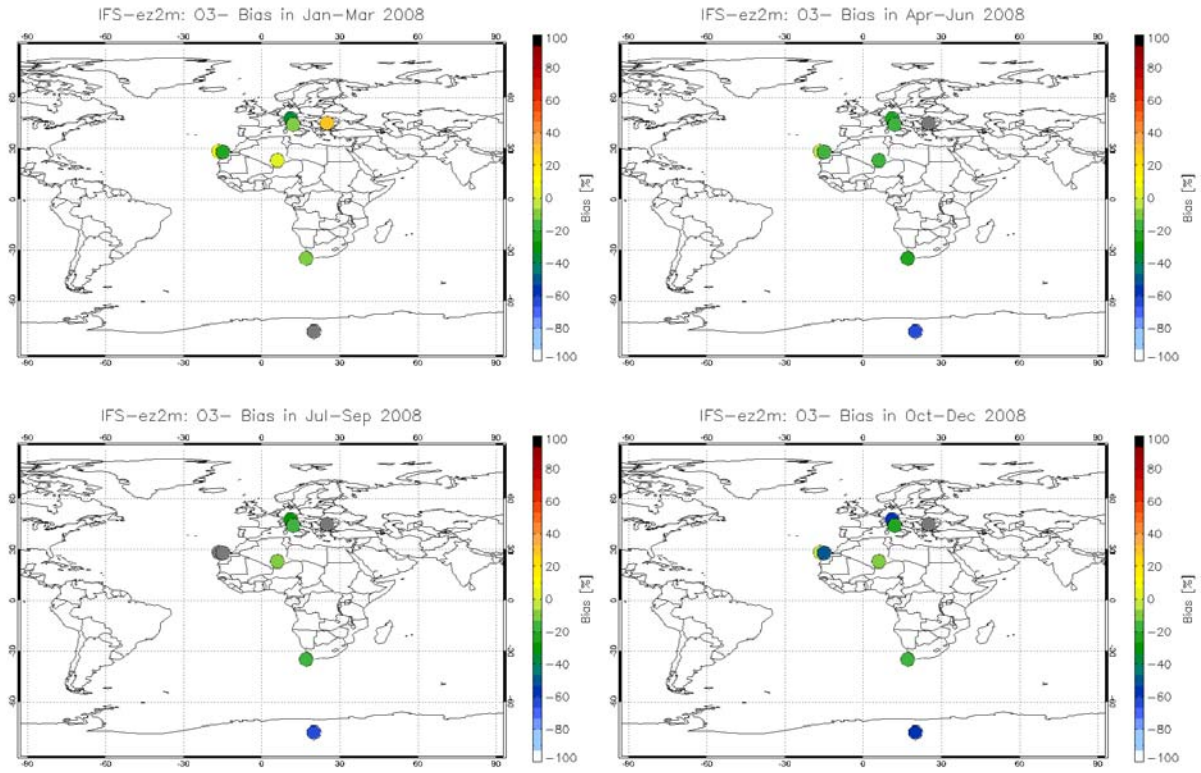


Figure 6.14: Relative bias of O_3 simulations from run ez2m with respect to NRT GAW surface station observations for the four quarters of 2008. Grey colour indicates no data.

The evaluation of the first GEMS NRT run eyr0 showed equivalent results like it's successor ez2m. As example, we show the time series for Oct 07. It came up that the data filtering for unrepresentative local contaminations at CPT was even too restrictive as it removed CO peaks which were reproduced by the global CTM.

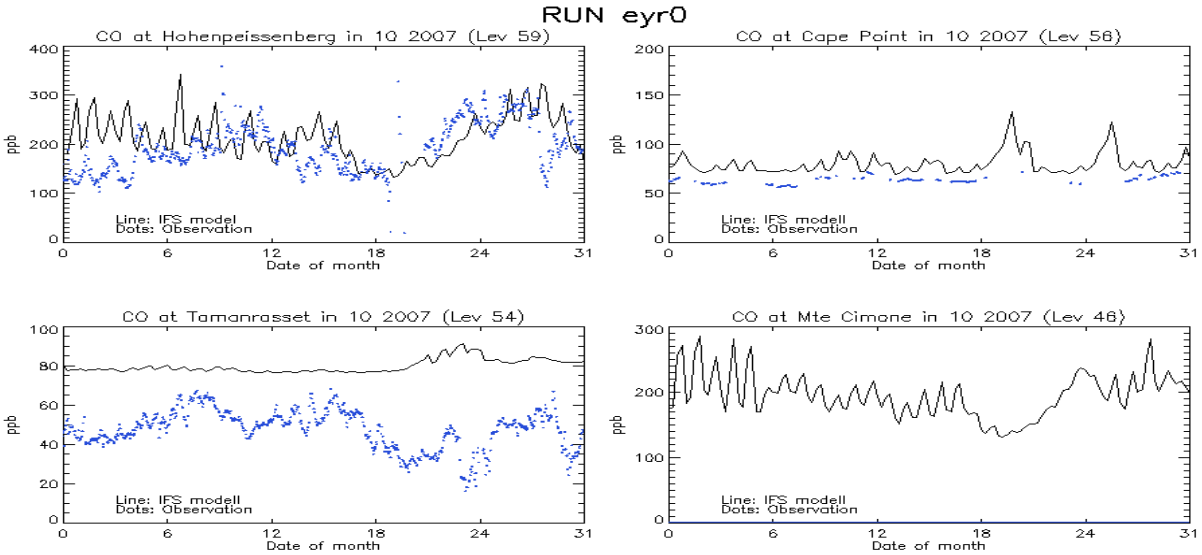


Figure 6.15: Carbon Monoxide (CO) simulations from run eyr0 and GAW near real time observations for the 4th quarter of year 2007.

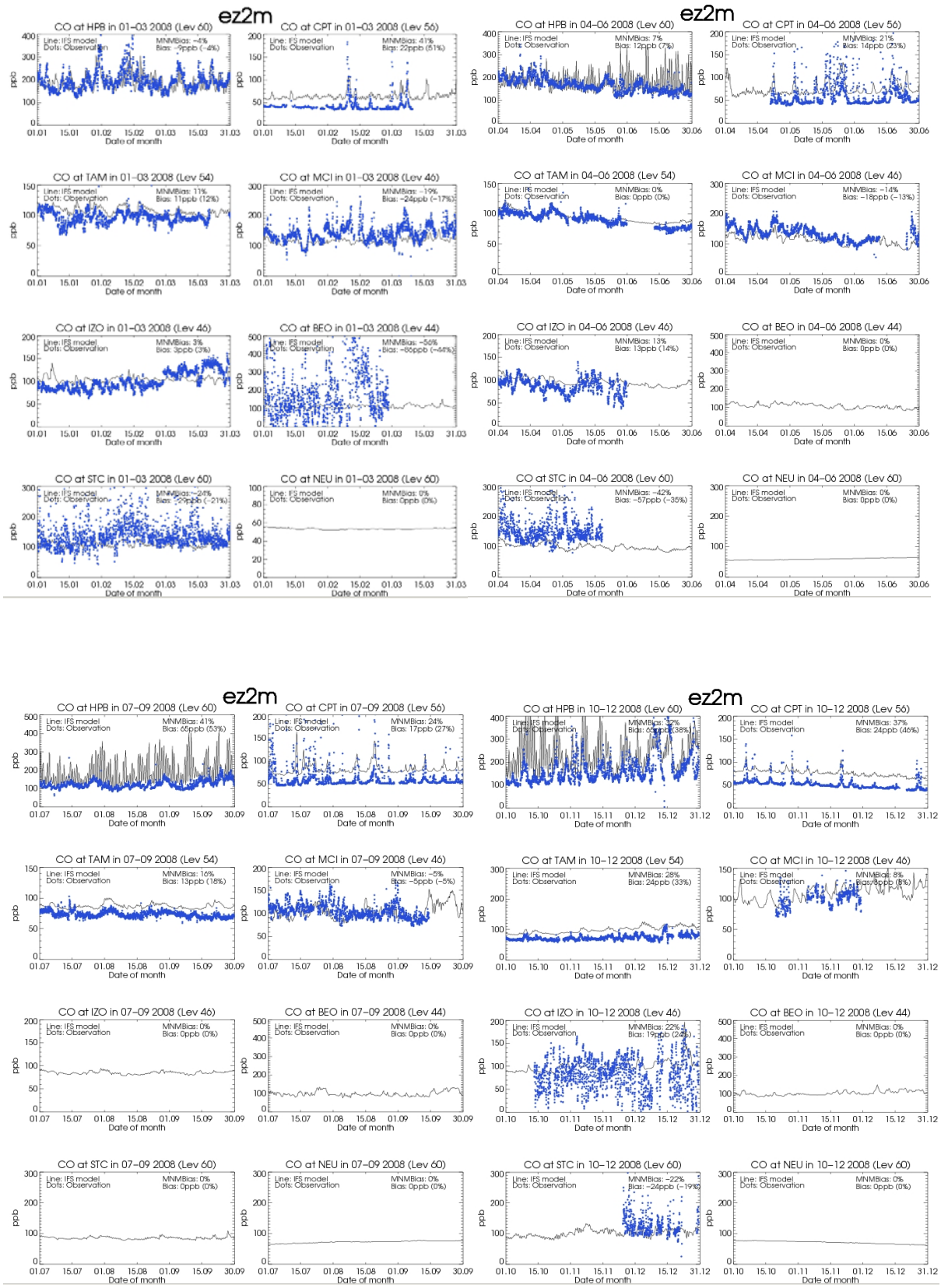


Figure 6.16: Carbon Monoxide (CO) simulations from run ez2m and GAW near real time observations for the 4 quarters of year 2008.

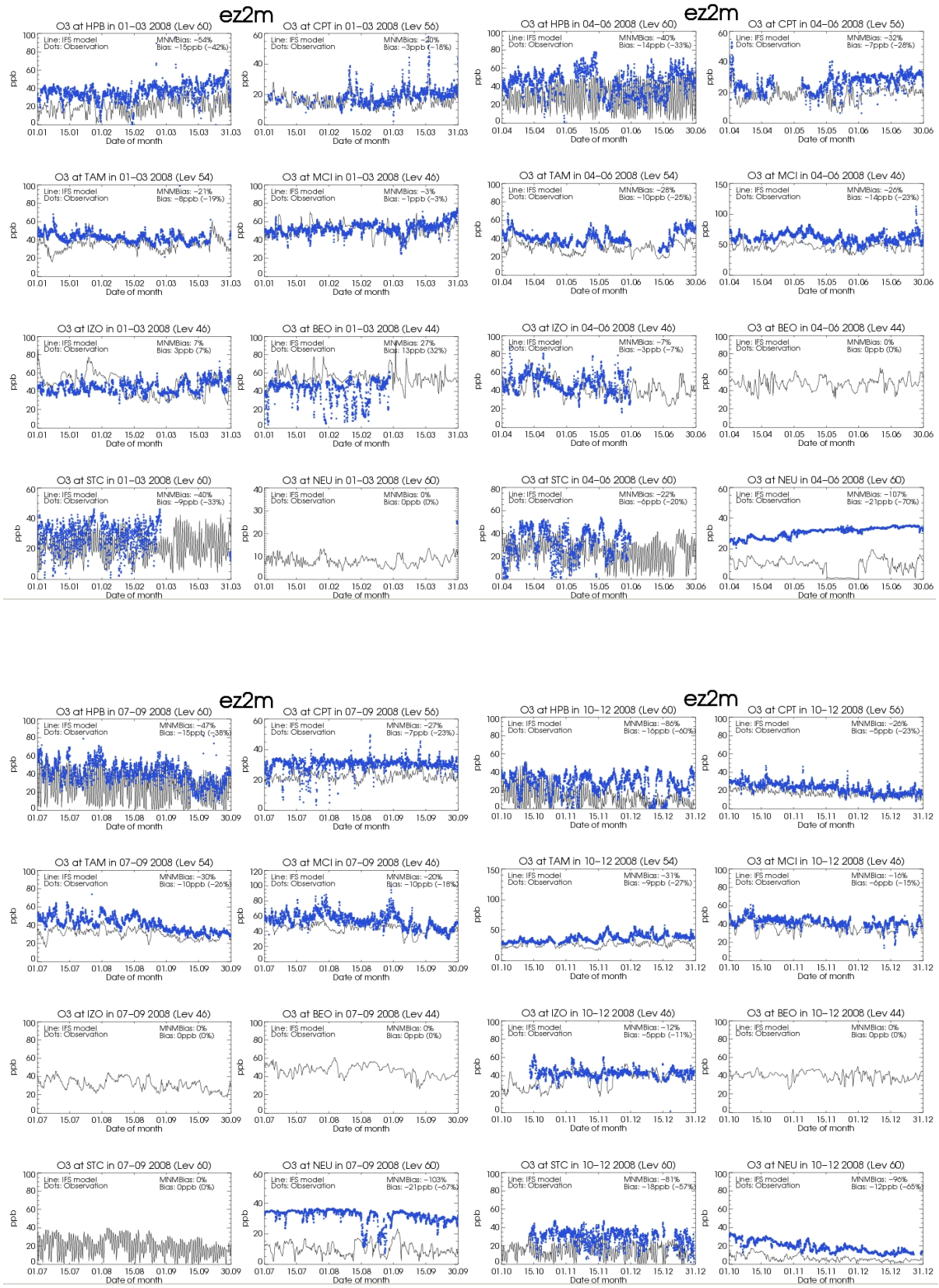


Figure 6.17: Ozone (O_3) simulations from run ez2m and GAW near real time observations for the 4 quarters of year 2008.

In contrast to the diurnal cycle, the variability on few-days timescale is mostly caused by transport and depends on the transport scheme of the driving meteorological model. Generally, the pattern of intra-week variability is often reproduced, but every now and then distinct events are missed, evidently when the corresponding emissions are lacking. Based on 24-h moving averages (discarding the diurnal cycle), Taylorplots indicate only moderate correlation which is not systematically better for any of the stations. The different positions in the diagrams reflect the substantially different conditions at the individual stations. Clearly the global model underestimates temporary CO maxima at stations like CPT and MCI which result from occasional emissions and sub-gridscale transport. Thus CPT is closer to the reference in SH-summer than in SH-winter. The fact that most stations exhibit less CO and O_3 variability than the model expects (stdev < 1) corresponds to their background characteristics. The simulated ozone variability at the remote mountain station Izana (IZO) is too high in winter, maybe due to intrusions of upper tropospheric air (Jan 2008), orographic transport or it's position relative to the PBL top.

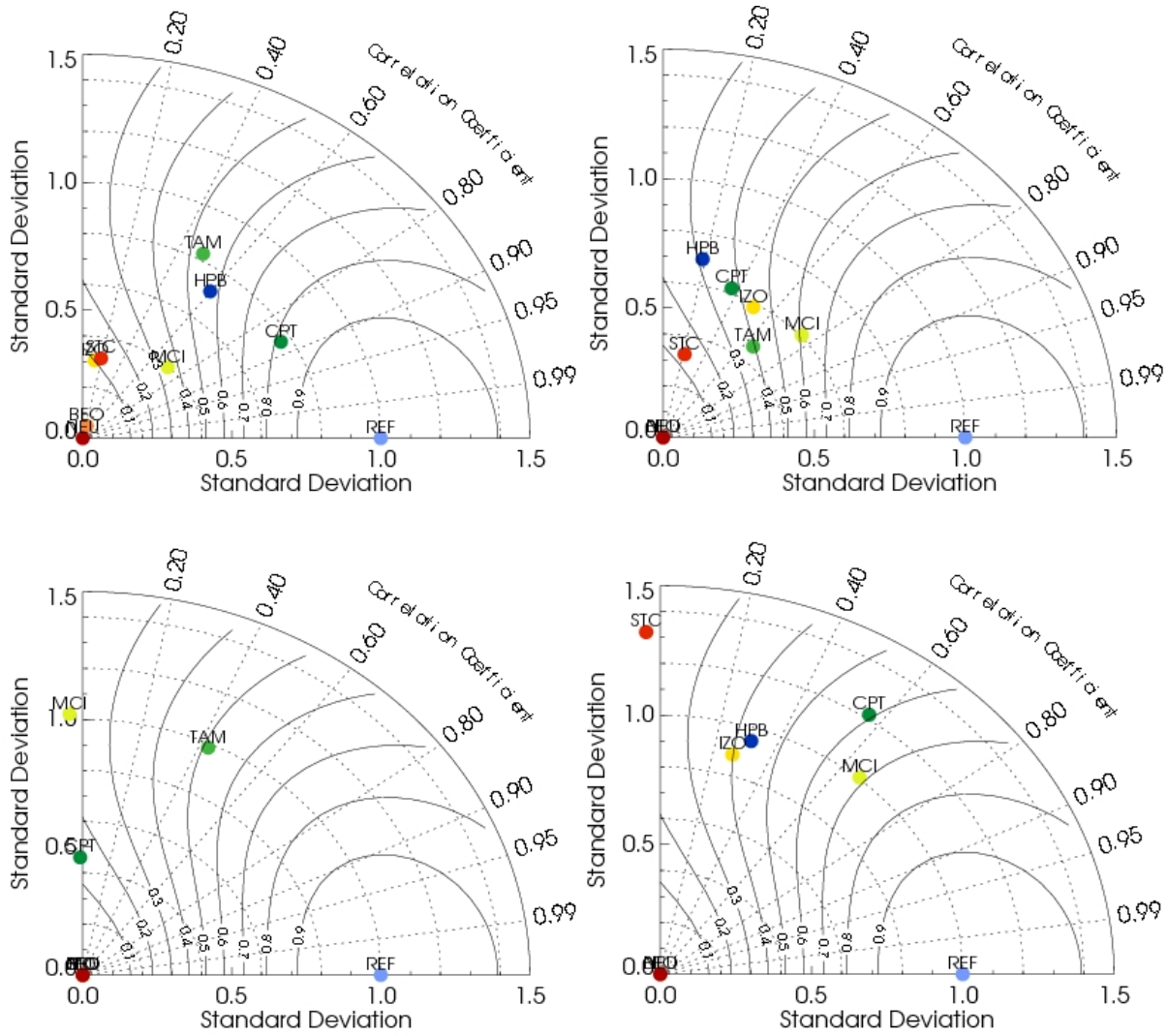


Figure 6.18: Taylor Diagrams MOZART-IFS (run ez2m) CO for quarters of 2008, based on 24-h moving averages (Jan-Mar, Apr-Jun, Jul-Sep, Oct-Dec).

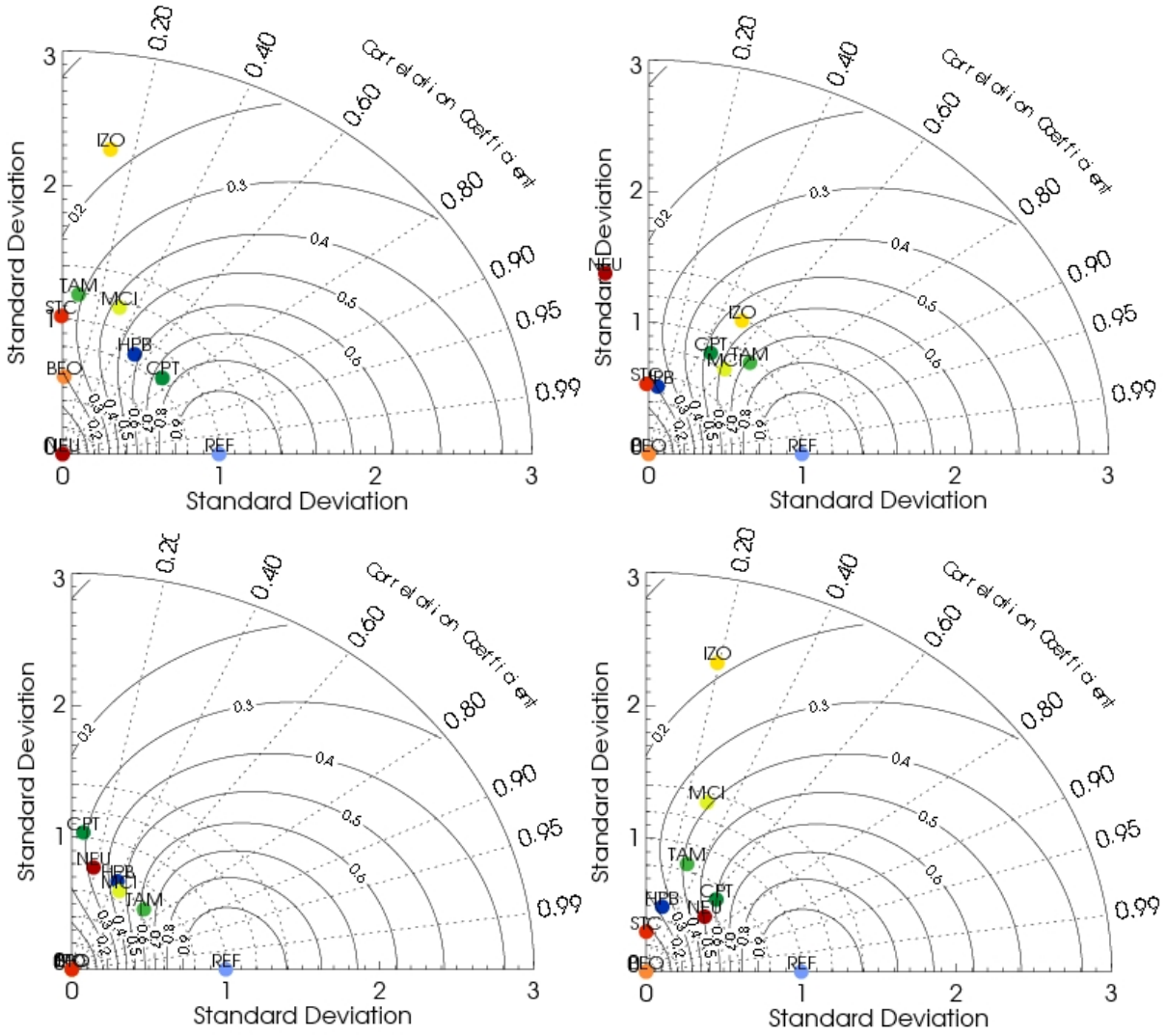


Figure 6.19: Taylor Diagrams for MOZART-IFS (run ez2m) O_3 for quarters of 2008, based on 24-h moving averages (Jan-Mar, Apr-Jun, Jul-Sep, Oct-Dec).

6.5 Conclusions and recommendations

Albeit the initial (offline) versions of the 3 CTMs already showed most of the skill which is now found in the coupled operational forecasts, and each retained its typical characteristics, there have also been several significant improvements with respect to surface level trace gas distributions due to:

- compilation of temporally higher resolved emission and wildfire inventories which improved the model-observation covariance and background levels (especially TM5)
- technical bug-fixes and improvements to the chemical scheme which shifted the concentrations of NO, NO_2 and SO_2 more to what is observed
- improvement of vertical diffusion and transport
- and the coupling to ECMWF's very accurate transport scheme

The offline versions of MOZART, MOCAGE and TM5 as well as the two coupled models that have been evaluated (TM5, MOZART) by GAW surface data are capable to produce global concentration fields in reasonable agreement with surface observations, regarding the known principal limitations imposed by representativity and computational issues. Given that the global fields shall (mainly) provide boundary conditions as a basis for higher resolved regional models, the current system fulfils what was originally

proposed, i.e. to monitor and forecast air quality relevant for people. Still somewhat off background concentrations can be expected to be further improved by assimilation of additional data significant for the lower troposphere.

The present status of the system is defined by the performance of the operational forecasts which have been evaluated with data from 8 GAW surface stations from Oct 2007 onward. In summary, ozone is mostly underestimated, particularly during night and in winter, partly due to low concentrations (emission/formation) of precursor gases, partly due to NO titration. The low nocturnal values indicate a too strong surface (dry deposition) sink in the model. In Europe in summer it appears that CO-formation from organic compounds may be too high while depletion by OH may be too low. At Cape Point station (marine southern hemisphere) the CO background is too low and regional sources (fires, industrial areas) are captured but underestimated (dilution). The model is capable to reproduce the day-to-day covariance for CO in most cases, for O_3 occasional significant deviations occur particularly during photosmog periods. The latter could be shown to improve by running the MOZART model with higher spatial resolution. The amplitude of the diurnal cycle however is generally simulated too high in summer in Europe which may regarded as representative for other source regions as well. However, there is a considerable uncertainty imposed by the selection of the appropriate model level which represents the local conditions of the surface station, in particular for mountain sites. This is certainly an issue for future evaluation in MACC. Ozone is strongly underestimated over the Antarctic continent (Neumayer station) indicating that relevant processes are not yet captured there.

From the surface data perspective, potential for improvements can be identified both in the model's periphery as well as in evaluation strategy/infrastructure:

- the assimilation of ozone columns has not lead to noticeable improvements in the lower troposphere due to missing relevant information at these levels – profiles would help
- interaction between aerosol/radiation and chemical parameters should enter the focus. A real cross-cutting interpretation with the aerosol fields is not yet feasible due to lack of essential parameters in the archived aerosol fields (e.g. particle extinction, profiles)
- evaluations need to be extended to other gases like NO, NO_2, SO_2 in order to tackle model/parametrisation shortcomings but these are routinely reported only by few (almost only European) stations yet:
- Concerning the 'near real time' evaluation strategy of the operational forecasts, more GAW stations are required, particularly in flat terrain (avoiding level sensitivity) and in other continents outside Europe. On the GAW side, efforts are currently pushed, to commit at least the GAW global stations to participation in fully automated evaluation. A closer collaboration with the regional modelling RAQ sub-project is neccessary.
- The format, pre-processing and archiving/retrieval of model data should finally be harmonized as an overarching standard between all model runs as already done for the sub-set of GRG IFS runs.
- The information content of the different evaluation metrics with respect to climatological (monitoring) or case-related (forecasting) application has not yet been fully exploited.
- Generally, a closer collaboration between the sub-groups (in particular GRG, AER, RAQ) shall be established.

6.6 References

1. WMO/GAW Strategic Plan: 2008-2015 - A Contribution to the Implementation of the WMO Strategic Plan: 2008-2011 (WMO TD No. 1384), http://www.wmo.int/pages/prog/arep/gaw/gaw-reports.html
2. GAW letters of the German Meteorological Agency: http://www.dwd.de/bvbw/appmanager/bvbw/dwdwwwDesktop
3. Flemming, J. A. Dethof, H. Flentje, V. Huinen, P. Moinat, M.G. Schultz and O. Stein, Coupling global chemistry transport models to ECMWF's integrated forecast system for forecasts and data assimilation, ECMWF Technical Memorandum 5xx (http://www.ecmwf.int/publications/library/do/references/list/14), t.b. submitted to ACP in 2009.

7. Evaluation with EMEP ground-based data

Contributors:Katragkou Eleni, Melas Dimitri, Zerefos Christos (NKUA, Athens, Greece), and the modeller teams

This work has been accomplished in the framework of the GEMS project (Global Earth-system Monitoring using Space and in-situ data). It concerns validation of three global chemistry transport models (CTM), MOCAGE, MOZART and TM5 which have performed simulations for the Global Reactive Gases (GRG) subproject of GEMS. The GRG subproject of GEMS aimed at developing a pre-operational data assimilation system for chemically reactive gases within the ECMWF's Integrated Forecasting System (IFS) capable of providing global products for the troposphere and stratosphere on a daily basis [Hollingsworth et al, 2008].

The measurements used for validation were taken from the EMEP database and consist of hourly surface ozone time series. The runs validated are stand-alone runs performed for the reference years 2003 and 2004 and for selected IFS coupled model experiments.

In section 1 is given a short description of the three global models and the measurements that were used. Section 2 describes in detail the statistics that were calculated for model validation. In section 3 are presented in detail the calculated statistics for the three models for different experiments. The three separate Annex (MOCAGE, MOZART and TM5) that accompany this chapter provide monthly validation graphs for the three models. In section 4 are discussed the results of validation giving special emphasis on seasonal variations of model performance and day- vs. nighttime model performance. Annex A provides all available EMEP station used for validation and Annex B presents the runs that have been validated accompanied with a short description concerning their technical implementation.

The plots presented in the Annex and the calculations of statistics were produced by scripts written in IDL language.

7.1 Datasets and methodologies

7.1.1 Datasets

The European Monitoring and Evaluation Programme (EMEP) is a scientifically based and policy driven programme under the Convention of Long-range Transboundary Air Pollution (LRTAP) for international co-operation to solve transboundary air pollution problems. The EMEP website (http://www.emep.int/) is intended to facilitate the understanding and accessibility of the information compiled under the Co-operative Programme for Monitoring and Evaluation of the Long-range Transmission of Air Pollutants in Europe. The EMEP data on the European regional air quality concern acidification, eutrophication, ground level ozone, heavy metals, persistent organic compounds and atmospheric particles. Parties to the LRTAP perform monitoring at regional monitoring sites across Europe. The data are subject to national quality assessment prior to submission to the EMEP Chemical Coordinating Centre (CCC) at the Norwegian Institute of Air Research (NILU). The submitted data are further assessed by the EMEP-CCC in collaboration with the data originators before they are reported on an annual basis.

Surface ozone measurements used in this work were downloaded from the EMEP database for the reference years 2003/2004 and used for GEMS-GRG model evaluation. EMEP stations are all located in rural areas in order to be unaffected by local anthropogenic emission sources. They are spread over Europe, the coverage being denser over central and northern Europe (Figure 1). Annex A lists all available EMEP stations and their geographical location.

Measurement data are available for 131 stations of 27 different countries. All stations use the UV-absorption method to measure ozone for the reference years. Information about the ozone data quality, calibration and maintenance procedures as well as statistical summaries and geographical distributions are given in *Aas and Hjellbrekke, 2000* and *Hjellbrekke and Solberg, 2005*.

All EMEP stations measure surface ozone concentrations. Since model levels are terrain following, one would naturally use the surface model level O3 concentrations to compare with measurements. This methodology is acceptable for the majority of EMEP stations; however there are some considerations whether surface model levels should be used for comparison with measurements performed in stations located at higher altitudes. The main problem is that orography is not correctly resolved by the model: in the case of mount station Sonnblick, AT for example, the model topography indicates an altitude of 1150 m while the station has an actual altitude of 3106 m. One suggested way of account this problem is using pressure as an indicator of selecting a more suitable model level to sample.

This will not guarantee necessarily an improvement in metrics indicating model performance; however, it should be checked as an alternative to comparison by default to the surface model level. In the following analysis we extracted for every grid cell corresponding to a measurement station the model orography Z_{mod} and we compared it with its actual altitude of the station Z_{meas} and the corresponding pressures assuming a standard atmosphere (P_{mod}; P_{meas}). A model level was assigned to each stations so that $P_{mod} \approx P_{meas}$. This task has been performed using the MOZART model in the IFS coupled configuration. In Annex A is shown the assignment of each station to a model vertical level. Naturally, for a different model configuration, the assignment would have been different. To avoid this complication since we work with three different models, with different configurations, it was decided in the current analysis to include only those stations that were assigned to the lower model level, hereafter referred to as L60.

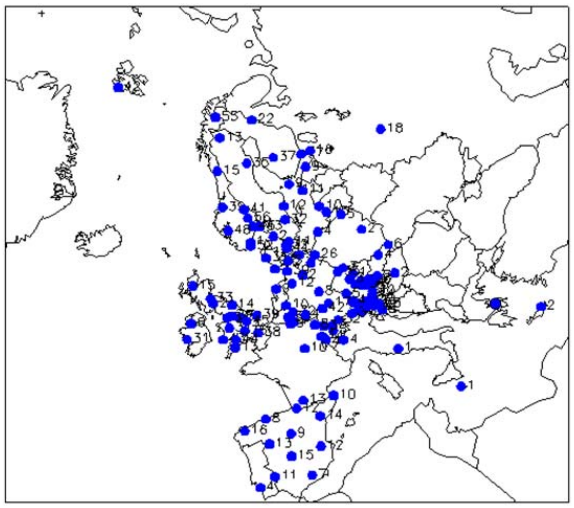


Figure 1 Map with EMEP stations

7.1.2 Methodologies

For the evaluation of model performance with respect to surface ozone in GEMS-GRG, we used the metrics described in Annex 7. The contribution of NKUA in the GEMS-GRG model validation is concerned with surface ozone. In this report were validated the TM5, MOCAGE and MOZART stand-alone runs for the reference years 2003 and 2004 as well as selected IFS-coupled model versions. At first stage observational data were downloaded from EMEP and were homogenized: They were all corrected to

correspond to GMT, missing values were treated separately and stations not fulfilling the criterion of 75 % data completeness were excluded from data analysis. All data were treated programs written in IDL language. All model data extracted for every EMEP station were downloaded from ECMWF servers. Model data were available either on txt- or ncdf format. Programming routines were prepared in IDL to extract model data and homogenize them as well and also calculate statistics. The graphics presented in this report were all prepared with IDL routines that have been developed so that they run semi-automatically, i.e. the user has to provide only the name of experiment and period to be investigated to produce the plots.

7.2 Offline simulations

In this chapter selected runs of GEMS-GRG are validated using the statistics presented in detail in chapter 2. The statistics are calculated for every month using mean daily values. Fractional Gross Error (FGE) was calculated also for day- and nighttime surface O3 to assess model capability to reproduce day- and nighttime ozone chemistry. All statistics are calculated for every selected EMEP station and are plotted over a European map for a more concise presentation of results. Annex TM5, MOZART and MOCAGE, show the plots for all evaluated model runs.

7.2.1 TM5

EXP: v3

Experiment v3 is the reference run for 2003. Fractional Gross Error (FGE) is calculated for every month (Jan-Dec) and provides a measure of the overall model performance. FGE is usually expressed as a percentage: when FGE < 35% the model performance is considered satisfactory. Using mean daily model and observation of surface ozone values, we calculate FGE for every month and every station separately. The results are plotted on a European map. With the help of a relevant color-scale (0-100%) it is easy to distinguish between regions with better/worse performance. As seen in Fig. 1_ ANNEX_TM5 the model performs in an acceptable way throughout the year, with the exception the months of November and December during which FGE is ~40 % on average. During warm months (May-September) the model performs somewhat better and monthly average FGEs have low standard deviations. On the contrary, for cold months deviations of the average monthly FGEs are larger, meaning that FGEs of different stations in one month vary more in winter- than in summer-time.

In order to examine whether the model captures correctly day- and nighttime O3 chemistry we calculate FGE again using only day- and nighttime O3 values (Fig. 2 and Fig. 3_ANNEX_TM5). We assume that the mean daytime O3 concentration is the average O3 between 12:00-15:00 while nighttime O3 the average between 00:00-3:00. These criteria apply to the whole European domain independently of geographical location and season. As indicated by FGE calculation the model performs slightly better when considering only daytime values while the metrics indicate unacceptable model performance when using only nighttime values. Interesting features are the O3 daytime under-prediction over all European surface stations during the spring months (March-April) and the severe surface ozone under-prediction in central Europe in August 2003, when exceptionally high ozone concentrations were recorded.

Modified Normalized Mean Bias (MNMB) is calculated for experiment TM5.v3. As shown in Fig. 4_ANNEX_TM5 ozone is over-predicted in November – month with the worst overall performance as indicated also by FGE calculation – for the majority of EMEP stations. For the rest of the year, ozone is over- or under-predict depending on season, with UK mostly over-predicting and central Europe mainly under-predicting. The same conclusions hold true when examining Mean Bias (Fig. 5_ANNEX_TM5).

Calculation of RMSE (Fig. 6_TM5_ANNEX) indicates errors that range between 8-10 ppb on average for all L60 stations, with no pronounced seasonal behavior.

EXP: v7

TM5.v7 is the second stand-alone run for the reference year 2003. As shown in Fig. 7_ANNEX_TM5 FGE is within acceptable limits for all months but November and December. FGE calculated using daytime ozone values is better or similar to FGE calculated with mean daily values, with the exception of January, November and December (Fig. 8_ANNEX_TM5). Monthly FGE values deteriorate dramatically when nighttime values are taken into account (Fig. 9_ANNEX_TM5). The simulation of night chemistry is worse than this of TM5.V3. Concerning TM5.v7 bias (Fig.10 and Fig. 11_ANNEX_TM5): October and November are months with mostly positive bias for the majority of EMEP stations. For the time period February to May surface ozone is mostly underestimated. The root mean square error increases from January to April to reach 12 ppb and then gradually falls remaining to around 9 ppb on average for the time period September to December. Due to lower concentrations of surface ozone, the relative error is getting very high during the last two months, although RMSE lower (Fig. 12_ANNEX_TM5).

EXP: v9

The stand-alone TM5.v9 run has been performed for the time period May to August 2004. FGE calculation indicates acceptable model performance throughout the period investigated, for almost every surface EMEP station (Fig. 13_ANNEX_TM5). Again using only daytime ozone concentration improves somewhat FGE (Fig. 14_ANNEX_TM5). In comparison to TM5.v3 run it has comparable results with slightly higher FGEs. The model performs clearly worse when taking into account only night-time surface ozone concentrations (Fig. 15_ANNEX_TM5). Calculation of MNMB indicates mostly underprediction of surface O3 during May, which however turns until August to slight over-prediction with the exception of central Europe, where ozone is under-predicted during the whole period examined (Fig. 16_ANNEX_TM5). When comparing the two experiments v3 vs v9, one should note that the two runs correspond to two different reference years. Mean Bias is also calculated and presented in Fig. 17_ANNEX_TM5 for the time period investigated. Although MB has high negative values over central Europe during warm months, when inter-comparing with MNMB plots we can conclude that the underprediction is not so severe since O3 concentrations are quite high. TM5.v9 exhibits the same behavior as TM5.v3 with respect to RMSE with values ranging between 8-10 ppb (Fig. 18_ANNEX_TM5).

7.2.2 MOZART

EXP: v1

The first model stand-alone run covers the time period January to December 2003. Based on calculation of the FGE of surface ozone MOZART.v1 has an acceptable performance from May to September with average FGE < 35 % and deteriorates during cold months. January and December yield an average FGE ~ 60 % (Fig. 1_ANNEX_MOZART). Concerning calculation of day- and nighttime surface ozone values: the metrics are comparable when taking into account the whole-day and only daytime concentrations (Fig. 2_ANNEX_MOZART) but deteriorate when performing calculations of nighttime ozone (Fig. 3_ANNEX_MOZART).In winter surface ozone is mostly underestimated as can be seen in Fig. 4 and Fig. 5_ANNEX_MOZART that show MNMB and MB for each month of the year. In summer some stations over Benelux and UK exhibit positive bias but otherwise the majority of L60 stations still underestimate surface ozone. Calculation of RMSE indicates higher errors for the time period January to May (12-16 ppb) which improve for the rest of the year to 9-11 ppb (Fig. 6_ANNEX_MOZART).

EXP: v9

MOZART.v9 is the reference stand-alone run for the year 2004. Calculation of FGE indicates acceptable model performance for the period May-September. During winter months FGE ranges between 40-60 % (Fig. 5_ANNEX_MOZART). In comparison to MOZART.v1 it performs slightly better during the warm period, however months Jan–Feb and Nov-Dec yield considerably higher FGE. During winter surface ozone is mostly underestimated over all European L60 stations while form June to August mostly overestimated (Fig. 10 and Fig. 11_ANNEX_MOZART). During the transition months (May, September, October) southern and western European stations over-estimate surface ozone, northern and eastern European stations underestimate surface ozone. Average RMSE indicates higher errors for the period January to March (16 ppb) and July to August (14 ppb). For the remaining year RMSE ranges between 10-12 ppb (Fig. 12_ANNEX_MOZART).

EXP: v10-2003

MOZART.v10 is performed as a reference stand-alone run for the year 2003. As seen also in experiment v9 months May-September have an acceptable performance with FGE < 30 %. During transition months the performance gradually deteriorates reaching the highest FGE during the period November to February (Fig. 13_ANNEX_MOZART). Daytime surface ozone is simulated slightly better than whole-day average ozone with the exception of months January, November and December (Fig. 14_ANNEX_MOZART). During night-time the model in this configuration seems to totally fail to reproduce surface ozone concentrations, during the whole year (Fig. 15_ANNEX_MOZART). The high errors calculated during colder months (January-April and October-December) are due to severe underestimation that reach up to 70-80 % MNMB (Fig. 16_ANNEX_MOZART) or -8 to -10 ppb MB (Fig. 17_ANNEX_MOZART). On the contrary, during summer months, ozone is almost over all EMEP stations overestimated with MNMB ~25% in July corresponding to MB of ~ 10 ppb. Obviously during colder months surface O3 concentrations are lower and 8 ppb of absolute bias correspond to 70-80 % relative bias, while during summer the same absolute bias is only ~25% of relative bias. Such results should be interpreted with caution, and apparently both absolute and relative values should be inspected. Calculation of RMSE shows that the error ranges during all year between 10-15 ppb which yields considerably higher winter relative error (Fig. 18_ANNEX_MOZART).

EXP: v10-2004

In comparison to experiment v9 for the year 2004, v10 has a comparable performance. May to September yield acceptable FGE, December and January fail to reproduce surface O3 concentrations (Fig 19_ANNEX_MOZART). Calculation of statistics using daytime O3 values produces the same results, with the exception of months January, November and December (Fig 20_ANNEX_MOZART). Night-time ozone is poorly reproduced by this model configuration as well (Fig 21_ANNEX_MOZART). From the beginning of the year up to April the model underestimates surface O3, during the transition period some stations underestimate (northern Europe) some overestimate O3, and during summer O3 is overestimated in the majority of stations (Fig. 22 and Fig 24_ANNEX_MOZART). RMSE ranges between 8-15 ppb, with central Europe and Scandinavia having the highest errors (Fig. 25_ANNEX_MOZART).

7.2.3 MOCAGE

EXP: 60LEV02

This experiment is the reference run for the year 2003. It has a satisfactory performance for the whole period except of months November and December for which FGE > 35 % (Fig. 1_ANNEX_MOCAGE). According to FGE calculation which can be considered the index of the overall model performance, the

model is simulating better surface O3 during summer months (May-September). When calculating FGE taking into account only daytime ozone values the statistics are getting slightly better (Fig. 2_ANNEX_MOCAGE). On the contrary nighttime chemistry does not work sufficiently well and FGE during nighttime are for every month higher than 40 % reaching up to 65 % in November (Fig. 3_ANNEX_MOCAGE). Concerning model bias: The sign of the bias for each station changes throughout the year, meaning the every station may over- or underestimate ozone depending on season (Fig. 4 and Fig. 5_ANNEX_MOCAGE). Strong negative bias is seen over central Europe in summer. During March the bias is getting mostly negative while in October mostly positive. RMSE ranges between 10-14 ppb which gives worse relative statistics for winter months (Fig. 6_ANNEX_MOCAGE).

EXP: HTAP_ES1

Experiment HTAP_ES1 is the MOCAGE reference stand-alone run for the year 2004. It covers the time period June-September 2004. FGE indicates acceptable model performance (FGE < 35 %) which is stable throughout the period examined (Fig. 7_ANNEX_MOCAGE). Fractional gross error calculated for daytime surface ozone is almost identical to the one calculated for the daily averages (Fig. 8_ANNEX_MOCAGE). On the contrary the model does not simulate in a satisfactory way nighttime ozone (FGE > 50%, Fig. 9_ANNEX_MOCAGE). Model bias differs for every EMEP station, it is mostly negative for the majority of central European and Scandinavian EMEP stations (Fig. 10 and Fig. 11_ANNEX_MOCAGE). The average RMSE over all EMEP stations ranges between 11-12 ppb (Fig. 12_ANNEX_MOCAGE).

7.3 GEMS-GRG reanalysis

7.3.1 MOZART

EXP: eyih

The first reanalysis experiment covers the time period from May to December 2003. Fractional gross error is below 30-35 % for the time period May to September and increases gradually from October to December, as indicated in Fig. 25_ANNEX_MOZART. Daytime surface ozone gives on average lower FGE, in contrary to nighttime ozone that yields even 30 % higher FGE (Fig. 26 and Fig. 27_ANNEX_MOZART). Surface ozone is mostly underestimated for the whole period as shown in Fig. 28 and Fig. 29_ANNEX_MOZART. Average MB is ranging between -6 to -9 ppb. Root mean square error mostly ranges between 10-12 ppb for the period studied (Fig. 30_ANNEX_MOZART).

EXP: eyq6

The experiment eyq6 is the control experiment for eyih, i.e. it is identical to eyih without O3 and CO assimilation. Therefore comparing the two experiments, we quantify the impact of assimilation in the coupled system. As shown in Fig. 31 – Fig. 34_ANNEX_MOZART, the metrics yield almost identical results with eyih, thus the impact of assimilation on surface O3 is negligible. Hourly time-series of surface O3 are as well almost identical for the eyih and eyq6 experiments (not shown here).

EXP: f026

The highest FGE are found during transitional months (March-April) (Fig. 35_ANNEX_MOZART) and especially for central European and north Scandinavian EMEP stations. Since ozone concentrations have been extracted on a 6-hourly basis, the day- and nighttime ozone has been extracted as the midday and midnight ozone respectively. The model performance is getting considerably better when including in the analysis only daytime concentrations and dramatically deteriorates when using only nighttime values (Fig.

36 and Fig. 37_ANNEX_MOZART). This model version severely underestimates surface ozone during the whole simulated period (Fig. 38 and Fig. 39_ANNEX_MOZART). During March and April MB reaches -20 ppb, and ranges between 7 – 15 ppb for the rest of the period examined. Root mean square error has the same seasonal behavior (Fig. 40_ANNEX_MOZART)

7.4 GEMS-GRG forecast

7.4.1 TM5

EXP: f1d7

The coupled TM5 system in the forecast mode was tested for the time-period June-July 2004. This model version is based on TM5.v9. A comparison of the stand-alone TM5.v9 and the coupled forecast TM5.f1d7 shows that TM5 performs slightly better in the stand-alone version as indicated by FGE calculations (Fig. 19_ANNEX_TM5). For daytime ozone the model seems to perform somewhat better (Fig. 20_ANNEX_TM5) than when taking into account the whole day averages. As indicated in (Fig. 21_ANNEX_TM5) nighttime O3 in TM5.f1d7 yields considerably higher errors. In June surface ozone is mostly under-estimated in TM5.f1d7 with the exception of a few stations in Northern-Europe including UK, while during July surface ozone is mostly overestimated with the exception of some stations in central Europe (Fig. 22 and Fig. 23 ANNEX_TM5).

7.5 Analysis of results

7.5.1 Evaluation of stand-alone model experiments

Figure 3 shows the Taylor plots for the TM5.v3, TM5.v7 MOZART.v1, MOZART.v10 and MOCAGE-V2 stand-alone runs performed for the reference year 2003. Every dot in the Taylor plot corresponds to one EMEP station. As explained in detail in paragraph 'Statistics and Plots' the azimuthal position gives the correlation coefficient between model and observation fields. The radial distances from the origin to the points, are proportional to the normalized standard deviations. Contours of constant skill score, as defined by Eq. 6 of chapter 2, are also plotted in each Taylor diagram. The statistics were calculated using yearly time-series of mean daily ozone.

For the MOZART.v1 run the majority of EMEP stations have normalized standard deviations σ_n 0.8-1.2 meaning that model and observation fields have comparable amplitude of variation, either moderately under-or overestimating it. Correlation coefficients mostly vary between 0.6-0.8 and the maximum R is found to be 0.88. The skill is decreasing with increasing unbiased-RMSE but at low correlation stations with low variability are penalized. According to the color scale, total RMSE ranges between 8 to 16 ppb.

The statistics look quite different for MOZART.v10, the most striking feature being the exaggeration in the variation of amplitude. In contrast to MOZART.v1 for which the majority of stations is close to the reference point, most EMEP stations exhibit $1.2< \sigma_n <1.7$. Correlation coefficients are also lower and the skill score for the majority of station is 0.9-0.7.

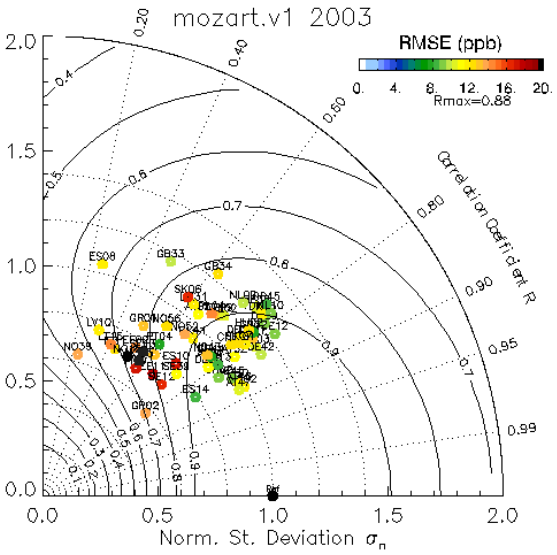
The difference in phase between model and observed values seems to be the major problem for MOCAGE-V2 experiment. Correlation coefficients can be as low as 0.3 and do not get higher than 0.76. The model cannot capture in several stations correctly the variation in daily O3 rather seems to overestimate it. Root mean square error is similar with the other model experiments around 12 ppb on average. The stations with σ_n close to unity are awarded with a high skill score despite the relative lower R.

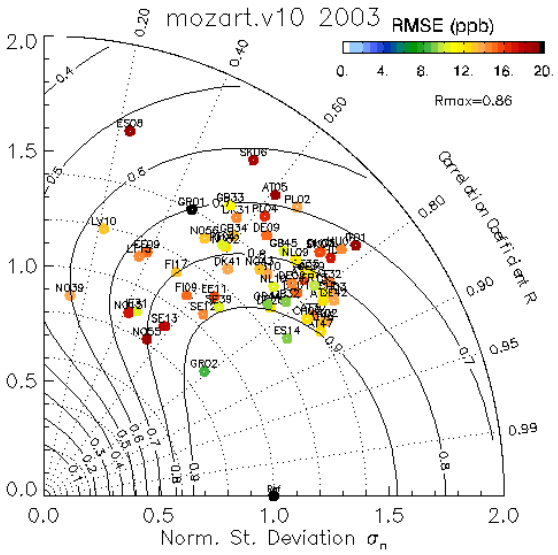
TM5.v3 has a totally different behavior underestimating the amplitude of variation in all EMEP stations since σ_n <1. Correlation coefficients do not get higher than 0.86 and RMSE is lower than all other experiments with an average of 9 ppb. TM5.v7 has similar performance with TM5.v3 with slightly higher average RMSE.

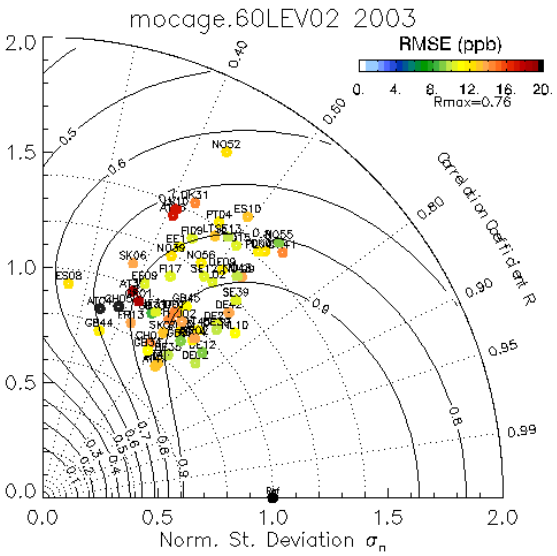
Summarizing the performance of four different runs performed for the reference year 2003 the following conclusions could be reached:

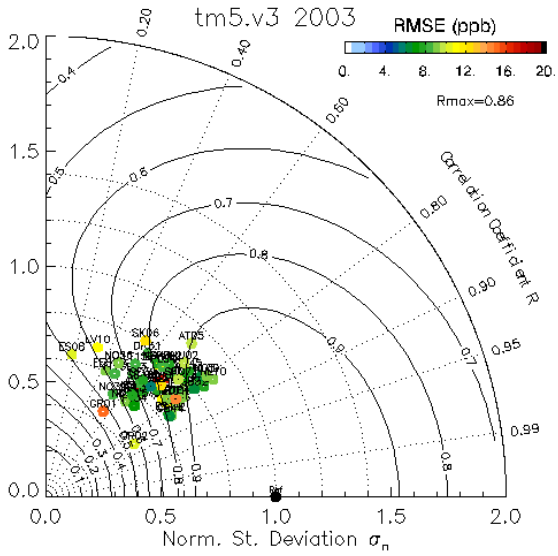
- The best overall performance for reference year 2003 based on the combination of S, R and σ_n is MOZART.v1 which has the higher skill scores and correlation coefficients besides the most comparable amplitude of variation with measurements.
- The major problem with TM5.v3 and TM5.v7 is the underestimation of the amplitude of variation that penalized many EMEP stations with lower skill scores. Its best skill is the low RMSE suggesting simulations with lower bias. Average RMSE is higher for TM5.v7 than TM5.v3
- MOZART.v10 mostly overestimates the amplitude of variation and has also higher RMSE than other experiments, so model results are prone to bias.
- MOCAGE-V2 has relatively lower correlation coefficients than other experiments, captures the amplitude of variations for several stations (mostly the mainland) for the remaining stations there is an overestimation in amplitude.

For the year 2004 there are only two complete runs available: MOZART.v9 and MOZART.v10. The two experiments produce almost identical Taylor plots suggesting that changes in the model configuration hardly affect surface ozone concentrations (Fig. 4). The model behavior is similar to the MOZART.v10 for the year 2003 which is quite reasonable: Taylor plots are meant to present in a concise way the model-dependent pattern statistics which should be independent of the reference year.









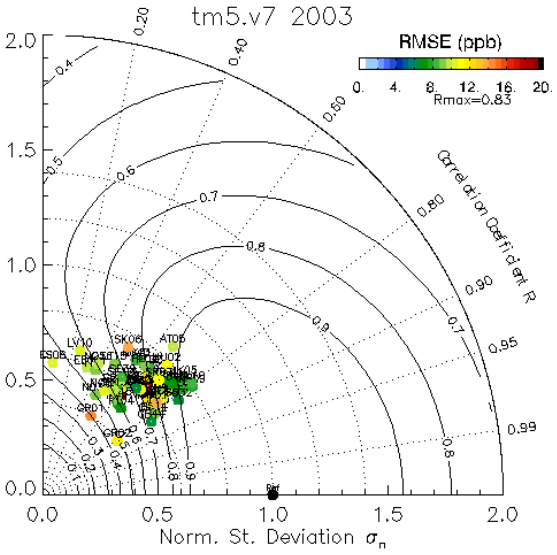


Figure 2 Taylor plots of the stand-alone runs for the reference year 2003. Each dot corresponds to one EMEP station. The statistics have been calculated for the complete year using daily averages of surface ozone. For details see text.

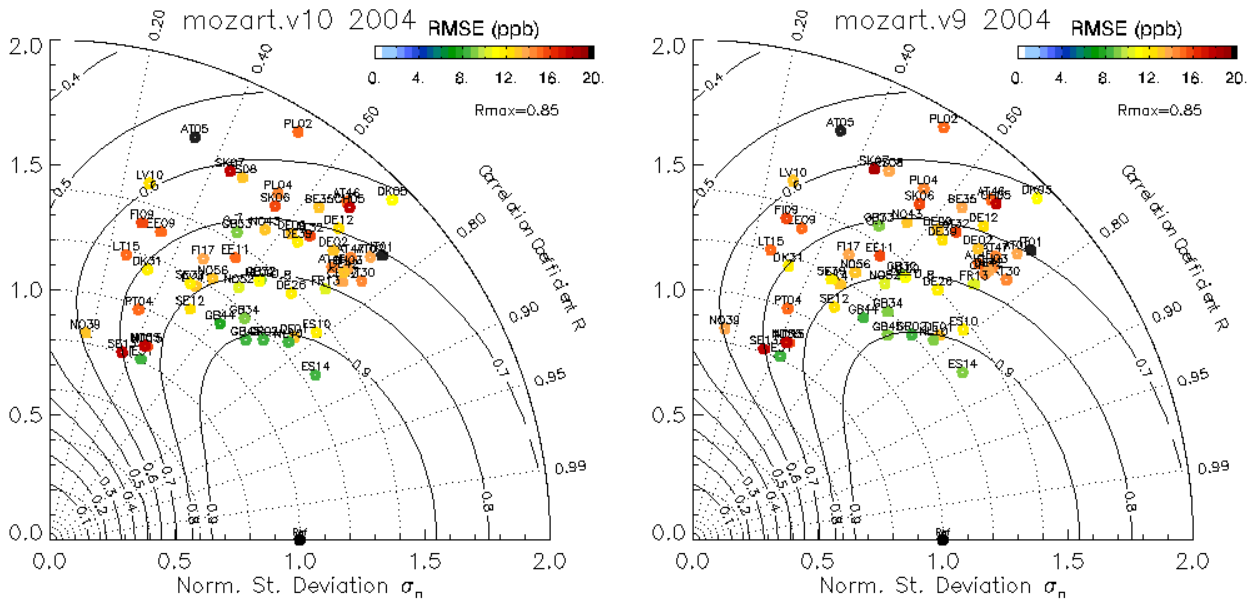


Figure 3 Taylor plots of the stand-alone runs for the reference year 2004. Each dot corresponds to one EMEP station.

7.5.2 Evaluation of coupled assimilation model experiments

The version of the coupled-assimilation system used here consists of the Integrated Forecasting System (IFS) coupled to MOZART. IFS supplies meteorological data at 1-h temporal resolution to the coupled chemistry model. The configuration of the coupled IFS/MOZART system used in this study assimilates CO column data from the MOPITT instrument and O3 data from SCIAMACHY, MIPAS, GOME and SBUV. CO and O3 modelled fields are constrained by observations from the mentioned satellite instruments within each data assimilation window. The changes in concentration due to these observational constraints can then be fed back to the coupled CTM at the start of the next assimilation window.

Experiment f026 concerns the coupled IFS/MOZART system with assimilation (see ANNEX B for description of model runs). This run is the 2nd reanalysis and comparison with the stand-alone runs reveals the impact of coupling and assimilation on model performance. Figure 5 shows a Taylor plot for experiment f026 performed for the year 2003. The model performance is quite different for the stand-alone runs (v1 and v10 for year 2003). Firstly, for the majority of EMEP stations $\sigma_n < 1$ indicating underestimation of amplitude variation. Correlation coefficients are for all stations below 0.8 which is another indication of model performance deterioration. Root mean square errors also are for all stations higher then the stand-alone experiments, the average RMSE of all stations being ~ 15 ppb. As seen in detail in chapter 3.2.2 f026 suffers from gross underestimation of surface ozone during all year. The Taylor plot summarizes the bad model performance by penalizing almost half of the stations with lower skill scores.

As already discussed in chapter 3.2.2 the *impact of assimilation on surface O3 has been negligible*. Experiments eyih and eyq6, the latter being the same as eyih without assimilation, yielded almost identical statistics. *Thus, if we assume that assimilation has no impact on surface ozone the changes of model performance between the Mozart.coupled assimilation version and the Mozart.stand-alone versions should be mainly due to coupling of Mozart to IFS. The impact of coupling has been a deterioration of all calculated statistics. However, calculation of a skill score taking into account σ_n and R as suggested by Taylor 2001, brings the coupled assimilated run in the second position after Mozart.v1 (see below).*

7.5.3 Monthly variation of statistics

In Figures 6 and 7 are shown calculated FGE and RMSE for all stand-alone runs performed for the reference year 2003. It is suggested to examine absolute and relative statistics together since inspection of relative statistics may lead to misleading results, especially when comparing different seasons with large differences in absolute concentrations. **TM5.v3** has the lowest RMSE than all other stand-alone models which mostly

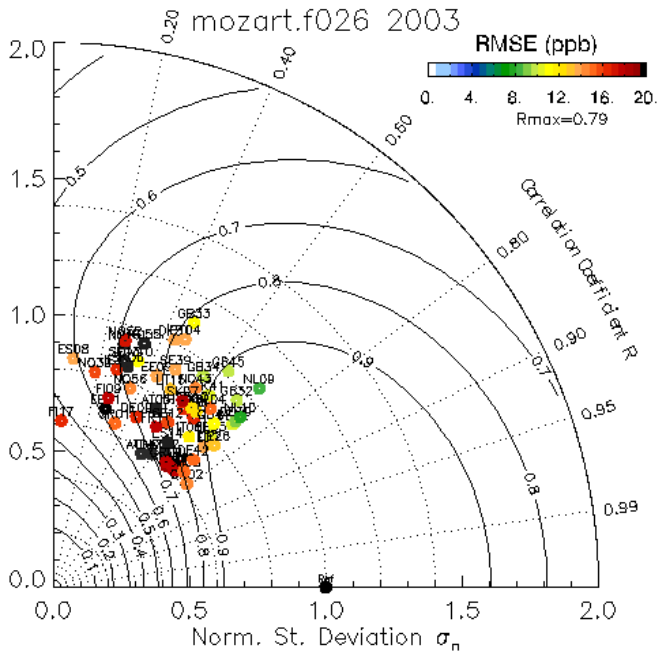


Figure 4 Taylor plots of the MOZART IFS-coupled experiment. Each dot corresponds to one EMEP station. The statistics have been calculated for the complete year using daily averages of surface ozone.

ranges between 8-10 ppb. Calculation of FGE classifies the model performance as acceptable throughout the year with the exception of months November and December for which FGE> 35 %. **MOZART.v1** follows with an average RMSE of 11.5 ppb for the year 2003, the highest errors found from January to April (12-16 ppb). The combination of lower ozone concentrations in colder months and the high absolute error yield very high FGE for months January-April and October-December. During these periods the MOZART.v1 does not pass the threshold set for an acceptable model performance. The pattern is even worse for **MOZART.v10**: January to March yield even higher RMSE (14-18 ppb) all other months are similar to MOZART.v1 with the exception of month July when again RMSE and FGE get higher than those calculated for MOZART.v1. The difference of **MOCAGE-V2** is that maintains higher RMSE even during warm months when all other models perform better. During the colder period MOCAGE is doing better than MOZART and is a little worse than TM5.

Concerning the monthly variation of statistics we could conclude that:

- TM5 performs better than other models with respect to RMSE and FGE for both cold and warm season

- MOZART performs worse than TM5 and MOCAGE during colder months
- MOCAGE performs worse that TM5 and MOZART during warm months
- The higher RMSE during the period January-April is a common feature of all four stand-alone simulations being more pronounced for MOZART and least pronounced for TM5.

In Figures 8 and 9 are shown calculated RMSE and FGE respectively for all stand-alone runs performed for the reference year 2004. TM5.v9 run is performed only for months May to August and MOCAGE.HTAP_ES1 for the time period June to September. As in year 2003 TM5 exhibits the lowest RMSE for the summer period, followed by MOCAGE and MOZART. Tables 1 and 2 summarize RMSE for all performed runs on a seasonal basis, for the years 2003 and 2004. The given RMSE is the average of all EMEP stations examined.

Figure 10 shows monthly variation of the average RMSE of surface ozone for the MOZART/IFS coupled experiments with assimilation. Coupled experiments are presented as lines and stand-alone runs as dots for an easier comparison. Starting with experiment MOZART.f026 which was performed for the whole year, we find as previously that the first months of the year have higher RMSE. The monthly variation of RMSE for MOZART.f026 resembles more the monthly behaviour of MOZART.v1 than of MOZART.10 stand-alone. The absolute RMSE values are however much higher than both MOZART stand-alone runs performed for the year 2003. The changes in surface ozone concentrations are due to changes in model versions, the IFS coupling and the impact of assimilation. Table 3 summarizes the seasonal variation of RMSE for the coupled experiments with assimilation.

Equation Nr. 6 (page 14) defines a skill score for the evaluation of the overall model performance. We average the skill score for all EMEP stations and thus we present in Table 4 the final score for every model version run for a complete year. As the model variance approaches the observed variance and as the correlation coefficient approached the maximum attainable value, the skill approaches unity. Under this definition, skill decreases toward zero as the correlation becomes more and more negative or as the model variance approaches either zero or infinity. For fixed variance the skill increases linearly with correlation. Note also that for small model variance, skill is proportional to the variance, and for large variance, skill is inversely proportional to the variance (Taylor 2001).

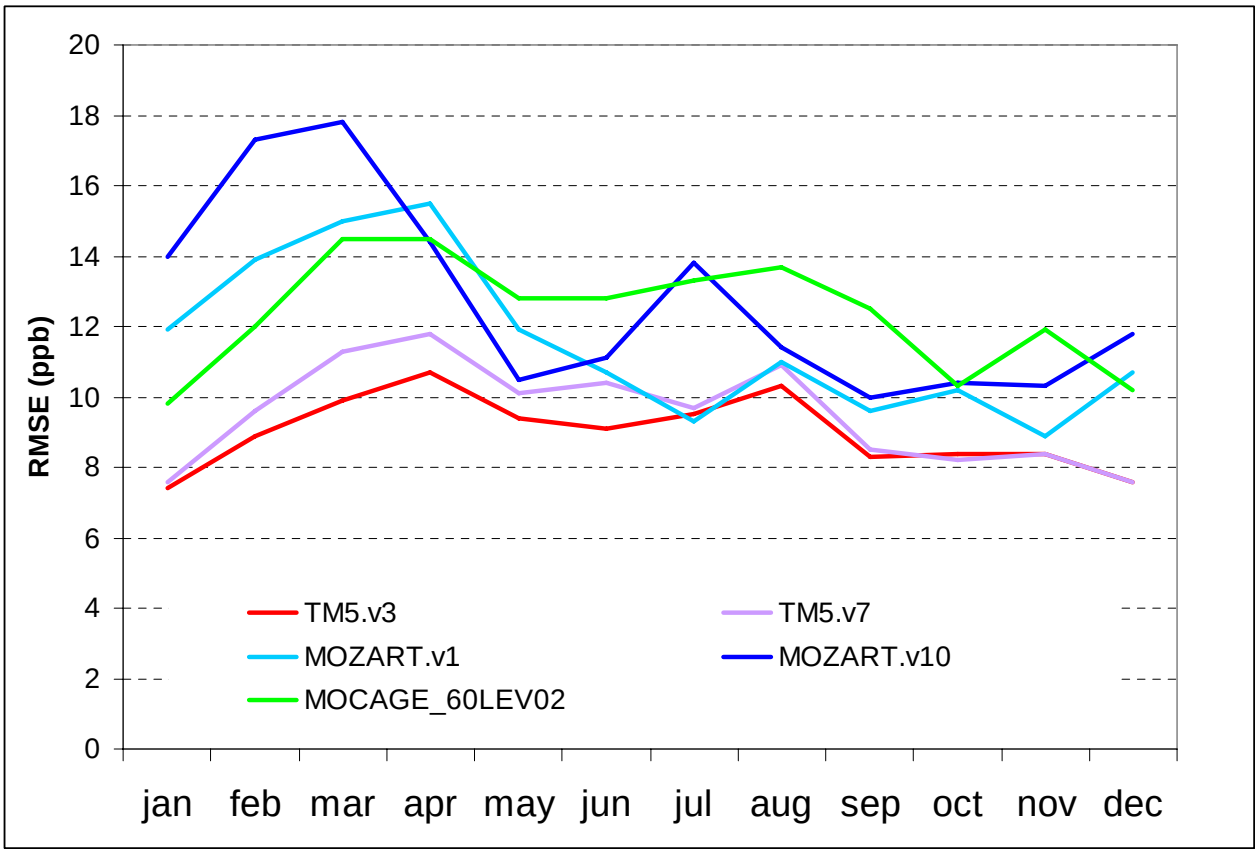


Figure 5 Comparison of RMSE for the TM5/MOZART/MOCAGE stand-alone runs for the year 2003.

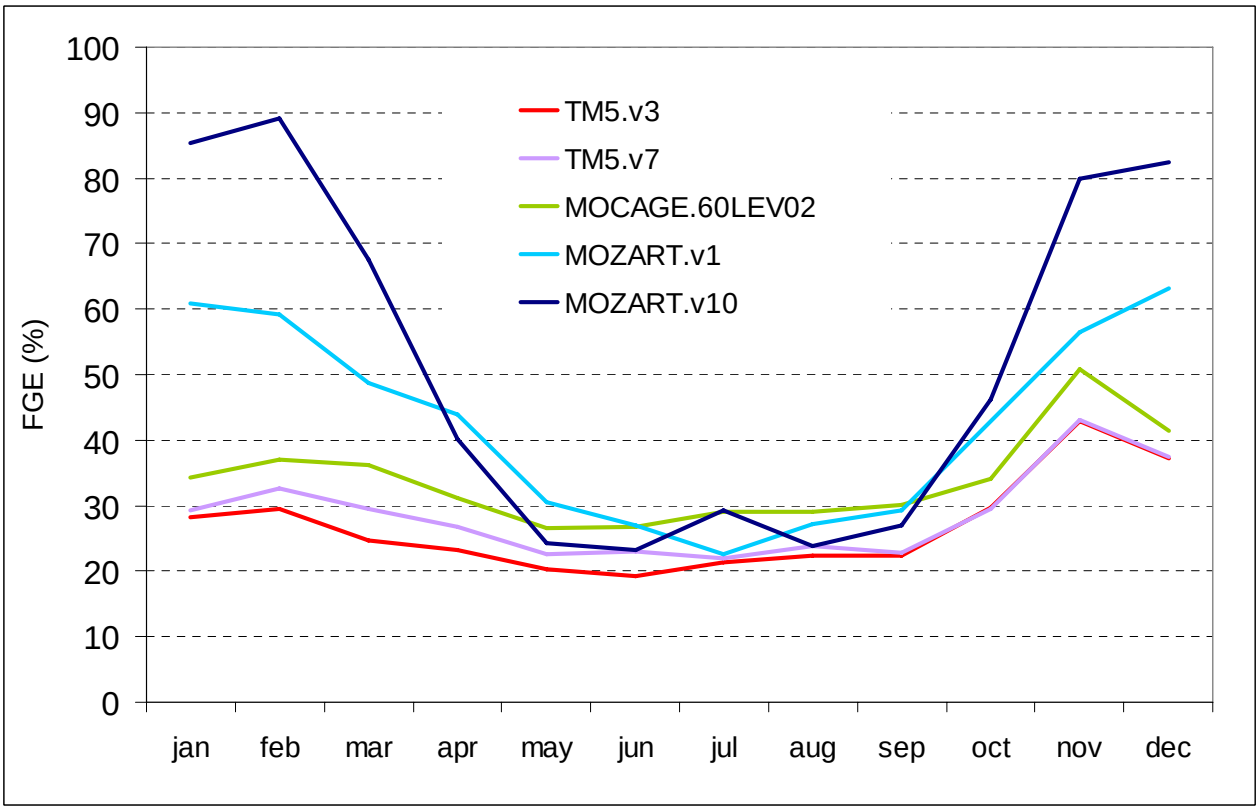


Figure 6 Comparison of FGE for the TM5/MOZART/MOCAGE stand-alone runs for the year 2003

Table 1. Seasonal RMSE (ppb) for the stand-alone runs for the reference year 2003. Numbers correspond to the average of all EMEP stations.

2003	Winter	Spring	Summer	Autumn
TM5.v3	8.0	10.0	9.6	8.4
TM5.v7	8.3	11.1	10.3	8.4
MOZART.v1	12.2	14.1	10.3	9.6
MOZART.v10	14.4	14.2	12.1	10.2
MOCAGE-V2	10.7	13.9	13.3	11.6

Table 2 Seasonal RMSE (ppb) for the stand-alone runs for the reference year 2004. Numbers correspond to the average of all EMEP stations.

2004	Winter	Spring	Summer	Autumn
TM5.v9	-	9.1	-	-
MOZART.v9	14.3	13.3	13.5	10.1
MOZART.v10	14.1	13.2	13.3	10.1
MOCAGE.HTAP_ES1	-	11.4	-	-

Table 3 Seasonal RMSE (ppb) for the IFS coupled runs with assimilation for the reference year 2003. Numbers correspond to an average of all EMEP stations.

2003	Winter	Spring	Summer	Autumn
MOZART.f026	13.2	20.7	14.4	10.8
MOZART.eyih	-	-	12.1	9.2
MOZART.eyq6	-	-	12.2	9.4

Table 4 Calculation of a skill score averaged over all EMEP stations for the experiments run for a complete year. For details see text.

Model	Experiment	Score
MOZART	v1	0.894
MOCAGE	60LEV02	0.860

TM5	v10	0.831
MOZART	f026	0.829
MOZART	v10.2003	0.820
TM5	v3	0.799
MOZART	v10.2004	0.769
MOZART	v9	0.763
TM5	v7	0.750

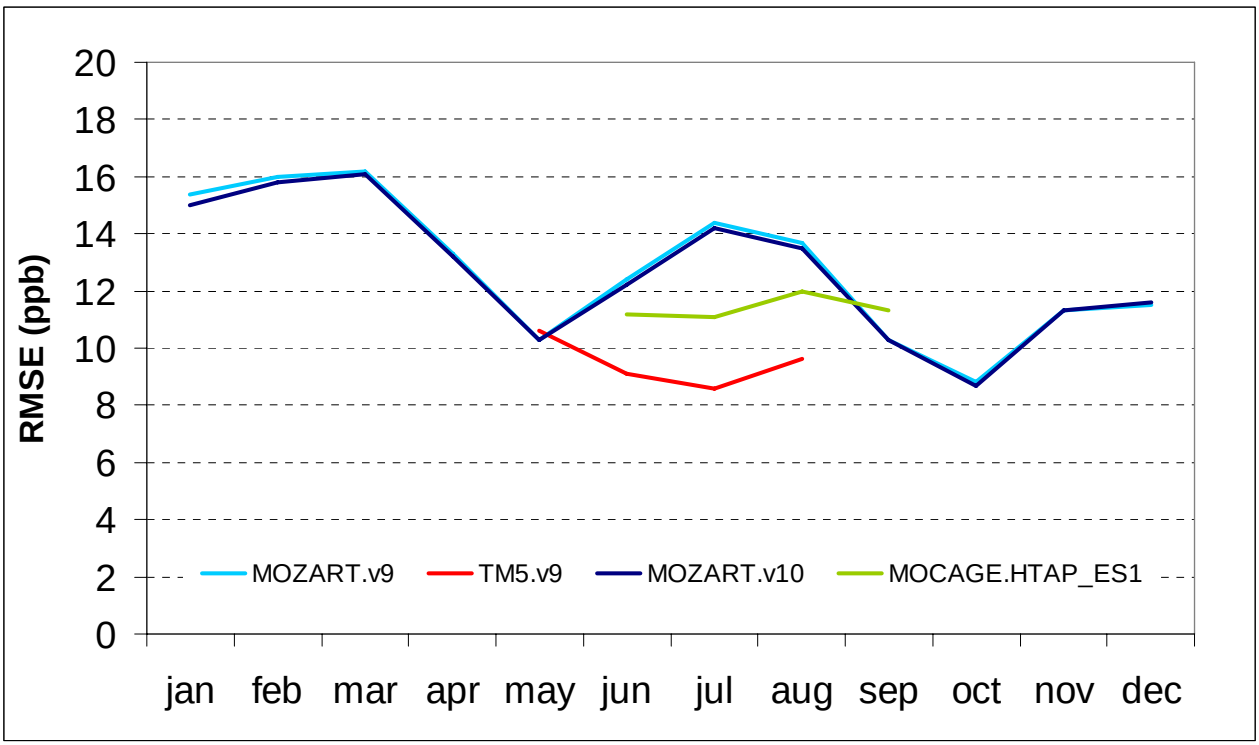


Figure 7 Comparison of RMSE for the second TM5/MOZART/MOCAGE stand-alone runs for the year 2004.

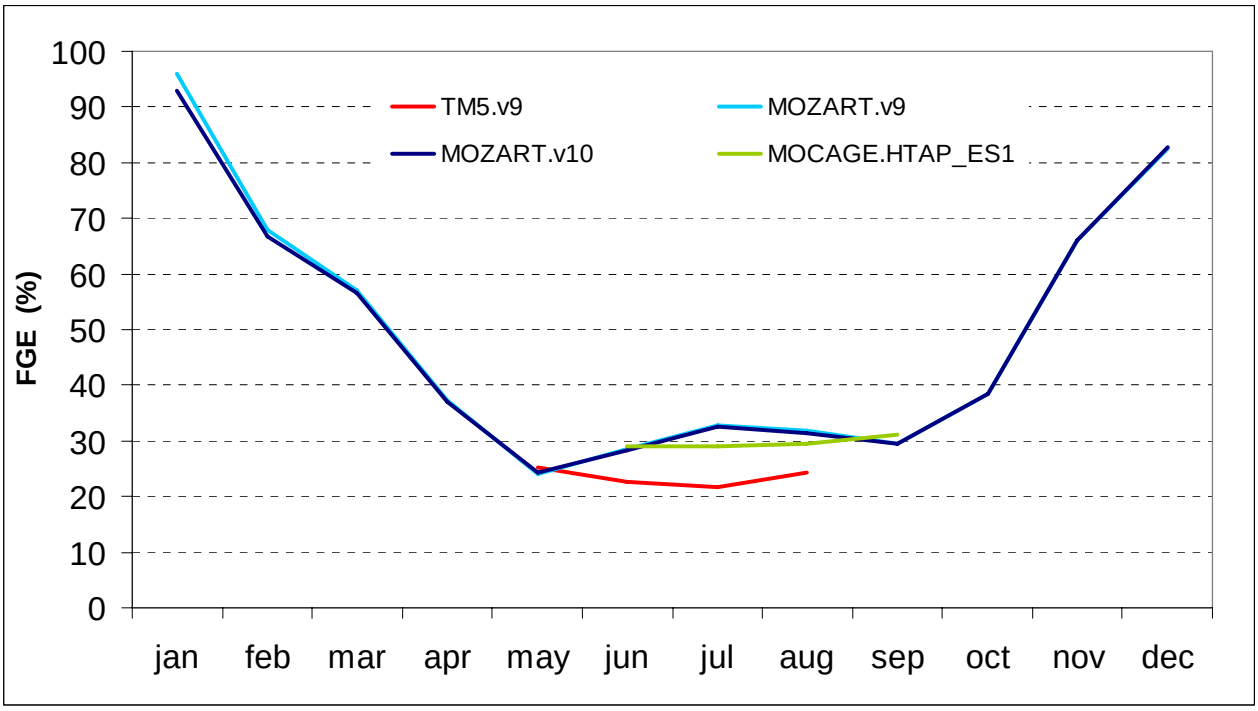


Figure 8 Comparison of FGE for the TM5/MOZART/MOCAGE stand-alone runs for the year 2004

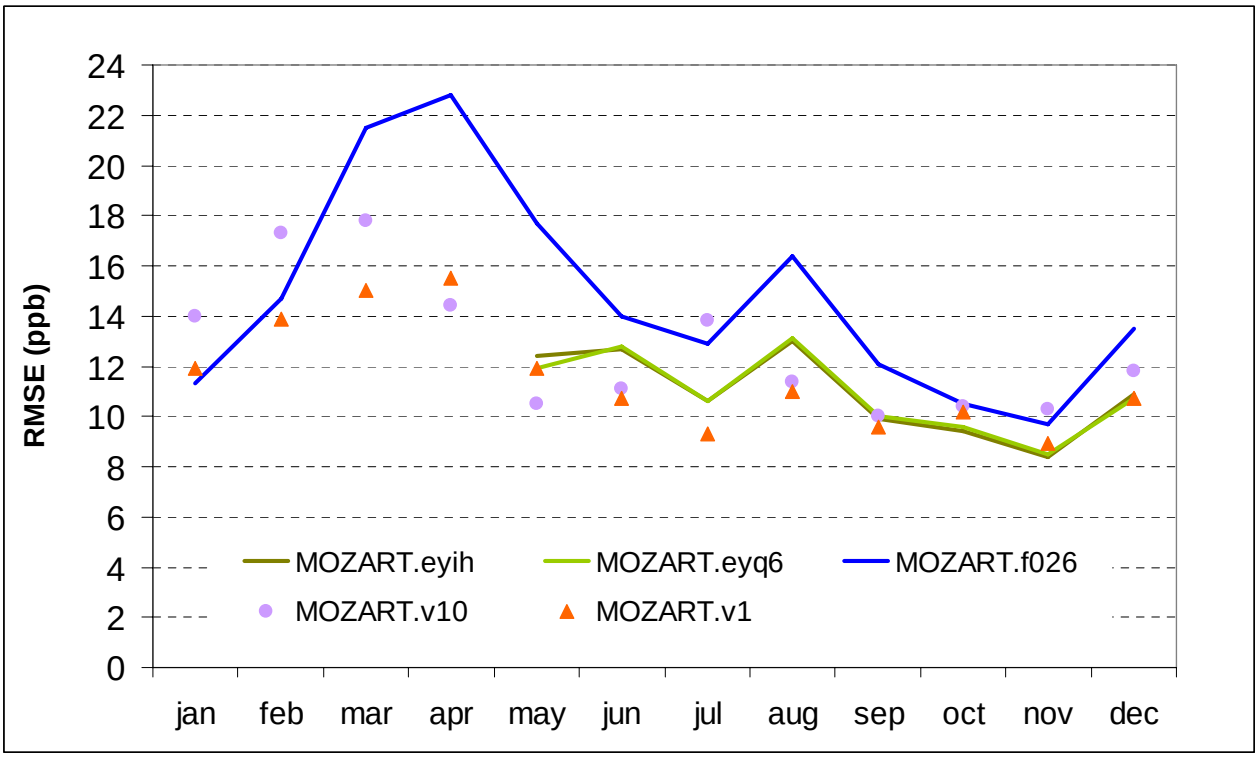


Figure 9 Comparison of RMSE for the IFS-coupled experiments with assimilation (solid lines). Two stand-alone experiments for the same year (2003) have been added as symbols for comparison.

7.5.4 Day versus night surface ozone

The statistics presented in previous chapters were calculated using mean daily values. Calculation of the FGE was performed again using only day- and nighttime ozone values. As mentioned in chapter 3.1.1 the mean daytime O3 concentration is taken as the average O3 between 12:00-15:00 while nighttime as the average between 00:00-3:00. In Fig. 11 are presented the monthly mean FGEs of all stand-alone runs performed for the reference year 2003. Daytime monthly values are the solid lines, nighttime values are the dashed lines and the dots represent the mean daily values. For every experiment the FGE is considerable higher during nighttime, when compared to daytime values, while the latter are somewhat better or comparable to daily FGE. Since FGE> 35 % for every calculated nighttime monthly mean, *all models fail to capture the night-time ozone chemistry*. Summarizing the features showed in Fig. 11:

- TM5 has the lower relative error in simulation of surface ozone during night.
- MOCAGE-V2 has the higher relative error during warm months. Indeed the RMSE during nighttime reaches up to 18 ppb during summer (not shown here).
- MOZART.v1 simulations yield lower FGE for day, night and mean daily surface ozone values than MOZART.10 for the year 2003. The difference in performance concerns mostly the cold months (January-March, November-December).

Figure 12 shows monthly mean FGEs for the reference year 2004. The same conclusions are reached for the runs of he second reference year. Notice that FGE is almost the same for MOZART.v9 and v10 when calculating FGE using daily means (compare symbols), while FGE is different for daytime surface ozone during cold months: MOZART.v9 has lower FGE during daytime than MOZART.V10

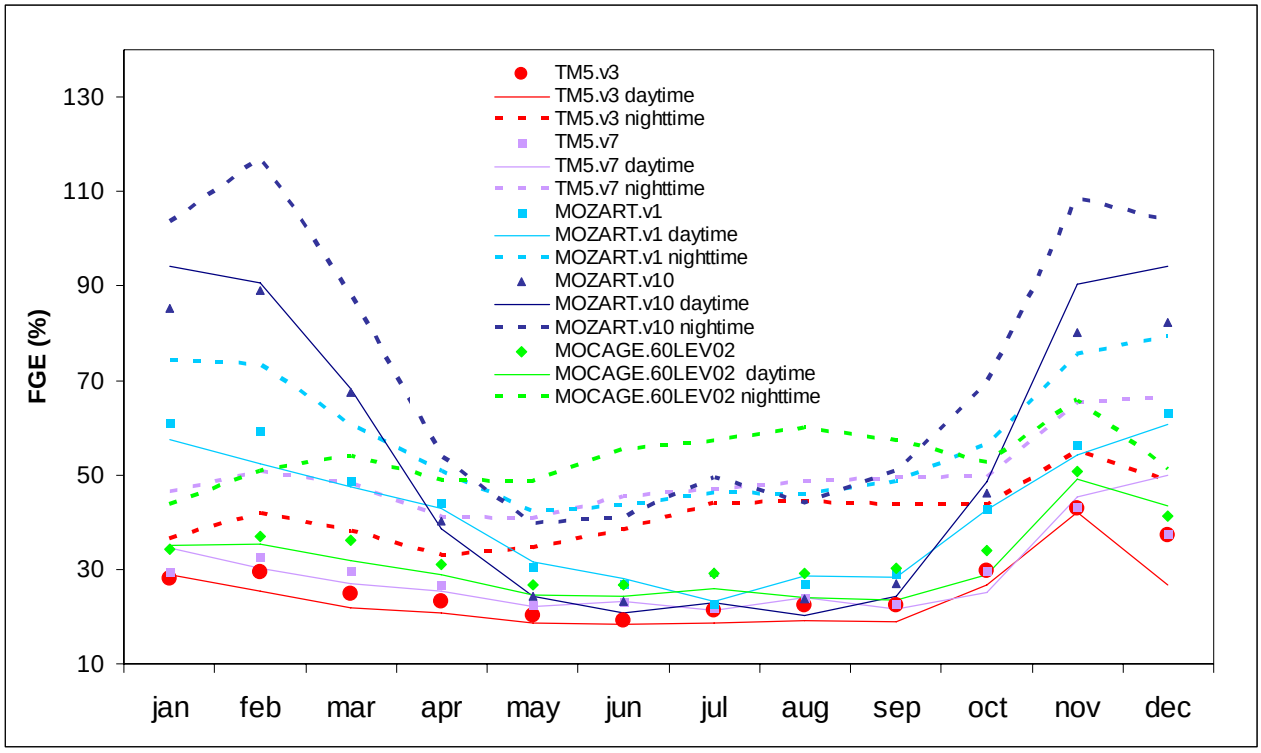


Figure 10 Comparison of FGE for the stand-alone runs for the reference year 2003. Solid lines correspond to monthly means of daytime surface ozone, dashed lines to monthly means of nighttime surface ozone, and points to monthly means of 24h means of surface ozone. The data are averaged over all selected EMEP stations.

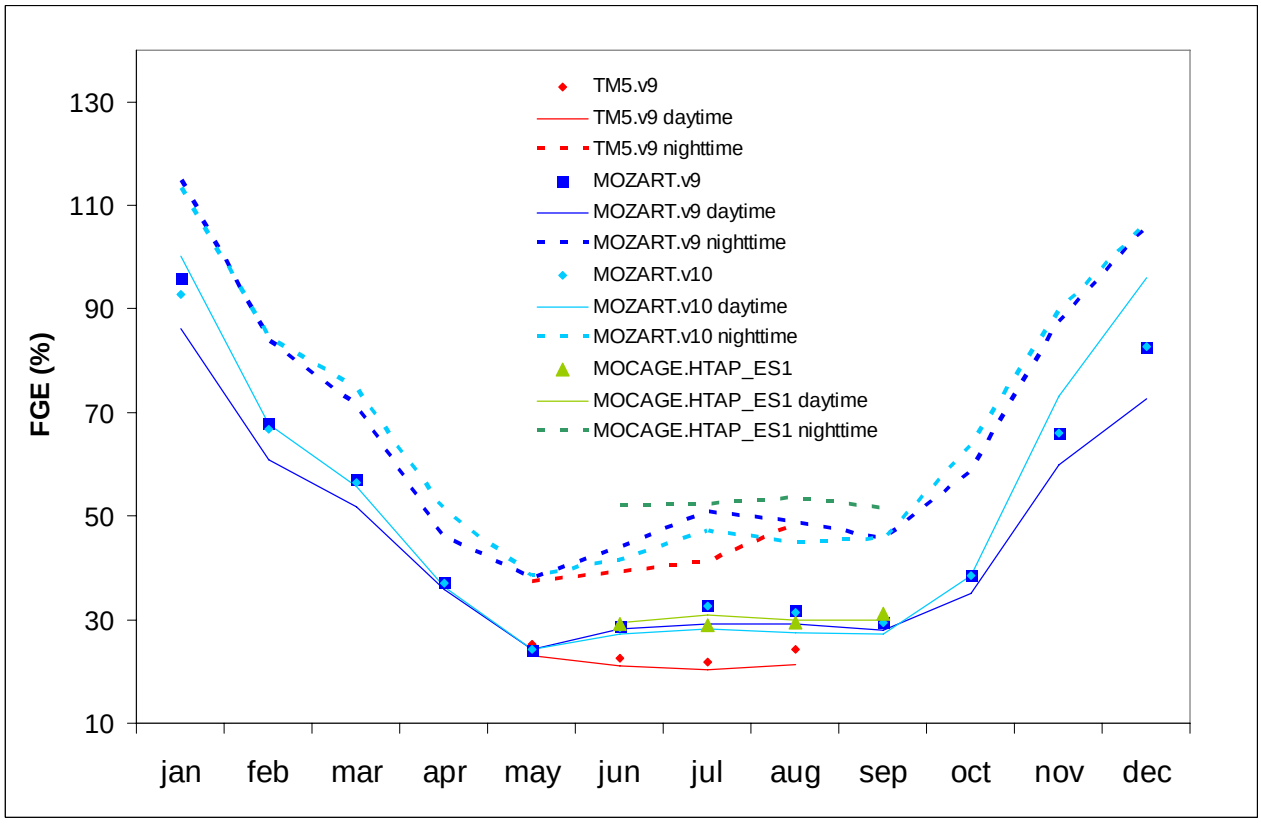


Figure 11 Comparison of FGE for the stand-alone runs for the reference year 2004. Solid lines correspond to monthly means of daytime surface ozone, dashed lines to monthly means of nighttime surface ozone, and points to monthly means of 24h means of surface ozone. The data are averaged over all selected EMEP stations.

7.6 Conclusions and recommendations

This work is concerned with validation of three global chemistry transport models (CTM), MOCAGE, MOZART and TM5 which have performed simulations for the Global Reactive Gases (GRG) subproject of GEMS. The measurements used for validation were taken from the EMEP database and consisted of hourly surface ozone time series. The runs validated were stand-alone runs performed for the reference years 2003 and 2004 and for selected IFS coupled model experiments. The overall model performance based on the calculation of a skill score proposed by Taylor, 2001 which takes into account the correlation coefficient and the normalized standard deviation, brings first MOZART.v1 first (0.894) followed by MOCAGE-V2 (0.86) and TM5.v10 (0.831). The worse performance of all is TM5.v7 (0.750), penalized heavily for the low variation of amplitude (low normalized standard deviation). Higher agreement in a temporal scale would be necessary for better agreement with measurements for all models, especially with the MOCAGE-V2 for which were recorded the lowest correlation coefficients. TM5 in general yielded the lowest bias and error in comparison to other model experiments. MOZART, on the other hand, yielded higher errors and bias for all seasons. The impact of assimilation on surface ozone appeared to be negligible for all station data examined. Coupling of MOZART to IFS leads to deterioration of performance in comparison to MOZART.v1 (skill score for MOZART.f026 is 0.829), however the overall performance was better than this of MOZART.v9 and V10. A common feature in all models is that they fail to reproduce in a satisfactory way night surface ozone levels in rural areas. Better representation of near-surface nighttime chemistry would be, thus, necessary for all models.

7.7 References

1. Aas, W. and A_G. Hjellbrekke, Data quality 2003, quality assurance and field comparisons, Norwegian Institute for Air Research, EMEP/CCC-Report 6/2005, Reference O-95024, 2005, available at http://www.emep.int/
2. Bechtold, P., et al. (2001), A mass flux convection scheme for regional and global models, Q. J. R. Meteorol. Soc., 127, 869–886.
3. N. Bousserez, J. L. Attie, V. H. Peuch, M. Michou, G. Pfister, D. Edwards,L. Emmons, C. Mari,1 B. Barret,S. R. Arnold, A. Heckel, A. Richter, H. Schlager, A. Lewis, M. Avery, G. Sachse, E. V. Browell, and J. W. Hair, Evaluation of the MOCAGE chemistry transport model during the ICARTT/ITOP experiment, JGR, 112, D10S42, doi:10.1029/2006JD007595, 2007
4. Brasseur, G. P., D. A. Hauglustaine, S. Walters, P. J. Rasch, J.-F. Muller, C. Granier, and X.-X. Tie (1998), MOZART: A global chemical transport model for ozone and related chemical tracers: 1. Model description, J. Geophys. Res., 103, 28,265–28,289.
5. Cammas J.P., F. Eddounia, H. Flentje, C. Ordonez, C. Textor, Summary of global evaluation for GEMS-GRG, GEMS-GRG report, 2007.
6. Dentener F., M. van Weele, M. Krol, S. Houweling, and P. van Veldhoven (2003). Trends and inter-annual variablity of methane emissions derived from 1979-1993 global ctm simulations. Atmos. Chem. Phys., 3:73–88.
7. Dentener, F., et al. (2004), The impact of air pollutant and methane emission controls on tropospheric ozone and radiative forcing: CTM calculations for the period 1990– 2030, Atmos. Chem. Phys., 4, 1551– 1564
8. Emmons, L. K., et al. (2003), Budget of tropospheric ozone during TOPSE from two chemical transport models, J. Geophys. Res., 108(D8), 8372, doi:10.1029/2002JD002665.
9. Giorgi, F., and W. L. Chameides (1986), Rainout lifetimes of highly soluble aerosols and gases as inferred from simulations with a general circulation model, J. Geophys. Res., 91, 14,367– 14,376.

10. Hauglustaine, D. A., G. P. Brasseur, S. Walters, P. J. Rasch, J-.F. Mueller, L. K. Emmons, and M. A. Carroll (1998), MOZART: A global chemical transport model for ozone and related chemical tracers: 2. Model results and evaluation, J. Geophys. Res., 103, 28,291– 28,335.
11. Heimann M., P. Monfray, and G. Polian (1998). Long-range transport of 222Rn – a test for 3D tracer models. Chem. Geol., 70:98–98.
12. Hjellbrekke A-G. and S. Solberg, Ozone Measurements 2003, Norwegian Institute for Air Research, EMEP/CCC Report 4/2005, Reference, O-99074, 2005, available at http://www.emep.int/
13. Hollingsworth, A, R. J. Engelen, C. Textor, A. Benedetti, O. Boucher, et al. The Global Earth-system Monitoring using Satellite and in-situ data (GEMS) Project: Towards a monitoring and forecasting system for atmospheric composition, BAMS, doi: 10.1175/2008BAMS2355.1, 2008.
14. Horowitz, L. W., et al. (2003), A global simulation of tropospheric ozone and related tracers: Description and evaluation of MOZART, version J. Geophys. Res., 108(D24), 4784, doi:10.1029/2002JD002853.
15. Houweling S., F. Dentener, and J. Lelieveld (1998). The impact of non-methane hydrocarbon compounds on tropospheric photochemistry. J. Geophys. Res., 103(D9):10673–10696 doi:10.1029/97JD03582.
16. Huijnen, V., Model documentation of TM5 with focus on GEMS settings, 2008, available at http://gems.ecmwf.int/documents/index.jsp#general
17. Huijnen V., H. Eskes, J. Marshall, D. Ceburnis, A. Gross, M. Schulz, P. Agnew, S. Kinne, A. Ghelli, Scoring approaches for the GEMS project, GEMS_VAL , deliverable D1.1, 2008
18. Jolliff, J.K., Kindle, J.C., Shulman, I., Penta, B., Friedrichs, M.A.M., Helber, R., Arnone, R.A. Summary diagrams for coupled hydrodynamic-ecosystem model skill assessment (2008) Journal of Marine Systems, . Article in Press.
19. Josse, B., P. Simon, and V.-H. Peuch (2004), Radon global simulations withthe multiscale chemistry transport model MOCAGE, Tellus, Ser. B, 56,339– 356.
20. Kinnison, D.E., Brasseur, G.P., Walters, S., Garcia, R.R., Marsh, D.R., Sassi, F., Harvey, V.L., Randall, C.E., Emmons, L., Lamarque, J.F., Hess, P., Orlando, J.J., Tie, X.X., Randel, W., Pan, L.L., Gettelman, A., Granier, C., Diehl, T., Niemeier, U., Simmons, A.J. (2007) Sensitivity of chemical tracers to meteorological parameters in the MOZART-3 chemical transport model,Journal of Geophysical Research D: Atmospheres, 112 (20), art. no. D20302.
21. M.C. Krol, S. Houweling, B. Bregman, M. van den Broek, A.J. Segers, P. van Velthoven, W. Peters, F. Dentener, and P. Bergamaschi. The two-way nested global chemistry–transport zoom model TM5: algorithm and applications. Atmos. Chem. Phys., 5(2):417–432, February 2005.
22. Lin, S. J., and R. B. Rood (1996), A fast flux form semi-Lagrangian transport scheme on the sphere, Mon. Weather Rev., 124, 2046– 2070.
23. Louis, J.-F. (1979), A parametric model for vertical eddy-fluxes in the Atmosphere, Boundary Layer Meteorol., 17, 187– 202.
24. Liu, H., D. J. Jacob, I. Bey, and R. M. Yantosca (2001), Constraints from210Pb and 7Be on wet deposition and transport in a global three-dimensional chemical-transport model driven by assimilated meteorological fields, J. Geophys. Res., 106, 12,109– 12,128
25. Mari, C., D. J. Jacob, and P. Bechtold (2000), Transport and scavenging of soluble gases in a deep convective cloud, J. Geophys. Res., 105, 22,255– 22,267.
26. Michou, M., and V.-H. Peuch (2002), Surface exchanges in the MOCAGE multi-scale chemistry and transport model, J. Water Sci, 15, 173– 204
27. Michou, M., et al. (2005), Measured and modeled dry deposition velocities over the ESCOMPTE area, Atmos. Res, 74(1–4), 89– 116.
28. Nho-Kim, E.-Y., et al. (2004), Parameterization of size dependent particle dry deposition velocities for global modeling, Atmos. Environ., 38(13), 1933– 1942.
29. Rasch, P. J., N. M. Mahowald, and B. E. Eaton (1997), Representations oftransport, convection, and the hydrological cycle in chemical transport models: Implications for the modeling of short-lived and soluble species, J. Geophys. Res., 102, 28,127– 28,138.

30. Rotman, D. A., et al. (2004), IMPACT, the LLNL 3-D global atmospheric chemical transport model for combined troposphere and stratosphere: Model description and analysis of ozone and other trace gases, J. Geophys. Res., 109, D04303, doi:10.1029/2002JD003155.
31. Sander S.P., R.R. Friedl, A.R. Ravishankara, D.M. Golden, C.E. Kolb, M.J. Kurylo, M.J. Molina, G.K. Moortgart, H. Keller-Rudek, B.J. Finlayson-Pitts, P.H. Wine, R.E. Huie, and V.L. Orkin. Chemical kinetics and photochemical data for use in atmospheric studies, evaluation no. 15. Technical report, JPL, 2006.
32. Taylor, K.E., 2001. Summarizing multiple aspects of model performance in a single diagram. Journal of Geophysical Research 106, 7183–7192.
33. Wesely, M. L. (1989), Parameterization of surface resistances to gaseous dry deposition in regional-scale numerical models, Atmos. Environ., 23,1293–1304
34. Williams J.E. and T.C.P. van Noije. On the upgrading of the modified carbon bond mechanism iv for use in global chemistry transport models. Technical Report TR-2008-02, KNMI, De Bilt, The Netherlands, May 2008.

7.8 Annex

List of EMEP stations with geographical information and suggested assignment to model levels

Station	Name	Latitude	Longitude	Altitude (m)	model level
AT02	Illmitz	47.77	16.77	117	L60
AT04	StKoloman	47.65	13.2	851	L60
AT05	Vorhegg	46.67	12.97	1020	L60
AT30	Pillersdor	48.72	15.93	315	L60
AT32	Sulzberg	47.52	9.92	1020	L60
AT33	Stolzalpe_	47.12	14.2	1302	L54
AT34	Sonnblick	47.05	12.95	3106	L47
AT37	Zillertale	47.13	11.87	1970	L52
AT38	Gerlitzen	46.68	13.9	1895	L50
AT40	Masenberg	47.33	15.87	1170	L52
AT41	Haunsberg	47.97	13	730	L52
AT42	Heidenreic	48.87	15.03	570	L58
AT43	Forsthof	48.1	15.92	581	L57
AT44	Graz_Platt	47.1	15.47	651	L58
AT45	Dunkelstei	48.37	15.53	320	L60
AT46	Gaenserndo	48.33	16.72	161	L60
AT47	Stixneusie	48.05	16.67	240	L60
AT48	zobelboden	47.83	14.43	899	L58
BE01	Offagne	49.87	5.2	430	L56
BE32	Eupen	50.62	6	295	L57
BE35	Vezin	50.5	4.98	160	L60
BG53	Rojen_Peak	41.7	24.73	1750	L49
CH02	Payerne	46.82	6.95	510	L60
CH03	Taenikon	47.48	8.9	540	L60
CH04	Chaumont	47.05	6.97	1130	L55

CH05	Rigi	47.07	8.45	1030	L60
CZ01	Svratouch	49.73	16.03	737	L54
CZ03	Kosetice	49.58	15.08	534	L57
DE01/DEUB001	Westerland	54.92	8.3	12	L60
DE02	Langenbrue	52.8	10.75	74	L60
DE03/DEUB004	Schauinsla	47.9	7.9	1205	L52
DE04	Deuselbach	49.75	7.05	480	L56
DE05	Brotjacklr	48.82	13.22	1016	L54
DE07/DEUB030	Neuglobsow	53.17	13.03	62	L59
DE08/DEUB29	Schmuecke	50.65	10.77	937	L52
DE09/DEUB28	Zingst	54.43	12.73	1	L60
DE12	Bassum	52.85	8.7	52	L60
DE26	Ueckermuen	53.75	14.07	1	L60
DE35	Lueckendor	50.83	14.77	490	L56
DE39	Aukrug	54.07	9.78	15	L60
DE42	Oehringen	49.23	9.43	283	L60
DE45	Schorfheid	52.97	13.65	70	L59
DK05	Keldsnor	54.73	10.73	10	L60
DK31	Ulborg	56.28	8.43	10	L60
DK41	Lille_Valb	55.68	12.12	10	L60
EE09	Lahemaa	59.5	25.9	32	L60
EE11	Vilsandy	58.38	21.82	6	L60
ES07	Viznar	37.23	-3.53	1265	L52
ES08	Niembro	43.43	-4.85	134	L60
ES09	Campisabal	41.27	-3.13	1360	L54
ES10	Cabo_de_Cr	42.32	3.32	23	L60
ES11	Barcarrola	38.47	-6.92	393	L57
ES12	Zarra	39.08	-1.1	885	L55
ES13	Penausende	41.28	-5.87	985	L57

ES14	Els_Torms	41.4	0.72	470	L60
ES15	Risco_Llan	39.52	4.35	1241	L49
ES16	O_Savinyao	43.22	-7.68	506	L57
FI09	Utoe	59.77	21.37	7	L60
FI17	Virolahti_	60.52	27.68	4	L60
FI22	Oulanka	66.32	29.4	310	L57
FI37	Ahtari_II	62.58	24.18	180	L58
FR08	Donon	48.5	7.13	775	L54
FR09	Revin	49.9	4.63	390	L56
FR10	Morvan	47.27	4.08	620	L54
FR12	Iraty	43.03	-1.08	1300	L51
FR13	Peyrusse_V	43.37	0.1	236	L60
FR14	Montandon	47.18	6.5	746	L58
FR15	La_Tardier	46.65	-0.75	143	L58
FR16	Le_Casset	45	6.47	750	L60
GB02	Eskdalemui	55.3	-3.2	243	L58
GB06	Lough_Nava	54.43	-7.87	126	L58
GB13	Yarner_Woo	50.58	-3.7	119	L58
GB14	High_Muffl	54.33	-0.8	267	L57
GB15	Strath_Vai	57.73	-4.77	270	L57
GB31	Aston_Hill	52.5	-3.02	370	L56
GB32	Bottesford	52.92	-0.8	32	L60
GB33	Bush	55.85	-3.2	180	L60
GB34	Glazebury	53.45	-2.45	21	L60
GB35	Great_Dun_	54.68	-2.43	847	L52
GB36	Harwell	51.57	-1.32	137	L59
GB37	Ladybower_	53.38	-1.75	420	L55
GB38	Lullington	50.78	0.17	120	L58
GB39	Sibton	52.28	1.45	46	L59

GB43	Narberth	51.23	-4.7	160	L57
GB44	Somerton	51.22	-3.03	55	L60
GB45	Wicken_Fen	52.28	-0.28	5	L60
GR01	Aliartos	38.37	23.08	110	L60
GR02	Finokalia	35.5	26.17	0	L60
HU02	K-puszta	46.97	19.58	125	L60
IE31	Mace_Head	53.17	-9.00	340	L55
IT01	Montelibre	42.1	12.63	48	L60
LT15	Preila	55.35	21.07	5	L60
LV10	Rucava	56.22	21.22	5	L60
MT01	Giordan_li	36.1	14.2	160	L57
NL09	Kollumerwa	53.33	6.27	1	L60
NL10	Vredepeel	51.53	5.85	28	L60
NO01	Birkenes	58.38	8.25	190	L58
NO15	Tustervatn	65.83	13.92	439	L60
NO39	Karvatn	62.78	8.88	210	L60
NO41	Osen	61.25	11.78	440	L60
NO42	Spitsberge	78.9	11.88	474	L56
NO43	Prestebakk	59	11.53	160	L60
NO45	Jelya	59.43	10.6	5	L60
NO48	Voss	60.6	6.53	500	L60
NO52	Sandve	59.2	5.2	15	L60
NO55	Karasjok	69.47	25.22	333	L60
NO56	Hurdal	60.37	11.07	300	L60
PL02	Jarczew	51.82	21.98	180	L60
PL03	Sniezka	50.73	15.73	1603	L49
PL04	Leba	54.75	17.53	2	L60
PL05	Diabla_Gor	54.15	22.07	157	L58
PT04	Monte_Velh	38.08	-8.8	43	L60

RU16	Shepeljovo	59.97	29.12	4	L60
RU18	Danki	54.9	37.8	150	L60
SE11	Vavihill	56.02	13.15	175	L57
SE12	Aspvreten	58.8	17.38	20	L60
SE13	Esrange	67.88	21.07	475	L60
SE14	Rao	57.4	11.92	10	L60
SE32	Norra-Kvil	57.82	15.57	261	L57
SE35	Vindeln	64.25	19.77	225	L58
SE39	Grimsoe	59.75	15.58	120	L60
SI08	Iskrba	45.57	14.87	520	L57
SI31	Zarodnje	46.42	15	770	L56
SI32	Krvavec	46.28	14.53	1740	L50
SI33	Kovk	46.12	15.1	600	L58
SK02	Chopok	48.93	19.58	2008	L48
SK04	Stara_Lesn	49.15	20.28	808	L54
SK06	Starina	49.05	22.27	345	L60
SK07	Topolniky	47.95	17.85	113	L60

8. Evaluation during the 2003 European heat wave

Contributors: Ordonez Carlos, Elguindi Nellie, Thouret Valérie, Athier Gilles, Cammas Jean-Pierre (CNRS-LA, Toulouse, France), Katragou Eleni, (NKUA, Athens, Greece), Flentje Harald (DWD, Hohenpeissenberg, Germany), and the modeller teams

The objective of this chapter is to evaluate the distribution of both carbon monoxide (CO) and Ozone (O3) over the European domain in August 2003 as given by GEMS-GRG models, against MOZAIC aircraft tropospheric vertical profiles and data from surface stations (EMEP and GAW). The summer 2003 was extremely dry and warm in Europe. Tressol et al. [2008] showed MOZAIC aircraft profiles above Frankfurt in July–August 2003 with strong temperature anomalies (exceeding 4° C) throughout the lower troposphere, with respect to a 11-year MOZAIC climatology. Similarly to Trigo et al. [2005], the highest positive temperature anomalies and the strongest negative anomalies of both wind speed and relative humidity were found in the MOZAIC profiles in the same period 2–14 August 2003..Less pronounced anomalies were found for the periods before (16–31 July 2003) and after (16–31 August 2003) the heat wave. Tressol et al. [2008] also showed the strong O3 and CO anomalies observed in the MOZAIC tropospheric vertical profiles.. The European summer heat wave provides a unique opportunity to evaluate the GEMS-GRG model's under such extraordinary conditions.

This part of the report is extracted from an extended study by Ordonez et al. which will be submitted to ACPD.

8.1 Datasets and methodologies

8.1.1 Datasets

Observational data set:

- MOZAIC: **M**easurements of **OZ**one, water vapour, carbon monoxide and nitrogen oxides by **A**irbus **I**n-service air**C**raft, http://mozaic.aero.obs-mip.fr/web/

The characteristics of the MOZAIC observations used for evaluation (precision, accuracy, spatial and temporal scales representativeness) are described in Chapt. 4.

In the present study, measurements of O_3, CO, temperature, wind and relative humidity of MOZAIC ascent and descent vertical profiles above European airports were used from July – August 2003. Three airports with relatively high observation frequency during the period of interest have been selected: These are Paris (2.56º E, 49.00º N), Frankfurt (8.56º E, 50.03º N), and Vienna (16.57º E, 48.11º N). MOZAIC raw data (4 s time resolution) are averaged over 150 m height intervals. For all analyses including MOZAIC data, all available measured MOZAIC profiles from 9 UTC to 18 UTC were averaged for each day during the 16 July – 31 August period. All days with MOZAIC measurements were similarly weighted, independent of the number of profiles per day. The use of only daytime data allows analysing the hours with strongest photochemical activity and the possibly lowest relative contribution of deposition processes. The 1-h model output is interpolated in space and time to the measured profiles and the location of corresponding airports. Similarly to the MOZAIC measurements, all modelled daytime profiles were averaged for each day. Our analyses presented here are mainly focused on the evolution of gas pollutants above Frankfurt, the airport with best MOZAIC data coverage. EMEP measurements are described in Chapter 7, while GAW measurements are described in Chapter 6.

8.1.2 Methodologies

The scoring methods used for the evaluation are described in Annex 7.

8.2 Offline simulations and GEMS-GRG reanalysis

Results from MOZART, TM5 and MOCAGE stand-alone runs as well as from the coupled IFS/MOZART system are shown in the following. A complete description of the models and their different versions is given in Chapter 1. Since there are only coupled runs for this period available, these are presented and discussed in this section.

All model configurations discussed here are built on 60 vertical levels from the surface up to 0.1 hPa. The MOCAGE, MOZART-v1 and TM5-v3 stand-alone runs were initially performed at horizontal resolution of 2°x2°, 1.875º x 1.895º and 3º x 2º, respectively. The coupled IFS/MOZART simulations were without and with data assimilation at a horizontal resolution of 0.5º x 0.5º. In the following, these global runs are named MOZART, TM5-HWGL, COUPL and COUPL-ASSIM, respectively. A number of sensitivity runs have also been performed with TM5, in order to investigate the impact of different processes on the ozone and CO concentrations:

- Sensitivity to higher horizontal resolution. Two runs are available:
  - MOZART t106: MOZART-v1 stand-alone run at an improved horizontal resolution of 1.125º X 1.125º
  - TM5-HWHR: Same as TM5-HWGL, but with output fields at 1° x 1° horizontal resolution over the European domain. This improved horizontal resolution results from the two-way nested zoom capability of TM5 as described in Krol et al. [2005]. Since other sensitivity runs of TM5 were evaluated at 1° x 1°, the high resolution TM5-HWHR run was considered as a control run.
- Sensitivity to anthropogenic emissions in and outside Europe. Two runs were compared with TM5-HWHR, in order to investigate the impact of emission inventories, the recirculation of pollution in Europe and the transboundary transport on pollution levels during and before/after the heat wave period:
  - TM5-HWEE: As TM5-HWHR, but with 25% additional anthropogenic emissions of NO_x, CO, and non-methane VOCs (NMVOCs) over Europe.
  - TM5-HWEN: As TM5-HWHR, but without anthropogenic emissions of NO_x, CO, and NMVOCs outside Europe.
- Sensitivity to the reduced dry deposition of O_3, as a consequence of the dryness of the vegetation accumulated during spring and summer. One available run:
  - TM5-HWDN: As TM5-HWHR, but decreasing the soil wetness factor to 0.1.

Table 1 provides only the list of models versions used in this study (including the name of the sensitivity tests done with TM5).

Name	Horiz. Resol.	Details
MOZART	1.875º x 1.895º	Stand-alone MOZART-v1 run
TM5-HWGL	3º X 2º	Stand-alone TM5 run
MOCAGE	2° x 2°	Stand-alone MOCAGE run
COUPL	0.5º x 0.5º	Coupled IFS / MOZART system
COUPL-ASSIM	0.5º x 0.5º	As COUPL but with data assimilation for CO and O_3
MOZART t106	1.125º X 1.125º	As MOZART but with improved horizontal resolution
TM5-HWHR	zoom to 1° x 1°	As TM5-HWGL but with zoom over the European domain
TM5-HWEE	zoom to 1° x 1°	As TM5-HWHR but with 25% extra anthropogenic emissions over Europe
TM5-HWEN	zoom to 1° x 1°	As TM5-HWHR but without anthropogenic emissions outside Europe
TM5-HWDN	zoom to 1° x 1°	As TM5-HWHR but with soil wetness factor set to 0.1

Table 1. Summary of analysed model runs. CTMs stand-alone runs are driven by 6-h ECMWF reanalysis, while the temporal resolution of meteorological fields in the coupled system is hourly. Colours used for the model names in this table correspond to those used in the plots.

8.2.1 Meteorology during the heat wave

A detailed analysis of the meteorology during the period of interest can be found in Tressol et al [2008]. We analyse the same periods as considered in that work – before (16–31 July 2003), during (2–14 August 2003) and after (16–31 August 2003) the heat wave – in order to test the model performance under moderate and extreme meteorological conditions.

The strong meteorological anomalies in the lower troposphere during the 2–14 Aug 2003 period are illustrated in Figure 1, showing the evolution of temperature, relative humidity and wind speed at 850 hPa above Frankfurt from 15 Jul to 31 Aug 2003. There is an overall good agreement between measured and modelled meteorological fields, with the exception of lower temperatures in MOZART over Frankfurt (Fig 1, top) but not over Paris (not shown). Since both MOZART and TM5 are driven by the same meteorological data, these results confirm that the main meteorological features found in the MOZAIC data in summer 2003 are typically well reproduced by the 2003 ECMWF reanalysis providing the meteorological input for the CTMs. However, the interpolation of non-collocated coarse grid cells from different models may lead to differences for some locations.

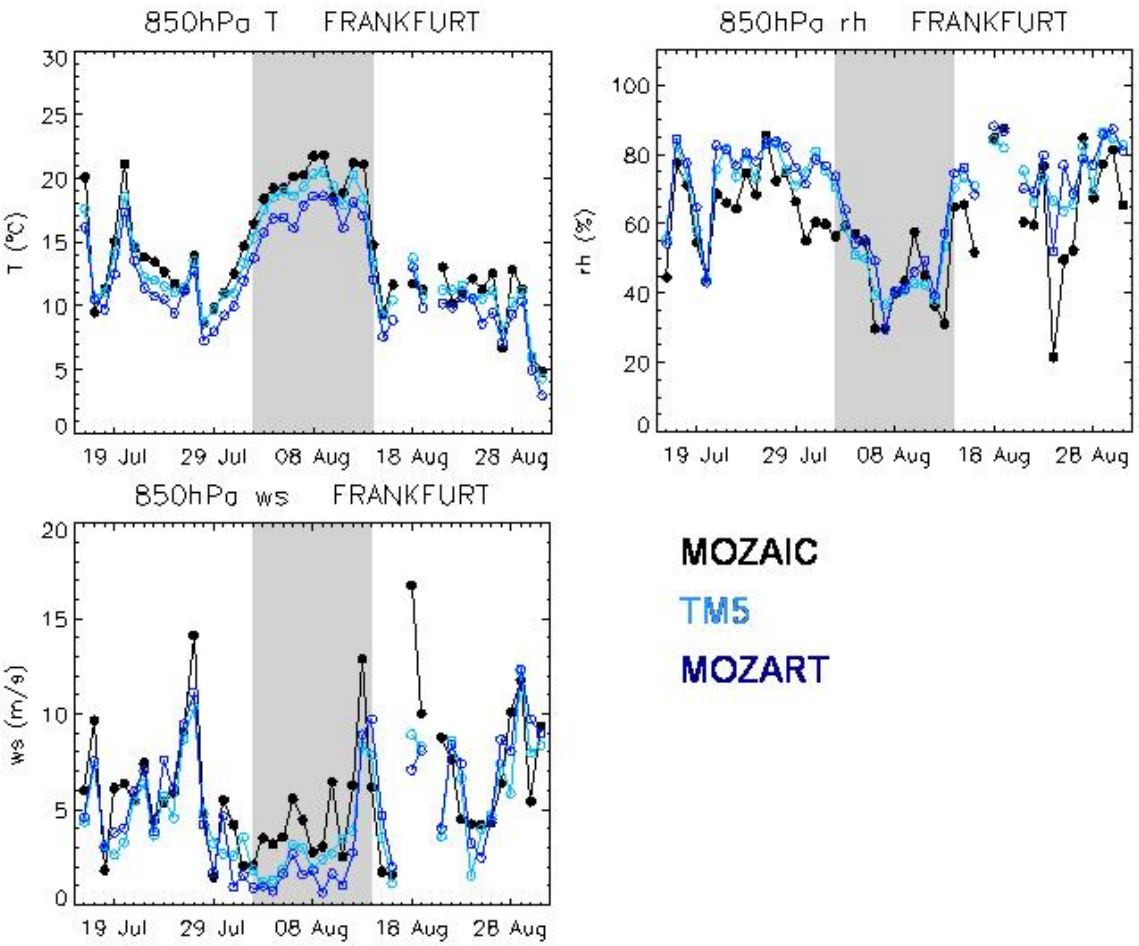


Fig 1. Time series of daytime (9 – 18 UTC) average temperature (top), relative humidity (middle) and wind speed (bottom) at 850 hPa above Frankfurt for the period 15 July – 31 August 2003. The area shaded in grey denotes the heat wave period (2-14 Aug). Black lines represent MOZAIC measurements; dark blue is used for MOZART output and light blue for TM5.

The MOZART and TM5 model output as well as the MOZAIC measurements of temperature, relative humidity (RH) and wind speed were analysed at different atmospheric pressure levels – 950, 850, 700, 500 and 300 hPa for favourable conditions for photochemical smog formation in July and August 2003. Elevated temperatures are found from the ground up to 700 hPa for Paris and Frankfurt, but not for Vienna. During the same period (2-14 Aug 2003), RH is low in the boundary layer and up to 850 hPa for the three airports, and low wind speed is observed throughout the whole troposphere above Paris and Frankfurt. The simultaneous decreases in relative humidity and wind speed reflect the dryness and stagnation of air masses over Central/Western Europe during that fortnight.

8.2.2 Air pollution during the heat wave

A first comparison of modelled ozone from the reference MOZART, TM5 and MOCAGE runs with GAW/EMEP surface measurements shows that models have some difficulties to reproduce the elevated ozone mixing ratios during July and August 2003, particularly over central Europe in August (Fig. 2). Surface ozone fields are similar for both MOZART and TM5, although ozone levels in MOZART are somewhat lower over parts of Scandinavia and the Northern European domain, as well as over the Mediterranean. MOCAGE fields are especially higher over Central Europe, the Atlantic

Ocean and the Mediterranean regions. A detailed picture of the evolution of modelled daytime surface O_3 and CO fields for the 3 sub-periods is shown for MOZART in Figure 3. The high temperatures and stagnation of air masses during the heat wave period resulted in enhanced surface O_3 mixing ratios over western/central Europe. This is also found in TM5 and MOCAGE (not shown) results. Although highest surface ozone levels in TM5 are found over the Mediterranean when both daytime and night-time data were considered for the whole month of August (Fig. 2), TM5 simulates even more ozone over the continent during daytime, probably because of a lower relative contribution of deposition processes. The stagnation of air masses also led to high levels of some primary pollutants in the boundary layer, as for example seen in the MOZART surface CO mixing ratios (Fig. 3). The strongest CO sources during the heat wave were found over areas affected by forest fires (e.g. Portugal and Italy) and densely populated areas (e.g. England and the Netherlands).

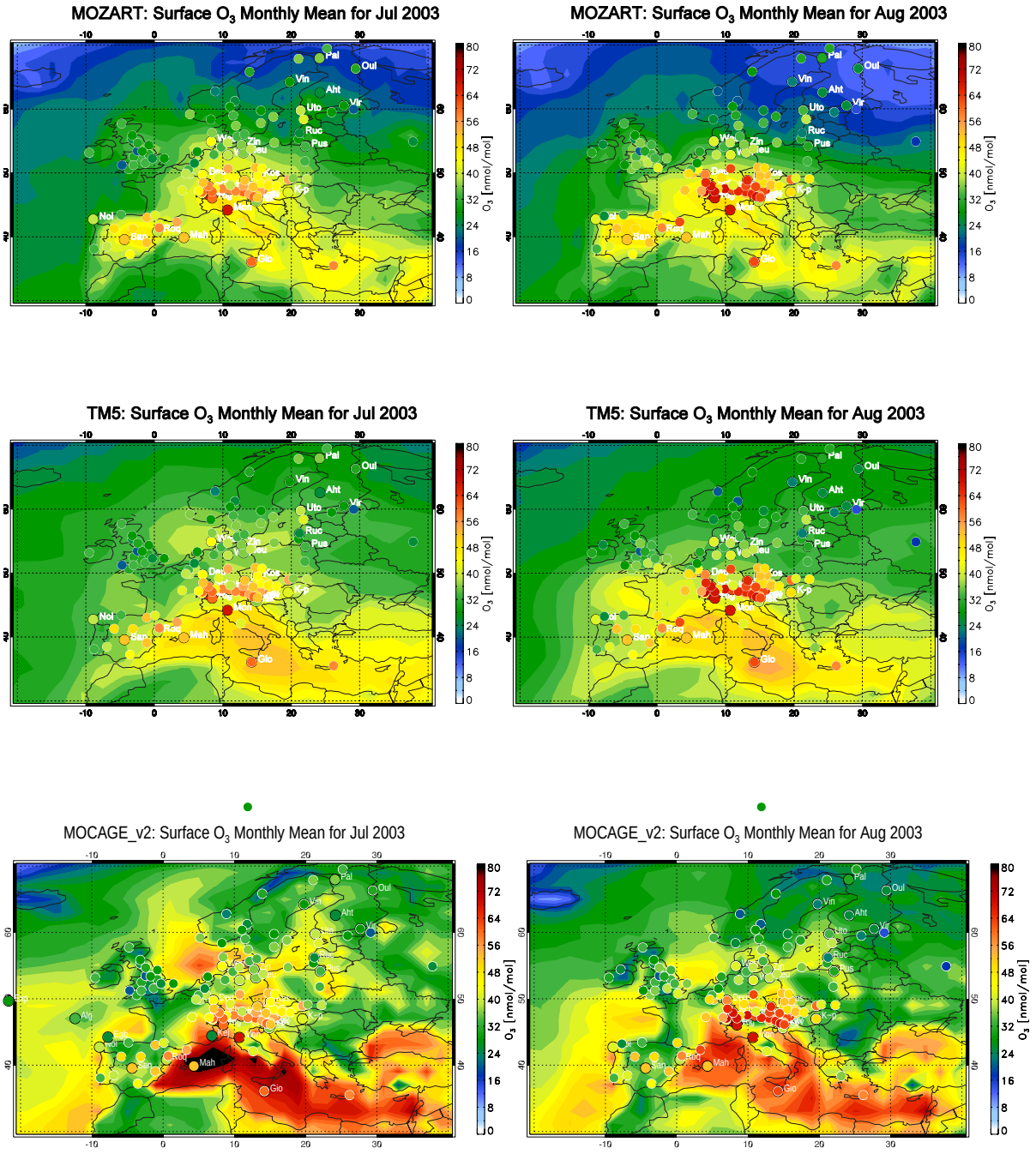


Fig 2. Monthly averaged surface ozone in (left) July and (right) August as modelled by (top) MOZART, (middle) TM5-HWGL and (bottom) MOCAGE. Measurements from the GAW and EMEP networks are overplotted using the same colour scale.

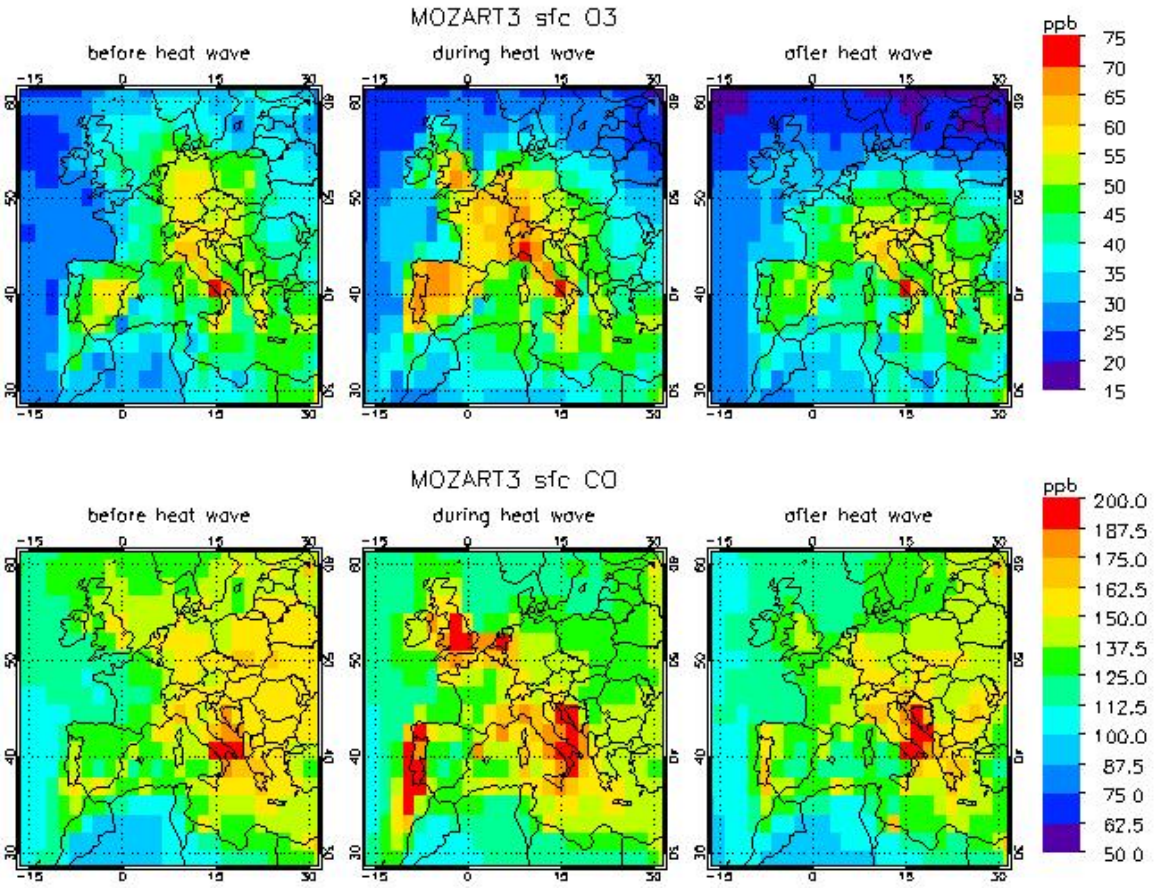


Figure 3. Evolution of average surface (top) O_3 and (bottom) CO simulated by MOZART at daytime (9 – 18 UTC) for the 3 periods of interest: before (16-31 Jul), during (02-14 Aug) and after (16-31 Aug) the heat wave.

Time series of MOZAIC measurements and stand-alone CTMs output illustrate that the highest ozone and CO mixing ratios in the lower troposphere over Frankfurt are found during the heat wave period (Fig 4). The figure also shows that CTM stand-alone runs (MOZART, MOZART t106 at high resolution, TM5-HWGL and TM5-HWHR at high resolution) underestimate corresponding MOZAIC measurements of O_3 and CO. MOCAGE is the exception, with an O3 overestimation throughout the period. Despite the overall underestimation (or overestimation), all models capture very well the mainly synoptically driven evolution of O_3, but not the measured O_3 levels during the heat wave. The coarse resolution TM5 (light blue) and MOZART (dark blue) runs perform similarly for O_3, with typically lower biases for TM5 and particularly over Frankfurt at the end of the heat wave period. Nevertheless, TM5 clearly has problems in reproducing measured CO levels. The decrease in both O_3 and CO in the middle of the heat wave period, which is a consequence of the ventilation by a low-pressure system, is reproduced by the models. However, they have difficulties in reproducing the previous strong increase of CO, which is at least partly associated with the transport from Portuguese fires [Hodzic et al., 2007; Tressol et al., 2008]. This was expected since the model simulations are based on monthly averaged emission data without the strong CO sources caused by Portuguese fires. Above Frankfurt, there is a significant improvement in MOZART t106 with respect to the coarse resolution MOZART run, while no apparent improvement is typically found for TM5 at increased resolution. This is because the MOZART t106 run has higher horizontal resolution than the coarse MOZART run for the whole globe, while TM5-HWHR only zooms to a higher resolution (compared to TM5-HWGL) over a limited area. In fact, the TM5 model running on a global 3° x 2° gid already includes some features at higher resolution: the surface processes (emissions and depositions) are evaluated on a global 1° x 1° resolution first and subsequently reduced to a global resolution of 3° x 2°. As a consequence, the differences with and without zooming are generally small.

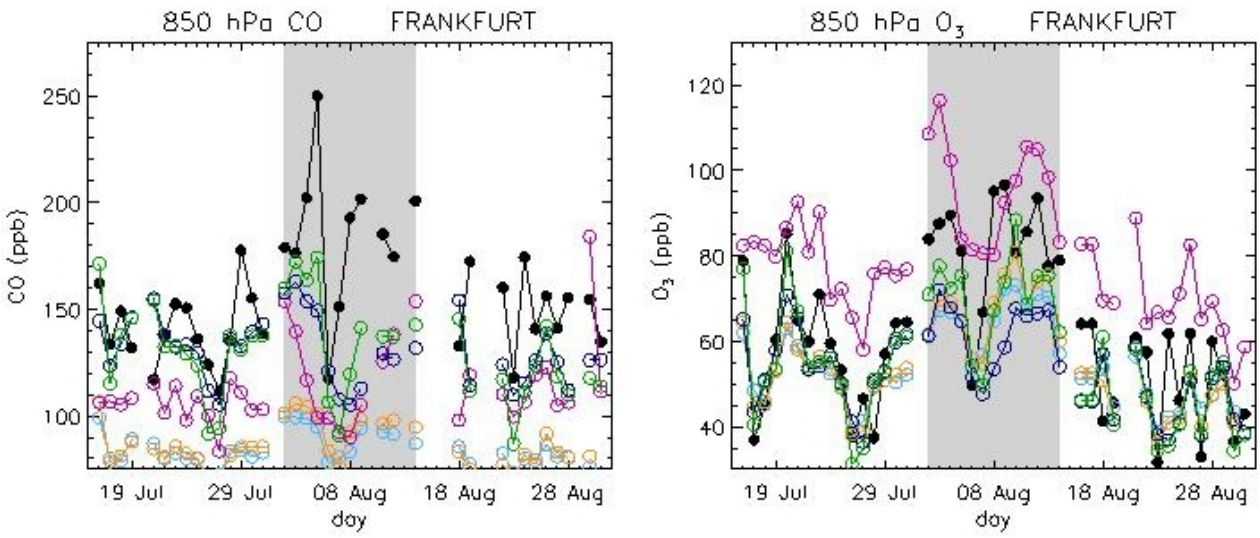


Figure 4. Time series of daytime average (top) O_3 above Paris, (middle) O_3 above Frankfurt and (bottom) CO above Frankfurt at 850 hPa for the period 15 July – 31 August 2003. The gray-shaded area denotes the heat wave period (2-14 Aug). Black lines represent MOZAIC measurements, light blue is used for TM5-HWGL, orange for TM5-HWHR, dark blue for MOZART, green for MOZART t106, and magenta for MOCAGE.

Figures 5 and 6 provide a detailed evaluation of O_3 and CO profiles above Frankfurt. Figure 5 shows the average vertical profiles of O_3 and CO measured by MOZAIC and simulated by various model runs for the three periods before, during and after the heat wave period. Figure 6 illustrates the modified normalised mean biases for the same periods and for 3 atmospheric layers: the planetary boundary layer (p>850 hPa), the free troposphere (850–650 hPa) and the mid-troposphere (650–300 hPa). Results from the main MOZART-v1 (MOZART, MOZART t106), TM5 (TM5 HWGL, TM5 HWHR) and MOCAGE runs as well as from the coupled IFS/MOZART runs (COUPL and COUPL-ASSIM) are shown.

All model runs, **with the exception of MOCAGE,** underestimate O_3 throughout almost the entire troposphere during the heat wave. However, modelled and measured O_3 are close together within their range of uncertainties, both in the mid- and free troposphere but not in the boundary layer (Fig 5, left). Similar negative biases are found for all models in the mid- and free troposphere during the three periods (Fig 6, left) although the coupled system performs less good than the CTM stand-alone runs. A different picture is found for p > 850 hPa, where the strongest model underestimation during the heat wave period was observed while biases are small and positive (except for TM5 before the heat wave) in the other two periods. One of the most remarkable features is the improved modelled O_3 for the lowest levels during the heat wave in MOZART t106 (B_n'□-20%), with respect to the coarser MOZART run (B_n'□-40%). This is most probably due to a better simulation of the accumulation and recirculation of pollution within the lowest levels at improved horizontal resolution. In addition, the changed shape of the vertical profile at higher resolution leads to a much better correlation in MOZART t106 (r=0.82) as compared to MOZART (r=0.23) during the heat wave period. Note the lack of improvement in the modelled O_3 including the coupling, with the exception of the lower levels during the heat wave (compare COUPL in violet with MOZART in dark blue). The differences between the two coupled runs with and without assimilation are small for all periods, which is due to the reduced sensitivity of satellite measurements to O_3 in the troposphere.

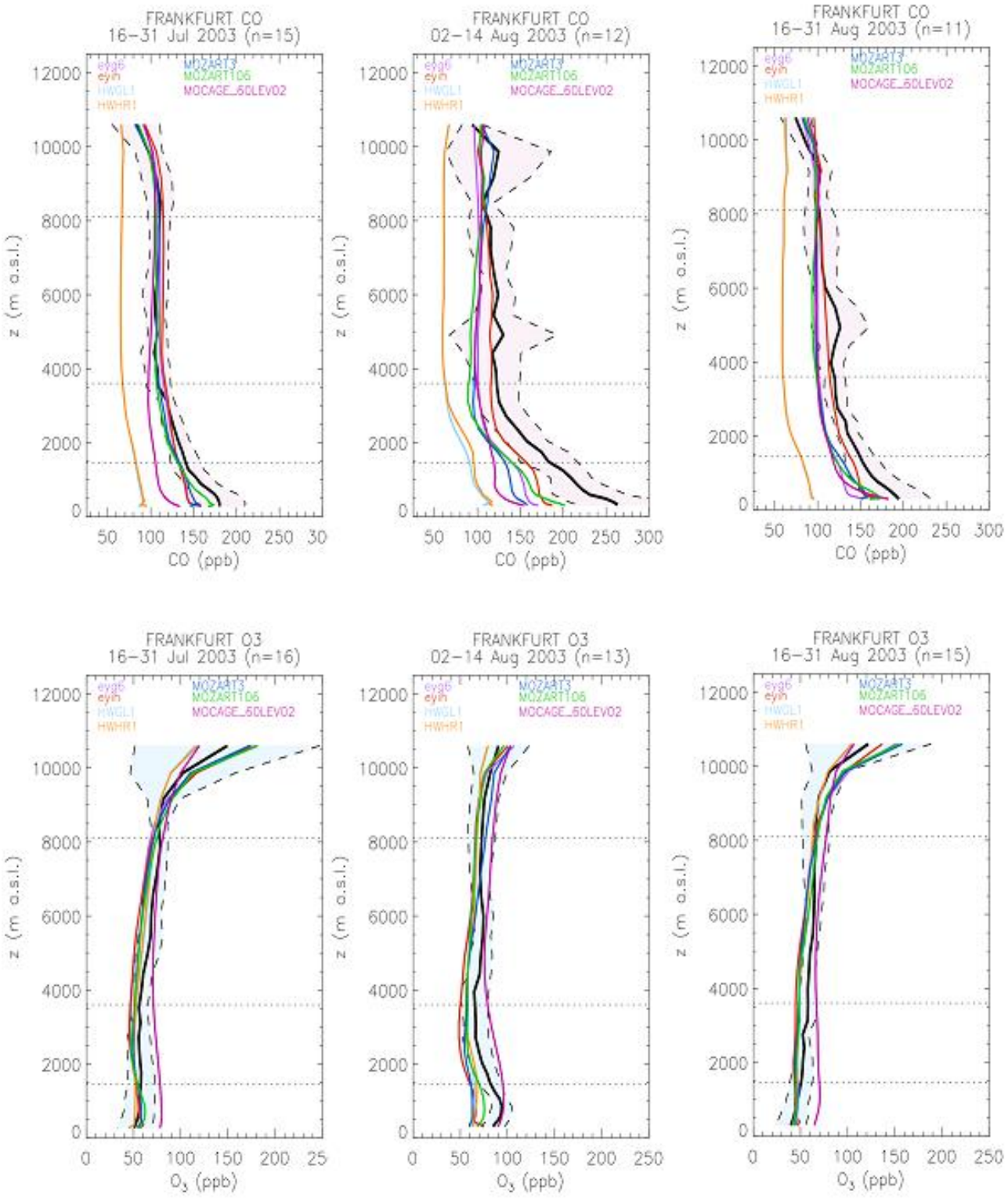


Figure 5. Averages (lines) and standard deviations (shaded area in the case of MOZAIC) of (bottom panel) O_3 and (top panel) CO mixing ratios above Frankfurt during the three periods before, during and after the heat wave. MOZAIC measurements in black, TM5-HWGL in light blue, TM5-HWR in orange, MOZART in dark blue, MOZART t106 in green, MOCAGE in magenta, COUPL in violet and COUPL-ASSIM in red. Only daytime data (9–18 UTC) were considered. The number of days

with data is indicated on top of the plots. The horizontal dotted lines represent, from bottom to top, the 850, 650 and 300 hPa pressure levels, respectively.

All model runs underestimate CO in the troposphere during the heat wave period, particularly in the free and the lower troposphere (Fig. 5, right). Although TM5 reproduces the measured profile shape nicely (~0.98 correlation coefficient for TM5-HWGL and ~0.94 for MOZART), the largest biases are found for TM5 throughout the troposphere. The zoomed version (TM5-HWHR) leads only to marginal improvements. MOZART has more problems to reproduce CO levels in the lower troposphere than elsewhere, particularly during the heat wave period where the negative bias is larger than -40% (Figs. 5 and 6). Similarly to what was found for O_3, the MOZART run with improved resolution (MOZART t106) clearly reduces the negative biases during the heat wave, but only in the lower troposphere. This may be because the higher horizontal resolution of this run is better simulating transport and residence of air masses close to the PBL. Lower biases are also found for the coupled IFS/MOZART (COUPL, red) model compared to the basic MOZART stand-alone run (dark blue). Since the chemistry in the coupled system is the same as in the MOZART runs, the improved results are most probably due tothe improved meteorology and transport phenomena (fully modelled meteorology in the coupled model compared to 6-h meteorological feedback in MOZART). In addition, the assimilation of CO columns from MOPITT further improves the comparison with MOZAIC data, reducing negative biases to less than 20 % for all tropospheric levels and periods but not in the lower troposphere during the heat wave.

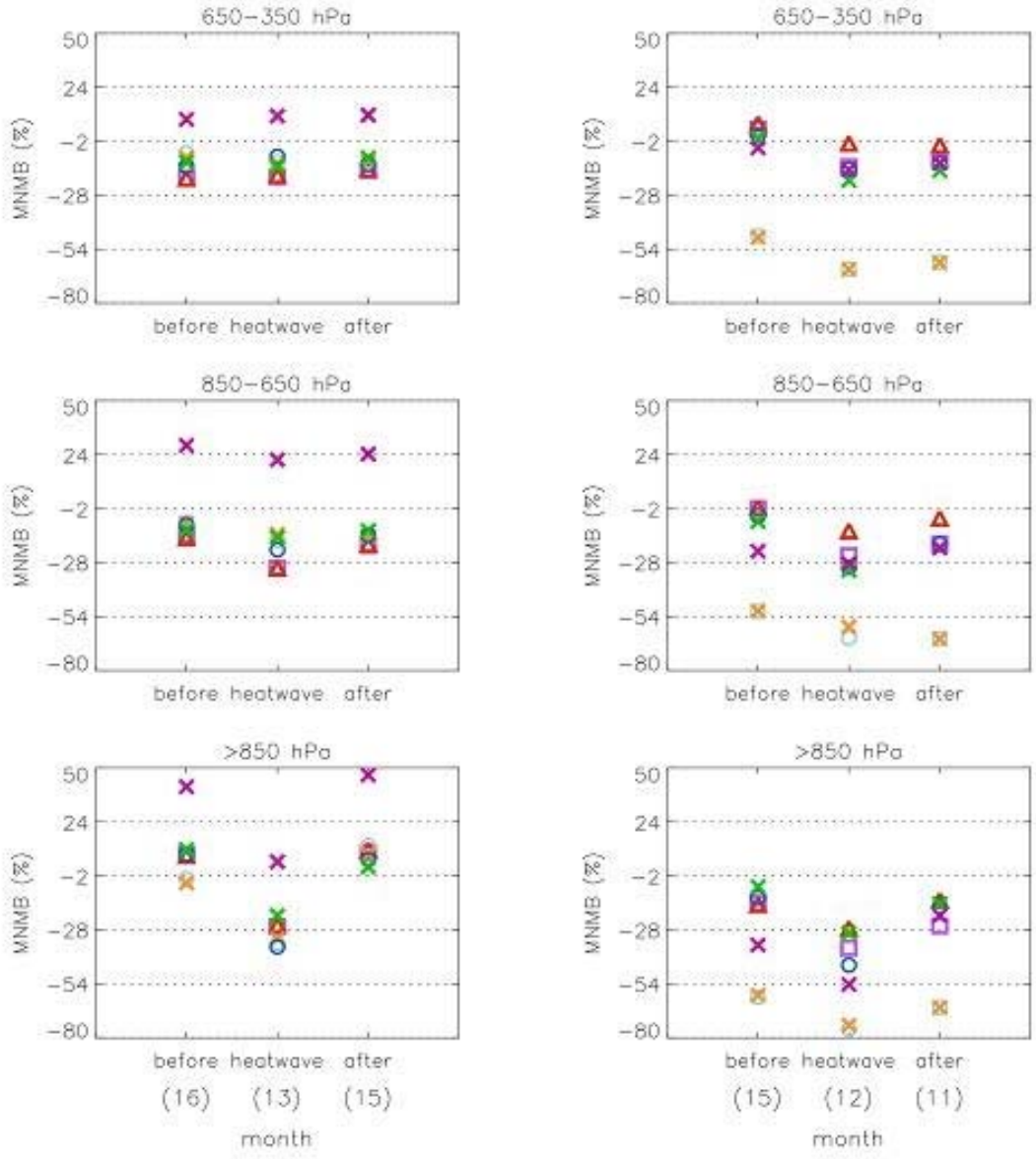


Figure 6. Modified normalised mean biases of (left) O_3 and (right) CO mixing ratios for 3 atmospheric layers above Frankfurt during the 3 periods of interest. The total number of days of each period is shown in brackets. TM5-HWGL in light blue, TM5-HWR in orange, MOZART in dark blue, MOZART t106 in green, MOCAGE in magenta, COUPL in violet and COUPL-ASSIM in red.

8.3 Sensitivity Experiments

Sensitivity studies with respect to the horizontal resolution and a comparison to MOZAIC measurements were shown in the section above. Some further results on the resolution issue are shown in this section. Further sensitivity runs on emissions as well as on the influence of soil wetness on the dry deposition of ozone have been performed with TM5 (see summary in Section 8.2 and Table 2). Results from these sensitivity runs are not presented here, but Figure 7 provides the main conclusions. Details can be found in Ordonez et al., (in preparation).

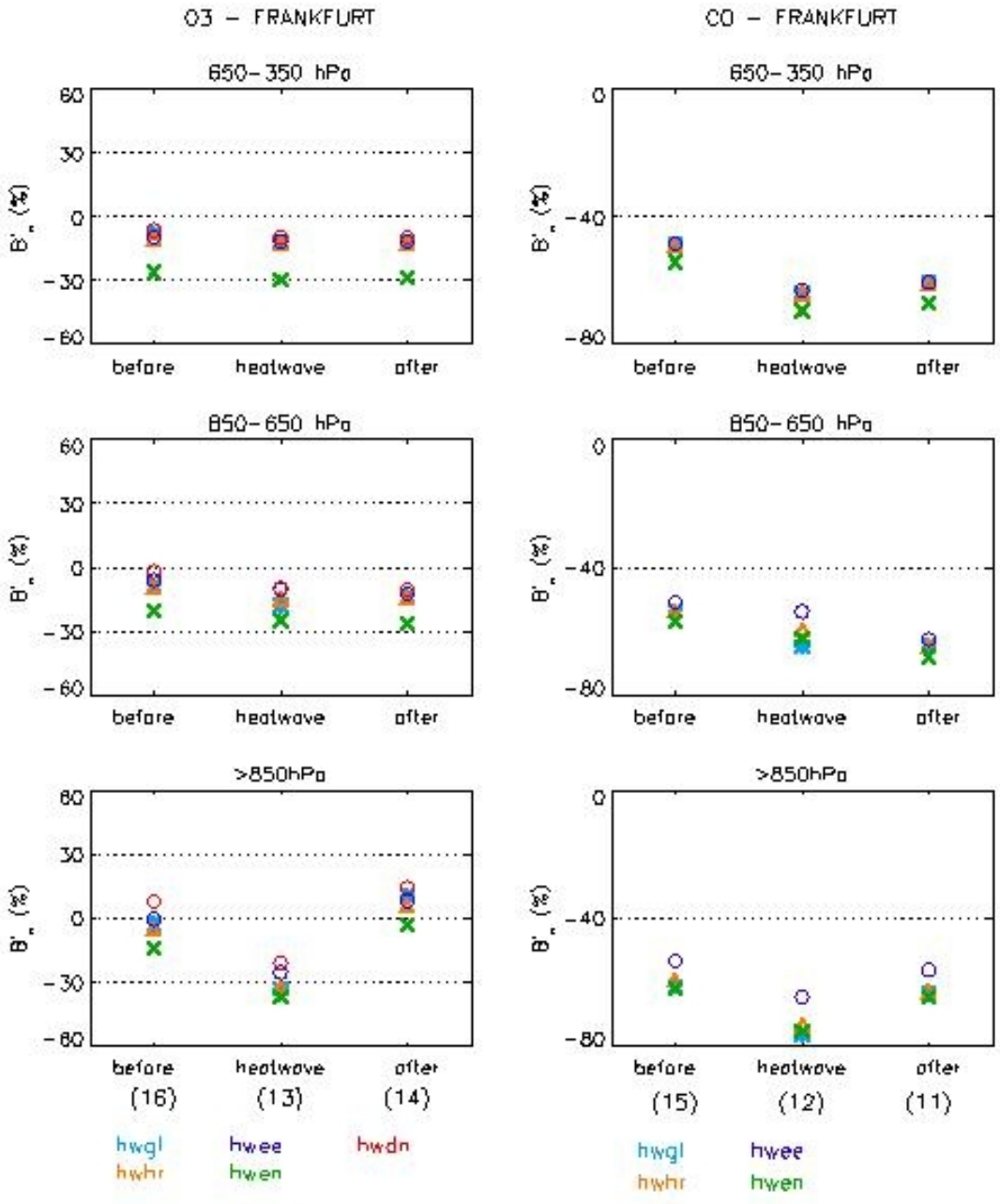


Figure 7. Modified normalised mean biases for daytime (left) O_3 and (right) CO mixing ratios from different TM5 runs (see summary in Table 2) for three atmospheric layers above Frankfurt during the three periods of analysis. The total number of days per period are shown in brackets.

It seems that something is missing here…at least a description of what is seen in the images. It might be better to resolve this section and to add it simply at the end of the section before.

8.4 Summary

We performed a comparison of three reference CTM stand-alone runs (TM5, MOCAGE and MOZART) with MOZAIC measurements during the summer heat wave in August 2003. We further analyzed the performance of two reanalysis runs using the coupled IFS/MOZART system and we carried out sensitivity studies with respect to the horizontal (model) resolution, emission rates and the dry deposition at the surface. Three periods were analyzed in detail: before the heat wave (16–31 July

2003), during the heat wave (2–14 Aug 2003) and after the heat wave (16–31 Aug 2003). Largest meteorological anomalies and highest pollutant mixing ratios were found during the heat wave period..

The following conclusions are drawn from the reference CTM stand-alone runs and the reanalysis runs:

- The global CTMs (MOZART and TM5) reasonably well reproduce the meteorological features as well as the cycles of O_3 and CO during the analysed periods. However, these two models tend to underestimate the real pollution levels, particularly in the lower troposphere during the heat wave period.

- MOCAGE is the only model slightly overestimating O3 during the heat wave in the lower troposphere. This lowest bias actually reflects the general overestimation by MOCAGE (see Chapter 4), which is only reduced here due to the anomalous high ozone event during the heat wave.

- Especially TM5 has difficulties in reproducing measured CO mixing ratios.

- Lack of improvement in modelling O_3 by the coupled run (MOZART/IFS coupled run, COUPL) compared to the reference MOZART stand-alone run, with the exception of the lower atmospheric levels during the heat wave. The differences between the two coupled runs with (COUPL) and without (COUPL-ASSIM) satellite data assimilation are also small, which is due to the reduced sensitivity of satellite UV measurements to O_3 absorption in the lower troposphere.

- Lower CO biases are found for the coupled IFS/MOZART run (COUPL) compared to the base MOZART stand-alone run. Since the model chemistry in the coupled system is the same as in the MOZART runs, the improved results might be due to both improvements in the meteorology (full meteorological modelling in IFS compared to the only 6-h meteorological feedback in MOZART) and a better treatment of transport process following the improved horizontal resolution (0.5° x 0.5° in IFS compared to 1.875° X 1.895° in MOZART). In addition, the assimilation of CO columns from MOPITT (COUPL-ASSIM run) further improves the comparison with MOZAIC data, reducing negative biases to less than 20 % for all tropospheric levels and periods, with the exception of the lower troposphere above Frankfurt during the heat wave. Therefore, the assimilation efficiently overrules some of the deficiencies in the transport, chemistry or emissions in the model.

Conclusions from the sensitivity runs:

- Results from the high resolution MOZART run (MOZART t106) suggest that the coarse resolution of the global CTMs can partly explain the differences relative to the observations. It is likely that the increased spatial resolution improves the simulation of the horizontal/vertical transport which is then responsible for the enhanced concentration of pollutants within the lowest levels in this run.

- The TM5 runs with increased emissions in Europe (HWEE) and without anthropogenic emissions outside Europe (HWEN) indicate that both uncertainties in European emissions and limited long-range transport in the models have a small effect and cannot completely account for the underestimation of CO and O_3 by the models.

- The TM5 sensitivity run with reduced wetness factor (HWDN) leads to the strongest reduction in the negative O_3 biases but still underestimates ozone concentrations in the PBL during the heat wave period. Low dry deposition as a consequence of the dryness of vegetation is likely to have contributed to the elevated levels of both ozone and other relevant species.

Some of the outstanding issues:

- The factors and processes mentioned here (horizontal resolution, emissions and influence of the soil stress on surface dry deposition) have a significant impact on the simulation of pollution during the heat wave period. However, there may be other mechanisms playing an important role in explaining the large O_3 concentrations during the heat wave.

- Some of the processes that might be insufficiently represented by the models are the horizontal / vertical transport, the chemistry, the anthropogenic emissions (monthly emissions are used and therefore daily and monthly cycles are not included) as well as the effect of high temperatures on biogenic emissions and on the evaporation of anthropogenic VOCs. Together with moderately high NO_x concentrations the latter may increase ozone levels.

- The different magnitude and even different sign of the O_3 biases in the PBL over Frankfurt, Paris and Vienna outside the heat wave period may be related to the limited representativeness of these airports. Some local effects at these airports, particularly in the lower tropospheric levels, are likely to be at least partly responsible for those differences, e.g. local anomalies due to surface emissions, city plumes, ….

It is important to bear in mind that the model runs analysed here have a coarse horizontal resolution, and all of them use monthly emissions at 1°x1° which have been interpolated to the size of the corresponding model grid cells. This surely has an impact on the recirculation and accumulation of pollution within the boundary layer, and therefore on the chemical processes and especially on the ozone levels over central/western Europe during and outside the heat wave period. The meteorological and photochemical modelling of such an extreme episode would require higher spatial resolution and temporally better resolved emission data. It is beyond the scope of GEMS to further develop the global CTMs and the coupled system in a way that such extreme cases can be analysed quantitatively. The objective is to develop a coupled system providing forcing fields (initialisation and boundary conditions) that will include signatures of anomalies which are further propagated to regional air quality models on a daily basis. This is ongoing (and very promising) work within the RAQ (Regional Air Quality) sub-project of GEMS and other running EU framework projects.

8.5 References

1. Hodzic, A., Madronich, S., Bohn, B., Massie, S., Menut, L., and Wiedinmyer, C.: Wildfire particulate matter in Europe during summer 2003 - meso-scale modeling of smoke emissions, transport and radiative effects, Atmos. Chem. Phys., 7, 4043-4064, 2007.
2. Krol, M., Houweling, S., Bregman, B., van den Broek, M., Segers, A., van Velthoven, P., Peters, W., Dentener, F., and Bergamaschi, P.: The two-way nested chemistry-transport zoom model TM5: Algoritm and Applications, Atmos. Chem. Phys., 5(2), 417-432, 2005.
3. Ordonez et al., Global model simulations of air pollution during the European 2003 heat wave. GEMS project. In préparation.
4. Trigo, R. M., García-Herrera, R., Díaz, J., Franco Trigo, I., and Valente, M. A.: How exceptional was the early August 2003 heatwave in France?, Geophys. Res. Lett., 32, L10701, doi: 10.1029/2005GL022410, 2005.
5. Tressol, M., Ordonez, C., Zbinden, R., Brioude, J., Thouret, V., Mari, C., Nedelec, P., Cammas, J.-P., Smit, H., Patz, H.-W., and Volz-Thomas, A.: Air pollution during the 2003 European heat wave as seen by MOZAIC airliners**,** Atmos. Chem. Phys., 8, 2133-2150, 2008

9. Evaluation of boreal fires impact with routine aircraft data

Contributors: Elguindi Nellie, Ordonez Carlos, Thouret Valérie, Athier Gilles, Cammas Jean-Pierre (CNRS-LA, Toulouse, France), and the modeller teams

The first objective of this chapter is to evaluate the global distribution of carbon monoxide (CO) given by GEMS-GRG models against MOZAIC aircraft tropospheric vertical profiles. Wildfires which occurred over North America during summer 2004 provide a unique opportunity to evaluate the GEMS-GRG model's ability to be used in process studies, such as the long-range transport of biomass fire plumes. The second objective of this chapter is to assess the ability of the IFS coupled assimilation system and standalone CTMs to simulate and transport CO originating from wildfires and to assess the satellites data assimilation procedures used in the GEMS-GRG subproject.

This part of the report is extracted from an extended study by Elguindi et al. which will be submitted to ACPD.

9.1 Datasets and methodologies

9.1.1 Datasets

- MOZAIC: **M**easurements of **OZ**one, water vapour, carbon monoxide and nitrogen oxides by **A**irbus **I**n-service air**C**raft, http://mozaic.aero.obs-mip.fr/web/

The characteristics of the MOZAIC observations used for evaluation (precision, accuracy, spatial and temporal scales representativeness) are described in Chapter 4.

9.1.2 Methodologies

9.2 Offline simulations

	CTM	**version**	**available period**
Stand-alone runs	MOZART	mozart3.1/run_era1_v10	20030101-20041231
	TM5	V10	20030701-20041231
	MOCAGE	simu2003/60lev01 simu2003/60lev02	20030101-20031231 20021101-20031231 20040601-20040930

Table 9.1: Details of the standalone CTM runs available for evaluation.

In this section we compare modelled estimates of monthly averaged CO from the stand-alone CTMs (Table 9.1, see also Annex 6) to the observed MOZAIC CO data measured near several airports during

the year 2004 (Figures 9.1-5). The modified normalized mean bias (MNMB) for CO are calculated at different atmospheric levels for each month using daily averaged profiles from various airports. The different atmospheric levels are defined as follows: surface layer (>950 hPa), boundary layer (950-850 hPa), free troposphere (850 hPa up to 1 km below tropopause), and upper troposphere (1 km below the tropopause up to the tropopause, where tropopause is defined as the highest level with a lapse rate less than 2K/km). In order to examine the effect the time of day the observations were taken has on model biases, we calculate the biases for both daytime (9:00 – 18:00 local time) and nighttime (0:00 – 9:00 and 18:00 – 0.00 local time) hours.

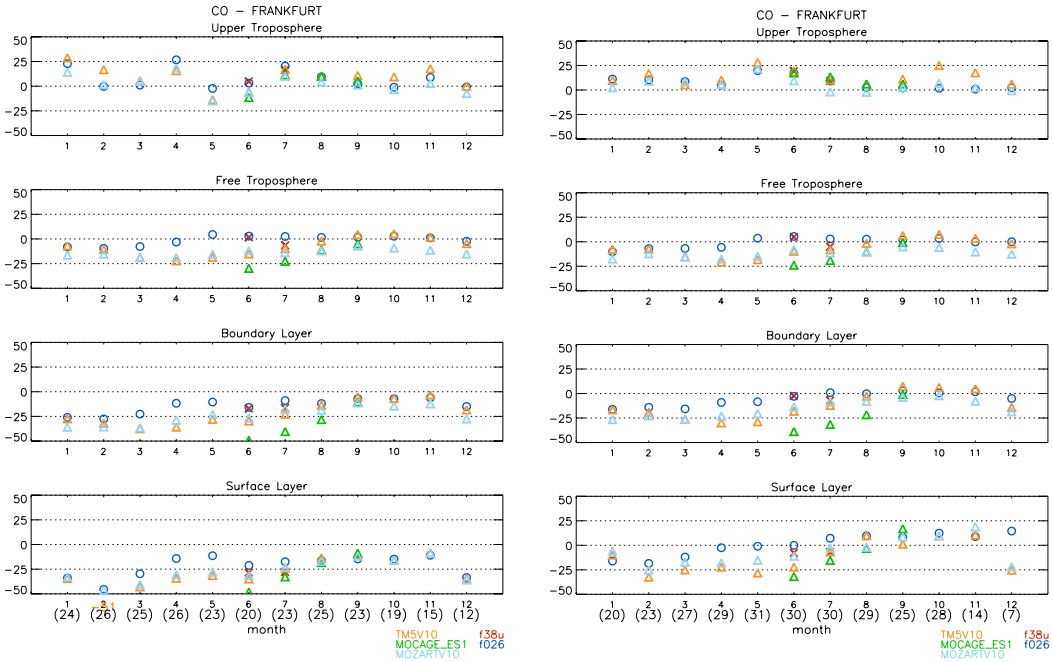


Figure 9.1: Monthly Modified Normalized Mean Bias (MNMB in %) based on daytime (left) and nighttime (right) daily averaged profiles over Frankfurt. The numbers in parenthesis at the bottom of the graph represent the number of days with available profiles.

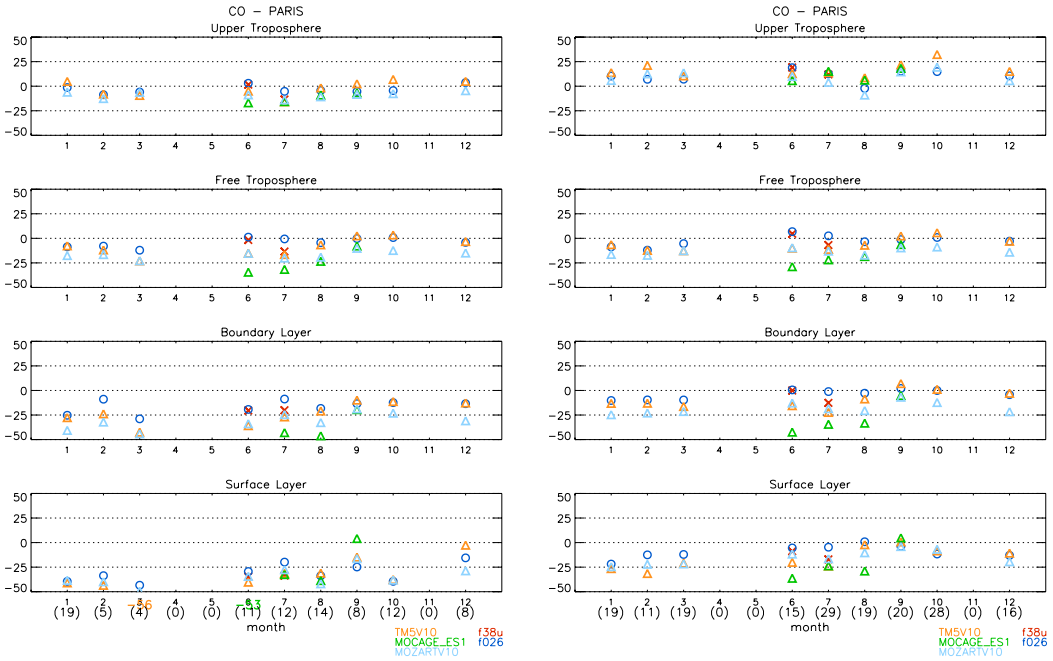


Figure 9.2: Same as Figure 9.1 except over Paris.

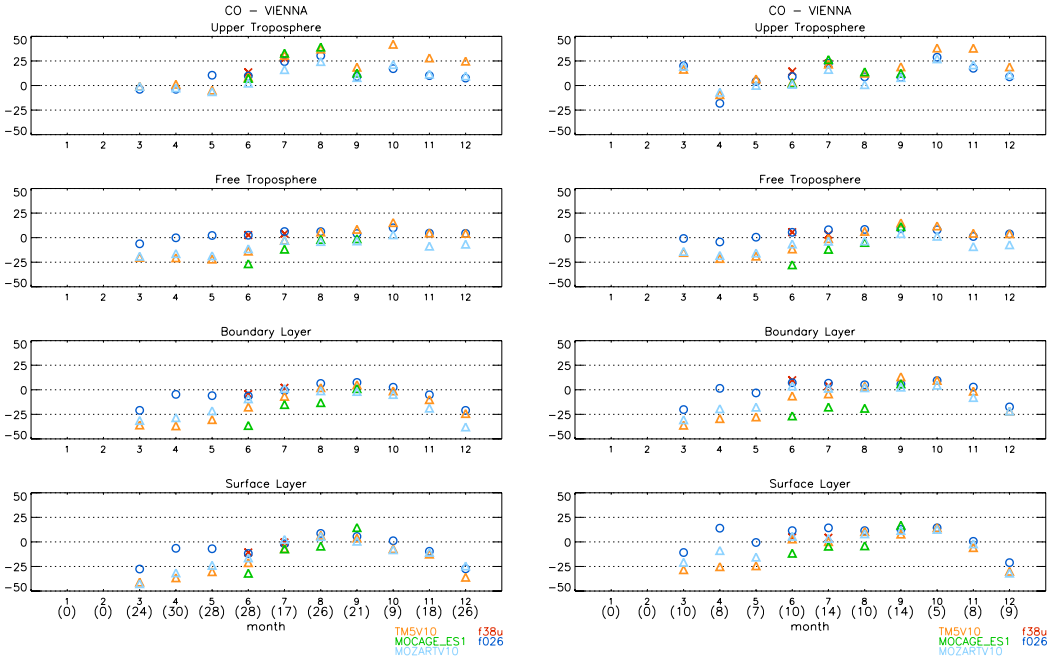


Figure 9.3: Same as Figure 9.1 except over Vienna.

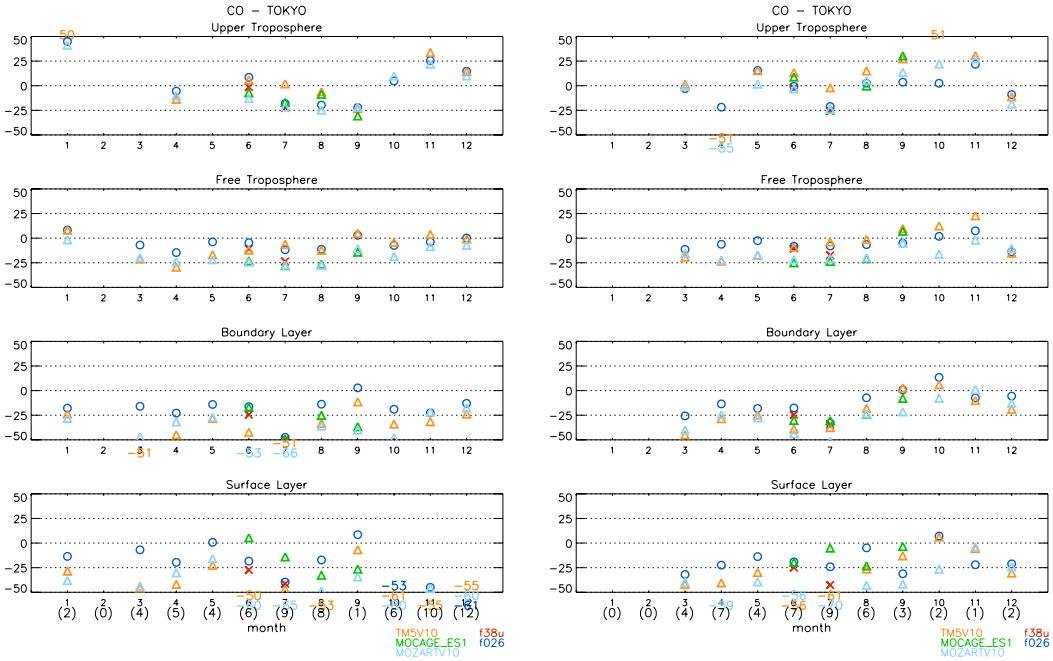


Figure 9.4: Same as Figure 9.1 except over Tokyo.

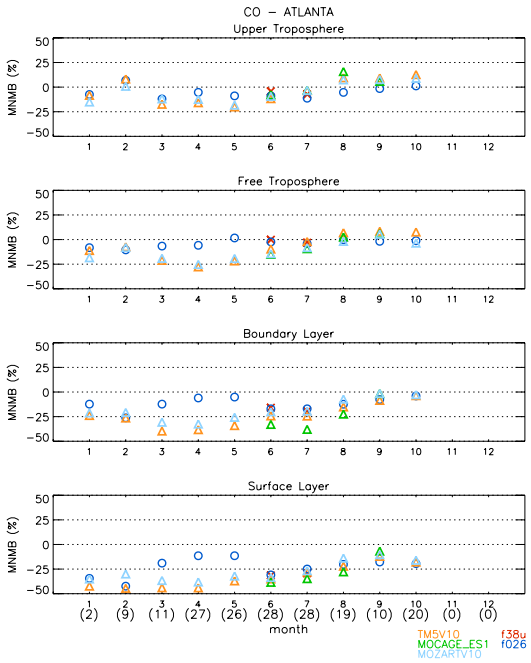


Figure 9.5: Same as Figure 9.1 except over Atlanta for daytime measurements.

In general, the model biases are smallest in the free and upper troposphere, mostly between +/- 25%. In the free troposphere, the models tend to underestimate CO, especially in the winter and spring. Similarly, in the surface and boundary layers CO tends to be underestimated, with significantly larger biases ranging from +/-50%.

There are some noticeable differences in the biases between daytime and nighttime hours at the surface and boundary layer. Namely, the nighttime biases tend to be lower than daytime biases. Because diurnal changes are not represented in the emission data, this indicates that the emissions actually represent the minimum concentrations. Over Frankfurt, Paris and Vienna the nighttime biases are up to 25% less than during the daytime in the surface layer and up to 15% less in the boundary layer. There is almost no difference in the troposphere, however in the upper troposphere there are some discrepancies that may be explained by disparities in the density of data. As well, the differences between daytime and nighttime biases over Tokyo are ambiguous as in some cases the nighttime bias is lower and in other cases the inverse is true.

A coupled of representative examples of daily averaged CO timeseries are presented in figure 9.6 in order to show how well the stand-alone CTMs reproduce the day-to-day variability. Overall, the models do reasonably well in reproducing the variability, although they tend to underestimate the concentrations, particularly at low levels (850 and 700 hPa). The CTM biaises are somewhat better at the upper levels (500 and 300 hPa). In general at the lower levels, MOCAGE tends to have the largest biases, MOZART the smallest and TM5 falling somewhere in between. For example, over Frankfurt in July at 750 hPa MOZART has a MNMB of -8%, while TM5 and MOCAGE have MNMBs of -14% and -31%, respectively. And over Atlanta at 850 hPa in June MOZART has a MNMB of -18%, while TM5 and MOCAGE have biaises of -19% and -29%, respectively. At the upper levels, MOZART generally still has the smallest biases, while MOCAGE and TM5 tend to perform more equally. The IFS coupled system with assimilation, f026, does a better job in reproducing both the variability and the concentrations than the stand-alone CTMs. However, still there are some peaks that even the IFS coupled system is not able to reproduce, such as on 22 July over Frankfurt, which might be due to a couple of reasons. It generally takes the MOPITT satellite about 4 days to get full coverage of the

earth, thus specific events such as this may not be captured by MOPITT. Another possible reason is that meso-scale CO plumes may not be seen by MOPITT, so are therefore not assimilated. In general, the control simulation, f38u, does not capture the daily variability in CO as well as the f026 simulation. This is especially evident over Frankfurt where CO peaks during the latter half of August are reproduced by the f026 simulation but not the f38u simulation. This is clear indication of the improvements gained by the assimilation process.

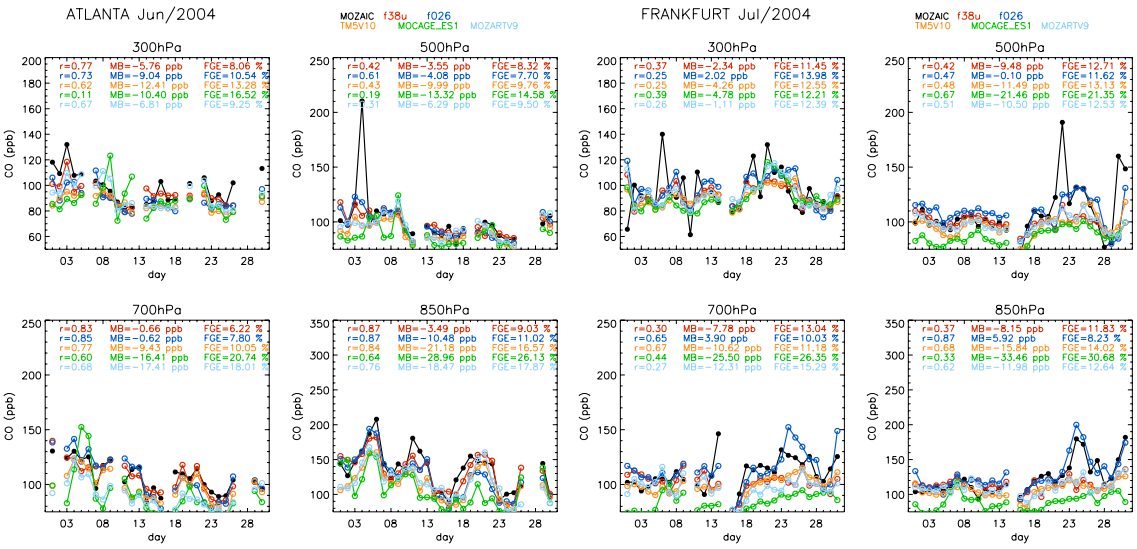


Figure 9.6: Timeseries of CO (ppbv) for June 2004 daily averaged profiles over Atlanta (left) and Frankfurt (right) at different levels in the atmosphere. Note: "r" is the correlation coefficient, "MB" is the modified normalized mean bias and "FGE" is the fractional growth error.

We find that the stand alone CTM's biases range from 0 to +/- 25% in the free troposphere and 0 to +/- 50% in the surface and boundary layers, with a tendency towards underestimating CO. Table 9.2 shows the monthly average biases within the three lowest layers. On the yearly mean, TM5 and MOZART have approximately equal performances with an absolute bias of 20% in the surface and boundary layers, decreasing to 10% in the free troposphere. MOCAGE seems to perform worse than the other CTMs, except during autumn for which all standalone CTMs have the lowest biases (<10%).

Table 9.2: Monthly Modified Normalized Mean Bias (MNMB in %) based on daily averaged profiles over Frankfurt (daytime and nighttime observations) and yearly absolute mean values.

Month	Layer	f38u	f026	TM5	MOCAGE	MOZART	Days with profiles
JAN	SL		-28	-27		-20	29
	BL		-22	-23		-32	
	FT		-08	-07		-16	
FEB	SL		-35	-44		-39	27
	BL		-23	-27		-31	
	FT		-08	-08		-13	
MAR	SL		-20	-34		-29	30
	BL		-18	-31		-31	
	FT		-07	-17		-16	
APR	SL		-07	-26		-22	29
	BL		-10	-32		-25	
	FT		-05	-21		-17	
MAY	SL		-04	-30		-19	31
	BL		-09	-28		-21	
	FT		+03	-18		-15	
	SL	-16	-08	-31	-40	-20	

	BL	-08	-08	-23	-45	-19	
	FT	+04	+05	-11	-25	-09	
JUL	SL	-13	-00	-13	-20	-09	
	BL	-10	-03	-17	-35	-12	30
	FT	-05	+03	-09	-20	-12	
AUG	SL		-01	-01	-10	-08	
	BL		-05	-08	-26	-12	30
	FT		+02	-01	-11	-11	
SEP	SL		-04	-08	+03	-01	
	BL		-03	+02	-05	-07	29
	FT		+01	+05	-03	-06	
OCT	SL		+04	+02		-00	
	BL		-01	+02		-06	
	FT		+03	+07		-06	
NOV	SL		-01	+02		+06	
	BL		-05	-00		-10	20
	FT		-00	+02		-12	
DEC	SL		-20	-33		-33	
	BL		-13	-16		-26	12
	FT		-02	-04		-14	
YEAR	SL		11	21		17	
	BL		10	17		19	
	FT		04	8		12	

9.3 GEMS-GRG reanalysis

In this section we compare how well the IFS/MOZART coupled model with assimilation performs in comparison to the stand-alone models. This allows us to assess how much the better transport brought about by the IFS dynamical model and the data assimilation process improve the simulation of CO. As a sensitivity test of the data assimilation process, we analyze both a coupled IFS/MOZART simulation with full data assimilation (f026) and a control run with no data assimilation (f38u). Because we are most interested in assessing how the assimilation of CO and O3 data affect the simulation, ideally our control simulation would exclude only the assimilation of these variables. However, due to lack of computer resources the control simulation used in this study excludes the assimilation of meteorological variables as well, thus caution must be taken when interpreting the results. A further limitation of the control simulation is that it was only run for the months of June and July so only a summer comparison is possible.

	run	**description**	**available period**
IFS/MOZART coupled model	f026	-with assimilation -injection height at the surface	20030101-20071231
	f3xo	-no assimilation -tracers injected at surface, 6 and 8 km	20040601-20040731
	f51b	-no assimilation -tracers injected at surface, 6 and 8 km -daily emission inventory	20040601-20040731

Table 9.3: Details of the IFS/MOZART coupled runs available for evaluation.

Monthly Modified Normalized Mean Bias (MNMB in %) based on daily averaged profiles (Fig. 9.1-5) show that the largest improvements due to the MOPITT assimilation occurs during the springtime where biases improve by much as 25% with the coupled IFS/MOZART system (f026) with assimilation. For example, CTM biases during the winter and spring months over Frankfurt are generally between -20 and -25% in the free troposphere. However, during March through May the f026 model biases are only between 0 and +/-10%. Similar improvements are also seen in the surface and boundary layers with the f026 model during the spring where CTM biases are around -25 to -50%. The largest improvements made by the IFS/MOZART coupled system also occur during the springtime over the other cities as well.

The smaller biases produced by the f026 simulation are in fact due to the poor performance of the stand-alone CTMs during the winter and spring months rather than an improved performance by the coupled model with assimilation which performs equally as good throughout the year except during the winter months. One explanation for this is that the CTMs may consume CO too fast during the springtime because of too intense activity of photochemistry and this deficiency is compensated for in the IFS coupled model by the MOPITT assimilation. Another point to keep in mind is that during the springtime there is less activity of extratropical lows thus more cloud-free days which increases the efficiency of the MOPITT assimilation. This might contribute to the fact that the f026 simulation performs better during the springtime and rest of the year than during the wintertime.

Because there is no control simulation during the spring months, it is difficult to assess how much of the improvement in the f026 simulation can be attributed to the CO assimilation versus the better transport provided by the IFS coupled dynamical model. During the summer months when the control simulation is available the IFS/MOZART coupled system also performs better than the stand-alone CTMs, although the improvements in biases are not as large as during the springtime. In the free troposphere, biases from the f026 simulation range only between +/- 5% for all airports, while the stand-alone CTMs tend to under-estimate CO in the free troposphere by approximately 15-25%. During July over Frankfurt and Paris, the control IFS coupled simulation without assimilation (f38u) performs slightly better than CTMs but not as good as the f026 simulation with assimilation, except in the upper troposphere where the MOPITT signal is very weak so there is no clear impact of the assimilation. In June, the only improvement due to assimilation is seen at the surface layer. There are also noticeable improvements due to assimilation over Tokyo, while the impact of assimilation over Vienna and Atlanta are less clear.

The fact that in many of the cases the biases from the control f38u simulation are slightly better than the CTMs, but not as good as the f026 simulation with assimilation suggests that the better performance of the f026 simulation is partly due to the better transport brought about by the IFS dynamical model, but also largely due to the assimilation. However it is still difficult to explain how valid this improvement is because there are many influencing factors, such as the sensitivity of the MOPITT sensor, the method to assimilate CO MOPITT tropospheric columns into vertical profiles, and the 4DVAR procedure itself.

Firstly, MOPITT CO data have been generally regarded as not very sensitive to the boundary layer. The MOPITT data products used for assimilation are based on thermal-infrared measurements near 4.7 microns, and its sensitivity to the lower troposphere is debatable (Deeter et al. 2007). Hence, changes in the PBL can not be efficiently constrained by MOPITT data assimilation (This will be even more true when we will use averaging kernels). However, recent studies have shown that MOPITT sensitivity varies considerably depending on the thermal contrast conditions (Deeter et al. (2007); Clerbaux et al. (2008)). These studies show that to certain degree, MOPITT is sensitive to the lower atmosphere in conditions where there is a strong temperature gradient between the surface and lower atmosphere. This may be well the case during summers in the northern hemisphere and it emphasizes the need to evaluate the impact at all seasons, and explore the difference of impact using either day- or night-time data.

Secondly, as MOPITT's averaging kernels are not used in the assimilation process (f026 run), the information on the vertical distribution of CO is lost and the low sensitivity in the lowest troposphere

is not taken into account. This may allow the 4DVAR technique to change the concentration profile predominantly in the PBL where the CO variability is the larger. This emphasizes the need to further explore the impact of the assimilation in case studies which have identical CO tropospheric columns except for either the presence of a mid-tropospheric biomass burning plume which has been transported over a non-polluted maritime PBL and/or either a strongly polluted continental PBL is present below a nonpolluted mid-troposphere.

Thirdly, in the 4DVAR context, a simplified model (but still something "physical") is used to make sure that the profiles are modified in a way that their total column (i.e. vertical integral) better fits the MOPITT total columns. The freedom of the 4DVAR technique to change the concentration profiles depends on the assumed background errors statistics. The higher this error is the bigger are the allowed changes. The background error is often linked to the variability. And because variability is often high in the PBL, the values in this region are more likely to be changed.

In summary, the f026 simulation has significantly lower biases (by up to 25%) in the free troposphere, surface and boundary layers, especially during the springtime. However, since the control simulation f38u without assimilation is only available during the summer months it is difficult to assess how much of the improvement is due to assimilation versus better transport by provided by the coupled dynamical model. In the current setup, a change in the PBL can help to bend the profile towards MOPITT tropospheric columns and at the same time compensate for deficiences in emission inventories. In such cases, the general improvment of bias in the PBL may be the result of the cleverness of the 4DVAR technique. However, there is more than one way for the CO profile to change, especially in presence of mid-tropospheric biomass burning plumes. This is why including more information about the profile (averaging kernel, profile retrievals) would be more beneficial. With averaging kernels, the impact of CO assimilation in the PBL will be weaker, however, the PBL concentrations may still change to a large extent in order to provide a "physical" process to increase values in the free troposphere.

9.4 Case studies and sensitivity experiments

Here, we identify several case studies in which CO plumes originating from the Alaskan wildfires were transported downwind as far as the eastern United States and across the Atlantic Ocean to Europe. We evaluate the ability of the IFS coupled system with assimilation, as well as the stand-alone CTMs to simulate the long-range transport and evolution of CO emitted from the fires.

9.4.1 Long range transport of biomass burning plumes

In this section we examine how well the stand alone CTMs and the IFS coupled system with assimilation (f026) can capture long-range transport of CO plumes originating from biomass burning during the Alaskan wildfires at downwind locations, e.g. Paris and Frankfurt. As in the previous section, we include the coupled IFS/MOZART control simulation with no data assimilation (f38u) in our analysis in order to provide some insight on the sensitivity of the assimilation process. The selection of case studies shown on Figure 9.7 is guided by the presence of CO plumes in the vertical sounding over Europe and by the demonstration (not shown here, see Elguindi et al., in preparation) with backward trajectory simulations of the Flexpart model (Stohl et al., 2005) that the CO plumes originate from the wild fires region over North America. Note that the vertical profiles of tracer mass shown on Fig. 9.7 are discussed in section 9.4.2.

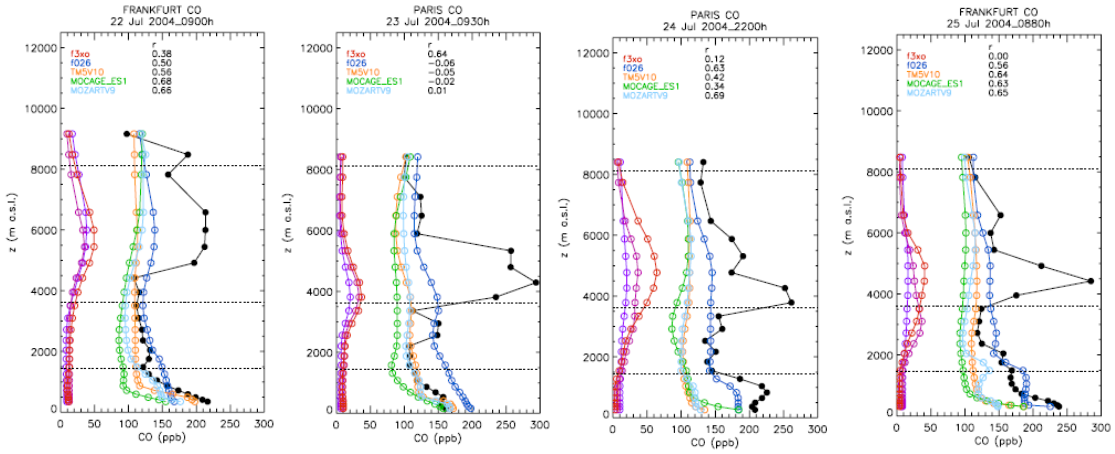


Figure 9.7: Vertical profiles of modelled and observed CO over Frankfurt and Paris in July 2004, and vertical profiles of tracer mass injected at the surface (blue), at 6 and 8 km altitude (purple and red, respectively.

None of the standalone CTMs are able to reproduce the layering of CO in the vertical profiles over Frankfurt and Paris. As stated previously, all CTMs generally under-estimate CO, especially MOCAGE for the examples shown, while TM5 and MOZART perform about the same. Several factors contribute to the CTMs' poor representation of, or lack thereof, the CO plumes. One contributing factor is the 8-daily temporal resolution of the fire emissions data used by the models which is inadequate to capture all of the fires. Another contributing factor might be the parameterization of injection heights above the fires. Often, emissions from these boreal fires are injected high into the troposphere, sometimes even reaching the stratosphere. In the FLEXPART simulations (not shown here), the long-range transport of CO originating from the fires can be quite sensitive the injection height used by the model, thus it is likely that using an injection height parameterization which accounts for the intense convection associated with boreal forest fires would improve the representation of the CO plumes in the CTMs.

Although the concentrations are significantly weaker than those observed in the MOZAIC, the f026 runs (IFS/MOZART coupled system with assimilation) generally perform better than the standalone CTMs and better than the f38u control run (IFS/MOZART coupled system without assimilation). It clearly indicates that the assimilation of MOPITT CO data in the f026 simulation is helping to improve the representation of the CO plumes. However, in the lower troposphere (below 2.5 km) the f026 model sometimes over-estimates CO by about 50 ppb in comparison to the MOZAIC data (as it is the case over Paris on July 23), whereas the biases for the f38u model are substantially smaller. It is possible that this is an effect of the simplified assimilation process in the IFS coupled system in which CO is assimilated evenly throughout all levels in the models. This assimilation process is not necessarily correct because MOPITT sensitivity to CO concentrations in the lower troposphere varies widely (Deeter et al. 2007). A more realistic approach to the assimilation process would be to determine from the analysis of MOPITT avering kernels where and when the measurements offer useful sensitivity to lower tropospheric CO.

Overall, based on these few case studies, the coupled run with assimilation better represents the long range transport of pollution than the coupled IFS/MOZART control simulation without assimilation (f38u) or the stand-alone CTMs. In all of the cases, the assimilation process in the f026 simulation improves the representation of the CO plumes, in large part by compensating for the lack of adequate wildfire emissions data. However, the CO plumes in the f026 simulation are still weak and not always at the correct levels in comparison to the MOZAIC observed profiles.

As further sensitivity tests will show, an improved injection height parameterization scheme in the coupled IFS/MOZART system and the utilisation of a daily emissions inventory could lead to some improvements. Also, the fact that the depth of the CO plumes are not well represented in the

troposphere, and that in some cases the CO appears to be over-compensated for in the PBL (i.e. Case over Paris on July 23) suggests that some improvements could be made in the assimilation process, e.g. using averaging kernels for a better allocation of CO increments across altitude.

9.4.2 Sensitivity experiments

Several studies have shown that emissions from the boreal fires in N. America during the summer of 2004 were emitted as high into the atmosphere as the upper troposphere and lower stratosphere (Jost et al., 2004 ; Damoah et al., 2006). Modelling studies have further shown that long-range transport of the emissions are better simulated when a high injection is used (de Gouw et al., 2006 ; Turquety et al., 2007)). De Gouw et al. (2006) and Cammas et al. (2008) used the atmospheric transport model FLEXPART to describe the transport of the emissions from the 2004 North American fires and found that injecting the emissions to a height of up to 10 km produced the best results in comparison to observations. Using the GEOS-Chem CTM, Turquety et al. (2007) found good results when a portion of the emissions were injected into the upper troposphere.

For the models used in this study, emissions were either injected at the surface or at relatively low heights in the atmosphere. In order to determine how much of an effect the injection height has on the ability of the models to transport CO long distances, we performed an injection height sensitivity test using the IFS model. A tracer mass is added to the physics tendencies at a chosen model level at every model time step (every 1800 seconds). For this sensitivity test, we inject different tracers at several different model levels, namely the surface and at approximately 1, 2, 3, 4, 6 and 8 km. There is little difference in the long-range transport of the tracers emitted between the surface and 4 km, thus we only present the results for the tracers emitted at the surface and at 6 and 8 km.

Profiles of the tracers over the Frankfurt and Paris airports on various days during July 2004 are shown in Figure 9.7. Whatever is the injection height at the source, the tracer concentration maximises nearby the altitude of the CO plume at the receptor site. It seems to show that the cloud convection and the biomass fire emissions occur at the same time in the same grid mesh of the model, and that convection is contributing to the vertical transport. Otherwise the differential advection with height will make larger differences in long-range transport between the tracer emitted at the surface and the ones emitted at higher altitudes.

There is considerable uncertainty associated with the injection height of emissions from boreal fires, as the heights vary with the intensity of the fire and the present synoptic conditions. Given the temporal and spatial variability of the injection height, a parameterization would be most accurate.

Sensitivity experiments have also been performed on the temporal resolution of emissions data. Runs with the tracer mass were repeated with injection height at 8 km altitude and using the 8-day inventory emission and a daily inventory emission made by Turquéty et al., 2007 (Figure 9.8). In the example shown, the results using the daily emission inventory considerably improve the representation of the long-range transport of biomass fire plumes from Alaska to over Europe.

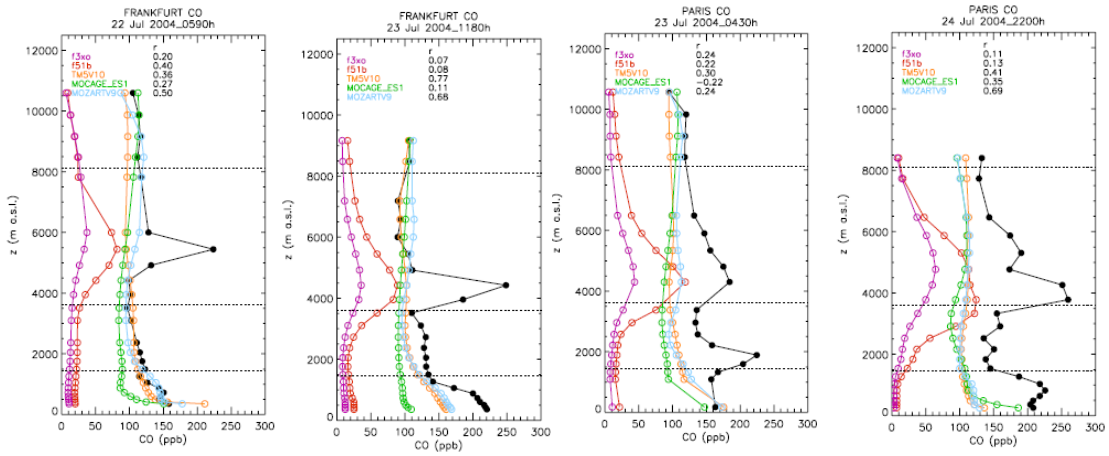


Figure 9.8: Vertical profiles of modelled and observed CO over Frankfurt and Paris in July 2004, and vertical profiles of tracer mass injected at 8 km altitude using the 8-day inventory emission (purple line) and a daily inventory emission made by Turquéty et al., 2007 (red line).

9.5 Summary

Globally, the model biases are between +/- 25% in the free and upper troposphere and +/- 50% in the surface and boundary layers. By in large, the models have a tendency to underestimate CO everywhere. Nighttime biases are generally smaller than daytime biases indicating that the observations represent the minimum concentrations. Overall, TM5 and MOZART perform equally while MOCAGE performs slightly worse, although it should be kept in mind that only fours months of the MOCAGE simulation was available.

The IFS/MOZART coupled model with assimilation (f026) is cleary better able to simulate the global scale distribution of CO than the stand-alone models. This is most evident during the springtime when the CTMs perform poorly, perhaps because it is a transition season. This improvement is partly due to the better transport brought about by the IFS dynamical model, but also largely due to the assimilation.

In evaluating how well the models are able to simulate the long-range transport of CO plumes originating from biomass burning during the Alaskan and Canadian wildfires, we found that the stand-alone models were either unable to simulate the CO plumes or did a very a poor job. While the IFS/MOZART coupled model with assimilation did a better job in simulating the CO plumes, the concentrations were significantly weaker than those observed in the MOZAIC data. Sensitivity tests indicate that the simulation of the CO plumes could be much improved by the use of a parameterized injection height and daily emissions inventory.

9.6 References

1. Clerbaux, C., D. Edwards, M. Deeter and L. Emmons, 2008: Carbon monoxide pollution from cities and urban areas observed by the TERRA/MOPITT mission. Geophysical Research Letters, , 35, L03817, doi:10.1029/2007GL032300.
2. Cammas, J.-P., Brioude J., Chaboureau J.-P., Duron J., Mari C., Mascart P., Nédélec P., Smit H., Pätz H.-W., Volz-Thomas A., Stohl A., and Fromm M., Injection in the lower stratosphere of biomass fire emissions followed by long-range transport: a MOZAIC case study. *Atmos. Chem. Phys., 2009, Vol.9, pp. 5829-5846, SRef-ID: 1680-7324/acp/2009-9-5829.*

3. Deeter, M., D. Edwards and J. Drummond, 2007: Sensitivity of MOPITT observations to carbon monoxide in the lower troposphere. Journal of Geophysical Research, 112, D24306, doi:10.1029/2007JD008929.
4. Damoah, R., N. Spichtinger, R. Servranckx, M. Fromm, E. Eloranta, I. Razenkov, P. James, M. Shulski, C. Forster and A. Stohl, 2006. A case study of pyroconvection using transport model and remote sensing data. Atmospheric Chemistry and Physics, 6, 173- 185.
5. de Gouw, J., C. Warneke, A. Stohl, A. Wollny, C. Brock, O. Cooper, J. Holloway, M. Trainer, F. Fehsenfeld, E. Atlas, S. Donnelly, V. Stroud and A. Lueb, 2006: Volatile organic compounds of merged and aged forest fire plumes from Alaskan and western Canada. Journal of Geophysical Research, 111, D10303, doi:10.1029/2005JD006175.
6. Jost, H.-J., K. Drdla, A. Stohl, L. Pfister, M. Loewenstein, J. Lopez, P. Hudson, D. Murphy, D. Cziczo, M. Fromm, T. Bui, J. Dean-Day; C. Gerbig, M. Mahoney, E. Richard, N. Spichtinger, J. Pittmann, E. Weinstock, J. Wilson and I. Xueref, 2004: In-situ observations of mid-latitude forest fire plumes deep in the stratosphere. Geophysical Research Letters, 31, L11101, doi:10.1029/2003GL019253.
7. Kaiser, J.W., M. Suttie, J. Flemming, J.-J. Morcrette, O. Boucher and M.G. Schultz, 2008: Global real-time fire emission estimates based on space-borne fire radiative power observations. The Proceedings of the International Radiation Symposium.
8. Stohl, A., C. Forster, A. Frank, P. Seibert, and G. Wotawa, 2005: Technical Note : The Lagrangian particle dispersion model FLEXPART version 6.2. Atmos. Chem. Phys. 5, 2461-2474.
9. Turquety, S., J. Logan, D. Jacob, R. Hudman, F.Y. Leung, C. Heald, R. Yantosca, S. Wu, L. Emmons, D. Edwards and G. Sachse, 2007. Inventory of boreal fire emissions for North America in 2004: Importance of peat burning and pyroconvection injection. Journal of Geophysical Research, 112, 109, D21301, doi:10.1029/2004JD004821.

10. Evaluation of boreal fires impact with satellite data

Contributors: Khokhar Fahim, Law Kathy, Granier Claire (SA-UPMC, Paris, France), and the modeller teams

GRG Task 4.3.2 aims to analyze the GRG chemistry transport models (GRG-CTMs) for regions affected by strong boreal forest fire emissions. The main focus is on the evaluation of specific pollution events in terms of location, trans-boundary transport, chemical evolution during the transport and the chemical composition of the boundary layer and free troposphere during these fire events. In this section, we analyze the models for regions affected by strong boreal forest fire emissions by presenting two case studies over the Siberian fire region during summer 2003 and the Alaska fire region during summer 2004.

10.1 Datasets and methodologies

10.1.1 Datasets

MOPITT (Measurements Of Pollution In The Troposphere)

MOPITT is an instrument onboard NASA's EOS Terra spacecraft launched in December 1999. It is primarily designed for measuring the global distributions of carbon monoxide (CO) and methane (CH_4) in the troposphere. MOPITT operates by sensing infrared radiation from either the thermal emission/absorption at 4.7 μm for CO profiles, or reflected sunlight at about 2.2-2.4 μm for CO and CH_4 column measurements in daylight. The use of solar channels enhances the instrument sensitivity to the atmospheric boundary layer. The satellite is deployed in a polar synchronous orbit with 10:30 am local equator crossover time. Spatial resolution is 22 km x 22 km, and cross-track scanning achieves approximate global coverage in 3 days. MOPITT retrievals are reported on 7 vertical levels (surface, 850, 700, 500, 350, 250, and 150 hPa), and as a total column, for all cloud-free scenes. The measurement technique relies on thermal contrast between the surface and the atmosphere, leading to a retrieval dependence on surface temperature, and little sensitivity to CO in the boundary layer.

The MOPITT retrieval uses a fixed “global” a priori profile. This a priori profile is generated from a master set of 525 in-situ profiles measured from aircraft during eight atmospheric chemistry field campaigns and at two fixed sites.

In order to compare model output and satellite retrievals, all data are re-grided to the same spatial resolution. In our case, MOZART, MOCAGE and MOPITT data (with 60 vertical levels) are interpolated to TM5 spatial resolution (2° latitude x 3° longitude). MOPITT averaging kernels are calculated for a given time frame and over area of interest. After this step, the model output fields are convoluted with average kernels of MOPITT and results (new model CO profiles resolved at 7

MOPITT levels) are compared with MOPITT CO profiles. Results are presented in the following sections.

10.1.2 Methodologies

GRG - CTMs

The following GRG-CTMs experiments are analysed in this study.

MOZART Experiments

:: **V10:** is the latest (December 2008) stand-alone reference run for the year 2003. GRG anthropogenic emission inventories and 8 day GFEDv2 for fire emissions inventories were used and confined to the lower MOZART level. The new JPL-06 evaluation of the reaction CO + OH leads to a reaction rate which is about 10-20% lower than before in the troposphere, is introduced in V10 simulations.

:: **F026:** 2nd reanalysis run is with fixes (i-IFS convection and diffusion for CO, CTM convection and diffusion for GO_3, NOx, coupled to MOZART. ii-two-way coupling for CO and GO_3. iii- NOx and HCHO initialized from CTM in every forecast. iv- Bug fix for CTM fluxes) from **eyih** (1st reanalysis run – with CO (MOPITT) and GEMS ozone assimilation (SCIAMACHY, MIPAS, GOME, SBU). This is first GEMS-GRG reanalysis together with GHG and AER. This run is available for 2003 -2007. Monthly GFEDv2 for fire emissions inventories are incorporated in MOZART- F026 simulations and confined to the lower MOZART level.

MOCAGE Experiments

:: **60LEV02 >> V2:** Formally, "60LEV02" is the name for stand-alone reference run for the year 2003. However, for simplicity in this script/analysis "60LEV02" is changed with **V2**. It is available since June 2008. Monthly GEMS RETRO wildfire emissions for 2003 are used and implemented as a vertical profile (fire CO percentage at certain levels- see Table 1).

:: **60LEV02 >> V2:** is the latest stand-alone reference run for the year 2004 and is available for only (June –September 2004). It is available since June 2008. Vertical profiles of 8 day GFEDv2 for CO fire emissions inventories are used for 2004 (see table 1).

TM5 Experiments

:: **V7:** is the latest stand-alone reference run for the year 2003 and available since May 2008. GRG anthropogenic emission inventories and monthly GFEDv2 for fire emissions inventories are used and are injected in different amount to certain height levels as given by Table 1.

:: **V10:** is the latest stand-alone reference run for the year 2004. GRG anthropogenic emission inventories and 8 day GFEDv2 for CO fir emissions inventories are used and are injected in different amount to certain height levels as given by Table 1.

GRG-CTMs Performance

The method for the evaluation of the GRG-CTMs (agreed upon GRG-subgroup meeting and afterwards on-line discussions) is based on scoring index. The score is based on median values, is defined as:

$$S = 1 - \text{avg}\left(\text{abs}\left(\text{median\_bias}^* / 2(\text{median\_obs}^*)\right)\right) \qquad \text{Eq (1)}$$

where median_bias = median(Mi – Oi)/ Ô

and Ô is evaluated as:

Ô = max((median(Oi)), error_scale) where Oi and Mi stands for observations and model output per grid cell (i) , respectively. Additionally, error_scale is the value of median_obs (i.e. local median of observations). It is limited to a minimum value (i.e. 0.5 molec/cm²) because the individual retrievals can be quite small / negative (e.g. in case of SCIAMACHY NO_2 –see Chapter 2). This is not the case for MOPITT CO observations and it is set to 0 in this case. For further details about score calculations see *Vincent et al.,* scoring document available at http://gems.ecmwf.int/do/get/Documents). All GRG-CTMs are evaluated by calculating monthly mean scoring index. The monthly mean score is obtained from the medians of daily CO total columns from the GRG-CTMs simulations and MOPITT observations over respective regions (see Equation 1). The performance of each GRG-CTMs during both case studies is discussed on monthly basis in the following sub-sections.

Table1: Fraction of CO fire emissions at injected heights by GRG-CTMs

**: this information about injected height are only for Alaska fire region for both year 2003 and 2004*

for Siberian fire region it might vary in numbers but follow the same procedure for vegetation type

Injection height (m)	MOZART		TM5		MOCAGE* 2003 and 2004	
	2003	2004	2003	2004	forest	shrubs
0-100	100%	100%	20%	10%	3%	40%
100-500	-	-	20%	10%	3%	40%
500-1000	-	-	20%	10%		
1000-2000	-	-	40%	10%	6%	20%
2000-3000	-	-	-	20%	13%	-
3000-4000	-	-	-	40%	16%	-
4000-5000	-	-	-	-	19%	-
5000-6000	-	-	-	-	19%	-
6000-7000	-	-	-	-	13%	-
7000-8000	-	-	-	-	5%	-
8000-9000	-	-	-	-	2%	-
9000-10000	-	-	-	-	1%	-
10000-11000	-	-	-	-	-	-
Resolution	1.9°x 1.9°		3° x 2°		2° x 2°	

10.2 Siberia Fire Region

Boreal forest and agricultural fires (see Figure 1) contribute significantly to the global budgets of a number of atmospheric trace gases and particles. Year 2003 experienced a very large number of fire events, particularly in the boreal region of Russia. A region between 40°-65°N and 100°-140° E referred to as **Siberian fire Region** is selected for the GEMS analysis. Off-line simulations performed with the CTMs (MOZART3, MOCAGE and TM5) in the framework of GRG sub-project of GEMS, are compared with ground-based data and satellite observations. MOPITT level 2 data version MOP03M.L3V91.01 (downloadable form MOPITT site ftp://l4ftl01.larc.nasa.gov/MOPITT/ MOP03M.003) was used in this analysis. CTMs simulation results used are the latest available off-line simulations for year 2003 (e.g. MOZART-V10, TM5-V7, MOCAGE-V2)

Advance Along Track Scanning Radiometer (AATSR) data for fire counts is shown in Figure 1 with very large number of fire events during the May-July 2003 period over boreal forest area within selected region (magenta box).

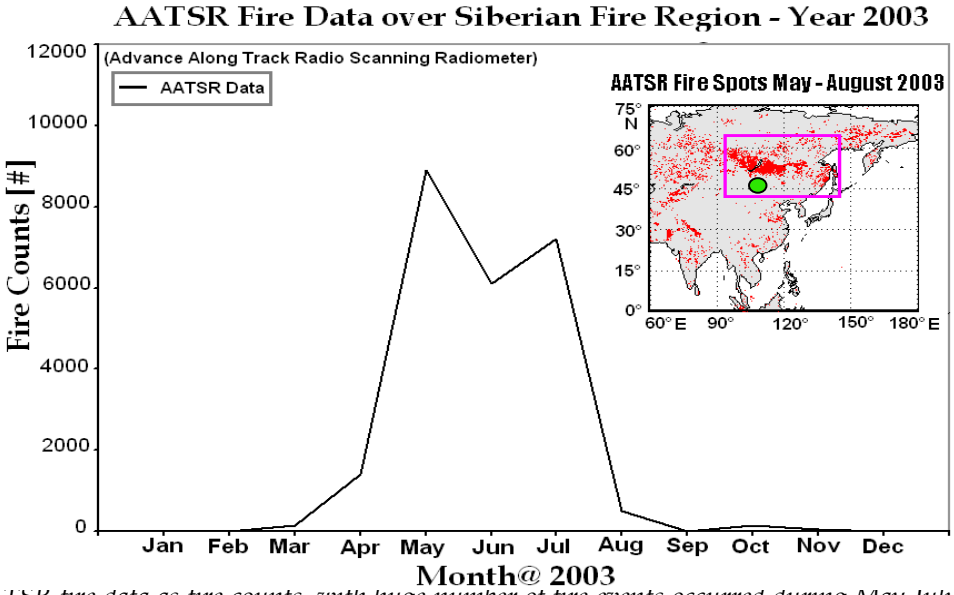


Figure 1: *AATSR fire data as fire counts, with huge number of fire events occurred during May-July 2003 over boreal forest area. Siberian fire region - Magenta rectangle - [40°-65° N, 100°-140° E] with extensive fire density over Siberia is selected for further analysis (Thumbnail Fig. 1). Green dot in thumbnail figure indicates the location of UUM NOAA/GMD surface station (formerly known as CMDL - Climate Monitoring and Diagnostics Laboratory)*

As a first step, we compared CO concentrations calculated by the CTMs with ground based stations close to or downwind of the area of interest. The green dot (thumbnail figure) shows the location of the UUM (Ulaan Uul Mongolia- 44.27°N and 111.6° E) NOAA/GMD (formerly known as NOAA/CMDL) surface station in Mongolia situated within Siberian fire region. A comparison between the surface CO measured by the NOAA/GMD (www.esrl.noaa.gov/gmd/) network at the UUM Mongolia (green dot – Fig. 1) and concentrations calculated by all of three GRG-CTMs for that particular site are shown in Figure 2.

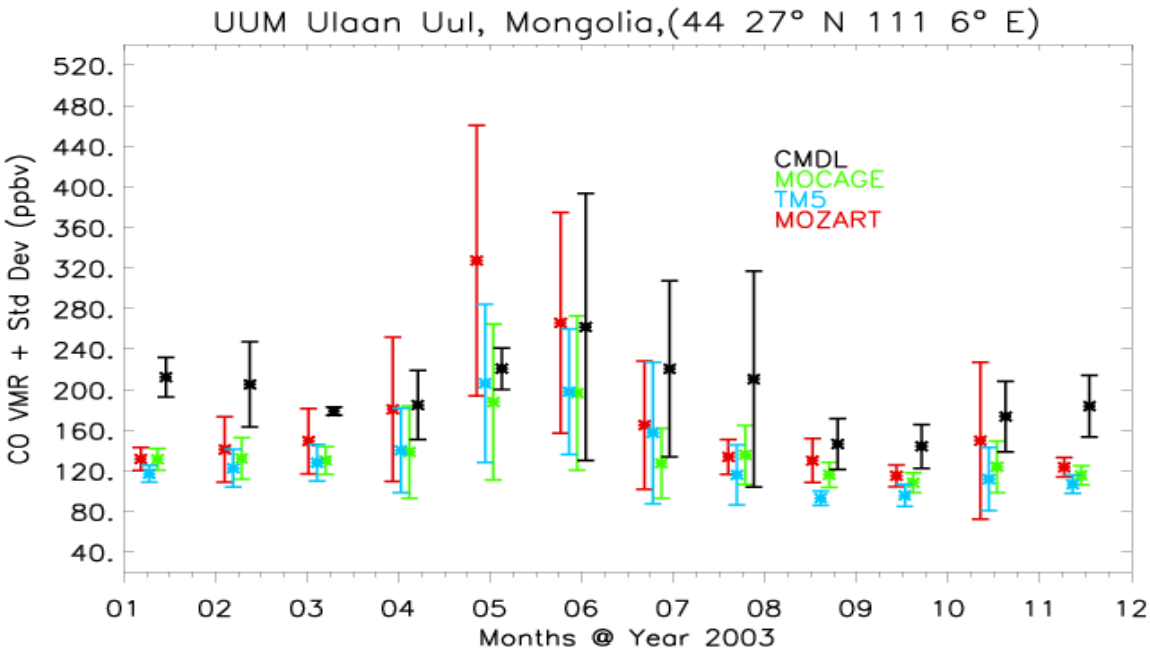


Figure 2: *Comparison between the surface CO concentrations (CMDL - event data) by the NOAA/GMD network at the UUM Mongolia (green dot – Fig. 1) with concentrations calculated by all of three CTM [MOZART-V10 (red), MOCAGE-V2 (green) and TM5-V7 (cyan)]. Vertical bars give the temporal variations of CO concentration over a month period.*

CMDL observations and CTM simulations results show a substantial increase in CO concentrations during the fire season with maximum values during May – Aug. 2003 (Fig. 2). Data shown in Figure 2 are monthly means of CO concentrations from all GRG-CTMs simulations and NOAA/GMD monitoring station (CMDL CO - event data: flasks containing air samples from fixed stations are collected four times in a month and data is referred as event data, for details see chapter 5). Vertical error bars account for the standard deviation calculated as the temporal variations in CO concentrations over a one-month period. The three GRG-CTMs underestimate CO concentrations during the whole year except for months of June and July 2003. Furthermore, during the month of June MOZART-V10 shows large temporal variability and overestimated CO while TM5-V7 and MOCAGE-V2 are in fair agreement with CO observation data from NOAA/GMD station UUM. During the month of June, MOZART-V10 outputs are in a good agreement while TM5-V7 and MOCAGE-V2 underestimate CO concentrations over UUM surface station.

Even though GRG-CTMs underestimate CO concentrations when compared with observations, they show similar seasonality in CO concentrations when compared to CDML observations over the whole year. The inter-model variations may be due to various factors such as: different reaction rates [$CO+OH+M \rightarrow HCHO+M$ (R1) and $CO+OH+M \rightarrow CO_2+H+M$ (R2)], different fire emissions inventories and injection heights (see Table 1), anthropogenic emissions and different model grids used by GRG-CTMs.

As a second step, comparison of CO vmr (volume mixing ratios in ppbv) over the Siberian fire region from MOPITT observations (black), MOZART-V10 stand-alone run for year 2003 in red,

TM5-V7 stand-alone for 2003 in cyan, MOCAGE-V2 stand-alone run for 2003 in green and of MOZART-F026 2nd reanalysis run for 2003 in violet are presented in Figure 3.

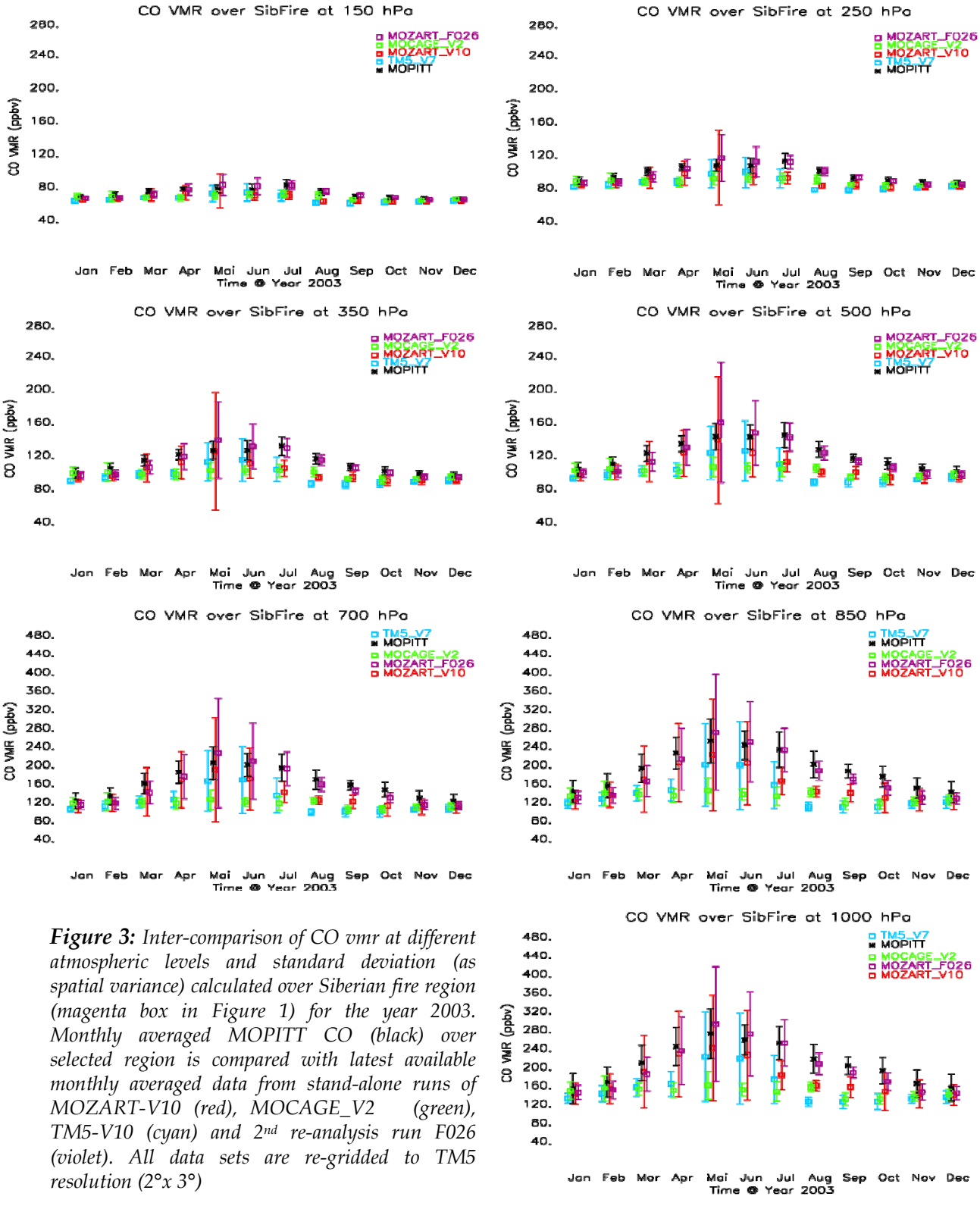


Figure 3: *Inter-comparison of CO vmr at different atmospheric levels and standard deviation (as spatial variance) calculated over Siberian fire region (magenta box in Figure 1) for the year 2003. Monthly averaged MOPITT CO (black) over selected region is compared with latest available monthly averaged data from stand-alone runs of MOZART-V10 (red), MOCAGE_V2 (green), TM5-V10 (cyan) and 2nd re-analysis run F026 (violet). All data sets are re-gridded to TM5 resolution (2°x 3°)*

Monthly mean of CO concentrations over Siberian fire region at 7 pressure levels 1000, 850, 700, 500, 250 and 150 hPa from MOPITT are compared with CTM CO concentrations after applying MOPITT averaging kernels. Seasonality (larger CO peaks during summer months – see Figure 3) in the observed MOPITT CO concentrations is subjected to emissions from agricultural and forest fires in Siberian fire region, and is consistent to fire data shown in Figure 1. Additionally, CO vertical profiles averaged over Siberian fire region for May-Aug. 2003 are presented in Fig. 4. Standard deviations (as spatial variation) are plotted as vertical and horizontal bars in the Figure 3 and 4 respectively. At almost all atmospheric levels MOZART- F026 is very well correlated with MOPITT CO data and shows a slight overestimation during May –June 2003.

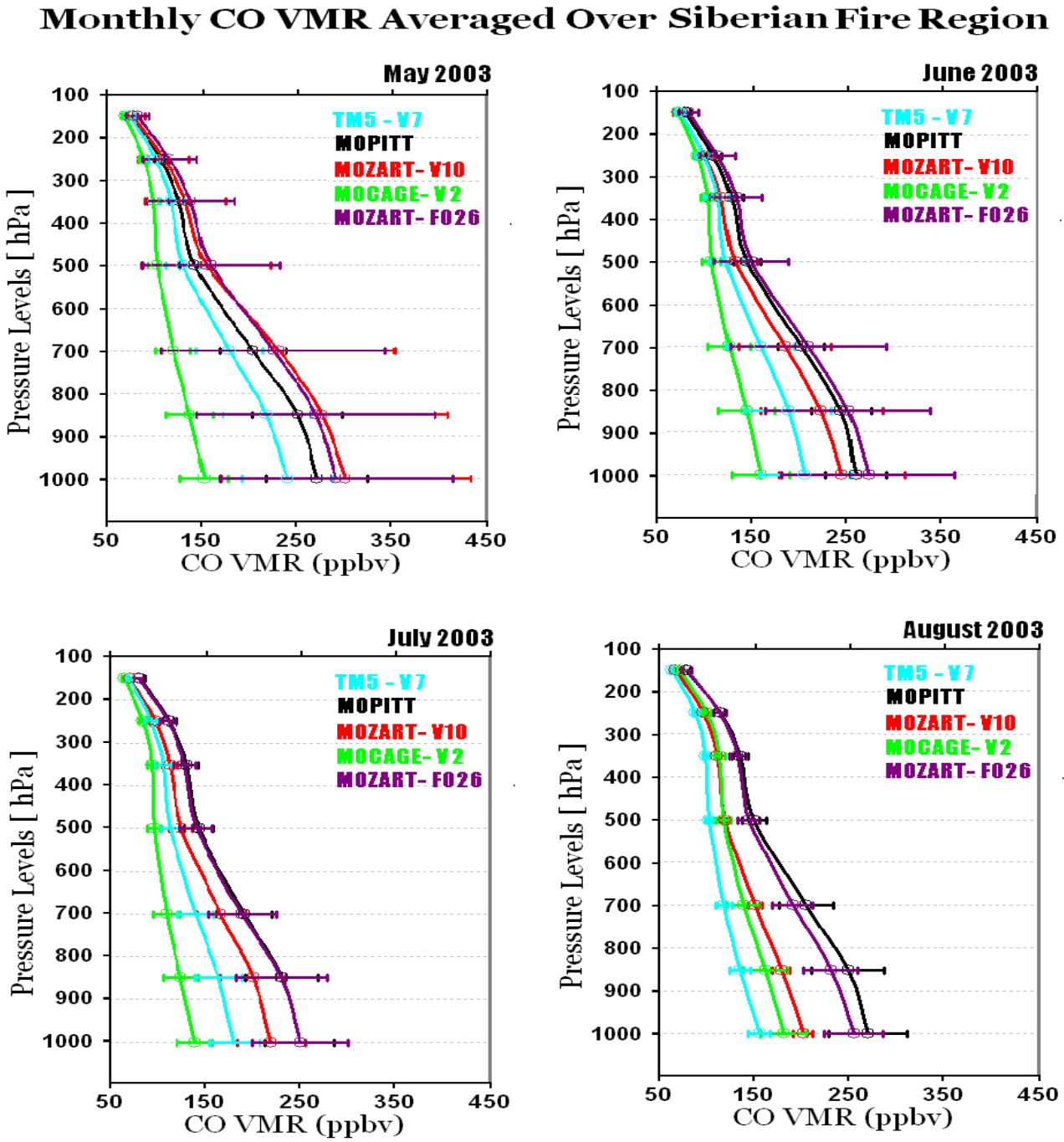


Figure 4: *CO Vertical profiles averaged over Siberian fire region (magenta box, Fig.1) from May to August 2003. GRG-CTM data sets [MOZART-V10 (red), MOCAGE (green), TM5 (cyan) and MOZART-F026*

(violet)] are interpolated to MOPITT vertical levels after applying averaging kernels. All data sets are re-gridded to TM5 resolution (2°x 3°). Horizontal bars give spatial variance of CO profile over Siberian fire region.

Maximum overestimation in CO concentration (20 ppbv during May and 8 ppbv during June) in MOZART-F026 data is found at 1000 hPa level. MOZART-V10 overestimates CO concentrations during May and underestimates CO concentration in June – August. The fair agreement with MOZART-F026 is mainly because MOZART-F026 simulations used assimilated MOPITT CO. Even so, MOZART-V10 is also quite good.

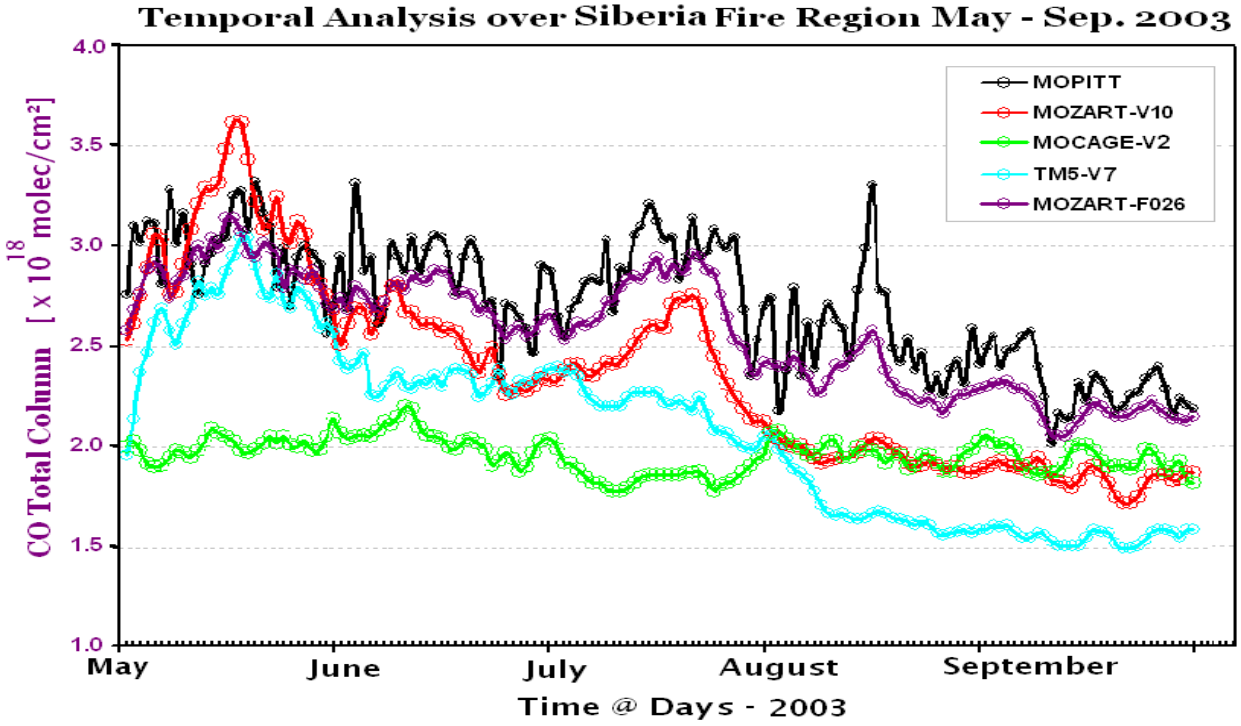


Figure 5: *Temporal analysis of CO total column amounts averaged over Siberian fire region (magenta box, Fig.1) on daily basis. MOPITT CO total columns (Black) are compared with the GRG-CTMs total columns [MOZART-V10 (red), MOCAGE-V2 (green), TM5-V7 (cyan) and [MOZART-F026 (violet)]. All data sets are re-gridded to TM5 resolution (2°x 3°)*

MOCAGE-V2 data is relatively poorly correlated and shows underestimated CO concentrations particularly during spring to autumn months. TM-V7 also underestimates CO concentrations throughout the 2003, although it does capture increases in CO concentrations due to summer fire events.

Similar behavior of all GRG-CTM is reflected in a temporal analysis of CO total column amounts over Siberian fire region compared with CO total columns from MOPITT observations. Results presented in Figure 5, show a comparison of CO total columns from all GRG- CTMs and MOPITT observations on a daily basis from May to September 2003. CO total column from MOZART-V10 are overestimated during May and underestimated during June-September 2003. TM5-V7 and MOCAGE-V2 underestimate CO total columns during the whole period while, MOZART-F026 is in a good agreement with MOPITT CO total columns during May – June and shows a slight underestimation during August and September. This good agreement is mainly because MOPITT

CO is assimilated in MOZART-F026 simulations. In general, the temporal evolution of CO total column from MOZART-F026 and V10 are in a fair agreement with both MOPITT CO and AATSR fire data. The inter-model differences especially during fire episodes are mainly due to different injection height of fire CO in CTMs simulations (see Table 1). Particularly, in MOZART-V10 and f026 all fire emissions are injected at lowest model levels as compared to TM5-V10 and MOCAGE-V2 simulations which used vertical profile of fire CO injected in different amounts at certain altitudes. Furthermore, in case of MOCAGE-V2 simulations, it further depends on the vegetation types (forest and shrubs) within the selected region. The fire CO injection profile of MOCAGE-V2 given in Table 1 is only representative for Alaska fire region and it may slightly differ over Siberian fire region depending on index of vegetation type (for further details see *Rast et al.,* submitted to JGR Nov. 2008).

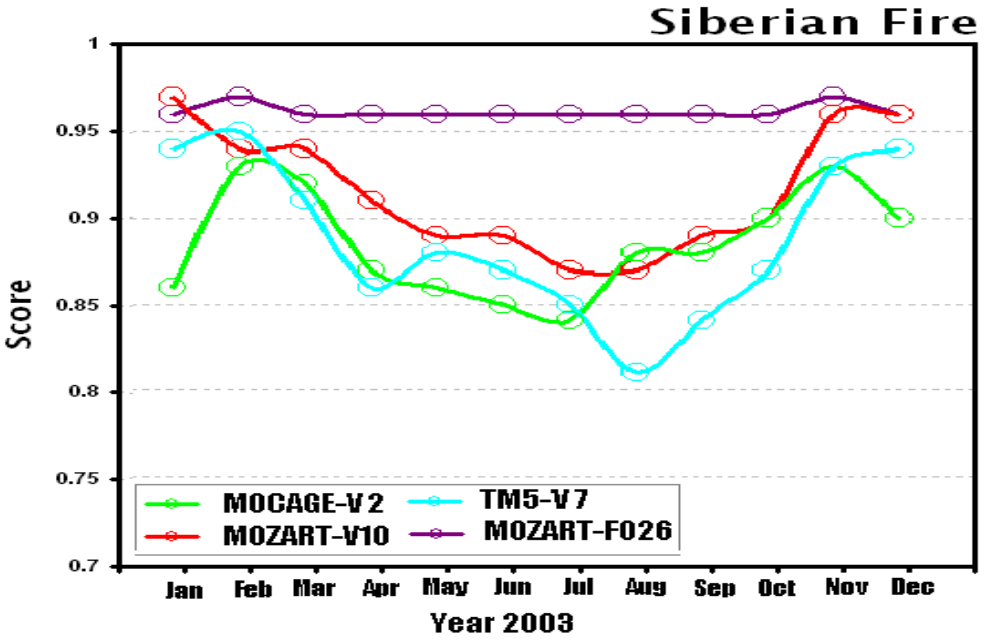


***Figure 6**: Monthly scores, averaged over Siberian fire region, for GRG-CTMs [[MOZART-V10 (red), MOCAGE-V2 (green), TM5-V7 (cyan) and [MOZART-F026 (violet)] based on daily CO total column during the year 2003.*

The performance of each model based on monthly averaged scoring index over Siberian fire region for the entire year 2003 is presented in Figure 6. Monthly scoring index is calculated from difference of medians of CO total columns from all GRG-CTMs simulations and MOPITT observations on a daily basis (as indicated in Eq.1). The 2nd reanalysis run F026 of MOZART (with assimilated MOPITT CO) performed best of all, showing a nearly consistent performance during the year 2003. MOCAGE-V2 exhibited least performance, particularly during the summer months of year 2003. TM5-V7 performed better than MOCAGE–V2 except for autumn months of year 2003. In general, all GRG-CTMs performances are above 80% for temporal analysis over Siberian fire region during the whole year 2003.

10.3 Alaska Fire Region

During summer 2004, large number of fire events took place in Alaska and Canada (termed as Alaska Fire region- magenta box [55°-70° N, 105°-165°W] - Thumbnail Fig. 7). The major goals of this study are to quantitatively evaluate the capabilities of the GRG–CTMs involved in the TF-HTAP (Task Force on Hemispheric Transport of Air Pollution - for details see http://htap.icg.fz-juelich.de/data) and to compare with results from MOPITT CO observations in 2004. Figure 7 shows the location and number of active fires as given by the AATSR (Advanced Along Track Scanning Radiometer) instrument: a very large number of daily fire events were detected over the boreal forest area during the June - September 2004 period in Canada and Alaska, USA. Times series of daily fire events within the selected region (magenta box in Figure 1 (Thumbnail)) is shown. Daily fire data show a large number of fire events starting at the end of June and continuing into July 2004. There was a break in fire activities during the first fifteen days of August and fire episodes returned until the first week of September 2004.

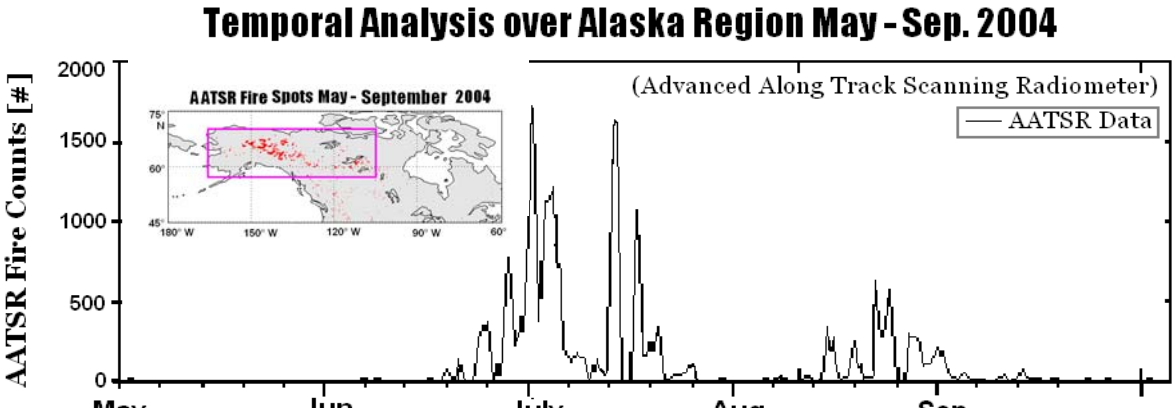


Figure 7: *AATSR data for daily fire events during May - Sep. 2004 over Alaska fire region. Thumbnail picture shows the map of fire hot spots during the given time period and Magenta box is the region termed as Alaska fire region. Time series shows maximum number of fire events took place during July 2004*

Intercomparison of CO concentrations over the Alaska fire region from MOPITT observations (black), MOZART-V10 stand-alone run for year 2004 in red, TM5-V10 stand-alone for 2004 in cyan, MOCAGE-V2 stand-alone run for 2004 in green and of MOZART-F026 2nd reanalysis run for 2004 in violet are presented in Figure 8. The concentrations of fire CO (ppbv) from off-line simulations of GRG-CTMs are compared at 7 different atmospheric levels after application of MOPITT average kernels as described in the previous section.

All GRG-CTMs are consistent with MOPITT observations with respect to temporal patterns (seasonality). However, MOZART-F026 shows slightly overestimated CO concentrations at all levels and TM5-V10 is in good agreement with MOPITT CO observations at upper atmospheric

levels (above 500hPa). The model results exhibit underestimated CO concentrations at lower atmospheric levels (below 500 hPa). CO concentrations from TM5-V10 and MOZART-V10 are less underestimated when compared to MOCAGE CO data. The vertical bars give the spatial variance of the CO concentrations within Alaska fire region over the time period of one month. Additionally, CO vertical profiles averaged over the Alaska fire region for June - September 2004 are presented in Figure 9. Vertical transport of fire CO is high during July 2004 in MOPITT data. The comparison shows that MOZART-F026 overestimates during the month of June and July, while fair agreement with MOPITT CO profiles is found during August and September 2004 (with slightly lower CO at 1000 hPa). MOZART-F026 CO concentrations are slightly overestimated with a maximum of 12 ppbv at 350 hPa level during June and with 9 ppbv at 850 hPa in July 2004.

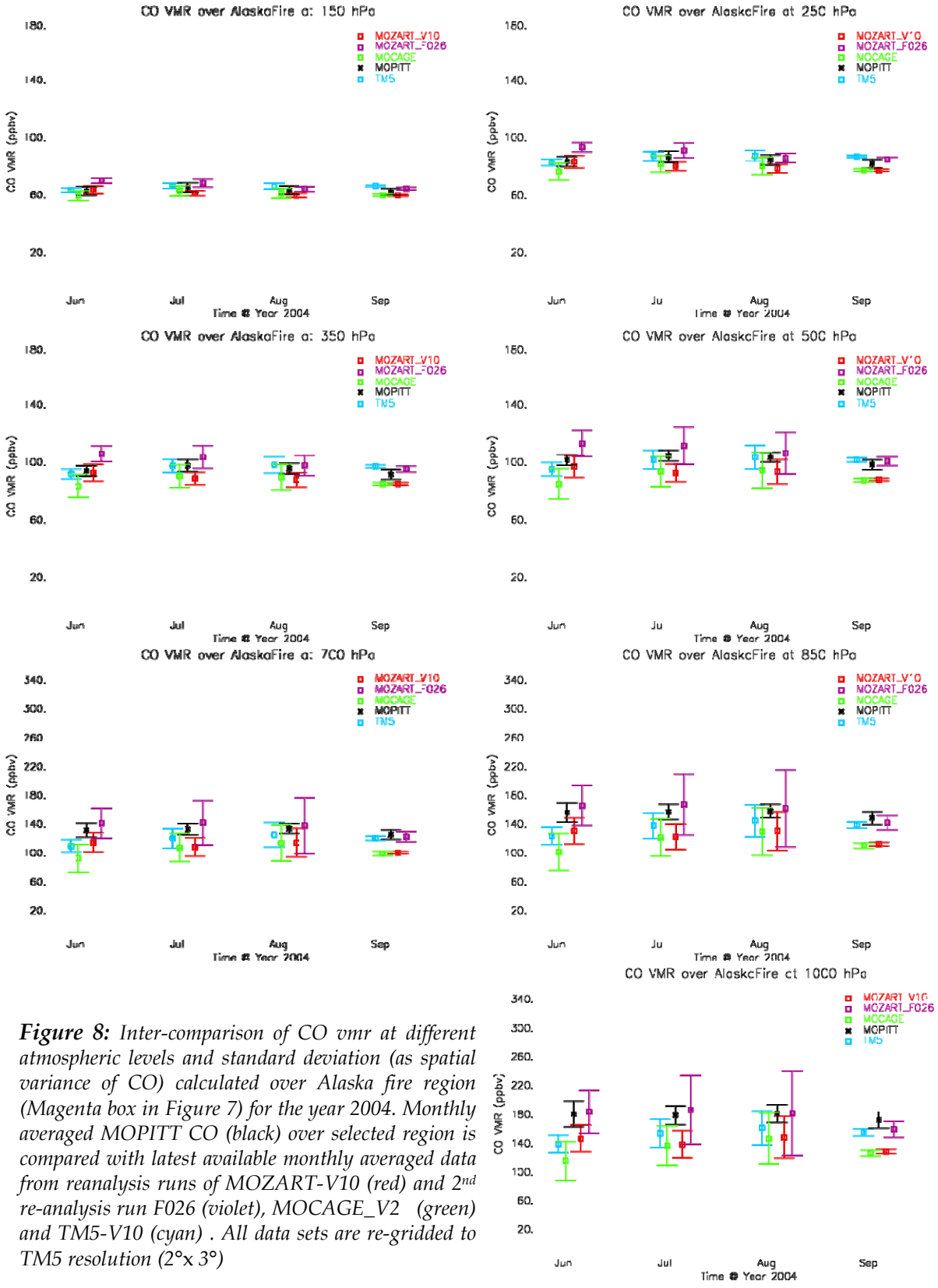


Figure 8: *Inter-comparison of CO vmr at different atmospheric levels and standard deviation (as spatial variance of CO) calculated over Alaska fire region (Magenta box in Figure 7) for the year 2004. Monthly averaged MOPITT CO (black) over selected region is compared with latest available monthly averaged data from reanalysis runs of MOZART-V10 (red) and 2nd re-analysis run F026 (violet), MOCAGE_V2 (green) and TM5-V10 (cyan) . All data sets are re-gridded to TM5 resolution (2°x 3°)*

MOCAGE-V2 CO concentrations are underestimated at all levels and during the whole period (June- August), while a good agreement for June 2004 is obtained at upper atmospheric levels

(>500 hPa). The possible explanation for this inter-model difference could be due to different injection heights of fire emissions (see Table 1). Fire CO injection heights used in MOCAGE depend on the vegetation type and the region of emissions. For the northern part of America, the injection height for forests fires and shrub land fires are given in Table 1. The profile used is a mix of these two profiles, depending on the respective fraction of each vegetation type within selected region (for further details see *Rast et al.,* submitted to JGR, Nov 2008).

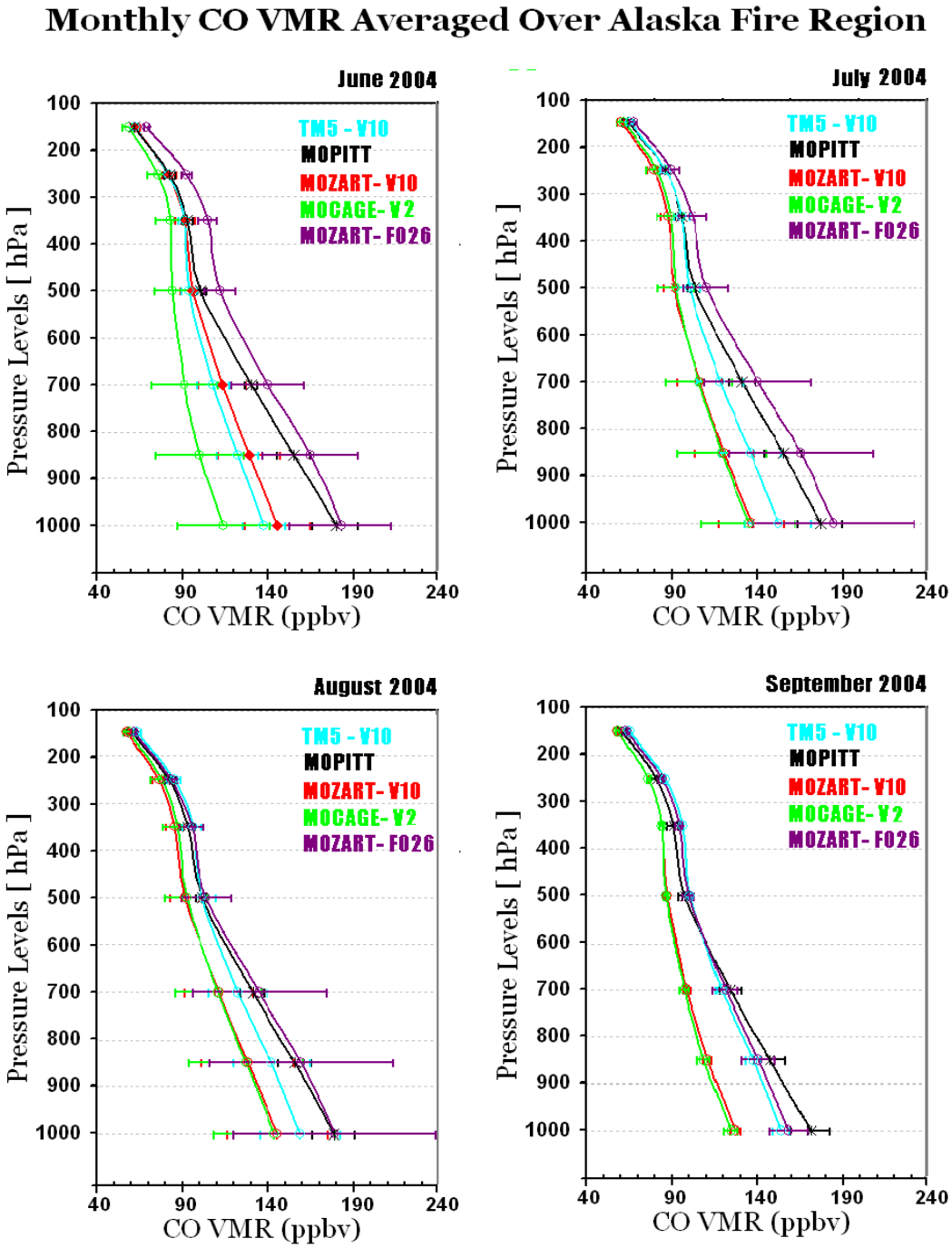


Figure 9: *Monthly mean CO vertical profiles averaged over Alaska fire region for MOPITT (black), MOZART –V10 (red), MOCAGE–V2 (green), MOZART-F026 (violet) and TM5 –V10 (cyan) after application of MOPITT average kernel for June – September 2004. Horizontal bars give the spatial variance of CO profile over the Alaska fire region.*

MOPITT observations show large CO concentrations in the lower tropospheric levels during the fire season. The maximum CO concentrations at 1000 hPa during June and August 2004 (180 ppbv) can be compared to values of about 170 ppbv in September 2004 with relatively less fire events.

MOZART-V10 fire CO emissions were injected only in the lowest model level. MOZART-V10 CO profile for June 2004 is in a good agreement with MOPITT CO profiles at upper levels. However, MOZART-V10 underestimates CO during the whole period at lower levels. TM5-V10 simulations used different injection heights for fire CO emissions compared to MOCAGE-V2 and MOZART-V10 simulations (see Table 1). TM5-V10 results show a good agreement with MOPITT CO profile at upper levels for all months.

Furthermore, CO column amounts from MOPITT observation and all GRG-CTMs simulations results for the time period of June to September 2004 are considered and averaged over the Alaska fire region on a daily basis. An intercomparison of CO total column is presented in Figure 10. Maxima in daily MOPITT CO total column during July- August 2004 show the same temporal evolution with AATSR fire data. In contrast, almost all GRG-CTMs show disagreement in CO column amounts over Alaska fire region. MOZART-F026 slightly overestimates CO total column during June–July, with fair agreement during August and slight underestimation during September 2004. TM5-V10, MOZART-V10 and MOCAGE-V2 underestimate CO total column during the whole period except for occasional event.

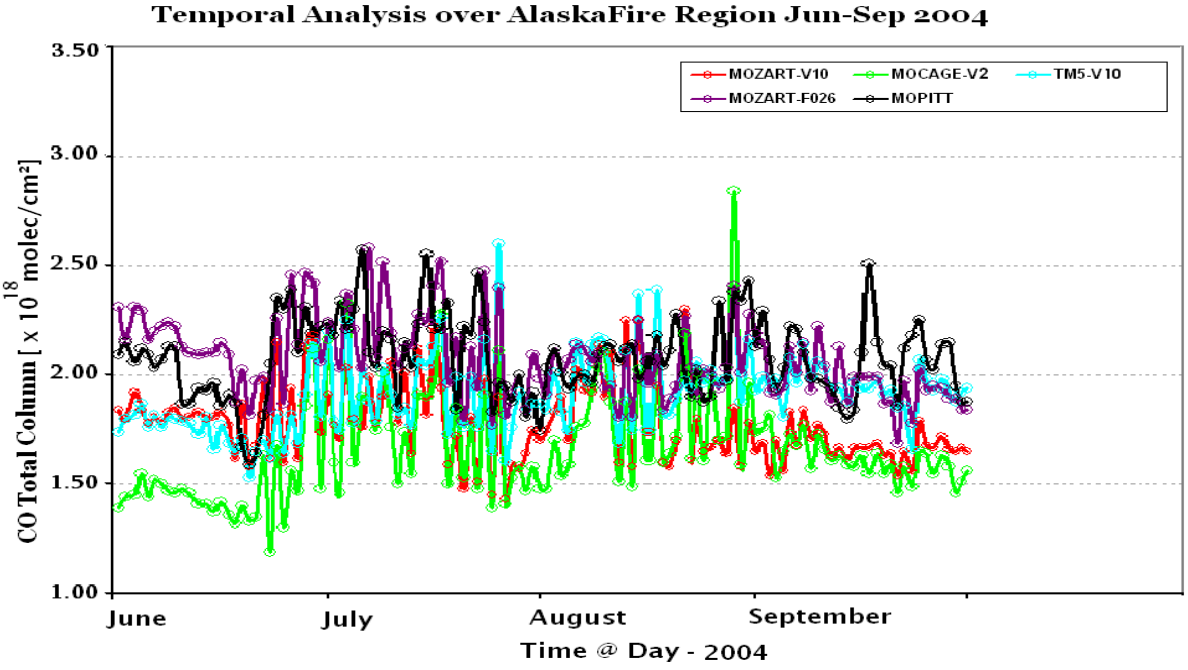


Figure 10 : Temporal data for daily CO total columns (molec/cm²) averaged over Alaska fire region from MOPITT (black), MOZART–V10 (red), MOCAGE–V2 (green), MOZART-F026 (violet) and TM5–V10 (cyan) Jun- Sep. 2004 over Alaska fire region. High MOPITT CO total column observed during July 2004 are consistent to large number of fire events within selected region.

However, during July 2004 all CTMs captured the increase in CO total columns related to fire events. Additionally, a temporal increase (after applying linear fits) in CO total columns from June is observed in MOCAGE-V2 and TM5-V10 outputs. This could be due to fact that CO emissions extracted from GFEDv2 inventory used by GRG-CTMs over Alska fire region exhibited maximum

CO emissions during August 2004 (see Figure11), inconsistent to AATSR fire data with maximum during July 2004 [Turquety et al., 2004].

Further differences among models could be due to the fact that all GRG–CTMs did not use same parameters, neither have similar chemistry schemes nor the same anthropogenic emission inventories (as only TM5-V10 and MOZART-V10 used GRG anthropogenic emissions inventories). It may probably be due to the use of different fire emission inventories: MOZART-V10, MOCAGE-V2 and TM5-V10 used the GFEDv2 8 day inventory, while monthly GFEDv2 fire emissions were used in MOZART-F026 simulations. Furthermore, GFEDv2 8 day fire emissions in TM5-V10 and MOCAGE-V2 simulations are interpolated to different pressure levels with a certain percentage of CO emissions. While in MOZART-V10 and F026 simulations all fire emissions are injected at the lowest model level. Figure 12 compares the average performance of each model based on monthly averaged scoring index over the Alaska fire region during time period of June – September 2004, calculated from CO total column from GRG-CTMs simulations and MOPITT observations on a daily basis. The 2nd reanalysis run MOZART-F026 performed best of all mainly because of assimilated MOPITT CO, showing relatively larger scores for the months of August and September as compared to scores in June and July 2004 (slightly overestimated CO vertical profiles).

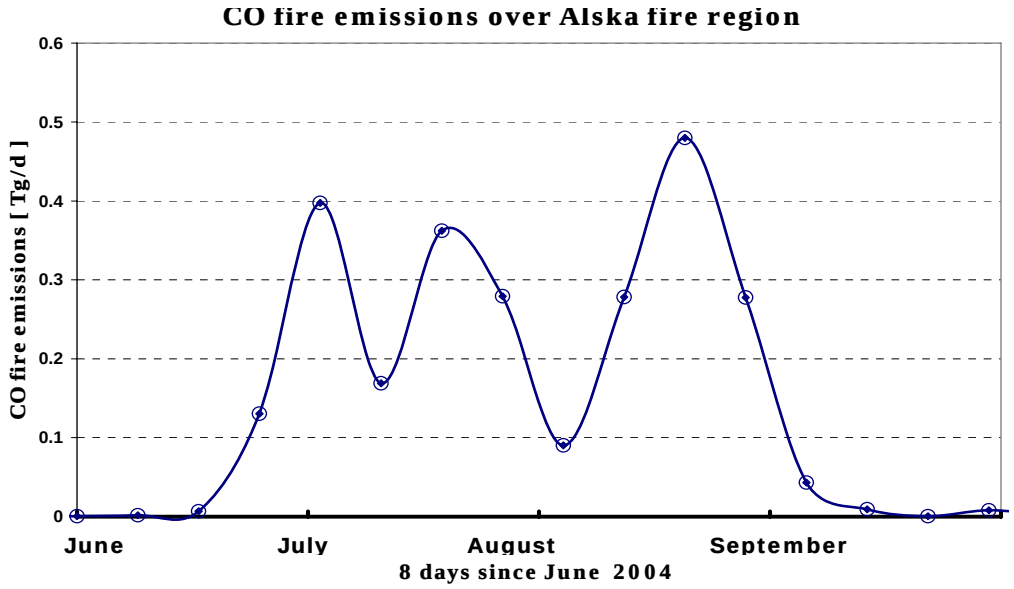


Figure 11: *CO emissions extracted from 8day GFEDv2 fire emission inventory over Alaska fire region during the time period of June-September 2004. CO emissions are in Tg per day.*

MOCAGE-V2 and MOZART-V10 exhibited the poorest performance particularly during July 2004. Besides of concurrence in temporal peaks of CO total columns from CTMs simulations and MOPITT CO total columns during July 2004; all GRG-CTMs performance is however higher than 80%.

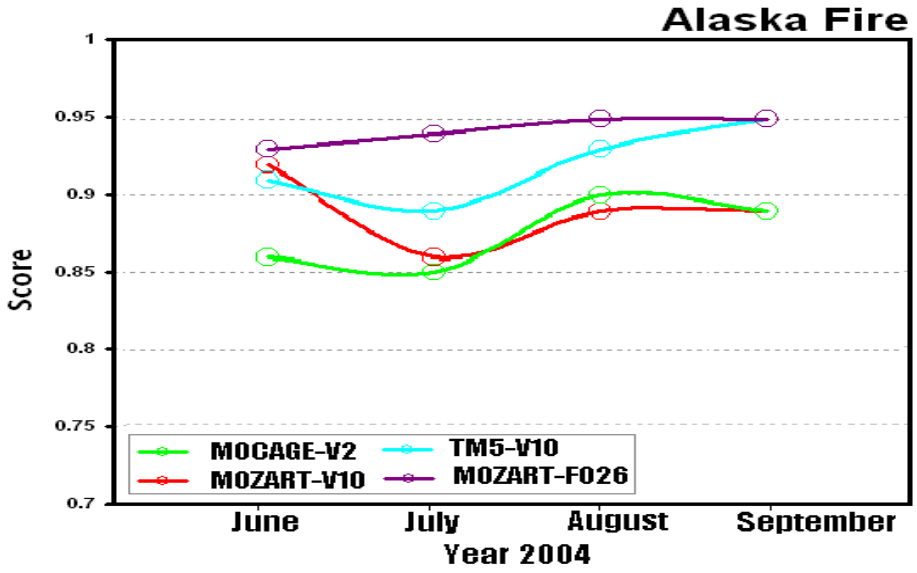


Figure 12: Scores for all GRG-CTMs (MOZART –V10 (red), MOCAGE–V2 (green), MOZART-F026 (violet) and TM5 –V10 (cyan)) calculated after implementing relation given in Equation 1. Score describes the model performance over the given time period and selected region on monthly basis.

10.4 Conclusions and recommendations

The intercomparison of CO model simulations (stand-alone and reanalysis) with MOPITT measurements was performed over Siberian and Alaska fire regions for the years 2003 and 2004, respectively. In general, GRG-CTMs were able to capture the correct seasonal behaviour in CO concentrations within the selected regions. The reanalysis run of MOZART-F026 compares better with MOPITT CO measurements during both case studies. For the monthly mean scoring indices calculated over Siberian and Alaska fire regions (presented in Figure 6 and 12 respectively) for MOZART-F026 always obtains higher values primarily because MOPITT CO was assimilated in these simulations. MOZART-V10 (with overestimated and underestimated CO concentrations - Figure 5) performed well over Siberian fire region as compared to TM5-V7 and MOCAGE-V2 stand-alone simulations. However, over the Alaska fire region, the TM5-V10 performance is better than MOZART-V10 and MOCAGE-V2 . The poor performance of GRG-CTMs (stand-alone runs) over Siberian and Alaskan fire regions seems to be mainly because of the use of inappropriate fire CO emission inventories inconsistent with AATSR fire data. Furthermore, inter-model differences can be due to different reaction rates, anthropogenic and fire emission inventories, horizontal girds and fire CO injection heights that were used in the different simulations.

10.5 References

1. Rast et al.: submitted to JGR, Nov. 2008.
2. Turquety, S., et al.: Inventory of boreal fire emissions for North America in 2004: Importance of peat burning and pyrocovective injection, J. Geophys. Res., 112, D12S03, doi:10.1029/2006JD007281, 2007
3. Huijnen et al., …

11. GEMS Experimental forecasts during POLARCAT campaigns 2008

Contributors: Khokhar Fahim, Law Kathy, Granier Claire (SA-UPMC, Paris, France), Jones Luke, Flemmeing Johannes (ECMWF, Reading, UK), Schultz Martin (FZJ, Juliech, Germany), and the modeller teams

11.1 POLARCAT campaigns 2008

11.1.1. Spring campaign

As a part of international polar year (IPY), major POLARCAT-France measurement campaign carried out at Kiruna Sweden during spring 2008 (see www.polarcat.no). POLARCAT-France (coordinated by K. Law et al) made 12 flights of ATR42 carrying MOZART instrument starting from 27 March and continued till 12 April 2008.
The main objectives for the spring campaign were to study the Arctic Haze phenomenon and its climate impact by performing air and ship borne measurements. The campaign involved aircraft, ship, and Arctic surface station observations. Also, forecasts from different chemistry models were used for flight planning to study the Arctic haze, anthropogenic pollution outflows, European agricultural and Siberian forest fire plumes.

11.1.2 Summer campaign

A second set of POLARCAT measurement campaigns were also carried out at Kangerlussuaq, Greenland (67° N, 50° W) during summer 2008 (see www.polarcat.no) from June 30th to July 14th 2008 as well as by NASA in Canada and YAK in Siberia. The main objectives for the summer campaign were:

- to study the inter-continental transport of Siberian and Canadian fire plumes and the chemical compositions
- to study North American and European anthropogenic pollution transported to Greenland and their impact on chemical composition over Greenland
- to validate satellite observations over Greenland (IASI, CALIPSO etc).

As part of our recommendations in the report of POLARCAT campaign spring 2008, fields for CO_BB_NRT (CO tracer from biomass burning emissions derived from near real time MODIS and SEVRI fire radiative power products) were included for this campaign in the forecast system

11.2 Data sets

11.2.1 GRG- forecasts

As a part of the GEMS project, we actively participated in flight planning for POLARCAT campaigns by using simulation results (termed as GRG-experimental forecasts) from MOZART3-IFS coupled system. MOZART3-IFS consists of the ECMWF *(European Centre for Medium range Weather Forecasts)* Integrated Forecast System model (IFS Cycle 32R3 http://www.ecmwf.int/research/ifsdocs) which is coupled to the MOZART3 chemistry transport model – (CTM). MOZART3 is designed to represent chemical and physical processes from the troposphere through the lower thermosphere. It is designed in a framework similar to the earlier

versions of MOZART but includes a more detailed formulation of physical processes in the troposphere (for details see [Kinsenet el., 2007]).
Fields for O_3, CO, NO_x, PAN, NO_y, CH_2O and C_2H_6 (as volume mixing ratios) with a resolution of 1.9° × 1.9°/ 60 vertical levels from the CTM and CO_EU, CO_EA, CO_SA, CO_US, CO_GL and CO_BB (tracers from different regions- for details see Table 9.1) from the IFS at a resolution of 125 x 125 km² and 60 atmospheric levels were provided. The forecasts were initiated each day at 00:00 h GMT and run for 3 days (for 4 days since October 2008). After processing and product generation the web products were made available at 00:00 h GMT on the base date + 1 day (t+24 hr).

Table 1: Description of GRG forecast products for POLARCAT campaigns in 2008

Species	Details	Model
O_3	Ozone volume mixing ratio	CTM
CO	Carbon monoxide volume mixing ratio	CTM
NO_x	Nitrogen Oxides ($NO+NO_2$) volume mixing ratio	CTM
NO_y	Total reactive nitrogen ($NO+NO_2+PAN+HNO_3+NO_3+N_2O_5$) volume mixing ratio	CTM
PAN	Peroxyacetylnitrate volume mixing ratio	CTM
CH_2O	Formaldehyde volume mixing ratio	CTM
C_2H_6	Ethane volume mixing ratio	CTM
CO_GL	Volume mixing ratio of a passive (50 days lifetime) tracer driven by anthropogenic sources of carbon monoxide	IFS
CO_EU	Volume mixing ratio of a passive (50 days lifetime) tracer driven by anthropogenic sources of carbon monoxide with sources in Europe	IFS
CO_US	Volume mixing ratio of a passive (50 days lifetime) tracer driven by anthropogenic sources of carbon monoxide with sources in North America	IFS
CO_EA	Volume mixing ratio of a passive (50 days lifetime) tracer driven by anthropogenic sources of carbon monoxide with sources in East Asia	IFS

CO_SA	Volume mixing ratio of a passive (50 days lifetime) tracer driven by anthropogenic sources of carbon monoxide with sources in Southern Asia	IFS
CO_BB_NRT $,@	Volume mixing ratio of a passive (50 days lifetime) tracer driven by biomass burning emissions derived from near-real time global MODIS and SEVIRI fire radiative power products. Yesterday's daily average of the emissions is used for the forecast	IFS

$: Fields for CO_BB_NRT are available since July 2008
@: Fire emissions are injected at lowest model level

11.2.2 NRT run "ez2m"

GRG- NRT forecast analysis run "ez2m" is available since 1 March 2008. The "ez2m" simulations are run with corrected fluxes, improved cloud corrections and with improved vertical velocity fields, and outputs are available at 3 hours temporal resolution and for 4 days. Hindcast evaluation compared to output from ez2m (NRT) forecast runs with IASI onboard METOP and MOZART photometer instrument onboard ATR42- French aircraft during the POLARCAT-France summer 2008 campaign are presented in section 9.4.

11.3 Snap shot Analysis

The GRG forecasts were compared to different forecast models such as LMDZ-INCA, MOZART4-GFS, GEOS-5 and FLEXPART. Although, all of these forecast models have many vertical levels (e.g. MOZART3 uses 60 vertical levels), output for only six different pressure levels e.g. surface, 925, 700, 500, 300, and 250 hPa, were compared. Details about different parameters of these forecast models used in forecast analysis are summarized in Table 2. A snapshot analysis of GRG (MOZART3-IFS) forecasts in comparison to the above mentioned forecast models is presented in this section.

Table 2: Specifications of different forecast models used for POLARCAT flight planning

Models	Range*/ Level	Forecast tendency	Resolution	CO	CO_EU	CO_GI	CO_BB	NO_X	NO_Y	PAN	O_3	SO_2 / SO_4	BC	Wind fields
MOZART3 – IFS	0 – 11km / 6	3 days	1.9°x 1.9°	x	x	x	x $	x	x	x	x	–	–	ECMWF
FLEXPART	0 – 13km / 13	5days	1° x 1°	x	x	x	x	x	–	–	–	x	x	ECMWF
MOZART-4/GFS ARCTAS	0 – 9km / 4	3days	2.8° x 2.8°	x	x	x	x	x	–	–	x	–	–	NCEP/ GFS

LMDZ-INCA	0 - 11km / 6	2days	3.75°x 2.5°	x	x	x	x	x	-	-	x	x	x	ECMWF
GEOS-5 Chem	0-31.2km / 27	5days	0.5°x 0.62°	x	x	x	x	-	-	-	-	x	x	GMAO

**Range means the for which forecasts are available, however, actual range (vertical domain of models)may be larger than this*

$: Fields for CO_BB_NRT are available since July 2008

11.3.1. MOZART3-IFS and LMDZ-INCA

LMDZ-INCA model (for details see Hauglustaine et al., (2004)) forecasts were available with 3.75°x 2.5° horizontal resolution in vertical domain of 11 km with 6 steps and 3 hour temporal resolution. LMDZ-INCA uses 6-hourly ECMWF wind fields interpolated onto the LMDz grid. Below are the x-sectional forecast plots for European CO tracer at 20°E for 2 April based on 1 April 2008. The forecasts compared in Figure 1 are x-sectional plots for latitudes from 45° to 90°N (left) while LMDZ-INCA forecast are from -90° to 90°N (right). Some discrepancies are found which are mainly because of differences in the resolution of model grids. In particular, European tracer CO concentrations are more confined to lower levels in the case of MOZART3-IFS system compared to LMDZ-INCA. Maximum CO concentrations (180 ppbv) predicted by MOZART-IFS forecast around 55°N are compatible with the LMDZ-INCA forecast around 55°N (>120 ppbv) close to the surface. Although there are some minor differences which are mainly because both CTMs have different spatial and vertical resolutions (Table 2). An overall fair agreement has been observed between the forecasts of LMDZ-INCA and GRG during POLARCAT spring 2008 campaign.

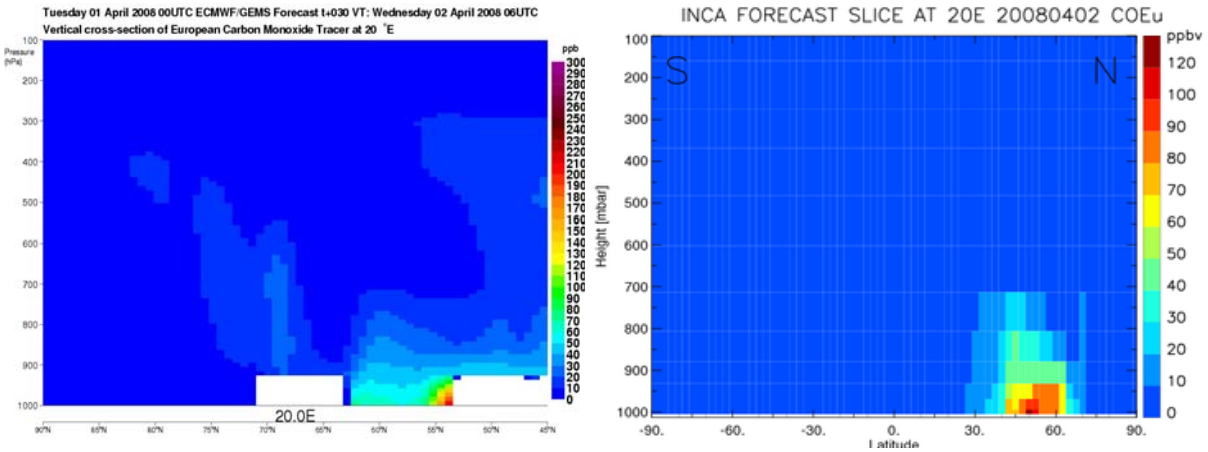


Figure 1: Latitudinal x-sections of MOZART3-IFS forecasts for CO European tracer (left) at 20° E compared with LMDZ-INCA forecast (right) for 02 April based on 01 April 2008. White gaps in MOZART3-IFS plots are due to areas with pressure less/more than pressure of given plot

11.3.2 MOZART3-IFS and MOZART-4/GFS

MOZART-4/GFS ARCTAS (Arctic Research of the Composition of the Troposphere from Aircraft and Satellites, for details see: [Pfister et al., 2008]) is an updated version of MOZART2, including tropospheric aerosols and other improvements. Forecasts are from model simulations with horizontal resolution of 2.8°x 2.8° available for a vertical range of 9 km (in 4 steps) and temporal resolution of 6 hours. Meteorological data for wind fields from NCEP/GFS (National Centre for Environmental Predictions/ Global Forecast System) are used in MOZART-4/GFS simulations with temporal resolution of 3 hours.

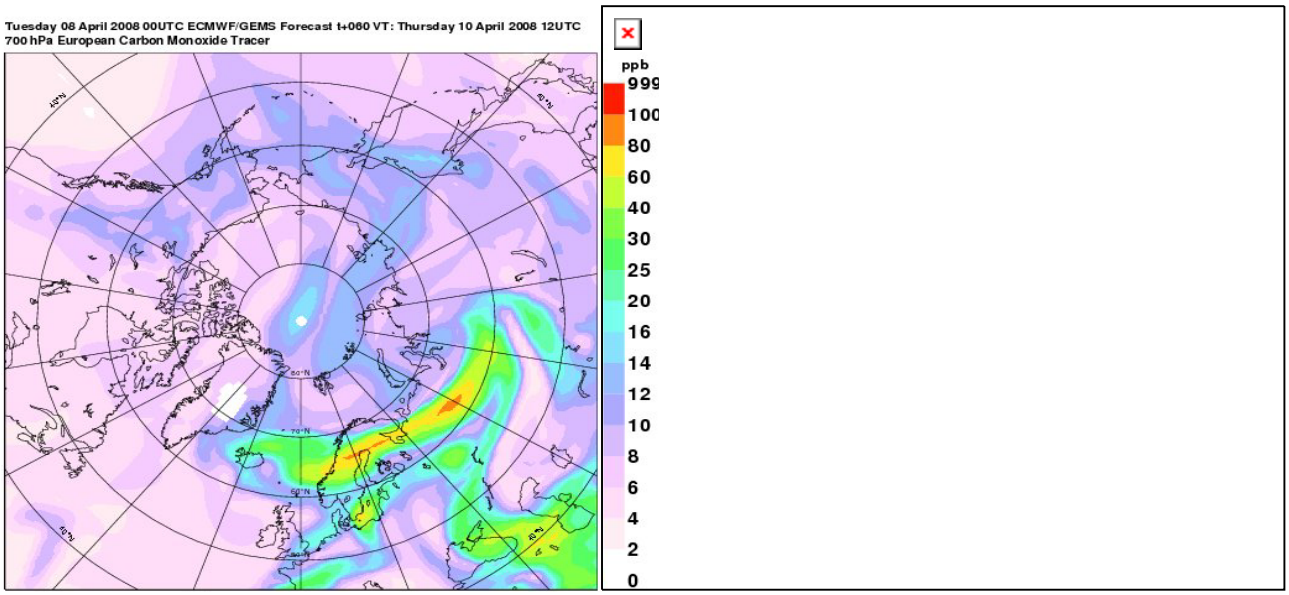


Figure 2: Inter-comparison between fields of CO as European tracer (CO_EU) at 700 hPa from MOZART3-IFS (left) and MOZART-4/GFS (right) on 10 April based on 08 April 08.

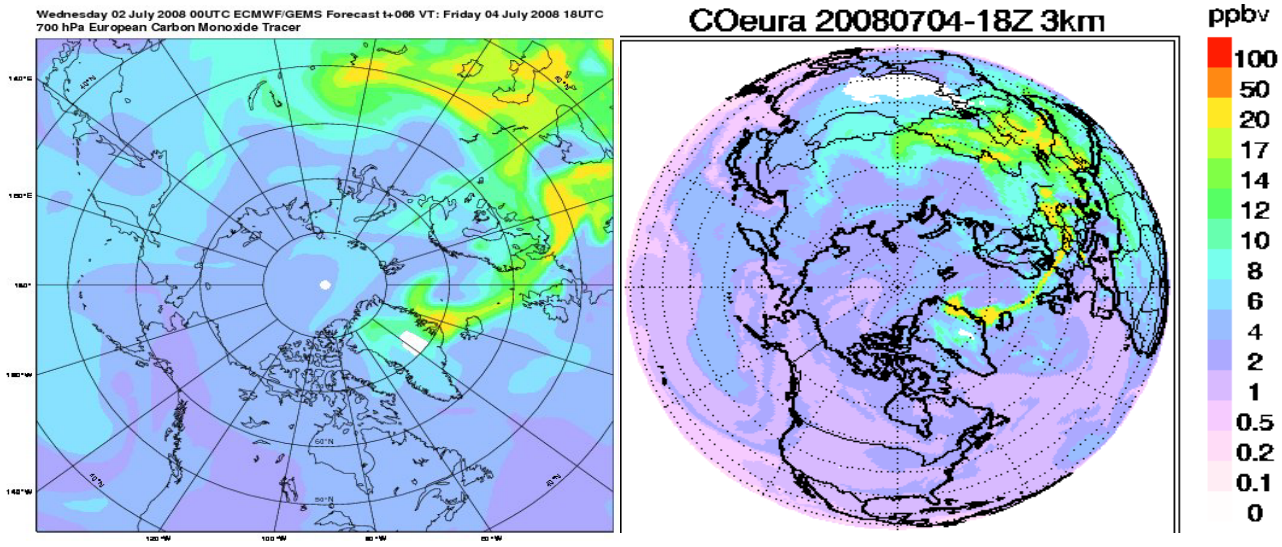


Figure 3: MOZART3-IFS forecasts for CO European tracer (left) at 700 hPa compared with MOZART-4/GFS forecast (right) for 04 July 2008 based on 02 July 2008.

An inter-comparison between CO_EU fields (forecasts) from both versions i.e. MOZART3-IFS (left panel) and MOZART-4/GFS (right panel) in the context of POLARCAT campaign spring 2008 is presented in Figure 2. The plot shows forecasts from 10 April 2008 based on 08 April 2008. Although there are some minor differences which are probably due to differences between the model grids, overall there is still a good consistency in the results. Plume patterns and predicted range of concentrations (e.g. maximum CO concentrations) are seen around 60°E and 63°N in both models (~80 ppbv - MOZART3-IFS with slightly higher), and are in a fair agreement with each other. Minor difference in CO concentrations and plume distributions are due to the fact that both forecast systems differ in temporal resolution, horizontal and vertical grids. Another comparison between CO European tracer forecasts from both versions (MOZART3-IFS (left panel) and MOZART-4/GFS (right panel)) during POLARCAT campaign summer 2008 is presented in Figure

3. Plots are forecasts from both models for 04 July 2008 based on 02 July 2008. There is an overall good consistency in the results. In particular, predicted carbon monoxide concentrations and its distributions (e.g. maximum CO concentrations from the plume reaching to eastern coasts of Greenland) are observed around 70° N latitude along 20°W (~30 ppbv) and are consistent with CO concentrations from MOZART-IFS. A similar behaviour was found during the both POLARCAT campaigns which demonstrate a strong inter-version consistency among the MOZART forecast systems.

11.3.3 MOZART3-IFS and GEOS-5

GEOS-5 forecast were available with 0.5° x 0.625° horizontal resolution and with a vertical domain of 32 km in 27 steps and 3 hour based temporal resolution. GEOS-5 uses 6-hourly wind fields (A6) from Global Modelling and Assimilation Office (GMAO). Bey et al. [2001] and Park et al. [2004] provide a detailed description of GEOS-5 model.

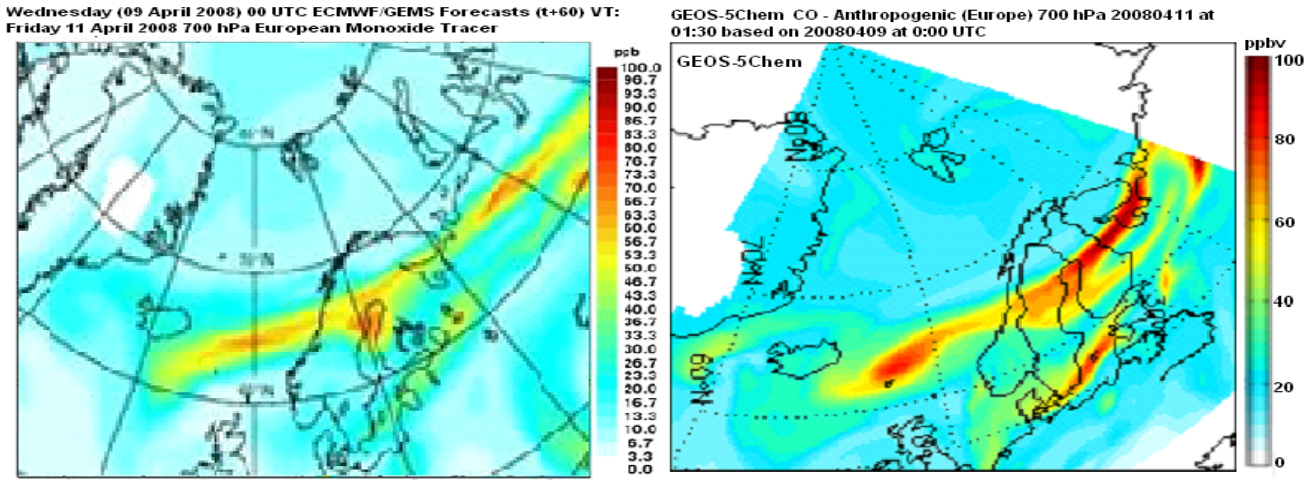


Figure 4: MOZART3-IFS forecasts for CO European tracer (left) at 700 hPa compared with GEOS-5 forecast (right) for 11 April 2008 based on 09 April 2008. MOZART-IFS plot (left) are zoomed out and indicate approximate region equal to GEOS-5 forecast region (right).

The forecasts from MOZART3-IFS compared with GEOS-5 model in Figure 4 for 11 April 2008 are based on 09 April 2008. Plots show the CO European tracer concentrations at 700 hPa over the domain of the POLARCAT campaign. There exists some agreement between both forecasts, especially, CO transport patterns and plume extensions over different regions (e.g. over Iceland and Scandinavian region). The minor differences in the CO plume patterns and predicted amount of CO concentrations; 70 ppbv for MOZART3-IFS and > 80 ppbv for GEOS-5 could arise due to differences in the time of forecast as MOZART3-IFS forecast is for 12:00 while GEOS-5 forecast is for 13:30. Additionally, the GEOS-5 plot (only representative for region given by magenta lines on MOZART3-IFS forecast) is more zoomed over the small area, and also has a much finer resolution (0.5° x 0.65°) which may contribute to these differences.

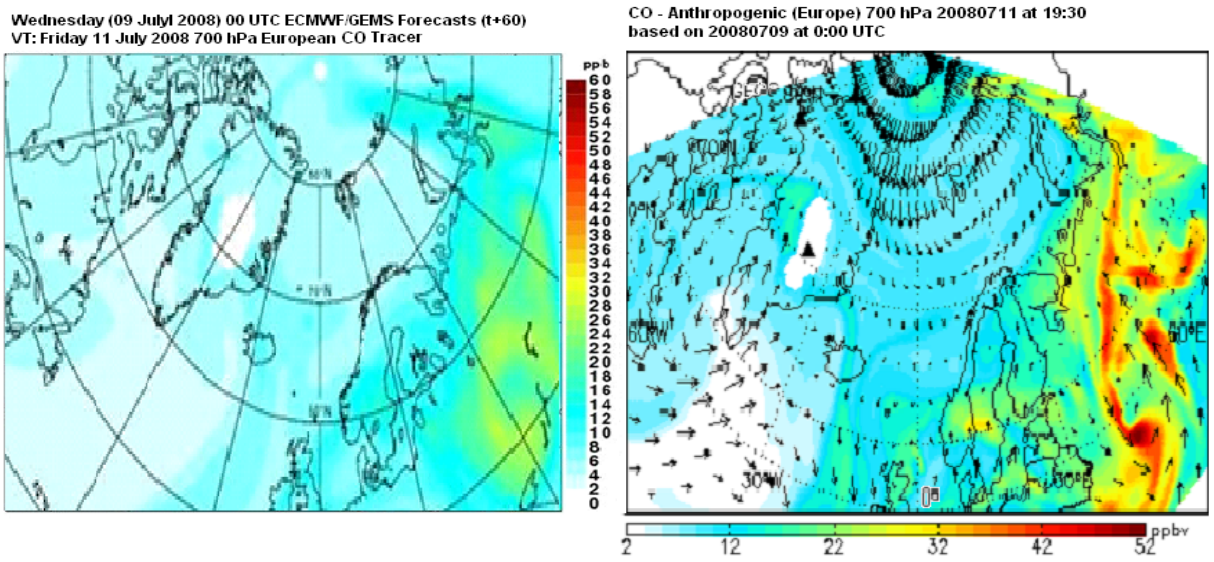


Figure 5: MOZART3-IFS forecasts for CO European tracer (left) at 700 hPa compared with GEOS-5 forecast (right) for 11 July 2008 based on 09 July 2008. Black arrows in right panel figure indicate wind direction.

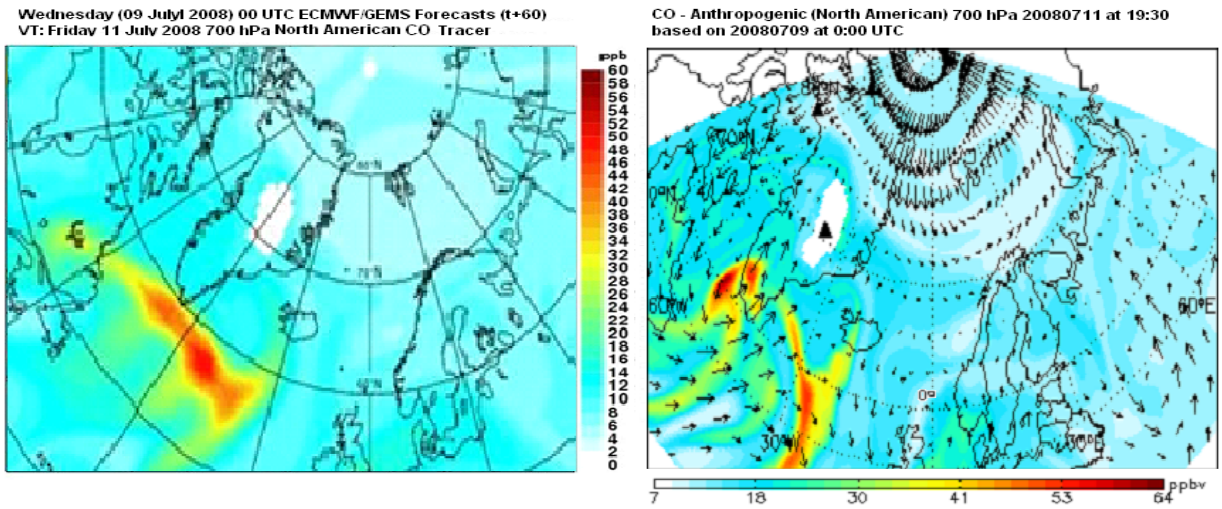


Figure 6: MOZART3-IFS forecasts for CO North American tracer (left) at 700 hPa compared with GEOS-5 forecast (right) for 11 July 2008 (t+60 hrs) based on 09 July 2008. Black arrows in right panel figure indicate wind direction.

Figure 5 shows comparison of forecasts between MOZART3-IFS (left) and GEOS-5 model (right) for 11 July 2008 based on 09 July 2008 (t+66 hr). The plots show the CO European tracer concentrations at 700 hPa over the domain of the POLARCAT-France measurement campaign. Both forecasts show CO transport patterns and plume extensions over North Pole, Iceland, Siberia, Eastern Europe and Scandinavian region. The predicted amounts of CO are less in case of MOZART3-IFS compared to the GEOS-5 forecast by about 15 ppbv. One reason could be due to difference in time of forecast as MOZART3-IFS forecast is for 18:00 while GEOS-5 forecast is for 19:30. Black arrows in GEOS-5 forecast are wind vectors indicating the wind direction.

A different behaviour was observed during the inter-comparison of CO North American tracer forecasts for the same day during POLARCAT summer campaign given in Figure 6. There was almost complete inconsistency observed with respect to CO plume transport patterns. For example, MOZART3-IFS forecasts (left-Fig.6) exhibited North American outflow mainly over

Atlantic Ocean, passing through south of Greenland and reaching south of Iceland, while GEOS-5 (right-Fig.6) exhibited a totally different outflow reaching south western parts of Greenland with three different fringes. Also, black arrows on GEOS-5 forecast demonstrate northward wind flow, almost perpendicular to North American plume patterns in MOZART3-IFS forecast.

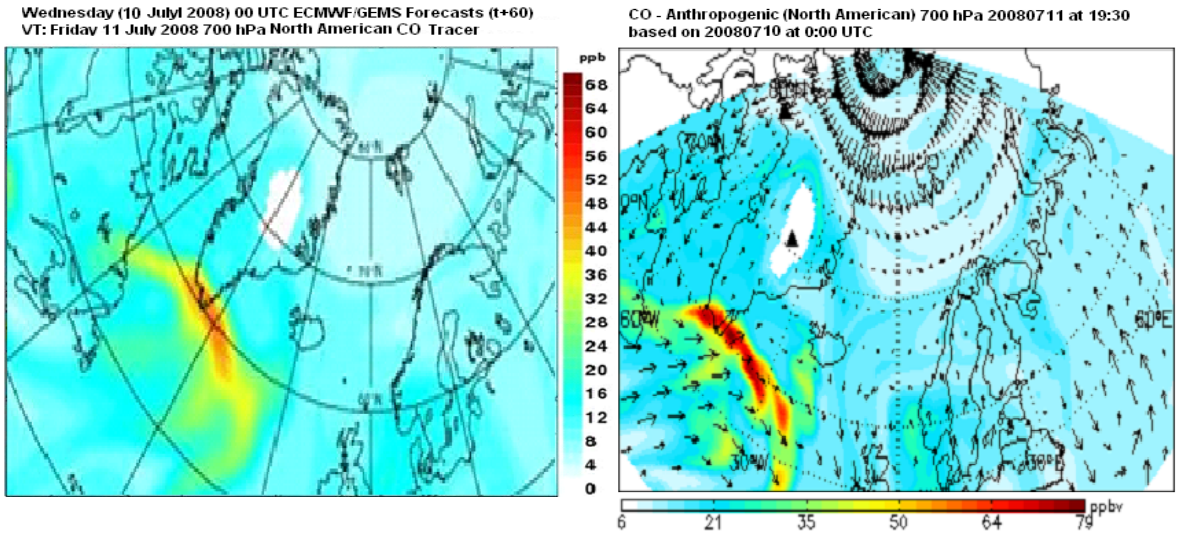


Figure 7: MOZART3-IFS forecasts for CO North American tracer (left) at 700 hPa compared with GEOS-5 forecast (right) for 11 July 2008 (t+42 hrs) based on 10 July 2008. Black arrows in right panel figure indicate wind direction.

This has been investigated further. Forecasts for the same day (11 July 2008) but based on 10 July 2008 (t+42 hrs) are shown in Figure 7. The inter-comparison presented for base time t+42 hrs exhibits a more consistent behaviour between both forecasts. Especially, plume patterns and predicted CO concentrations MOZART3-IFS (~70 ppbv) are in agreement with GEOS-5 (~70ppbv). Our analysis demonstrates the efficiency and consistency in the performance of MOZART3-IFS system, and that the discrepancy observed in Figure 6 was due to the GEOS-5 forecast.

11.3.4 MOZART3-IFS and FLEXPART

FLEXPART is an atmospheric trajectory and a particle dispersion model and uses meteorological input data from a variety of global and regional models, most commonly from ECMWF (for details see Stohl et al., 2005). FLEXPART forecast used in this analysis have a horizontal resolution of 1°x1° globally, 0.25°x 0.25° over Europe with vertical range of 13 km in 16 steps and a temporal resolution of 3 hours. Wind fields from ECMWF with temporal resolution of 6 hours are used in FLEXPART simulations.

MOZART3-IFS forecasts (left) were compared with FLEXPART forecast (right) for 11 April 2008 based on 9 April 2008 (08 April 2008 at 18:00 in case of FLEXPART) in Figure 8. Plots presented are CO European tracer concentrations at 700 hPa.
There is a disagreement found in the patterns of CO with higher CO concentrations in MOZART3-IFS and slightly narrower plume dimensions compared to FLEXPART. This could be explained by the fact that FLEXPART forecasts are CO values above CO back-ground levels while MOZART3-IFS forecasts are a passive (50 day lifetime) CO tracer driven by anthropogenic European sources of carbon monoxide. Also, there is a difference of 6 hours in the base time (+6

hours for MOZART3-IFS base time). Both forecasts for t+60hr time show a dense CO plume most probably originating from north-western Russia and Finland extending over Scandinavia transported over the Atlantic Ocean as far as south of Iceland and even to Greenland. Additionally, in the MOZART3-IFS plot (left) a smaller plume of CO originating from central France and reaching to Poland is missing in FLEXPART forecasts.

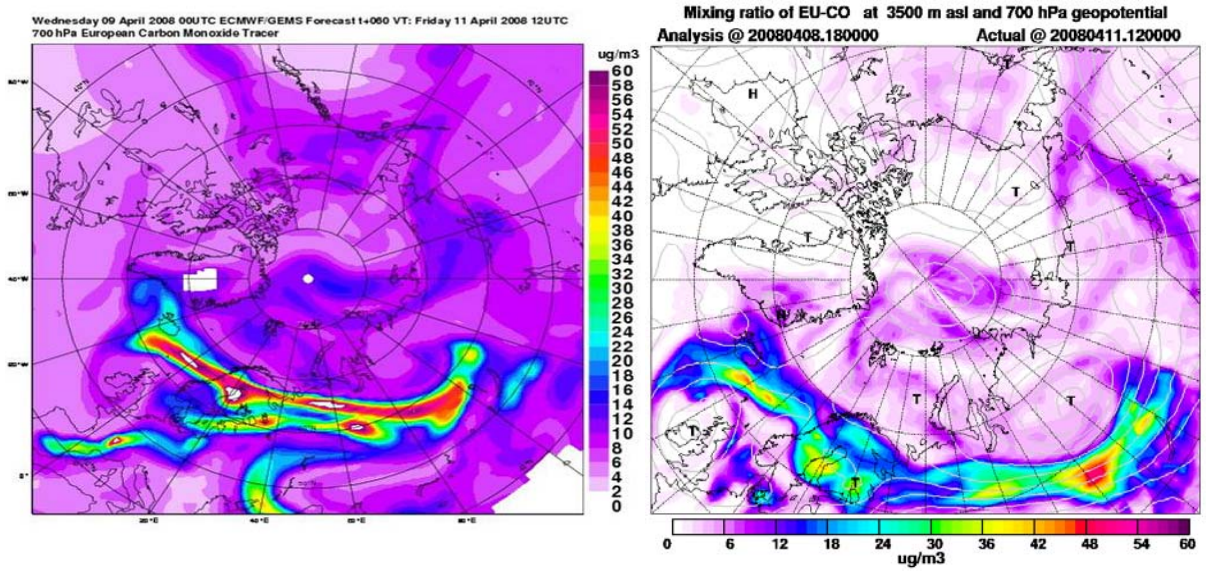


Figure 8: MOZART3-IFS forecasts for CO European tracer (left – t+60 hrs) at 700 hPa compared with FLEXPART forecast (right – t+66 hrs) for 11 July 2008 based on 09 July 08.

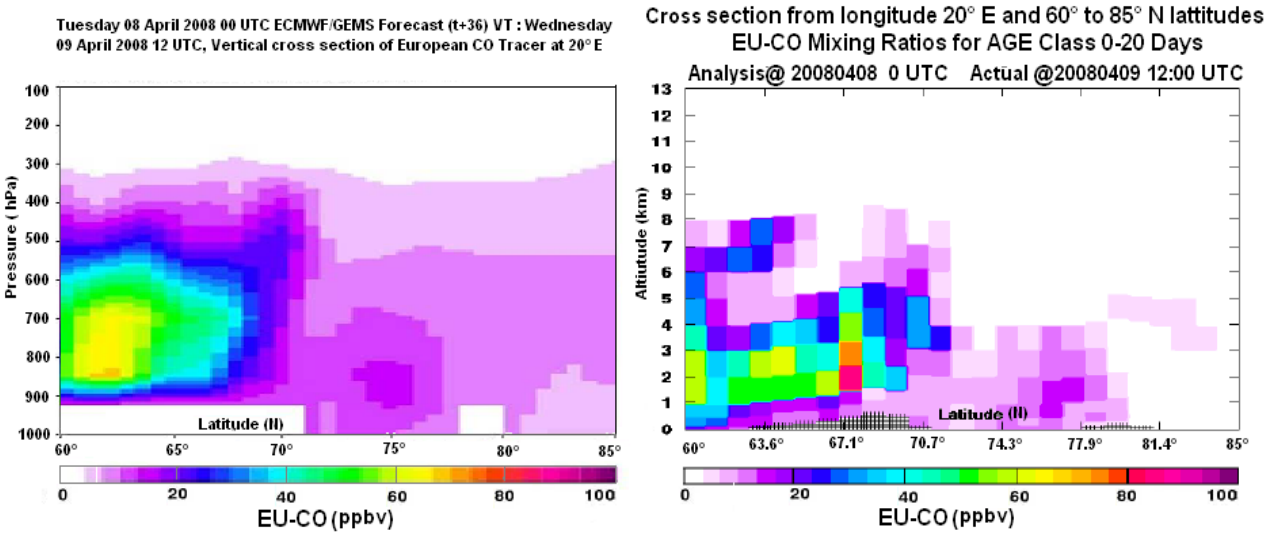


Figure 9: Latitudinal x-sections of MOZART3-IFS forecasts for CO European tracer (left) at 20° E compared with FLEXPART forecast (right) for 09 April based on 08 April 2008. White gaps in MOZART3-IFS plots are due to areas with pressure less/more than pressure of given plot

Additionally, we compared forecasts for 9 April 2008 based on 8 April 2008 as latitudinal x-sections at 20°E which are shown below. The FLEXPART plot shows x-sectional forecast for latitude range from 60° to 85°N is compared to GRG forecast. There still exist some discrepancies with FLEXPART European tracer CO concentrations showing a maximum around 67° N below 2 km while GRG forecasts show a maximum around 63°N and above 2 km altitude (there is no difference in base time in this case). Particularly the difference in predicted CO concentrations around 67° N is large; more than 80 ppbv in FLEXPART (right) while MOZART3-IFS only predicted 40 ppbv (left) around these latitudes. This discrepancy is further investigated in the

following section. An inter-comparison of forecasts from FLEXPART with GRG forecast used for the POLARCAT campaign summer 2008 is presented in Figure 10. This time we compared forecasts for 11 July 2008 based on 10 July 2008 (r t+42 hrs). Inter-comparison presented in Figure 10 is showing latitudinal x-sections at 50° W for North American CO tracer.

The FLEXPART plot shows x-sectional forecast (right- Fig10) for latitude range from 55° to 80°N as compared to MOZART3-IFS forecast (left- Fig10). Both forecasts predicted similar plume patterns with maximum concentrations above 55° N and below 5 km altitude. However, the plume in FLEXPART forecasts is well defined because of higher spatial resolution as compared to MOZART3-IFS forecast plot. Overall we have seen a reasonable agreement between both forecasts.

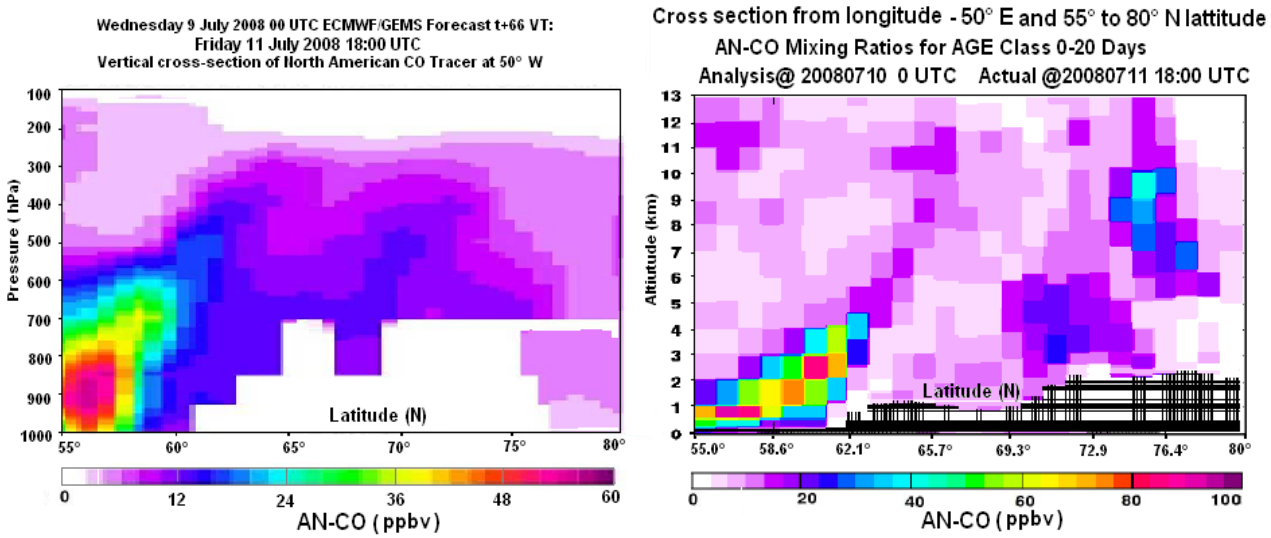


Figure 10: MOZART3-IFS forecasts for CO north American tracer (left) compared with GEOS-5 forecast (right) for 11 July 2008 based on 10 July 2008 for t+42 hrs. Shown are the curtain plots at 50 ° W longitude (Plots ranges from 45°-90° N (MOZART3-IFS) and 55°-80°N (FLEXPART)

A new CO passive tracer (CO_BB_NRT) driven by NRT fire emission derived from the MODIS-NRT radiative fire power product was introduced in MOZART3-IFS forecast system during the POLARCAT summer 2008 campaign.

During summer 2008 a large number of boreal forest fires took place in Canada and Siberia, Russia. Black dots in Figure 11 (right panel - FLEXPART forecast map) indicate the location of fire events. Fields for CO_BB_NRT from MOZART3-IFS were available from 3 July 2008.

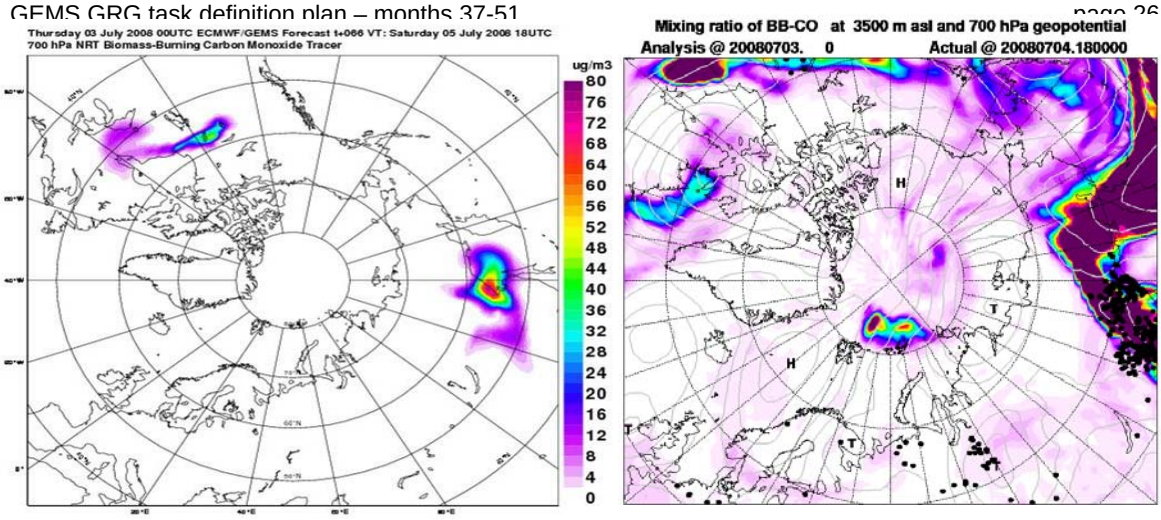


Figure11: MOZART3-IFS forecasts for CO biomass burnings tracer (left) compared with FLEXPART CO fire tracer (right) for 4 July 2008 based on 03 July 2008 at 700 hPa. Black dots in FLEXPART map indicate the location of European agricultural and Siberian forest fires during summer 2008.

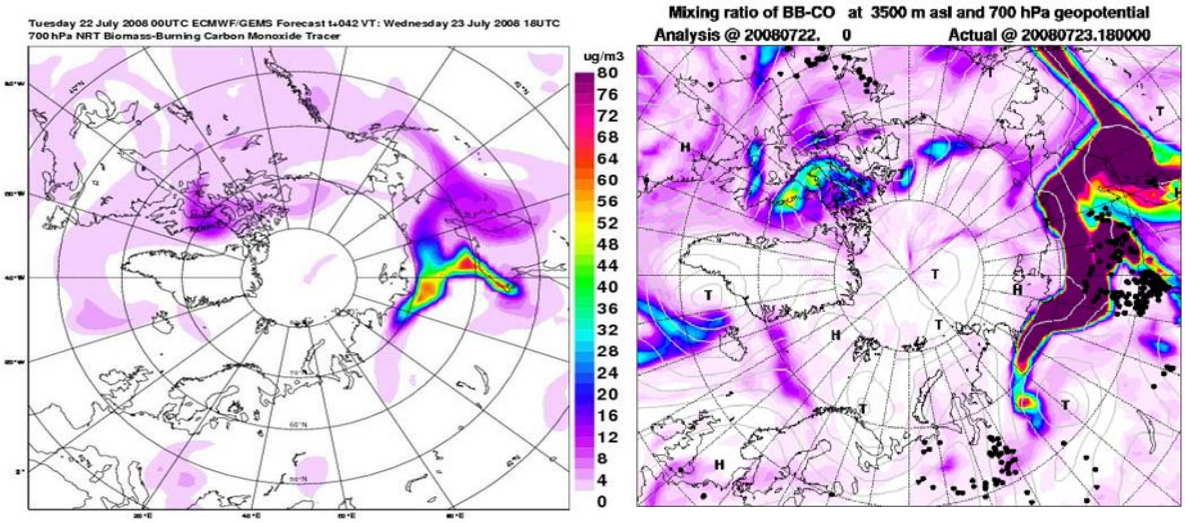


Figure 12: MOZART3-IFS forecasts for CO emissions from biomass burnings (left) compared with FLEXPART CO fire tracer (right) for 23 July 2008 based on 22 July 2008 at 700 hPa. Black dots indicate the location of fires.

Figure 11 (left panel) shows the CO_BB_NRT forecast for 4 July 2008 based on 03 July 2008 at 700 hPa. As compared to FLEXPART forecast for CO as a fire tracer for same day, two smaller plumes are only visible from the MOZART-IFS forecasts. However, MOZARTS3-IFS started forecasts for CO_BB_NRT from an empty atmosphere only on 3 July 2008 and so the forest fire tracer did not have time to build up.

It was expected that tracer mass would gradually increase in the first month but reasonable transport patterns would be visible after a couple of days. That can be clearly seen from next example presented in Figure 12. The forecast compared in Figure 12 are for CO tracer from biomass burning for 23 July 2008 based on 22 July 2008 at 18:00 hr and show fair agreement with FLEXPART CO forecasts. The predicted CO patterns and its transport to and around the Arctic Circle, North America and over the Atlantic Ocean are consistent in both plots. Minor differences in predicted concentrations could be due to the fact that FLEXPART CO concentrations are only

above background levels. However, results demonstrate the efficiency of the GRG experimental forecast system, particularly for newly introduced tracer CO_BB_NRT.

11.4 Hind cast Analysis

GRG- NRT forecast analysis run "ez2m" is available for 15 November 2007 onward. The "ez2m" simulations outputs at T+00 hrs are compared with observations of CO and O_3 by MOZART photometer onboard ATR42 French aircraft during POLARCAT-France field campaign at Kangerlussaque, Greenland.

11.4.1 ATR42 – French aircraft

POLARCAT-France conducted two measurement campaigns at Kiruna, Sweden and Kangerlussuaq, Greenland during spring and summer 2008, respectively. During the POLARCAT summer campaign from June 30 to July 14 with payload onboard ATR42 French aircraft based in Kangerlussuaq (western Greenland, 67N, 50W). In total 12 scientific flights were made over the area of interest. We, as a part of Paris team, participated by using GRG-forecasts in flight planning to have better coverage of pollution transport. A preliminary analysis of CO and O_3 observation data from one of POLARCAT flights (data is available for internal use only) are presented in Figure 13. We compared MOZART3-IFS simulations (ez2m) with MOZART photometer onboard ATR42 French aircraft (referred as POLARCAT data) for CO and O_3 concentrations on 5 July 2008. O_3 and CO from the MOZART3-IFS (line) along the POLARCAT flight-path (dashed) were compared. The POLARCAT CO (red - dashed) and O_3 (blue –dashed) observations were averaged over 1-minute intervals. The MOZART3-IFS output (red – line) shown in the line plot (Fig.13- A) is simply the data from the nearest grid-box at the nearest output step (every 3 hours). The flight-path: pressure, altitude and distance from the start-point, is also shown in Fig.13 (panel-B). The flights only cross a few MOZART3-IFS grid-boxes, and so the model cannot capture the observed horizontal variability. However, mean values look reasonable. Much of the variability shown in the MOZART3-IFS fields comes from vertical motion of the plane. The 3D flight trajectory plots coloured according to observed O_3 and CO concentrations along with lat-lon plots of MOZART3-IFS forecast on 5 July 2008 for O_3 and CO over the area of interest are presented in Figure 13, panel C and D, respectively. The peak in CO observations on 5 July 2008 around 12:20 caused by transport of CO from North America reaching the west coast of Greenland is well captured by both POLARCAT observations and MOZART3-IFS simulations. In general, we see a reasonable agreement between POLARCAT field observations and MOZART3-IFS NRT run ez2m.

11.4.2 IASI NRT comparison

The Infrared Atmospheric Sounding Interferometer (IASI) is a Michelson Interferometer onboard METOP satellite launched on 19 October 2006 in polar orbit at an altitude of about 800 km. It measures the spectral distribution of the atmospheric radiation in spectral range from 15.5 to 3.62 μm. This kind of resolution would enable us to derive contents of chemical components present in small concentrations in the atmosphere but which play a major role in the greenhouse effect (i.e., ozone, methane, nitrous oxide and carbon monoxide). The IASI instrument observes the Earth up to an angle of 48.5° perpendicular to both sides of the satellite track, which corresponds to a swath width of about 2,000 km on the ground so that global coverage will be achieved in 12 hours. For further details about retrieval algorithm see (Turquety et al, 2004; Clerbaux et al., 2009). During the POLARCAT summer campaign, IASI NRT analysis was available on daily basis for a flight planning in addition to chemical model forecasts

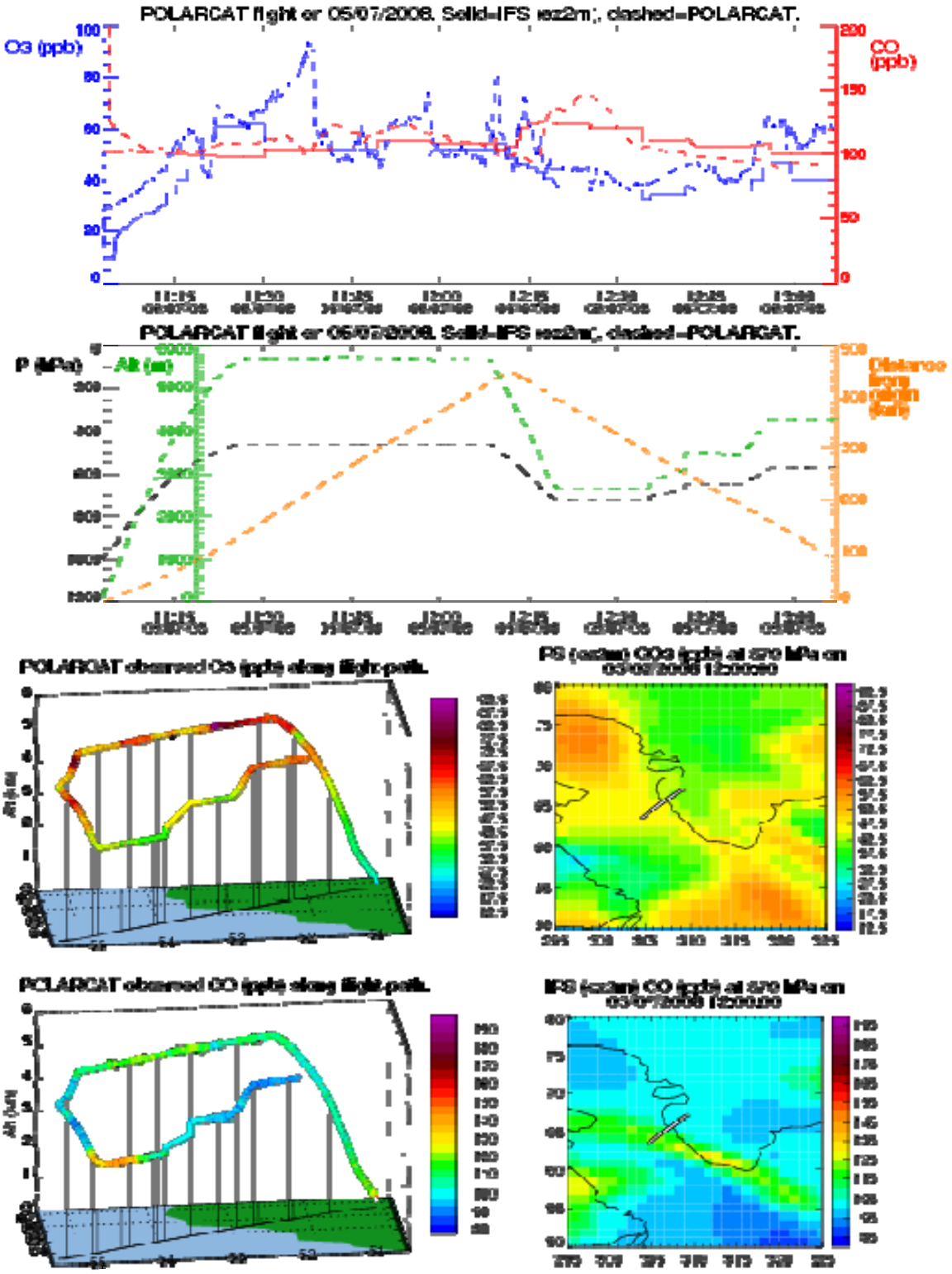


Figure 13: shows intercomparison (A) of MOZART3-IFS simulations with POLARCAT flight data for CO and O_3 on 5 July 2008. Information about flight altitude and distance from starting point (B). MOZART3-IFS forecasts for O_3 and CO over Kangerlussuaq, Greenland and 3D plot showing POLARCAT flight data for O_3 and CO at flight altitude is given by panel C and D respectively.

A MOZART3-IFS NRT run (ez2m) map of CO at 286 hPa for 10 July based on 09 July 2008 (t+38hr) at 9 km altitude and 14:00 hrs is presented in Figure 14 (left). MOZART3-IFS exhibited CO emissions from Siberian fire transported towards Polar Regions, along the Arctic circle passing over northern Canada and Alaska and finally reaching to the north of Norway. Another plume was observed over central parts of Greenland splitting further into two narrow plumes one leading to southeast of the Greenland and other reaching north of the UK after passing over Iceland. Additionally, CO outflow from north-eastern USA and its transport to Europe was observed. We compared ez2m outputs with IASI near real time analysis of CO concentrations for 10 July 2008 at 8.5 km altitude level and 13:57 hrs (IASI overpass time 13:57 UTC).

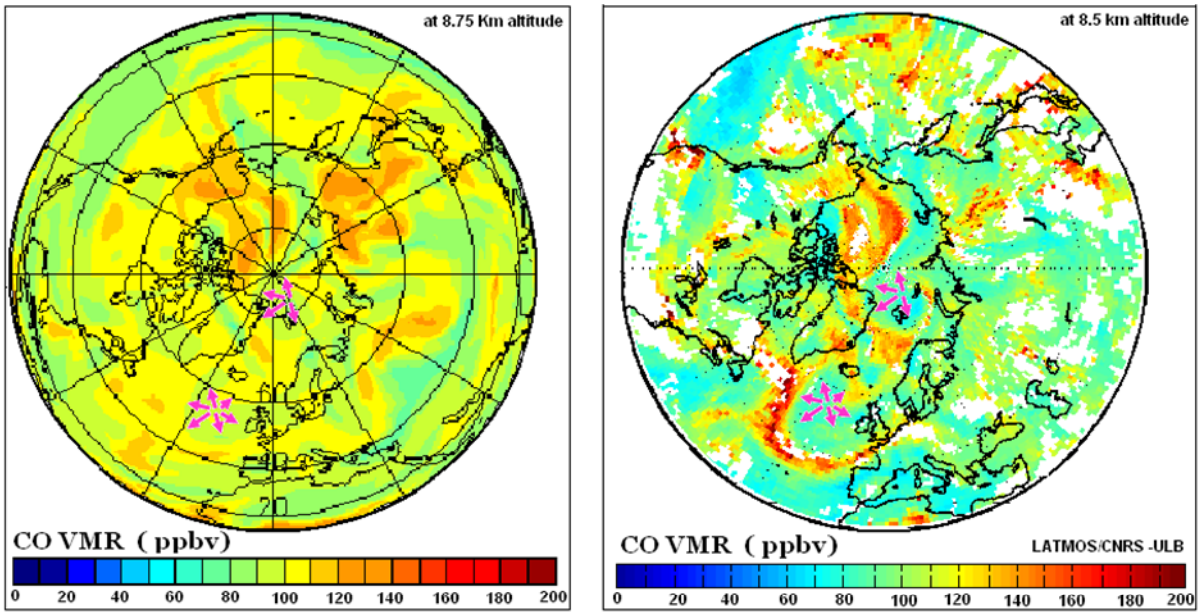


Figure 14: MOZART3-IFS forecast analysis run for global CO tracer (left) at 8.75km altitude for 10 July 2008 based on 09 July 2008 compared with IASI-METOP map (right) showing NRT analysis of CO concentrations at 8.5 km on 10 July 2008 (*courtesy of M. Pommier, C. Clerbaux and S. Turquety from LATMOS-UPMC France*). Magenta arrows indicating the patterns of CO plume consistent in both plots. White gaps in IASI plot are due excluded cloudy pixels

Inter-comparison presented in Figure 14 exhibited an overall good agreement between MOZART3-IFS analysis run and NRT IASI CO concentrations at 8.5 km. Regions indicated with magenta arrows show that inter-continental transport of Siberian plumes and North American (NA) outflow over Atlantic Ocean and its transport to Europe show a reasonable agreement. However, CO concentrations at 8.75 km in MOZART3-IFS plot are less than the IASI CO concentrations at the 8.5 km level. These differences may in part be due to the fact that IASI CO concentrations at 8.5 km are compared to MOZART3-IFS CO concentrations at 8.75 km. Discrepancies in IASI CO total columns for latitudes above 45°N can be up to 30% during fires events [George et al., 2009]. IASI CO total columns are converted into CO concentrations by applying IASI averaging kernels (for detail see Clerbaux et al., 2009; George et al., 2009).

Furthermore, IASI averaging kernels may play a role in the observed differences because they show less sensitivity to the lower surface and maximum sensitivity at lower troposphere [George

et al., 2009] and MOZART-IFS CO concentrations are compared without application of IASI averaging kernels. White gaps in IASI plot are because IASI observations from cloudy pixels are not included in the analysis. In general we see a reasonable agreement between MOZART-IFS outputs and IASI CO data, however, for a qualitative and quantitative comparison we should take into account the afore mentioned effect.

11.5 Conclusions and recommendations

GRG (MOZART3-IFS) forecast products are not operational as they are prepared only on experimental basis for the POLARCAT campaign spring and summer 2008. Nevertheless, forecasts show a good consistency when compared to other model forecasts, ATR42 aircraft measurement and IASI NRT analysis. However, there is still the need to further improve the capability of MOZART3-IFS. The following are some recommendations:

- Product list can be extended by including forecast of other species like SO_2, SO_4, BC etc.
- Run at higher resolution
- Investigate whether the current delay (longer than other products) leads to less accurate forecasts. If possible, decrease the lag time of the forecast products
- Improve the forest fire treatments (e.g proper injection heights for fire CO etc.)

Acknowledgment

Authors fully acknowledge IASI evaluation team (C. Clerbaux, M.Pommier, S. Turquety) at LATMOS, UPMC Paris, France for their cooperation and providing IASI NRT analysis during the POLARCAT summer 2008 campaign. We also acknowledge the POLARCAT-France (coordinated by K. Law and G. Ancellet) for providing with ATR42 flight data.

11.6 References

1. Bey, I., M. G. Schultz, D. J. Jacob, R. M. Yantosca, J. A. Logan, B. D. Field, A. M. Fiore, Q. Li, H. Y. Liu, and L. J. Mickley: Global modeling of tropospheric chemistry with assimilated meteorology: Model description and evaluation, J. Geophys. Res., 106(D9), 23,073 – 23,096, 2001
2. Clerbaux, C., Boynard, A., Clarisse, L., George, M., Hadji-Lazaro, J., Hurtmans, D., Herbin, H., Pommier, M., Razavi, A., Turquety, S., Wespes, C., and Coheur, P.-F.: Monitoring of atmospheric composition using the thermal infrared IASI/METOP sounder, submitted to Atmos. Chem. Phys. Discuss., IASI Special Issue, 2009
3. George M., C. Clerbaux, D. Hurtmans, S. Turquety, P.-F. Coheur, M. Pommier, J. Hadji-Lazaro, D. P. Edwards, H. Worden, M. Luo, C. Rinsland and W. McMillan: Carbon monoxide distributions from the IASI/METOP mission: evaluation with other space-borne remote sensors, acp-2009-103 submitted to ACPD, 2009
4. Hauglustaine, D. A., F. Hourdin, S. Walters, L. Jourdain, M.-A. Filiberti, J.-F. Larmarque, E. A. Holland: Interactive chemistry in the Laboratoire de Météorologie Dynamique general circulation model: description and background tropospheric chemistry evaluation, J. Geophys. Res., 109, D04314, doi:10.1029/3JD003957, 2004. (© 2004 by the AGU), 2004
5. Kinnison, D. E., G. P. Brasseur, S. Walters, R. R. Garcia, D. R. Marsh, F. Sassi, V. L. Harvey, L. Emmons, J. F. Lamarque, P. Hess, J. J. Orlando, X. X. Tie, W. Randel, L. L. Pan, A. Gettelman, C. E. Randall, C. Granier, T. Diehl, U. Niemeier, A. J. Simmons: Sensitivity of chemical tracers to meteorological parameters in the MOZART-3 chemical transport model, J. Geophys. Res., 112, D20302, doi:10.1029/2006JD007879, 2007
6. Park, R. J., D. J. Jacob, B. D. Field, R. M. Yantosca, and M. Chin: Natural and transboundary pollution influences on sulfate-nitrate- ammonium aerosols in the United States: implications for policy, J. Geo- phys. Res., 109, 15204, doi:10.1029/2003JD004473, 2004

7. Pfister, G. G., L. K. Emmons, P. G. Hess, J.-F. Lamarque, J. J. Orlando, S. Walters, A. Guenther, P. I. Palmer, and P. J. Lawrence: Contribution of isoprene to chemical budgets: A model tracer study with the NCAR CTM MOZART-4, J. Geophys. Res., 113, D05308, doi:10.1029/2007JD008948, 2008
8. Stohl, A., C. Forster, A. Frank, P. Seibert, and G. Wotawa, Technical Note: The Lagrangian particle dispersion model FLEXPART version 6.2. Atmos. Chem. Phys. 5, 2461-2474, 2005
9. Turquety, S., Hadji-Lazaro, J., Clerbaux, C., Hauglustaine, D. A., Clough, S. A., Cassé, V., Schlüssel, P., and Mégie, G.: Operational trace gas retrieval algorithm for the Infrared Atmospheric Sounding Interferometer, J. Geophys. Res., 109, D21301, doi:10.1029/2004JD004821, 2004

12. Evaluation of stratospheric ozone

Contributors: Lefever Karolien, Chabrillat Simon, Daerden Frank (BIRA, Brussels, Belgium), and the modeller teams

The evaluation of the GRG model and data assimilation system in respect to stratospheric ozone was a major topic during the GEMS project. In this section we present the evaluation of

- Offline runs with MOCAGE, MOZART and TM5
- GEMS-reanalysis with IFS-MOZART
- GEMS-forecast with IFS-MOZART

Special emphasis was put on the year 2003 because ozone analyses by the BASCOE system, which assimilates MIPAS data, were available as a reference. The 2003 Antarctic ozone hole was also the second largest ozone hole area ever recorded (11.1 million square miles on September, 24th).

Therefore, in addition to the two tropospheric case studies (being the 'Summer 2003 European heat wave' and the 'Impact of boreal fires', cfr. Sections 8, 9 and 10), a third stratospheric case study was set up, in order to evaluate the GEMS-GRG model performance for 2003. Data produced by these models were compared compared with measurements of stratospheric O_3 retrieved from several satellite instruments: EP-TOMS and POAM. We present the results for both columns and profiles.

12.1 Datasets and methodologies

12.1.1 Datasets

Global evaluations for total ozone columns have been made both with BASCOE/MIPAS analyses and the independent TOMS total ozone column data. We also compared the ozone volume mixing ratio at 5 selected levels: 4 pressure levels (around 1 hPa, 10 hPa, 54 hPa and 100 hPa, corresponding to levels 5, 14, 22 and 25 in the set of 61 ECMWF levels) and 1 isentropic level at potential temperature 475K. The latter allows us to follow ozone through an adiabatic vertical flow. The 475K-level is situated in the lower stratosphere at an altitude of about 22 km, which is the altitude of maximal chemical destruction of the ozone layer. It lies very close to the pressure level at about 54 hPa. We used polar satellite profiles from POAM for an independent comparison.

Independent assimilation system:

- BASCOE v4q30: **B**elgian **A**ssimilation **S**ystem of **C**hemical **O**bservations from **E**nvisat

Observational data sets:

Data set used for the evaluation of model runs:

- TOMS: **T**otal **O**zone **M**apping **S**pectrometer
- POAM III: **P**olar **O**zone and **A**erosol **M**easurement

In the following tables (Table 12.1 and 12.2), we give an overview of the characteristics of each of the comparison datasets.

Table 12.1: Overview of the characteristics of the different observational data sets, used for the evaluation of stratospheric ozone.

obs. dataset	method measurement	vert. resol.	horiz. resol. (lat x lon)	period	time density
TOMS	total ozone mapping spectrometer (Mc Peters et al., 1998) measures total O_3 columns	none	1.00º x 1.25º	20021001 till 20031231	daily average columns
POAM III	visible/NIR solar occultation photometer onboard SPOT-4 (Lucke et al., 1999) measures stratospheric constituents O_3, HNO_3, H_2O, and aerosol extinction	1 km 61 levels	polar regions latitudes vary throughout the year: ± 54º - 72º N ± 62º - 88º S lon-spacing: ± 25º	20021101 till 20041231	on average 14 profiles per day (at most 15) daily profiles located at a fixed latitude on each hemisphere

Table 12.2: Overview of the characterstics of BASCOE, the assimilation system used as a reference in the evaluation of stratospheric ozone.

assimilation system	method measurement	vert. resol.	horiz. resol. (lat x lon)	period	time density
BASCOE v4q30 (CTM: v4q23)	4D-var assimilation system based on MIPAS data (Errera et al., 2008) models 57 chemical species (*stratospheric* chemistry only)	37 levels from surface to 0.1 hPa (28 upper levels identical to the stratospheric ECMWF levels, 9 lower levels subset of the ECMWF tropospheric levels (in 61 level product))	3.75º x 5º	20021101 till 20040331	5 steps per day, every 6h

The model performance varies seasonally and regionally. One of our goals is to investigate this variability. Therefore, we need a good temporal and spatial coverage of the observation sample. Each of the above mentioned instruments/satellites has specific characteristics making them appropriate to probe different parts of time and space. Due to the orbit of the satellite or the specific design of the instrument, the data are sometimes a bit sparse in time and space, hampering the deduction of spatial and seasonal patterns. However, in general, the coverage is quite satisfying.

TOMS measures backscattered UV radiances with daily near-global coverage. Due to the lack of sunlight at the very high latitudes, there are no data available from March till September in the southern high latitudes and from October till March in the northern high latitudes. *POAM* only allows assessing ozone in the Polar Regions, so no global patterns can be retrieved. Envisat follows a helio-synchronous polar orbit, thus allowing *MIPAS* to sample globally. The coverage is quite uniform in time, with only minor variations in latitude. On the other hand, MIPAS only observes around 10h30 in the morning and the evening, so that no detailed diurnal variations can be mapped.

Due to the relatively low horizontal resolution of BASCOE, it is difficult to assess large horizontal gradients and to capture realistically, e.g., the dynamical evolution during the vortex split.

12.1.2 Methodologies

The average deviations between two datasets are captured by the bias parameter. Throughout the entire evaluation, we use the modified normalized mean bias (MNMB) B'_n, as defined in Annex 7, to quantify the bias between the CTM/GEMS-GRG runs on the one hand and the observational comparison datasets and BASCOE/satellite data on the other hand. When interpreting the bias plots, one needs to keep in mind the theoretical bias and uncertainties of BASCOE itself w.r.t. the `truth´ (see Table 12.3). Errera et al. (2008) clearly illustrated that in their Fig. 5, shown below as Fig. 12.1, for the period September - October 2003. POAM-III errors are below 5% throughout the polar stratosphere (Randall et al., 2003).

To compare the performance of the different models among each other, and the evolution of one model in the coarse of time, scores were introduced. The applied scoring method is defined as agreed within the GEMS-GRG team and described by Vincent Huijnen in the document “Scoring European tropospheric NO_2” (see also Annex 7). In what follows, we describe our procedure for **total ozone columns only** (slightly adapted to our purposes):

- We interpolate the O_3 columns at 12h of the model and the observations to a 1° x 1°grid (lat x lon)
- For each day and for each latitude, we average the O_3 columns over the longitude
- For each latitude, we calculate the median of the observations over a month, $\text{median}_{\text{month}}(O_i)$, and the median of the bias between the model and observations over a month, $\text{median}_{\text{month}}(M_i - O_i)$
- For each latitude, we can now compute the ratio $\dfrac{|\text{median}_{\text{month}}(M_i - O_i)|}{2\ \text{median}_{\text{month}}(O_i)}$
- We calculate the area weighted average of this ratio within a pre-specified latitude band, with the weights defined as

$$w_i = \frac{\cos(lat_i)}{\sum_i \cos(lat_i)}$$

and define the score as

$$S = 1 - avg_{region}\left(w_i \,.\, \frac{|\text{median}_{\text{month}}(M_i - O_i)|}{2\ \text{median}_{\text{month}}(O_i)} \right).$$

The factor of 2 was included to give the score an intuitive character: a score of 50% would then mean that the model error is of the size of the observations.

Table 12.3: Estimate of the quality of BASCOE ozone analyses on the combined bias and standard deviation against HALOE and POAM-III. A positive bias indicates that BASCOE globally overestimates HALOE and POAM-III. If not specified, units are in percent (biases and uncertainties for BASCOE, taken from Errera et al., 2008).

	Altitude	**Poles** poleward of 60ºN and 60ºS	**Mid Latitudes** 30ºN-60ºN and 60ºS-30ºS	**Tropics** 30ºS-30ºN	**2003 ozone hole** poleward of 60ºS
BASCOE	0.5 hPa	-15 ± 15	-7 ± 10	-10 ± 7	/
	10 hPa	+3 ± 7	+3 ± 5	+2 ± 5	/
	70 hPa	+11 ± 12	+10 ± 13	+4 ± 10	$-3 \pm 3 \times 10^{11}$ molec/cm^3
	Tropopause 100 hPa at the Tropics, 200 hPa at the Extra-Tropics	+30 ± 30	+40 ± 80	+45 ± 70	/

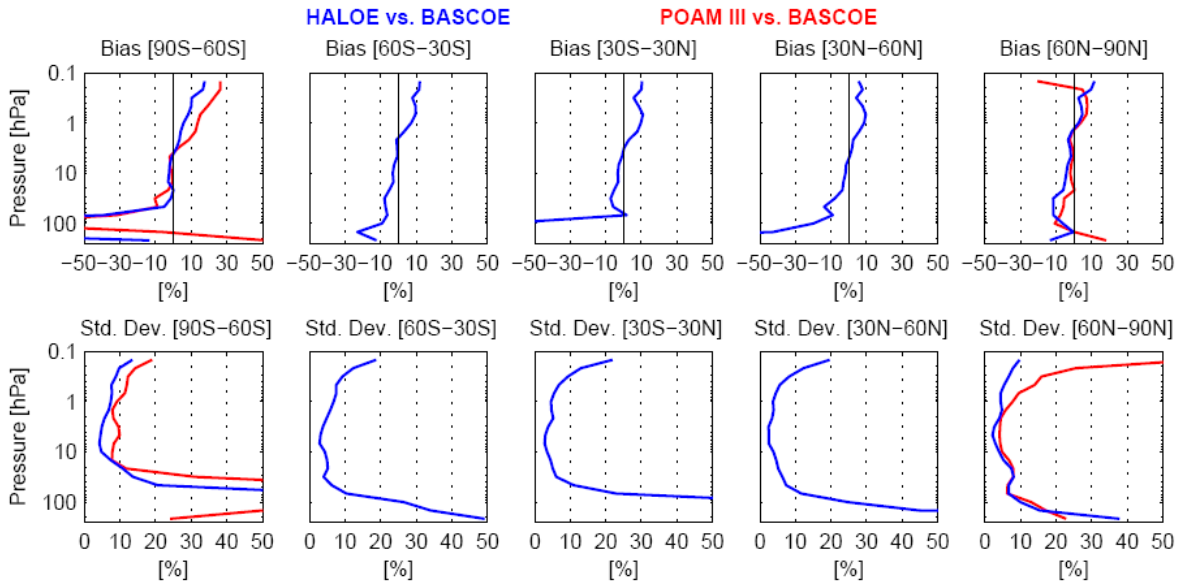


Fig 12.1: Comparison between BASCOE O_3 analyses and independent observations of HALOE and POAM-III for the period September-October 2003. Blue: HALOE vs. BASCOE, red: POAM-III vs. BASCOE. A positive value for the bias indicates that BASCOE underestimates the HALOE and POAM-III data. (Fig. 5 of Errera et al., 2008)

12.2 Offline simulations

	run	version	available period
stand-alone CTM	MOZART	mozart3.1/run_era1_v7 mozart3.1/run_era1_v10	20030101-20031231 20030101-20031231
	TM5	v3	20021101-20031231
	MOCAGE	simu2003/60lev01 simu2003/60lev02	20030101-20031231 20021101-20031231

12.2.1 Comparison plots and quantitative measures

Total columns

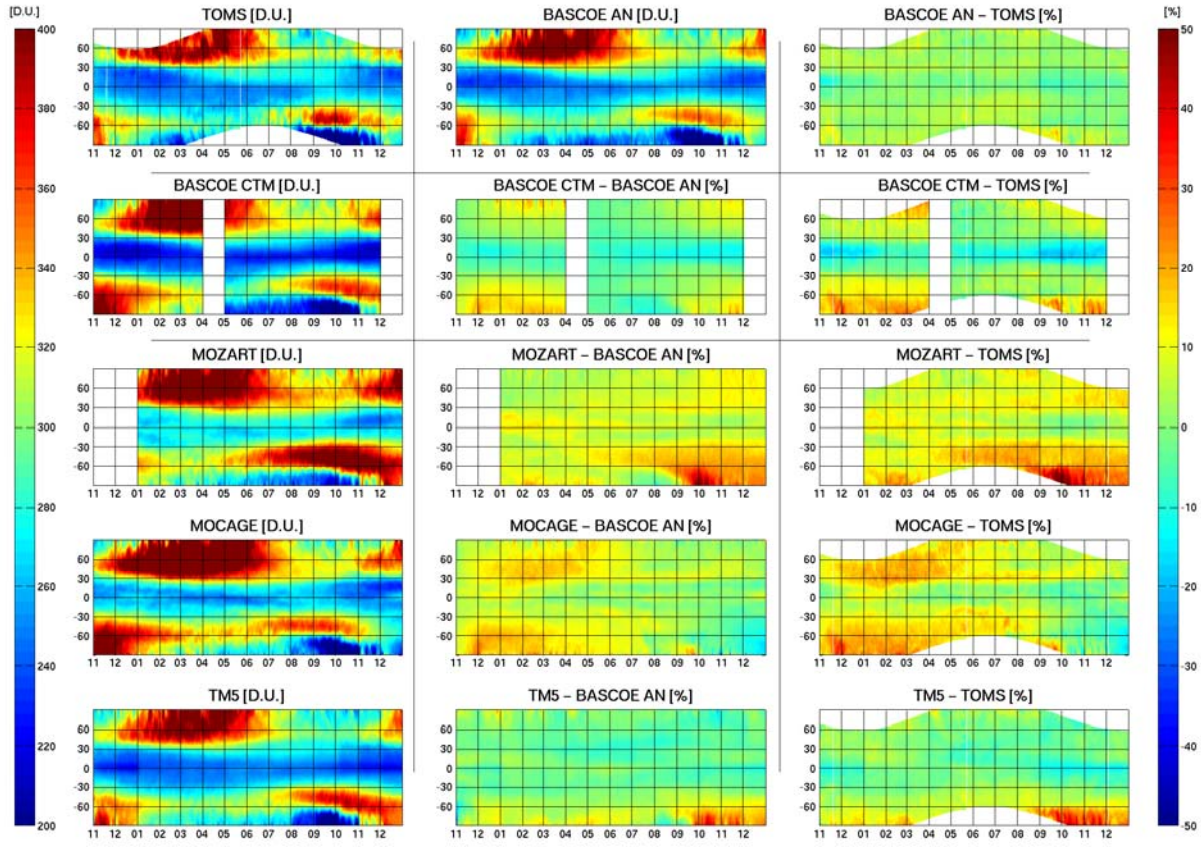


Figure 12.2: Evolution of the total ozone columns (expressed in Dobson Units) for TOMS, BASCOE CTM (free model run, without assimilation), BASCOE AN (resulting from the MIPAS offline assimilation), and the offline CTMs (MOZART-v10, MOCAGE-v2 , TM5-v3), with their respective normalized mean biases (expressed in %), for the period November 2002-December 2003.

Ozone zonal means of selected levels

1) Pressure levels: ±1hPa (lev5), ±10hPa (lev14), ±54hPa (lev22), ±100hPa (lev25)

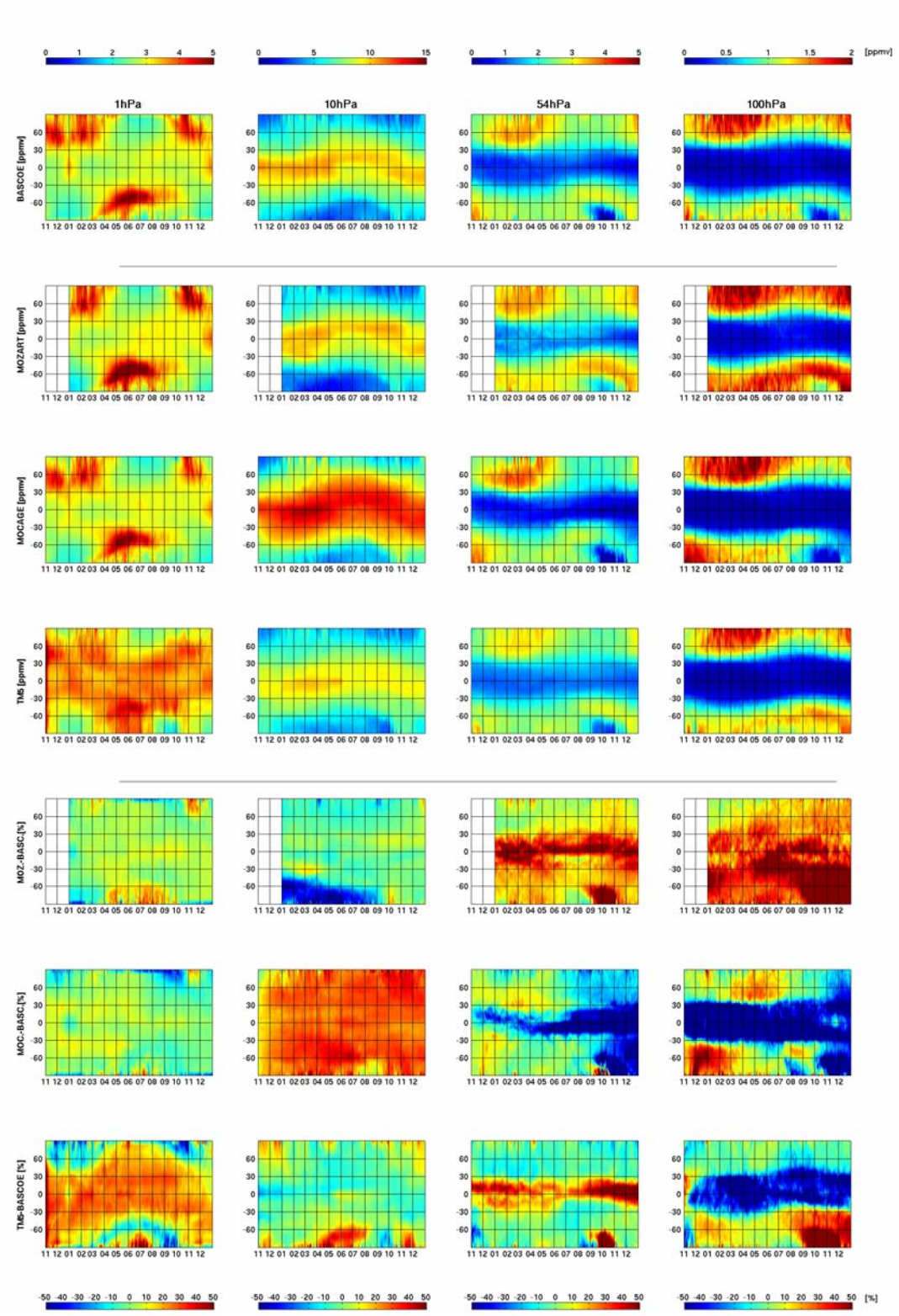


Figure 12.3: Evolution of ozone (expressed in ppmv) at selected pressure levels (from left to right: ± 1, 10, 54, 100 hPa) for (from top to bottom) BASCOE (AN, resulting from the MIPAS offline assimilation), the three offline CTMs (MOZART-v10, MOCAGE-v2 , TM5-v3) and their respective normalized mean biases (expressed in %), for the period November 2002-December 2003. Mind the different scaling axes of the ozone content at each level (top color bars).

2) Isentropic level 475K

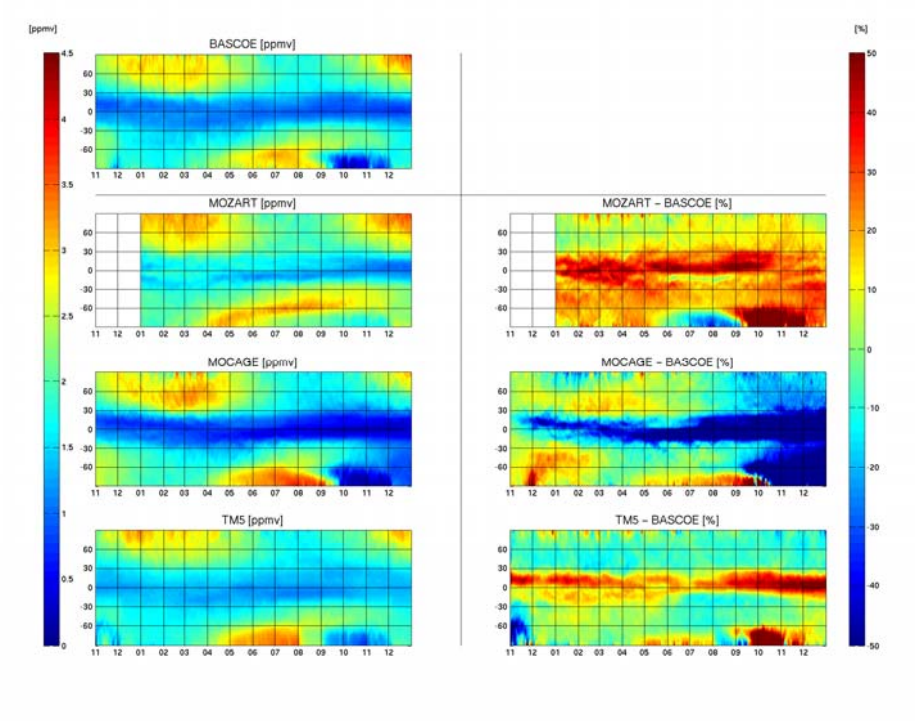


Figure 12.4: Zonal mean evolution of ozone (expressed in ppmv) at the 475K isentropic level for BASCOE (AN, resulting from the MIPAS offline assimilation), the three offline CTMs (MOZART-v10, MOCAGE-v2 , TM5-v3), and their respective normalized mean biases (expressed in %), for the period November 2002-December 2003.

<u>Ozone profile evolution (5d averages): comparison to independent satellite data (POAM)</u>

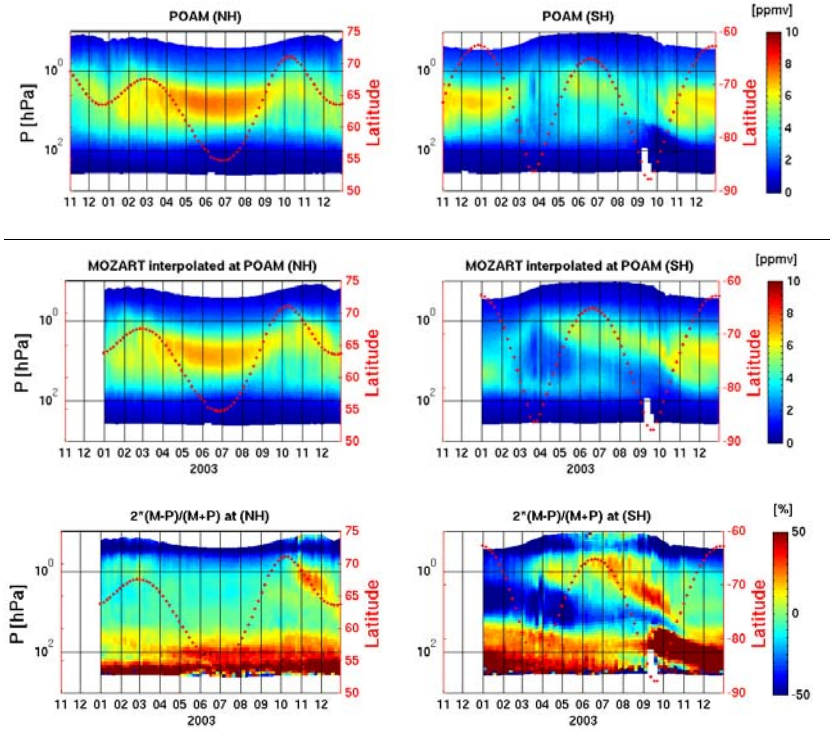


Figure 12.5: Comparison between the POAM satellite data and the co-located MOZART-v10 offline run profiles and the corresponding normalized mean biases, for the period November 2002-December 2003.

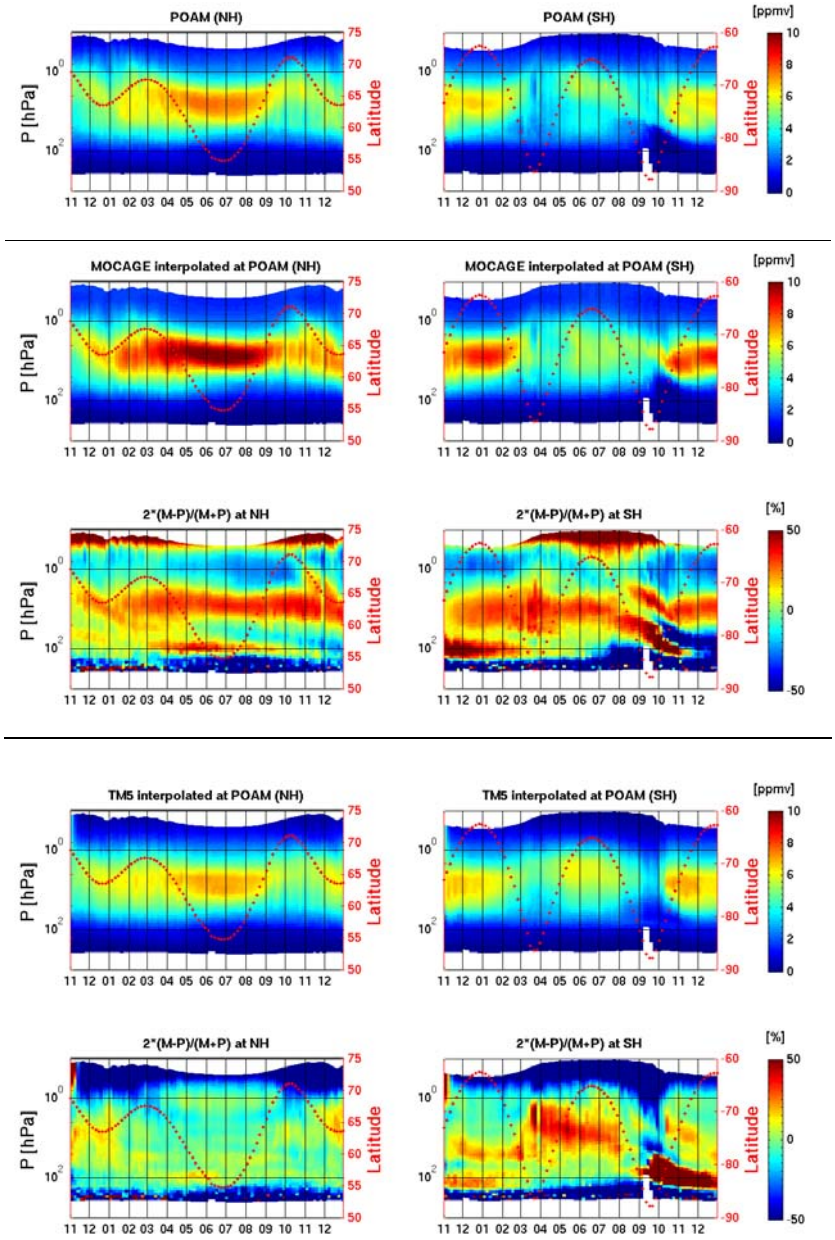


Figure 12.6: Comparison between the POAM satellite data and the co-located MOCAGE-v2 and TM5-v3 offline profiles, with the respective normalized mean biases, for the period November 2002-December 2003.

12.2.2 Discussion on the model performance

MOZART-v10

MOZART-v10 total ozone columns overestimate BASCOE and TOMS specifically in the southern hemisphere from May on. The discrepancies are generally not larger than 25%, except for the South Pole (SP) region during the main ozone hole period (from mid September till mid October), where discrepancies up to 50% and higher are found. From Fig 12.3, it is clear that this is mainly due to the serious overestimation in the lower levels, which points towards a problem in the transport. Before September 2003, MOZART-v10 performs very well in the SP region as far as the total columns are concerned (Fig. 12.2). This is the result of compensation between a serious underestimation (up to 50%) of ozone in the higher levels (around 10hPa) and an overestimation of ozone in the lower levels (mainly close to 100hPa), as can be seen in Fig. 12.3. Besides the mentioned underestimation, MOZART-v10 performs very well in the higher levels.

Comparison with POAM profiles (Fig. 12.5) shows that MOZART-v10 performs very well in the Northern Hemisphere (NH) and that the vertical distribution is almost perfectly represented. Fig. 12.5 shows a general overestimation of ozone in the Antarctic vortex. The bias is not limited to the lower stratosphere (ozone depletion due to heterogeneous chemistry) but extends to the whole stratosphere. Again, this points to a problem in the modeling of transport processes. During the first part of the year, we observe a general underestimation in the middle stratosphere. The performance is improving from May onwards, with only one downward trend of overestimation during the ozone hole period. Again, serious problems are mainly situated in the lower stratosphere from September on, where biases are always larger than 50%.

The Antarctic ozone hole is clearly present (Fig 12.2) but it ends too soon and is still underestimated in depth from mid September on. From the ozone levels at 54 hPa, 100 hPa and 475K, it is clear that MOZART-v10 exhibits ozone depletion during the ozone hole period, but at the wrong altitude. It is also not strong enough and ends too early (see Fig. 12.3 and 12.4, months 6-12).

The comparison with the POAM profiles confirms this (see Fig. 12.5). The ozone hole (blue area in the SH, between 10 and 100 hPa, from August until November) is reproduced in the MOZART-v10 run, but is not deep enough, which results in the dark red area in the bias plot. This area of positive bias slowly sinks to lower levels as ozone is transported downwards.

MOCAGE-v2

MOCAGE-V2 shows an overestimation of up to 20% in ozone column both at northern and southern midlatitudes and at the polar regions (mainly SP) until July 2003. At first sight, it seems like the simulation provides better columns afterwards (Fig. 12.2) and the Antarctic ozone hole is even nicely reproduced in the total column output. However, taking a closer look at ozone at the different levels (Fig. 12.3) reveals that the overestimation of the 10 hPa ozone maximum (by ~3 ppmv) lasts all year long but after July is compensated by an underestimation in the lower levels (at 54 and 100 hPa). To summarize, it seems that the model has a problem with photochemistry (~10hPa) all year long as well as with dynamics/transport (54-100hPa) in the second half of the year.

The comparison with the POAM profiles (Fig 12.6) confirms these findings. It additionally shows that the ozone hole is shifted in time, i.e. its onset and ending periods are both delayed by about one month resulting in the red and blue part of the SH-plot between the months August till December (Fig. 12.6).

TM5-v3

The climatology-based TM5-v3 ozone columns are realistic everywhere except near the South Pole region, where the ozone column is overestimated by ~10% most of the year and by more than 25% during the Spring season (ozone depletion). At 1hPa, TM5 displays a seasonal variation of its bias

w.r.t. BASCOE analyses, that is the overestimation extends polewards in summertime. Note though that the contribution of this level to the total ozone column is so small that this bias is not detectable in the total column output. At the 10hPa level, TM5 generally performs quite well. At the lower levels, we observe a large bias in the Tropics, but with completely opposite signs: an overestimation of more than 50% at the 54hPa level and an underestimation of about the same amount at the 100hPa level. As for the post-July MOCAGE simulation, TM5 appears to simulate well the ozone column but with a wrong vertical distribution. Comparison with POAM profiles shows that the early ozone hole period (until mid-September) is represented correctly but ozone depletion ends too early, resulting in an overestimation of 50% during October and November. One should keep in mind that the 2003 ozone hole event was exceptionally intense and therefore difficult to capture by a climatology.

Since we are looking at stand-alone CTM runs, we have included in the total column comparison (Fig. 12.2) the output of an unconstrained run of the BASCOE CTM (v4q23). It delivers no bias except for an underestimation of ~15% in the Tropics and an overestimation of ~20% during South Pole summer.

In conclusion, a simple comparison of ozone columns designates TM5 as the best-performing model. But upon closer inspection at different pressure levels, it appears that all CTMs suffer from some problem(s) with transport. Since the differences w.r.t. analyses look very different for each model, the problem is probably different in each case, i.e. not due to incorrect dynamical fields. Finally, it appears that TM5 has trouble simulating ozone in the upper stratosphere (1hPa) and MOCAGE in the mid-stratosphere (10hPa). Since ozone in these regions is controlled by photochemistry, this points to possible deficiencies in the corresponding modules of these CTMs.

12.2.3 Evolution of model performance during the project

MOZART v3 - v7 – v10

From the evaluation of the stand-alone MOZART simulation v7 it turned out that the Antarctic ozone hole 2003 was almost missing in this simulation and other simulations from v3 on (v7 being even worse than v3). This could be traced back to a severe underestimation of most halogen source gases which has been introduced during the stratospheric initialisation process. A corrected stand-alone MOZART simulation (v10) has been produced for the years 2003 and 2004 and is now the basis simulation for evaluation. As can be seen from Figure 12.7, MOZART-v10 has improved a lot compared to MOZART-v7. MOZART-v3 and v7 perform poor mainly south of 30°S and in particular at the SP region during the ozone hole period. The ozone concentrations in the tropical regions are very well reproduced in all runs. MOZART-v10 seems to perform overall a lot better than the previous versions. The improvement is especially striking in the South Pole regions, but it is also clearly noticeable in the northern hemisphere. MOZART-v10 now clearly shows the Antarctic ozone depletion during the ozone hole season, although this is still underestimated by 50 to 100 DU which is in line with the results from Kinnison et al. (2007) and, in addition, it ends too soon.

MOCAGE 60lev01 - 60lev02

When compared to BASCOE, the MOCAGE-v1 run seems to perform better than the later 60lev02 run (Fig. 12.8). The same is true from the point of view of the vertical distribution (figures not included). Especially in the first months of the case study, ozone is overestimated by some 15% in the MOCAGE-V2 run, both in the northern and southern midlatitudes and South Pole region. This overestimation is present at all levels. MOCAGE-v1 performed well at these regions. Only in the northern midlatitudes an overestimation of not more than 10% was detected.

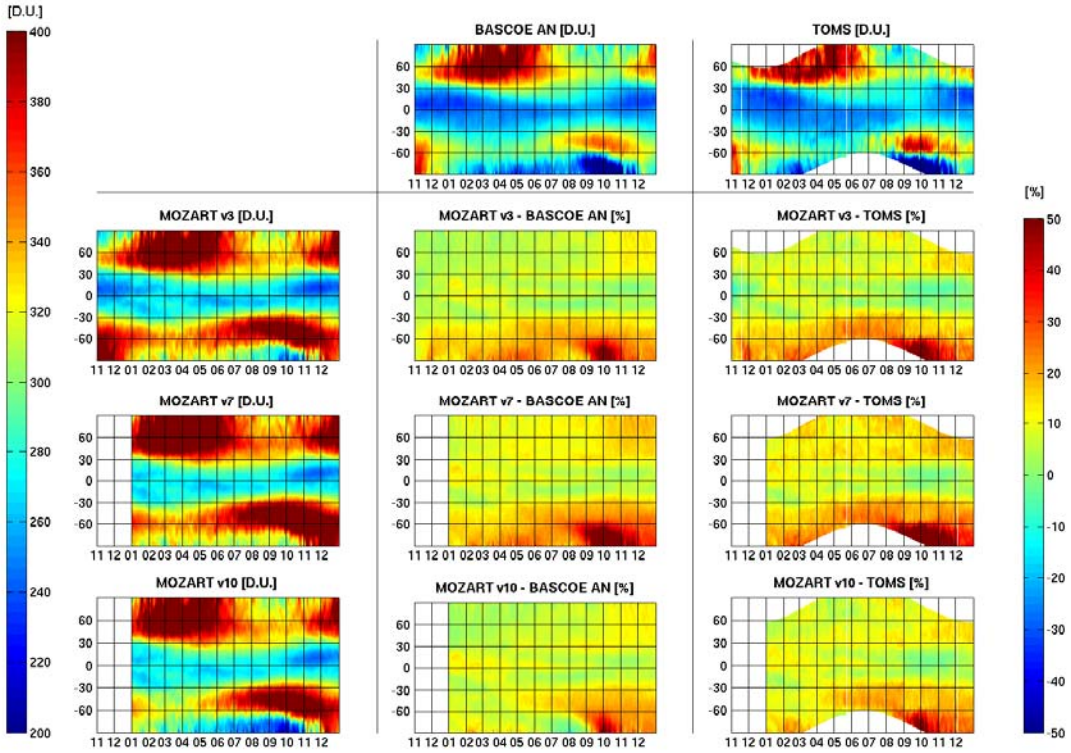


Fig 12.7: Comparison between the total ozone columns of the different runs of MOZART: v3, v7 and v10 for November 2002-December 2003. Biases w.r.t. BASCOE AN are shown.

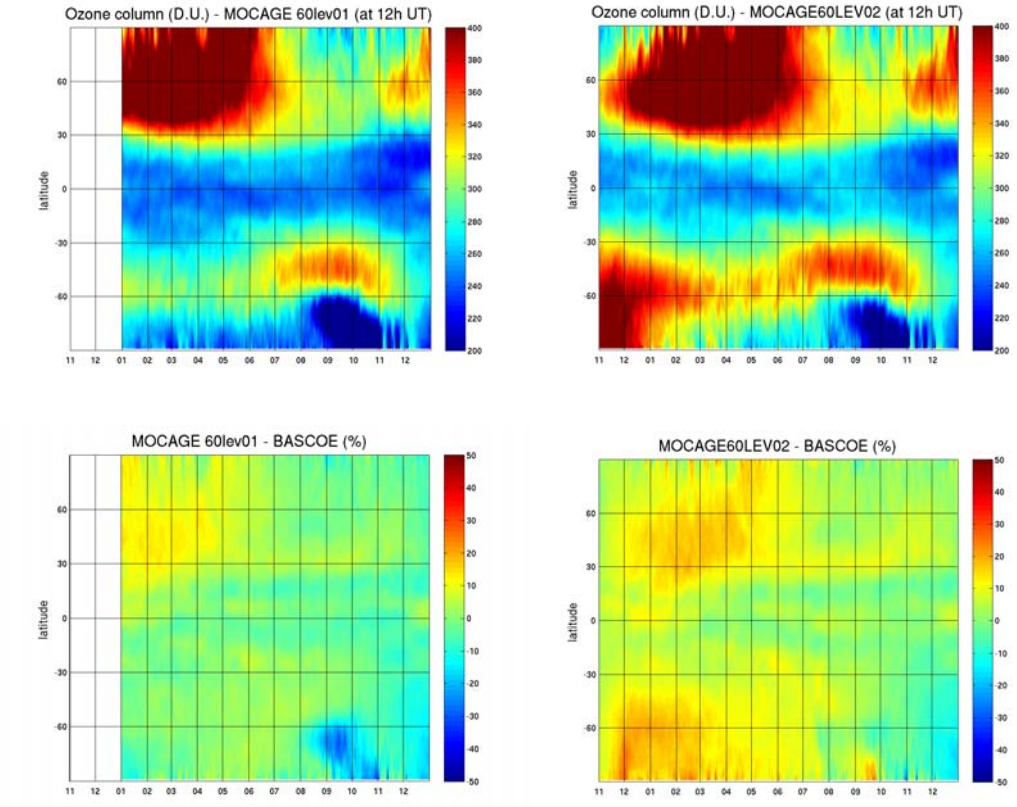


Fig 12.8: Comparison between the total ozone columns of the two most recent runs of MOCAGE: 60lev01 and 60lev02. Bottom: biases with BASCOE AN are shown.

12.3 GEMS-GRG reanalysis

run		version	available period
1st REA	**first GEMS reanalysis with assimilation** coupled system IFS+MOZART	exqx	20030101-20030430
		eyih	20030501-20031231
2nd REA	**Fully integrated second reanalysis (GRG+AER+GHG)** Active assimilation of O_3, CO, CO_2, CH_4 and aerosol Passive monitoring of NO_2 and HCHO	f026	20030101-20041231 (assimilated only SBUV in the period of f171)
		f171	20040101-20040501

12.3.1 Comparison plots and quantitative measures

Total columns

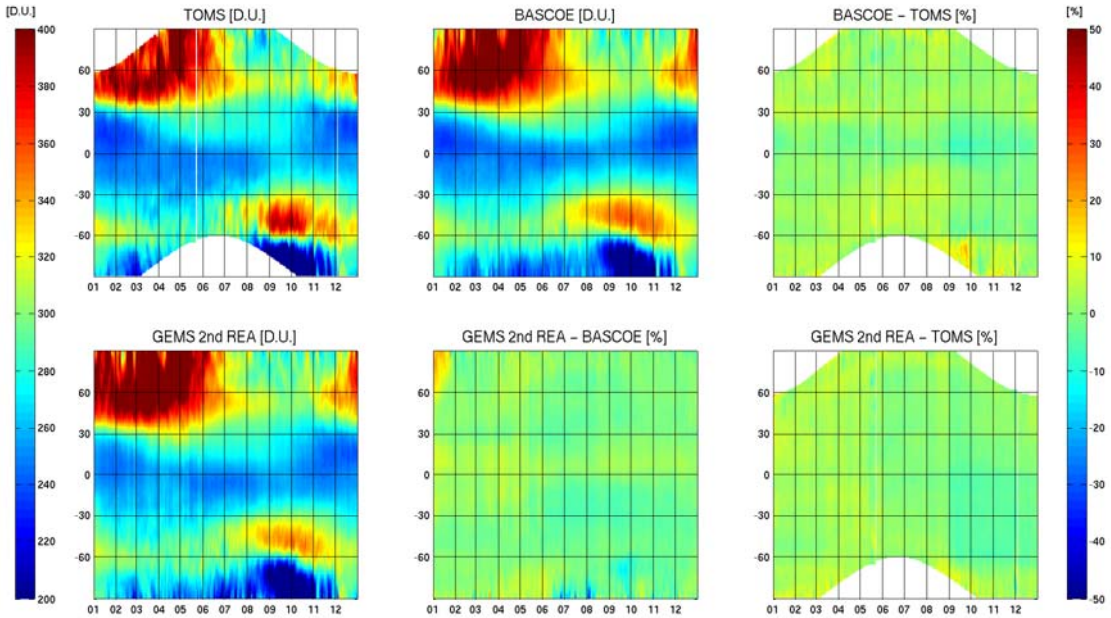


Figure 12.9: Evolution of the total ozone columns for TOMS, BASCOE (AN, resulting from the MIPAS offline assimilation), and the integrated GEMS reanalysis (2nd REA - f026) with the respective normalized mean biases for case study year 2003.

Ozone zonal means of selected levels

1. ***Pressure levels: ±1hPa (lev5), ±10hPa (lev14), ±54hPa (lev22), ±100hPa (lev25)***

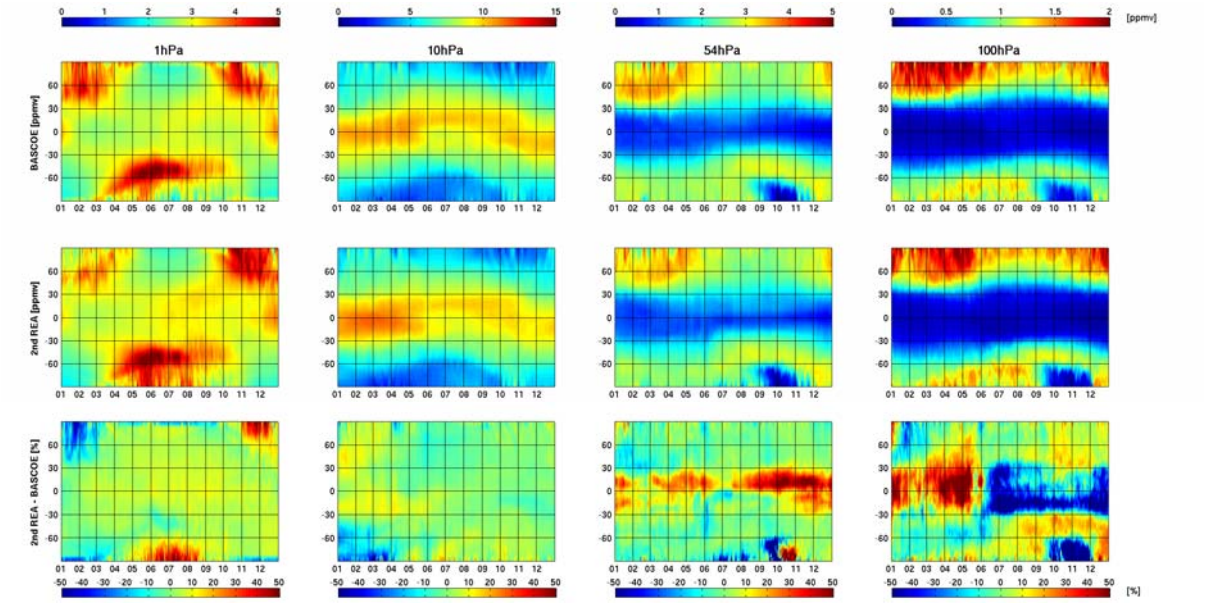


Figure 12.10: Evolution of ozone at selected pressure levels (from left to right: ± 1, 10, 54, 100hPa) for (from top to bottom) BASCOE AN, the integrated GEMS reanalysis (2nd REA i.e. f026) and the normalized mean biases for the case study year 2003. Mind the different scaling axes of the ozone content at each level (top color bars).

2. ***Isentropic level 475K***

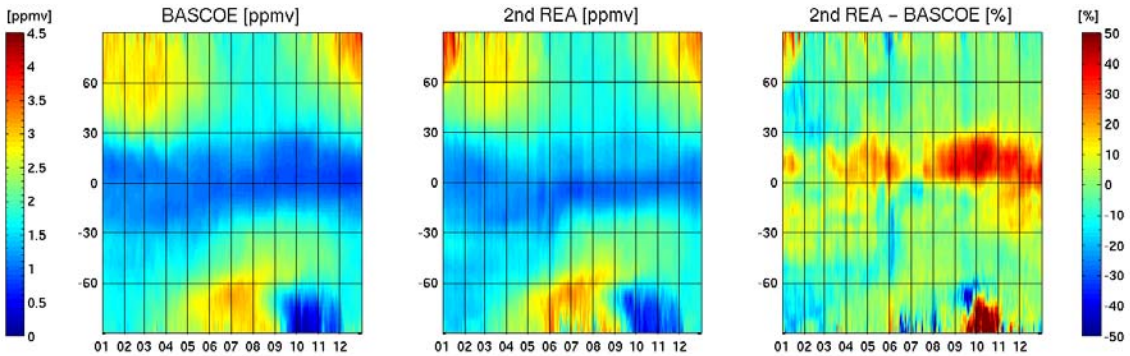


Figure 12.11: Evolution of ozone at the 475K isentropic level for BASCOE AN, the integrated GEMS reanalysis (f026), and the normalized mean biases for 2003.

Ozone profile evolution (5d averages): comparison to independent satellite data (POAM)

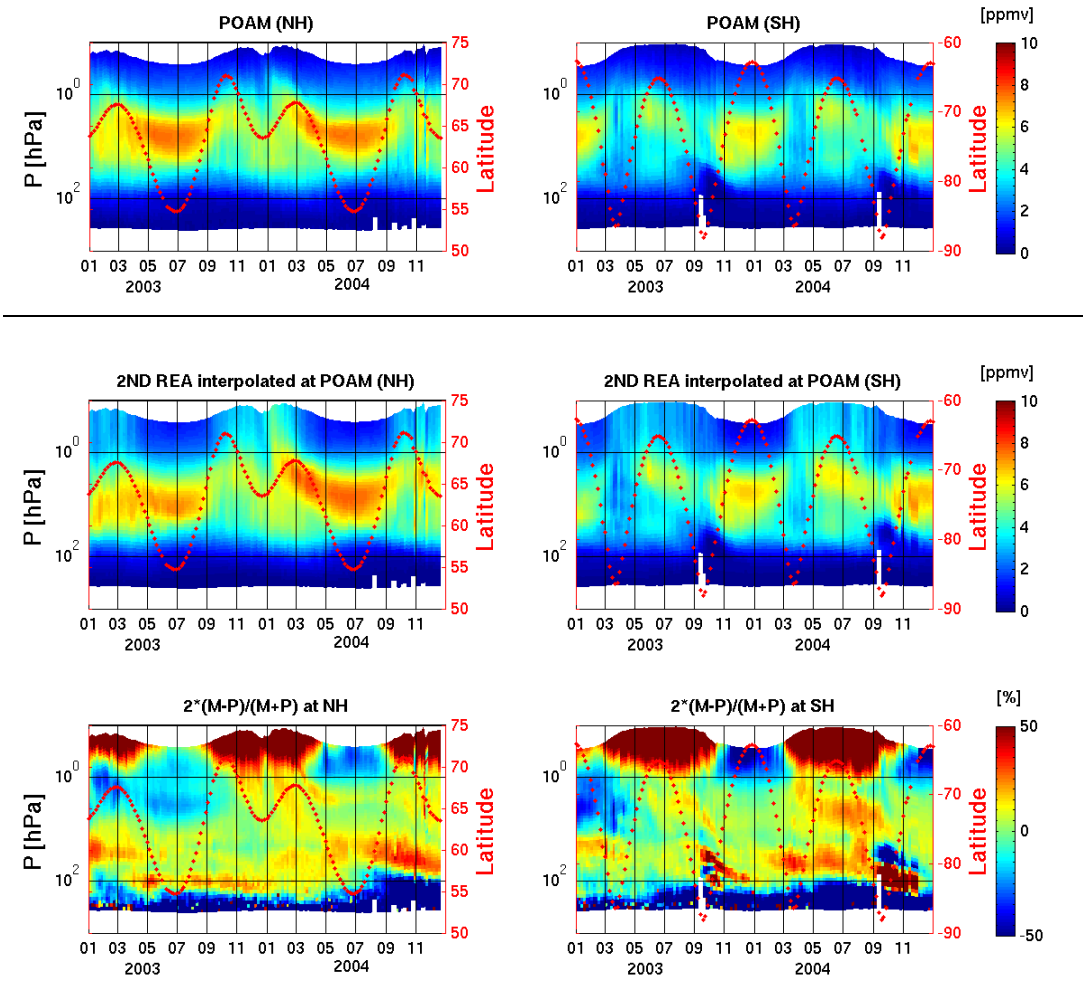


Figure 12.12: Comparison of the evolution of the total ozone column between POAM satellite profiles and co-located integrated GEMS reanalyses (f026) for the northern (left, NH) and southern (right, SH) hemisphere during the years 2003 and 2004, and the normalized mean biases. The red dotted profiles represent the latitudinal trajectory of the satellite.

12.3.2 Discussion on the model performance

Integrated GEMS reanalysis: f026

In Fig. 12.9, we compare the integrated (GRG+GHG+AER) GEMS reanalysis (f026, improved model) with the analyses of BASCOE assimilation of MIPAS offline data (v4.61-2). The reanalysis started from too high total ozone columns at high northern latitudes (20% too high) on January 1, 2003. In the course of weeks the influence of the data assimilation became visible and before the end of January the reanalysis already provided overall realistic ozone column fields. This was confirmed by comparison with the independent TOMS total column data.

From Fig. 12.10, it is clear that, even though the total ozone columns compare well, the vertical distribution of ozone is not always well-represented.

At 1 hPa, there is a large underestimation of more than 50% for the high northern latitudes (> 40ºN) from January till the end of February 2003. As BASCOE itself generally underestimates satellite data by some 10% at this pressure level (see Fig. 12.1), the underestimation of the integrated REA is worse

than that presented on Fig. 12.10. A large positive bias is observed at the same high northern latitudes during the months November and December 2003 and at high southern latitudes (< -60ºS) from May till August 2003. These biases coincide with the northern vortex in NH winter and the southern vortex in SH winter, and is therefore probably due to dynamical properties related to the vortex. The results for the 10hPa level are overall very satisfying. Only at the most southern latitudes (< -70ºS) an underestimation of ozone is observed for the first four months of the assimilation. The bias is not larger than 10%. Ozone is seriously overestimated at the 54 hPa level in the region just north of the Tropics (0-30ºN), and at the most southern latitudes (< -70ºS) during the whole month of October. During the last part of September 2003, we observe exactly the opposite. At the 100 hPa level, we clearly distinguish two periods, the transit being at the beginning of June. This is probably due to the fact that this run assimilates GOME profile data from January till May 2003, but after May there are no more GOME data available. During the first five months of 2003 ozone is overestimated at the Tropics, extending even towards higher latitudes from March till the end of May by 40% and higher. Since at this pressure level BASCOE itself generally overestimates ozone (see Fig. 12.1), the overestimation of the 2nd REA is even worse than shown on Fig. 12.10. From June onwards, the opposite happens and a clear underestimation of ozone at the Tropics is observed. The observed negative bias is probably due to the overestimation of BASCOE in the lower stratosphere (see Errera et al., 2008), rather than to an underestimation of the GEMS reanalysis. Ozone at the isentropic level 475K (Fig 12.11) follows a similar behavior as at the closeby 54hPa level.

Even though we see that differences at the separate levels are sometimes quite large, they do seem to compensate well in the total column output. This could be due to the assimilation of total column observations, which are expected to scale the vertical profiles but preserve their shape.

Figure 12.9 additionally shows that the description of the 2003 Antarctic ozone hole is in very good agreement with compared to the BASCOE analyses. TOMS data are missing to correctly judge the onset of the ozone hole. However, the BASCOE analyses have been independently evaluated using POAM-III and HALOE observations, so we can trust the good quality of the ozone hole extent and timing in the reanalyses.

Comparison of the integrated GEMS reanalysis with the POAM ozone profiles (Fig. 12.12) reveals a serious overestimation of mesospheric ozone during autumn and winter, both at the northern and the southern hemisphere. Stratospheric ozone, on the other hand, is very well reproduced, with biases generally not larger than 20%. The POAM profiles confirm that the ozone hole (and in particular the onset) is quite well reproduced in the reanalysis. The only shortcoming is in the vertical distribution of ozone depletion, which occurs a bit too low in the stratosphere in 2003 and a bit too high in 2004.

12.3.3 Evolution of the model performance during the project

First and second GEMS reanalysis

Comparison of the 2nd, integrated, reanalysis (f026) with the 1st reanalysis run (exqx + eyih), illustrates that the integrated version definitely performs much better. The improvement is hardly detectable in the total column output (Fig. 12.13), although the small biases detected in the total columns of the 1st reanalysis have become even smaller in the 2nd reanalysis run. The difference between both reanalysis runs becomes especially apparent in the vertical distribution (Fig. 12.14).

In the first GEMS reanalysis run (exqx + eyih), severe underestimations were detected in the 1hPa level during winter time at both hemispheres. The biases coincide with the northern vortex in NH winter and the southern vortex in SH winter, and is therefore probably due to dynamical properties related to the vortex. This was (over)corrected in the 2nd GEMS reanalysis, as the underestimations turned into overestimations, however, now over a much more concise region and for a shorter period. Underestimations detected at the 10hPa level in the southern latitudes have become less pronounced in the months January till the end of March and have disappeared completely from September till the end of 2003. The 1st reanalysis clearly displayed a bug at the 54 hPa level, with an abrupt discontinuity at the equator. This was corrected in the 2nd reanalysis, but at this level both runs still overestimate

Northern Tropical ozone by 30 to 50%. Where ozone is now quite well represented at the lowest levels (recall that the underestimation at the Tropics is due to the overestimation by BASCOE), it was not the case in the 1st GEMS reanalysis. Extreme positive biases were detected between -35ºS and 35ºN.

Comparison of the 1st GEMS reanalysis with the POAM ozone profiles (Fig. 12.15) reveals a serious underestimation of mesospheric ozone both at the northern and the southern hemisphere, all year long. The transit from the mesosphere to the stratosphere is quite abrupt. Between 1 and 10 hPa, the southern hemisphere displays an extended triangular-shaped area of low ozone between July and October. This is probably due to the too large injection of NO_x into the vortex, which is consequently transported downwards into the stratosphere. The same triangular-shaped area between July and October was observed in the old stand-alone MOZART run (run_era1_v3), which was used for the first coupled reanalysis run. It disappeared in later MOZART runs (e.g. v7 or v10, see Fig. 12.5) with improved upper boundary conditions, and consequently in the second integrated GEMS reanalysis run as well (see Fig. 12.12).

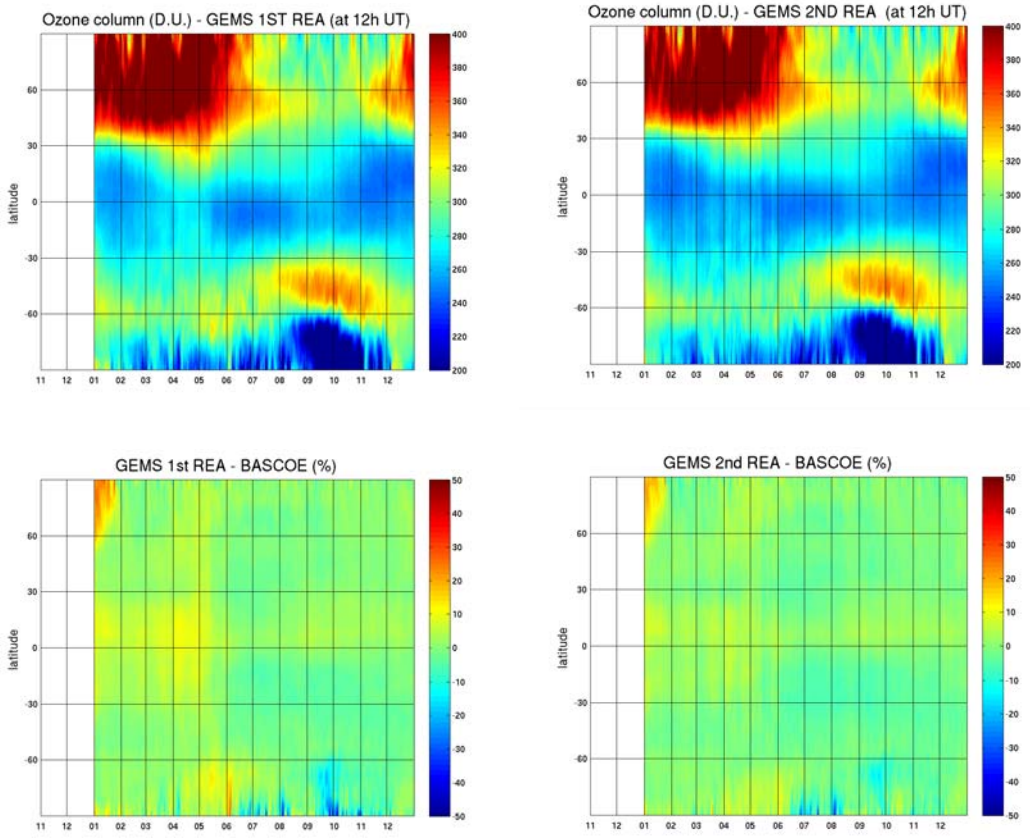


Figure 12.13: Top: Comparison between the total ozone columns of two GEMS reanalysis runs: exqx + eyih (the first reanalysis run) and f026 (the second reanalysis run) for the year 2003. Bottom: biases with BASCOE AN are shown.

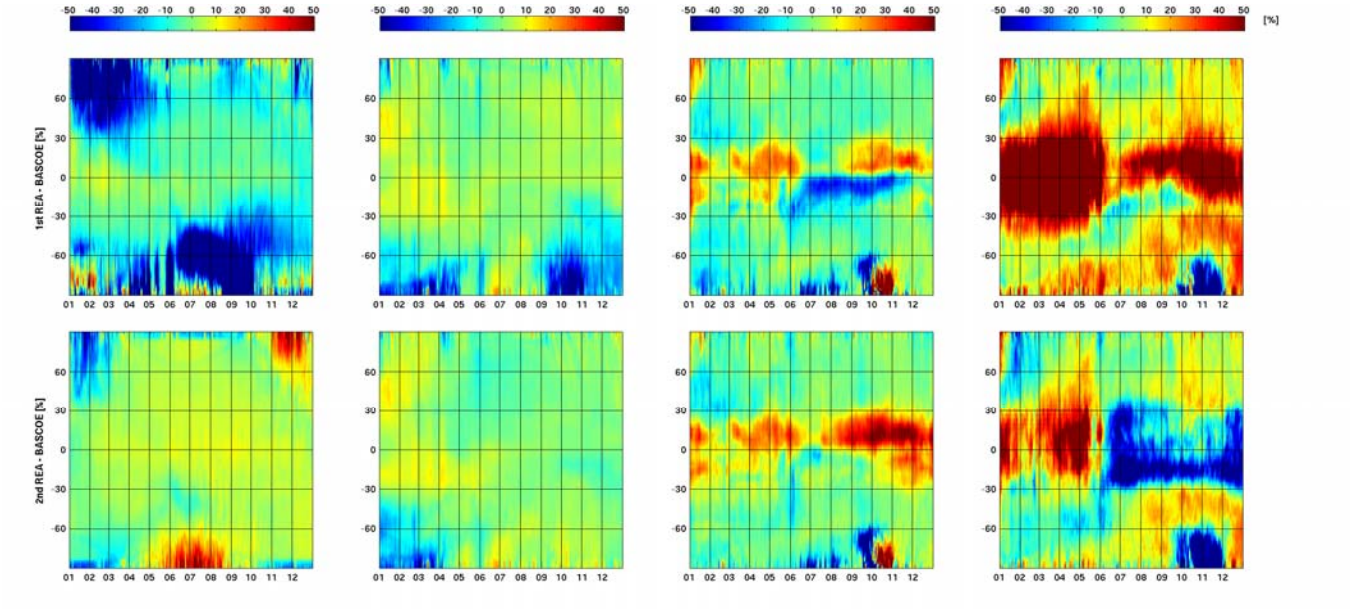


Figure 12.14: Evolution of the ozone bias of the first (exqx+eyih) and second (f026) GEMS reanalysis runs compared to BASCOE AN, at four selected pressure level (from left to right: ± 1, 10, 54, 100hPa) for the year 2003.

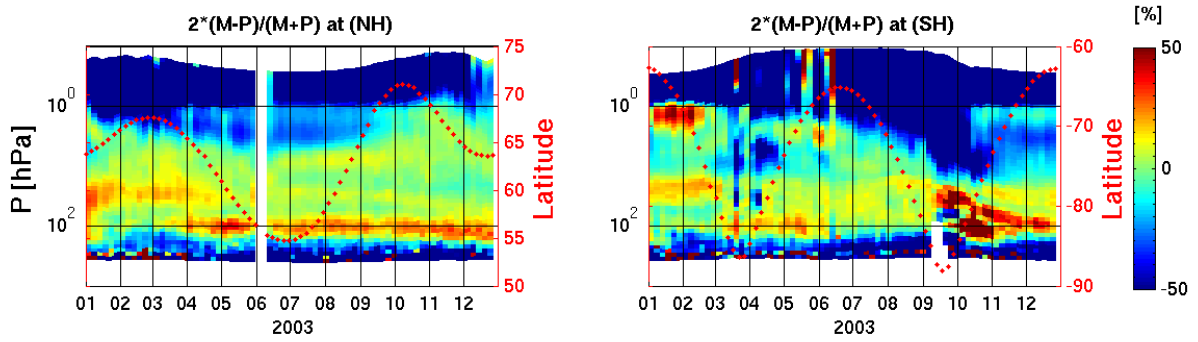


Figure 12.15: Normalized mean bias between POAM satellite profiles and the co-located first GEMS reanalysis profiles (exqx + eyih) for the year 2003. Left: NH, right: SH.

12.4 GEMS-GRG forecast

run		version	available period
FC	**GEMS forecast** coupled system IFS+MOZART	eywm	20021101-20031231 IFS: 20021101-20030228
		euas	20030228-20031231 (complementation of *eywm* IFS)

12.4.1 Comparison plots and quantitative measures

Total columns

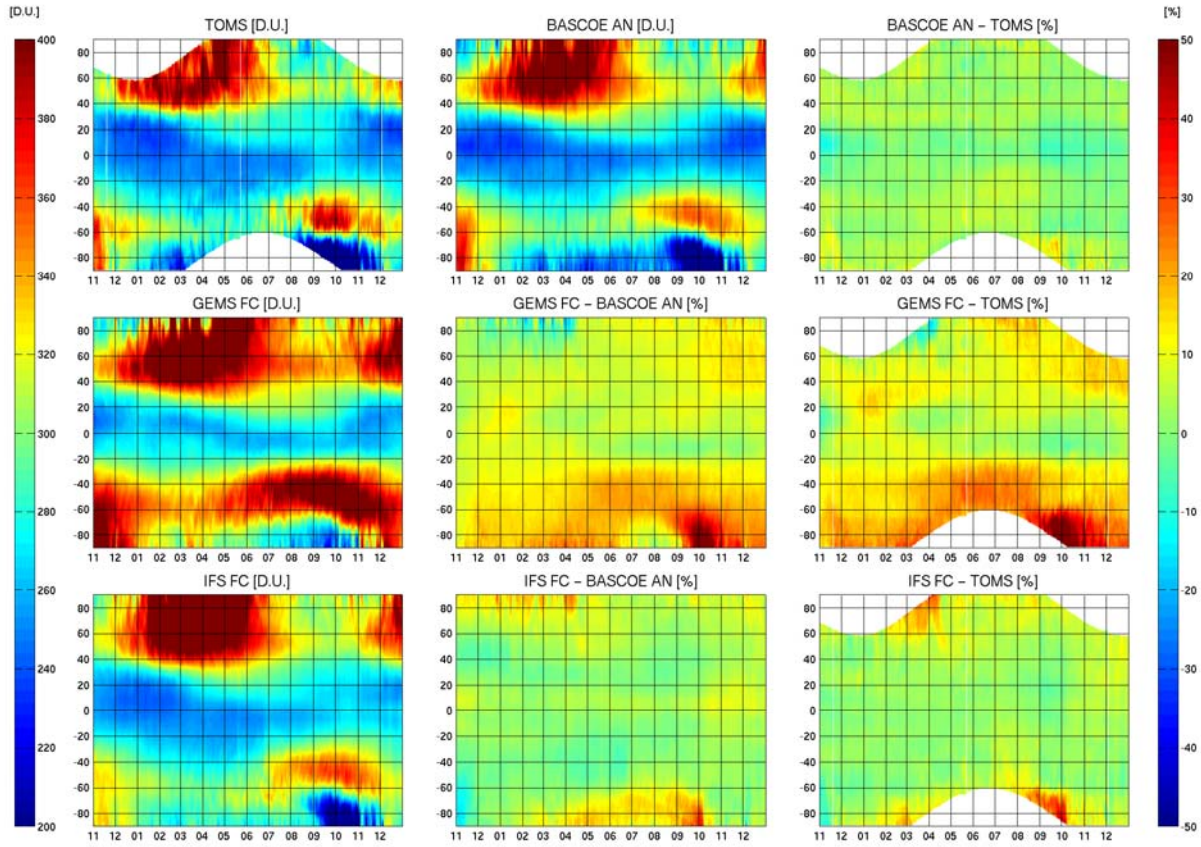


Figure 12.16: Zonal mean evolution of the total ozone column for TOMS, BASCOE AN, and the GEMS coupled system forecast (FC) as well as the IFS ozone forecast (using the Cariolle scheme), and the respective normalized mean biases, for the period November 2002-December 2003.

Ozone zonal means of selected levels

1. *Pressure levels: ±1hPa (lev5), ±10hPa (lev14), ±54hPa (lev22), ±100hPa (lev25)*

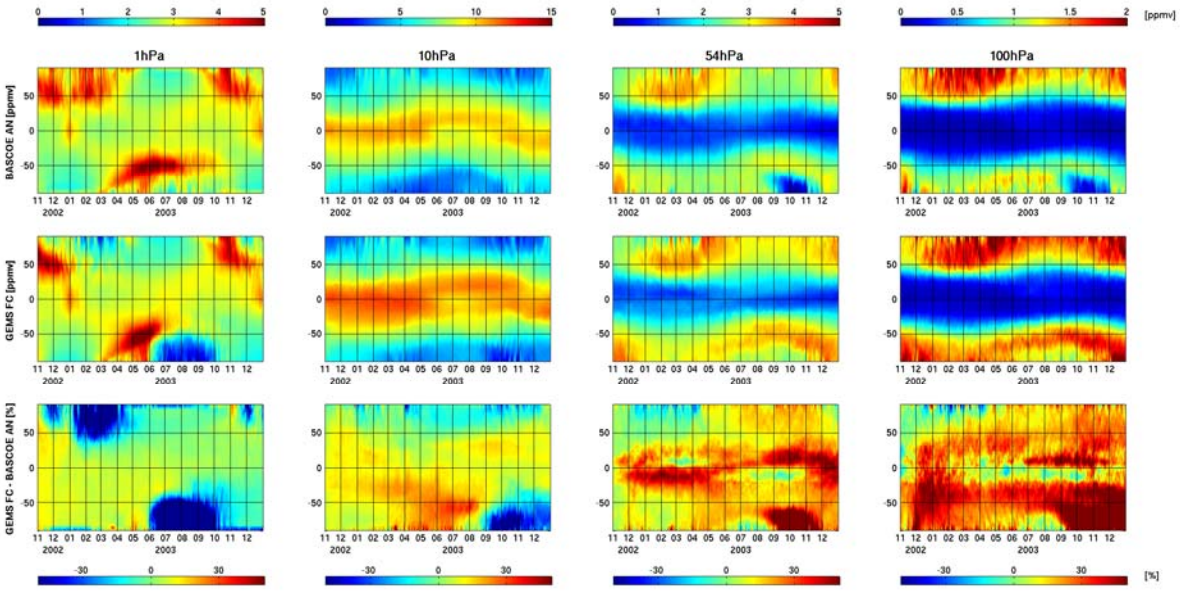


Figure 12.17: Evolution of the ozone at selected pressure levels (± 1, 10, 54, 100 hPa) for the GEMS coupled system forecast (FC) and BASCOE AN, and the respective normalized mean biases. Mind the different scaling axes of the ozone content at each level.

2. *Isentropic level 475 K*

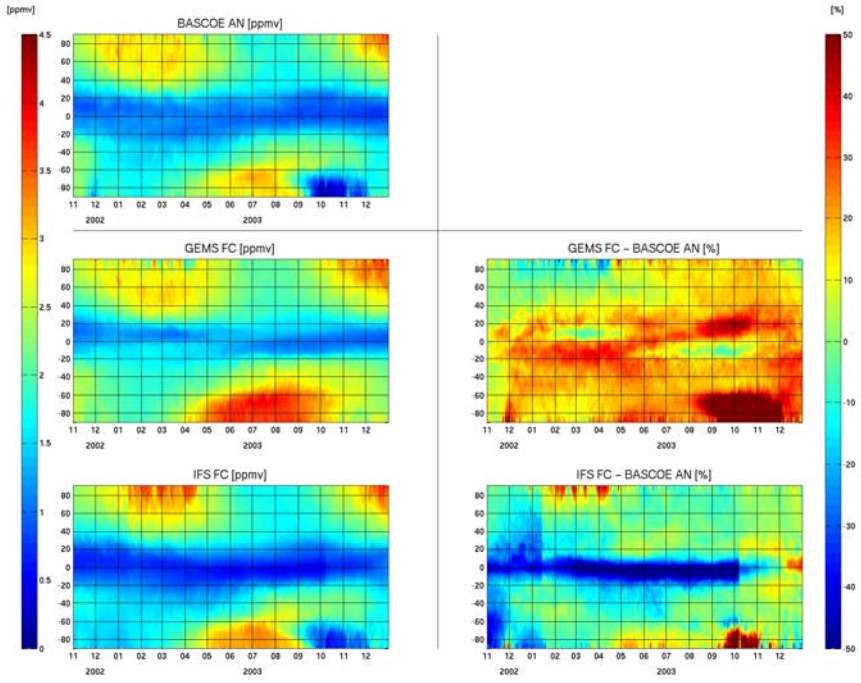


Figure 12.18: Evolution of ozone at the 475K isentropic level for the GEMS coupled system forecast (FC) and BASCOE AN, and the respective normalized mean biases, for the period November 2002-December 2003.

Ozone profile evolution (5d averages): comparison to independent satellite data (POAM)

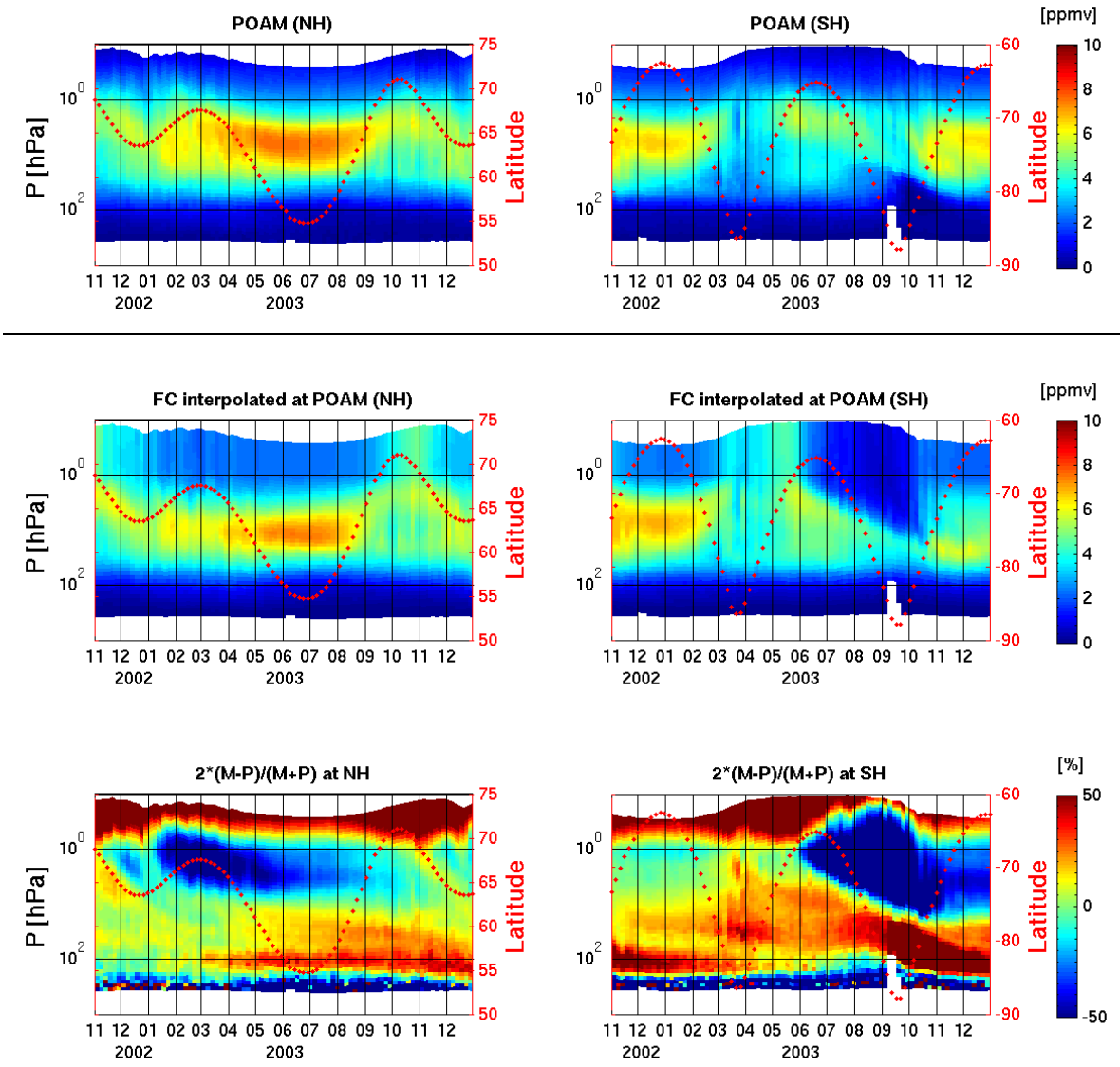


Figure 12.19: Comparison of the evolution of the total ozone column between POAM satellite profiles and the co-located forecast run for the northern (left) and southern (right) hemisphere, and their respective normalized mean biases for the period November 2002-December 2003. The red dotted profiles represent the latitudinal trajectory of the satellite.

12.4.2 Discussion on the model performance

GEMS forecast run: eywm

First of all, it is important to stress the fact that the BASCOE version we used to compare the forecast with is the reanalysis version of BASCOE and therefore we cannot start from BASCOE as the perfect representation of a forecast run. Moreover, for the coupled system forecasts, IFS is still coupled to the old MOZART-v1 run and fully follows this CTM run. It is restarted from the coupled run of the previous day. As the new MOZART-v10 run has improved a lot, also a new coupled system forecast (IFS+MOZART-v10) may represent a large improvement. We draw the attention to a systematic overestimation of total ozone of some 15 to 20% in the northern midlatitudes and of 20 to 30% in the southern hemisphere.

Even though the dynamical model is identical, the GEMS and IFS forecasts give very different results for both the total ozone columns and the ozone at the levels. This is not surprising as the IFS forecast, using the Cariolle scheme, is started every day from the operational analysis of IFS ozone, i.e. it already contains observations.

As the forecast run is coupled to the (old) MOZART-v1 CTM, the ozone profiles behave almost exactly the same and the POAM comparison consequently gives the same trends as for the old MOZART-v1/v3 run: triangular shaped low ozone between 0.5 and 10 hPa between mid June and the start of October due to problems in the mesosphere. During NH summer 2003, a lot of NO_x was created in the mesosphere and transported downwards in the stratospheric Antarctic vortex (Funk et al. 2005). Also ozone was far too high.

From our experience with BASCOE, the chemical model (MOZART-v1) seriously underestimates the ozone depletion. There is an underestimation of ozone depletion in the IFS forecast of 10 to 30% in the southern hemisphere and even a complete absence of ozone depletion in the ozone hole period in the GEMS forecast.

12.5 Summary

12.5.1 Comparison plots and quantitative measures

For the following summarizing figures, we have defined 5 latitude zones as follows:

- South Pole: $-90^\circ < lat < -60^\circ$
- Midlatitude South: $-60^\circ < lat < -30^\circ$
- Tropics: $-30^\circ < lat < 30^\circ$
- Midlatitude North: $30^\circ < lat < 60^\circ$
- North Pole: $60^\circ < lat < 90^\circ$

This choice is justified by several criteria. It respects the natural atmospheric circulation regimes and climate types, it maximizes the number of observations in each latitude bin and optimizes the coverage and representativeness.

In this way, we find the monthly scores (as defined in Section 12.1.2) for each model within each latitude band, which allows us to study how the model performance changes throughout the year and from latitude zone to latitude zone (Fig 12.20). Averaging the model score over the whole year allows for showing which model performs best in the separate latitude zones, and even globally (Fig 12.21).

12.5.2 Discussion on the model performance

The resulting scores never go below 85%. But rather than looking at the absolute score value, one should look at the trends and the relative behavior of the different models as the absolute score fully depends on the proportionality factor in the score's definition.

General conclusion: point of view of the separate models

MOCAGE:

MOCAGE-v1 scored very well during the period Jan'03-July'03, but performed bad during the ozone hole period afterwards. With different initial conditions, the quality during the first semester of 2003 lies a lot lower in the 60lev02 run, but is improving over time with a steeper increase from June 2003 onwards. The ozone hole in the total column output is now very well represented. Recall that this is in agreement with the results discussed in Section 12.2.2, where we have also shown that this statement is not valid for the vertical distribution. The long period of lower scores in the 60lev02 run gives more weight than the relatively short period of strong biases during the ozone hole period in the 60lev01 run, which explains why, on the average, MOCAGE-v1 performs better than the newer run MOCAGE-v2 .

MOZART:

MOZART-v7 is the model that performs worst. The quality is constantly decreasing between the start and the end of the year, and is even worse than this general trend between August and September. This is mainly due to the absence of an ozone hole due to the wrong initialization of halogen gases. This bug was corrected in MOZART-v10 and consequently MOZART-v10 scores a lot better. The scores are more or less constant during the first semester of the year, even still improving during July and August, which shows that the onset of the ozone hole is very well predicted. But then, the ozone hole is no longer deep enough and ends too early, which results in low scores for September, October and the following months.

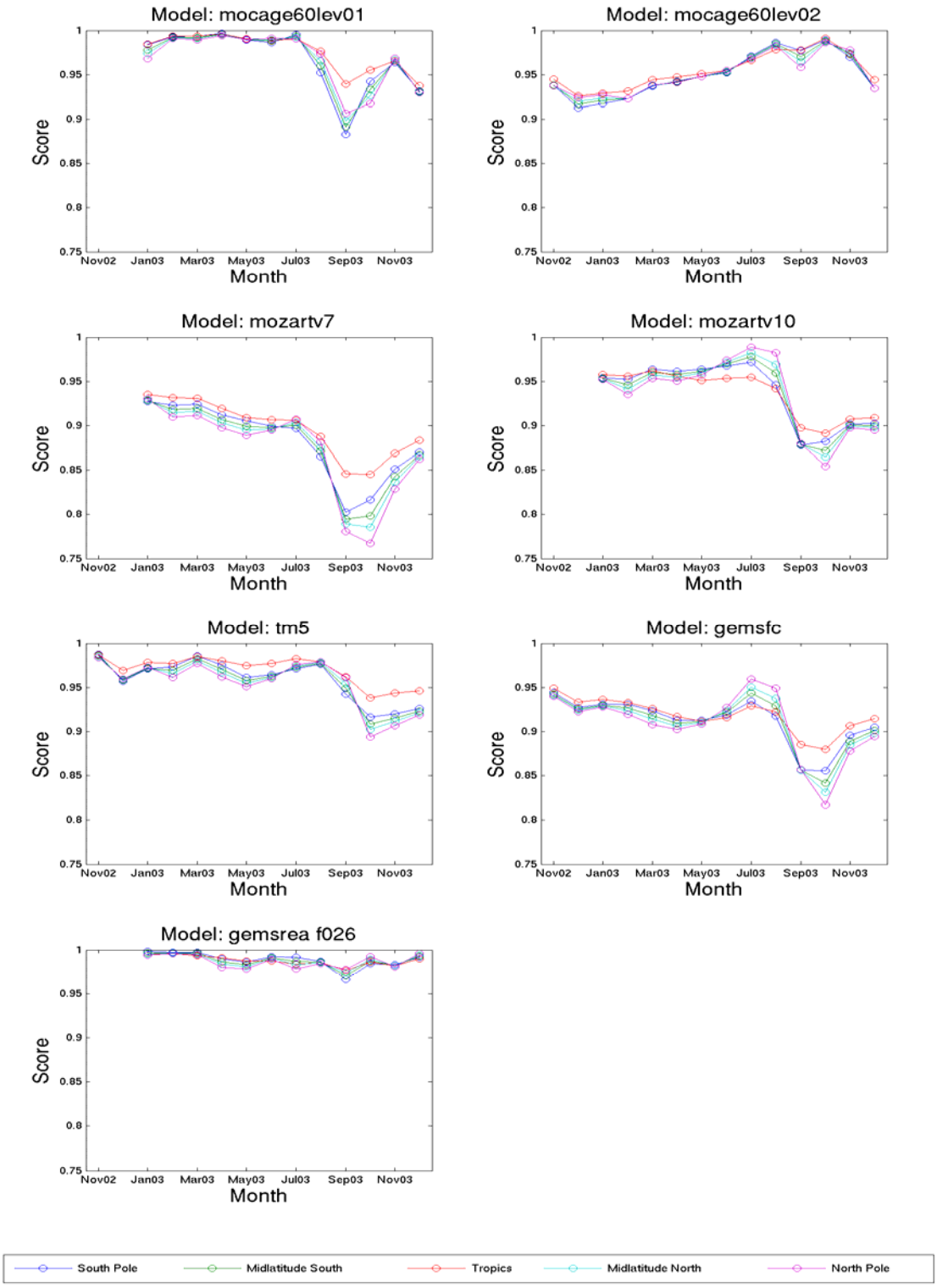


Figure 12.20: Monthly evolution of the model scores for the five latitude bands (blue- SP, green – Midlatitude South, red – Tropics, cyan – Midlatitude North and magenta – NP) for MOCAGE-v1 and 60lev02, for MOZART-v7 and v10, for TM5, GEMS FC (eywm) and GEMS integrated REA (f026).

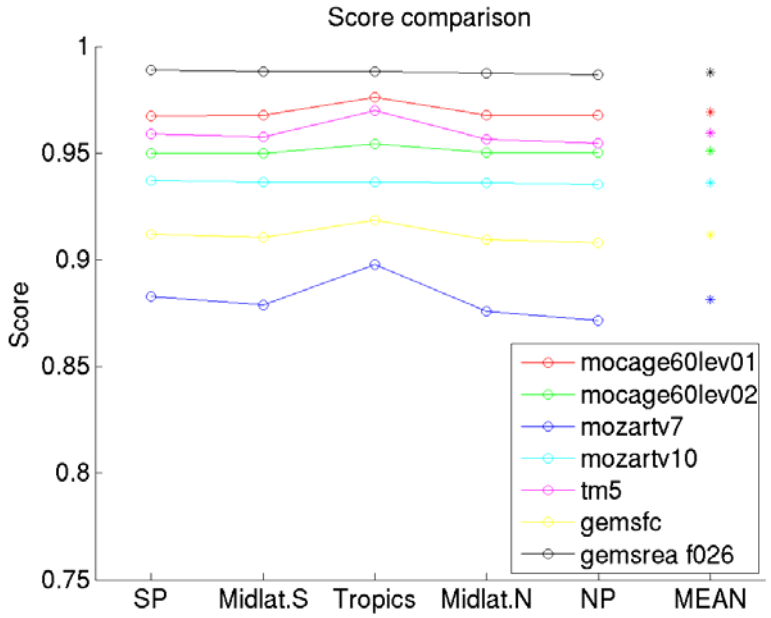


Figure 12.21: Global model performance (presented in scores) over the whole year in 5 different latitude zones (colors indicate different models). Asterisks give a global mean score for each model, which allows a quick intercomparison of model performance.

TM5:

TM5 performs overall very well and stable. Only the ozone depletion is underestimated during the months October till December 2003, which leads to low scores in this period for the SP.

GEMS REA:

The integrated GEMS assimilation run performs extremely well in all latitude zones over the whole year. Scores are generally higher than 98%.

GEMS FC:

The GEMS FC run (eywm) follows the chemistry of the old MOZART-v1 run and is therefore not really representative for the quality that may be obtained when it were coupled to the new MOZART-v10 run, which, in the meanwhile, has improved w.r.t. several aspects.

General conclusion: point of view of the different latitude zones

The performance of all models is more or less constant over the different latitude zones, except for the Tropics, where all models attain their best results, with scores generally higher than 90%.

The same is generally true for the highest levels (1-10hPa), except for TM5 and MOCAGE. Where TM5 overestimates ozone throughout the year with more than 20% at 1 hPa and performs much better at 10hPa, the opposite is true for MOCAGE. MOCAGE performs very well at the 1hPa level, but overestimates ozone at 10hPa, typically with some 30%. The situation for MOCAGE is even worse at the lower levels (54-100hPa), where the performance is good at the start of the run, but gradually decreases towards the end of 2003. There is a large underestimation of ozone between 35 and 60% in the lower stratosphere for MOCAGE.

On the whole, biases to BASCOE increase for all models towards lower levels.

Midlatitudes:

Total columns at the Midlatitude regions are generally slightly overestimated by all models (except for the GEMS REA), with a positive bias not larger than 25% for the FC run, and not larger than 15-20% for all other runs. The performance of MOCAGE even seems to improve in the course of the year. The evolution plots of ozone at the specified levels, however, contradict this. They illustrate how, from August 2003 on, ozone underestimations in one level (10 hPa) are nicely compensated by ozone overestimations in another level (54-100 hPa), and actually MOCAGE performs a lot worse during this period.

During ozone depletion season, the FC run seriously underestimate ozone at 1 hPa at Midlatitude regions (at the northern Midlatitude during months 1-4, and at the southern midlatitudes during months 6-10). Especially at southern midlatitudes, there is a sudden (too) strong depletion from May onwards. At lower levels, the FC run always overestimates ozone, meticulously following the MOZART-v1/v3 trends. The 2nd GEMS REA performs overall the best. Ozone columns are perfectly reproduced and also the ozone levels almost perfectly reproduce the BASCOE ozone levels.

Poles:

Outside ozone hole conditions, almost all models perform worst near the South Pole, except for MOCAGE-v1 and MOZART-v10. Trends over the year at the poles are generally very similar to midlatitude trends, but biases are more pronounced. The same accounts for the ozone levels: deviations occurring already at midlatitudes are even more pronounced at the Polar Regions. Except for MOCAGE-v1 and MOZART-v10, which follow very similar trends, the NP region is quite well represented in all models.

Ozone hole conditions:

Almost all models display the largest bias with BASCOE during ozone hole conditions, but one should of course keep in mind that the 2003 ozone hole event was exceptionally intense. The exception on this large bias is MOCAGE-v2 and the integrated GEMS reanalysis run, which have an almost perfect reproduction of the ozone hole, eventhough this is not true for the vertical distribution of MOCAGE-v2. The MOZART-v10 ozone hole ends too soon, which is why the model scores tend to be low from September 2003 on. As the GEMS FC run is a simulation based on the current GRG coupled system IFS + MOZART-v1/v3, the GEMS FC run is fully controlled by MOZART-v1/v3 and the absence of ozone depletion in this CTM consequently propagates into a serious underestimation of the ozone depletion (up to almost 40%) in the GEMS FC. TM5 correctly represents the early ozone hole until mid-September but ozone depletion also ends too early, resulting in a bias of 20% from October 2003 on.

12.6 Conclusions and recommendations

Table 12.2 lists the final scores for all models from best to worst. The integrated (GRG+GHG+AER) reanalysis run (f026) performs overall the best and excels the three stand-alone CTM. It is followed by the old MOCAGE-v1 run and TM5 (which is relaxed to climatology above 50hPa) strikingly perform better than MOCAGE-v2 and MOZART-v10.

Average model performance is constant in the different latitude zones, except for the Tropics, where it is generally better. Largest biases (i.e. worst scores) are generally observed during ozone hole conditions. O3 columns are generally quite well reproduced but the vertical distribution is incorrect.

Table 12.2: Final listing of the models following their performance, compared to BASCOE (from best to worst).

Model run	**Score (in %)**
GEMS REA – f026	98.78
MOCAGE-60lev01	*96,93*
TM5	95.95
MOCAGE-v2	95.09
MOZART-v10	93.62
GEMS FC	91.16
MOZART-v7	*88.13*

12.7 References

1. McPeters, R.D, Krueger, A.J., Bhartia, P.K., Herman, J.R., Wellemeyer, C.G., Seftor, C.J., Jaross, G., Torres, O., Moy, L., Labow G., Byerly, W., Taylor, S.L., Swissler, T., Cebula, R.P., "Earth Probe Total Ozone Mapping Spectrometer (TOMS) Data Products User's Guide", NASA Reference Publication, 1998-206895, 1998

2. Lucke, R. L., Korwan, D. R., Bevilacqua, R. M., Hornstein, J. S., Shettle, E. P., Chen, D. T., Daehler, M., Lumpe, J. D., Fromm, M. D., Debrestian, D., Neff, B., Squire, M., König-Langlo, G., Davies, J., "The Polar Ozone and Aerosol Measurement (POAM) III instrument and early validation results", Journal of Geophysical Research, vol. 104, no. D15, p. 18785-18800, 1999

3. Russell, James M., III, Gordley, Larry L., Park, Jae H., Drayson, S. R., Hesketh, W. D., Cicerone, Ralph J., Tuck, Adrian F., Frederick, John E., Harries, John E., Crutzen, Paul J., "The Halogen Occultation Experiment", Journal of Geophysical Research, vol. 98, no. D6, p. 10777-10797, 1993

4. Fischer, H., and Oelhaf, H., Remote sensing of vertical profiles of atmospheric trace constituents with MIPAS limb emission spectrometers, Appl. Opt., 35(16), 2787-2796, 1996.

5. Errera, Q., Daerden, F., Chabrillat, S., Lambert, J. C., Lahoz, W. A., Viscardy, S., Bonjean, S., Fonteyn, D., "4D-Var assimilation of MIPAS chemical observations: ozone and nitrogen dioxide analyse", Atmospheric Chemistry and Physics, Volume 8, Issue 20, 2008

6. Randall, C. E., Rusch, D. W., Bevilacqua, R. M., Hoppel, K. W., Lumpe, J. D., Shettle, E., Thompson, E., Deaver, L., Zawodny, J., Kyrö, E., Johnson, B., Kelder, H., Dorokhov, V. M., König-Langlo, G., Gil, M., "Validation of POAM III ozone: Comparisons with ozonesonde and satellite data", Journal of Geophysical Research Atmospheres, vol. 108, no. D12, p. ACH 6-1, CiteID 4367, 2003

7. Kinnison, D. E., Brasseur, G. P., Walters, S., Garcia, R. R., Marsh, D. R., Sassi, F., Harvey, V. L., Randall, C. E., Emmons, L., Lamarque, J. F., Hess, P., Orlando, J. J., Tie, X. X., Randel, W., Pan, L. L., Gettelman, A., Granier, C., Diehl, T., Niemeier, U., Simmons, A. J., "Sensitivity of chemical tracers to meteorological parameters in the MOZART-3 chemical transport model", Journal of Geophysical Research, vol. 112, no. D20, 2007

13. Summary and recommendations

The GEMS-GRG subproject has set up a global modelling system for analyses, forecasts and reanalyses of stratospheric and tropospheric ozone and ozone precursor species. Three state-of-the-art chemistry transport models (MOZART, TM5 and MOCAGE) were operated in standalone modes and coupled to the ECMWF Integrated Forecast and Assimilation System. Several simulations performed both with the coupled models and with the individual standalone models were scored against independent data sets in order to identify specific model uncertainties and provide a foundation for model improvements. The 4-dimensional variational data assimilation of ECMWF has been successfully used for ozone and carbon monoxide satellite data and its impact has been extensively assessed in the GEMS reanalyses.

The evaluation work carried out in these GEMS-GRG analyses identified a number of specific issues that are summarized below.

Tropospheric nitrogen dioxide (NO2):

Compared to SCIAMACHY and OMI products, the modelled tropospheric NO_2 columns are underestimated in polluted regions in winter, due to a lack of seasonality and intensity for anthropogenic emissions, especially over East Asia. Modelled tropospheric NO_2 columns are too large over biomass burning regions ; however the seasonality is correct. Results from the NO_2 evaluation support the provision of boundary limit conditions from global to regional models.

Tropospheric ozone (O3):

Globally, models have good scores for ozone, showing that the CTMs are able to reproduce correctly and integrate all the non-linear processes leading to the tropospheric ozone distribution (photochemical production/destruction, deposition and import from the stratosphere as main contributors).

At the surface and in comparison with the GAW and EMEP networks, all model simulations are capable to produce global concentration fields in reasonable agreement, i.e. within the objective to provide boundary conditions for regional models. Strong biases occur during winter and at night. All models fail to reproduce night surface ozone in rural areas. It may be partly due to low concentrations (emission/formation) of precursor gases, and partly due to too strong surface sink (dry deposition, NO titration). In Europe in summer it appears that CO-formation from organic compounds may be too high while depletion by OH may be too low. For mountain sites, there is a considerable uncertainty in the evaluation exercise which is imposed by the selection of the appropriate model level which represents the local conditions of the surface station. This limitation is certainly an issue for future evaluation in MACC.

In the boundary layer and in comparison with balloon soundings biases are generally negative. With regards to routine airborne MOZAIC profiles, there is no systematic behaviour (except for MOCAGE that has a tendency to overestimate).

In the free troposphere, there is no clear pattern or tendency for the models. Compared to ozone balloon soundings, all CTMs have negative bias in the troposphere. Compared to routine airborne MOZAIC profiles, modelled ozone concentrations are either slightly overestimated (MOZART-v10, TM5-v10, and even more by MOCAGE-v2), or slightly underestimated (MOZART-v1, TM5-v7).

The assimilation of ozone columns has not lead to noticeable improvements in the lower troposphere and in the boundary layer due to missing relevant information at these levels (reduced sensitivity of satellite UV measurements to O_3 absorption in the lower troposphere).

The coupled models don't perform especially better than the free CTMs. Interestingly, both reanalyses (IFS_f026_2003 and IFS_f026_2004) tend to slightly underestimate when CTMs used to slightly overestimate ozone. Improvements due to assimilation are less evident than with CO, likely because O3 is a secondary pollutant. In general, models perform better over Europe and US.

Tropospheric carbon Monoxide (CO):

Globally, models have lower scores for CO than for ozone. It reveals some deficiencies in inventory emissions, especially noticeable over Asia and developing countries.

Compared to data from the surface CMDL network, stand-alone models capture the CO seasonal variation, although it is quite under-estimated, especially in northern hemisphere. Better results are got in the southern hemisphere.

Compared to MOZAIC profiles, all CTMs have negative bias in the troposphere. Largest negative biases are in the surface and boundary layers. CO is slightly overestimated in the UT, indicating that the vertical gradient is not well reproduced. Models perform better over Europe and US. Beijing and Caracas have the overall worst scores. It is mainly due to either deficiencies in emission inventories (anthropogenic and biomass fires) or to the representativity of the MOZAIC profiles compared to the modelled grid resolution. Eventually, there may be deficiencies due to the chemical schemes, but it would require other kind of comparisons or some dedicated sensitivity experiments to demonstrate it.

Compared to routine aircraft vertical profiles of CO, the coupled runs generally perform better than the stand-alone runs. Data assimilation of CO from MOPITT improves mid- and upper-tropospheric modelled values. This is most evident during the springtime when the CTMs perform poorly, perhaps because it is a transition season. This improvement is partly due to the better transport brought about by the IFS dynamical model, but also largely due to the assimilation. However, data assimilation fails to improve to reproduce the high CO values in the boundary layer over most of large cities visited by MOZAIC. It is likely due to the poor degree of information of CO retrievals brought in addition to the a priori profile itself used in the retrieval process. Note also that data assimilation does not use the CO MOPITT averaging kernels. This is planned to be done within MACC.

In evaluating how well the models are able to simulate the long-range transport of CO plumes originating from biomass burning during the Alaskan and Canadian wildfires, it is shown that the stand-alone models were either unable to simulate the CO plumes or did a very a poor job. The IFS/MOZART coupled model with assimilation did a better job in simulating such plumes, however, the concentrations were significantly weaker than those observed in the MOZAIC data. Sensitivity studies have shown large sensitivities to the injection height for biomass burning plumes and to the time resolution of inventory emissions for biomass burning (8-days versus daily).

Stratospheric nitrogen dioxide (NO2)::

Compared to SCIAMACHY and OMI products, the stratospheric NO2 columns are well reproduced by stand-alone models (except MOCAGE with an underestimation).

Stratospheric ozone (O3): Compared to ozone balloon soundings, all CTMs have positive bias in the stratosphere and then do not reflect the Antarctic ozone hole conditions. Largest bias are in the tropopause region, with the ozone vertical gradient either too steep or to loose. Data assimilation improves the stratospheric ozone.

Comparisons of total ozone columns made both with BASCOE/MIPAS analyses and with the independent TOMS total ozone column data rank the coupled-model first, then CTMs. The performance of all models is more or less constant over the different latitude zones, except for the Tropics, where all models attain their best representativeness. On the whole, biases to BASCOE

increase for all models towards lower levels. Total columns in the midlatitude regions are generally slightly overestimated by all CTMs. Over the Poles, and outside ozone hole conditions, almost all models perform worst near the South Pole, except for MOCAGE-V1 and MOZART-v10. In ozone hole conditions, almost all models display the largest bias with BASCOE (one should of course keep in mind that the 2003 ozone hole event was exceptionally intense), except for MOCAGE-V2 and the integrated GEMS reanalysis run, which have an almost perfect reproduction of the ozone hole (even though this is not true for the vertical distribution of MOCAGE-V2).

Recommendations:

- Emissions:
  - Include general changes of emissions in the last years in the major urban centers and over developing countries (e.g. East Asia) in emission inventories. Include better time resolution in inventories.
  - Use of appropriate fire CO emission inventories consistent with up-to-date methodologies based on satellite detection
- Model set up:
  - Increase the horizontal and vertical resolution of CTMs simulations: improve the representation of transport processes, particularly at the tropopause region
  - Have the interactions between aerosol/radiation and chemical parameters in CTMs. A real cross-cutting interpretation with the aerosol fields is not yet feasible due to lack of essential parameters in the archived aerosol fields (e.g. particle extinction, profiles)
  - There is a need to assimilate surface observations and in-situ profiles in the future in order to constrain the boundary layer concentrations
- Needed parameterisations:
  - Include a parameterization for injection heights of biomass fire plumes (pyro-convection)
  - Have a better representation of near surface nighttime chemistry
  - Parameterisation needed to reproduce low dry deposition as a consequence of the dryness of vegetation during heat waves.
  - Parameterisation needed to reproduce the effect of high temperatures on biogenic emissions and on the evaporation of anthropogenic VOCs.
- Evaluation set up:
  - It is important to continue to build standardized validation using many different data sets and model intercomparisons. It is very useful to identify problems and improve coupled systems.
  - Extend the evaluation to other gases, like NO, NO_2, SO_2 in order to tackle model/parametrisation shortcomings
  - Have more data available in NRT (e.g. from GAW stations over flat terrains, from future IAGOS aircraft, …)

This report may be considered for defining the future automatic evaluation methodology to set up within MACC.

ANNEX 1: Summary of the different CTM specifications

The three CTMs are MOCAGE by Meteo France (Josse et al., 2004 ; Bousserez et al.? 2007), MOZART-3 by NCAR, MPI Hamburg, FZ Juelich (Horowitz et al., 2003 ; Kinnisson et al., 2007) and TM5 by KNMI (version KNMI-cy3-GEMS, Krol et al., 2005). Three CTMs were selected because previous model intercomparison studies showed considerable spread of results, and a 3-model ensemble can provide some guidance with respect to the robustness of the simulation results. A summary of the different CTM specifications can be found in Table 1.

	MOZART-3	TM5 (KNMI-cy3-GEMS)	MOCAGE
Horizontal resolution	1.875°*1.875°	2°*3°	2°*2°
Vertical resolution	60 layers up to 0.1 hPa	as MOZART-3	as MOZART-3
Meteorological fields	Basic fields, Heat fluxes	as MOZART-3 and precipitation, clouds, convective mass fluxes and surface properties	as MOZART-3
Advection	Flux form semi-lagrangigain (Lin and Rood, 1996)	Slopes scheme (Russel and Lerner, 1981)	Semi-implicit, semi-lagrangian (Williamson and Rasch, 1989)
Convection scheme	Hack (1994) for shallow and mid-level convection, Zhang & McFarlane (1995) for deep convection	Tiedtke (1989)	Bechtold et al. (2001), completed by Mari et al. (2000)
Diffusion scheme	Holstlag and Boville (1993)	Holtslag and Moeng (1991) for near surface, Louis (1979) for free troposphere	Louis (1979)
Chemical mechanism	JPL-03 and JPL-06 (Sander et al.., 2003, 2006) as described in Kinnison et al. (2007) with updates to JPL-06 (Sander et al., 2006), $SO_x/NH_3/NH_4$ mechanism from MOZART-4 (Emmons et al. 2009., in prep.) (115 species, 325 reactions)	CBM4 scheme as described in Houweling et al. [(1998]) for troposphere, updated reaction rates according to JPL-03 and JPL-06 (Sander et al.., 2003, 2006) stratospheric O_3 climatology, Fortuin- and Kelder (1998) HNO3 climatology from UARS (55 species, 85 reactions)	REPROBUS (Lefèvre et al., 1994) scheme included in the RACMOBUS scheme (Carslaw et al. 1995) for heterogeneous stratospheric chemistry (118 species and 350 reactions)

Emissions	RETRO (Schultz et al., 2009), GFED (GFEDv2 (Van der Werf et al., 2006; Randerson et al., 2006)	as MOZART-3	as MOZART-3

Tabble 1: CTM specifications.

ANNEX 2: Model documentation of the MOZART CTM as implemented in the GEMS system

Olaf Stein

Forschungszentrum Jülich, ICG2

o.stein@fz-juelich.de

March 4, 2009

Overview

MOZART (Model for Ozone And Related Tracers) has been chosen as one of the global chemistry transport models (CTM) to be coupled to the ECMWF Integrated Forecast system (IFS) in the GEMS project. The documentation refers to the MOZART version which is actually implemented in the GEMS forecast system. This version is also available as a stand-alone CTM and differs slightly from previously released MOZART versions.

MOZART in GEMS is based on the MOZART3 model code (Kinnison et al. 2007) which itself is an extension of the troposphere model MOZART2 (Horowitz et al. 2003) to the stratosphere and mesosphere. Some features have been added from MOZART4 (Pfister et al. 2008, Emmons et al. 2009). In the following the MOZART stand-alone model as used for reanalysis and near-realtime forecast simulations in GEMS is described, main features of the coupled version MOZART-IFS are introduced in chapter 6. The coupling is described in more detail in Flemming (2008).

Meteorology and resolution

In stand-alone mode MOZART is driven by meteorological fields from ECMWF. This can be either reanalysis or forecast runs from IFS, currently updated every 6 hours. The variables needed from MOZART are displayed in table 1. MOZART uses the same 60 vertical hybrid layers as the IFS model reaching from the surface to 0.1 hPa. Horizontally the IFS fields are interpolated on a T63-Lin-Rood grid (Lin & Rood 1996), which has 192 x 96 grid points (lon/lat) corresponding to a resolution of 1.875° x 1.895°. The temporal resolution for this setting is 15 minutes. In principle the horizontal resolution is flexible, one shorter simulation has been done with enhanced resolution of 320 x 160 grid points (1.125°).

Chemistry scheme

MOZART simulates the mixing ratios of 115 species from the surface to the mesosphere (table 2). There are three lumped hydrocarbons (representation of physical and chemical properties bold): TOLUENE represents the aromatic species benzene, **toluene**, and xylene, BIGALK represents the alkanes containing four or more carbons (**butane**, pentane, hexane, and higher), BIGENE represents the alkenes containing four or more carbons (**butene** and higher). Invariant species of the system are N2 and $N(^2D)$. The system of chemical reactions consists of 71 photolysis reactions, 223 gas phase reactions and 21 heterogeneous reactions. Most of these reactions are described in Kinnison et al. (2007). 37 gas phase reactions have been updated to JPL-06 (Sander et al. 2006). These reactions are described in table 3. Details of the isoprene degradation scheme can be found in Pfister et al. (2008). The MOZART3 chemical mechanism has been complemented by a Radon and lead scheme from MOZART2 (Horowitz et al. 2003) and by the $SO_x/NH_3/NH_4$ chemistry scheme from MOZART4 (Emmons et al. 2009). The latter contains the additional species SO_2, SO_4, DMS, NH_3, NH_4, and NH_4NO_3. New reactions are:

- $SO_2 + OH \rightarrow SO_4$
- $DMS + OH \rightarrow SO_2$
- $DMS + OH \rightarrow .5 * SO_2 + .5 * HO_2$
- $DMS + NO_3 \rightarrow SO_2 + HNO_3$
- $NH_3 \rightarrow NH_4$
- $NH_3 + OH \rightarrow H_2O$
- $NH_3 + HNO_3 \rightarrow NH_4NO_3$

in-cloud:

- $SO_2 + H_2O_2 \rightarrow SO_4$
- $SO_2 + O_3 \rightarrow SO_4$

The chemical and photochemical processes are expressed at each grid point by a system of time-independent ordinary differential equations as described in Kinnison et al. (2007). The system of equations for the 18 longer-lived species is solved by an explicit forward Euler method, while for species with short lifetimes and strong forcings an implicit backward Euler method is chosen. A detailed description of the photolytic approach can be found in Kinnison et al. (2007): For wavelengths from 120 nm to 200 nm the transmission function is calculated explicitly, while for wavelengths from 200 to 750 nm a flux look-up table approach based on STUV is used. The parameterization of J(NO) follows Minschwaner & Siskind (1993), $J(O_3)$ is based on Koppers & Murtagh (1996) and Chabrillat & Kockarts (1997, 1998).

Physical parameterizations

MOZART includes parameterizations for advection, convective transport, boundary layer mixing, wet and dry deposition as described in Horowitz et al. (2003):

Tracer advection uses the flux-form semi-Lagrangian advection scheme of Lin & Rood (1996). Convective muss fluxes are calculated using the Hack (1994) scheme for shallow and mid-level convection and the Zhang & MacFarlane (1995) scheme for deep convection. Vertical diffusion in the boundary layer is represented by the parameterization of Holtslag & Boville (1993).

Parameterizations for the hydrological cycle, heterogeneous stratospheric chemistry processes, gravity waves, molecular diffusion follow Kinnison et al. (2007): H_2O is advected partitioned into condensed and gas phase. Basic tropospheric cloud processes as well as evaporation and precipitation are represented in the model. These processes are replaced by heterogeneous chemistry processes for high latitudes (>60° N/S) at pressures < 300hPa in order to enable the formation of aerosols from

heterogeneous activation of bromine and chlorine radicals. The heterogeneous chemistry scheme following Considine et al. (2000) consists of 18 processes taking place on four aerosol types: liquid binary sulphate (LBS), supercooled ternary solution (STS), nitric acid tri-hydrate (NAT), and water ice. STS and NAT can be formed for temperatures below 200 K allowing MOZART to include heterogeneous processes on liquid surface aerosols and polar stratospheric clouds (PSC, type 1a, 1b, and 2) and HNO_3 uptake on solid aerosol.

Transport by breaking gravity waves is based on a parameterization from Lindzen (1981) as described in Kinnison et al. (2007).

Boundary conditions

Upper and lower tracer boundary conditions can be read in from netcdf files if necessary. In the GEMS framework all upper tracer boundaries are set to zero flux. At the surface species volume mixing ratios can be prescribed from observations or species can be emitted as emission fluxes into the lowest model level. The actual lower boundary settings are displayed in table 2. All boundary condition fields are interpolated from the original resolution to the MOZART resolution.

The emission inventories produced for the GEMS project are described in detail in Schultz & Stein (2006). Anthropogenic and natural emissions are provided as monthly mean fields representing the year 2003 derived from the emission inventory of the RETRO project (Schultz et al. 2005) while wildfire emissions from GFEDv2 (Randerson et al. 2006) are available in monthly resolution for the years 1997-2007 and in 8day resolution for 2001-2006. Emissions of SO_2, NH_3, and DMS are not part of the RETRO inventory and have been taken from various sources: SO_2 anthropogenic emissions are from EDGAR-FT2000 (Olivier et al. 2005), SO_2 wildfire emissions from GFEDv2, SO_2 volcano emissions from GEIA, NH_3 emissions from EDGAR2 (Olivier et al. 1996), and DMS emissions from Kloster et al. (2005). For the simulation MOZART_V10 RETRO ship emissions have been replaced by estimates based on Corbett et al. (2003) and East Asian anthropogenic emissions have been replaced by the REAS inventory (Ohara et al. 2007) but keeping the original RETRO seasonality.

The model also accounts for in-situ production of tracers: NO_x from lightning is dependent on the distribution of convective clouds, following a parameterization of Price et al. (1997). Aircraft emissions of NO_x and CO are also included in the model (Horowitz et al. 2003).

MOZART contains a detailed representation of both wet and dry deposition for the species displayed in table 2. Dry deposition velocities used in the model have been calculated offline using a resistance-inseries-scheme (Wesely 1989, Hess et al. 2000) driven by ten years of NCEP reanalysis. Wet deposition is represented as a first-order loss process, with additional in-cloud scavenging and below cloud washout for soluble species (Horowitz et al. 2003).

Coupling strategy

In the framework of GEMS the MOZART CTM is coupled to the ECMWF integrated forecast model (IFS) building the MOZART-IFS model system. For a coupled simulation both models are running in parallel and exchange several two- and three-dimensional fields every hour using the OASIS4 coupling software developed in the PRISM project (Valcke & Redler 2006): IFS provides meteorological data to MOZART as described in table 1. Data assimilation and transport of the GEMS species O_3, CO, NO_x, HCHO, and SO_2 takes place in IFS, while the whole chemical system is calculated in MOZART. At exchange time MOZART provides updated tendency terms for chemistry, emission and deposition sources and sinks for the GEMS species. IFS returns the updated mixing ratio fields for these species to MOZART. Currently the coupled reanalysis and forecast experiments run in

the CTM-constrained mode, which means that the feedback of tracer concentrations is only done at the beginning of a new IFS forecast run, typically every 24 hours. In feedback-mode updated GEMS tracer concentrations are provided to MOZART every hour. The coupled system is described in detail in Flemming (2008).

References

1. Chabrillat, S., and G. Kockarts (1997), Simple Parameterization of the Absorption of the Solar Lyman-Alpha Line, *Geophys. Res. Lett.*, 24(21), 2659–2662.
2. Chabrillat, S., and G. Kockarts (1998), Correction to "Simple Parameterization of the Absorption of the Solar Lyman-alpha line", *Geophys. Res. Lett.*, 25(1), 79.
3. Considine, D. B., A. R. Douglass, P. S. Connell, D. E. Kinnison, and D. A. Rotman (2000), A polar stratospheric cloud parameterization for the global modeling initiative three-dimensional model and its response to stratospheric aircraft, *J. Geophys. Res.*, 105(D3), 3955–3973.
4. Corbett, J. J., and H. W. Koehler (2003), Updated emissions from ocean shipping, J. Geophys. Res., 108(D20), 4650, doi:10.1029/2003JD003751.
5. Emmons, L. et al. (2009), Sensitivity of chemical tracers to meteorology in MOZART-4, in preparation
6. Flemming, J. (2008), Technical description of the coupled forecast system IFS-CTM for global reactive gases forecast and assimilation in GEMS, available at http://gems.ecmwf.int/do/get/PublicDocuments
7. Hack, J. J. (1994), Parameterization of moist convection in the NCAR community climate model (CCM2), *J. Geophys. Res.*, *99*, 5551–5568.
8. Hess, P., S. Flocke, J.-F. Lamarque, M. Barth, and S. Madronich (2000), Episodic modeling of the chemical structure of the troposphere as revealed during the spring MLOPEX 2 intensive, J. Geophys. Res., 105(D22), 26809-26839.
9. Hollingsworth, A., et al. (2008), Toward a monitoring and forecasting system for atmospheric composition: The GEMS project, Bull. Am. Meter. Soc., 89, 1147-1164, doi:10.1175/2008BAMS2355.1
10. Holtslag, A., and B. Boville (1993), Local versus nonlocal boundary-layer diffusion in a global climate model, *J. Clim.*, *6*, 1825–1842.
11. Horowitz, L. W., et al. (2003), A global simulation of tropospheric ozone and related tracers: Description and evaluation of MOZART, version 2, *J. Geophys. Res.*, 108(D24), 4784, doi:10.1029/2002JD002853.
12. Kloster, S, J. Feichter, E. Maier-Reimer, K. D. Six, P. Stier, P. Wetzel (2005): DMS cycle in the marine ocean-atmosphere system - a global model study, Biogeosciences Discussions; 2, 1067-1126.
13. Koppers, G. A. A., and D. P. Murtagh (1996), Model studies of the influence of O_2 photodissociation parameterizations in the Schumann-Runge bands on ozone related photolysis in the upper atmosphere, *Ann. Geophys.*, *14*, 68–79.
14. Lin, S. J., and R. B. Rood (1996), A fast flux form semi-Lagrangian transport scheme on the sphere, *Mon. Weather Rev.*, *124*, 2046–2070.
15. Lindzen, R. S. (1981), Turbulence and Stress Owing to Gravity Wave and Tidal Breakdown, *J. Geophys. Res.*, 86(C10), 9707–9714.
16. Minschwaner, K., and D. E. Siskind (1993), A new calculation of nitric oxide photolysis in the stratosphere, mesosphere, and lower thermosphere, *J. Geophys. Res.*, *98*, 20,401–20,412.
17. Ohara, T., Akimoto, H., Kurokawa, J., Horii, N., Yamaji, K., Yan, X., and Hayasaka, T. (2007), An Asian emission inventory of anthropogenic emission sources for the period 1980–2020, Atmos. Chem. Phys., 7, 4419-4444.
18. Olivier, J.G.J., Bouwman, A.F., Van der Maas, C.W.M., Berdowski, J.J.M., Veldt, C., Bloos, J.P.J., Visschedijk, A.J.H., Zandveld, P.Y.J. and Haverlag, J.L. (1996). *Description of EDGAR Version 2.0: A set of global emission inventories of greenhous gases and ozone-depleting substances for all anthropogenic and most natural sources on a per country basis and on 1ox1o grid*. National Institute of Public Health and the Environment (RIVM) report no. 771060 002 / TNO-MEP report no. R96/119.
19. Olivier, J.G.J., Van Aardenne, J.A., Dentener, F., Ganzeveld, L. and J.A.H.W. Peters (2005). Recent trends in global greenhouse gas emissions: regional trends and spatial distribution of key sources. In: "Non-CO2 Greenhouse Gases (NCGG-4)", A. van Amstel (coord.), page 325-330. Millpress, Rotterdam, ISBN 90 5966 043 9.
20. Pfister, G., L. K. Emmons, P. G. Hess, J.-F. Lamarque, S. Walters, A. Guenther, P. I. Palmer, and P. Lawrence (2008), Contribution of isoprene to chemical budgets: A model tracer study with the NCAR CTM MOZART-4, *J. Geophys. Res.*, *113*, D05308, doi:10.1029/2007JD008948.

21. Price, C., J. Penner, and M. Prather (1997), NO_x from lightning 1. Global distribution based on lightning physics, *J. Geophys. Res.*, 102(D5), 5929–5941.
22. Randerson, J. T., G. R. van der Werf, L. Giglio, G. J. Collatz, and P. S. Kasibhatla (2006), Global Fire Emissions Database, Version 2 (GFEDv2). Data set. Available on-line [http://daac.ornl.gov/] from Oak Ridge National Laboratory Distributed Active Archive Center, Oak Ridge, Tennessee, U.S.A. idoi:10.3334/ORNLDAAC/834.
23. Sander, S.P., et al., *Chemical Kinetics and Photochemical Data for Use in Atmospheric Studies, Evaluation Number 15, JPL Publication 06-02*, Jet Propulsion Laboratory, Pasadena, Calif., 2006.
24. Schultz, M.G., T. Pulles, R. Brand, M. van het Bolscher, and S. B. Dalsøren (2005), A global data set of anthropogenic CO, NOx, and NMVOC emissions for 1960-2000 (paper in preparation), data available from http://retro.enes.org/data_emissions.shtml
25. Schultz, M.G., and O. Stein (2006), *GEMS (GRG) emissions for 2003 reanalysis simulations*. Technical report, MPI-M Hamburg, 2006.
26. Valcke, S., and R. Redler (2006), OASIS4 User Guide (OASIS4_0_2). PRISM Support Initiative Report No 4, 60 pp.
27. Wesely, M. L.**,** Parameterization of surface resistance to gaseous dry deposition in regional-scale numerical models, *Atmos. Environ.*, *23*, 1293–1304, 1989.
28. Zhang, G. J., and N. A. McFarlane, Sensitivity of climate simulations to the parameterization of cumulus convection in the Canadian climate centre general circulation model, *Atmos. Ocean*, *33*, 407–446, 1995.

Variable	long name	Unit	dim		GRIB Code Nr.
T	temperature	K	3 D	Required	130
U	zonal wind component	m/s	3 D	Required	131
V	meridional wind component	m/s	3 D	Required	132
Q	specific humidity	kg/kg	3 D	Required	133
PS	surface pressure	Pa	2 D	Required	134
TS	surface temperature	K	2 D	Desired	235
ORO	land / sea / ice flag		2 D	Desired	172
PHIS	surface geopotential	m^2/s^2	2 D	Required	129
SHFLX	surface sensible heat flux	W/m^2	2 D	Required	146
QFLX	surface latent heat flux	$kg/(m^2*s)$	2 D	Required	147
TAUX	X-component of surface stress	N/m^2	2 D	Required	180
TAUY	Y-component of surface stress	N/m^2	2 D	Required	181
OMEGA	vertical wind velocity	Pa/s	3 D	Desired	135
SD	snow depth	m	2 D	Desired	141
SSRD	surface solar radiation downw.	W/m^2	2 D	Desired	169

Table 1: Dynamic input variables to be read into MOZART-3 from ecmwf:

Table 2: MOZART transported species:

advected species	long name	solver	Lower boundary	Deposition	Assimilation in GEMS
O3	ozone	implicit		dry	x
O		implicit			
O1D	O^1D	implicit			
O2		implicit			
N2O	nitrous oxide	explicit	fixed		
N		implicit			
NO	nitric oxide	implicit	emission		

NO2	nitrogen dioxide	implicit		dry	x
NO3	nitrate radical	implicit			
HNO3	nitric acid	implicit		dry/wet	
HO2NO2	peroxynitric acid	implicit		dry/wet	
N2O5	nitric pentoxide	implicit			
H		implicit			
OH	hydroxyl radical	implicit			
HO2	hydroperoxy radical	implicit			
H2O2	Hydrogen Peroxide	implicit		dry/wet	
H2		explicit	emission		
CL	chlorine	implicit			
CL2		implicit			
CLO	chlorine monoxide	implicit			
OCLO	chlorine dioxide	implicit			
CL2O2	ClO dimer	implicit			
HCL	hydrochloric acid	implicit		wet	
HOCL	hypochlorous acid	implicit		wet	
CLONO2	chlorine nitrate	implicit		wet	
BR	bromine	implicit			
BRO	bromine monoxide	implicit			
HBR	hydrogen Bromide	implicit		wet	
HOBR	cypobromous acid	implicit		wet	
BRONO2	bromine nitrate	implicit		wet	
BRCL	bromine monochloride	implicit			
CH4	methane	explicit	fixed		
CH3O2	methylperoxy radical	implicit			
CH3OOH	methyl hydroperoxide	implicit		dry/wet	
CH3OH	methanol	implicit	emission	dry/wet	
CH2O	formaldehyde	implicit	emission	dry/wet	x
CO	carbon monoxide	explicit	emission	dry	x
C2H4	ethylene	implicit	emission		

C2H6	ethane	implicit	emission		
C2H5O2	etylperoxy radical	implicit			
C2H5OOH		implicit		dry/wet	
CH3CO3	acylperoxy radical	implicit			
CH3COOH	acetic acid	implicit		dry/wet	
CH3CHO	acetaldehyde	implicit	emission	dry/wet	
C2H5OH	ethanol	implicit	emission	dry/wet	
GLYALD	glycolaldehyde	implicit		dry/wet	
GLYOXAL		implicit			
CH3COOOH	peracetic acid	implicit		dry/wet	
EO2	HOCH2CH2O2	implicit			
EO	HOCH2CH2O	implicit			
PAN	peroxyacetylnitrate	implicit		dry	
C3H6	propene	implicit	emission		
C3H8	propane	implicit	emission		
C3H7O2		implicit			
C3H7OOH	propionic acid	implicit		dry/wet	
CH3COCH3	acetone	implicit	emission	dry	
PO2	C3H6OHO2	implicit			
POOH	C3H6OHOOH	implicit		dry/wet	
HYAC	hydroxyacetone	implicit		dry/wet	
RO2	peroxy radical	implicit			
CH3COCHO	methylglyoxal	implicit		dry/wet	
ROOH	hydroperoxide	implicit		dry/wet	
BIGENE	lumped alkenes (**butene** and higher)	implicit	emission		
BIGALK	lumped alkanes (**butane**/pentane/hexane and higher)	implicit	emission		
MEK	methyl ethyl ketone	implicit			
ENEO2	C4H9O3	implicit			
MEKO2	C4H7O3	implicit			

MEKOOH	MEK peroxide	implicit		dry/wet	
MCO3	CH2CCH3CO3	implicit			
MVK	methyl vinyl ketone	implicit		wet	
MACR	methacrolein	implicit		wet	
MACRO2		implicit			
MACROOH		implicit		dry/wet	
MPAN	peroxymethacryloyl nitrate	implicit		dry	
ONIT	lumped organic nitrate	implicit		dry/wet	
ISOP	isoprene	implicit	emission		
ALKO2	C5H11O2	implicit			
ALKOOH	acetone dimethyl acetal	implicit		dry/wet	
BIGALD		implicit			
HYDRALD	lumped unsaturated hydroxycarbonyl	implicit		dry/wet	
ISOPO2	peroxy radical from OH + Isoprene	implicit			
ISOPNO3	peroxy radical from NO3 + Isoprene	implicit		wet	
ONITR	reactive organic nitrates	implicit		dry/wet	
XO2	halogenated peroxy radicals	implicit			
XOOH	halogenated hydroperoxides	implicit		dry/wet	
ISOPOOH	unsaturated hydroxyhydroxyperoxide	implicit		dry/wet	
TOLUENE	lumped aromatics (benzene/**toluene**/xylene)	implicit	emission		
CRESOL		implicit			
TOLO2	C7H9O3	implicit			
TOLOOH	C7H10O3	implicit		dry/wet	
XOH	C7H10O4	implicit			
C10H16	terpenes	implicit	emission		
TERPO2	C10H17O3	implicit			
TERPOOH	C10H18O3	implicit		dry/wet	
CH3CL	methyl chloride	explicit	fixed		
CH3BR	methyl bromide	explicit	fixed		
CFCL3	CFC-11	explicit	fixed		

CF2CL2	CFC-12	explicit	fixed		
CFC113	CFC-113	explicit	fixed		
HCFC22	HCFC-22	explicit	fixed		
CCL4	carbon tetrachloride	explicit	fixed		
CH3CCL3	methyl chloroform	explicit	fixed		
CF3BR	halon 1301	explicit	fixed		
CF2CLBR	halon 1211	explicit	fixed		
CO2	carbon dioxide	explicit	fixed		
H2O	water vapor	implicit			
Rn	radon	explicit	emission		
Pb	lead	explicit		dry/wet	
SO2	sulphur dioxide	implicit	emission	dry/wet	x
SO4	sulphate	implicit		dry/wet	
DMS	dimethylsulfide	implicit	emission		
NH3	ammonia	implicit	emission	dry/wet	
NH4NO3	ammonium nitrate	implicit		dry/wet	
NH4	ammonium	implicit		dry/wet	
H2SO4	sulphuric acid	explicit		dry/wet	

Table 3: Changes to the MOZART3 chemical mechanism after comparison to JPL-2006

reaction	A-Factor	-E/R	k(298 K)	Comments
$O(^1D) + O_2 \rightarrow O + O_2$	3.3×10^{-11}	55		JPL06
$O(^1D) + H_2O \rightarrow OH + OH$	1.63×10^{-10}	60		JPL06
$O(^1D) + N_2 \rightarrow O + N_2$	2.15×10^{-11}	110		JPL06
$O(^1D) + N_2O \rightarrow N_2 + O_2$	4.7×10^{-11}	20		JPL06
$O(^1D) + N_2O \rightarrow NO + NO$	6.7×10^{-11}	20		JPL06
$H + HO_2 \rightarrow 2\ OH$			7.2×10^{-11}	JPL06

$H + HO_2 \rightarrow H_2 + O_2$			6.9×10^{-12}	JPL06
$H + HO_2 \rightarrow O + H_2O$			1.6×10^{-12}	JPL06
$OH + H_2 \rightarrow H_2O + H$	2.8×10^{-12}	-1800		JPL06
$OH + OH \rightarrow H_2O + O$			1.8×10^{-12}	JPL06
$OH + H_2O_2 \rightarrow H_2O + HO_2$			1.8×10^{-12}	JPL06
(1) $HO_2 + HO_2 \rightarrow H_2O_2 + O_2$	(1) 3.5×10^{-13} (2) 1.7×10^{-33}	(1) 430 (2) 1000		JPL06
(2) $HO_2 + HO_2 + M \rightarrow H_2O_2 + O_2$	$f_c = 1. + 1.4\times10^{-21} * [H_2O] * \exp(2200/T)$ $k(T) = (k_1 + k_2 * [M]) * f_c$			
$HO_2 + NO_3 \rightarrow$ products			3.5×10^{-12}	JPL06
$O + NO_2 \rightarrow NO + O_2$	5.1×10^{-12}	210		JPL06
$O + ClO \rightarrow Cl + O_2$	2.8×10^{-11}	85		JPL06
$Cl + H_2 \rightarrow HCl + H$	$3.05\text{x}10^{-11}$	-2270		JPL06
$Cl + CH_4 \rightarrow HCl + CH_3$	$7.3\text{x}10^{-12}$	-1280		JPL06
$OH + BrO \rightarrow$ products	$1.7\text{x}10^{-11}$	250		JPL06
$OH + HBr \rightarrow H_2O + Br$	$5.5\text{x}10^{-12}$	200		JPL06
$HO_2 + Br \rightarrow HBr + O_2$	$4.8\text{x}10^{-12}$	-310		JPL06
$HO_2 + BrO \rightarrow$ products	4.5×10^{-12}	460		JPL06
$Cl + CH_3Cl \rightarrow HO_2 + CO + 2\,HCl$	$2.17\text{x}10^{-11}$	-1130		JPL06, with products HO2 + CO + 2HCl
$OH + CH_3CCl_3 \rightarrow CH_2CCl_3 + H_2O$	1.64×10^{-12}	-1520		JPL06
$OH + CH_3OH \rightarrow$ products	2.9×10^{-12}	-345		JPL06
$OH + H_2CO \rightarrow H_2O + HCO$	5.5×10^{-12}	125		JPL06

OH + $CH_3C(O)CH_3$ → products	k(T) = 1.33×10^{-13} + 3.82×10^{-11} exp(-2000/T)			changed user defined reaction for this rate constant. See note D27 in JPL06
$CH_3C(O)O_2$ + $CH_3C(O)O_2$ → products	2.9×10^{-12}	500		JPL06
OH + C_3H_8→ products	8.7×10^{-12}	-615		JPL06
CH_3O_2 + $CH_3C(O)CH_2O_2$ → products	7.5×10^{-13}	500		JPL06
OH + $C_2H_5C(O)OH$ → products			1.2×10^{-12}	JPL06
OH + HC(O)C(O)H→ products			1.15×10^{-11}	JPL06
OH + CH_3SCH_3 → SO_2 (c)	1.1×10^{-11}	-240		JPL06; products as in MZ4
NO_3 + CH_3SCH_3 → SO_2 + HNO_3 (c)	1.0×10^{-12}	500		JPL06; products as in MZ4

(c) additional reaction taken from MOZART4

reaction	k_o^{300}	N	k_∞^{300}	m	f
H + O_2 + M -> HO_2 + M	4.4×10^{-32}	1.3	4.7×10^{-11}	0.2	0.6
NO_2 + OH + M →$HONO_2$ + M	1.8×10^{-30}	3.0	2.8×10^{-11}	0.0	0.6
NO_2 + HO_2 + M →HO_2NO_2 + M	2.0×10^{-31}	3.4	2.9×10^{-12}	1.1	0.6
(k1) OH + CO + M ⟶ HOCO + M (k2) OH + CO + M ⟶ CO_2 + H + M create user defined reaction with k = k1 + k2 for k2 use the expression for chemical activation reactions from JPL06 (see note D1)	(k1) 5.9×10^{-33} (k2) 1.5×10^{-13}	1.4 -0.6	1.1×10^{-12} 2.1×10^{9}	-1.3 -6.1	0.6 0.6

JPL06 updates in the above troe reaction rate constants, except for f, which is kept at 0.6 (this is done for all troe reactions).

ANNEX 3: TM5 model description

Model Documentation of TM5-KNMI-cy3-GEMS

Vincent Huijnen

KNMI, The Netherlands

email: huijnen@knmi.nl

November 5, 2008

Abstract

This report provides a documentation of the global chemistry Transport Model version 5 (TM5), as applied in the GEMS subproject on Global Reactive Gases (GRG) for the period up to 2008. This includes a discussion of the grid handling (the zoom option), transport modeling and the chemical mechanism. In addition specific settings for the GEMS project are provided. Throughout the document information on the specific settings of the reference model runs and sensitivity runs as performed in GEMS-GRG subproject are described.

Introduction

The aim of the EU GEMS subproject on Global Reactive Gases (GRG) is to setup an operational assimilation and forecast system for atmospheric, chemically reactive trace gases. In this context it is chosen to couple a number of chemical transport models (CTM's) to the ECMWF Integrated Forecast System (IFS).

In the GEMS-GRG project TM5 (Krol et al. [17]) serves as one of the independent chemical transport models. TM5 is a global chemistry transport model, developed in a cooperation of a number of institutes. Parametrizations as applied in TM5 are mainly similar as in version TM3, Dentener et al. [8] and references therein. The focus in TM5 is on modeling the global, tropospheric trace gases. The chemical mechanism that is used is based on a modified CBM4 mechanism, Houweling et al. [15]. Recently, the reaction rates have been updated, e.g. the inclusion of reccomendations by Jet Propulsion Laboratory according to Sander et al. [31], as described in Williams and van Noije [40].

For the evaluation of the regional atmospheric chemical composition, the effects of long range transport are important. This requires the application of a global model, which includes an exchange with the stratosphere. On the other hand, emission sources will influence the regional composition. Also measurements taken at sites that are still affected by anthropogenic influences requires a simulation that is representative locally. This requires a high spatial / temporal resolution. However, the evaluation of a global model on a local scale is computationally highly demanding whilst being relatively inefficient. In TM5 this problem is solved by the application of regional zooming, as described in Krol et al. [17]. TM5 documentation on the web can be found at http://www.phys.uu.nl/~tm5/ .

in the GEMS-GRG project TM5 is applied in two different modes. In the stand-alone mode the model uses meteorological data delivered from either reanalyses or operational forecasts of the IFS. The second mode is the application of TM5 in the coupled IFS-CTM system. In this setting the IFS provides online meteorological data to the CTM on a high temporal resolution, and tendency terms for chemical sources and surface fluxes from a limited number of trace gases is provided by the CTM to IFS. This is used in online forecast experiments and in data-assimilation runs.

In the next sections details of the TM5 model will be addressed, with focus on settings of TM5 in the GEMS-GRG project (version TM5-KNMI-cy3-GEMS). During the project the TM5 code has been updated several times, therefore in practise several runs have been made available for validation. The different versions of the TM5 code are denoted by version V3 until V10. First, the grid handling is discussed. Next, information on the transport models and chemistry models is provided, including a description of the applied parametrizations and emissions. Then, coupling aspects from the viewpoint of TM5 are described.

The grid, zoom regions and time stepping

The discretization of TM5 is based on a cartesian grid. Apart from the definition of a global, three-dimensional grid and a two-dimensional surface grid, an arbitrary number of zoom regions can be defined. The zooming strategy adopted in TM5 allows a nesting of regions with an increasing resolution, Krol et al. [17]. In this way predefined regions can be evaluated simultaneously at various resolutions, with the coarsest region providing boundary conditions to the nested regions. The zooming algorithm has been developed by Berkvens et al. [2], who provided the mathematical background for the mass-conserving transport algorithm.

With respect to the time stepping, TM5 adopts an operator splitting algorithm. This consists of the following steps: Advection in X,Y,Z-directions, parameterization of sub-grid scale mixing by deep convection and vertical diffusion (V), chemistry (C) and sources/sinks (S). This is different from the algorithm presented in Krol et al. [17], where iterations to fulfill the CFL criterion were performed after each substep of the algorithm, in the current setting iterations are performed over the whole operator sequence

$(XY\ Z\ V\ SC\ CSV\ ZY\ X)^n$

In this expression *n* denotes the number of iterations needed to fulfill the CFL criterion. Notice that in this procedure the time-stepping does not need to be constant. In case that there are zoom regions defined, the time stepping then becomes nested. The time step for the nested domain reduce accordingly, while the number of iterations increase. Within the GEMS project, a global resolution of $3^\circ \times 2^\circ$ has been applied, while for specific model studies, (e.g. the 2003 European Heat Wave period and the 2003 Siberian Fire case study) refinements over the European region and Siberian fire region of $1^\circ \times 1^\circ$ is introduced. At the surface layer the resolution for many of the processes (e.g. emission, deposition) is set standard to $1^\circ \times 1^\circ$. After evaluation of these surface parametrizations the information is coarsened to the working horizontal resolution, if necessary.

To avoid a very small time stepping near the poles due to the converging grid, a reduced grid has been applied in the x-direction, Petersen et al. [25]. This means that grid cells are combined. For the vertical distribution, 60 hybrid sigma layers are adopted with the same distribution as employed by the ECMWF model, from the surface up to 0.1hPa.

Transport schemes

In TM5, the so-called first-order slopes Russell and Lerner [29] and second-order moments, Prather [26] advection schemes are implemented. In the GEMS setup the slopes scheme is used, which is appropriate for the troposphere, Krol et al. [17]. It should be noted that use of the second-order moments scheme within zoomed regions is prohibited due to CFL violations. Moreover, applying second-order moments throughout the troposphere severely degrades the model performance.

In TM5 two sets of convective mass fluxes are available: the offline computed convective mass fluxes, based on Tiedtke [36], and the archived convective fluxes provided by ECMWF for ERA40. The latter ones are used in GEMS for the years that are available, i.e. until 2003. For the 2004 stand-alone runs the scheme by Tiedtke is applied. One of the problems with using offline computed convective fluxes is a mis-match with archived convective precipitation. In the coupled runs the convection is also taken directly from IFS. The representation of the model winds has been updated to ensure mass conservation [33] which has made significant improvements to the resulting tracer fields, Bregman et al. [5].

Additional literature on the behavior of convective schemes can be found in van Noije et al. [38], which discusses the enhanced Brewer-Dobson circulation in ERA40, and in Bregman et al. [6], which focuses on the iterative time stepping in advection, as well as the impact of reduced grid at the poles and the impact of temporal interpolation of meteo.

The vertical diffusion is computed online in TM5 (see Olivié et al. [24] and Olivié [23]). This scheme is originally based on Louis [22], and improved by Holtslag and Boville [14]. Notice that the parameterizations are similar to those that have been employed in an earlier version of the ECMWF model. Olivié [24] show a comparison between offline computed diffussion coefficients with archived diffusion from ECMWF in the

ERA40 reanalysis. It was found that the ERA40 archived coefficients are very similar, but still give an overall slightly stronger diffusion. The offline computed diffusion is however very well able to reproduce the archived fields. The validation of the transport scheme in TM5 is given in Krol et al. [17].

The chemistry modeling in TM5

The modified CBM4 chemistry scheme

The chemical scheme in TM5 is based on a modified version of CBM4, Houweling et al. [15]. This mechanism is a lumped mechanism where non-methane organic species are included as functional groups in order to represent differences in reactivity. In combination with the Eulerian Backward Iterative (EBI) scheme Hertel et al. [13] this provides a computationally efficient method for accounting for chemical processing in a global context. Other species include NO^{-3} , SO_2^{-4} , and aerosols (included as 3 size segregated particle bins, Jeuken et al. [16], de Meij et al. [7]). In the current TM5 setup for the GEMS runs the aerosols do not have interaction with any of the other chemical parameters. The tracers NO, NO_2, NO_3, HNO_4, N_2O_5 make up NO_x. Notice that only NO_x is transported; the components of which it consists are only evaluated in the chemical scheme.

In total, in the tropospheric TM5 version applied in GEMS there are 55 individual tracers, with 39 being transported and 16 being non-transported. In the TM5-coupled mode 5 trace gases can be exchanged with IFS: CO, O3, CH2O, SO2, NOx. The transported aerosols include three bins of sea salt (number / mass density) and similarly two bins dust. Also two non-reactive trace gases are modelled, which can be used for transport-studies: rn222 and pb210. A comprehensive list of the chemical species is provided in table 1. During the GEMS project the chemistry scheme is updated in the later runs (version V5 and higher), according to the latest recommendations (e.g. JPL [31], [1]). A comprehensive description of the updated scheme with focus on the differences compared to the old scheme can be found in Williams and van Noije [40].

In earlier versions of the TM5 code, up to V5 and also in the early sensitivity studies the surface CH_4 concentration had been fixed to a single value of 1.76 ppm (following IPCC reference). This is relatively low for the 2003 reanalysis period. Starting from model version V6 the methane concentrations are scaled using a latitudinal surface distribution, which changes as a function of month, and slowly increases per year. The reference year is 1984; Methane surface concentrations growth rate for 1984-1999 is based on NOAA South Pole FLASK measurements. For 2000-2010 a growth rate of 0.5 % per year is assumed. The new surface concentrations reach up to about 1.9 ppm, depending on the latitude.

Stratosphere

At pressure levels higher than 50 hPa ozone concentrations are relaxed to the Fortuin-Kelder climatology [9], which is prescribed by a zonal and monthly mean ozone column measurements, with prescribed vertical distribution from an ozone climatology representative for the 1980s or 1990s. This is a monthly- based two-dimensional dataset (latitude, height). The modeled ozone is nudged to this climatology, such that the total ozone column in the stratosphere matches to that of the climatology. The stratospheric ozone has large impact on the radiation field in the troposphere. Therefore in a later model version (V10) the stratospheric ozone columns are updated to a more realistic values by the use of assimilated datasets based on GOME/SCIAMACHY (available for the years 1996-2007) or OMI (2005-present) measurements. For HNO3 the UARS climatology is applied and scaled to ozone. Concerning the other trace gases the same chemistry as for the troposphere is used. These considerations imply that the current (GEMS-)version of TM5 is a tropospheric model, rather than a tropospheric/stratospheric model.

Table 1: Table containing transported / non-transported chemical tracer in GEMS version of TM5

tracer name	transported	CBM-IV	comment
ozone chemistry			
O3	T		coupled to IFS
O3s	T		marked stratosph. ozone

NOx, NOy			
NOx	T		coupled to IFS
NO			
NO2			
NO3			
HNO4			
N2O5			
HNO3	T		
PAN	T		Peroxyacetyl nitrate and higher PANs
ammonium chemistry			
NH3	T		
NH4	T		
NH2			
HOx chemistry			
CH2O	T	c	coupled to IFS
H2O2	T		H-OOH
CH3O2H	T		CH3-OOH
CH3O2			CH3-OO.
OH			HO.
HO2			HOO.
lower carbon chemistry			
CH4	T		fixed at surface
CO	T		coupled to IFS
higher carbon chemistry			
par	T	c	paraffinic carbon atoms
eth	T	c	
Ole	T	c	
ald2	T	c	acetaldehyde and higher aldehydes

mgly	T	c	methylglyoxal
ROOH	T		lumped organic peroxide
ISOP	T		isoprene
orgntr	T		lumped organic nitrates, except PAN
C2O3			peroxyacetyl radical
ROR			
RXPAR			PAR budget corrector
XO2			NO to NO2 operator
XO2N			NO to alkyl nitrate operator
sulfur chemistry			
SO2	T		coupled to IFS
SO4	T		
DMS	T		di-methyl-sulfide
msa	T		methyl-sulfonic-acid
radio-active tracers			
Rn222	T		
Pb210	T		

Photolysis

Currently an offline parameterization of photolysis scheme is used that is based on the work of Landgraf and Crutzen [20] and Krol and van Weele [19]. This uses a loop-up table for the respective fluxes indexed with respect to temperature and pressure, using pre-defined atmospheres scaled to the respective ozone column. Coefficients are provided for the individual trace gases which are photolytically active. New scaling ratios are applied for orgntr and mgly in line with [40]. In general the parameters used for the absorption characteristics of (e.g.) NO_2 originate from the JPL 2000 recommendations, [30]. The effective photo-dissociation rates had been used for the contributions made by TM5 for the fourth assessment of the IPCC as well as the many intercomparison papers subsequently published (e.g. Stevenson et al. [35])

Deposition

Deposition fluxes are split in the contributions due to wet and dry deposition. Wet deposition by precipitation is a three-dimensional sink term whereas the dry deposition takes place at the surface only. In TM5 the dry deposition scheme based on Weseley [39] is applied. Detailed descriptions on the parametrization in TM5 can be found in Ganzeveld et al. [10]. The applied wet deposition scheme is outlined in Guelle et al. [11] (for aerosol scavenging) and Roelofs and Lelieveld [28] (for scavenging of soluble gases).

GEMS-emissions inventory

For the GEMS-GRG project the emissions that are applied in the different CTM's (MOZART, MOCAGE and TM5) have been aligned to give the same total global emission fluxes. The emissions are devided into three groups: the anthropogenic, biogenic (natural) and wildfire (biomass burning).

Anthropogenic and natural emissions

Anthropogenic emissions as applied in GEMS are mainly based on the RETRO inventory for the year 2000. Biogenic emissions (soil/oceans) are derived from GEIA (1985) and the ORCHIDEE model Latiére et al. [21]. The GEMS-emission inventory has a spatial resolution of 0.5 degree and a monthly time resolution. Note that some data had not been available on this resolution, but has been regridded. During the GEMS project the emission inventory as applied in TM5 has been updated several times, due to improvements in the emission inventory itself (for instance by using the REAS scaling factors for east-Asian emissions in version V10) and by performing bug-fixes (including biogenic CO emissions in V7, removing double-counted biofuel and biomass burning emissions for NO_x in V10). Most surface emissions are injected in the lowest two model layers.

Biomass burning from wildfires

Fire emissions from the GFED version 2, Randerson et al. [27] emission database are included, based on analysis of MODIS fire data, are included. The GFEDv2 emission inventory is available and used on a monthly time scale, as well as on an 8-day time scale. Total (natural, antropogenic and wildfire) emissions for the different compounds can be found in Schultz and Stein [32]. In the 60-layer version of TM5 as applied in the GEMS project the anthropogenic emissions are injected in the lowest two layers. The biomass burning emissions are distributed over different altitude ranges. Currently a single injection height distribution is applied globally. For the 2003 run the maximum injection height was set to 2000 meter, while for 2004 this was increased to 6000 m, to be compliant with the requests from HTAP.

Additional emissions

For aircraft NO_x emissions the ANCAT emission set is used, applied with a year-dependent scaling factor. This adds up to about 0.5 Tg(N)/year. In model version V10 this inventory has been replaced by the GREWE emissions (0.7 Tg(N)/year) from the RETRO project. NO_x production from lightning is calculated using a linear relationship between lightning flashes and convective precipitation. The total annual production is approximately 5 Tg(N)/yr.

Anthropogenic (incl. biomass burning) and soil/ocean emissions for NH_3 are based on van Aardenne et al. [37]. TM5 applies a refinement and distribution over months these using data from Bouwman et al. [4] and [3].

Anthropogenic emissions for SO_2 are based on IIASA, 2000-BAU, and are distributed over two layers. Biomass burning emissions based on GFED 2000. In V10 the GFED emissions for the appropriate model year are used. Emissions from volcanoes are based on Aerocom (distributed per grid cell between two heights).

Meteo fields

In the current TM5-IFS coupled system meteo-data is provided online through the coupler, standard on an hourly basis. The current timing of the coupled system is setup such that CTM-tendency data is provided at the start of the coupling interval. The IFS system provides its meteo-data to TM5 at half the coupling interval (i.e. after 30 min model time), which is valid for the full coupling period (which is standard one hour). For the convection data, mass fluxes and detrainment rates are coupled to TM5, which are then converted into TM5-variables (entrainment / detrainment updraft/downdraft). This replaces the parameterization scheme as outlined in section 3.

A complete list of the meteorological parameters used in the current version of TM5 is given in table 2. These parameters are used as input for parameterizations in TM5. The basic temporal treatment in the offline mode is to keep fields constant during 6 hourly intervals. Exceptions are surface pressure that should be interpolated at staggered 6 hourly resolution, and surface fields that should be constant during 3 or 24 hours. In the coupled system the meteo is provided on an hourly basis.

Table 2: Meteo-data from ECMWF used in TM5.

field	Unit	Resolution	Temporal	Coupled system	description
sp	Pa	3D glb3x2	6h	1h	surface pressure
mfu	kg/s	3D glb3x2	6h	1h	eastwards mass flux
mfv	kg/s	3D glb3x2	6h	1h	northwards mass flux
mfw	kg/s	3D glb3x2	6h	1h	upwards mass flux
T K	K	3D glb3x2	6h	1h	temperature
q	kg/kg	3D glb3x2	6h	1h	specific humidity
lwc	kg/kg	3D glb3x2	6h	1h	cloud liquid water content
iwc	kg/kg	3D glb3x2	6h	1h	cloud ice water content
cc	0-1	3D glb3x2	6h	1h	fractional cloud cover
entu	kg/m^2s	3D glb3x2	6h	1h	entrainment updraft
entd	kg/m^2s	3D glb3x2	6h	1h	entrainment downdraft
detu	kg/m^2s	3D glb3x2	6h	1h	detrainment updraft
detd	kg/m^2s	3D glb3x2	6h	1h	detrainment downdraft
oro	m^2/s^2	2D glb1x1	const	const	surface geopotential (orography)
lsm	%	2D glb1x1	const	const	land/sea mask
ci	0-1	2D glb1x1	24	24h	sea ice
sst	K	2D glb1x1	3h	24h	surface temperature
u10m	m/s	2D glb1x1	3h	1h	10m u wind
v10m	m/s	2D glb1x1	3h	1h	10m v wind
src	m	2D glb1x1	3h	1h	skin reservoir content
d2m	K	2D glb1x1	3h	1h	dewpoint temperature
t2m	K	2D glb1x1	3h	1h	2 meter temperature
sshf	W/m^2	2D glb1x1	3h	1h	surface sensible heat flux

slhf	W/m^2	2D glb1x1	3h	1h	surface latent heat flux
sstr	m/s	2D glb1x1	3h	1h	surface stress
cp	m/s	2D glb1x1	3h	1h	convective precipitation
lsp	m/s	2D glb1x1	3h	1h	large scale stratiform precipitation
ssr	W/m^2	2D glb1x1	3h	1h	surface solar radiation
sd	m	2D glb1x1	3h		snow depth
swvl1	m3/m3	2D glb1x1	3h	1h	volumetric soil water layer 1
tv	%	2D glb1x1	6h	archive	vegetation type
cvl	0-1	2D glb1x1	24h	archive	low vegetation cover
cvh	0-1	2D glb1x1	24h	archive	high vegetation cover
albedo	0-1	2D glb1x1	24h	1h	albedo
sr	m	2D glb1x1	24h	1h	surface roughness (ecmwf,ncep)
srols	m	2D glb1x1	monthly	archive	surface roughness (olsson)

Literature

1. R. Atkinson, D.L. Baulch, R.A. Cox, J.N . Crowley, R.F. Hamson, R.G. Hynes, M.E. Jenkin, M.J.
2. Rossi, and J. Troe. Evaluated kinetic, photochemical and heterogeneous data for atmospheric chemistry: gas phase reactions of organic species. Atmos. Chem. Phys., 6:3625 – 4055, 2006.
3. P. J. F. Berkvens, M. A. Botchev, W. M. Lioen, and J. G. Verwer. A zooming technique for wind transport of air pollution. In R. Vilsmeier, D. Hanel, and F. Benkhaldoun, editors, Finite Volumes for Complex Applications, Proceedings of the 2nd International Symposium on Finite Volumes for Complex Applications (Duisburg, Germany, July 19-22, 1999), pages 499–506. Hermes Publisher, 1999.
4. A.F. Bouwman, L.J.M. Boumans, and N.H. Batjes. Estimation of global nh3 volatilization loss from synthetic fertilizers and animal manure applied to arable lands and grasslands. Global Biogeochem. Cycles, 16(2):1024, 2002.
5. A.F. Bouwman, D.S. Lee, W.A.H. Asman, F.J. Dentener, K.W. Van Der Hoek, and J.G.J Olivier. A global high-resolution emission inventory for ammonia. Global Biogeochem. Cycles, 11(4):561–588, 1997.
6. A. Bregman, A.J. Segers, M. C. Krol, E. Meijer, and P. van Velthoven. On the use of mass-conserving wind fields in chemistry–transport models. Atmos. Chem. Phys., 3(2):447–457, 2003.
7. B. Bregman, E.Meijer, and R. Scheele. Key aspects of stratospheric tracer modeling using assimilated winds. Atmos. Chem. Phys., 6(12):4529–4543, october 2006. Paper in ACP.
8. A. de Meij, P. Thunis, C. Cuvelier, E. Vignati, F. Dentener, and M. Krol. The sensitivity of aerosol in europe to two different emission inventories and temporal distribution of emissions. Atmos. Chem. Phys., 6:4287–4309, 2006.
9. F. Dentener, M. van Weele, M. Krol, S. Houweling, and P. van Veldhoven. Trends and inter-annual variablity of methane emissions derived from 1979-1993 global ctm simulations. Atmos. Chem. Phys., 3:73–88, 2003. Paper in ACP.

10. J.P.F. Fortuin and H. Kelder. An ozone climatology based on ozonesonde and satellite measurements. J. Geophys. Res., 103:31709–31734, 1998. Paper in JGR.
11. L. Ganzeveld, J. Lelieveld, and G.-J. Roelofs. A dry deposition parameterization for sulfur oxides in a chemistry and general circulation model. J. Geophys. Res., 103(D5):5679–5694, 1998. doi:10.1029/97JD03077.
12. W. Guelle, Y. J. Balkanski, M. Schulz, F. Dulac, and P. Monfray. Wet deposition in a global size-dependent aerosol transport model, 1, Comparison of a 1 year Pb^{210} simulation with ground measurements. J. Geophys. Res., 103(D10):11429–11446, 1998. doi:10.1029/97JD03680.
13. M. Heimann, P. Monfray, and G. Polian. Long-range transport of 222Rn – a test for 3D tracer models. Chem. Geol., 70:98–98, 1988.
14. O. Hertel, R. Berkowicz, and J. Christensen. Test of two numerical schemes for use in atmospheric transport-chemistry models. Atmos. Environ., 27A,16:2591–2611, 1993.
15. A. A. Holtslag and B. A. Boville. Local versus nonlocal boundary-layer diffusion in a global climate model. J. Climate, 10:1825–1842, 1993.
16. S. Houweling, F. Dentener, and J. Lelieveld. The impact of non-methane hydrocarbon compounds on tropospheric photochemistry. J. Geophys. Res., 103(D9):10673–10696, 1998. doi:10.1029/97JD03582.
17. A. Jeuken, J.P. Veefkind, F. Dentener, S. Metzger, and C.R. Gonzalez. Simulation of the aerosol optical depth over europe for august 1997 and a comparison with observations. J. Geophys. Res., 106(D22):28295–28312, 2001.
18. M.C. Krol, S. Houweling, B. Bregman, M. van den Broek, A.J. Segers, P. van Velthoven, W. Peters, F. Dentener, and P. Bergamaschi. The two-way nested global chemistry–transport zoom model TM5: algorithm and applications. Atmos. Chem. Phys., 5(2):417–432, February 2005.
19. M.C. Krol, J. Lelieveld, D.E. Oram, G.A. Sturrock, S.A. Penkett, C.A.M. Brenninkmeijer, V. Gros, J. Williams, and H.A. Scheeren. Continuing emissions of methyl chloroform from Europe. Nature, 421:131–135, January 2003.
20. M.C. Krol and M. van Weele. Implications of variations in photodissociation rates for global tropospheric chemistry. Atmos. Environ., 31:1257–1273, 1997.
21. J. Landgraf and P.J. Crutzen. An efficient method for online calculations of photolysis and heating trates. J. Atmos. Sci., 55:863–878, 1998.
22. J. Lathiére, D. Hauglustaine, N. De Noblet-Ducoudré, G. Krinner, and G.A. Folberth. Past and future changes in biogenic volatile organic compound emissions simulated with a global dynamic vegetation model. Geophys. Res. Lett., 32:L20818, 2004.
23. J. F. Louis. A parametric model of vertical eddy fluxes in the atmosphere. Boundary Layer Meteorology, 17:187–202, 1979.
24. D. J. L. Olivié. On the role of convection and turbulence for tropospheric ozone and its precursors. PhD thesis, Technical University of Eindhoven, The Netherlands, March 2005.
25. D. J. L. Olivié, P. F. J. van Velthoven, and A. C. M. Beljaars. Evaluation of archived and off-line diagnosed vertical diffusion coefficients from ERA-40 with ^{222}Rn simulations. Atmos. Chem. Phys., 4(9/10):2313–2336, 2004.
26. A. C. Petersen, E. J. Spee, H. van Dop, and W. Hundsdorfer. An evaluation and intercomparison of four new advection schemes for use in global chemistry models. J. Geophys. Res., 103(D15):19,253–19,269, 1998.
27. M. Prather. Numerical advection by conservation of second-order moments. J. Geophys. Res., 91:6671–6681, 1986.
28. J.T. Randerson, G. R. van der Werf, G .J. Collatz, L. Giglio, C. J. Still, P. Kasibhatla, J. B. Miller, J. W. C. White, R. S. DeFries, and E. S. Kasischke. Fire emissions from c3 and c4 vegetation and their influence on interannual variability of atmospheric co2 and d13co2. Global Biogeochem. Cycles, 19:GB2019, 2005.
29. G.-J. Roelofs and J. Lelieveld. Distribution and budget of o_3 in the troposphere calculated with a chemistry-general circulation model. J. Geophys. Res., 100:20983–20998, 1995.
30. G. L. Russell and J. A. Lerner. A new finite-differencing scheme for the tracer transport equation. J. Appl. Meteorol., 20:1483–1498, 1981.
31. S.P. Sander, R.R. Friedl, W.B. DeMore, A.R. Ravishankara, D.M. Golden, C.E. Kolb, M.J. Kurylo, R.F. Hampson, R.E. Huie, M.J. Molina, and G.K. Moortgart. Chemical kinetics and photochemical data for use in atmospheric studies, evaluation no. 13. Technical report, JPL, 2000.
32. S.P. Sander, R.R. Friedl, A.R. Ravishankara, D.M. Golden, C.E. Kolb, M.J. Kurylo, M.J. Molina, G.K. Moortgart, H. Keller-Rudek, B.J. Finlayson-Pitts, P.H. Wine, R.E. Huie, and V.L. Orkin. Chemical kinetics and photochemical data for use in atmospheric studies, evaluation no. 15. Technical report, JPL, 2006.
33. M. Schultz and O. Stein. Gems (grg) emissions for 2003 reanalysis simulations. Technical report, MPI-M, 2006.

34. A.J. Segers, P. van Velthoven, B. Bregman, and M.C. Krol. On the computation of mass fluxes for Eulerian transport models from spectral meteorological fields. In P.M.A. Sloot, C.J. Kenneth Tan, J.J. Dongarra, and A.G. Hoekstra, editors, Proceedings of the 2002 International Conference in Computational Science, volume 2330/2002 of Lecture Notes in Computer Science (LNCS), p. 767–776. Springer Verlag, april 2002.
35. P. Spiro, D. Jacob, and J. Logan. Global inventory of sulfur emissions with 1 x 1 resolution. J. Geophys. Res., 97:6023–6036, 1992.
36. D. S. Stevenson et al. Multi-model ensemble simulations of present-day and near-future tropospheric ozone. J. Geophys. Res., 111:D08301, 2006. doi:10.1029/2005JD006338.
37. M. Tiedtke. A comprehensive mass flux scheme for cumulus parameterization in large-scale models. Mon. Weather. Rev., 117(8):1779–1800, 1989.
38. J. van Aardenne, J.F. Dentener, J.G.J. Olivier, C.G.M. Klein Goldewijk, and J. Lelieveld. A 1 x 1 resolution dataset of historical anthropogenic trace gas emissions for the period 1890 - 1990. Global Biogeochem. Cycles, 15(4):909 – 928, 2001.
39. T. P. C. van Noije, H. J. Eskes, M. van Weele, and P. F. J. van Velthoven. Implications of the enhanced Brewer-Dobson circulation in European Centre for Medium-Range Weather Forecasts reanalysis ERA-40 for the stratosphere-troposphere exchange of ozone in global chemistry transport models. J. Geophys. Res., 109(D19):D19308, October 2004. doi:10.1029/2004JD004586.
40. M.L. Weseley. Parameterization of surface resistance to gaseous dry deposition in regional numerical models. Atmos. Env., 16:1293–1304, 1989.
41. J.E. Williams and T.C.P. van Noije. On the upgrading of the modified carbon bond mechanism iv for use in global chemistry transport models. Technical Report TR-2008-02, KNMI, De Bilt, The Netherlands, may 2008.

ANNEX 4: Model documentation of MOCAGE with a focus on GEMS settings

Philippe Moinat, May 15, 2009

Introduction

The MOCAGE 3D multi-scale Chemistry and Transport Model has been designed for both research and operational applications in the field of environmental modelling. Since 2000, MOCAGE allows to cover a wide range of topical issues: chemical weather forecasting, tracking and backtracking of accidental point source releases, trans-boundary pollution assessment, assimilation of remote sensing measurements of atmospheric composition, studies of the impact of anthropogenic emissions of pollutants on climate change, with over 20 references in the international literature.

MOCAGE is able to consider several domains simultaneously, allowing to "zoom" on some particular areas of the globe. Extending from the surface up to the middle or upper stratosphere (depending on the vertical levels chosen) and the main model domain being global, MOCAGE is able to provide its own time-dependent chemical boundaries over "zoom" domains.

At Météo-France, MOCAGE has been daily run since 6 years. During the 2003 heatwave, it provided 3-day air quality forecasts over Europe showing that ozone peak events overlap a large part of France and of Western Europe. In 2004, Météo-France joined the operational platform "Prév'Air" (http://www.prevair.org) in charge of the pollution monitoring and forecasting over France, lead by the Ministery of Environment. MOCAGE also provides operational UV index forecasts since 2002.

MOCAGE includes a comprehensive data assimilation package via the coupling with the PALM software developed at CERFACS (Massart et al. 2005), allowing for variational assimilation of in-situ or profile chemical data. MOCAGE-PALM performed well in the first international intercomparison of ozone profile analyses (Geer et al., 2006), organised during the ASSET european project (Lahoz et al., 2007).

MOCAGE has originally been developed and optimized for the operational super-computers of Météo-France. Since recently, these computers had a vector architecture and for this reason MOCAGE has been highly vectorized. At the beginning of the GEMS project, a great effort has been accomplished in order to introduce some parallelization and reach more acceptable performances on the ECMWF parallel computer. But as this parallelization is only partial, the performances of MOCAGE at ECMWF remain insufficient for an extensive use. For this reason, MOCAGE stand-alone runs have been performed on Météo-France computer and latter transferred at ECMWF while the coupled runs have covered only short periods.

In the GEMS GRG project, MOCAGE was used in the CTM mode without data assimilation, on the global domain only, forced by external meteorological data obtained from IFS reanalysis or forecast, read from files or communicated through the coupling software OASIS4. The same version of MOCAGE is used for the stand-alone runs and for the coupled model but with a particular CPP key activated at the compilation in order to include the coupling definitions and routines. In this case, MOCAGE execution is synchronized with the IFS by the coupler in order to allow the various data exchanges at the appropriate times.

Horizontal grid and vertical levels

The horizontal grid used in the GEMS GRG project is a regular latitude/longitude grid with a 2° resolution. All the input fields (meteorological forcing, emissions, deposition velocities…) have been projected on this grid. No sub-domain has been considered in the GRG runs but a fine resolution sub-domain covering Europe has been used in the GEMS Regional Air Quality subproject.

In the vertical, two different level distributions have been used in GRG. In the first MOCAGE run (identified as 47LEV), 47 vertical levels were used, which is the present operational version of MOCAGE. The 47 hybrid (s,P) levels go from the surface up to 5 hPa, with approximately 7 levels in the Planetary Boundary Layer (PBL), 16 in the free troposphere and 24 in the stratosphere. In this case, a monthly climatology of chemical compounds is applied above 4 hPa, in order to take into account the Brewer-Dobson circulation in the upper stratosphere.

In the most recent MOCAGE runs (60LEV01, 60LEV02 and HTAP) which have been more extensively evaluated, the domain is extended up to the lower mesosphere, up to 0.07 hPa (approximately 70 km) and 60 vertical levels were used with the same distribution as employed by the ECMWF model. In this case, a simple zero-gradient boundary condition is applied at the top.

Time stepping

The CTM MOCAGE is driven by meteorological inputs (winds, temperature, humidity, and pressure) provided by external NWP models, such as the analyses or forecasts of the ECMWF system. When run in off-line mode, the meteorological fields are read from files every 3 or 6 hours and are then linearly interpolated to yield 1-hour values, which is the time step for advection. When run in coupled mode, MOCAGE can receive hourly meteorological data from the NWP running in parallel.

While the meteorological data are kept constant during each 1-hour step, shorter time steps are used for physical processes and chemistry: 15 minutes for the emission, deposition and chemistry models and 30 minutes for the parameterized transport (convection and diffusion). The chemical scheme solves the various timescales of the chemical processes by fractioning the 15 minute time step and using an iterative method. Concerning the physical and chemical parameterizations, an operator splitting approach is used. Parameterizations are called alternatively in forward and reverse order, with the objective to reduce systematic errors.

Meteorological forcing and advection scheme

The meteorological forcing

The CTM MOCAGE needs winds, temperature, humidity and pressure to feed the advection scheme, as well as the physical and chemical parameterizations. A procedure has been established in order to retrieve and use in MOCAGE the reanalysis or forecasts products provided by the ECMWF operational Numerical Weather Prediction model (IFS) and stored on the MARS archive. The meteorological forcing fields are projected on the horizontal grid used by MOCAGE.

The 3D pressure at the model levels is easily computed by MOCAGE from the surface pressure using the vertical coordinate parameters.

The surface pressure, temperature, specific humidity and horizontal winds are directly derived from the quantities processed by the NWP model but getting the W field (Pa/s) requires some calculations based on the continuity equation.

Two approaches have been used during the GEMS project. In the first approach adopted for the 47LEV and 60LEV01 MOCAGE runs, the calculation makes use of the divergence of the horizontal winds directly provided by the NWP and of the horizontal pressure gradients accurately computed in

the spectral space before being transformed as grid point fields. The W field computed this way is very close to the one available from the MARS archive and computed by ECMWF routines. In the second approach adopted for the 60LEV02 and HTAP runs, all the computations are based on the retrieved grid-point U, V and surface pressure fields at the MOCAGE horizontal resolution. This latter approach is expected to be less accurate but more consistent with the numerical approximations made by MOCAGE.

The W fields obtained by the two methods are very close at the top levels but the differences become noticeable close to the ground. This is normal since W at a given model level is obtained by a top-down vertical integration from the top to this level which results in error accumulation at the bottom levels. Differences are also stronger close to the poles due to the difficulty to accurately compute the divergence in these regions.

These differences on W have an impact on the results produced by MOCAGE. For example, the mixing due to the vertical advection can be different close to the ground where most species are emitted. The parameterized transport can also be changed as the triggering criteria include the W values.

In the coupled MOCAGE-IFS model, MOCAGE receives the surface pressure, temperature, specific humidity and horizontal winds from the IFS through the coupling software OASIS4. But W is not available and has to be computed with the second approach described above using the received grid point fields.

The advection scheme

Large scale transport of chemical constituents or tracers is ensured by a semi-Lagrangian advection scheme (Williamson and Rasch, 1989) which has two main advantages in term of computation efficiency: the scheme permits large time steps without stability restrictions and, on the contrary to Eulerian schemes, it allows to treat simultaneously a large number of tracers since back-trajectories are computed once for all chemical species. But the drawback is that the scheme used does not conserve mass, as soon as the grid is irregular. In MOCAGE, a global uniform correction is applied, considering that the whole mass of tracer has to remain constant during the advection process.

Chipperfield (2006) indicated that many CTMs seem to give reasonable simulations without concern over mass conservation in advection schemes or the balance of winds but on the other hand several studies show possible spurious mass transfers either due to non conservation or improper correction (see for example Bregman et al., 2003).

Physical parameterizations

Turbulent diffusion

Turbulent mixing is treated in MOCAGE following (Louis, 1979). Horizontal diffusion is neglected, while the vertical diffusion coefficient K depends on height, wind shear and atmospheric stability. The more unstable the atmosphere, the greater K. The model has been slightly modified to allow a deeper application within the boundary layer, in an attempt to enhance the exchanges between the surface and the upper boundary layer.

Convection and precipitations

The Bechtold et al. (2001) mass-flux type convection scheme has been adopted in MOCAGE after a number of studies and evaluations (details in Josse et al., 2004). Bechtold's convection is based on a mass flux scheme, where downdrafts are taken into account, as well as freezing and melting. All computations are one-dimensional. To trigger (or not) convection in a column, a mixed air parcel is lifted from the ground to its lifting condensation level. If the difference between its virtual temperature and that of environment is sufficiently high, then convection can be triggered off. The ability of the parcel to produce sufficient cloud depth is added to this condition. Shallow convection shall give at least a 500m high cloud, and deep convection shall extend on 3 km.

Fractional entrainment and detrainment rates are set constant. The thermodynamic characteristics of the updraft are computed assuming that, except from precipitation processes, enthalpy and total water mixing ratio are conserved. Finally, the intensity of the convection is controlled by a closure assumption. It is based on the removing of all Convective Available Potential Energy during an adjustment period, set to 3 hours for shallow convection and between 0.5 and 1 hour for deep convection.

Wet removal by precipitation is included in MOCAGE. In convective clouds, it is parameterized according to Mari et al. (2000); convective transport and scavenging are therefore computed simultaneously. In large-scale precipitation clouds, removal follows the first-order scheme of Giorgi and Chamedeis (1986). Below clouds, the recommendations of Liu et al. (2001) (again a first-order scheme) are used. Wet removal has been evaluated with simulations of 210Pb, a highly soluble tracer, by comparing model outputs with both climatologies and fine temporal resolution observations. The 3D large-scale precipitations fluxes are reconstructed from the 2D precipitation fluxes at ground read from the surface files and obtained from the Météo-France Arpège NWP model output.

Chemistry model

The chemistry is described with the RACMOBUS scheme which includes 118 species and 350 thermal reactions. 89 of these species are transported (tracer species) while the remaining 29 are assumed at instantaneous chemical equilibrium (short lived species), as described in Brasseur and Solomon (1986). Except for the water content, RACMOBUS makes no explicit distinction between the troposphere and the stratosphere, the activation of a reaction at a given location depending only on the availability of the species. The same model with the same species and same reaction rates is used whatever the model resolution is.

The RACMOBUS scheme is in fact a combination of the tropospheric RACM scheme of Stockwell et al. (1997) and of the stratospheric REPROBUS scheme of Lefèvre et al. (1994) that includes the heterogeneous stratospheric chemistry on Polar Stratospheric Clouds described in Carslaw et al. (1995).

Radiation is taken into account by MOCAGE via photolysis rates. These photolysis rates have been computed off-line from the solar radiation with the Tropospheric Ultraviolet-Visible model version 4.0 (see Madronich and Flocke, 1998). The impact of clouds on the photolysis rates is calculated on-line according to Brasseur et al. (1998), increasing photolysis rates above clouds and weakening it below.

At the crossroads between dynamics, physics and chemistry, MOCAGE uses the water vapour mixing ratios provided by the meteorological forcing model up to the 380 K isentropic level. Above this level, H2O is calculated by the chemical scheme of MOCAGE and advected by its semi-lagrangian transport scheme. Prescribing the water vapour field between the surface and the 380 K level allows MOCAGE to benefit from the advanced modelling of the physical processes in the troposphere and the UTLS region included in the NWP model.

In all MOCAGE GRG runs except for the very first one, the methane mixing ratio has been set to a constant value (1.76 ppm, following IPCC reference) throughout the troposphere while it results from the chemical scheme in the upper layers. A 150 ppb ozone mixing ratio criterion is used to locate the tropopause.

Surface processes

All surface data are read from files prepared off-line, including the model orography, the land/sea mask, the precipitation fluxes used to reconstruct the 3D large-scale precipitation fluxes and of course the dry deposition velocities and surface emissions. The meteorological data provided through these files or used to evaluate some the quantities come from ARPEGE (Météo-France GCM model) reanalysis.

A description of MOCAGE surface exchanges module is presented in Michou and Peuch, 2002, and in Michou et al., 2004.

Dry deposition velocities

The dry deposition velocities are computed on the basis of a modified Wesely (1989) "big-leaf" resistance approach (Michou and Peuch, 2002), from three resistances in series, aerodynamic, laminar, and surface. The original surface resistance scheme was modified with the introduction of a specific parameterization for the stomatal resistance depending upon environmental factors; it is based on Noilhan and Mahfouf (1996), and follows the Jarvis-type meteorological approach that attempts to modify a minimum stomatal resistance defined a priori through external factors, such as moisture and radiation availability.

Emissions

MOCAGE need emissions for the following species:

ALD	RETRO voc 22 (CH3CHO and other alkanals)	KET	RETRO voc 23 (acetone)
CO	RETRO	NO	RETRO 1/10*NOx
CSL	RETRO voc 17 (Other aromatics)	NO2	RETRO 9/10*NOx
ETE	RETRO voc 7 (C2H4)	NOx	RETRO
ETH	RETRO voc 2 (C2H6)	OLI	RETRO voc 12 (Other alkenes)
HC3	RETRO voc 3 (C3H8) + voc 4 (Butanes) + voc 9 (C2H2)	OLT	RETRO voc 8 (C3H6)
HC5	RETRO voc 5 (C5H12)	TOL	RETRO voc 13 (Benzene) + voc 14 (Toluene)
HC8	RETRO voc 6 (Hexanes and higher)	XYL	RETRO voc 15 (Xylene) + voc 16 (Trimethylbenzene)
HCHO	RETRO voc 21		

API	IPCC	DMS	IPCC
CFC11	GEIA	ISO	GEIA
CFC12	GEIA	LIM	IPCC
CFC113	EDGAR	N2O	GEIA
CH3CCl3	EDGAR	SO2	IPCC
CH4	IPCC	SULF	IPCC

The table also indicates the emission datasets used for each specie and the way some emission data are lumped together for a given MOCAGE specie.

The RETRO dataset used for most of the species corresponds to the anthropogenic and biogenic emissions provided to the GEMS GRG group by O.Stein and M.Schultz and is based on the emission dataset completed for the year 2000 in the frame of the RETRO project (http://retro.enes.org). The GEIA, IPCC and EDGARD inventories are used for the other species (see Michou and Peuch, 2002, for details). The RETRO datasets include the GFED-v2 wildfire emissions.

Several emissions sources are considered

- inc, Fossil fuel use / Industry (hourly profiles)
- exf, Fossil fuel use / Production and distribution
- tra, Fossil fuel use / Transport road (hourly profiles)
- pow, Fossil fuel use / Other
- res, Fossil fuel use / Other
- ships, Fossil fuel use / Other
- sol, Industrial processes / Solvents
- agr, Landuse and waste treatment / Agriculture
- soil, Natural sources / soils and animals
- was, Landuse and waste treatment / Waste treatment
- biogenic, Natural sources / continents and oceans
- ocean, Natural sources / continents and oceans
- aircraft, Fossil fuel use / Other
- wildfires, Landuse and waste treatment / Biomass burning

but all these sources are finally mixed together into the MOCAGE input emission files, at the exception of the aircraft and wildfire emissions when injected in altitude. A time profile accounting for the daily variability is applied to some of the sources.

The version of MOCAGE used in GEMS GRG does not include NOx emissions from lightning (total estimated to 7 Tg(N) per year). A parameterization of these emissions has been recently implemented in MOCAGE and first results reported in Bousserez et al. (2007).

The RETRO and other emission data sets used in all MOCAGE stand-alone runs for the year 2003 describe monthly averaged emissions. But, the GFED-v2 wildfire emissions with 8 day variability (http://www.geo.vu.nl/users/gwerf/GFED/data/) have also been implemented and used for the 2004 HTAP runs.

Emissions are not only applied in the surface layer but distributed in the eight lowest levels of the model (that correspond on average to a layer of 600 m), in order to avoid too strong vertical gradients, as proposed in Josse et al. (2004).

Wildfire emissions receive a particular treatment: they are distributed in the vertical according to the vertical profiles available from ftp://ftp.retro.enes.org/pub/emissions/aggregated/fires/ (Rast et al., submitted to JGR). These biomass burning injection profiles depend on the vegetation type (wood, forest, grass). For example, in the case of boreal forest fires, the emissions profiles extend up to an altitude of approximately 9 km.

Model outputs

MOCAGE reads (restart file, meteorological data, surface data) and produces files (restart files) in the Arpege/Aladin format. At the beginning of the GEMS project, the possibility to output a selection of fields at a defined time step in standard netcdf files (CF-1.0 convention) has been added. Early in the GRG project a common model output format for evaluation of the CTM runs has been defined (Tasks 4.3 and 4.4) in order to facilitate the evaluation process and make sure that all data needed for the evaluation would be available.

The MOCAGE outputs have been adapted to these specifications and the following files can be directly produced by a MOCAGE run, all outputs being specified through a namelist.:

- Model grid definition
- 3D 3-hourly fields for Thermo-dynamical Parameters
- 3D 3-hourly fields for a selection of Chemistry Species (volume mixing ratios)
- 3D 1-hourly fields for a selection of Chemistry Species (volume mixing ratios)
- Daily dry and wet deposition fields for a selection of Chemistry Species.
- Ozone budgets including chemical production and destruction and surface deposition
- 3h (or 1h) Tendency fields due to various processes (emission-diffusion-deposition, chemistry production, chemistry loss, convection)

In the coupled model, some of the computed fields are sent to the IFS via the OASIS4 coupler at a defined time step. In addition to the concentrations of a short list of species (ozone, NOx, SO2, CO, CH2O), several hourly tendencies can be sent, individually or combined together, as specified in the MOCAGE name list: the total tendency, the tendencies due to dry deposition, convection and diffusion, emissions.

In addition to the output files described above, dry deposition velocity fields can be extracted from the surface files provided as an input for MOCAGE. Moreover, several post-processing scripts and programs have been developed in order to extract additional files required for the comparisons with the observations from the MOCAGE output files:

- Monthly model output (mixing ratios & meteo fields)
- Surface station data
- Profile data (ozone sondes, MOZAIC ascents/descents above airports, …)
- MOZAIC data and DUCHESS data (records along flight paths)

The evaluators have also developed their own scripts to extract other specific outputs, for example for comparing MOCAGE results to satellite data.

Lately, new post-processing scripts have been developed in order to produce similar plots as the ones proposed by the evaluators and shown in the GRG evaluation reports, for example:

- Ozone and other species zonal mean at selected levels (pressure or model) time evolution
- Ozone and other species total/tropospheric/stratospheric columns zonal mean time evolution
- Seasonality curves of regional means (Europe, North America…) of CO, NO2 and other species total column or value at selected levels with standard deviation.

References

1. Bechtold P., E. Bazile, F. Guichard, P. Mascart, and E. Richard, A mass flux convection scheme for mesoscale and global models, Quart. J. Roy. Meteor. Soc., 126, 865-889, 2001.
2. Bousserez, N., J.-L. Attié, V.-H. Peuch, M. Michou, G. Pfister, D. Edwards, M. Avery, G. Sachse, E. Browell and E. Ferrare, 2007: Evaluation of MOCAGE chemistry and transport model during the ICARTT/ITOP experiment, J. Geophys. Res., 112, D10S42, doi:10.1029/2006JD007595.
3. Brasseur, G. P. and Solomon, S. , 1986: Aeoronomy of the middle atmosphere, D. Reidel Publishing Company, 452 pp..
4. Brasseur, G. P., Hauglustaine, D. A., Walters, S., Rasch, P. J., Müller, J.-F., Granier, C., and Tie, X. X., 1998: MOZART, a global chemical transport model for ozone and related chemical tracers. 1. Model description, J. Geophys. Res., 103, 28 265–28 289.
5. Bregman B., A. Segers, M. Krol, E. Meijer, P. Va Velthoven, 2003: on the use of mass-conserving wind fields in chemistry-transport models, Atmos.Chem.Phys., 3, 447-457.
6. Carslaw, K. S., Luo, B., Peter, T., and Clegg, S. L. , 1995: Vapour pressures of H2SO4/HNO3/HBr/H2O solutions to low stratospheric temperatures, Geophys. Res. Lett., 22, 247–250.
7. Chipperfield M. P. , 2006: New version of the TOMCAT/SLIMCAT offline chemical transport model: Intercomparison of stratospheric tracer experiments, Q. J. Roy. Meteor. Soc. , 132, 1179-1203.
8. Dufour, A., M. Amodei, G. Ancellet and V.-H. Peuch, 2004 : Observed and modelled "chemical weather" during ESCOMPTE, Atmos. Res., 74 (1-4), 161-189.
9. El Amraoui, L., V.-H. Peuch, P. Richaud, S. Massart, J. Urban, N. Semane, H. Teyssedre, D. Cariolle and F. Karcher, 2008: Ozone loss in the 2002/03 Artic vortex deduced from the assimilation of O3 and N2O measurements: N2O as a dynamical tracer, Q. J. R. Meteorol. Soc, 134, 217-228.
10. El Amraoui L., N. Semane, V.-H. Peuch and M. L. Santee, 2007 : Investigation of dynamical and chemical processes in the polar stratospheric vortex during the unusually cold winter 2004/2005. Geophysical Research Letters 35(3): L03803.
11. Geer, A.J. , W.A. Lahoz, S. Bekki, N. Bormann, Q. Errera, H.J. Eskes, D. Fonteyn, D.R. Jackson, M.N. Juckes, S. Massart, V.-H. Peuch, S. Rharmili and A. Segers, 2006 : The ASSET intercomparison of ozone analyses : method and first results, Atmos.Chem.Phys., 6, 5445-5474.
12. Giorgi, F. and Chamedeis, W. L. , 1986: Rainout lifetimes of highly soluble aerosols and gases as inferred from simulations with a general circulation model, J. Geophys. Res., 91, 14 367–14 376.
13. Josse B., Simon P. and V.-H. Peuch, 2004 : Rn-222 global simulations with the multiscale CTM MOCAGE, Tellus, 56B, 339-356.
14. Lahoz W.A., A.J. Geer, S. Bekki, N. Bormann, S. Ceccherini, H. Elbern, Q. Errera, H.J. Eskes, D. Fonteyn, D.R. Jackson, B. Khattatov, S. Massart, V.-H. Peuch, S. Rharmili, M. Ridolfi, A. Segers, O. Talagrand, H.E. Thornton, A.F. Vik et T. Von Clarman, 2007 : The Assimilation of Envisat data (ASSET) project, Atmos. Chem. Phys., 7, 1773-1796.
15. Lefbèvre, F., Brasseur, G. P., Folkins, I., Smith, A. K. and P. Simon, Chemistry of the 1991-1992 stratospheric winter: three-dimensional model simulations, J. Geophys. Res., 1994, 99(D4), 8183-8195.
16. Liu, H., Jacob, D. J., Bey, I., and Yantosca, R. M. , 2001: Constraints from 210Pb and 7Be on wet deposition and transport in a global three dimensional chemical-transport model driven by assimilated meteorological fields, J. Geophys. Res., 106, 12 109–12 128.
17. Louis J.F., A parametric model of vertical eddy fluxes in the atmosphere, Bound.-Layer Meteor., 1979, 17, 197-202.
18. Madronich, S. and Flocke, S. , 1998: The role of solar radiation in atmospheric chemistry, in: Handbook of Environmental Chemistry, edited by: Boule, P., Springer-Verlag, New York, 1–26.

19. Mari, C., Jacob, D. J., and Bechtold, P. , 2000: Transport and scavenging of soluble gases in a deep convective cloud, J. Geophys. Res., 105, 22 255–22 267.
20. Massart, S., D. Cariolle and V.-H. Peuch, 2005 : Vers une meilleure représentation de la distribution et de la variabilité de l'ozone atmosphérique par l'assimilation des données satellitaires, C. R. Géosciences, doi:10.1016/j.crte.2005.08.001.
21. Michou M. et V.-H. Peuch, 2002 : Surface exchanges in the multi-scale chemistry and transport model MOCAGE, Water Sci. Rev., 15 special issue, 173-203.
22. Michou M., P. Laville, D. Serça, A. Fotiadi, P. Bouchou and V.-H. Peuch, 2004 : Measured and modeled dry deposition velocities over the ESCOMPTE area, Atmos. Res., 74 (1-4), 89-116.
23. Noilhan, J. and Mahfouf, J.-F. , 1996: The ISBA land surface parameterisation, Glob. Plan. Change, 13, 145–159.
24. Semane, N., V.-H. Peuch, L. El Amraoui, H. Bencherif, S. Massart, D. Cariolle, J.-L. Attié and R. Abida, (2007), An observed and analysed stratospheric ozone intrusion over the high Canadian Arctic UTLS region during the summer of 2003, Q. J. R. Meteorol. Soc, 113, 171-178.
25. Stockwell, W. R., Kirchner, F., Khun, M., and Seefeld, S. , 1997: A new mechanism for regional atmospheric chemistry modelling, J. Geophys. Res., 102, 25 847–25 879.
26. Teyssèdre H., M. Michou, H. L. Clark, B. Josse, F. Karcher, D. Olivié, V.-H. Peuch, D. Saint-Martin, D. Cariolle, J.-L. Attié, P. Nédélec, P. Ricaud, V. Thouret, R. J. van der A, A. Volz-Thomas, and F. Chéroux, 2007: A new tropospheric and stratospheric Chemistry and Transport Model MOCAGE-Climat for multi-year studies: evaluation of the present-day climatology and sensitivity to surface processes, Atmos. Chem. Phys., 7, 5815-5860.
27. Wesely, M. L. , 1989: Parameterization of surface resistances to gaseous dry deposition in regional-scale numerical models, Atmos. Environ., 23, 1293–1304.
28. Williamson, D. L. and P. J. Rash. Two-dimensional semi-lagrangian transport with shape-preserving interpolation, Mon. Wea. Rev.,1989, 117, 102-129.

ANNEX 5: Coupled system description

Documentation of IFS coupled system

Johannes Flemming

ECMWF, Reading, UK

email: Johannes.Flemming@ecmwf.int

Description of the Coupled System

Forecast and assimilation of global reactive gases are generated by a two-way coupled system, which links the IFS to any of the global chemistry transport models (CTM) MOCAGE, MOZART-3 or TM5 (version KNMI-cy3-GEMS).

Our development effort in GEMS was to upgrade the IFS with more complex non-linear chemistry schemes for troposphere and stratosphere. In particular, the IFS should be enabled to also simulate tropospheric O_3, carbon monoxide (CO), nitrogen oxides (NO_x), formaldehyde (HCHO) and sulphur dioxide (SO_2) because these species – besides playing a key role in atmospheric chemistry – are observable from space. The respective satellite observations could therefore be assimilated by the IFS.

The idea of the coupled system is that the IFS computes only the transport of the reactive gases while the tendencies due to chemical conversion, deposition and emission injection are provided by the coupled CTM. The CTM itself is driven by meteorological data which are transferred at high temporal resolution from the IFS to the CTM. For example, the call of a subroutine for chemical conversion in an integrated chemistry-global-circulation-model code is substituted in the coupled system by a call to the coupler software requesting the respective total tendencies from the CTM. The tendencies are then applied to the concentration fields in the IFS at every time step to account for the local concentration changes.

Motivation for the design of a two-way coupled system

An extension of an earth-system model can follow two approaches: (i) directly integrating subroutines or modules in one unified model or (ii) coupling independent models by a means of dedicated coupler software. Direct integration - often referred to as on-line coupling - is normally pursued when complex chemistry schemes are included in meteorological models.

Coupling independent models with coupler software is often applied when the models cover different domains of the earth-system such as ocean and atmosphere. We decided to develop a coupled system (IFS-CTM) in which the IFS and a CTM are run in parallel because of the envisaged high development cost to integrate and test complex chemical mechanisms as integrated part of the IFS. We also concluded that benefits from using ECMWF's operational data assimilation system and the associated infrastructure for observation processing would be difficult to maintain if we developed a new data assimilation system around an existing CTM. Another advantage of the coupled system is the possibility to easily couple different CTM to the IFS and therefore to be more flexible in the choice of the applied chemistry schemes. A coupled system of independent components can also better benefit from the ongoing development work of the stand-alone versions of the CTMs since the CTMs stay independent models.

Data exchange and experiment setup

The coupled system is a three-dimensional two-way coupled system consisting of the IFS, one of the CTMs MOZART-3, TM5 and MOCAGE and the coupler software OASIS4 (Valcke, S. and Redler, R. 2006). In the coupled system, the IFS simulates the advection, vertical diffusion and convection of selected chemical tracers (CO, NO_x, HCHO, SO_2 and O_3) and applies tracer tendencies calculated by the coupled CTM to account for sink and source processes such as emission, chemical conversion and deposition. The prognostic tracer variables are also part of the control variables of the data assimilation mode in IFS. Figure 12 shows a schematic of the data flow in the coupled system. Every hour, the IFS provides meteorological fields to drive the CTMs and receives the CTM tendencies. The details of the application and formulation of the tendency terms is given in section 0.

The choice of the exchanged meteorological fields depends on the requirements of the CTM. All CTMs receive fields of humidity, temperature, wind components, or divergence and vorticity in spectral representation, and sensible and latent heat flux. MOZART-3 and MOCAGE simulate their own hydrological cycle whereas TM5 also requires the IFS's precipitation and cloud fields, surface properties and convective mass fluxes.

A further coupling pathway is the exchange of concentration fields form the IFS to the CTM or vice versa at the start of each coupled run. The tracer concentrations are used as initial conditions for the coupled run. The experiments with the coupled system are organized as a sequence of several 12 hour runs in data assimilation mode or - in forecast mode - as runs over up to 4 days starting every 24 hours. In data assimilation mode, the length of the coupled run is given by the length of the 4D-VAR assimilation window, which is normally 12 hours. A more detailed description of the functionality in data assimilation of the coupled system is given in section 0. In forecast mode, the meteorological fields in the IFS need to be initialized at least every 24 hours with a meteorological analysis in order to avoid a drift from the observed state of the atmosphere. At the start of each coupled run, the initial conditions of the coupled tracer in the IFS and the CTM are set to the same values: Either the CTM fields replace the IFS tracer initial conditions fields (hereafter CTM-IC mode) or the IFS tracer fields replace the respective initial conditions of the CTM (hereafter IFS-FB mode).

In CTM-IC mode, the IFS takes the initial tracer conditions from the CTM at the start of each new run. The CTM itself gets the whole set of initial conditions from the previous CTM run. In this configuration the concatenated CTM output of IFS-CTM is equivalent to the normal continuous CTM off-line run.

In IFS-FB mode, CTM fields are used as IFS initial conditions only at the very first coupled run. In all subsequent runs, the IFS's coupled tracers are initialized from the previous IFS run and the respective CTMs fields are replaced by the IFS fields. The un-coupled CTM species are initialised from the previous CTM run as in CTM-IC mode. The IFS-FB mode is applied in data assimilation because the IFS tracer fields contain the assimilated information of the observations. The IFS-FB mode can also be applied to impose vertical tracer profiles simulated by the IFS on the CTM.

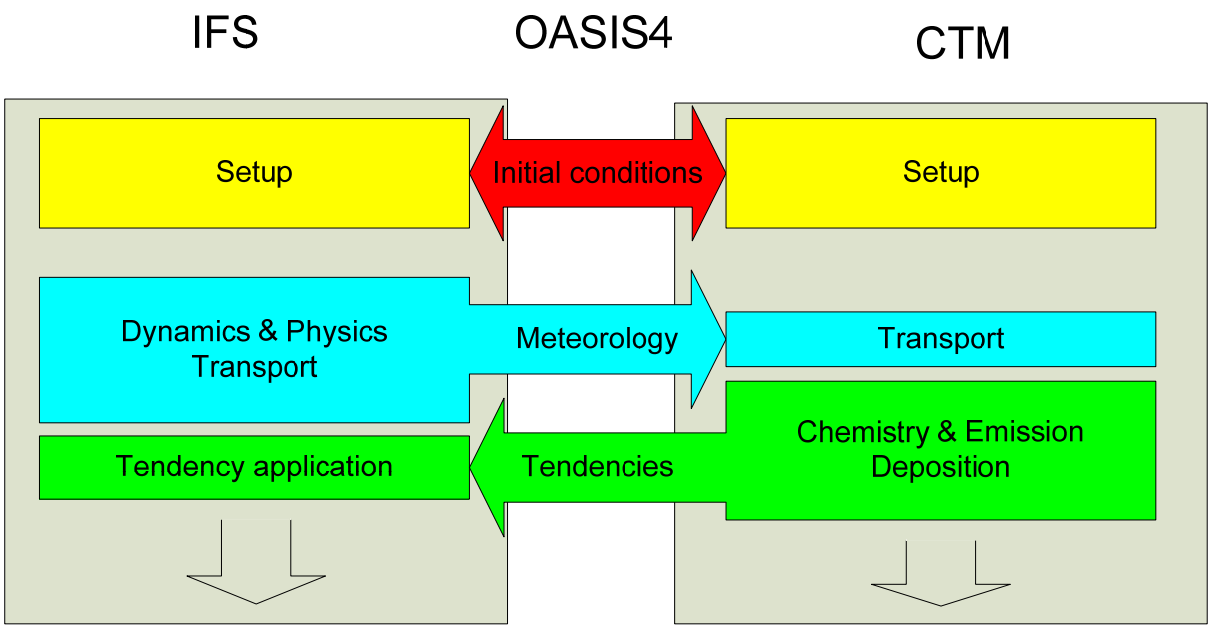


Figure 12 Schematic of the data flow (setup and first time step) in the coupled system consisting of the IFS and CTM.

Formulation of tendency terms and their application in the IFS

The exchange of concentration tendencies is a unique feature of the coupled system. The formulation of the tendency terms has to reflect (i) operator splitting and time stepping in both the CTMs and the IFS, (ii) relative size and spatial structure of the tendency fields, and (iii) the computational cost of the exchange.

The CTMs use an operator-splitting approach in which advection, chemistry, emission injection, diffusion and deposition are called in sequence, and the concentrations are updated directly within each operator subroutine. The IFS computes semi-lagrangian advection, diffusion and convection of the tracers based on unperturbed concentration field values from the previous time step (Beljaars, 2004) and updates the concentration values with the accumulated tendency of all sink and source processes at the end of the time step.

The total CTM tendencies T [kg/(kg*s)] are given by the sum of chemical loss L_C and production P_C, production due to emission injection P_E and loss L_D due to deposition:

$$T = P_C - L_C + P_E - L_D$$

The injection of surface emissions is integrated in the MOZART-3 diffusion scheme, whereas TM5 and MOCAGE distribute the injected mass in a fixed ratio over selected layers near the surface and apply their diffusion operator after the injection. P_E is therefore a combination of the emission injection and the tendencies due to vertical turbulent diffusion. Since P_E already contains the diffusion tendencies, its application in the IFS requires that the IFS diffusion must not be applied again to the respective tracer fields. In order to also use the IFS diffusion scheme for the tracer transport within the coupled system, we determined the effective surface flux $\Phi_{E\text{-}D}$ from emissions and dry deposition by calculating the total columns of the surface contribution of P_E and the fraction of L_D representing dry deposition. $\Phi_{E\text{-}D}$ is then presented as a surface flux to the IFS diffusion scheme and the components P_E and L_D are excluded from T_{All} leaving T_{Air}.

Deposition L_D and chemical loss L_C are almost always proportional to the tracer concentration x. A relative formulation $L = l\,x$, i.e. a loss rate l, would have linked tendency and concentration values and would have helped to avoid negative concentrations after the application of the CTM tendencies in the

IFS. However, we decided against the relative formulation of tendencies because (i) it would have been be more difficult to distinguish chemical loss and production from the output arguments of the chemistry routines, which directly only provide the total change, and (ii) because a separate interpolation of production and loss tendencies, which often almost compensate each other, could have caused imbalances when the two fields are combined again in the IFS.

After consideration of the above arguments we decided to transfer and apply the process-specific tendencies of the CTM in one of the following two modes:

1. one three-dimensional tendency field T_{All} containing all sources and sinks as well as diffusion (hereafter total-tendency mode)
2. on three-dimensional tendency field T_{Air} and the effective $\Phi_{E\text{-}D}$ surface flux (emission and deposition) (hereafter surface-flux mode)

The coupled system in data assimilation mode

The implementation of the incremental solution for the 4D-VAR data assimilation in the ECMWF integrated forecast system is structured into an "outer" and "inner" loop. The outer loop is an IFS run with the complete model at high resolution to determine the increments between the model and the observation. The "inner loop" is run with a linearized model version and its adjoint formulation at a lower resolution. The simplified model used in the inner loop acts on the increments to solve the minimization problem of the 4D-VAR cost function. The sequence of outer and inner loop is repeated two or more times to determine the final analysis. The time length of the assimilation window is 12 h, and analyses valid every 6 hours are produced. A forecast runs started from the previous analysis provides the starting point for the assimilation in the next 12h window.

The implementation of the data assimilation algorithm in the coupled system follows the assimilation of the IFS ozone with the parameterized chemistry. The outer loops are run coupled in IFS-FB mode, i.e. the IFS fields replace the assimilated O_3 and CO fields in the CTM. The inner loops are run uncoupled, i.e. without tendency application. The simplified linearized model of the inner loop represents tracer advection, diffusion and convection of the increments between observed and simulated concentration fields.

The data assimilation of the global reactive gases observation with the coupled system is part of the assimilation of all meteorological observations to obtain the analysis of the meteorological fields as well as the chemistry fields. In principle, the observations of trace gases could influence the analysis of the meteorological fields. However, given the uncertainty of these observations and the lack of observational constrains of variables such as wind or temperature in the stratosphere and above, a possible influence of the observations of atmospheric composition on the meteorological fields is currently suppressed. The meteorological observations however influence the tracer fields because of the change in transport patterns and temperature.

The assimilation of the reactive gases is statistically uni-variant in the sense that no background error covariance terms between the chemical tracers or the meteorological fields are defined. Since also no chemical mechanism is included in the minimization, a possible impact of the CO or ozone assimilation on the rest of the chemical species appears only as a consequence of the chemical interaction in the 12 h CTM runs. The CTM run starts from initial conditions of the previous 12h window, and only the assimilated CO or O_3 fields from the IFS replace the initial fields of the two species.

CTM and IFS specifications

In the coupled system, the IFS runs in a T159 spectral resolution and the grid point space is represented in the reduced Gaussian grid, which has a grid box size of about 125 km. The CTMs use a

regular latitude-longitude grid of about 2° - 3° grid box length. The coupler performs horizontal interpolations for which a bi-linear mode is applied. The IFS runs - at most parts of the globe - at a finer horizontal resolution than the CTMs because it improves (i) the quality of the meteorological forecasts and (ii) the acceptance of high resolution observations within data assimilation mechanism.

The vertical coordinates for the IFS and all CTMs are the same 60 hybrid sigma-pressure levels, reaching up to 0.1 hPa. The identical vertical structure in the IFS and the CTM avoids the need for vertical interpolation. The minimum coupling interval is 3600 s which is the largest acceptable time step for the IFS at a T159 resolution, and also the time step of some of the CTMs. An overview of the CTM resolution and parameterisations is given in Annex 1, Table 1.

ANNEX 6: List of GEMS-GRG experiments

MOZART stand-alone runs:

Run ID	Run date	Start – End date	Data location	Notes	Purpose
MOZART_V1	2006	20030101 20031231	ec:/hc0/mozart3.1/run_era1		First stand-alone run. It is still the actual reference for 2003 troposphere.
MOZART_V2	2006	20030601 20030831	FZJ	Stand-alone run for July/August 2003 in T106 resolution which had been run in Juelich.	High resolution run T106L60
MOZART_V3	2007	20021101 20031231	ec:/hc0/mozart3.1/run_era1_v3	Stratospheric initialization from BASCOE during October 2002 using a nudging scheme. Includes an error in tropospheric emissions. Initialized with the GFEDv2 8-day wildfire emissions BB emissions spread equally across PBL SOx chemistry provisional.	Second stand-alone run designed for the stratospheric case study; troposphere erroneous
MOZART_V7	2008	20021101 20040630	ec:/hc0/mozart3.1/run_era1_v7	BASCOE initialisation for the stratosphere, improved upper boundary conditions (zero-flux upper boundary conditions for several species), SOx chemisty, GFEDv2 8day emissions, Updated anthropogenic emissions (RETRO merged- with REAS), Chemical reactions updated to JPL06.	Reference for 2003 stratosphere, error in anthropogenic emissions and in rate constant for CO + OH
MOZART_V9	2008	20040101 20041231	ec:/hc0/mozart3.1/run_era1_v9	As MOZART-V7, but corrected anthropogenic emissions and rate constants for CO + OH	Reference for 2004, meteorology from f026
MOZART_NRT	2007	20070402 20070928	ec:/hc0/mozart3.1/run_ewgo	Reduced outputs.	NRT run for comparison to evnj and ewgo

MOZART-V10	2008	20030101 20041231		Debugged for ozone hole missing (wrong initialization of halogen source gases from run V3 on).	Reference run for 2003 & 2004

V10 is the latest stand-alone reference run for the year 2003. GRG anthropogenic emission inventories and 8days GFEDv2 for fire emissions inventories are used and confined to the lower MOZART level. The simulation results are available since December 2008. The new JPL-06 evaluation of the reaction CO + OH leads to a reaction rate which is about 10-20% lower than before in the troposphere, is introduced in V10 simulations.

TM5 stand-alone runs:

Run ID	Run date	Start – End date	Data location	Notes	Purpose
TM5_V3	2006	20030101 20031231	ec:/nl5/gems/grg/TM5/V3	No stratospheric chemistry.	First stand-alone run.
TM5_HWHR1	2006	20030601 20030901	ec:/nl5/gems/grg/TM5/HWHR1	Zoom region above Europe	European heat wave study (Ordonez et al., 2008)
TM5_HWEE				As for TM5-HWHR1 but with 25% extra anthropogenic emissions of NOx, CO and non-methane VOCs over Europe	
TM5_HWEN				As for TM5-HWHR1 but without anthropogenic emissions of NOx, CO and non-methane VOCs outside Europe	
TM5-HWDN				As for TM5-HWHR1 but decreasing the soil wetness factor to 0.1	
TM5_SFHR1	2006	20030401 20030701	ec:/nl5/gems/grg/TM5/SFHR1	Zoom region above Siberia.	Siberian forest fire study
TM5_V7	2008	20030101 20031231	ec:/nk9/gems/grg/TM5/V7	Improved CO fields compared to TM5-V3	TM5 reference stand-alone run for 2003
TM5_V8	2008	20040101 20041231	ec:/nk9/gems/grg/TM5/V8Y04	As for TM5_V7.	First stand-alone run for 2004 with GFEDv2 8-daily emissions. Some bugs were identified in this run.

TM5_V9		20040415 20040901	ec:/nk9/gems/grg/TM5/V9Y04	Debugged V8 version.	TM5 reference stand-alone run for 2004
TM5_V9ES2	2008	20040415 20040901	ec:/nk9/gems/grg/TM5/V9ES2	Specifications as in TM5-V9	HTAP sensitivity study ES2
TM5_V10	2008	20030701 20041231	ec:/nk9/gems/grg/TM5/V10	Available for July 2003 onwards, on 34 layer vertical resolution. This serves as a spinup for 2004 run. Improved (debugged) CO-emissions Improved NOx emissions, including biomass burning injection heights for all emissions. New boundary conditions for ozone in stratosphere: nudging to assimilated GOME - SCIAMACHY data for ozone in stratosphere	TM5 reference stand-alone run for 2004

MOCAGE stand-alone runs:

Run ID	Run date	Start – End date	Data location	Notes	Purpose
MOCAGE	2006	20030101 20031231	ec:/tok/ SIMU2003/ 47LEV	Bug found	First stand-alone run
MOCAGE_strato	2007	20020111 20030414	ec:/tok/ SIMU2003/ STRATO	Stratospheric initialization from BASCOE during October 2002 using a nudging scheme	Second stand-alone designed for stratospheric case study
MOCAGE_60lev_1 or MOCAGE-v1	2008	20021001 20031231	ec:/tok/SIMU2003/60LEV01	Initialized with climatology Improved setup (diffusion, convection, fire emission, CO initialisation)	Stand-alone run for 2003
MOCAGE_60lev_2 or MOCAGE-v2	2008	20021001	ec:/tok/SIMU2003/60LEV02	Stratospheric initialization from BASCOE during October 2002 using a nudging	Stand-alone run for 2003

		20031231		Modifications compare to 60LEV01: i) ozone deposition multiplied by 2 over sea, ii) reaction rate of CO with OH decreased by 10% (longer lifetime for CO, lower OH concentrations), iii) change in the computation of vertical velocity from ECMWF grid point fields	
MOCAGE_HTAP_ ES1	2008	20040101 20041001	ec:/tok/HTAP_exp3/ ES1 (june to september 2004)	Initialized from 60LEV02. GFED 8 day wildfire emissions. HTAP output formats.	Standard simulation ES1 for HTAP experiment 3
MOCAGE_HTAP_ ES3	2008	20040501 20041001	ec:/tok/HTAP_exp3/ ES3 (June to September 2004)	Initialized from ES1 North American biomass burning emissions reduced by 20%	Sensitivity run ES3 for HTAP experiment 3
MOCAGE_HTAP_ ES4	2008	20040501 20041001	ec:/tok/HTAP_exp3/ ES4 (june to september 2004)	Initialized from ES1. Biomass burning emissions restricted to the boundary layer	Sensitivity run ES4 for HTAP experiment 3

MOCAGE V2 is the name for stand-alone reference run for the year 2003 and 2004. It is available since June 2008. 8days GFEDv2 for CO fire emissions inventories are used.

IFS + MOZART coupled forecasts:

Run ID	Rundate	Start & End date	Data location	Notes	Purpose
euas	1/2007	20030101 20031231	class=rd expver=euas stream=oper type=fc	CTM constrained mode. CTM vertical transport	1st coupled system forecast run
exoz	9/2007	20030701 20030831	class=rd expver=exoz stream=oper	CO and O3 feedback mode IFS vertical	Fluxes corrected

			type=fc	transport	
eydy	10/2007	20030701 20030831	class=rd expver=excd stream=oper type=fc	CTM constrained mode CTM vertical transport	Fluxes corrected + improved cloud restart
eyg0	11/2007	20070401- 20070831	class=rd expver=eyg0 stream=oper type=fc	CTM constrained mode CTM vertical transport 48 h Forecast	GEMS NRT forecast. Fluxes corrected + improved cloud restart - still no correct ozone hole
eywm	02/2008	20021101 20031231	class=rd expver=eywm stream=oper type=fc	CTM constrained mode CTM vertical transport IFS ozone only available until 28/03/2003 ; for the period after that, use IFS ozone data of the euas run. Most noticeable technical changes: - smoothing of horizontal winds in order to get more realistic vertical velocity field, - new IFS cycle 32r3 (decreased diffusion in PBL, increased convection fluxes)	eywm is a coupled system forecast and, in essence, an improved version of the euas run. The run has been initialised with BASCOE ozone fields. eywm runs in CTM constrained mode without feedback. The setup for this run is meant to be used in the 2nd re-analysis, then of course with feedback for CO and Ozone.
eyr0	12/2007	20070401- 20071131	class=rd expver=eyr0 stream=oper type=fc	CTM constrained mode CTM vertical transport	GEMS NRT forecast. Fluxes corrected + improved cloud

				24 h Forecast	restart + Improved vertical velocity fields - ozone hole !
ez2m	15.11.2007- present	20070301 20071214	class=rd expver=evnj stream=oper Type=fc	CTM constrained mode CTM vertical	GEMS NRT forecast. Running, continues eyr0
f38u	2008	20040601- 20040731	Class=rd Expver=f38u Stream=oper type=fc		Kind of control fc run for f026

The MOZART-IFS forecast run (experiment *ez2m*) has run continuously since beginning of 2008, delivering global forecasts of trace gases up to three days ahead. This experiment is based on a free-running coupled system, i.e. without data assimilation.

The chemical solver is based on the MOZART-tracer model. It applies a gaussian grid with a resolution of about 1.9 deg lat/lon. MOZART applies a distribution of 60 layers, with the top layer at 10 hPa. The emission inventory is based on the RETRO-2003 inventory created for the GEMS-GRG project. For the biomass burning inventory a ten-year averaged (1997-2006) database is used for the forecast-run.

F026 second reanalysis run is with fixes: i) IFS convection and diffusion for CO, CTM convection and diffusion for GO_3, NOx, Coupled to MOZART, ii) Two-way coupling for CO and GO_3, iii) NOx and HCHO initialized from CTM in every forecast, iv) Bugfix for CTM fluxes from **eyih** (1st reanalysis run – with CO (MOPITT) and GEMS ozone assimilation (SCIAMACHY, MIPAS, GOME, SBU) and it is first 1st GEMS reanalysis (GRG together with GHG and AER). This run is available for 2003 - 2007. Monthly GFEDv2 for fire emissions inventories are incorporated in MOZART- *F026* simulations and confined to the lower MOZART level.

IFS + TM5 coupled forecasts:

Run ID / Info	Rundate	Start- End-date	Data location	Notes	Purpose
eybl	12/2007	20030701	class=rd	CTM constrained	- first trial

		20030830	expver=eybl stream=oper type=fc	mode CTM vertical transport 24 h Forecast	
ez3h	12/2007	20030701 20030830	class=rd expver=eybld stream=oper type=fc	CO feedback IFS vertical transport 24 h Forecast	-first trial.
f1d7	6/2008	20040601 20040731	class=rd expver=f1d7 stream=oper type=fc	O3 feedback IFS vertical transport 24 h Forecast	First successfull run with fully coupled meteo. TM5 model identical to V9

IFS + MOCAGE coupled forecasts:

Run ID / Info	Rundate	Start- End-date	Data location	Notes	Purpose

IFS + MOZART coupled reanalyses with assimilation:

Run ID	Rundate	Start & End date	Data location	Notes	Purpose
exqx	2007	20030101 20030430	Class=rd Expver=exqx Stream=DA Type=an	IFS convection and diffusion for CO. CTM convection and diffusion for GO3, NOx. Coupled to MOZART. Two-way coupling for CO and GO3. Bug: no SCIAMACHY data assimilated in IFS O3. Bug: CTM fluxes..	1st GRG reanalysis. CO (Mopitt) and GEMS ozone assimilation (SCIAMACHY, MIPAS, GOME, SBUV). Only used by BIRA.
eyih	2007	20030501- 20031231	Class=rd Expver=eyih Stream=DA Type=an	IFS convection and diffusion for CO. CTM convection and diffusion for GO3, NOx. Coupled to MOZART. Two-way coupling for CO and GO3. NOx and HCHO	1st GRG reanalysis. CO (Mopitt) and GEMS ozone assimilation (SCIAMACHY, MIPAS, GOME, SBUV).

				initialized from CTM in every forecast. Bugfix for CTM fluxes.	
eyq6	2007	20030501 20031231	Class=rd Expver=eyq6 Stream=DA Type=an	CO data and GEMS ozone data passive.	Control run for eyih.
f026	2008	20030101 20041231	Class=rd Expver=f026 Stream=DA Type=an	All fixes from eyih SBUV, MIPAS, SCIAMACHY ozone CO-Mopitt	1st GEMS reanalysis (GRG together with GHG and AER)
f171	2008	20040101 20040501	Class=rd Expver=f171 Stream=DA Type=an	As f026 but with more ozone from 20040101	Rerun of f026
f1kd	2008	20080705-	Class=rd Expver=f1kd Stream=DA Type=an	Assimilation of OMI ozone data and MODIS aerosols, monitoring of IASI CO	NRT GEMS analysis

IFS + TM5 coupled reanalyses with assimilation:

Run ID / Info	Rundate	Start- End-date	Data location	Notes	Purpose
Improvements: bug fixes with respect to the use of emissions data					

In progress	2008.11				

IFS + MOCAGE coupled reanalyses with assimilation:

Run ID / Info	Rundate	Start- End-date	Data location	Notes	Purpose
Improvements: identification and elimination of an intricate bug which had in the past rendered coupled MOCAGE simulations impossible					
??					

ANNEX 7: Scoring methods

For the evaluation of model performance with respect to observations used in GEMS-GRG, we used the following metrics, as suggested in the GEMS-GRG global evaluation report (Cammas et al., 2007) and GEMS-VAL report (Huijnen et al, 2008).

The traditional metrics used for model evaluation include the mean bias which is defined as:

$$\mathrm{MB} = \frac{1}{N}\sum_{i}^{N}\left(M_i - O_i\right) \qquad (1.1)$$

where Mi the model mean daily values, Oi the mean daily observations and N the number of pairs used.

A further desirable step is to normalize these metrics so that they give a relative error. The usual choice is to use the observed values for normalisation, giving the normalised mean bias (NMB).

$$\mathrm{NMB} = \frac{1}{N}\sum_{i}^{N}\frac{M_i - O_i}{O_i} \qquad (1.2)$$

One problem with the use of NMB is that there is an asymmetry between the case of under and over-prediction. For over-prediction NMB can grow to very high values much greater than unit, whilst for under prediction it is limited to -1. It is desirable to use a metric which treats both of these model deficiencies in a symmetric manner. A solution is to employ a normalisation comprised of the arithmetic mean of the observed and forecast value, giving a modified normalized mean bias:

$$\mathrm{MNMB} = \frac{2}{N}\sum_{i}^{N}\frac{M_i - O_i}{M_i + O_I} \qquad (1.3)$$

MNMB ranges for -2 to 2 and is usually expresses as a percentage %. For an acceptable model performance should range between 15-20%.

Fractional Gross Error (FGE) is defined as:

$$\mathrm{FGE} = \frac{2}{N}\sum_{i}^{N}\left|\frac{M_i - O_i}{M_i + O_i}\right| \qquad (1.4)$$

where Mi the model mean daily values, Oi the mean daily observations and N the number of pairs used. FGE ranges from 0 to 200% and can be also expresses as a percentage. For FGE < 30-35% we can assume an acceptable model performance. This metric is used as an indicator of the overall model performance.

The statistic most often used to quantify differences in two fields is the root mean square error (RMSE).

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i}(M_i - O_i)^2} \qquad (1.5)$$

The traditional rationale for employing the RMSE as an indicator of overall forecast error is two fold: (i) by squaring the errors before combining, this measure removes any cancellation of under and over-prediction; (ii) in cases where the spread of errors approximates to a well-known distribution (e.g.

normal, binomial, Poisson etc.) the RMSE can attributed with a physical significance. In the present case the errors are not expected to conform to any well-known distribution. In addition, the RMSE suffers from similar deficiencies as the mean bias, with respect to asymmetry. A further issue is that the RMSE gives added weight to those errors having greater magnitude, as a consequence of the squaring operation.

The Taylor diagram is a polar coordinate diagram that assigns the angular position to the inverse cosine of the correlation coefficient, R (*Taylor, 2001*). A correlation coefficient of 0 is thus 90° away from a correlation coefficient of 1.

$$R = \frac{\frac{1}{N}\sum_{n=1}^{N}\left(m_n - \bar{m}\right)\left(r_n - \bar{r}\right)}{\sigma_m \sigma_r} \qquad (1)$$

The correlation coefficient R provides information about the differences in phase between model and observations.The radial (along-axis) distance from the origin is assigned to the normalized standard deviation, σ_n.

$$\sigma_n = \frac{\sigma_m}{\sigma_r} \qquad (2)$$

The reference field point is indicated for the polar coordinates (1.0,0.0). The model to reference comparison points may then be gauged by how close they fall to the reference point. This distance is proportional to the unbiased Root-Mean- Square Error (RMSE′), as defined by:

$$RMSE' = \left(\frac{1}{N}\sum_{n=1}^{N}\left[\left(m_n - \bar{m}\right) - \left(r_n - \bar{r}\right)\right]^2 \right)^{0.5} \qquad (3)$$

where the overbars indicate the mean, m the model and r the reference (observation) fields. The unbiased RMSE may be conceptualized as an overall measure of the agreement between the amplitude (σ) and phase (R) of two temporal patterns. The term unbiased is used herein to emphasize that Eq. (1) removes any information about the potential bias (B), which is defined as the difference between the means of the two fields:

$$B = \bar{m} - \bar{r} \qquad (4)$$

In other words, the unbiased RMSE (RMSE′) is equal to the total RMSE if there is no bias between the model and reference fields. This may be verified given the quadratic relationship between the unbiased RMSE, the bias, and the total RMSE:

$$RMSE^2 = B^2 + RMSE'^2 \qquad (5)$$

A relatively skilful model should be able to accurately simulate both the amplitude and pattern of variability. Which of these factors is more important depends on the application and to a certain extent must be decided subjectively. Thus it is not possible to define a single skill score that would be universally considered most appropriate. Any defined score, however, should increase monotonically with increasing correlation, and for any given correlation the score should increase as the modelled variance approaches the observed variance. Traditionally, skill scores have been defined to vary from zero (least skilful) to one (most skilful). Note that in the case of relatively low correlation the inverse of the RMSE does not satisfy the criterion that skill should increase as the simulated variance approaches the observed. Thus a reduction in the RMSE may not necessarily be judged an

improvement in skill. One of the least complicated scores that fulfils the above requirements is defined as:

$$S = \frac{4(1+R)}{(\sigma_n + 1/\sigma_n)^2 (1+R_o)} \qquad (6)$$

where Ro is the maximum correlation attainable (*Taylor, 2001*). As σn -> 1 and R -> Ro the skill score approaches unity.

The Talyor plots presented in this report have been modified adding a total RMSE colour scale, which indicates the range of minimum to maximum total RMSE using a spectral (rainbow) colour scaling increment.

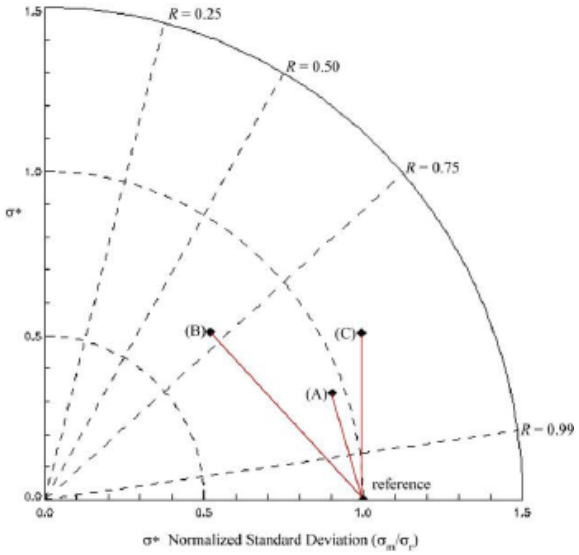


Figure 13 Example of a Taylor diagram. As explained in text, the radial distance is proportional to the normalized standard deviation, σ_n and the angular position corresponds to the linear correlation coefficient (R). The distances between the labeled points (A,B,C) and the reference point are proportional to the unbiased RMSE (Eq3). *Jolliff, et al., 2008.*

List of abbreviations:

FGE	Fractional Gross Error
GRG	Global Reactive Gases
IFS	Integrated Forecasting System
MB	Mean Bias
MNMB	Modified Normalized Mean Bias
R	Correlation Coefficient
RMSE	Root Mean Square Error
σ_n	Normalized Standard Deviation

ANNEX 8: Detailed model scores for the NO_2 evaluation (Chapter 2)

In chapter 2, in the end of each section regarding the evaluation of the models, the annual averaged scores are presented for each model version for all the stratospheric and tropospheric regions defined in section 2.1. Although these values already reveal the strong and weakness points of model runs, they also, in part, omit some of the problematic questions such as simulation of NO_2 over Siberia and Alaska fire regions. Therefore, when aiming for better understanding of model drawbacks (e.g., which months the model fails to simulate stratospheric fields over the South Pole) one should then look into the more complete evaluation and descriptive results. Those values are presented in the following graphs with the monthly scores averaged for those regions selected for the evaluation of model results.

a) Offline simulations

Stratosphere

MOCAGE

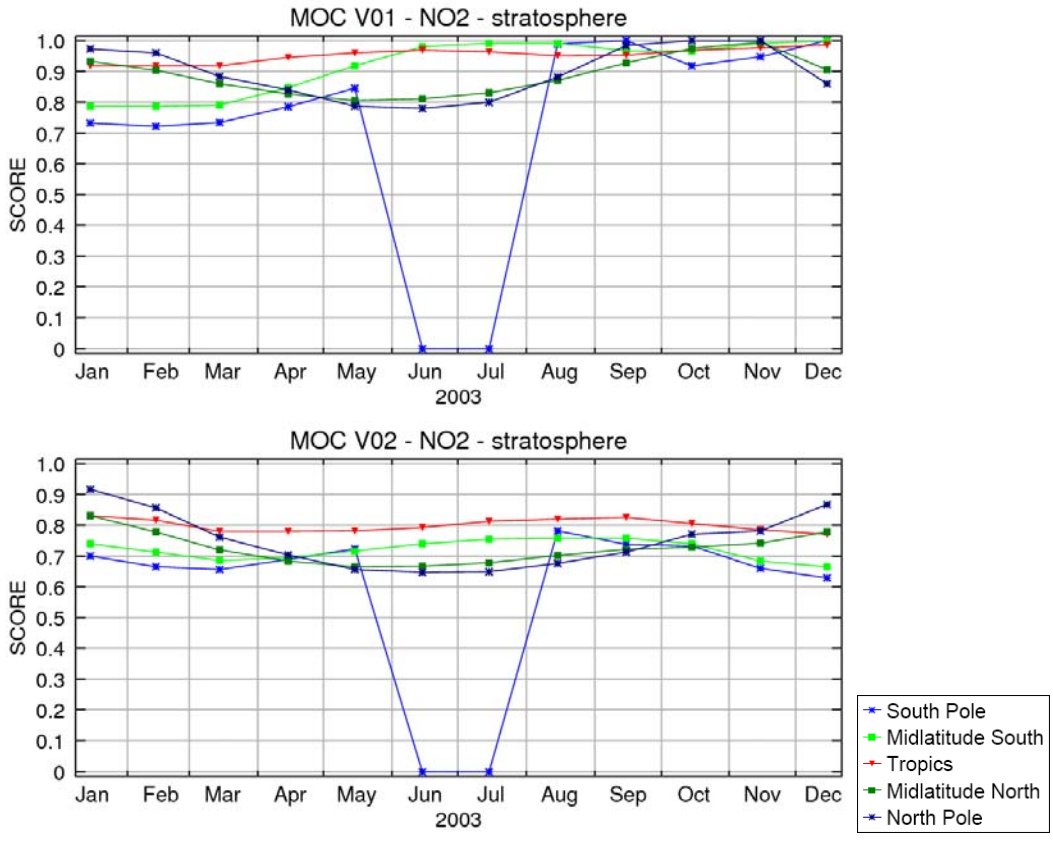


Figure 1: NO_2 monthly scores, averaged for the stratospheric regions, for MOCAGE stand-alone runs V01 (top) and V02 (bottom), for the year 2003.

MOZART

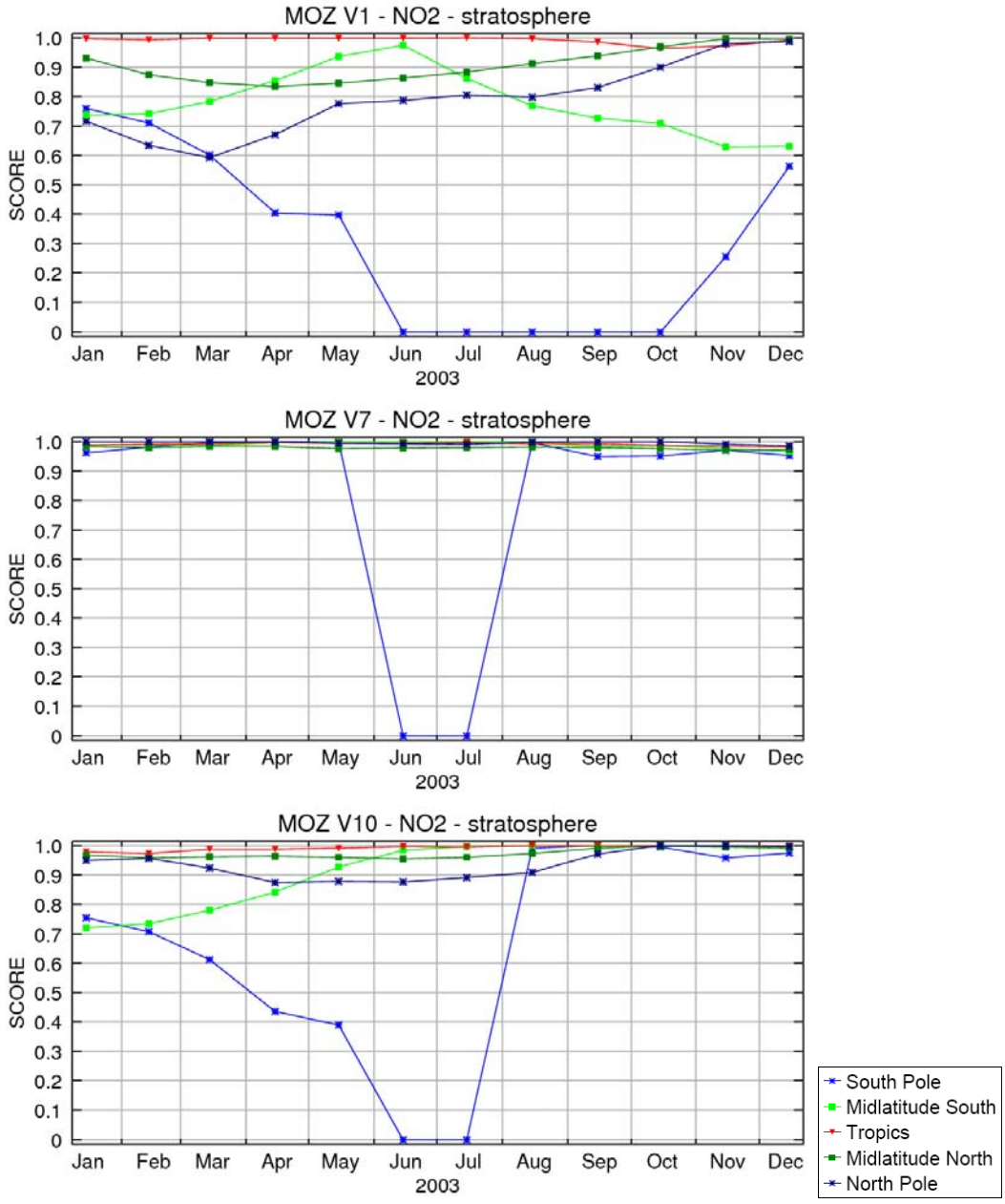


Figure 2: NO_2 monthly scores, averaged for the stratospheric regions, for MOZART stand-alone runs V1 (top), V7 (middle) and V10 (bottom), for the year 2003.

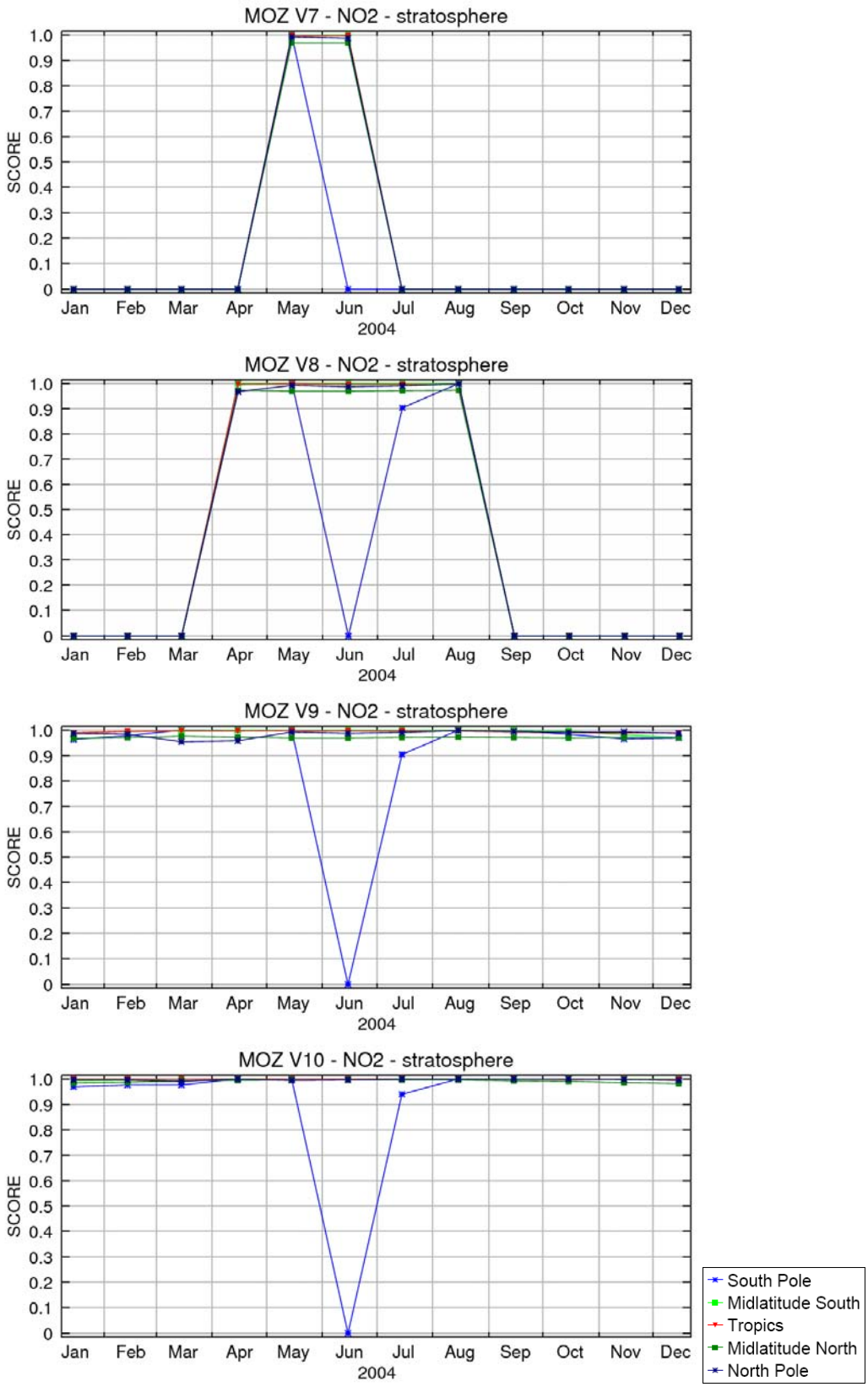


Figure 3: NO_2 monthly scores, averaged for the stratospheric regions, for MOZART stand-alone runs (from top to bottom) V7, V8, V9 and V10, for the year 2004.

TM5

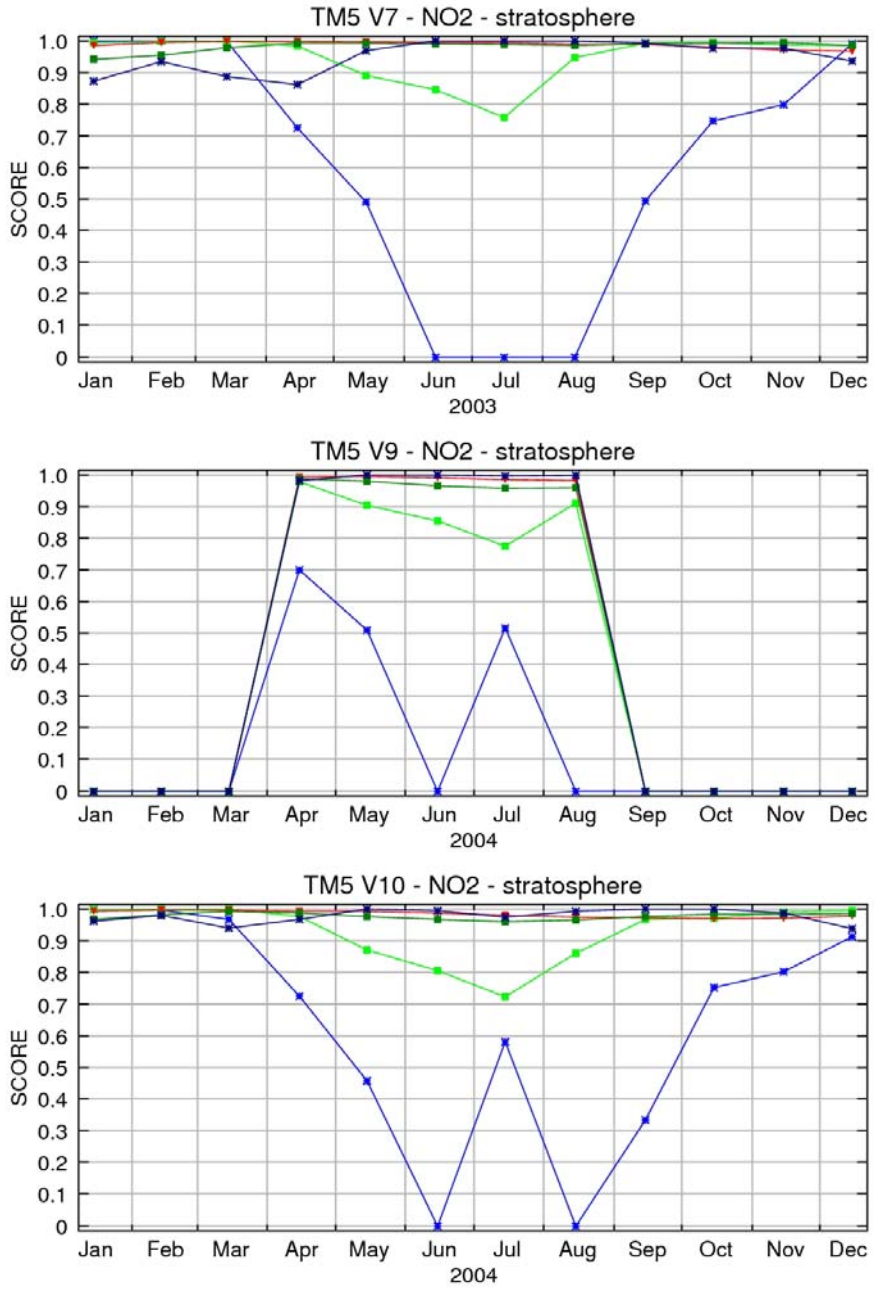


Figure 4: NO_2 monthly scores, averaged for the stratospheric regions, for TM5 stand-alone runs V7 (top), for the year 2003, and V9 (middle) and V10 (bottom), for the year 2004.

Troposphere

MOCAGE

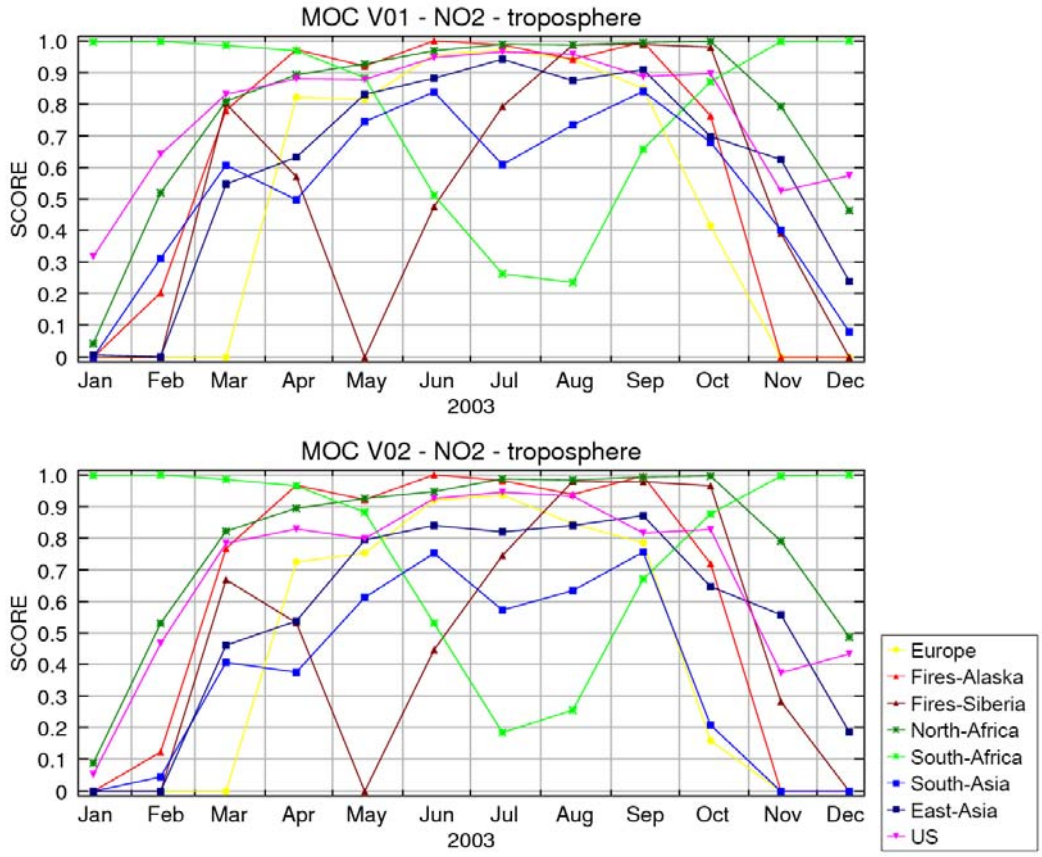


Figure 5: NO_2 monthly scores, averaged for the tropospheric regions, for MOCAGE stand-alone runs V01 (top) and V02 (bottom), for the year 2003.

MOZART

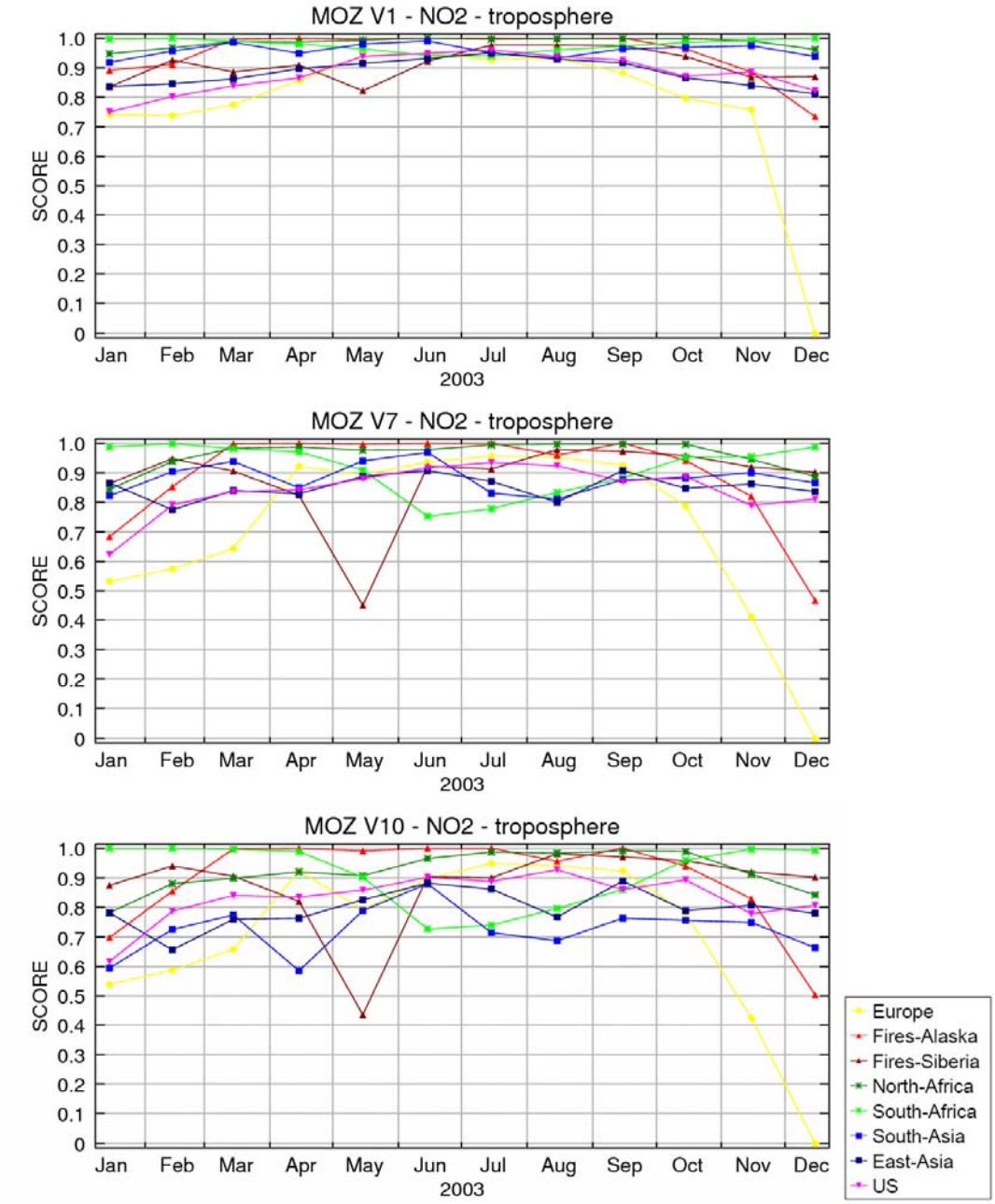

MOZ V1 - NO2 - troposphere
SCORE
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec
2003
MOZ V7 - NO2 - troposphere
SCORE
Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec
2003
MOZ V10 - NO2 - troposphere
SCORE
Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec
2003
Europe
Fires-Alaska
Fires-Siberia
North-Africa
South-Africa
South-Asia
East-Asia
US

Figure 6: NO_2 monthly scores, averaged for the tropospheric regions, for MOZART stand-alone runs V1 (top), V7 (middle) and V10 (bottom), for the year 2003.

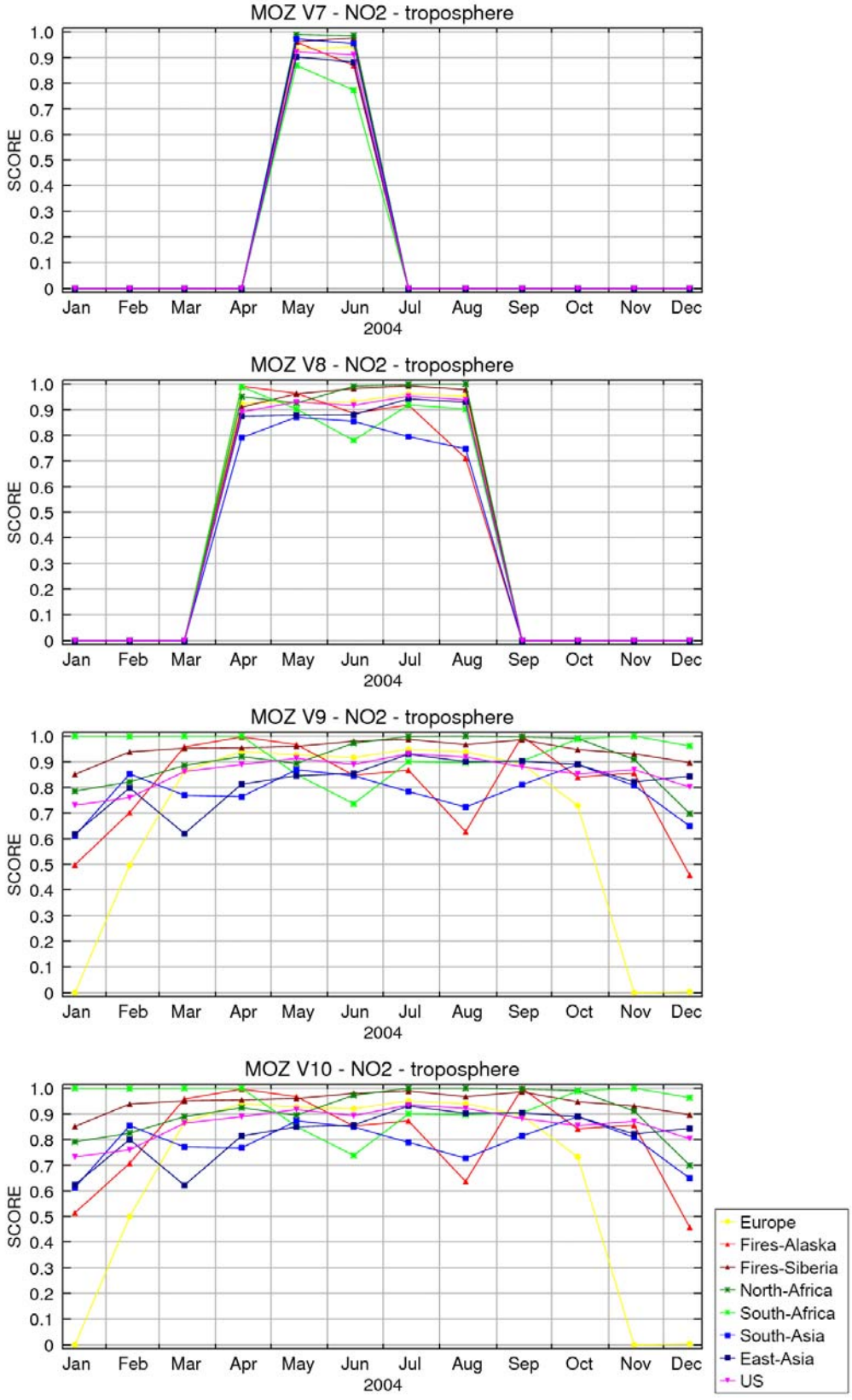


Figure 7: NO_2 monthly scores, averaged for the tropospheric regions, for MOZART stand-alone runs (from top to bottom) V7, V8, V9 and V10, for the year 2004.

TM5

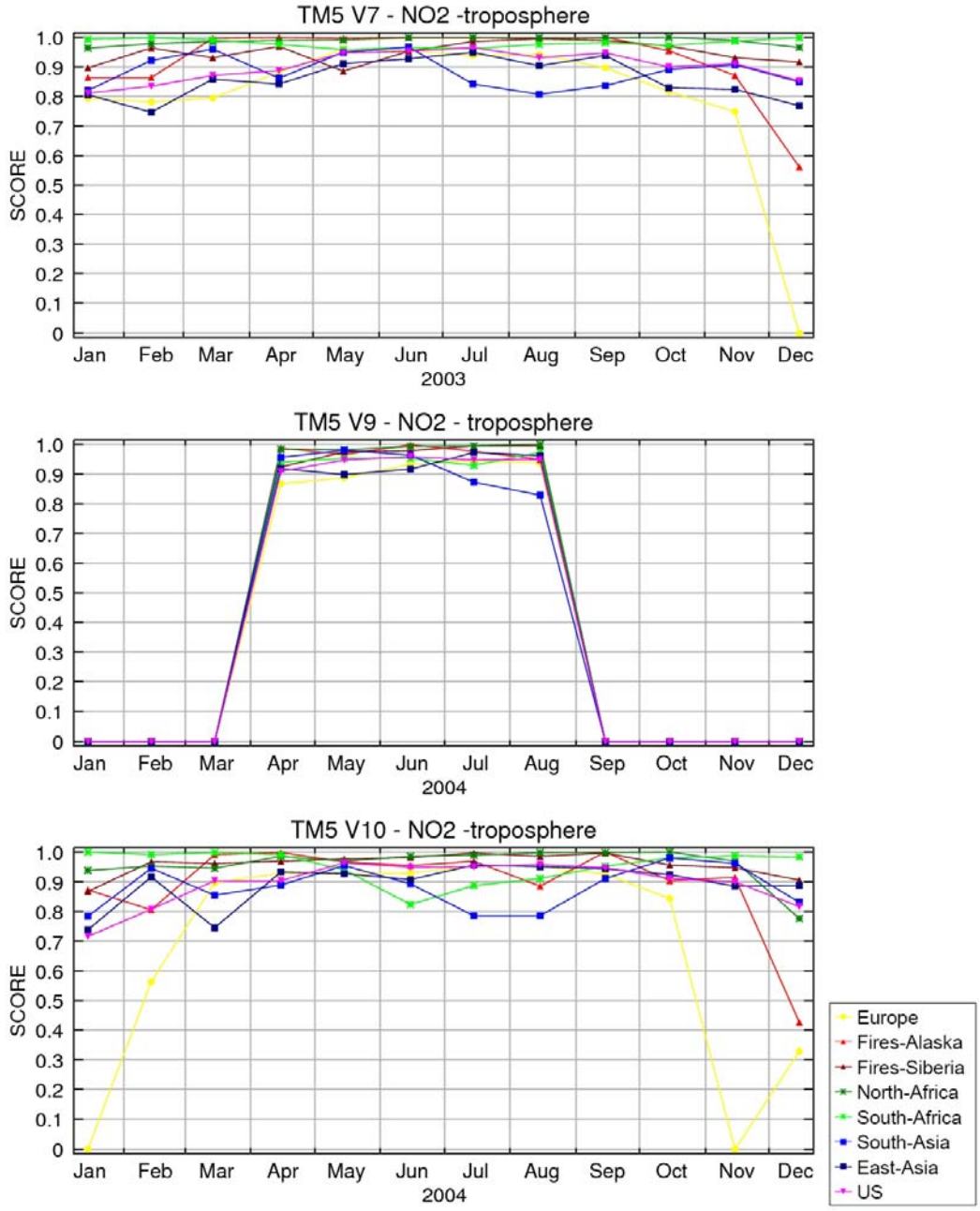


Figure 8: NO_2 monthly scores, averaged for the stratospheric regions, for TM5 stand-alone runs V7 (top), for the year 2003, and V9 (middle) and V10 (bottom), for the year 2004.

b) GEMS-GRG reanalysis – MOZART f026

Stratosphere

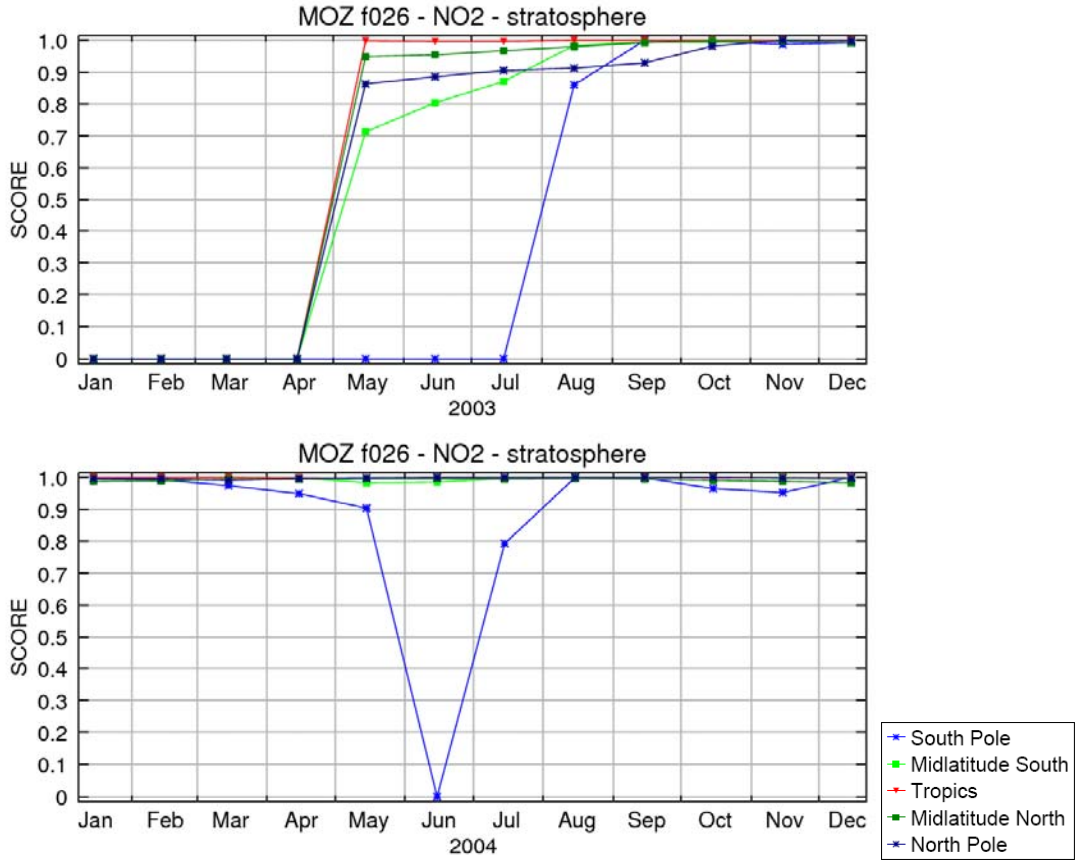


Figure 9: NO_2 monthly scores, averaged for the stratospheric regions, for MOZART reanalysis run f026 for the years 2003 to 2006 (top to bottom).

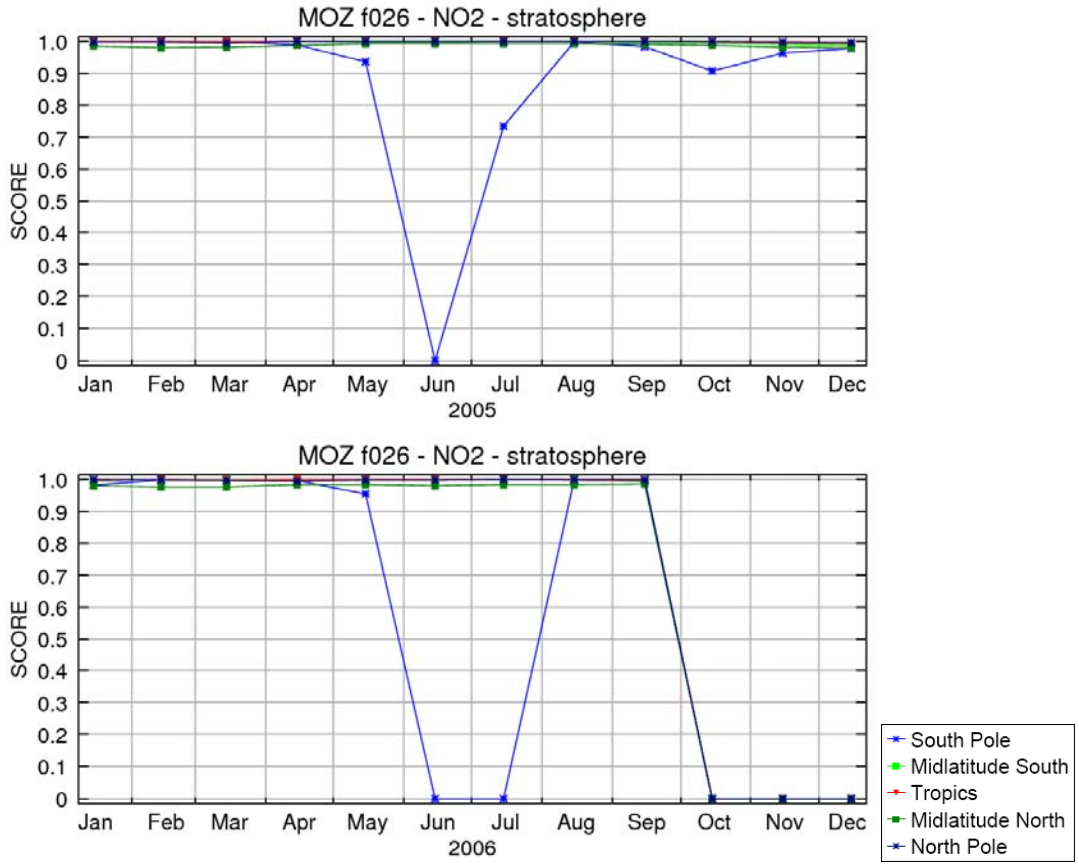
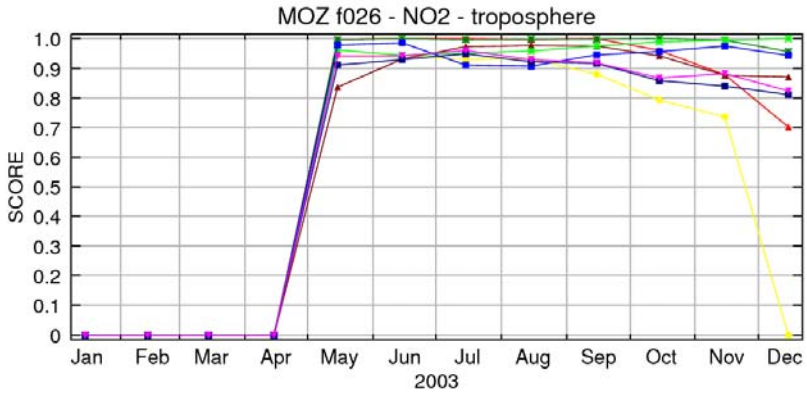
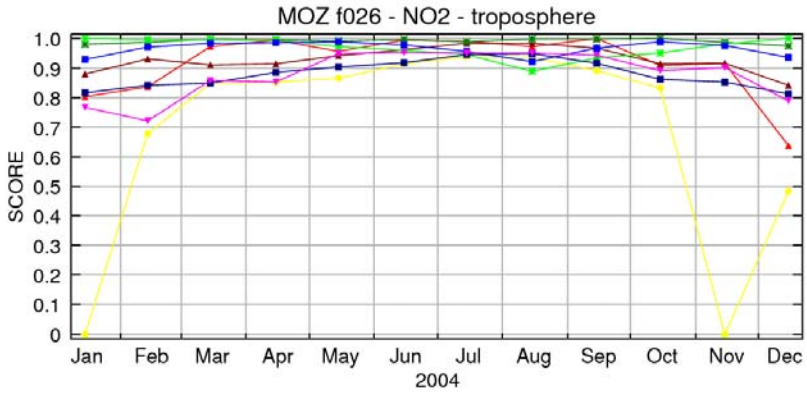


Figure 9: cont.

Troposphere





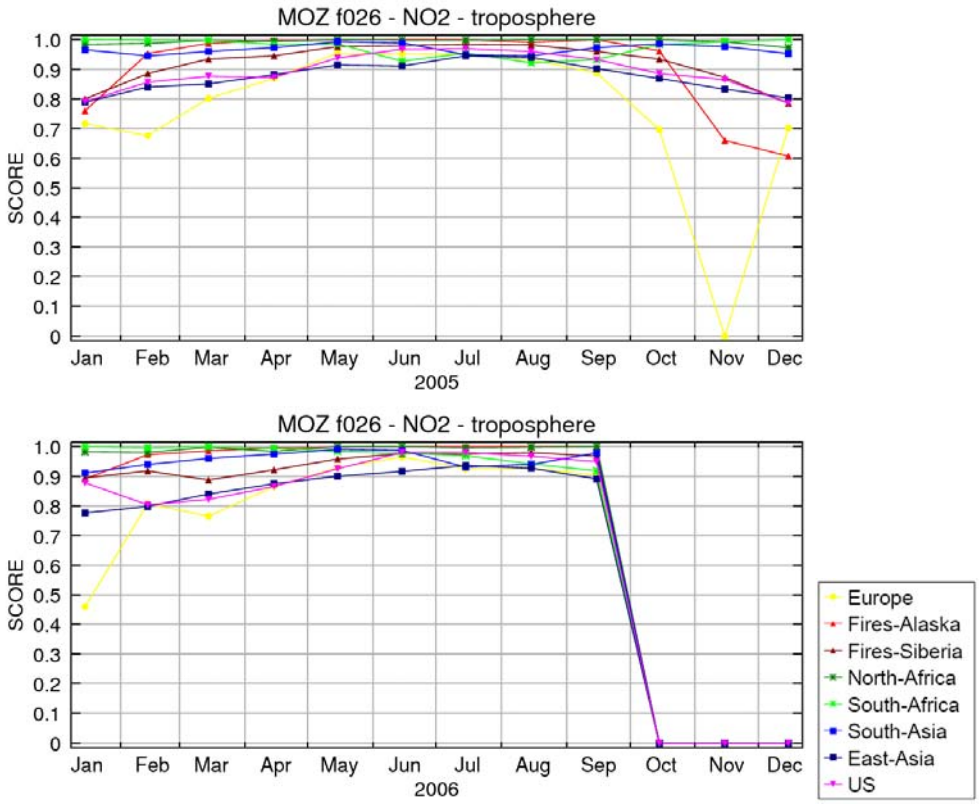
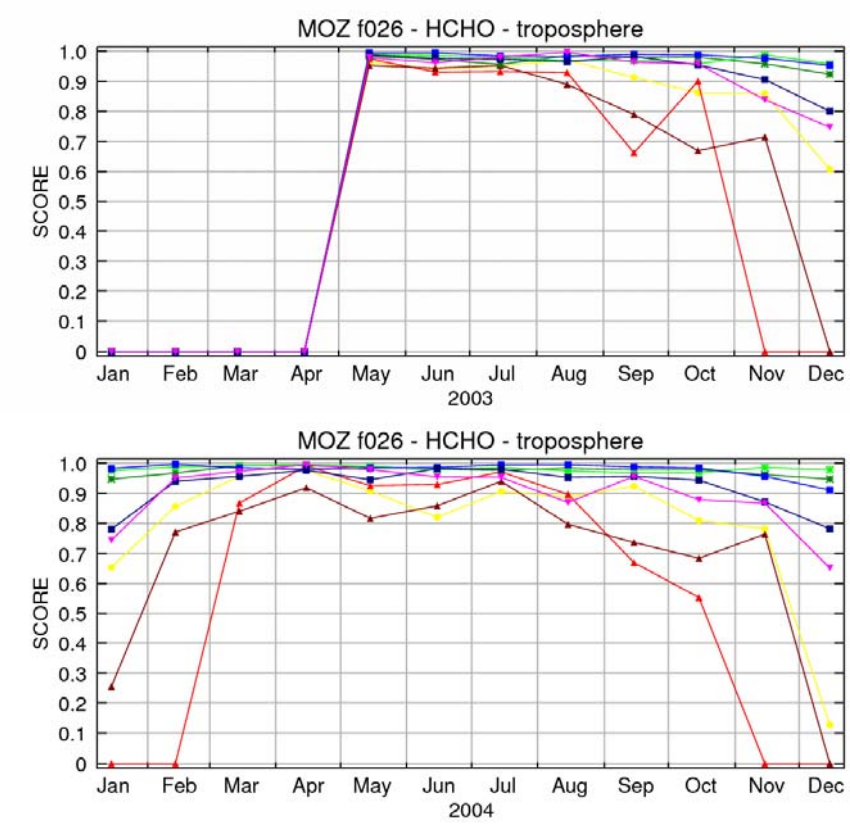


Figure 10: NO_2 monthly scores, averaged for the tropospheric regions, for MOZART reanalysis run f026 for the years 2003 to 2006 (top to bottom).



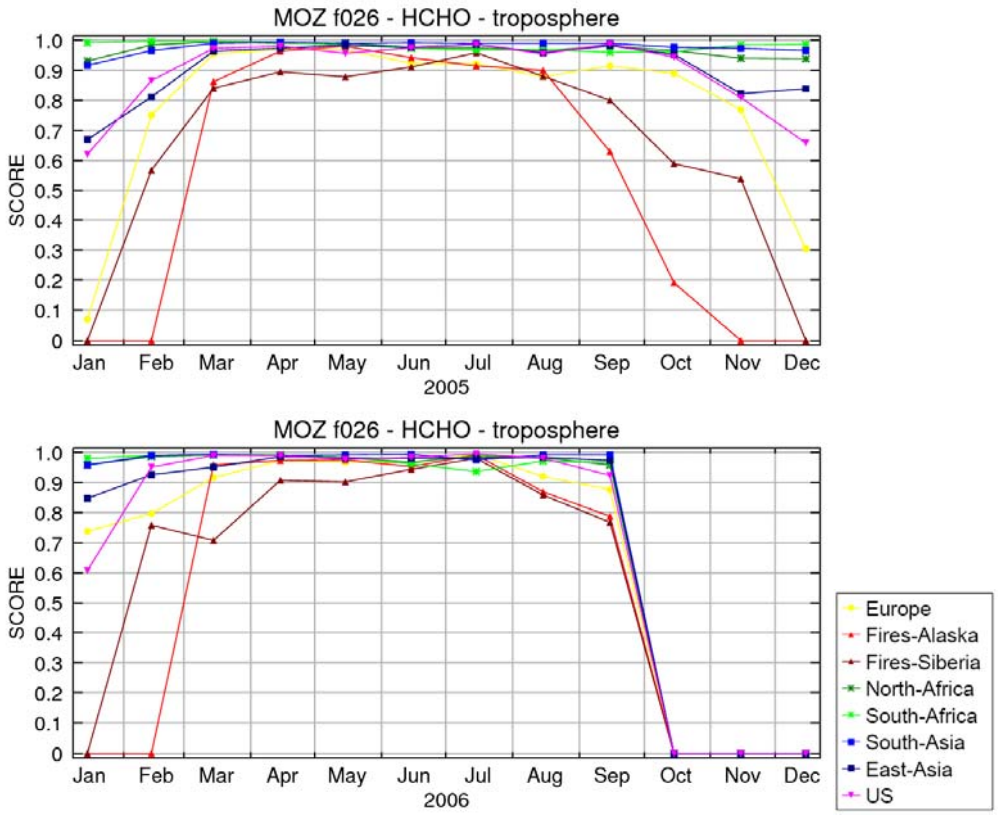


Figure 11: HCHO monthly scores, averaged for the tropospheric regions, for MOZART reanalysis run f026 for the years 2003 to 2006 (top to bottom).

c) GEMS-GRG forecast – MOZART ez2m

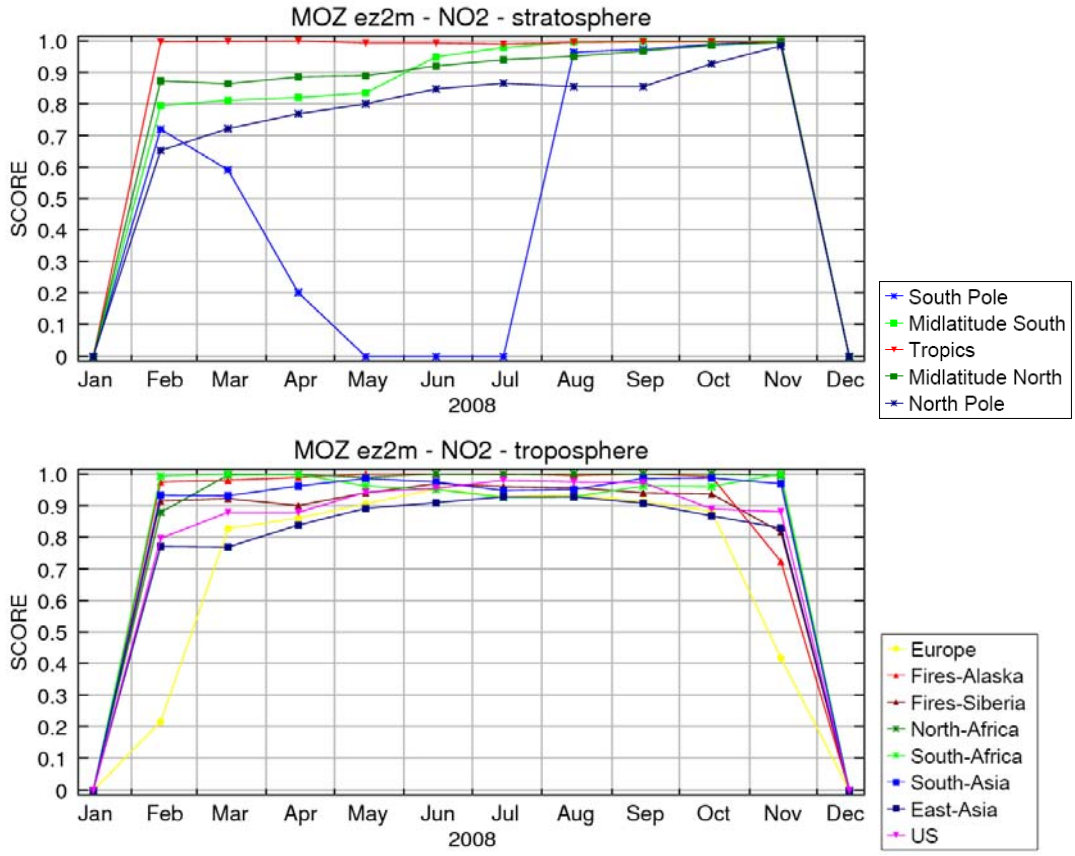


Figure 12: NO_2 monthly scores, averaged for the stratospheric (top) and tropospheric (bottom) regions, for MOZART forecast run ez2m for the year 2008.

ANNEX 9: List of contributors

Name, Firstname	**Institute**	**Email**
Athier Gilles	CNRS-LA	Gilles.Athier@aero.obs-mip.fr
Cammas Jean-Pierre	CNRS-LA	Jean-Pierre.Cammas@aero.obs-mip.fr
Chabrillat Simon	BIRA	Simon.Chabrillat@aeronomie.be
Daerden Frank	BIRA	Frank.Daerden@aeronomie.be
Elguindi Nellie	CNRS-LA	Nellie.Elguindi@aero.obs-mip.fr
Flemming Johannes	ECMWF	Johannes.Flemming@ecmwf.int
Flentje Harald	DWD	Harald.Flentje@dwd.de
Granier Claire	SA-UPMC	claire.granier@aero.jussieu.fr
Huijnen Vincent	KNMI	Vincent.Huijnen@knmi.nl
Inness Antje	ECMWF	Antje.Inness@ecmwf.int
Jones Luke	ECMWF	Luke.Jones@ecmwf.int
Katragkou Eleni	NKUA – AUT	katragou@auth.gr
Khokhar Fahim	SA-UPMC	khokhar@aero.jussieu.fr
Kins Lucia	DWD	Lucia.Kins-Hoefken@dwd.de
Law Kathy	SA-UPMC	kathy.law@aero.jussieu.fr
Lefever Karolien	BIRA	Karolien.Lefever@aeronomie.be
Leitao Joana	IUP	jleitao@iup.physik.uni-bremen.de
Melas Dimitris	AUT	melas@auth.gr
Moinat Philippe	METEO-F	Philippe.Moinat@cnrm.meteo.fr
Ordonez Carlos	CNRS-LA + UKMO	Carlos.Ordonez@metoffice.gov.uk
Peuch Vincent-Henri	METEO-F	Vincent-Henri.Peuch@meteo.fr
Reich Gabriele	DWD	Gabriele.Reich@dwd.de
Schultz Martin	FZJ	M.Schultz@fz-juelich.de

Stein Olaf	FZJ	O.Stein@fz-juelich.de
Thouret Valérie	CNRS-LA	Valerie.Thouret@aero.obs-mip.fr
Werner Thomas	DWD	Werner.Thomas@dwd.de
Zerefos Christos	NKUA	zerefos@geol.uoa.gr

Institutes:

- AUT : Laboratory of Atmospheric Physics, Aristotle University of Thessaloniki, Greece
- BIRA: Belgian Institute of Space Aeronomy, Ringlaan, 3, B-1180 Brussels, Belgium
- CNRS-LA: Centre National de la Recherche Scientifique, Laboratoire d'Aérologie, 14, avenue E. Belin, 31400 Toulouse, France
- DWD: Deutscher Wetterdienst, Meteorologisches Observatorium Hpbg, D – 82383 Hohenpeißenberg, Germany
- ECMWF, European Centre for Medium-Range Weather Forecasts, Shinfield Park, Reading, RG2 9AX, United Kingdom
- IUP: Institut für UmweltPhysik, Universität Bremen, Otto-Hahn-Allee 1, P.O. BOX 33 04 40 D-28334 Bremen Germany
- KNMI: Royal Netherlands Meteorological Institute, PO Box 201, NL-3730 AE De Bilt, Netherlands
- METEO-F: Météo-France, Centre National de la Recherche Météorologique, 42 avenue Gustave Coriolis, 31057 Toulouse cedex 1
- NKUA: Laboratory of Climatology and Atmospheric Environment, National Kapodistrian University of Athens, Greece
- SA-UPMC : Service d'aéronomie, Tour 45, Couloir 45-46, 3ème et 4ème étage (boîte 102) Université Pierre et Marie Curie, 4 place Jussieu, 75252 Paris Cédex 05, France
- UK Met Office, Exeter, United Kingdom