On a new assessment method for long-term chemistry-climate simulations in the UTLS based on IAGOS data: application to MOCAGE CCMI-REFC1SD simulation

Yann Cohen[1,2,*], Virginie Marécal[1], Béatrice Josse[1], and Valérie Thouret[2]

[1]Centre National de Recherches Météorologiques, CNRS-Météo-France, UMR 3589, Toulouse, France

[2]Laboratoire d'Aérologie, Université de Toulouse, CNRS, UPS, France

[*]Now at Laboratoire des Sciences du Climat et de l'Environnement, LSCE-IPSL (CEA-CNRS-UVSQ), Université Paris-Saclay, 91191 Gif-sur-Yvette, France

Correspondence to: yann.cohen@lsce.ipsl.fr

Abstract. A wide variety of observation data sets are used to assess long-term simulations provided by chemistry-climate models (CCMs) and chemistry-transport models (CTMs). However, the upper troposphere–lower stratosphere (UTLS) has hardly been assessed in the models yet. Observations performed in the framework of IAGOS (In-service Aircraft for a Global Observing System) combine the advantages of *in situ* airborne measurements in the UTLS with an almost global-scale sampling,

5   a ∼20-year monitoring period and a high frequency. If a few model assessments have been made using IAGOS database, none of them took advantage of the dense and high-resolution cruise data in their whole ensemble yet. The present study proposes a method to compare this large IAGOS data set to long-term simulations used for chemistry-climate studies. For this purpose, a new software (named Interpol-IAGOS) projects all IAGOS data on the 3D grid of the chosen model with a monthly resolution, since generally chemistry-climate models provide 3D outputs as monthly means. This provides a new IAGOS data set (IAGOS-

10   DM) mapped at the model's grid and time resolution. As a first application, the REF-C1SD simulation generated by MOCAGE CTM in the framework of CCMI phase-I has been evaluated during the 1994–2013 period for ozone (O_3) and the 2002–2013 period for carbon monoxide (CO). This comparison is exclusively based on the grid cells sampled by IAGOS, thus the assessed model output (MOCAGE-M) is obtained by applying a corresponding mask onto the grid. First, climatologies are derived from the IAGOS-DM product. Good correlations are reported between IAGOS-DM and MOCAGE-M spatial distributions. As

15   an attempt to analyse MOCAGE-M behaviour in the upper troposphere (UT) and the lower stratosphere (LS) separately, UT

and LS data in IAGOS-DM were sorted according to potential vorticity. From this, we derived O_3 and CO seasonal cycles in eight regions well sampled by IAGOS flights in the northern mid-latitudes. They are remarkably well-reproduced by the model for lower-stratospheric O_3 and also good for upper-tropospheric CO. The data projection onto the model's grid is a necessary step for a more accurate assessment, as it allows to filter out biases only due to either spatial or temporal resolution.

5   Beyond the MOCAGE REF-C1SD evaluation presented in this paper, the method could be used by CCMI models for individual assessments in the UTLS and for model intercomparisons with respect to IAGOS data set.

1   Introduction

Chemistry-climate models (CCMs) and chemistry-transport models (CTMs) are essential tools for understanding the atmospheric composition, for providing information where measurements are lacking, and for predicting air composition future evolution. Assessing and reducing uncertainties on the processes controlling its past and future changes can be achieved by

10   comparing an ensemble of simulations from different models while using the same simulation setup. Among the model intercomparison projects, the main goal of the Chemistry-Climate Model Initiative (CCMI: Eyring et al., 2013) lies on the reduction of the uncertainties in the multi-model projections involving stratospheric ozone, tropospheric composition and climate change, but also in a better understanding of the atmospheric processes relevant for these thematics. CCMI has taken over from both SPARC CCMVal (Chemistry-Climate Model Validation: SPARC, 2010) focused on the stratosphere and IGAC ACCMIP

15   (Atmospheric Chemistry-Climate Model Intercomparison Project: Lamarque et al., 2013) dealing mainly with tropospheric composition. In this framework, a set of simulations has been designed to address its objectives. Among them, the REF-C1SD experiment aims at assessing the ability of the models to reproduce the actual atmospheric composition for the recent climate time period. The task for each participating model thus consisted in simulating as realistically as possible the tropospheric and

20   stratospheric compositions in the last decades (1980–2010), following a common protocol.

Several studies have assessed the ability of REF-C1SD experiments, or previous similar simulations of air composition under recent climate conditions, to reproduce the mean tropospheric and/or stratospheric composition, by the use of monthly mean climatologies from observation data sets as reference, mostly from space. Froidevaux et al. (2019) based the evaluation of the REF-C1SD run from the CESM1-WACCM model on zonal monthly means of the stratospheric ozone column, using

25   the Microwave Limb Sounder on Aura satellite (Aura-MLS) and the multi-satellite data set merged in the framework of the GOZCARDS (Global OZone Chemistry And Related trace gas Data records for the Stratosphere) project. As described in Young et al. (2018), tropospheric ozone fields provided by the ACCMIP participating models have been assessed, referring to zonally averaged mixing ratios from the Tropospheric Emission Spectrometer (Bowman et al., 2013), and tropospheric ozone

column from OMI-MLS (Young et al., 2013). Hu et al. (2017) also compared the OMI-MLS tropospheric ozone columns to a GEOS-Chem simulation. The observed carbon monoxide (CO) columns from Measurement Of Pollution In The Troposphere (MOPITT) instrument served as the reference in the assessment of modeled tropospheric CO, notably from the REF-C1SD simulation generated by the GMI CTM over the period 2000–2010 (Strode et al., 2016), and from the Community Earth System

5   Model (CESM1) CAM4-Chem model (Tilmes et al., 2016).

Only few studies compared observations (*in situ* measurements or from space) and CCMI REF-C1SD or similar simulations, focusing on the upper troposphere–lower stratosphere (UTLS). However, the latter is a key region regarding both the ozone (O_3) radiative forcing (Riese et al., 2012) and the stratosphere–troposphere exchanges (STEs) that substantially influence tropospheric ozone levels (e.g. Tao et al., 2019), albeit with a high uncertainty due to their different representations in

10   models (Stevenson et al., 2006). Smalley et al. (2017) referred to the Aura-MLS measurements in the assessment of a 21st century projection (REF-C2) from twelve CCMs, focusing on their lower-stratospheric water vapour fields, during the 2004–2014 time period. *In situ* measurements with ozonesondes as part of the World Ozone and Ultraviolet radiation Data Center (WOUDC) have been compared to the REF-C1SD simulations from CMAM and EMAC models during the 2005–2010 time period (Williams et al., 2019). In addition to ozonesondes, aircraft measurements from different campaigns were used in the

15   evaluation of the REF-C1SD simulations from the model CESM1 CAM4-Chem (Tilmes et al., 2016).

Among available observation data sets, the commercial aircraft measurements from the on-going IAGOS European Research Infrastructure (In-service Aircraft for a Global Observing System: Petzold et al., 2015, http://www.iagos.org) are well designed to study ozone and CO on the long term, notably in the UTLS (Cohen et al., 2018). IAGOS observations started in August 1994 for ozone and in December 2001 for CO. They are characterized by a high spatio-temporal resolution and a wide coverage with

20   most data gathered at cruise levels (9–12 km above sea level). Thus, IAGOS database is suited to assess long-term simulations in this altitude range. Recently, its ozone data have been used to evaluate simulations from the models CESM1 CAM4-Chem (Tilmes et al., 2016) and GEOS-Chem (Hu et al., 2017) during the periods 1995–2010 and 2012–2013, respectively. Tilmes et al. (2016) used the IAGOS measurements gathered in the vicinity of Narita airport (Japan) only, and the comparison made by Hu et al. (2017) only spread over 2 years, while IAGOS ozone data are available since 1994 and covering a wide area, especially

25   in the northern mid-latitudes from Western North America to East Asia. Gaudel et al. (2015) performed an evaluation of the MACC (Monitoring Atmospheric Composition and Climate) reanalysis over Europe during 2003–2010, using IAGOS O_3 and CO measurements. However, this comparison was led using frequent simulation outputs, thus their methodology is not adapted to the assessment of the 3D outputs from the REF-C1SD simulations, which are monthly averages. Consequently, the IAGOS

cruise data in the UTLS have been used neither as a whole ensemble nor to derive a monthly climatology for the evaluation of long-term chemistry-climate simulations. This is what we propose in the present paper.

To compare REF-C1SD simulations against IAGOS data, interpolating the simulation outputs onto the high-resolution observations is not possible because of the coarse spatio-temporal resolution of the REF-C1SD outputs. It would be very expensive

5 computationally and not meaningful to interpolate monthly-mean model data onto very high frequency (a few seconds) IAGOS measurement locations. Alternatively, the comparison could be performed after mapping the high resolution IAGOS data on the model grid, on a monthly basis. Several gridding methods already exist for *in situ* measurements. Some of them consist in interpolating the neighbouring measurements points onto each gridpoint (e.g. New et al., 2000). However, it requires to memorize all the measurement locations for a whole month. It is thus convenient for measurements with regular locations only, but

10 their use on the IAGOS database would be expensive computationally as well. Variational methods are also widely employed (e.g. Bourassa et al., 2005) but they concern integration, which is not our purpose. The present study aims at providing a new methodology designed to generate a gridded monthly data set from the IAGOS measurements, in order to evaluate REF-C1SD types of simulations. We also propose a set of relevant diagnostics for the model evaluation against IAGOS data mapped on the model grid. These diagnostics originate from Cohen et al. (2018) that studied climatologies and trends in ozone and CO,

15 based on the analysis of the quasi-totality of the IAGOS database. The use of such a high spatial and temporal resolution data set allows to account for inter-regional differences that could not be highlighted with zonal means. Its projection into a model grid suits well to the constraint of working on monthly outputs from multi-decadal simulations like REF-C1SD. In order to demonstrate the interest of the new methodology and its associated diagnostics, we perform the assessment on one of the REF-C1SD simulations, that of the MOCAGE CTM model.

20 In Sect. 2, we describe briefly the IAGOS observations, the CCMI model intercomparison project, the MOCAGE CTM that we use in this study, and its configuration for the REF-C1SD simulation. In Sect. 3, we present the methodology proposed to map the IAGOS data set on the model grid on a monthly resolution, the chosen statistical metrics for models' evaluation and the different assessment diagnostics. In Sect. 4, we present a first application of this methodology on the evaluation of the MOCAGE REF-C1SD simulation. Strengths and weaknesses of the methodology and the chosen diagnostics are discussed.

25 Conclusions are given in Sect. 5.

2   Observations and simulation

2.1   The IAGOS observations

The European Research Infrastructure IAGOS (Petzold et al., 2015, http://www.iagos.org) provides *in situ* measurements on board several commercial aircrafts. The observations used hereafter have been performed in the framework of the on-going
5   IAGOS-Core program that followed the MOZAIC program (Marenco et al., 1998). Ozone (resp. CO) measurements started in August 1994 (resp. December 2001), based on an UV (resp. IR) absorption technology, with an accuracy of 2 ppb (resp. 5 ppb), a precision of 2% (resp. 5%) and a time resolution of 4 s (resp. 30 s). Further information about the instruments can be found in Marenco et al. (1998) and Thouret et al. (1998) for O_3, and in Nédélec et al. (2003) for CO. Nédélec et al. (2015) present a more recent evaluation of both ozone and CO instruments in the frame of IAGOS.

10 The IAGOS observations (referring to the IAGOS-Core database hereafter) sample frequently the whole troposphere nearby airports, measuring vertical profiles during ascent and descent phases, and the UTLS during the cruise phases, mostly in the northern mid-latitudes where most of the flight observations are gathered. In these latitudes, a recent analysis of O_3 and CO climatologies and trends based on almost two decades of IAGOS cruise measurements has been performed in Cohen et al. (2018). In addition to global climatologies, the same analysis also focused on eight well-sampled regions in the UT and the LS
15 separately. In order to generate comparable results with the latter, this study focuses on the same time period (1994–2013) and, where relevant, on the same regions.

2.2   The CCMI project and the REF-C1SD experiment

CCMI phase-1 gathers a community of 18 chemistry-climate models (CCMs) and two CTMs, which description is given in the review of Morgenstern et al. (2017). A series of experiments have been designed to model tropospheric and stratospheric air
20 composition for past, present and future climates. For each experiment, a common protocol is recommended to all participating models. Amongst the CCMI simulations, the REF-C1SD reference experiment aims at modelling as realistically as possible the day-to-day tropospheric and stratospheric compositions in a recent climate, using specified dynamics (SD). For this purpose, as described in Eyring et al. (2013), the simulations are driven by (or nudged towards) dynamical reanalyses data sets (typically ERA-Interim or MERRA), and expanding from 1980 until 2010. For this long-term simulation, the 3D outputs fields of species
25 concentrations are archived as monthly means.

2.3 The MOCAGE model and the simulation set-up

The MOCAGE model (MOdèle de Chimie Atmosphérique à Grande Echelle: Josse et al., 2004; Guth et al., 2016) is an offline global chemistry-transport model (CTM). The chemical scheme is composed by the coupling of the RACM (Regional Atmospheric Chemistry Mechanism: Stockwell et al., 1997) and the REPROBUS (REactive Processing Ruling the Ozone BUdget

5 in the Stratosphere: Lefèvre et al., 1994) schemes, corresponding to tropospheric and stratospheric chemistry, respectively. MOCAGE REF-C1SD simulation is run using a global domain at a $2° \times 2°$ horizontal resolution, and 47 vertical levels, in σ-hybrid pressure, distributed from the surface up to $\sim$5 hPa. The simulation is driven by the meteorological fields from the ERA-Interim reanalysis. The biomass burning and anthropogenic emissions come from the GFEDv2 and MACCity inventories, respectively. The latter is characterized by a 10-year resolution and a linear interpolation is applied to derive yearly emissions.

10 The period spreads from August 1994 until December 2013, consistently with Cohen et al. (2018). The first 14 years come from MOCAGE REF-C1SD simulation originally produced for CCMI project. For the years out of the period covered by the experiment, the MOCAGE REF-C1SD run has been extended until December 31, 2013 using the same code and inputs as in the original MOCAGE CCMI REF-C1SD simulation.

3 Methodology

15 The objective of the proposed methodology is to make possible the comparison between the full IAGOS database and the 3D monthly mean volume mixing ratios from CTMs and CCMs simulations. Our approach consists in distributing the IAGOS observations, performed every 4 s, on a given model grid. A first application is proposed on MOCAGE REF-C1SD run, characterized by a $\sim$200 km horizontal resolution in the mid-latitudes, and a $\sim$800 m vertical resolution in the UTLS. In order to account for the eccentricity of the measurements inside one given cell, we chose a reverse linear interpolation at the first

20 order, as described in Sect. 3.1 and illustrated in Fig. 1. The subsequent gridded monthly means are derived using weighted averages, as described in Sect. 3.2, and are directly comparable to the model monthly mean outputs.

In a first step, this is used for a statistical evaluation of MOCAGE REF-C1SD climatologies on a hemispheric scale over the periods December 1994–November 2013 for O_3 and December 2001–November 2013 for CO. The data processing used to produce the climatologies and the statistical metrics chosen are presented in Sect. 3.3. In a second step, we attempt to go further

25 in the assessment of MOCAGE simulation by evaluating separately the upper troposphere and the lower stratosphere. For this purpose, the discrimination between the grid points mostly representative of the UT or the LS is necessary. As in Cohen et al.
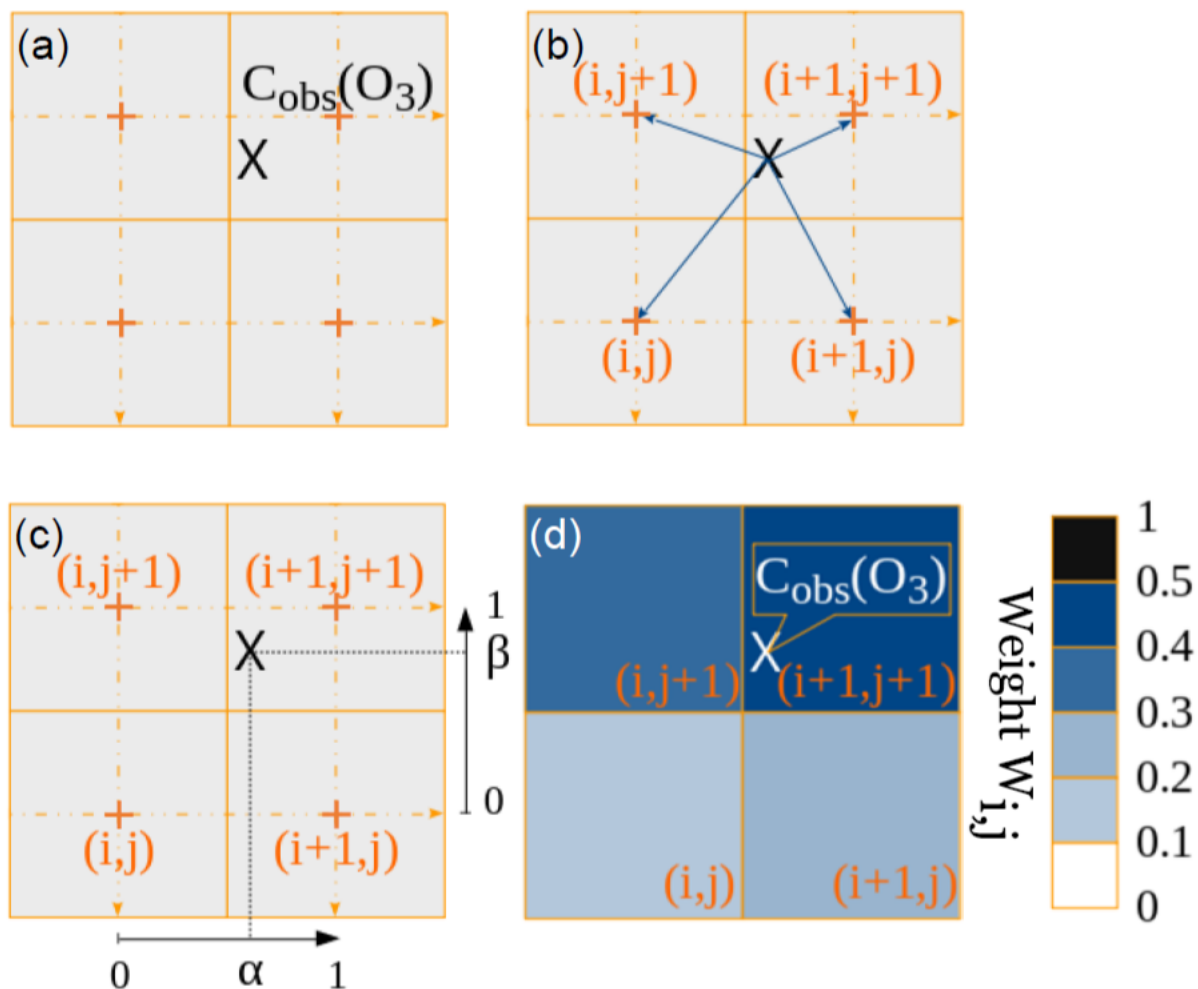
Figure 1. Schematic of the method used for the distribution of the observations on the model grid, represented in two dimensions for simplicity. (a) shows the location of a chosen measurement point in the model's grid. The steps of the method are (b) to locate the model's grid points (orange crosses) closest to this location; (c) to calculate a weighting coefficient for each dimension (α and β), depending on the distance between the measurement point and the "bottom-left" grid point; (d) the calculation of the weight for the four closest grid points. As indicated in the colour scale on the right, this weight ranges between 0 and 1.

(2018), this has been done with respect to Ertel potential vorticity (PV) and applied in eight northern mid-latitude regions selected because of their high level of sampling by IAGOS. The methodology used is explained in Sect. 3.4.

3.1 Reverse interpolation of a given measurement point on the model grid

At a given point where IAGOS measured a mixing ratio $C_{obs}(X)$ for an X species, the algorithm presented here locates its

5   position on the model grid defined by its longitude, latitude and σ-hybrid pressure coordinates. More precisely, we locate the

model grid point which is the closest west and south of, and below (in altitude) the observation point and which corresponds to the ith, j^{th} and k^{th} grid point coordinates respectively. As shown in Fig. 1c, a normalized weighting coefficient is then computed for each dimension (coefficients α, β, γ), increasing linearly with the distance between the measurement point and the (i, j, k) grid point. Note that the γ vertical coefficient is derived from log-pressure coordinates. Finally, a resulting 3D weight is

5   computed for each of the eight closest cells. By noting the variable indexes I, J and K belonging to the ensembles {i, i+1}, {j, j+1} and {k, k+1} respectively, we define the functions f_I, g_J and h_K which values depend on α, β and γ respectively, such as:

$$f_I(\alpha) = \begin{cases} 1-\alpha & \text{if } I = i \\ \alpha & \text{if } I = i+1 \end{cases} \quad ; \quad g_J(\beta) = \begin{cases} 1-\beta & \text{if } J = j \\ \beta & \text{if } J = j+1 \end{cases} \quad ; \quad h_K(\gamma) = \begin{cases} 1-\gamma & \text{if } K = k \\ \gamma & \text{if } K = k+1 \end{cases} \tag{1}$$

The resulting weight for each of the grid points surrounding the measurement location is thus defined as the following product:

10    $$W_{I,J,K}(\alpha,\beta,\gamma) = f_I(\alpha).g_J(\beta).h_K(\gamma) \tag{2}$$

In this way, as illustrated in Fig. 1d, for a given cell (I, J, K) amongst the eight closest ones, this weight decreases with the distance between the measurement point and the model grid point. Note that since the simulation outputs are monthly averages, we use the monthly mean surface pressure for determining the hybrid σ-pressure on the 47 vertical grid levels for a given model longitude/latitude. Although the surface pressure can show an important intra-monthly variability, we verified that a 30 hPa

15   change at surface would cause a variation weaker than 2 hPa on a given vertical grid level in the UTLS. Although caution is needed while treating low-altitude measurements, the monthly resolution on the surface pressure field thus has a negligible impact on the cruise IAGOS data distribution onto the model vertical grid.

3.2   Deriving the monthly mean values from observations

The weight coefficients defined above correspond to one single observation data point. To obtain monthly averages from the

20   observation data set, the last step consists in summing up all the values measured in the vicinity of the (i, j, k) grid point for each month. Thus, for a given grid point (i, j, k), we define n as the index for the measurement performed in its vicinity during the considered month, and the corresponding mixing ratio for the species X is noted $C_{obs,\,n}(X)$, and N the total amount of

measurements performed in this vicinity. The monthly value of the X mixing ratio at (i, j, k) is then derived with the equation:

$$X_{i,j,k} = \frac{\sum_{n=1}^{N} W_{i,j,k,n}(\alpha,\beta,\gamma).C_{\text{obs, n}}(X)}{\sum_{n=1}^{N} W_{i,j,k,n}(\alpha,\beta,\gamma)} \tag{3}$$

where the denominator is equivalent to the amount of measurement points performed in the (i, j, k) grid cell during the chosen month. Hereafter, we refer it as N_{eq}.

5 In the end, this method yields monthly fields of IAGOS O_3 and CO mixing ratios (or any other variable measured by IAGOS, e.g. water vapour) projected on the MOCAGE grid points where IAGOS data are available. This data set is named IAGOS-DM hereafter, the suffix DM referring to the distribution on the model grid. With this method, the cruise observation data are distributed on MOCAGE vertical levels from level 28 up to level 22 included, corresponding to the ∼360–175 hPa interval. Note that the measurement points on the MOCAGE vertical levels below level 28 (∼360 hPa) are considered as corresponding

10 to ascent or descent phases of the flights. These measurements are not processed, since they are only available in small areas close to airports. Levels 27 and 28 also correspond to these phases but include cruise measurements above elevated lands, since hybrid sigma-pressure levels tend to follow land elevation. In order to compare the observations and the model at the same locations and months, we apply a mask on the MOCAGE REF-C1SD simulation outputs that allows us to account only for the IAGOS-DM sampled grid points. The subsequent data set is named MOCAGE-M, the M letter referring to the mask. Thus,

15 IAGOS-DM and MOCAGE-M data sets are fully consistent and can be used to make gridpoint-by-gridpoint and month-by-month comparisons.

3.3 Methodology for the assessment of the climatologies

3.3.1 Filtering conditions

For the climatological part of this study, we chose to perform a seasonal and a yearly analysis. Avoiding sampling biases

20 where and when IAGOS-DM data (counted as N_{eq}) are not numerous enough requires that the seasonal sample N_{eq} reaches a minimum threshold to be selected (noted N_{thres}). We chose to set this N_{thres} limit depending on latitude to account for the varying gridbox area, and on the chemical tracer to account for the shorter period for CO measurements compared to O_3. N_{thres} therefore decreases with latitude following a cosine function, similarly to the model horizontal grid cell areas. The reference threshold $N_{\text{thres, ref}}$ corresponds to O_3 measurements for gridbox areas during a given season, over the whole period. It has

25 been set to $N_{\text{thres, ref}} = 100$ as a compromise between sampling robustness and a large-enough amount of data in IAGOS-DM sample. Accounting for the shorter CO measurement period compared to O_3, the corresponding N_{eq} threshold for this species

is derived by applying a factor 0.6, leading to 60. Last, the reference filter is defined seasonally. The filters defined here are thus quadrupled for yearly climatologies.

3.3.2 Statistical metrics for assessing the climatologies

Quantifying a simulation assessment requires the use of statistical parameters. This paragraph aims at defining the chosen

5   metrics, and at justifying this choice. Pearson's coefficient is a key result from linear regressions. It is used to quantify the correlation between two signals. If we call $(m_i)_{i\in[\![1,N]\!]}$ and $(o_i)_{i\in[\![1,N]\!]}$ the lists of modelled and observed values respectively, their correlation is defined as:

$$r = \frac{1}{N} \frac{\sum_{i=1}^{N}(m_i - \bar{m})(o_i - \bar{o})}{\sigma_m \sigma_o} \tag{4}$$

where $\bar{m}$ and $\bar{o}$ are the mean values and σ_m and σ_o their respective standard deviations. Quantifying total biases and mean

10   errors is also primordial in a model assessment. However, the use of the absolute mean bias and root mean square error (RMSE) may not be relevant for climatological purposes because of a strong influence from observed outliers. In our context, another inconvenient lies in the strong vertical O_3 gradient near and above the tropopause. It tends to induce a strong absolute bias with respect to the tropospheric mixing ratios, since it makes the O_3 absolute mean bias and RMSE mainly depending on the highest vertical grid cells. The normalized bias metric (and associated standard error) is chosen for a better representativeness

15   of biases for both low and high mixing ratios. The modified normalized mean bias (MNMB) and the fractional gross error (FGE) are respectively defined as:

$$MNMB = \frac{2}{N} \sum_{i=1}^{N} \frac{m_i - o_i}{m_i + o_i} \tag{5}$$

and

$$FGE = \frac{2}{N} \sum_{i=1}^{N} \left| \frac{m_i - o_i}{m_i + o_i} \right| \tag{6}$$

20   ## 3.4 Methodology for assessing the seasonal cycles in the UT and in the LS

A second part of this assessment targets the behaviour of the model in the UT and the LS. The diagnostics we use for this purpose are adapted from Cohen et al. (2018). In the latter study, based on Thouret et al. (2006), the tropopause for each IAGOS individual measurement was defined as the 2 PVU isosurface derived from the ECMWF operational analysis, with

a 3-hour resolution, before deriving monthly means in the two layers. The present work is based on monthly gridded fields, including potential vorticity (PV). Consequently, determining whether a given cell is mostly composed by tropospheric or stratospheric air masses is achieved with a monthly resolution. For this purpose, we use the PV from the dynamical field (based on ERA-Interim), yielding 6-hourly PV values which were averaged monthly to match the simulation outputs. A given

5   grid point is then considered as belonging to the UT if its monthly PV is lower than 2 potential vorticity units (PVU), and to the LS if the PV is greater than 3 PVU. The cells which PV ranges between 2 and 3 PVU are considered as belonging to the transition zone separating the two layers and are not selected. In order to enhance the distinction between the UT and the transition zone, the first model level below the 2 PVU threshold is also filtered out from the UT. The 2 PVU threshold is derived from a log-pressure interpolation between the grid points. We also filter out the grid boxes where this PV classification is not

10   consistent with observed O_3 mixing ratio, i.e. where O_3 level reaches 140 ppb in the UT and where it goes under 60 ppb in the LS. It avoids a supplementary bias based on errors in the dynamical field only. These thresholds on O_3 mixing ratio were chosen according to the O_3 seasonal cycles shown in Fig. 3.7 in Cohen et al. (2018), where the upper boundary linked to the interannual standard deviation in the UT is less than 100 ppb and where the lower boundary in the LS is greater than 100 ppb. We estimated that a supplementary 40 ppb interval would limit an exaggerated filtering of grid cells monthly values.

15   As in Cohen et al. (2018), we focus our analysis on the seasonal cycles for eight regions in the northern mid-latitudes that are well sampled by IAGOS. Their coordinates and their corresponding sampling are detailed in Table 1 in Cohen et al. (2018). Because of the $2° \times 2°$ horizontal grid resolution in the simulation, we applied a $1°$ eastward or northward shift on the odd-coordinated edges. The subsequent regions defined in this paper are shown in Fig. 2. For each of them, the monthly means are calculated by averaging the gridded monthly means separately in the UT and the LS. The latter values were defined as

20   described in Sect. 3.1 and 3.2.

In the previous study, the regional monthly means with N_{eq} lower than 300 were filtered out. Here, due to the loss of data caused by the monthly resolution, we lowered this minimum threshold to 150 in order to keep taking the less sampled regions into account, as Western North America and Siberia. Still, we kept the criterion from Cohen et al. (2018) which required at least 7 days between the first and last measurements, probably ensuring the averages to be representative of a synoptic timescale.

25 Following the same study, the computation of the seasonal cycles is based on the years exhibiting seven available months or more, distributed on three seasons at least. This criterion avoids biases linked to the inter-seasonal differences in the sampling, thus ensuring a good representativeness of the whole year.
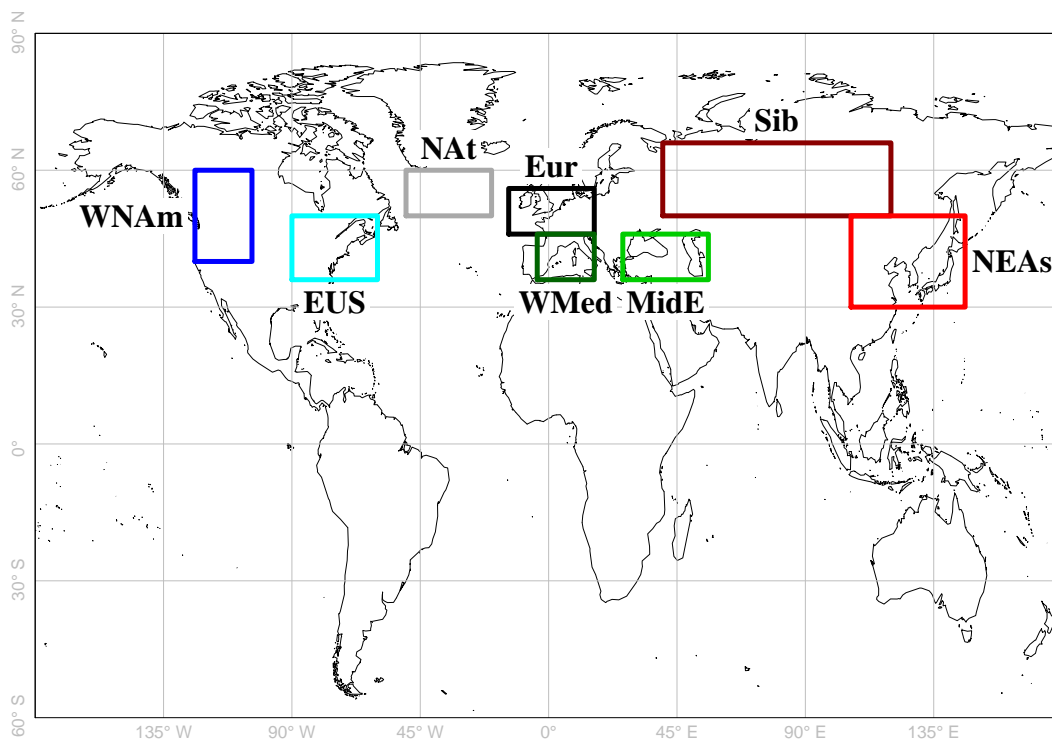
Figure 2. Map of the regions selected for this study based on Cohen et al. (2018), adapted to MOCAGE global grid. From West to East, the acronyms stand for Western North America, Eastern United-States, Northern Atlantic, western Europe, West Mediterranean basin, Middle East, central Siberia and Northeast Asia.

4 Results

4.1 Horizontal climatologies

Figures 3 and 4 show the yearly mean climatologies, respectively for O_3 and CO. The level 22 is seldom reached by the IAGOS measurements, and the levels 27 and 28 are sampled only in the vicinity of airports. Thus, only the levels 26 up to 23

5 are represented in these figures. Additionally, the seasonal mean climatologies are available in Appendix A.

In Fig. 3, IAGOS-DM and MOCAGE-M show similar geographical structures. In the tropics and subtropics, the O_3 amounts are close, with consistent poleward gradients. Both have maxima located above Northeast Canada. The O_3 mixing ratio in the northern mid-latitudes is underestimated in the model for the levels 24–26, and close to the observations for the level 23. The seasonal climatologies in Figs. A1–A4 show that this feature is representative of spring and fall, whereas ozone tends to
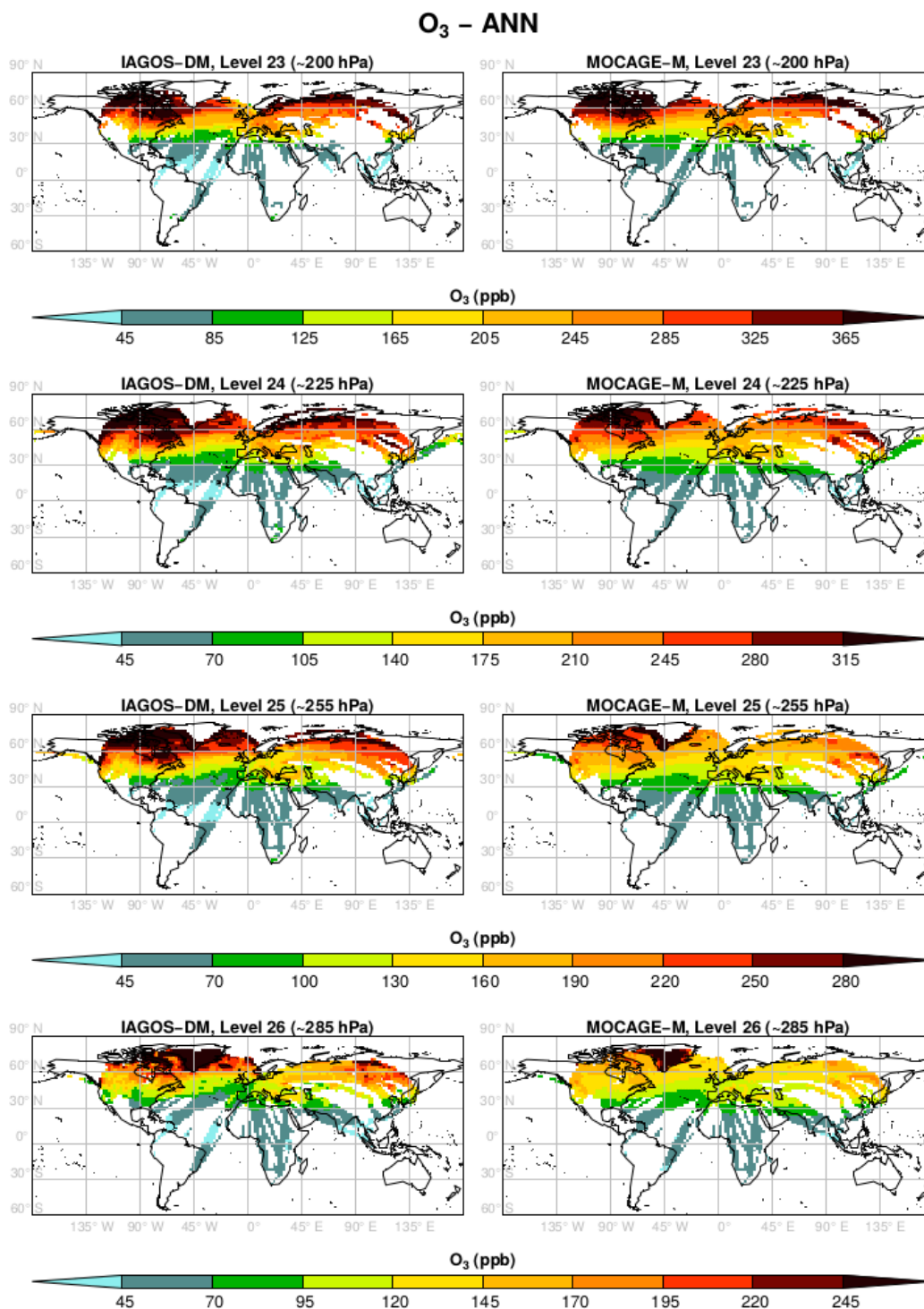
Figure 3. Mean horizontal distribution of the O_3 mixing ratio (ppb) from the model levels 26 (∼320 hPa, bottom panels) to 23 (∼200 hPa, top panels) in IAGOS-DM (left panels) and MOCAGE-M (right panels) during the period December 1994–November 2013.

be underestimated (resp. overestimated) in all vertical grid levels in summer (resp. winter). Note that the discontinuity over Greenland is due to its topography causing a steep elevation of the vertical grid levels.

In Fig. 4, CO also shows a good correlation between the two data sets, notably with the same maxima and minima locations. But the CO mixing ratio is generally overestimated by the model, especially over East Asia and India. In the northern mid-
5 latitudes, the seasonal climatologies in Figs. A5–A8 generally show an overestimation in winter and spring and a less-visible underestimation in summer and fall.

Figure 5 proposes a synthesis for the comparison between the yearly climatologies over the whole period. The same figures can also be found for each season in Appendix B. The linear regression parameters indicated in the graphs show a strong geographical correlation, its coefficient spreading from 0.74 up to 0.97. The correlation is better for O_3 (> 0.9), and at higher
10 levels for both species. Consequently, the geographical distributions in O_3 and CO are well reproduced in the simulation. Their stronger correlations at higher levels suggest a remarkably good consistence of the modelled stratospheric composition with the observations, showing its ability to simulate stratospheric chemistry and transport. The same feature is visible with the regression fit, showing a lower bias for O_3, and at highest levels. With respect to the 1:1 line, levels 25 and 26 are characterized by an overestimation of the lower part of the O_3 distribution (< 120 ppb) and by an underestimation of the higher part.
15 Concerning CO, the highest values (generally > 100 ppb) correspond to the strongly emitting and convective regions: South Asia, East Asia and tropical Africa. A supplementary test (not shown here) has identified the high mixing ratios close to the 1:1 line at tropical African points, whereas the high mixing ratios with a positive bias were associated both to South and East Asia areas. The latter can be due to an overestimation of convection in this region and/or an overestimation in the inventory for Asian emissions. On the contrary, CO above tropical Africa shows good results, indicating a realistic combination between
20 convection and emissions.

The method proposed in this paper to evaluate MOCAGE REF-C1SD against IAGOS data in the UTLS aims at being applied to other chemistry-climate simulations, like REF-C1SD simulations from other models. Since IAGOS is mapped onto the model vertical grid, the latter differing from one model to another, we also plotted a synthetic regression in Fig. 6, where all the points at all levels have been gathered into one single scatterplot. These summarized model performances
25 concerning mean spatial distributions are the final products of our evaluation methodology for horizontal climatologies. From the whole ensemble of ∼12,200 (∼10,500) sampled grid points for O_3 (CO), the correlation shows a good agreement between the simulation and the observations, especially for O_3 (r=0.95). Its regression fit is dominated by an overestimation for lower values (< 100 ppb) and an underestimation for higher values, especially between 200 and 300 ppb. Above 350 ppb, the balance between overestimated and underestimated O_3 values tends to be more equilibrated.
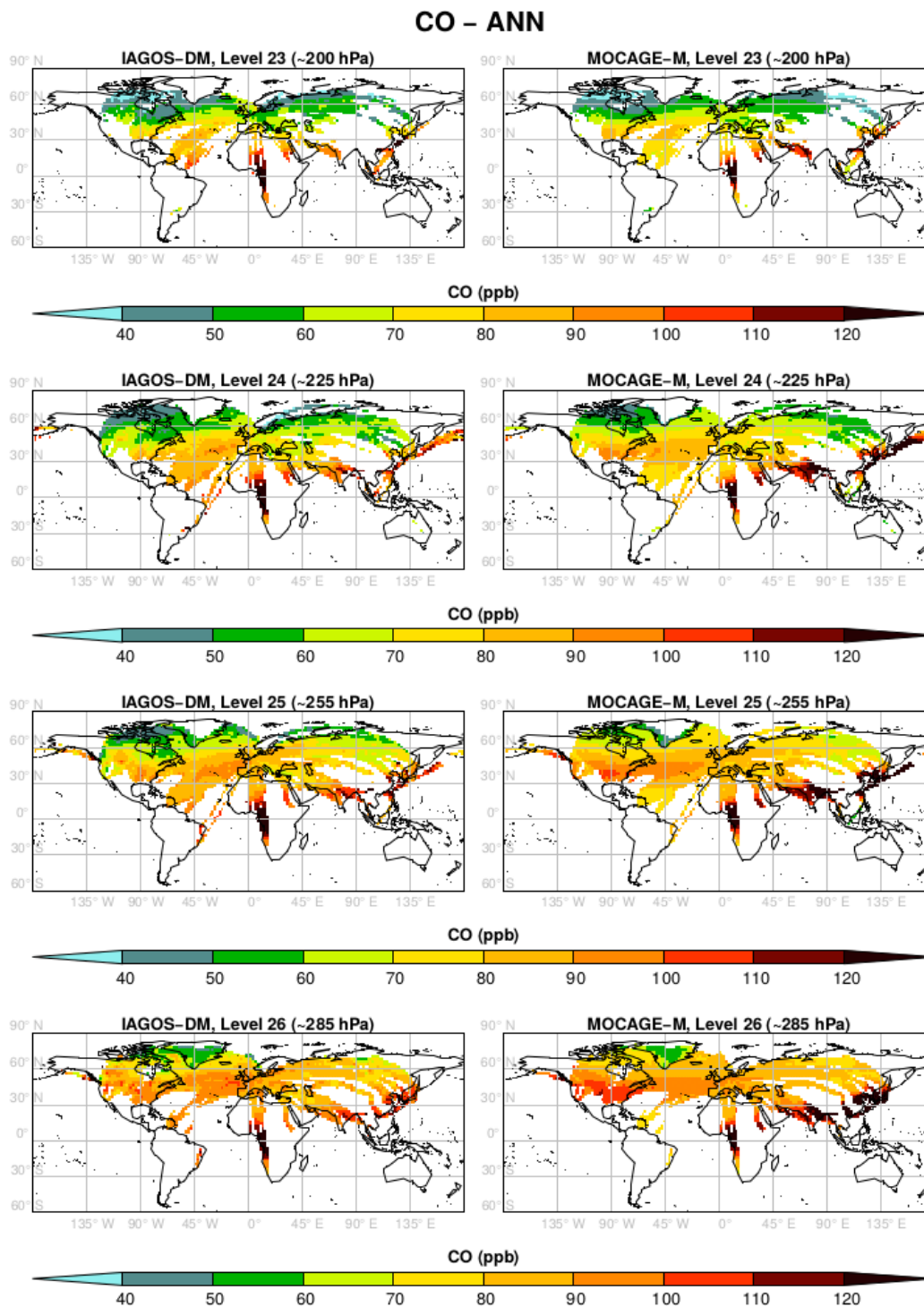
Figure 4. As in Fig. 3 for CO, during the period December 2001–November 2013.
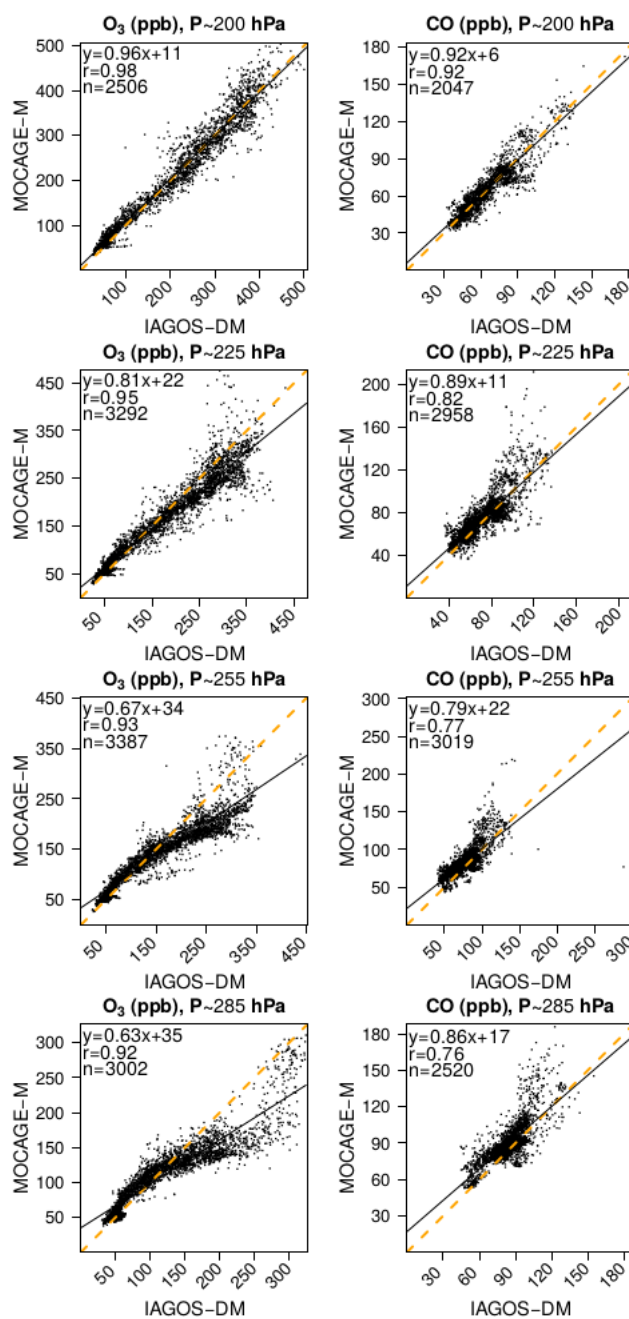
Figure 5. Scatterplots comparing the yearly mixing ratios from MOCAGE-M and IAGOS-DM, for O_3 (left panels) and CO (right panels), at each vertical grid level. The linear regression fit is represented by the solid black line, and the orange dashed line represents the $y = x$ reference line. The regression coefficients and the amount of data (n) are written in the top-left corner of each panel.
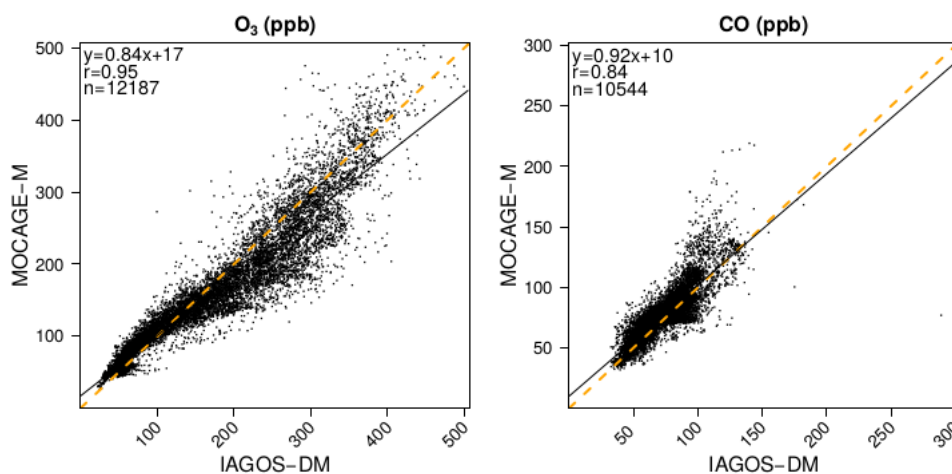
Figure 6. Same as Fig. 5, gathering the points from the four vertical levels.

Table 1. Seasonal and annual metrics synthesizing the assessment of the simulated O_3 climatology by IAGOS-DM, gathering all the vertical grid levels as in Fig. 6. From left to right: Pearson's correlation coefficient (r), modified normalized mean bias (MNMB), fractional gross error (FGE) and the sample size (N_{cells}).

Season	r	MNMB	FGE	N_{cells}
DJF	0.94	0.154	0.202	11,745
MAM	0.95	-0.030	0.168	11,608
JJA	0.92	-0.176	0.284	11,662
SON	0.89	0.028	0.188	12,115
ANN	0.95	-0.010	0.154	12,187

Table 2. Same as Table 1 for CO.

Season	r	MNMB	FGE	N_{cells}
DJF	0.81	0.106	0.164	9,726
MAM	0.84	0.112	0.164	9,622
JJA	0.72	-0.022	0.134	10,372
SON	0.76	0.024	0.122	10,426
ANN	0.84	0.052	0.108	10,544

The yearly MNMB equals -0.010 for O_3 and 0.052 for CO, demonstrating a very good estimation of these two species in the UTLS on a hemispheric scale, especially for O_3. More precisely, it shows a balance between positive and negative normalized biases. The yearly fractional gross error (FGE), corresponding to the averaged normalized bias absolute value is also low, with 0.154 and 0.108 for O_3 and CO respectively. The seasonal patterns show that metrics linked with CO biases (MNMB and

5 FGE) generally yield values closer to 0, compared to O_3. The O_3 seasonal behaviour is characterized by a balance between opposite seasons: the most positive (resp. negative) bias takes place in winter (resp. summer) and equals 0.154 (resp. -0.176), whereas the less negative (resp. positive) bias takes place in spring (resp. fall) and equals -0.030 (resp. 0.028). CO mixing ratio is slightly overestimated in winter and spring similarly (MNMB $\sim$ 0.11), with lower biases during summer (-0.022) and fall (0.024). Nevertheless, all MNMB and FGE are very low, showing good skills from the MOCAGE REF-C1SD simulation.

10 **4.2 Regional-scale analysis**

In this section, we attempt to evaluate the simulation in the UT and the LS separately, focusing on the seasonal cycles. For this, we sort both data sets between the two layers as explained in Sect. 3.4. As a first step, before comparing the simulation to the observations, we analyse the impact of the mapping method for IAGOS onto MOCAGE grid on a monthly basis. For this purpose, two versions of the IAGOS data set are used. Hereafter, IAGOS-HR refers to the high-resolved IAGOS data

15 synthesized in Cohen et al. (2018), where every single measurement was categorized as belonging to the UT (P_{TP} +15 hPa < P < P_{TP} +75 hPa), the transition layer or the LS (P < P_{TP} -15 hPa), and where regional monthly means were derived by averaging every value measured above the defined region. In contrast, IAGOS-DM refers to the new product presented in this paper, i.e. the IAGOS data distributed on the model's grid, directly comparable to the simulation. Note that IAGOS-HR seasonal cycles were computed on the original regions' coordinates, but the changes induced by some $1°$ difference are expected to be

20 negligible, based on the geographical sensitivity tests mentioned in Cohen et al. (2018).

The comparison between the two IAGOS products in matter of seasonal cycles is proposed in Figs. 7 and 8, respectively for O_3 and CO. They are shown with their corresponding interannual variability (IAV), defined as a year-to-year standard deviation. In Fig. 7, both IAGOS versions show a summertime O_3 maximum in the UT and a springtime maximum in the LS. A lessened contrast between the UT and the LS is observed in IAGOS-DM. In the UT, the O_3 volume mixing ratio and its

25 interannual variability are higher in IAGOS-DM than in IAGOS-HR for winter and fall seasons ($\sim 60 \pm 20$ ppb compared to $\sim 50 \pm 10$ ppb), whereas they are similar in spring and summer. In this layer, the most important differences between the two versions thus take place during lower-ozone seasons. In the LS, the O_3 amounts are lower in IAGOS-DM ($\sim$ 125–400 ppb) than in IAGOS-HR ($\sim$ 150–450 ppb) during the whole year. There are two main reasons that explain the lower O_3 amounts
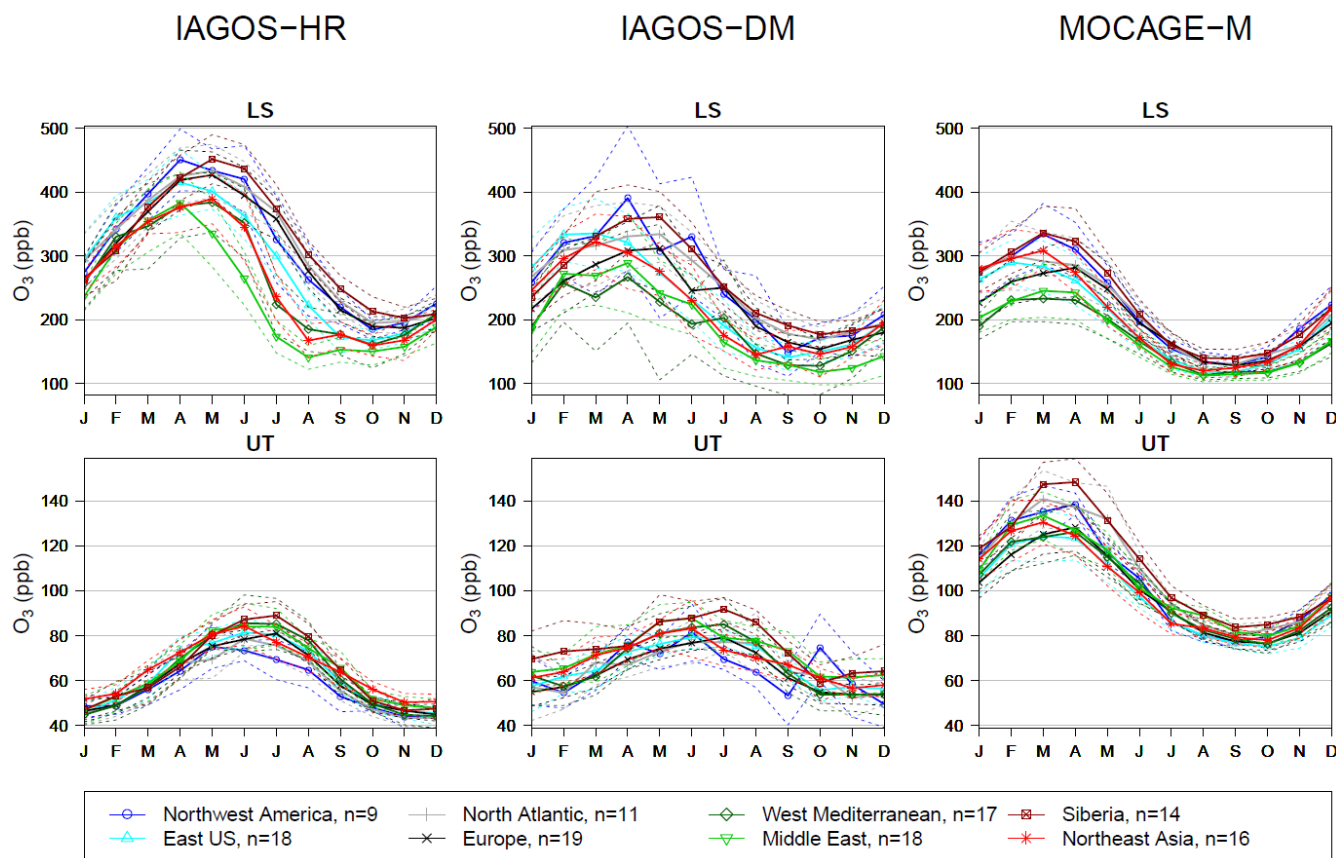
Figure 7. Mean seasonal cycles in ozone above the eight regions from 1995 until 2013, in the UT for the bottom panels, and in the LS for the top panels. Solid and dashed lines correspond respectively to mean values and their interannual variabilities, respectively. The left column corresponds to the high-resolution IAGOS data set (IAGOS-HR) presented in Cohen et al. (2018), and the column at the middle to IAGOS-DM. The right column represents the cycles derived from the simulation (MOCAGE-M), using the same grid points as in IAGOS-DM. The legend is shown at the bottom. For each region, the n integer indicates the amount of selected years contributing to the IAGOS-DM mean seasonal cycles in the UT.

in the LS and the higher amounts in the UT in IAGOS-DM compared to IAGOS-HR. First, the projection of IAGOS observations with a very fine vertical resolution onto MOCAGE vertical grid with a ~800 m vertical resolution. Second, the use of a monthly PV cannot provide the description of the day-to-day variations of the tropopause altitude, whereas the latter can be important to sort the data points between the two layers. In other words, by using the monthly mean PV from the simulation,

5   some of the IAGOS measurement points may be attributed to the LS while being in the UT (or in the tropopause layer) and vice-versa. Figure 7 also makes possible to compare the behaviour of each region. In the LS, the differences between northern and southern regions shown in IAGOS-HR are generally also visible in IAGOS-DM. The regional behaviours discussed in Cohen et al. (2018), i.e. the low summertime O_3 mixing ratio in the Northwest American UT and in the Middle East LS, remain

visible in IAGOS-DM, although the last one is substantially less pronounced. We also note the large O_3 value in October, in the Northwest American UT seen by IAGOS-DM only. We investigated the time series and found it was linked to a strong positive anomaly in October 2003 due to an upper-layer air mass that could not be differentiated to the UT, and weakly balanced by the average with few other years.

5   In Fig. 8, the CO seasonal cycles in the UT are consistent between IAGOS-HR and IAGOS-DM, with a generally low difference, a common springtime maximum, and a consistent inter-regional variability: a higher CO level in the two regions on the Pacific coast (Northwest America and Northeast Asia), higher summertime amounts in Siberia and Northeast Asia, and lower CO levels in the two southernmost regions (the West Mediterranean basin and Middle East). Note that the PV monthly resolution leads to a lessened sampling in the UT in IAGOS-DM. In the North Atlantic region where aircraft trajectories describe a

10 narrow altitude range, the resulting seasonal cycle was incomplete so that we chose to exclude it from the figure.

In the LS, the CO mixing ratio is always higher in IAGOS-DM (~50 ppb–95 ppb) than in IAGOS-HR (from ~40 ppb up to 65 ppb). In IAGOS-HR, a seasonal cycle is noticeable only in Middle East and Northeast Asia, whereas it is the case for almost every region in IAGOS-DM. The influence of the troposphere is increased in IAGOS-DM, with a high peak in May for the West Mediterranean basin, in June–July for Northeast Asia and in July for Siberia, likely related to the effects of boreal

15 biomass burning in the latter. Thus, mapping the observations on the model grid changes significantly the CO seasonal cycles in the LS.

As for O_3, the reason why the CO amounts in IAGOS-DM are higher in the LS and lower in the UT comes from the coarse vertical resolution in MOCAGE grid, and from the uncertainty when sorting the UT data from the LS data using a monthly mean modelled PV field. More generally, the comparison between IAGOS-HR and IAGOS-DM for O_3 and CO clearly shows

20 that the processing applied for mapping IAGOS high resolution data set onto MOCAGE coarse grid slightly modifies IAGOS characteristics. This processing, which enables a meaningful comparison between IAGOS long-term measurements and the REF-C1SD simulation, acts as a numerical filter. It is important to note that the seasonal cycles in IAGOS-DM generally show values comprised between the MOCAGE-M and the IAGOS-HR cycles. It confirms that the representation derived from IAGOS-HR cannot be reached by a model with the typical REF-C1SD resolution, especially for CO in the LS, although some

25 main characteristics mentioned above can still be used as criteria.

We now assess the MOCAGE-M seasonal cycles by comparing them to IAGOS-DM. As complements to Fig. 7, statistical results are given in Table 3. Note that inter-regional averages have been computed only to synthesize the assessment and to provide quantifications that confirm some features seen in the figures. As they are similar with zonal averages, they are not meant to have a geophysical signification. A qualitative summary is also provided in Table 4. In the UT, MOCAGE-M shows a

Table 3. Inter-regional averages derived from the seasonal cycles. Yearly mean values and interannual variabilities are shown for IAGOS-DM and MOCAGE-M, along with the correlation coefficient between these two data sets.

Species	Layer	IAGOS-DM	MOCAGE-M	$\bar{r}$
O_3 (ppb)	LS	229 ± 44	199 ± 25	0.87
	UT	68 ± 8	103 ± 6	0.17
CO (ppb)	LS	66 ± 10	73 ± 6	0.22
	UT	102 ± 11	102 ± 7	0.67

Table 4. Synthesis of MOCAGE-M ability to reproduce the main features from IAGOS-DM.

Species	Layer	Main features from IAGOS-DM	Reproduced by MOCAGE-M
O_3	LS	Springtime maxima	Yes
		Northward gradient	Yes
	UT	Summertime maxima	No
		Less O_3 in WNAm summer	No
		More O_3 in Sib	Yes
CO	LS	Summertime maxima	No
		More CO in WMed	No
		More CO in NEAs summer	Yes
	UT	Springtime maxima	Yes (winter-spring)
		More CO in WNAm and NEAs	Yes
		Spring-summer maximum in NEAs	No

springtime maximum and higher O_3 concentrations (from ~120 ppb up to 150 ppb), instead of the observed summertime maximum (season which O_3 values range between ~80 ppb and 115 ppb). Adding the fact that simulated O_3 levels are particularly strong in the northernmost regions (Western North America and Siberia) where the stratosphere is richer in O_3, it is likely that the stratospheric influence on the UT is overestimated in the simulation. The inter-regional averages shown in Table 3 con-

5 firms the significant difference between the two data sets in the UT, both from O_3 mixing ratio (103 ± 6 ppb in MOCAGE-M compared to 68 ± 8 ppb in IAGOS-DM) and from the seasonality ($\bar{r} = 0.17$). In the LS, the simulation reproduces well the cycles including the seasonality ($\bar{r} = 0.87$ as shown in Table 3), the magnitude, the amounts of ozone (199 ± 25 ppb compared to 229 ± 44 ppb from IAGOS-DM) and the inter-regional differences. The latter are characterized in both data sets by lower ozone levels in the two southernmost regions (West Mediterranean basin and Middle East) and higher ozone levels in the two

10 northernmost regions (Western North America and Siberia). Without the noisy signal characterising Western North America and the West Mediterranean basin in IAGOS-DM, the springtime interannual variabilities spread from ~200 ppb up to ~400 ppb in both data sets, showing another point well reproduced by the model. Though on a yearly basis, according to Table 3, the model tends to underestimate ozone IAV on average by a factor 1.8.
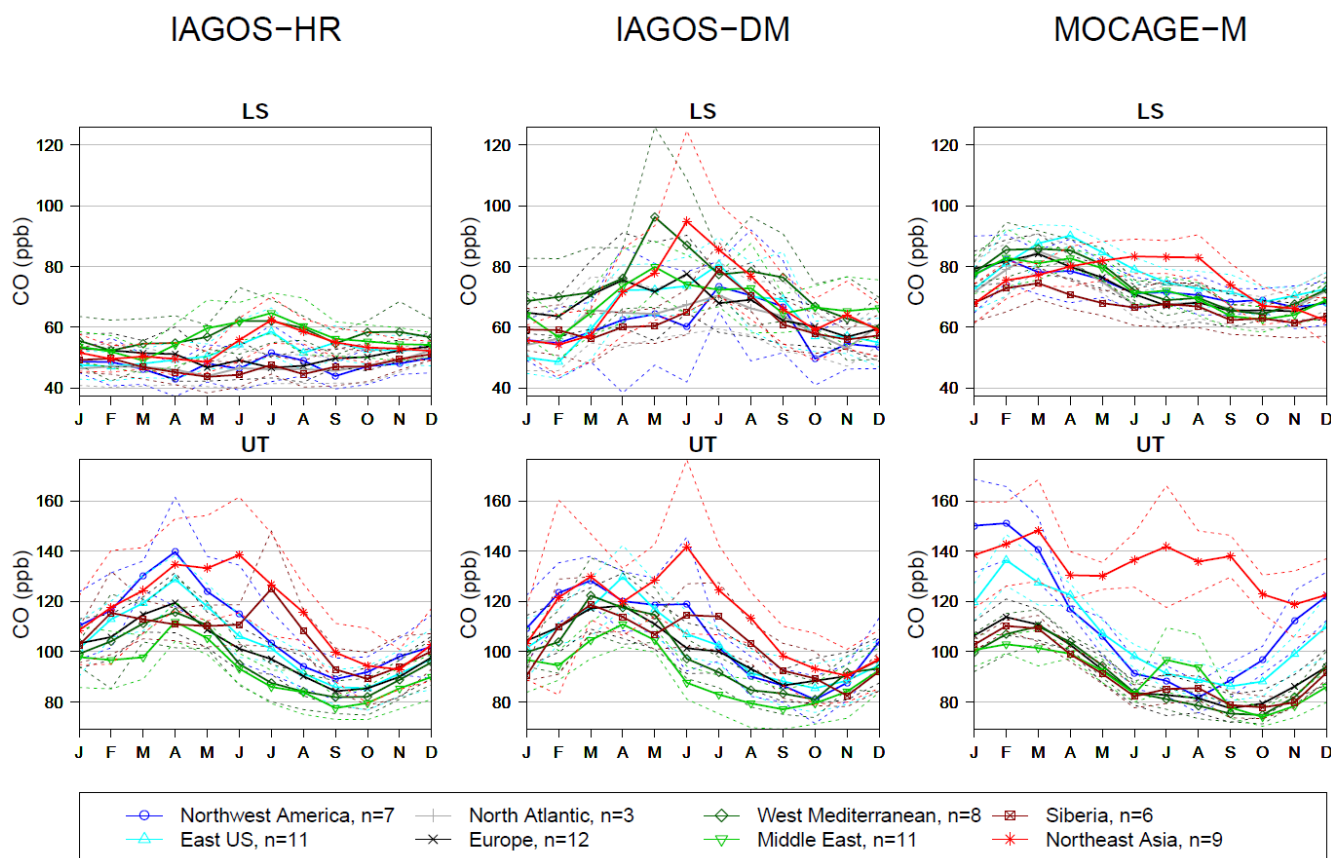
Figure 8. Same as Fig. 7 for carbon monoxide, from 2002 until 2013.

The modelled CO seasonal cycles (Fig. 8) in the UT show similarities with the observations (IAGOS-DM), including the higher concentrations in the two Pacific coast regions (Western North America and Northeast Asia), the strong summertime concentrations in Northeast Asia and also comparable mixing ratios between the model and the IAGOS-DM observations in most regions, as confirmed by Table 3. However, the simulation overestimates the CO mixing ratios in the two Pacific coast regions, and the seasonal maxima generally take place during late winter–early spring in the simulation, earlier than the observed middle-of-spring maxima. The seasonal minima are in phase with the observations. In the LS, the seasonal cycles magnitude is underestimated by the simulation but the overall bias remains relatively low, with a 73 ± 6 ppb average for MOCAGE-M compared to 66 ± 10 ppb for IAGOS-DM. In most regions, MOCAGE-M shows seasonal cycles in the LS in phase with the UT, thus contrasting with the observations and making the correlation drop from 0.67 in the UT to 0.22 in the LS. This suggests that the model simulation is affected in the LS by transport from the troposphere during springtime. In summer, the July maximum in Siberia is not represented in the simulation, showing an underestimation in the biomass burning inventory

or in convection in this region. Meanwhile, consistently with observations, MOCAGE-M shows a summertime maximum in Northeast Asia exclusively. Although part of this feature may originate from the positive bias in the UT, the fact that it only concerns the summer season, in contrast to the UT, suggests that summertime convection also plays a non-negligible role.

5 Summary and conclusions

We developed a methodology that makes the IAGOS database ready to assess chemistry-climate long-term model simulations
5 for recent decades, and particularly the REF-C1SD experiment produced in the frame of CCMI phase-I project. The current paper describes this methodology and its application on a chosen simulation (REF-C1SD simulation from MOCAGE-CTM), assessing modelled ozone and carbon monoxide monthly fields respectively during Aug. 1994–Dec. 2013 and Dec. 2001–Dec. 2013.

The first step consists in generating a gridded monthly IAGOS data set (IAGOS-DM), firstly by a linear-distanced reverse
10 interpolation on the chosen model grid on a monthly basis, and then by deriving weighted monthly means on each grid cell. The second step consists in deriving seasonal and annual climatologies for the well-sampled vertical grid levels, then to derive statistical scores for the simulation assessment. In the case of REF-C1SD simulation from MOCAGE, the yearly mean spatial distribution is well reproduced by the model, especially for O_3, and especially at the sampled highest levels too. It suggests a particularly good representation of the main stratospheric processes that affect O_3 in the UTLS. The extreme mean CO
15 mixing ratios observed above the strongly emitting and convective regions in the tropics and subtropics are also visible in the simulation, with a very low bias above tropical Africa and a significant positive bias above South and East Asia. Globally, the annual O_3 normalized mean bias is very low (MNMB = -0.010) and slightly higher seasonally. They are a bit higher in winter and summer (| MNMB | = 0.154–0.176) than in spring and fall (| MNMB | = 0.028–0.030), with quasi opposite values in each couple of antagonist seasons. The yearly bias in CO is positive (MNMB = 0.052), with highest values similarly in winter and
20 spring, and particularly low values in fall. The statistical metrics were applied for each vertical grid level separately in order to locate strengths and weaknesses of the model, but also for the whole UTLS grid cells for the purpose of a bulk comparison that could be reiterated on other model simulations.

Another step consists in a comparison of the seasonal cycles between IAGOS and MOCAGE simulation in the upper tropo-sphere (UT) and the lower stratosphere (LS). It lies on the use of a monthly mean calculated PV field to define a UT and a LS
25 separated by a transition layer, following the same principle as in Thouret et al. (2006). The mean seasonal cycles have been compared over the eight well-sampled regions defined and analysed in Cohen et al. (2018). The application to the assessment of this REF-C1SD experiment by MOCAGE is preceded by an analysis of the changes induced in IAGOS seasonal cycles

by the projection on the model monthly grid. As expected, going from IAGOS-HR to IAGOS-DM systematically leads to an increase (resp. decrease) in upper-tropospheric (resp. lower-stratospheric) O_3, to an increase in lower-stratospheric CO and generally to a slight decrease in upper-tropospheric CO. The use of a monthly mean PV field and the $\sim$ 800 m vertical resolution in the UTLS of MOCAGE onto which IAGOS observations are projected automatically result in an artificial increase of stratosphere–troposphere exchange, characterized as a decreased vertical gradient between UT and LS. Nevertheless, the

5   seasonal maxima and minima become less clear but remain visible in IAGOS-DM with respect to IAGOS-HR. The hierarchy between the regions is generally conserved from IAGOS-HR to IAGOS-DM, for both chemical species and both layers: in each of these cases, we find the same regions showing lowest/highest values between the two IAGOS representations. Also, some exclusive local behaviours mentioned in Cohen et al. (2018) remain visible in IAGOS-DM. Concerning O_3, we highlighted the consistency of the lowest quantities in the UT above Western North America and, substantially less significant, in the LS

10   above Middle East. Concerning CO, we showed the conservation of the spring-summer maximum in Northeast Asia in the UT and its summertime maximum in the LS.

     The evaluation of the MOCAGE REF-C1SD simulation (MOCAGE-M) with IAGOS-DM shows a good representation of O_3 in the LS in matter of seasonal cycle magnitudes and geographical variability, thus highlighting the well-reproduced main stratospheric processes. In the UT, for the eight regions, the model overestimates the O_3 mixing ratios and shows a typical

15   lower-stratospheric seasonality, suggesting an overestimation in the transport from the stratosphere. The modelled CO field shows similarities with the observations in the UT, with a one-month shift in the seasonal maxima. One possible reason is the decadal linear interpolation in anthropogenic emissions implemented in REF-C1SD, leading to a lack of year-to-year variability in modelled CO fields. In the LS, CO is generally higher in the simulation and shows a seasonal cycle in phase with the UT, in contrast to IAGOS-DM. It suggests an overestimated tropospheric influence in this layer during springtime.

20   The methodology shown in this paper has proven useful for assessing REF-C1SD experiment from MOCAGE in the UTLS, further highlighting the model strengths and weaknesses when compared to the densest *in situ* IAGOS data set in the UTLS. Particularly, the use of the IAGOS-DM product instead of IAGOS-HR systematically reduced the biases characterizing the simulation, thus avoiding an underestimation of the model abitilies to reproduce the chemical composition of the UT and the LS in a recent climate time period.

25   The present methodology could easily be applied to CCMI REF-C1SD simulations from other models, both for an inter-model comparison and for assessing CCMI products against IAGOS database, notably intermodel-averaged fields. To a greater extent, it can be used on a wide range of long-term simulations including CCMs free runs in order to perform climatological comparisons. Furthermore, the assessment illustrated in this study is based on two chosen applications of our methodology, i.e.

the analyses of long-term seasonal and yearly averages on different vertical grid levels and the mean seasonal cycles in the UT and the LS, but a wide diversity of complementary comparisons remains possible. We thus recommend this new product to the CCMI community.

Appendix A: Horizontal climatologies

A1 Ozone
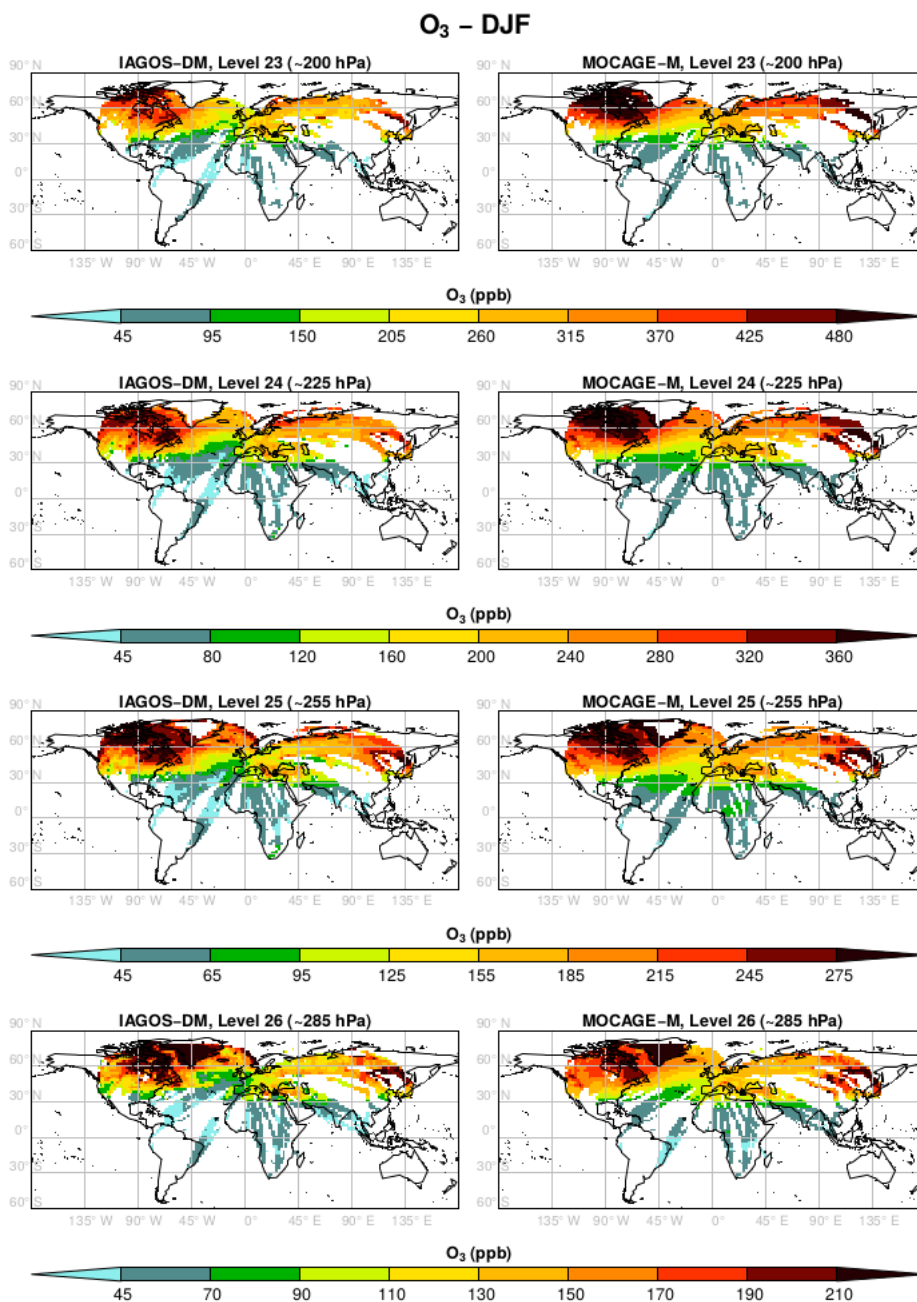
Figure A1. As Fig. 3 for boreal winter.
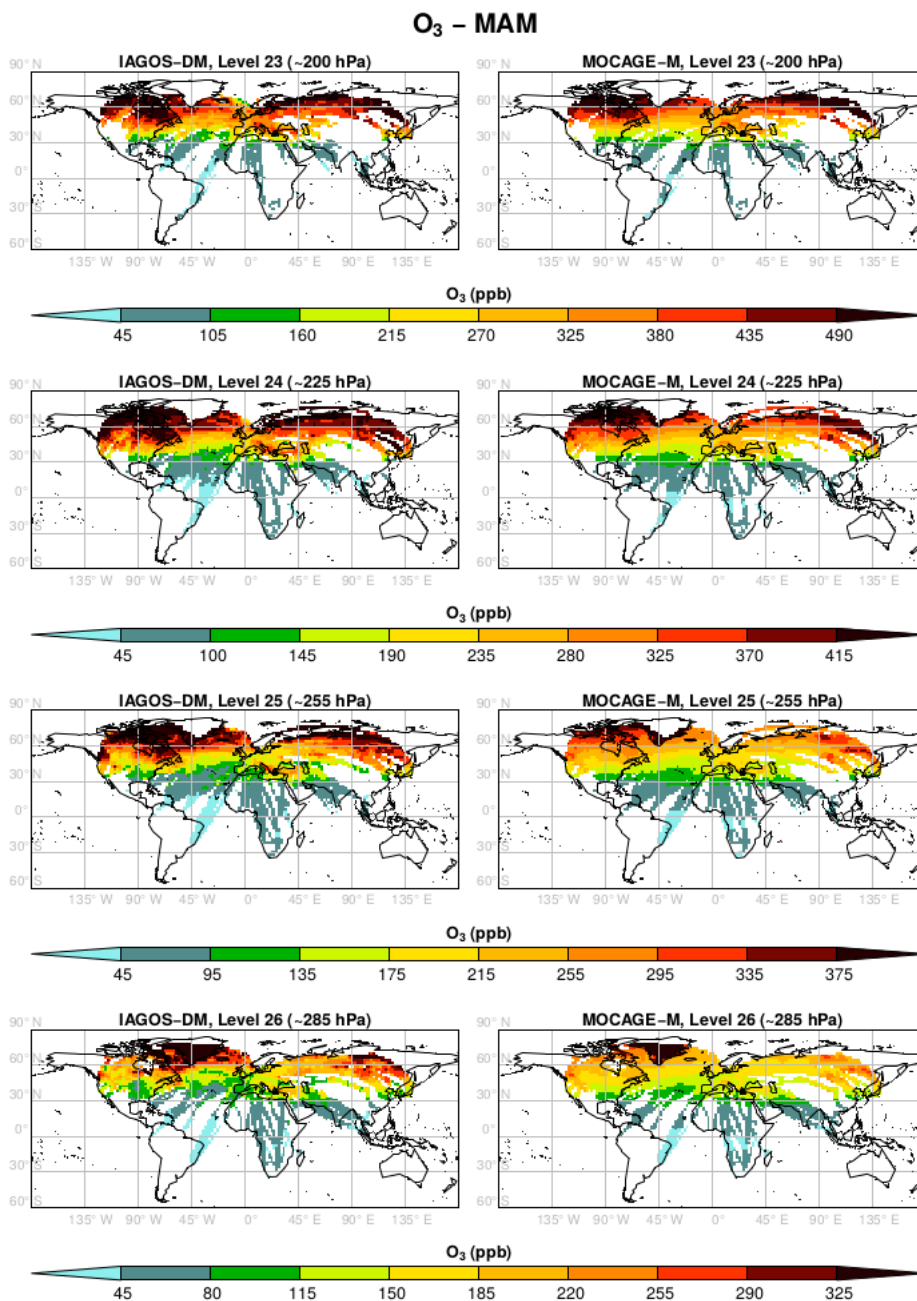
Figure A2. As Fig. 3 for boreal spring.
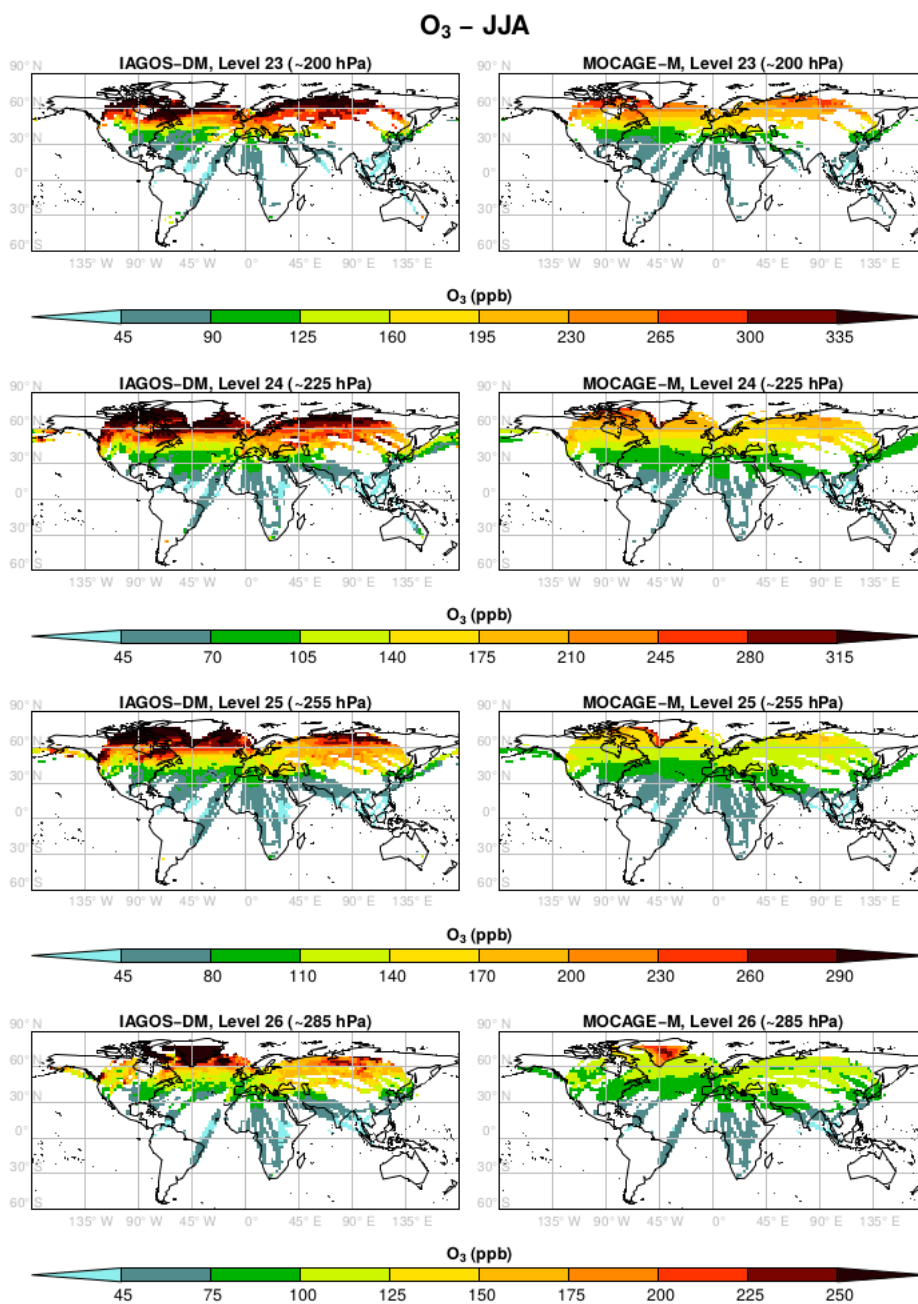
Figure A3. As Fig. 3 for boreal summer.
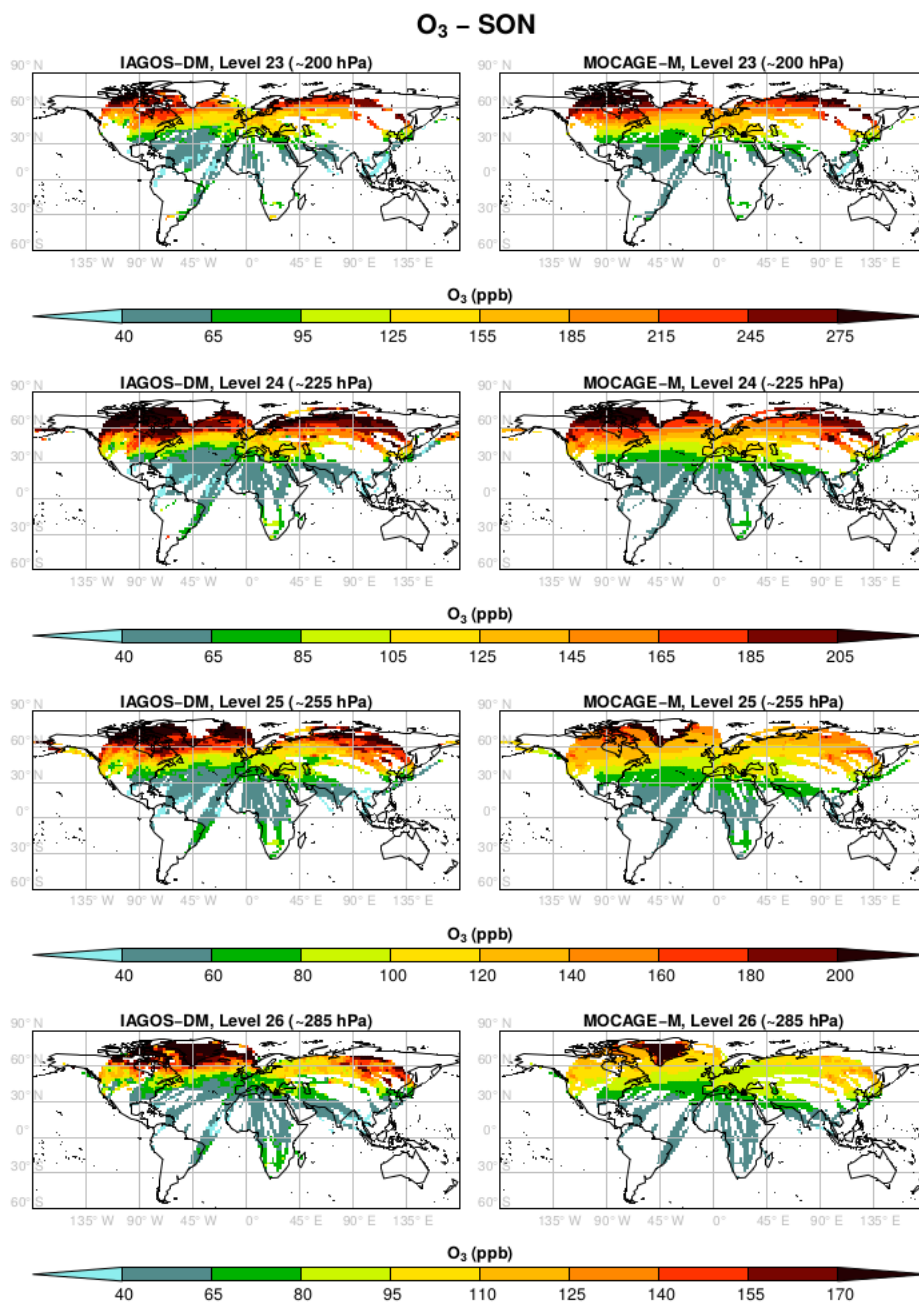
Figure A4. As Fig. 3 for boreal fall.

5 A2 Carbon monoxide
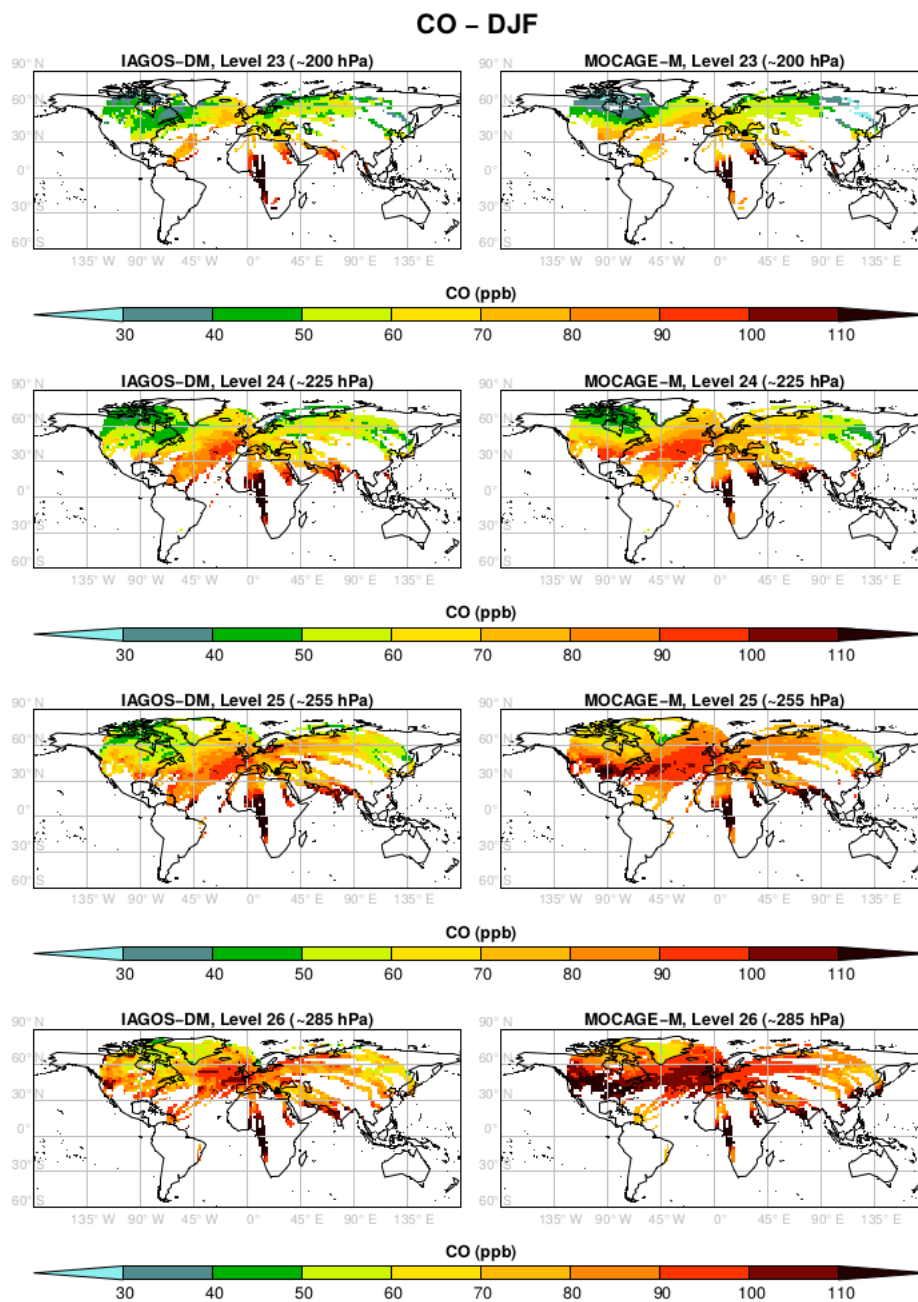
Figure A5. As Fig. 4 for boreal winter.
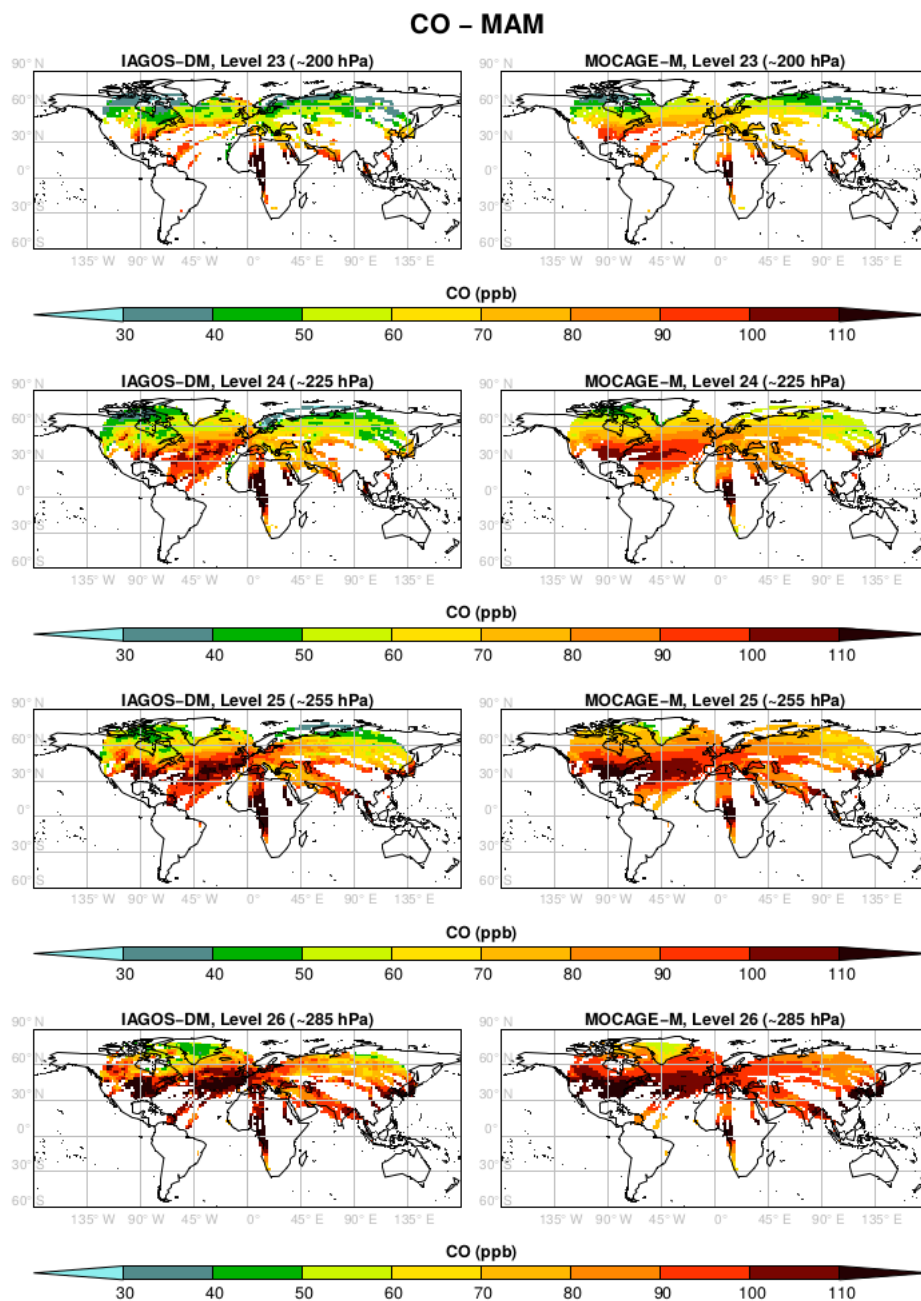
Figure A6. As Fig. 4 for boreal spring.
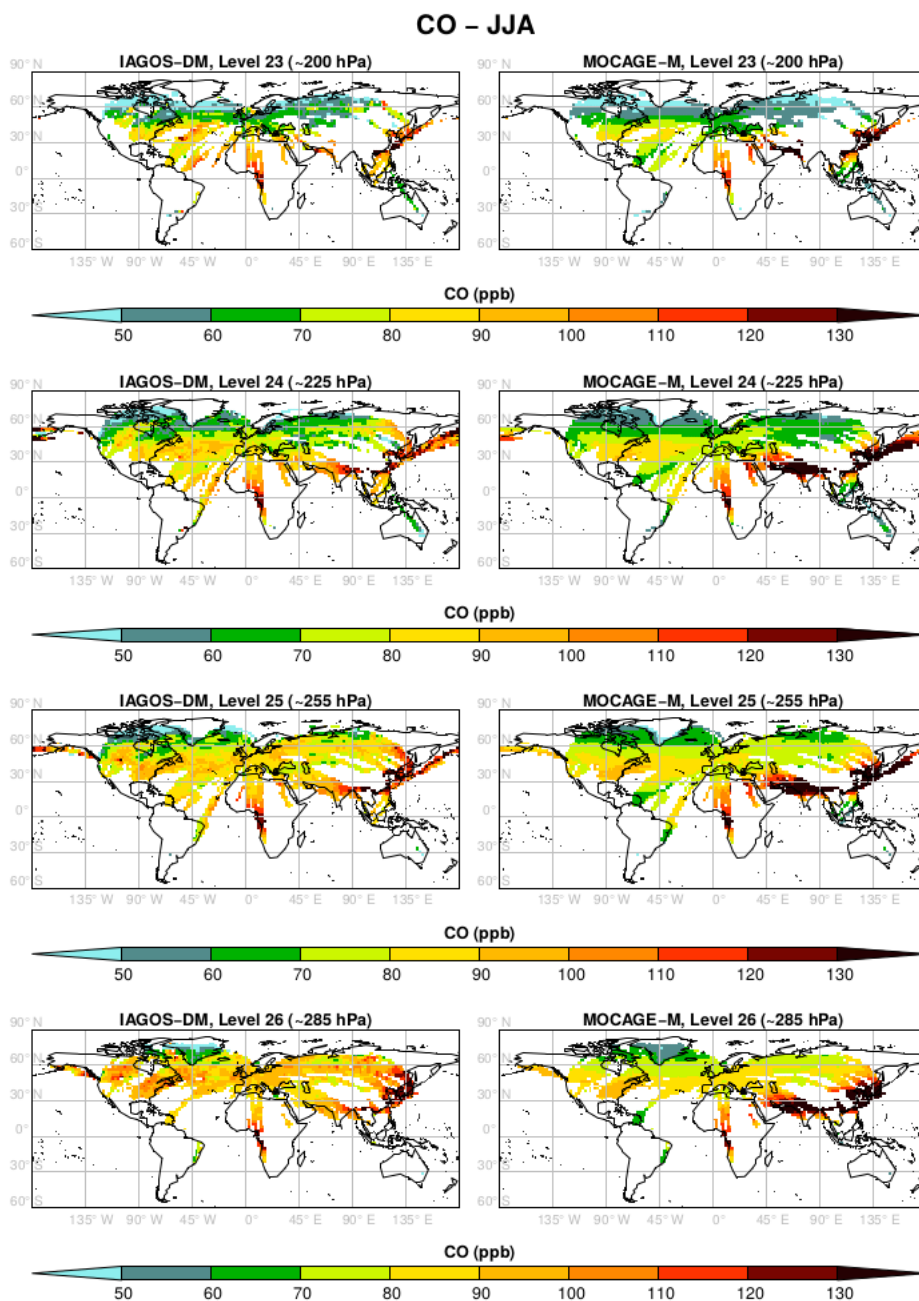
Figure A7. As Fig. 4 for boreal summer.
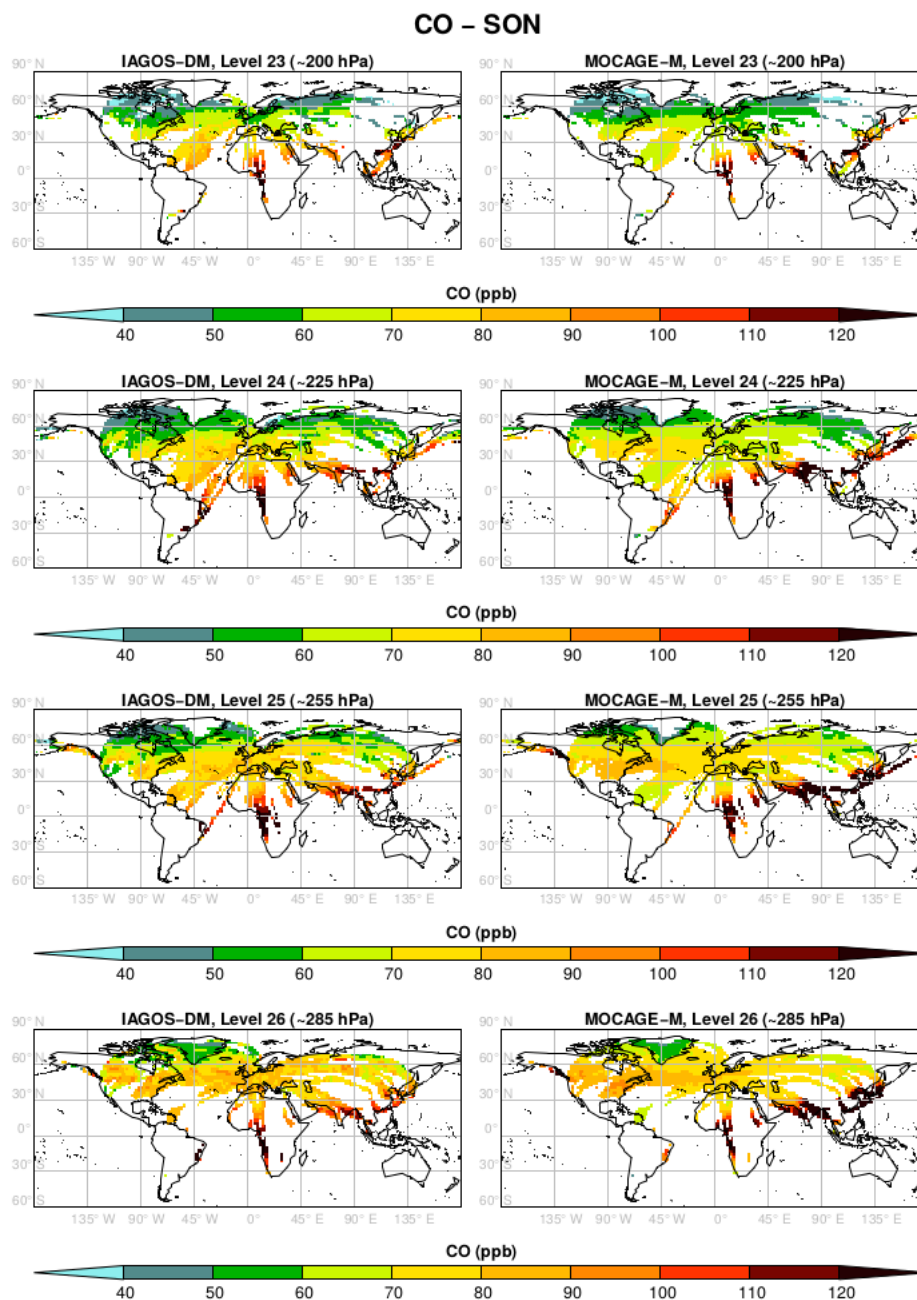
Figure A8. As Fig. 4 for boreal fall.

Appendix B: Scatterplots
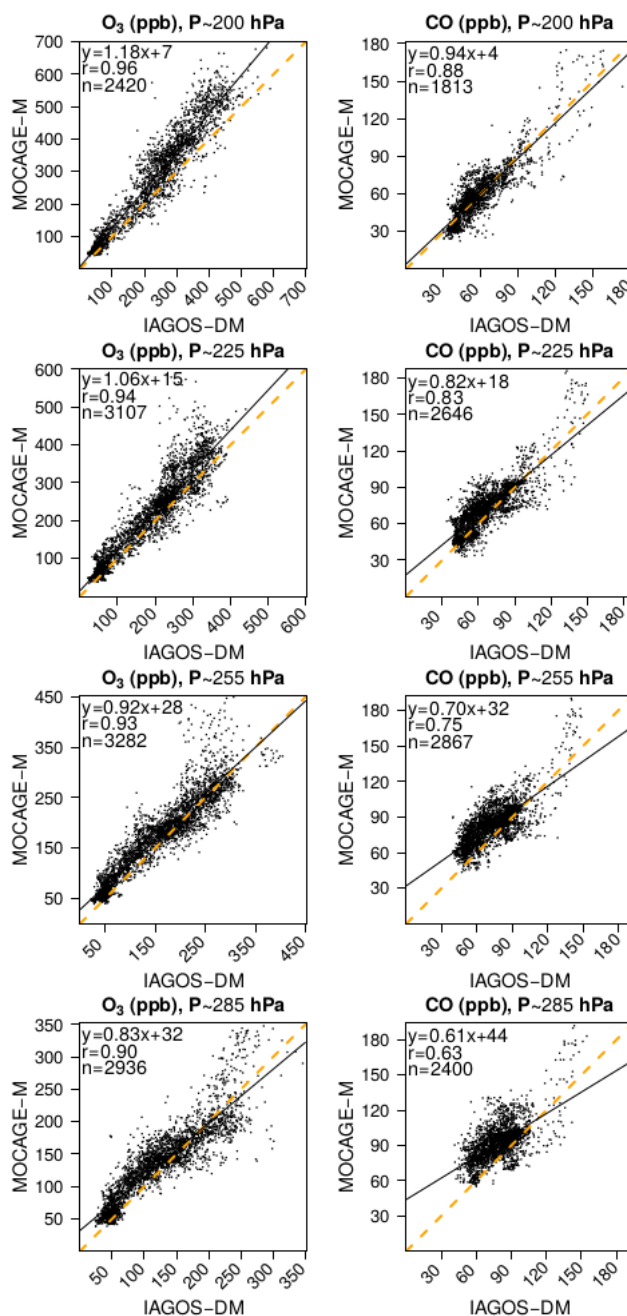
Figure B1. As Fig. 5 for boreal winter.
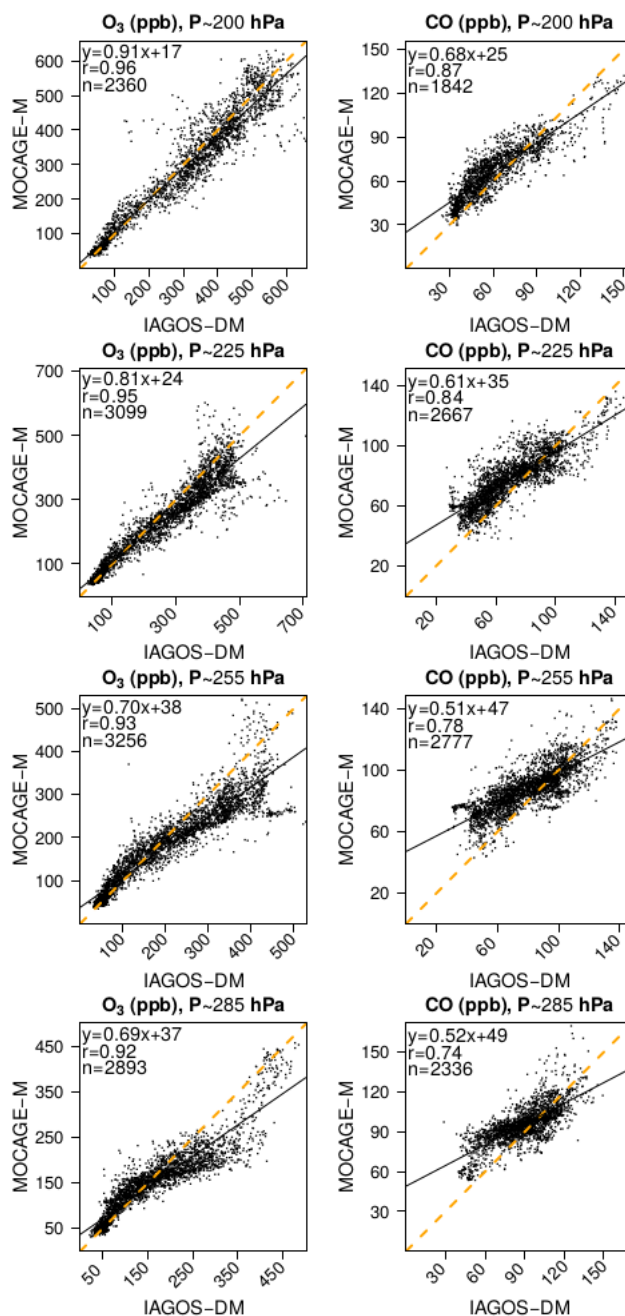
Figure B2. As Fig. 5 for boreal spring.
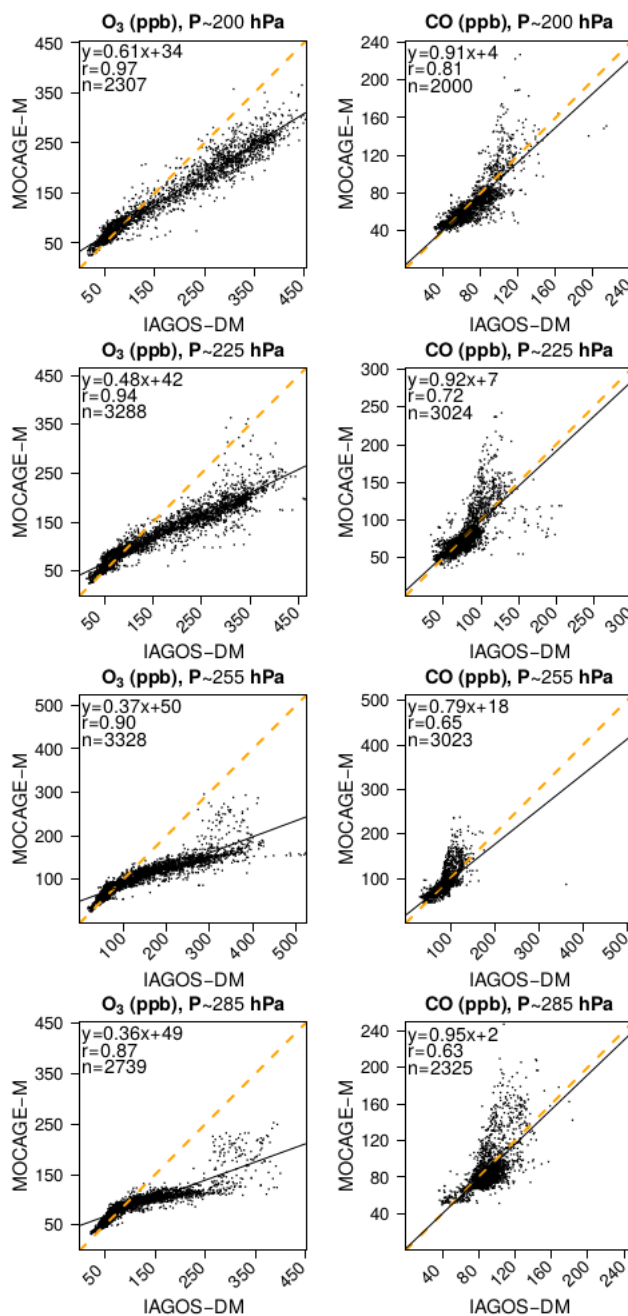
Figure B3. As Fig. 5 for boreal summer.
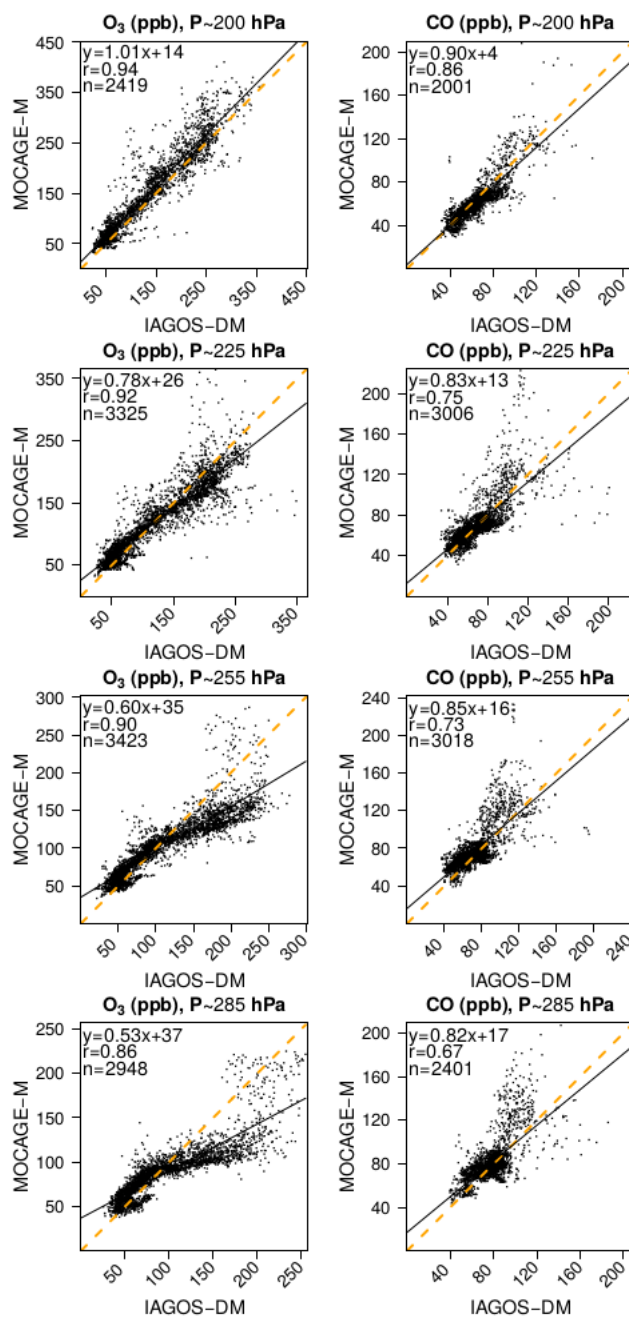
Figure B4. As Fig. 5 for boreal fall.

Code and data availability. The IAGOS ozone and carbon monoxide data are available at https://doi.org/10.25326/20 and more precisely, the time series data are found at https://doi.org/10.25326/06. The IAGOS-DM product adapted to the MOCAGE model grid (named IAGOS-DM-MOCAGE) is available at https://doi.org/10.25326/80 and the software (named Interpol-IAGOS) at https://doi.org/10.25326/81.

Author contributions. YC, VM, BJ and VT designed the study. The Interpol-IAGOS software was developped by YC. The simulation output was provided by BJ and VM, and the IAGOS data by VT. The paper was written by YC and reviewed by VM and VT, and commented,

5   edited and approved by all the authors.

Acknowledgements. The authors acknowledge the strong support of the European Commission, Airbus, and the Airlines (Lufthansa, Air-France, Austrian, Air Namibia, Cathay Pacific, Iberia and China Airlines so far) who carry the MOZAIC or IAGOS equipment and perform the maintenance since 1994. In its last 10 years of operation, MOZAIC has been funded by INSU-CNRS (France), Météo-France, Université Paul Sabatier (Toulouse, France) and Research Center Jülich (FZJ, Jülich, Germany). IAGOS has been additionally funded by the EU projects

10 IAGOS-DS and IAGOS-ERI. We also wish to acknowledge our colleagues from the IAGOS team in FZJ, Jülich for useful discussions. The MOZAIC–IAGOS database is supported by AERIS (CNES and INSU-CNRS). Data are also available via AERIS web site www.aeris-data.fr. Yann Cohen acknowledges the University of Toulouse for providing adminstrative support for his PhD.

Financial support. This research has been supported by the Occitanie region and Météo-France.

Competing interests. The authors declare that they have no conflict of interest.

15 References

Bourassa, M. A., Romero, R., Smith, S. R., and O'Brien, J. J.: A new FSU winds climatology, J. Clim., 18, 3686–3698, doi:10.1175/JCLI3487.1, 2005.

Bowman, K. W., Shindell, D. T., Worden, H. M., Lamarque, J.-F., Young, P. J., Stevenson, D. S., Qu, Z., de la Torre, M., Bergmann, D., Cameron-Smith, P. J., Collins, W. J., Doherty, R., Dalsøren, S. B., Faluvegi, G., Folberth, G., Horowitz, L. W., Josse, B. M., Lee, Y. H., MacKenzie, I. A., Myhre, G., Nagashima, T., Naik, V., Plummer, D. A., Rumbold, S. T., Skeie, R. B., Strode, S. A., Sudo, K., Szopa, S., Voulgarakis, A., Zeng, G., Kulawik, S. S., Aghedo, A. M., and Worden, J. R.: Evaluation of ACCMIP outgoing longwave radiation from tropospheric ozone using TES satellite observations, Atmos. Chem. Phys., 13, 4057–4072, doi:10.5194/acp-13-4057-2013, 2013.

Cohen, Y., Petetin, H., Thouret, V., Marécal, V., Josse, B., Clark, H., Sauvage, B., Fontaine, A., Athier, G., Blot, R., Boulanger, D., Cousin, J.-M., and Nédélec, P.: Climatology and long-term evolution of ozone and carbon monoxide in the upper troposphere–lower stratosphere (UTLS) at northern midlatitudes, as seen by IAGOS from 1995 to 2013, Atmos. Chem. Phys., 18, 5415–5453, doi:10.5194/acp-18-5415-2018, 2018.

Eyring, V., Lamarque, J.-F., Hess, P., Arfeuille, F., Bowman, K., Chipperfield, M. P., Duncan, B., Fiore, A., Gettelman, A., Giorgetta, M. A., Granier, C., Hegglin, M., Kinnison, D., Kunze, M., Langematz, U., Luo, B., Martin, R., Matthes, K., Newman, P. A., Peter, T., Robock, A., Ryerson, T., Saiz-Lopez, A., Salawitch, A., Schultz, M., Shepherd, T. G., Shindell, D., Staehelin, J., Tegtmeier, S., Thomason, L., Tilmes, S., Vernier, J.-P., Waugh, D. W., and Young, P. J.: Overview of IGAC/SPARC Chemistry-Climate Model Initiative (CCMI) Community Simulations in Support of Upcoming Ozone and Climate Assessments, SPARC Newsletter, pp. 48–66, 2013.

Froidevaux, L., Kinnison, D. E., Wang, R., Anderson, J., and Fuller, R. A.: Evaluation of CESM1 (WACCM) free-running and specified dynamics atmospheric composition simulations using global multispecies satellite data records, Atmos. Chem. Phys., 19, 4783–4821, doi:10.5194/acp-19-4783-2019, 2019.

Gaudel, A., Clark, H., Thouret, V., Jones, L., Inness, A., Flemming, J., Stein, O., Huijnen, V., Eskes, H., Nédélec, P., and Boulanger, D.: On the use of MOZAIC-IAGOS data to assess the ability of the MACC reanalysis to reproduce the distribution of ozone and CO in the UTLS over Europe, Tellus B, 67, 27 955, doi:10.3402/tellusb.v67.27955, 2015.

Guth, J., Josse, B., Marécal, V., Joly, M., and Hamer, P.: First implementation of secondary inorganic aerosols in the MOCAGE version R2.15.0 chemistry transport model, Geosci. Model Dev., 9, 137–160, doi:10.5194/gmd-9-137-2016, 2016.

Hu, L., Jacob, D. J., Liu, X., Zhang, Y., Zhang, L., Kim, P. S., Sulprizio, M. P., and Yantosca, R. M.: Global budget of tropospheric ozone: Evaluating recent model advances with satellite (OMI), aircraft (IAGOS), and ozonesonde observations, Atmos. Environ., 167, 323–334, doi:10.1016/j.atmosenv.2017.08.036, 2017.

Josse, B., Simon, P., and Peuch, V.-H.: Radon global simulations with the multiscale chemistry and transport model MOCAGE, Tellus B, 56, 339–356, doi:10.3402/tellusb.v56i4.16448, 2004.

30 Lamarque, J.-F., Shindell, D. T., Josse, B., Young, P. J., Cionni, I., Eyring, V., Bergmann, D., Cameron-Smith, P., Collins, W. J., Doherty, R.,
    Dalsoren, S., Faluvegi, G., Folberth, G., Ghan, S. J., Horowitz, L. W., Lee, Y. H., MacKenzie, I. A., Nagashima, T., Naik, V., Plummer,
    D., Righi, M., Rumbold, S. T., Schulz, M., Skeie, R. B., Stevenson, D. S., Strode, S., Sudo, K., Szopa, S., Voulgarakis, A., and Zeng, G.:
    The Atmospheric Chemistry and Climate Model Intercomparison Project (ACCMIP): overview and description of models, simulations
    and climate diagnostics, Geosci. Model Dev., 6, 179–206, doi:10.5194/gmd-6-179-2013, 2013.

5   Lefèvre, F., Brasseur, G. P., Folkins, I., Smith, A. K., and Simon, P.: Chemistry of the 1991–1992 stratospheric winter: Three-dimensional
    model simulations, J. Geophys. Res.-Atmos., 99, 8183–8195, doi:10.1029/93JD03476, 1994.

    Marenco, A., Thouret, V., Nédélec, P., Smit, H., Helten, M., Kley, D., Karsher, F., Simon, P., Law, K., Pyle, J., Poschmann, G., Von Wrede,
    R., Hume, C., and Cook, T.: Measurement of ozone and water vapour by Airbus in-service aircraft : The MOZAIC airborne programme,
    an overview, J. Geophys. Res., 103, 25 631–25 642, doi:10.1029/98JD00977, 1998.

10 Morgenstern, O., Hegglin, M. I., Rozanov, E., O'Connor, F. M., Abraham, N. L., Akiyoshi, H., Archibald, A. T., Bekki, S., Butchart, N.,
    Chipperfield, M. P., Deushi, M., Dhomse, S. S., Garcia, R. R., Hardiman, S. C., Horowitz, L. W., Jöckel, P., Josse, B., Kinnison, D.,
    Lin, M., Mancini, E., Manyin, M. E., Marchand, M., Marécal, V., Michou, M., Oman, L. D., Pitari, G., Plummer, D. A., Revell, L. E.,
    Saint-Martin, D., Schofield, R., Stenke, A., Stone, K., Sudo, K., Tanaka, T. Y., Tilmes, S., Yamashita, Y., Yoshida, K., and Zeng, G.:
    Review of the global models used within phase 1 of the Chemistry–Climate Model Initiative (CCMI), Geosci. Model Dev., 10, 639–671,
15 doi:10.5194/gmd-10-639-2017, 2017.

    Nédélec, P., Cammas, J.-P., Thouret, V., Athier, G., Cousin, J.-M., Legrand, C., Abonnel, C., Lecoeur, F., Cayez, G., and Marizy, C.: 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, doi:10.5194/acp-3-1551-2003, 2003.

    Nédélec, P., Blot, R., Boulanger, D., Athier, G., Cousin, J.-M., Gautron, B., Petzold, A., Volz-Thomas, A., and Thouret, V.: Instrumentation
20 on commercial aircraft for monitoring the atmospheric composition on a global scale: the IAGOS system, technical overview of ozone
    and carbon monoxide measurements, Tellus B, 67, 27 791, doi:10.3402/tellusb.v67.27791, 2015.

    New, M., Hulme, M., and Jones, P.: Representing Twentieth-Century space–time climate variability. Part II: Development of 1901–96
    monthly grids of terrestrial surface climate, J. Clim., 13, 2217–2238, doi:10.1175/1520-0442(2000)013<2217:RTCSTC>2.0.CO;2, 2000.

    Petzold, A., Thouret, V., Gerbig, C., Zahn, A., Brenninkmeijer, C., Gallagher, M., Hermann, M., Pontaud, M., Ziereis, H., Boulanger, D.,
25 Marshall, J., Nédélec, P., Smit, H., Friess, U., Flaud, J.-M., Wahner, A., Cammas, J.-P., and Volz-Thomas, A.: Global-scale atmosphere
    monitoring by in-service aircraft - current achievements and future prospects of the European Research Infrastructure IAGOS, Tellus B,
    67, doi:10.3402/tellusb.v67.28452, 2015.

    Riese, M., Ploeger, F., Rap, A., Vogel, B., Konopka, P., Dameris, M., and Forster, P.: Impact of uncertainties in atmospheric mixing on
    simulated UTLS composition and related radiative effects, J. Geophys. Res., 117, D16 305, doi:10.1029/2012JD017751, 2012.

30 Smalley, K. M., Dessler, A. E., Bekki, S., Deushi, M., Marchand, M., Morgenstern, O., Plummer, D. A., Shibata, K., Yamashita, Y., and
     Zeng, G.: Contribution of different processes to changes in tropical lower-stratospheric water vapor in chemistry–climate models, Atmos.
     Chem. Phys., 17, 8031–8044, doi:10.5194/acp-17-8031-2017, 2017.

     SPARC: CCMVal Report on the Evaluation of Chemistry-Climate Models, Tech. rep., SPARC, 2010.

     Stevenson, D. S., Dentener, F. J., Schultz, M. G., Ellingsen, K., van Noije, T. P. C., Wild, O., Zeng, G., Amann, M., Atherton, C. S., Bell, N.,
5   Bergmann, D. J., Bey, I., Butler, T., Cofala, J., Collins, W. J., Derwent, R. G., Doherty, R. M., Drevet, J., Eskes, H. J., Fiore, A. M., Gauss,
     M., Hauglustaine, D. A., Horowitz, L. W., Isaksen, I. S. A., Krol, M. C., Lamarque, J.-F., Lawrence, M. G., Montanaro, V., Müller, J.-F.,
     Pitari, G., Prather, M. J., Pyle, J. A., Rast, S., Rodriguez, J. M., Sanderson, M. G., Savage, N. H., Shindell, D. T., Strahan, S. E., Sudo,
     K., and Szopa, S.: Multimodel ensemble simulations of present-day and near-future tropospheric ozone, J. Geophys. Res.-Atmos., 111,
     doi:10.1029/2005JD006338, 2006.

10   Stockwell, W. R., Kirchner, F., Kuhn, M., and Seefeld, S.: A new mechanism for regional atmospheric chemistry modeling, J. Geophys.
     Res.-Atmos., 102, 25 847–25 879, doi:10.1029/97JD00849, 1997.

     Strode, S. A., Worden, H. M., Damon, M., Douglass, A. R., Duncan, B. N., Emmons, L. K., Lamarque, J.-F., Manyin, M., Oman, L. D.,
     Rodriguez, J. M., Strahan, S. E., and Tilmes, S.: Interpreting space-based trends in carbon monoxide with multiple models, Atmos. Chem.
     Phys., 16, 7285–7294, doi:10.5194/acp-16-7285-2016, 2016.

15   Tao, M., Konopka, P., Ploeger, F., Yan, X., Wright, J. S., Diallo, M., Fueglistaler, S., and Riese, M.: Multitimescale variations in modeled
     stratospheric water vapor derived from three modern reanalysis products, Atmos. Chem. Phys., 19, 6509–6534, doi:10.5194/acp-19-6509-
     2019, 2019.

     Thouret, V., Marenco, A., Logan, J. A., Nédélec, P., and Grouhel, C.: Comparisons of ozone measurements from the MOZAIC airborne
     program and the ozone sounding network at eight locations, J. Geophys. Res., 103, 25 695–25 720, doi:10.1029/98JD02243, 1998.

20   Thouret, V., Cammas, J.-P., Sauvage, B., Athier, G., Zbinden, R. M., Nédélec, P., Simon, P., and Karcher, F.: Tropopause referenced ozone cli-
     matology and inter-annual variability (1994–2003) from the MOZAIC programme, Atmos. Chem. Phys., 6, 1033–1051, doi:10.5194/acp-
     6-1033-2006, 2006.

     Tilmes, S., Lamarque, J.-F., Emmons, L. K., Kinnison, D. E., Marsh, D., Garcia, R. R., Smith, A. K., Neely, R. R., Conley, A., Vitt, F.,
     Val Martin, M., Tanimoto, H., Simpson, I., Blake, D. R., and Blake, N.: Representation of the Community Earth System Model (CESM1)
25   CAM4-chem within the Chemistry-Climate Model Initiative (CCMI), Geosci. Model Dev., 9, 1853–1890, doi:10.5194/gmd-9-1853-2016,
     2016.

     Williams, R. S., Hegglin, M. I., Kerridge, B. J., Jöckel, P., Latter, B. G., and Plummer, D. A.: Characterising the seasonal and geographical
     variability in tropospheric ozone, stratospheric influence and recent changes, Atmos. Chem. Phys., 19, 3589–3620, doi:10.5194/acp-19-
     3589-2019, 2019.

30   Young, P. J., Archibald, A. T., Bowman, K. W., Lamarque, J.-F., Naik, V., Stevenson, D. S., Tilmes, S., Voulgarakis, A., Wild, O., Bergmann,
     D., Cameron-Smith, P., Cionni, I., Collins, W. J., Dalsøren, S. B., Doherty, R. M., Eyring, V., Faluvegi, G., Horowitz, L. W., Josse, B.,

Lee, Y. H., MacKenzie, I. A., Nagashima, T., Plummer, D. A., Righi, M., Rumbold, S. T., Skeie, R. B., Shindell, D. T., Strode, S. A., Sudo, K., Szopa, S., and Zeng, G.: Pre-industrial to end 21st century projections of tropospheric ozone from the Atmospheric Chemistry and Climate Model Intercomparison Project (ACCMIP), Atmos. Chem. Phys., 13, 2063–2090, doi:10.5194/acp-13-2063-2013, 2013.

Young, P. J., Naik, V., Fiore, A. M., Gaudel, A., Guo, J., Lin, M. Y., Neu, J. L., Parrish, D. D., Rieder, H. E., Schnell, J. L., Tilmes, S., Wild, O., Zhang, L., Ziemke, J. R., Brandt, J., Delcloo, A., Doherty, R. M., Geels, C., Hegglin, M. I., Hu, L., Im, U., Kumar, R., Luhar, A., Murray, L., Plummer, D., Rodriguez, J., Saiz-Lopez, A., Schultz, M. G., Woodhouse, M. T., and Zeng, G.: Tropospheric Ozone Assessment Report: Assessment of global-scale model performance for global and regional ozone distributions, variability, and trends, Elem. Sci. Anth., 6, doi:10.1525/elementa.265, 2018.