



Spatially resolved evaluation of Earth system models with satellite column averaged CO₂

Bettina K. Gier^{1,2}, Michael Buchwitz¹, Maximilian Reuter¹, Peter M. Cox³, and Pierre Friedlingstein^{3,4},
Veronika Eyring^{2,1}

5 ¹University of Bremen, Institute of Environmental Physics (IUP), Bremen, Germany

²Deutsches Zentrum für Luft- und Raumfahrt (DLR), Institut für Physik der Atmosphäre, Oberpfaffenhofen, Germany

³College of Engineering, Mathematics and Physical Sciences, University of Exeter, Exeter, EX4 4QE, United Kingdom

⁴LMD/IPSL, ENS, PSL Université, École Polytechnique, Institut Polytechnique de Paris, Sorbonne Université, CNRS, Paris, France

10 *Correspondence to:* Bettina K. Gier (gier@uni-bremen.de)

Abstract. Earth System Models (ESMs) participating in the Coupled Model Intercomparison Project Phase 5 (CMIP5) showed large uncertainties in simulating atmospheric CO₂ concentrations. By comparing the simulations with satellite observations, in this study we find slight improvements in the ESMs participating in the new Phase 6 (CMIP6) compared to CMIP5. We utilize the Earth System Model Evaluation Tool (ESMValTool) to evaluate emission driven CMIP5 and CMIP6 simulations with satellite data of column-average CO₂ mole fractions (XCO₂). The satellite data are a combined data product covering the period 2003–2014 based on the SCIAMACHY/ENVISAT (2003–2012) and TANSO-FTS/GOSAT (2009–2014) instruments. In this study the Observations for Model Intercomparisons Project (Obs4MIPs) format data product version 3 (O4Mv3) with a spatial resolution of 5° x 5° and monthly time resolution has been used. Comparisons of XCO₂ time series show a large spread among the model ensembles both in CMIP5 and CMIP6, with differences in the absolute concentrations of up to approximately 20 ppmv relative to the satellite observations. The multi-model mean has a bias of approximately +10 and +2 ppmv in CMIP5 and CMIP6, respectively. The derived atmospheric XCO₂ growth rate (GR) is typically slightly overestimated in the models, with a slightly better average and lower spread for CMIP6. The interannual variability of the growth rate is well reproduced in the multi-model mean. All models capture the expected increase of the seasonal cycle amplitude (SCA) with increasing latitude, but most models underestimate the SCA. Most models from both ensembles show a positive trend of the SCA over the period 2003–2014, i.e. an increase of the SCA with XCO₂, similar to in situ ground-based measurements. In contrast, the combined satellite product shows a negative trend over this period. Any SCA derived from sampled data can only be considered an “effective” SCA, as sampling can remove the peaks or troughs. This negative trend can be reproduced by the models when sampling them as the observations. The average effective SCA in the models is higher when using the SCIAMACHY/ENVISAT instead of the TANSO-FTS/GOSAT mean data coverage mask, overall leading to a negative trend over the full period similar to the combined satellite product. Models with a larger difference in the average effective SCA between the two coverages also show a stronger trend reversal. Therefore, this trend reversal in the satellite data is due to sampling characteristics, more specifically the different data coverage of the two satellites that can be reproduced by the models



if sampled the same way. Overall, the CMIP6 ensemble shows better agreement with the satellite data in all considered quantities (XCO₂, GR, SCA and trend in SCA), with the biggest improvement in the mean XCO₂ content of the atmosphere.

35 This study shows that the availability of column-integral CO₂ from satellite provides a promising new way to evaluate the performance of Earth System Models on a global scale, complementing existing studies that are based on in situ measurements from single ground-based stations.

1 Introduction

The Intergovernmental Panel on Climate Change (IPCC) Fifth Assessment Report (AR5) concluded that since 1950 many of the observed changes in the climate system are unprecedented in the instrument record, confirming an unequivocal warming (IPCC, 2013). Increasing emissions of greenhouse gases (GHGs) are the key drivers of anthropogenic climate change. The most important anthropogenic greenhouse gas is carbon dioxide (CO₂), with CO₂ emissions contributing more than half of the total global radiative forcing in 2011 relative to 1750 (IPCC, 2013). It is therefore important to monitor the long-term changes in atmospheric CO₂ concentrations, to understand the sources and sinks of carbon, and to provide reliable projections of future CO₂ concentrations under various scenarios.

40

45

Plants absorb CO₂ while performing photosynthesis. This CO₂ is later partially released by autotrophic respiration, or after cycling through plant tissues, litter and soil carbon, is released through heterotrophic respiration by soil microbial decomposition, animals and other processes like forest fires. Photosynthesis is dominant in the growing season, resulting in a net uptake of atmospheric CO₂ in the northern hemispheric growing season. Conversely, atmospheric CO₂ concentrations rise throughout the northern hemispheric dormant season when there is a net release of CO₂ from the land due to decomposition of organic matter in soils. This uptake and release of carbon by the terrestrial biosphere throughout the year causes a seasonal cycle of atmospheric CO₂, with a minimum at the end of the northern hemispheric growing season and a maximum at the end of the northern hemispheric dormant season (Keeling et al., 1989). The seasonal cycle amplitude (SCA) has been increasing over the last 50 years, with higher increases in higher latitudes (Barnes et al., 2016; Graven et al., 2013; Yin et al., 2018; Keeling et al., 1995; Keeling et al., 1996; Myneni et al., 1997; Piao et al., 2018). A number of studies have explored the effects of CO₂ fertilization, land-use change and climate warming on the SCA (Bastos et al., 2019; Zhao et al., 2016; Fernández-Martínez et al., 2019). Although models do not agree unanimously, the dominant effect seems to be a positive trend in SCA due to the CO₂ fertilization combined with a negative trend due to climate warming. Some models however show a large positive trend due to climate warming (Zhao et al., 2016). Land-use is found to be a weaker effect in comparison to CO₂ fertilization and climate warming (Bastos et al., 2019; Fernández-Martínez et al., 2019).

50

55

60

Most long-term measurements of CO₂ are from ground-based stations. In situ ground-based measurements at Mauna Loa (Hawaii, USA) started in 1958, providing the first evidence that fossil fuel combustion leads to a measurable increase in atmospheric CO₂ concentrations (Keeling et al., 1976). Other observatories around the globe now also measure atmospheric CO₂, reporting an increase of about 45% since pre-industrial times (Ciais et al., 2013; Friedlingstein et al., 2019).



65 Satellite measurements of CO₂, with first near-infrared/short-wave-infrared (NIR/SWIR) nadir (downlooking) based satellite
retrievals starting in 2002, can complement the ground-based measurement network and provide regional and spatial
distributions of CO₂. The quantity obtained from measurements with NIR/SWIR satellite instruments is the column-average
dry-air mole fraction of atmospheric CO₂, denoted as XCO₂. XCO₂ is a dimensionless quantity defined as the vertical column
of CO₂ divided by the vertical column of dry air (i.e., all air molecules except water vapor) often given in ppmv (parts per
70 million per volume). An analysis of growth rates (GR) and seasonal cycle amplitude (SCA) from satellite data and their
sensitivity to growing season temperature anomaly presented in Schneising et al. (2014) shows a negative correlation between
SCA and growing season temperature anomaly for the period 2003–2011, which was confirmed by Yin et al. (2018) for SCA
anomaly in this timeframe. Satellite XCO₂ products are often used in combination with atmospheric transport inverse
modelling approaches to obtain information on surface fluxes by using a global or regional transport model with free fit
75 parameters (Basu et al., 2013; Houweling et al., 2015; Reuter et al., 2014; Chevallier et al., 2014). The satellite data can also
be used to constrain process parameters of a terrestrial biosphere model, e.g., as part of a Carbon Cycle Data Assimilation
System (CCDAS, e.g. Kaminski et al. (2013)), and have been used for the evaluation of chemistry-climate models (Hayman
et al., 2014; Shindell et al., 2013). In the last few years, satellite data have also been used in direct comparison to output from
climate models (e.g. Calle et al., 2019) characterizing rise and fall segments in seasonal cycles from GOSAT and comparing
80 them to model output.

A large ensemble of climate model simulations for different type of experiments under common forcings is provided by the
Coupled Model Intercomparison Project (CMIP), with output available for CMIP5 (Taylor et al., 2012) and more recently
Phase 6 (CMIP6, Eyring et al. (2016a)). ESMs produce a large range in projected atmospheric CO₂, as a result of uncertainties
in the future evolution of natural fluxes (Arora et al., 2013; Friedlingstein et al., 2006). Overall CMIP5 models overestimate
85 the carbon content of the atmosphere (Friedlingstein et al., 2014; Hoffman et al., 2014). The largest uncertainties are associated
with the response of the land carbon cycle to changes in climate and atmospheric CO₂ (Friedlingstein et al., 2014; Hajima et
al., 2014). The ability of ESMs to simulate the land and ocean contemporary carbon cycle has previously been investigated by
Anav et al. (2013). They showed that most models were able to correctly reproduce the main climatic variables and their
seasonal evolution, but found weaknesses in reproducing specific biogeochemical fields, such as a general overestimation of
90 photosynthesis. For CMIP6, Arora et al. (2019) analyzed simulations with a CO₂ increase of 1 % per year to quantify the
carbon-climate feedbacks. They found no significant change in behavior from CMIP5 to CMIP6, but lower absolute values
for models which included a nitrogen cycle.

In this paper we focus on evaluating the growth rate and the seasonal cycle amplitude of simulated CO₂, converted to XCO₂,
from CMIP ESMs which performed emission driven simulations with satellite observations in CMIP5 and CMIP6. The paper
95 is structured as follows: the data products used in this study are introduced in Section 2. Section 3 describes the methods used,
including the calculation of all derived quantities. A comparison between CO₂ flask measurements and XCO₂ measurements
at different locations is given in Section 4. The evaluation of CMIP simulations with satellite data is presented in Section 5,



divided into sections focusing at the models' ability to simulate XCO₂ time series, growth rate and seasonal cycle amplitude. A summary and conclusion is given in Section 6.

100 2 Data

2.1 Observational datasets

2.1.1 Satellite XCO₂

We use the Observations for Model Intercomparisons Project (obs4MIPs) version 3 (O4Mv3) XCO₂ satellite data (Buchwitz et al., 2017a; Buchwitz et al., 2018). Obs4MIPs hosts observationally based datasets which have been formatted according to the CMIP model output requirements (e.g. variable definitions, coordinates, frequencies) in order to facilitate an easier comparison between observations and models (Ferraro et al., 2015; Teixeira et al., 2014; Waliser et al., 2019). The satellite product used here is a gridded (Level 3) monthly data product with a 5° x 5° spatial resolution following the obs4MIPs format, produced as part of the Copernicus Climate Change Service (C3S). The O4Mv3 product is retrieved from the two satellite instruments SCIAMACHY/ENVISAT (Bovensmann et al., 1999; Burrows et al., 1995) and TANSO-FTS/GOSAT (Kuze et al., 2009).

This monthly mean XCO₂ satellite dataset covers a 14-year timespan (2003–2016). It is obtained by gridding the level 2 product (individual soundings) generated with the Ensemble Median Algorithm (EMMA, Reuter et al. (2013)), in this case EMMA version 3.0 (EMMAv3, Reuter et al. (2017)). EMMA combines several different XCO₂ level 2 satellite data products from SCIAMACHY/ENVISAT (2003–2012) and TANSO-FTS/GOSAT (2009–2016). This product was validated against Total Carbon Column Observing Network (TCCON, Wunch et al. (2011)) ground-based observations of XCO₂, revealing a +0.23 ppmv global bias, a relative accuracy (defined as standard deviation of the station-to-station biases) of 0.3 ppmv, and a very good stability in terms of a linear bias trend (-0.02 ± 0.04 ppmv yr⁻¹) (Buchwitz et al., 2017b). While the dataset ends in 2016, our evaluation only goes up to the year 2014 because the historical simulations for CMIP6 end in 2014 and scenarios from the emission-driven simulations that could be used to extend the runs are not yet available.

The number of observations depends significantly on the location with most points over locations with low cloud cover, high surface reflectivity and (at least) moderate to high sun elevation. Coverage over ocean is sparse as ocean retrievals are only included from GOSAT sun-glint mode observations - outside of glint conditions the reflectivity of water is very low in the NIR/SWIR spectral region. Figure 1 shows the mean monthly coverage of the dataset for 2003–2014. In Section 5 we will show that taking into account this sampling in the evaluation of ESMs is essential for a proper comparison.

The dataset also contains uncertainty estimates for each grid cell, with a mean value of 0.92 ppmv, accounting for both statistical uncertainties from the individual soundings and uncertainties from potential regional and temporal biases (Buchwitz et al., 2017a). However, the overall uncertainties are small compared to inter-model differences (see Section 3.1), and are therefore neglected in our analysis.



2.1.2 Surface CO₂ measurements

130 For the comparison of satellite XCO₂ and surface CO₂ data in Section 4 we have obtained surface flask measurements from the NOAA ESRL Carbon Cycle Cooperative Global Air Sampling Network (Dlugokencky et al., 2018). The sites were selected to cover different regions in the world and were chosen to have a maximum time overlap with the considered satellite period from 2003–2014. Furthermore, we excluded sites at locations with no available satellite data, which ruled out the four baseline observatories in Mauna Loa, Samoa, as well as the South Pole and Point Barrow sites. The selected sites are listed in Table 1.

135 2.2 Model simulations

We use monthly mean output data from ten CMIP5 and seven CMIP6 models which performed emission driven simulations, with three of the CMIP5 and four of the CMIP6 models including a nitrogen cycle. Tables 2 and 3 list all the CMIP5 and CMIP6 models used in this paper along with their atmosphere, land and ocean model component, respectively. Only models with interactive carbon cycle are able to perform the emission driven simulations, in which the emissions rather than the concentrations of the greenhouse gases are prescribed (Taylor et al., 2012; Eyring et al., 2016a). This allows the carbon cycle in the models to react to changes in climate and atmospheric CO₂, by adjusting their carbon fluxes to the new climate conditions and providing the atmospheric CO₂ concentration as an output (Friedlingstein et al., 2014). In order to facilitate the comparison between the satellite data and the CMIP5 emission driven simulations, the historical simulations (1850–2005) were extended beyond 2005 with simulations from the Representative Concentration Pathway (RCP) 8.5 (2006–2100), for which most ESM simulations are available. Since the period of observations only extends a decade beyond the historical runs, the choice of emissions scenario has a negligible impact on the results that we present below. For CMIP6 only the historical simulations are used, which end in 2014.

3 Methods

3.1 Sampling of observations and models

150 For comparison of model and satellite data, first the CO₂ data of the models were converted to XCO₂ data. The model data was interpolated to the grid of the satellite dataset using a bilinear interpolation scheme and grid cells with missing values in the satellite data were also set to missing values in the model fields. Further sampling considerations are discussed in more detail in Section 5.3.2 and in Appendix A.

Most analysis is carried out with regional averages covering several grid cells. Unless specifically stated otherwise, these are calculated by taking the arithmetic averages over all grid cells weighted by their area for each month.

3.2. Calculation of growth rate, seasonal cycle amplitude and growing season temperature anomaly

We compute the growth rate (GR) following the method described in Buchwitz et al. (2018). Monthly resolved annual growth rates are calculated by subtracting the XCO₂ value 6 months in the future from the one 6 months in the past. Then these monthly



resolved growth rates are averaged to a yearly GR for a calendar year, and any years with less than 7 months of data are flagged as missing. The addition of the 7-month data availability was introduced to be consistent with the constraint on SCA as explained below.

We define the SCA as the peak-to-trough amplitude in a calendar year of the detrended time series. The time series is detrended with the cumulative sum of monthly growth rates, using the annual mean growth rates as substitution for missing values where necessary. The SCA is calculated by subtracting the minimum from the maximum value for each year with a minimum data availability of 7 months. When investigating the seasonal cycle of observationally sampled simulations at higher latitudes, the maximum value of the time series was generally only accounted for if more than 7 months of data were available. We therefore introduce the cutoff of 7-month data availability to preserve as many peaks as possible without restricting the data too much. However, as peak preservation cannot be guaranteed when any missing values are present, we can only speak of an effective SCA. The absolute SCA is not as important in our comparison, because we use the same sampling for both the model and observations.

The growing season temperature anomaly ΔT is calculated from the GISTemp (Hansen et al., 2010) temperature anomaly map following Schneising et al. (2014). The data is masked to include only vegetated areas, using the MODIS Land Cover Classification (Friedl et al., 2010; Channan et al., 2014). Surface temperature anomalies are calculated with respect to their monthly climatologies. The data is averaged over the growing season if it covers only one hemisphere (April–September for the Northern Hemisphere, December to May for the Southern Hemisphere). Additionally, if the data covers both hemispheres, the whole year is taken into account. The growing season averages are taken because the temperature has a large influence on the plant growth and the resulting biospheric CO_2 fluxes, which in turn drive both the SCA and interannual variability of the GR (Schneising et al., 2014).

3.3 Earth System Model Evaluation Tool (ESMValTool)

All figures in this paper were produced with the Earth System Model Evaluation Tool (ESMValTool) version 2.0 (v2.0) (Righi et al., 2020; Eyring et al., 2019a; Lauer et al., 2020). Since its first release in 2016 (Eyring et al., 2016b) the ESMValTool has been further advanced facilitating analysis of many different ESM components, providing well-documented source code and scientific background of implemented diagnostics and metrics and allowing for traceability and reproducibility of results (provenance). ESMValTool v2.0 has been developed as a large community effort to specifically target the increased data volume of CMIP6 and the related challenges posed by analysis and evaluation of output from multiple high-resolution and complex ESMs. For this, the core functionalities have been completely rewritten in order to take advantage of state-of-the-art computational libraries and methods to allow for efficient and user-friendly data processing (Righi et al., 2020). Common operations on the input data such as regridding or computation of multi-model statistics are now centralized in a highly optimized preprocessor written in Python. The ESMValTool v2.0 includes an extended set of large-scale diagnostics for quasi-operational and comprehensive evaluation of ESMs (Eyring et al., 2019a), new diagnostics for extreme events, regional model and impact evaluation and analysis of ESMs (Weigel et al., 2020, in preparation), as well as diagnostics for emergent



constraints and analysis of future projections from ESMs (Lauer et al., 2020). For the study here, a new ESMValTool recipe has been developed that can be used to reproduce all plots of this paper.

4. Comparison of XCO₂ and surface CO₂

195 Until recent years, most model-observation comparisons have been carried-out using in situ surface CO₂ data (e.g. Wenzel et al., 2016). As such, it is interesting to compare the differences between XCO₂ and surface CO₂ at different locations. Figure 2 shows a comparison of time series between both kinds of observations and the multi-model mean for both XCO₂ and surface CO₂ for CMIP6 (top) and CMIP5 (bottom) models. The multi-model mean and satellite data are averaged between all grid cells covering a 5° x 5° radius around the stations, which results in a maximum of four grid cells to be considered. The multi-model mean is offset to have the same mean value as the satellite data, and this offset is noted above each time series panel. It is interesting to note that the offset appears to be larger at higher latitudes thus showing a different spatial gradient, indicating potential issues with surface fluxes or transport in the models. The mean and growth rate of XCO₂ and surface CO₂ are in very good agreement, while the multi-model mean overestimates both parameters at all sites, with the overestimated mean XCO₂ arising from the effect of higher growth rates over time. Furthermore, the offset from the modelled surface CO₂ is higher than
200 that of XCO₂, while this difference is smaller for CMIP5. This might be due to the fact that the CMIP5 offset for XCO₂ was larger overall with approximately 10 ppmv compared to the CMIP6 offset of approximately 2 ppmv.

SCA is higher at higher latitudes, and also generally higher at the surface compared to the column average. This is to be expected as the processes dominating the seasonal cycle, respiration and photosynthesis, take place at the surface. Mixing of air coming from lower latitudes with lower SCA dampens the SCA in the column compared to surface SCA. This is evident
210 in the increasing SCA difference between XCO₂ and surface CO₂ going from low latitude to high latitude stations, with no discernible seasonal cycle in the southern hemisphere due to the lack of land and vegetation. The multi-model mean follows this trend in the observations, although it underestimates the higher latitude SCA, with a larger underestimation at the surface while capturing the XCO₂ SCA relatively well. As this study aims at evaluating model simulations with satellite data, further analysis is restricted to XCO₂.

215 5. Evaluation of CMIP simulations with satellite data

5.1. XCO₂ time series

The globally averaged time series of XCO₂ is shown in Figure 3, with CMIP6 (top) and CMIP5 (bottom) models sampled as the satellite observations (see Section 3.1). The observational uncertainty is too small to be seen in this plot and is therefore neglected. As in Figure 2, an increase of XCO₂ with time and a pronounced seasonal cycle for all models can be seen. The
220 focus here is on the absolute values, as the trend (GR) and SCA are discussed in dedicated sections below. The CMIP5 models display a large range of absolute XCO₂ values, ranging from an underestimation by 15 ppmv (MRI-ESM2.0) to an overestimation by 20 ppmv (GFDL-ESM4). The models closest to the observations are NorESM2-LM and MPI-ESM1.2-LR



both overestimating XCO_2 by about 5 ppmv. The multi-model mean shows an overestimation by approximately 2 ppmv or equivalently a time-shift of 1 year. While the spread in the models has not decreased since CMIP5, the overestimation of the multi-model mean has decreased from 10 ppmv to 2 ppmv. Furthermore, CMIP6 models which have predecessors in CMIP5 show similar biases as their predecessors. The MPI model has the closest fit in both ensembles, while both MRI models underestimate XCO_2 significantly. The MRI-ESM1 model was the only model in CMIP5 to underestimate XCO_2 with respect to the observations, and this by about 20 ppmv. This model underestimates the historical warming, causing plant and soil respiration to be too low, which leads to a larger land sink and a reduced atmospheric CO_2 concentration (Adachi et al., 2013). This underestimation has been reduced by about 5 ppmv in CMIP6. The GFDL models show an overestimation of about 15 ppmv in both ensembles, and both CanESM models are 10 ppmv too high. A minor improvement can be seen for NorESM-LM over NorESM1-ME, with a decrease of the overestimation from 15 to 10 ppmv.

5.2 Growth Rate

Figure 4 shows the global mean GR of XCO_2 for 2003–2014 and its standard deviation over all years depicted as error bars, with the observations shown in black and the multi-model mean in red. The annual mean GR of the satellite data is 1.9 ± 0.4 ppmv yr^{-1} , while the CMIP5 models (right) range from 1.5 ± 0.4 ppmv yr^{-1} (MRI-ESM1) to 3.0 ± 0.9 ppmv yr^{-1} (CanESM2) with a multi-model mean of 2.4 ± 0.4 ppmv yr^{-1} . In CMIP6 (left), the multi-model mean is slightly lower at 2.3 ± 0.3 ppmv yr^{-1} and the spread has decreased by 0.6 ppmv yr^{-1} , with a range from 1.7 ± 0.4 ppmv yr^{-1} (MRI-ESM2.0) to 2.6 ± 0.7 ppmv yr^{-1} (GFDL-ESM4). As expected from Figure 3, the models - with the exception of MRI-ESM1 and MRI-ESM2.0 - overestimate the growth rate, leading to an increased XCO_2 level in the present-day atmosphere compared to observations. The interannual variability of the growth rate for the models is generally higher than that of the observations, but well reproduced in the multi-model mean.

The spatial variability of the GR is small, as CO_2 is long-lived and well mixed in the atmosphere with a 1 year mean interhemispheric crossing time. Thus there should be no significant regional changes on an annual level. Buchwitz et al. (2018) found the growth rate of the satellite dataset to be in agreement with those quoted by NOAA ESRLs global and Mauna Loa time series, as well as robust over several latitude bands. Our own analysis also shows only very small regional differences in the growth rate (not shown).

Emergent Constraints are relationships defined using an ensemble of models, between a measurable aspect of current or past climate and an aspect of Earth system feedback in the future, which can be constrained using observational data (Eyring et al., 2019b). Cox et al. (2013) developed an emergent constraint on the sensitivity of tropical land carbon to climate change using the sensitivity of the interannual variability (IAV) of CO_2 growth rate to the IAV of tropical temperature, which was later adapted by Wenzel et al. (2014) to CMIP5 models. Figure 5 shows the sensitivity of the IAV of XCO_2 growth rate to the tropical growing season temperature IAV for CMIP6 (left) and CMIP5 (right) models, both compared with observations. The observational temperature is taken from the GISTEMP v4 dataset (Lenssen et al., 2019) and the models use their own modeled temperature. We find an observational value of -0.23 ± 0.70 ppmv $yr^{-1} K^{-1}$ for the 2003–2014 period. However, when using the



full span of the satellite data until 2016, the slope increases to $0.75 \pm 0.6 \text{ ppmv yr}^{-1} \text{ K}^{-1}$ (not shown), as the additional years show both a high growing season temperature and GR IAV. This shows that the time period 2003–2014 is not sufficient to reproduce the emergent constraint, but it might be feasible once CMIP6 emission driven future simulations are available for a longer time overlap between models and observations. CMIP5 model values for the timeframe 2003–2014 range from $0.53 \pm 0.51 \text{ ppmv yr}^{-1} \text{ K}^{-1}$ (NorESM1-ME) to $3.14 \pm 0.63 \text{ ppmv yr}^{-1} \text{ K}^{-1}$ (MRI-ESM1), with only CESM1-BGC showing a negative trend of $-0.64 \pm 0.55 \text{ ppmv yr}^{-1} \text{ K}^{-1}$. The multi-model mean has a value of $0.23 \pm 0.70 \text{ ppmv yr}^{-1} \text{ K}^{-1}$. In CMIP6 the range is significantly decreased with a minimum of $1.14 \pm 0.56 \text{ ppmv yr}^{-1} \text{ K}^{-1}$ (ACCESS-ESM1.5) to a maximum of $2.07 \pm 0.33 \text{ ppmv yr}^{-1} \text{ K}^{-1}$ (CanESM5), with the multi-model mean at $1.36 \pm 0.32 \text{ ppmv yr}^{-1} \text{ K}^{-1}$.

5.3 Seasonal Cycle Amplitude

This section about the seasonal cycle amplitude (SCA) is divided into two subsections, with the first one taking a closer look at inter-model differences, while the second subsection is devoted to the impact of observational sampling.

5.3.1 Model differences

Figure 6 shows maps of the climatological mean SCA (2003–2014) for all models, with the global mean given in the top right. All CMIP6 models (Figure 6a) underestimate the SCA compared to satellite observations (Figure 7 middle) in the global mean, with the closest mean SCA being CanESM5. This underestimation was already present in CMIP5 (Figure 6b), with several studies revealing it for surface CO_2 SCA (Wenzel et al., 2016; Graven et al., 2013). In CMIP6 the multi-model mean has a globally averaged mean SCA of 3.45 ppmv, compared to 2.92 ppmv for CMIP5, while the observations show an effective SCA of 5.89 ppmv (Figure 7).

Both models and observations show the well-known increasing SCA with increasing latitude, due to the more pronounced seasonal cycle of the climate at higher latitudes. Tropical land regions in northern South America, Africa and south east Asia show increased SCA values compared to the ocean SCA at this latitude for the same model. While in the GFDL CMIP5 models this was so pronounced that these regions showed a higher SCA than the higher latitudes (Dunne et al., 2013), this is no longer the case for GFDL-ESM4 in CMIP6. Dunne et al. (2013) attributed the GFDL problem in CMIP5 to the seasonality of respiration in the northern latitudes and an Amazonian low-precipitation bias which reversed the seasonal cycle synchronizing it with the African and Oceanian rain forests.

The SCA in the MPI-ESM-LR model is on average twice as large as the observed one. The high SCA has been discussed in Giorgetta et al. (2013) where it was attributed to a combination of an overestimation of net primary productivity in ocean and land biology and uncertainties in atmospheric tracer transport. A particularly severe overestimation was seen in the Southern Hemisphere when comparing to station data. As shown in Figure 6, we additionally find a large overestimation in XCO_2 SCA in the Northern Hemisphere, in particular in the extra-tropics. For the CMIP6 successor model, MPI-ESM1.2-LR, the SCA is still the highest of the model ensemble, but is no longer twice as high as the other models. However, it now shows a more pronounced SCA over the tropical land regions mentioned above, which was not as dominant in CMIP5.



It is known that nitrogen limitations tend to suppress CO₂ fertilization (Reich et al., 2006). Of the four models with the lowest overall SCA in CMIP5 (CESM1-BGC, FIO-ESM, NorESM1-ME and BNU-ESM), three of them – BNU-ESM, CESM1-BGC, NorESM1-ME - include a nitrogen cycle. The SCA of NorESM1-ME and CESM1-BGC are very similar, which can be attributed to sharing the same land model (CLM4). FIO-ESM uses the predecessor CLM3.5 and is also comparable to the other two models. In CMIP6, ACCESS-ESM1.5, MPI-ESM1.2-LR, NorESM2-LM and UKESM1.0-LL include a nitrogen model but none of them share the same land model. While ACCESS-ESM1.5 has the lowest overall SCA, MPI-ESM1.2-LR has the highest, and thus the observation from CMIP5 models that N-cycle models feature a lower SCA no longer stands for CMIP6.

5.3.2 Influence of Sampling

There are a number of ways to compare model SCA to observational SCA, beginning with a grid box comparison. Figure 7 shows a comparison of the multi-model mean of CMIP6 (Figure 7a) and CMIP5 (Figure 7b) to observations. The top right shows the unsampled SCA. The top left panel shows the effective SCA when using observational sampling and the middle panel the satellite data's effective SCA. All numbers are given in ppmv. For an easier comparison the bottom panels show the absolute difference plots, with the left panel depicting the difference between sampled model and observations, and the right panel the difference between the sampled and unsampled model. Observational sampling slightly lowers the SCA, which is to be expected, as it could lead to masking out the peaks or troughs. While this effect was minimized by imposing the restriction of only computing the SCA of a year when at least 7 months of data are available, it is still a possibility. We therefore classify this SCA as an “effective SCA”. However, the SCA does not seem to decrease significantly through sampling and the difference does not follow a trend in latitude, so a grid box comparison seems feasible. This paves the way for more comprehensive spatial investigations, which previously relied on data from ground-based stations with sparse spatial coverage. While the stations provide data in higher latitudes that the satellite dataset does not cover, in the tropics and mid-latitudes the spatial coverage of the satellite is superior to the ground-based stations. A downside with this approach is the sparsity of the data when using observational sampling. Furthermore, this becomes a computationally expensive operation, as the SCA will need to be calculated for each grid box.

Another approach often used in model analysis is area averaging, e.g. over different latitude bands like the tropics or the northern mid-latitudes. Using surface flask measurements Wenzel et al. (2016) found an increased SCA with rising CO₂ concentration for CMIP5 – CO₂ fertilization -, and used this to establish an emergent constraint on the fertilization of terrestrial gross primary productivity (GPP). Figure 8 shows the SCA trend of CMIP6 models versus XCO₂ for 2003–2014 in the northern mid-latitudes (30–60° N), including a linear regression including slope, mean SCA and correlation. The left panel shows the trend in the unsampled models, while the right one shows the trend when following the sampling of the satellite data. The SCA was computed after a weighted area-average on the XCO₂ time series. While the unsampled models show an increasing SCA trend with increasing XCO₂, which is in agreement with the findings from Wenzel et al. (2016), the sampled model data instead shows a significant negative trend. Calculating the average with a zonal average before summing up the latitude bands does not change this result.



To investigate this reversal, Figure 9 shows the data coverage for different time periods, 2003–2008 for SCIAMACHY only measurements (top), the overlap between the two satellites in 2009–2012 (middle), and 2013–2014 for the GOSAT satellite only (bottom), with the pattern marking areas with a coverage of 50 % or above. Above 50° N SCIAMACHY measurements include more areas with 50 % or more coverage compared to GOSAT measurements. With a larger SCA in higher latitudes, it implies that SCIAMACHY would have a larger average SCA over this region compared to GOSAT, hence artificially generating a decreasing trend in the observed SCA, when moving from SCIAMACHY to GOSAT. Figure 10 shows the CMIP6 effective SCA trend with XCO₂ using SCIAMACHY and GOSAT masks obtained from masking out points with less than 50 % coverage. While the slopes remain largely the same, the mean effective SCA is higher in the models using the SCIAMACHY mask than when using the GOSAT mask. This mean SCA difference is larger than the total SCA difference within a model using the same sampling for the whole time period. Thus when considering the observational time series and its full sampling, the positive trend intrinsic to the model is dominated by the negative SCA difference going from the SCIAMACHY to the GOSAT data coverage and thus reversed to the negative trend seen in the observations. We can therefore attribute this trend reversal of the satellite to the influence of the different data coverage of the two satellites and are able to reproduce it with the models. This study on sampling also holds true for CMIP5 models, with the equivalent figures shown in Appendix B (Figures B1 and B2).

6. Summary and Conclusion

In this paper we have evaluated the performance of CMIP5 and CMIP6 ESMs with interactive carbon cycle (Tables 2 and 3) against column integral CO₂ data from satellite retrievals. Our analysis has compared ESM simulations to the 20032014 Obs4MIPs XCO₂ satellite dataset O4Mv3 retrieved from radiance spectra measured by the SCIAMACHY/ENVISAT (2003-2012) and TANSO-FTS/GOSAT (2009–2014) satellite instruments. The O4Mv3 data product has a spatial resolution of 5°x5° and monthly time resolution. For CMIP5, the historical simulations covering the period 2003–2005 were combined with simulations from the RCP 8.5 scenario (2006–2014) and for CMIP6 the historical simulation was used (2003–2014).

The evaluation of the CMIP models with the satellite data focused on the time series, the growth-rate (GR) and the seasonal cycle amplitude (SCA) of column integral CO₂ (XCO₂). All SCAs computed with a masked time series are considered to be “effective” SCAs due to the possibility of masking out peaks and troughs. The XCO₂ time series comparison shows that most models overestimate the carbon content of the atmosphere relative to the satellite observations in both model ensembles, with a lower overestimation for the CMIP6 models of 2 ppmv for the multi-model mean and a wide range of individual model differences of -15 ppmv to +20 ppmv. The CMIP5 models overestimate by 5 to 25 ppmv with the exception of the MRI-ESM1 model, which underestimates by 20 ppmv. The CMIP5 multi-model mean overestimates by 10 ppmv compared to the observations, which has also previously been found for surface comparisons (Friedlingstein et al., 2014; Hoffman et al., 2014). Overall, CMIP6 models follow the same trends as their CMIP5 counterparts, but with reduced systematic biases.

The XCO₂ annual mean growth rate is typically slightly overestimated compared to the observational value of 1.9 ± 0.4 ppmv yr⁻¹. CMIP6 models range from 1.7 ± 0.4 ppmv yr⁻¹ (MRI-ESM2.0) to 2.6 ± 0.7 ppmv yr⁻¹ (GFDL-ESM4) with a multi-model



mean of 2.3 ± 0.3 ppmv yr⁻¹. CMIP5 models have a slightly higher multi-model mean growth rate of 2.4 ± 0.4 ppmv yr⁻¹, and
 355 a larger spread, with the CMIP5 lowest model being MRI-ESM1 at 1.5 ± 0.4 ppmv yr⁻¹ and the highest CMIP5 growth rate
 shown by CanESM2 at 3.0 ± 0.9 ppmv yr⁻¹. We were able to reproduce the trend of increasing GR with increasing temperature
 anomaly in the tropics used by Cox et al. (2013) for the models, but the satellite observations showed a high variance in the
 considered timespan. No discernible sampling effect on GR or mean carbon content of the atmosphere was found, consistent
 with expectations.

360 All models capture the expected increase of the SCA with increasing latitudes, but most models underestimate the SCA to
 differing degrees in different regions. This result is in line with previous studies (Wenzel et al., 2016;Graven et al., 2013).
 Models with similar model components show similar behavior, with models including a nitrogen cycle generally showing a
 lower SCA in CMIP5, but this influence is not clear in CMIP6. Finally, the connection between SCA and XCO₂ was
 investigated in the northern midlatitudes. Most models from both ensembles show a positive trend, i.e., an increase of the SCA
 365 with XCO₂, consistent with findings for surface CO₂ (Wenzel et al., 2016). However, the satellite product shows a negative
 trend in contrast to the models and surface based observations. We have attributed this trend reversal to the sampling
 characteristics of the satellite products. The average effective SCA is higher for models sampled according to the
 SCIAMACHY/ENVISAT as opposed to the TANSO-FTS/GOSAT mean data coverage. As the early time series is based
 solely on the SCIAMACHY/ENVISAT data and the last years only use data from TANSO-FTS/GOSAT, this introduces an
 370 artificial negative trend which dominates the positive trend shown by the unsampled models. This demonstrates the importance
 of equal sampling of models and observations in model evaluation studies.

Overall, the CMIP6 ensemble shows an improved agreement with the satellite data in all considered quantities (mean XCO₂,
 growth rate, SCA and trend in SCA), with the biggest improvement shown in the mean XCO₂ content of the atmosphere. The
 paper demonstrates the great potential of satellite data for climate model evaluation as it allows to go beyond regional means
 375 or single point observations from in situ data, and also enables the investigation of regional effects on SCA, such as the increase
 in SCA at higher latitudes.



Appendix A. Calculation of XCO₂

380 Here we document the general procedure used to compute model XCO₂ for comparison with the satellite-based obs4MIPs product following the description in (Buchwitz and Reuter, 2016).

$$XCO_2 = \frac{\sum n_d \cdot c_{CO_2}}{\sum n_d} \quad (A1)$$

385 Here, c_{CO_2} represents the modeled CO₂ dry air mole fraction on model layers (i.e., layer centers or full levels) and n_d the number of dry air particles (air molecules excluding water vapor) within these levels. The summations are performed over all model layers. The number of dry air particles can be computed as follows:

$$n_d = \frac{N_a \cdot \Delta p \cdot (1 - q)}{m_d \cdot g} \quad (A2)$$

N_a is the Avogadro constant ($6.022140857 \cdot 10^{23} \text{ mol}^{-1}$) and m_d the molar mass of dry air ($28.9644 \cdot 10^{-3} \text{ kg mol}^{-1}$). Δp is the pressure difference (in hPa) computed from the model's pressure levels (i.e., layer boundaries or half levels) surrounding the model layers, q is the modeled specific humidity (in kg/kg), and g the gravitational acceleration approximated by:

390
$$g = \sqrt{g_0^2 - 2 \cdot f \cdot \phi} \quad (A3)$$

This includes the model's geopotential ϕ (in m^2s^{-2}) on layers, the free air correction constant $f = 3.0825959 \cdot 10^{-6} \text{ s}^{-2}$, and the gravitational acceleration g_0 on the geoid approximated by the international gravity formula depending only on the latitude φ :

$$g_0 = 9.780327 \cdot [1 + 0.0053024 \cdot \sin^2(\varphi) - 0.0000058 \cdot \sin^2(2\varphi)] \quad (A4)$$

395 Appendix B. Satellite data mean coverage impact on CMIP5 SCA trend

The analysis from Section 5.3.2. on the influence of the satellite data mean coverage on the trend of the SCA was also performed for CMIP5. Figures B1 and B2 are the CMIP5 equivalent of Figures 8 and 10. The CMIP5 models support the analysis of the CMIP6 models and show that the different satellite data coverage results in a different mean effective SCA, with a higher mean effective SCA for SCIAMACHY (2003–2012) than GOSAT (2009–2014) mean data coverage, which
 400 overshadows the positive trend and causes it to flip to a negative one in most models.

7. Data and Code Availability

The O4Mv3 XCO₂ data product is available via the Copernicus Climate Change Service (C3S, <https://climate.copernicus.eu/>) Climate Data Store (CDS) (<https://cds.climate.copernicus.eu/>), accessed Aug 2018. The surface flask measurements were obtained online (ftp://aftp.cmdl.noaa.gov/data/trace_gases/co2/flask/surface/) from Dlugokencky et al. (2018), accessed Aug-
 405 2018. Surface temperature anomalies were obtained from the GISTEMP Team, 2020: *GISS Surface Temperature Analysis*



(*GISTEMP*), version 4. NASA Goddard Institute for Space Studies. Dataset accessed 2020–0213 at data.giss.nasa.gov/gistemp/. The MODIS IGBP Land Cover Type Classification was obtained from <http://glcf.umd.edu/data/lc/> (accessed: 2018-01-31). CMIP5 data is available on various ESGF nodes (e.g. <https://esgf-data.dkrz.de/search/cmip5-dkrz/>) (Williams et al., 2009).

410 The corresponding recipe that can be used to reproduce the figures of this paper will be included in the ESMValTool v2.0 (Righi et al., 2020; Eyring et al., 2019a) as soon as the paper is published. The ESMValTool is released under the Apache License, VERSION 2.0. The ESMValTool code is available from the ESMValTool webpage at <https://www.esmvaltool.org/> and from github (<https://github.com/ESMValGroup/ESMValTool>).

8 Author contributions

415 BG led the writing and analysis of the paper. MB and MR provided the satellite dataset. VE, PC and PF contributed to the evaluation of the CMIP simulations. All authors contributed to the writing of the manuscript.

9 Competing interests

The authors declare that they have no conflict of interest.

10 Acknowledgements

420 This work has been supported by ESA Climate Change Initiative (CCI) Climate Modelling User Group (CMUG) project, the Horizon 2020 project Climate-Carbon Interactions in the Current Century (4C) funded by the EU (Grant agreement ID 821003) and the EVa4CMIP project funded by the Helmholtz Society. The generation of the satellite XCO₂ data product has received funding from ESA (GHG-CCI) and the EU (Copernicus Climate Change Service – C3S, led by ECMWF). We acknowledge the World Climate Research Programme (WCRP), which, through its Working Group on Coupled Modelling, coordinated and
425 promoted CMIP6. We thank the climate modeling groups (listed in Tables 2 and 3 of this paper) for producing and making available their model output, the Earth System Grid Federation (ESGF) for archiving the data and providing access, and the multiple funding agencies who support CMIP and ESGF.

References

- 430 Adachi, Y., Yukimoto, S., Deushi, M., Obata, A., Nakano, H., Tanaka, T. Y., Hosaka, M., Sakami, T., Yoshimura, H., Hirabara, M., Shindo, E., Tsujino, H., Mizuta, R., Yabu, S., Koshiro, T., Ose, T., and Kitoh, A.: Basic performance of a new earth system model of the Meteorological Research Institute (MRI-ESM1), *Papers in Meteorology and Geophysics*, 64, 1–19, 10.2467/mripapers.64.1, 2013.
- Anav, A., Friedlingstein, P., Kidston, M., Bopp, L., Ciais, P., Cox, P., Jones, C., Jung, M., Myneni, R., and Zhu, Z.: Evaluating the Land and Ocean Components of the Global Carbon Cycle in the CMIP5 Earth System Models, *J Climate*, 26, 6801–6843, 10.1175/Jcli-D-12-00417.1, 2013.
- 435



- Arora, V. K., Scinocca, J. F., Boer, G. J., Christian, J. R., Denman, K. L., Flato, G. M., Kharin, V. V., Lee, W. G., and Merryfield, W. J.: Carbon emission limits required to satisfy future representative concentration pathways of greenhouse gases, *Geophys Res Lett*, 38, n/a-n/a, Artn L05805 10.1029/2010gl046270, 2011.
- 440 Arora, V. K., Boer, G. J., Friedlingstein, P., Eby, M., Jones, C. D., Christian, J. R., Bonan, G., Bopp, L., Brovkin, V., Cadule, P., Hajima, T., Ilyina, T., Lindsay, K., Tjiputra, J. F., and Wu, T.: Carbon-Concentration and Carbon-Climate Feedbacks in CMIP5 Earth System Models, *J Climate*, 26, 5289-5314, 10.1175/Jcli-D-12-00494.1, 2013.
- Arora, V. K., Katavouta, A., Williams, R. G., Jones, C. D., Brovkin, V., Friedlingstein, P., Schwinger, J., Bopp, L., Boucher, O., Cadule, P., Chamberlain, M. A., Christian, J. R., Delire, C., Fisher, R. A., Hajima, T., Ilyina, T., Joetzjer, E., Kawamiya, M., Koven, C., Krasting, J., Law, R. M., Lawrence, D. M., Lenton, A., Lindsay, K., Pongratz, J., Raddatz, T., Séférian, R., Tachiiri, K., Tjiputra, J. F., Wiltshire, A., Wu, T., and Ziehn, T.: Carbon-concentration and carbon-climate feedbacks in CMIP6 models, and their comparison to CMIP5 models, *Biogeosciences Discuss.*, 2019, 1-124, 10.5194/bg-2019-473, 2019.
- 445
- Bao, Y., Qiao, F.-L., and Song, Z.: The historical global carbon cycle simulation of FIO-ESM, 6834, 2012.
- Barnes, E. A., Parazoo, N., Orbe, C., and Denning, A. S.: Isentropic transport and the seasonal cycle amplitude of CO₂, *J Geophys Res-Atmos*, 121, 8106-8124, 10.1002/2016jd025109, 2016.
- 450
- Bastos, A., Ciais, P., Chevallier, F., Rodenbeck, C., Ballantyne, A. P., Maignan, F., Yin, Y., Fernandez-Martinez, M., Friedlingstein, P., Penuelas, J., Piao, S. L., Sitch, S., Smith, W. K., Wang, X. H., Zhu, Z. C., Haverd, V., Kato, E., Jain, A. K., Lienert, S., Lombardozzi, D., Nabel, J. E. M. S., Peylin, P., Poulter, B., and Zhu, D.: Contrasting effects of CO₂ fertilization, land-use change and warming on seasonal amplitude of Northern Hemisphere CO₂ exchange, *Atmos Chem Phys*, 19, 12361-12375, 10.5194/acp-19-12361-2019, 2019.
- 455
- Basu, S., Guerlet, S., Butz, A., Houweling, S., Hasekamp, O., Aben, I., Krummel, P., Steele, P., Langenfelds, R., Torn, M., Biraud, S., Stephens, B., Andrews, A., and Worthy, D.: Global CO₂ fluxes estimated from GOSAT retrievals of total column CO₂, *Atmos Chem Phys*, 13, 8695-8717, 10.5194/acp-13-8695-2013, 2013.
- Bovensmann, H., Burrows, J. P., Buchwitz, M., Frerick, J., Noel, S., Rozanov, V. V., Chance, K. V., and Goede, A. P. H.: SCIAMACHY: Mission objectives and measurement modes, *J Atmos Sci*, 56, 127-150, Doi 10.1175/1520-0469(1999)056<0127:Smoamm>2.0.Co;2, 1999.
- 460
- Buchwitz, M., and Reuter, M.: Merged SCIAMACHY/ENVISAT and TANSO-FTS/GOSAT atmospheric column-average dry-air mole fraction of CO₂ (XCO₂) (XCO₂_CRDP3_001) - Technical Document, 2016.
- Buchwitz, M., Reuter, M., Schneising-Weigel, O., Aben, I., Detmers, R. G., Hasekamp, O. P., Boesch, H., Anand, J., Crevoisier, C., and Armante, R.: Product User Guide and Specification (PUGS) – Main document, Technical Report Copernicus Climate Change Service (C3S), 1-91, 2017a.
- 465
- Buchwitz, M., Reuter, M., Schneising-Weigel, O., Aben, I., Detmers, R. G., Hasekamp, O. P., Boesch, H., Anand, J., Crevoisier, C., and Armante, R.: Product Quality Assessment Report (PQAR) – Main document, Technical Report Copernicus Climate Change Service (C3S), 1-103, 2017b.
- 470
- Buchwitz, M., Reuter, M., Schneising, O., Noel, S., Gier, B., Bovensmann, H., Burrows, J. P., Boesch, H., Anand, J., Parker, R. J., Somkuti, P., Detmers, R. G., Hasekamp, O. P., Aben, I., Butz, A., Kuze, A., Suto, H., Yoshida, Y., Crisp, D., and O'Dell, C.: Computation and analysis of atmospheric carbon dioxide annual mean growth rates from satellite observations during 2003-2016, *Atmos Chem Phys*, 18, 1-22, 10.5194/acp-18-17355-2018, 2018.



- 475 Burrows, J. P., Holzle, E., Goede, A. P. H., Visser, H., and Fricke, W.: Sciamachy - Scanning Imaging Absorption Spectrometer for Atmospheric Chartography, *Acta Astronaut*, 35, 445-451, Doi 10.1016/0094-5765(94)00278-T, 1995.
- Calle, L., Poulter, B., and Patra, P. K.: A segmentation algorithm for characterizing rise and fall segments in seasonal cycles: an application to XCO₂ to estimate benchmarks and assess model bias, *Atmos. Meas. Tech.*, 12, 2611-2629, 10.5194/amt-12-2611-2019, 2019.
- 480 Channan, S., Collins, K., and Emanuel, W. R.: Global mosaics of the standard MODIS Vegetation Continuous Fields data, University of Maryland and the Pacific Northwest National Laboratory, College Park, Maryland, USA., 2014.
- Chevallier, F., Palmer, P. I., Feng, L., Boesch, H., O'Dell, C. W., and Bousquet, P.: Toward robust and consistent regional CO₂ flux estimates from in situ and spaceborne measurements of atmospheric CO₂, *Geophys Res Lett*, 41, 1065-1070, 10.1002/2013gl058772, 2014.
- 485 Cox, P. M., Pearson, D., Booth, B. B., Friedlingstein, P., Huntingford, C., Jones, C. D., and Luke, C. M.: Sensitivity of tropical carbon to climate change constrained by carbon dioxide variability, *Nature*, 494, 341-344, 10.1038/nature11882, 2013.
- Dunne, J. P., John, J. G., Adcroft, A. J., Griffies, S. M., Hallberg, R. W., Shevliakova, E., Stouffer, R. J., Cooke, W., Dunne, K. A., Harrison, M. J., Krasting, J. P., Malyshev, S. L., Milly, P. C. D., Phillips, P. J., Sentman, L. T., Samuels, B. L., Spelman, M. J., Winton, M., Wittenberg, A. T., and Zadeh, N.: GFDL's ESM2 Global Coupled Climate-Carbon Earth System Models. Part I: Physical Formulation and Baseline Simulation Characteristics, *J Climate*, 25, 6646-6665, 10.1175/Jcli-D-11-00560.1, 490 2012.
- Dunne, J. P., John, J. G., Shevliakova, E., Stouffer, R. J., Krasting, J. P., Malyshev, S. L., Milly, P. C. D., Sentman, L. T., Adcroft, A. J., Cooke, W., Dunne, K. A., Griffies, S. M., Hallberg, R. W., Harrison, M. J., Levy, H., Wittenberg, A. T., Phillips, P. J., and Zadeh, N.: GFDL's ESM2 Global Coupled Climate-Carbon Earth System Models. Part II: Carbon System Formulation and Baseline Simulation Characteristics, *J Climate*, 26, 2247-2267, 10.1175/Jcli-D-12-00150.1, 2013.
- 495 Dunne, J. P., Horowitz, L. W., Adcroft, A. J., Ginoux, P., Held, I. M., John, J. G., Krasting, J. P., Malyshev, S., Naik, V., Paulot, F., Shevliakova, E., Stock, C. A., Zadeh, N., Balaji, V., Blanton, C., Dunne, K. A., Dupuis, C., Durachta, J., Dussin, R., Gauthier, P. P. G., Griffies, S. M., Guo, H., Hallberg, R. W., Harrison, M., He, J., Hurlin, W., McHugh, C., Menzel, R., Milly, P. C. D., Nikonov, S., Paynter, D. J., Ploshay, J., Radhakrishnan, A., Rand, K., Reichl, B. G., Robinson, T., Schwarzkopf, D. M., Sentman, L. A., Underwood, S., Vahlenkamp, H., Winton, M., Wittenberg, A. T., Wyman, B., Zeng, Y., 500 and Zhao, M.: The GFDL Earth System Model version 4.1 (GFDL-ESM4.1): Model description and simulation characteristics., *Journal of Advances in Modeling Earth Systems*, submitted, 2019MS002008., 2019.
- Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., and Taylor, K. E.: Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization, *Geosci Model Dev*, 9, 1937-1958, 10.5194/gmd-9-1937-2016, 2016a.
- 505 Eyring, V., Righi, M., Lauer, A., Evaldsson, M., Wenzel, S., Jones, C., Anav, A., Andrews, O., Cionni, I., Davin, E. L., Deser, C., Ehbrecht, C., Friedlingstein, P., Gleckler, P., Gottschaldt, K. D., Hagemann, S., Juckes, M., Kindermann, S., Krasting, J., Kunert, D., Levine, R., Loew, A., Makela, J., Martin, G., Mason, E., Phillips, A. S., Read, S., Rio, C., Roehrig, R., Senftleben, D., Sterl, A., van Ulft, L. H., Walton, J., Wang, S. Y., and Williams, K. D.: ESMValTool (v1.0) - a community diagnostic and performance metrics tool for routine evaluation of Earth system models in CMIP, *Geosci Model Dev*, 9, 1747-1802, 510 10.5194/gmd-9-1747-2016, 2016b.
- Eyring, V., Bock, L., Lauer, A., Righi, M., Schlund, M., Andela, B., Arnone, E., Bellprat, O., Brötz, B., Caron, L. P., Carvalhais, N., Cionni, I., Cortesi, N., Crezee, B., Davin, E., Davini, P., Debeire, K., de Mora, L., Deser, C., Docquier, D., Earnshaw, P., Ehbrecht, C., Gier, B. K., Gonzalez-Reviriego, N., Goodman, P., Hagemann, S., Hardiman, S., Hassler, B.,



- 515 Hunter, A., Kadow, C., Kindermann, S., Koirala, S., Koldunov, N. V., Lejeune, Q., Lembo, V., Lovato, T., Lucarini, V.,
 Massonnet, F., Müller, B., Pandde, A., Pérez-Zanón, N., Phillips, A., Predoi, V., Russell, J., Sellar, A., Serva, F., Stacke, T.,
 Swaminathan, R., Torralba, V., Vegas-Regidor, J., von Hardenberg, J., Weigel, K., and Zimmermann, K.: ESMValTool v2.0
 - Extended set of large-scale diagnostics for quasi-operational and comprehensive evaluation of Earth system models in CMIP,
 Geosci. Model Dev. Discuss., 2019, 1-81, 10.5194/gmd-2019-291, 2019a.
- 520 Eyring, V., Cox, P. M., Flato, G. M., Gleckler, P. J., Abramowitz, G., Caldwell, P., Collins, W. D., Gier, B. K., Hall, A. D.,
 Hoffman, F. M., Hurtt, G. C., Jahn, A., Jones, C. D., Klein, S. A., Krasting, J. P., Kwiatkowski, L., Lorenz, R., Maloney, E.,
 Meehl, G. A., Pendergrass, A. G., Pincus, R., Ruane, A. C., Russell, J. L., Sanderson, B. M., Santer, B. D., Sherwood, S. C.,
 Simpson, I. R., Stouffer, R. J., and Williamson, M. S.: Taking climate model evaluation to the next level, Nature Climate
 Change, 9, 102-110, 10.1038/s41558-018-0355-y, 2019b.
- 525 Fernández-Martínez, M., Sardans, J., Chevallier, F., Ciais, P., Obersteiner, M., Vicca, S., Canadell, J. G., Bastos, A.,
 Friedlingstein, P., Sitch, S., Piao, S. L., Janssens, I. A., and Peñuelas, J.: Global trends in carbon sinks and their relationships
 with CO₂ and temperature, Nature Climate Change, 9, 73-79, 10.1038/s41558-018-0367-7, 2019.
- Ferraro, R., Waliser, D. E., Gleckler, P., Taylor, K. E., and Eyring, V.: Evolving Obs4MIPs to Support Phase 6 of the Coupled
 Model Intercomparison Project (CMIP6), Bulletin of the American Meteorological Society, 96, ES131-ES133, 10.1175/bams-
 d-14-00216.1, 2015.
- 530 Friedl, M. A., Sulla-Menashe, D., Tan, B., Schneider, A., Ramankutty, N., Sibley, A., and Huang, X. M.: MODIS Collection
 5 global land cover: Algorithm refinements and characterization of new datasets, Remote Sens Environ, 114, 168-182,
 10.1016/j.rse.2009.08.016, 2010.
- 535 Friedlingstein, P., Cox, P., Betts, R., Bopp, L., von Bloh, W., Brovkin, V., Cadule, P., Doney, S., Eby, M., Fung, I., Bala, G.,
 John, J., Jones, C., Joos, F., Kato, T., Kawamiya, M., Knorr, W., Lindsay, K., Matthews, H. D., Raddatz, T., Rayner, P., Reick,
 C., Roeckner, E., Schnitzler, K. G., Schnur, R., Strassmann, K., Weaver, A. J., Yoshikawa, C., and Zeng, N.: Climate–Carbon
 Cycle Feedback Analysis: Results from the C4MIP Model Intercomparison, J Climate, 19, 3337-3353, 10.1175/jcli3800.1,
 2006.
- Friedlingstein, P., Meinshausen, M., Arora, V. K., Jones, C. D., Anav, A., Liddicoat, S. K., and Knutti, R.: Uncertainties in
 CMIP5 Climate Projections due to Carbon Cycle Feedbacks, J Climate, 27, 511-526, 10.1175/Jcli-D-12-00579.1, 2014.
- 540 Friedlingstein, P., Jones, M. W., O'Sullivan, M., Andrew, R. M., Hauck, J., Peters, G. P., Peters, W., Pongratz, J., Sitch, S., Le
 Quéré, C., Bakker, D. C. E., Canadell, J. G., Ciais, P., Jackson, R. B., Anthoni, P., Barbero, L., Bastos, A., Bastrikov, V.,
 Becker, M., Bopp, L., Buitenhuis, E., Chandra, N., Chevallier, F., Chini, L. P., Currie, K. I., Feely, R. A., Gehlen, M., Gilfillan,
 D., Gkritzalis, T., Goll, D. S., Gruber, N., Gutekunst, S., Harris, I., Haverd, V., Houghton, R. A., Hurtt, G., Ilyina, T., Jain, A.
 K., Joetzjer, E., Kaplan, J. O., Kato, E., Klein Goldewijk, K., Korsbakken, J. I., Landschützer, P., Lauvset, S. K., Lefèvre, N.,
 545 Lenton, A., Lienert, S., Lombardozzi, D., Marland, G., McGuire, P. C., Melton, J. R., Metzl, N., Munro, D. R., Nabel, J. E. M.
 S., Nakaoka, S. I., Neill, C., Omar, A. M., Ono, T., Peregon, A., Pierrot, D., Poulter, B., Rehder, G., Resplandy, L., Robertson,
 E., Rödenbeck, C., Séférian, R., Schwinger, J., Smith, N., Tans, P. P., Tian, H., Tilbrook, B., Tubiello, F. N., van der Werf, G.
 R., Wiltshire, A. J., and Zaehle, S.: Global Carbon Budget 2019, Earth Syst. Sci. Data, 11, 1783-1838, 10.5194/essd-11-1783-
 2019, 2019.
- 550 Gent, P. R., Danabasoglu, G., Donner, L. J., Holland, M. M., Hunke, E. C., Jayne, S. R., Lawrence, D. M., Neale, R. B., Rasch,
 P. J., Vertenstein, M., Worley, P. H., Yang, Z. L., and Zhang, M. H.: The Community Climate System Model Version 4, J
 Climate, 24, 4973-4991, 10.1175/2011jcli4083.1, 2011.
- Giorgetta, M. A., Jungclaus, J., Reick, C. H., Legutke, S., Bader, J., Bottinger, M., Brovkin, V., Crueger, T., Esch, M., Fieg,
 K., Glushak, K., Gayler, V., Haak, H., Hollweg, H. D., Ilyina, T., Kinne, S., Kornblueh, L., Matei, D., Mauritsen, T.,



- 555 Mikolajewicz, U., Mueller, W., Notz, D., Pithan, F., Raddatz, T., Rast, S., Redler, R., Roeckner, E., Schmidt, H., Schnur, R., Segschneider, J., Six, K. D., Stockhause, M., Timmreck, C., Wegner, J., Widmann, H., Wieners, K. H., Claussen, M., Marotzke, J., and Stevens, B.: Climate and carbon cycle changes from 1850 to 2100 in MPI-ESM simulations for the Coupled Model Intercomparison Project phase 5, *J Adv Model Earth Sy*, 5, 572-597, 10.1002/jame.20038, 2013.
- Graven, H. D., Keeling, R. F., Piper, S. C., Patra, P. K., Stephens, B. B., Wofsy, S. C., Welp, L. R., Sweeney, C., Tans, P. P.,
 560 Kelley, J. J., Daube, B. C., Kort, E. A., Santoni, G. W., and Bent, J. D.: Enhanced seasonal exchange of CO₂ by northern ecosystems since 1960, *Science*, 341, 1085-1089, 10.1126/science.1239207, 2013.
- Hajima, T., Tachiiri, K., Ito, A., and Kawamiya, M.: Uncertainty of Concentration-Terrestrial Carbon Feedback in Earth System Models, *J Climate*, 27, 3425-3445, 10.1175/Jcli-D-13-00177.1, 2014.
- Hansen, J., Ruedy, R., Sato, M., and Lo, K.: Global Surface Temperature Change, *Rev Geophys*, 48, Artn Rg4004
 565 10.1029/2010rg000345, 2010.
- Hayman, G. D., O'Connor, F. M., Dalvi, M., Clark, D. B., Gedney, N., Huntingford, C., Prigent, C., Buchwitz, M., Schneising, O., Burrows, J. P., Wilson, C., Richards, N., and Chipperfield, M.: Comparison of the HadGEM2 climate-chemistry model against in situ and SCIAMACHY atmospheric methane data, *Atmos Chem Phys*, 14, 13257-13280, 10.5194/acp-14-13257-2014, 2014.
- 570 Hoffman, F. M., Randerson, J. T., Arora, V. K., Bao, Q., Cadule, P., Ji, D., Jones, C. D., Kawamiya, M., Khatiwala, S., Lindsay, K., Obata, A., Shevliakova, E., Six, K. D., Tjiputra, J. F., Volodin, E. M., and Wu, T.: Causes and implications of persistent atmospheric carbon dioxide biases in Earth System Models, *J Geophys Res-Bioge*, 119, 141-162, 10.1002/2013jg002381, 2014.
- Houweling, S., Baker, D., Basu, S., Boesch, H., Butz, A., Chevallier, F., Deng, F., Dlugokencky, E. J., Feng, L., Ganshin, A.,
 575 Hasekamp, O., Jones, D., Maksyutov, S., Marshall, J., Oda, T., O'Dell, C. W., Oshchepkov, S., Palmer, P. I., Peylin, P., Poussi, Z., Reum, F., Takagi, H., Yoshida, Y., and Zhuravlev, R.: An intercomparison of inverse models for estimating sources and sinks of CO₂ using GOSAT measurements, *Journal of Geophysical Research: Atmospheres*, 120, 5253-5266, 10.1002/2014jd022962, 2015.
- IPCC: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, Cambridge University Press, 1535, 10.1017/CBO9781107415324, 2013.
 580
- Ji, D., Wang, L., Feng, J., Wu, Q., Cheng, H., Zhang, Q., Yang, J., Dong, W., Dai, Y., Gong, D., Zhang, R. H., Wang, X., Liu, J., Moore, J. C., Chen, D., and Zhou, M.: Description and basic evaluation of Beijing Normal University Earth System Model (BNU-ESM) version 1, *Geosci Model Dev*, 7, 2039-2064, 10.5194/gmd-7-2039-2014, 2014.
- Kaminski, T., Knorr, W., Schurmann, G., Scholze, M., Rayner, P. J., Zaehle, S., Blessing, S., Dorigo, W., Gayler, V., Giering, R., Gobron, N., Grant, J. P., Heimann, M., Hooker-Stroud, A., Houweling, S., Kato, T., Kattge, J., Kelley, D., Kemp, S., Koffi, E. N., Kostler, C., Mathieu, P. P., Pinty, B., Reick, C. H., Rodenbeck, C., Schnur, R., Scipal, K., Sebal, C., Stacke, T., van Scheltinga, A. T., Vossbeck, M., Widmann, H., and Ziehn, T.: The BETHY/JSBACH Carbon Cycle Data Assimilation System: experiences and challenges, *J Geophys Res-Bioge*, 118, 1414-1426, 10.1002/jgrg.20118, 2013.
 585
- Keeling, C. D., Bacastow, R. B., Bainbridge, A. E., Ekdahl, C. A., Guenther, P. R., Waterman, L. S., and Chin, J. F. S.: Atmospheric Carbon-Dioxide Variations at Mauna-Loa Observatory, Hawaii, *Tellus*, 28, 538-551, 10.1111/j.2153-3490.1976.tb00701.x, 1976.
 590



- Keeling, C. D., Bacastow, R. B., Carter, A. F., Piper, S. C., Whorf, T. P., Heimann, M., Mook, W. G., and Roeloffzen, H.: A three-dimensional model of atmospheric CO₂ transport based on observed winds: 1. Analysis of observational data, in: Aspects of Climate Variability in the Pacific and the Western Americas, Geophysical Monograph Series, 165-236, 1989.
- 595 Keeling, C. D., Whorf, T. P., Wahlen, M., and Vanderpligt, J.: Interannual Extremes in the Rate of Rise of Atmospheric Carbon-Dioxide since 1980, *Nature*, 375, 666-670, DOI 10.1038/375666a0, 1995.
- Keeling, C. D., Chin, J. F. S., and Whorf, T. P.: Increased activity of northern vegetation inferred from atmospheric CO₂ measurements, *Nature*, 382, 146-149, 10.1038/382146a0, 1996.
- 600 Kuze, A., Suto, H., Nakajima, M., and Hamazaki, T.: Thermal and near infrared sensor for carbon observation Fourier-transform spectrometer on the Greenhouse Gases Observing Satellite for greenhouse gases monitoring, *Appl Opt*, 48, 6716-6733, 10.1364/AO.48.006716, 2009.
- Lauer, A., Eyring, V., Bellprat, O., Bock, L., Gier, B. K., Hunter, A., Lorenz, R., Pérez-Zanón, N., Righi, M., Schlund, M., Senftleben, D., Weigel, K., and Zechlau, S.: Earth System Model Evaluation Tool (ESMValTool) v2.0 – diagnostics for emergent constraints and future projections from Earth system models in CMIP, *Geosci. Model Dev. Discuss.*, 2020, 1-47, 10.5194/gmd-2020-60, 2020.
- 605 Law, R. M., Ziehn, T., Matear, R. J., Lenton, A., Chamberlain, M. A., Stevens, L. E., Wang, Y. P., Sribinovsky, J., Bi, D., Yan, H., and Vohralik, P. F.: The carbon cycle in the Australian Community Climate and Earth System Simulator (ACCESS-ESM1) – Part 1: Model description and pre-industrial simulation, *Geosci. Model Dev.*, 10, 2567-2590, 10.5194/gmd-10-2567-2017, 2017.
- 610 Lenssen, N. J. L., Schmidt, G. A., Hansen, J. E., Menne, M. J., Persin, A., Ruedy, R., and Zyss, D.: Improvements in the GISTEMP Uncertainty Model, *Journal of Geophysical Research: Atmospheres*, 124, 6307-6326, 10.1029/2018jd029522, 2019.
- Lindsay, K., Bonan, G. B., Doney, S. C., Hoffman, F. M., Lawrence, D. M., Long, M. C., Mahowald, N. M., Moore, J. K., Randerson, J. T., and Thornton, P. E.: Preindustrial-Control and Twentieth-Century Carbon Cycle Experiments with the Earth System Model CESM1(BGC), *J Climate*, 27, 8981-9005, 10.1175/Jcli-D-12-00565.1, 2014.
- 615 Mauritsen, T., Bader, J., Becker, T., Behrens, J., Bittner, M., Brokopf, R., Brovkin, V., Claussen, M., Crueger, T., Esch, M., Fast, I., Fiedler, S., Fläschner, D., Gayler, V., Giorgetta, M., Goll, D. S., Haak, H., Hagemann, S., Hedemann, C., Hohenegger, C., Ilyina, T., Jahns, T., Jimenéz-de-la-Cuesta, D., Jungclaus, J., Kleinen, T., Kloster, S., Kracher, D., Kinne, S., Kleberg, D., Lasslop, G., Kornbluh, L., Marotzke, J., Matei, D., Meraner, K., Mikolajewicz, U., Modali, K., Möbis, B., Müller, W. A., Nabel, J. E. M. S., Nam, C. C. W., Notz, D., Nyawira, S.-S., Paulsen, H., Peters, K., Pincus, R., Pohlmann, H., Pongratz, J., Popp, M., Raddatz, T. J., Rast, S., Redler, R., Reick, C. H., Rohrschneider, T., Schemann, V., Schmidt, H., Schnur, R., Schulzweida, U., Six, K. D., Stein, L., Stemmler, I., Stevens, B., von Storch, J.-S., Tian, F., Voigt, A., Vrese, P., Wieners, K.-H., Wilkenskeld, S., Winkler, A., and Roeckner, E.: Developments in the MPI-M Earth System Model version 1.2 (MPI-ESM1.2) and Its Response to Increasing CO₂, *J Adv Model Earth Sy*, 11, 998-1038, 10.1029/2018ms001400, 2019.
- 620 Myneni, R. B., Keeling, C. D., Tucker, C. J., Asrar, G., and Nemani, R. R.: Increased plant growth in the northern high latitudes from 1981 to 1991, *Nature*, 386, 698-702, 10.1038/386698a0, 1997.
- Piao, S., Liu, Z., Wang, Y., Ciais, P., Yao, Y., Peng, S., Chevallier, F., Friedlingstein, P., Janssens, I. A., Peñuelas, J., Sitch, S., and Wang, T.: On the causes of trends in the seasonal amplitude of atmospheric CO₂, *Global Change Biology*, 24, 608-616, 10.1111/gcb.13909, 2018.



- 630 Qiao, F. L., Song, Z. Y., Bao, Y., Song, Y. J., Shu, Q., Huang, C. J., and Zhao, W.: Development and evaluation of an Earth System Model with surface gravity waves, *J Geophys Res-Oceans*, 118, 4514-4524, 10.1002/jgrc.20327, 2013.
- Reich, P. B., Hobbie, S. E., Lee, T., Ellsworth, D. S., West, J. B., Tilman, D., Knops, J. M. H., Naeem, S., and Trost, J.: Nitrogen limitation constrains sustainability of ecosystem response to CO₂, *Nature*, 440, 922-925, 10.1038/nature04486, 2006.
- 635 Reuter, M., Bösch, H., Bovensmann, H., Bril, A., Buchwitz, M., Butz, A., Burrows, J. P., amp, apos, Dell, C. W., Guerlet, S., Hasekamp, O., Heymann, J., Kikuchi, N., Oshchepkov, S., Parker, R., Pfeifer, S., Schneising, O., Yokota, T., and Yoshida, Y.: A joint effort to deliver satellite retrieved atmospheric CO₂ concentrations for surface flux inversions: the ensemble median algorithm EMMA, *Atmos Chem Phys*, 13, 1771-1780, 10.5194/acp-13-1771-2013, 2013.
- 640 Reuter, M., Buchwitz, M., Hilker, M., Heymann, J., Schneising, O., Pillai, D., Bovensmann, H., Burrows, J. P., Bosch, H., Parker, R., Butz, A., Hasekamp, O., O'Dell, C. W., Yoshida, Y., Gerbig, C., Nehrkorn, T., Deutscher, N. M., Warneke, T., Notholt, J., Hase, F., Kivi, R., Sussmann, R., Machida, T., Matsueda, H., and Sawa, Y.: Satellite-inferred European carbon sink larger than expected, *Atmos Chem Phys*, 14, 13739-13753, 10.5194/acp-14-13739-2014, 2014.
- Reuter, M., Buchwitz, M., and Schneising-Weigel, O.: Product User Guide and Specification (PUGS) – ANNEX D for products XCO₂_EMMA and XCH₄_EMMA, 1-19, 2017.
- 645 Righi, M., Andela, B., Eyring, V., Lauer, A., Predoi, V., Schlund, M., Vegas-Regidor, J., Bock, L., Brötz, B., de Mora, L., Diblen, F., Dreyer, L., Drost, N., Earnshaw, P., Hassler, B., Koldunov, N., Little, B., Loosveldt Tomas, S., and Zimmermann, K.: Earth System Model Evaluation Tool (ESMValTool) v2.0 – technical overview, *Geosci. Model Dev.*, 13, 1179-1199, 10.5194/gmd-13-1179-2020, 2020.
- Schneising, O., Reuter, M., Buchwitz, M., Heymann, J., Bovensmann, H., and Burrows, J. P.: Terrestrial carbon sink observed from space: variation of growth rates and seasonal cycle amplitudes in response to interannual surface temperature variability, *Atmos Chem Phys*, 14, 133-141, 10.5194/acp-14-133-2014, 2014.
- 650 Seland, Ø., Bentsen, M., Seland Graff, L., Olivié, D., Toniazzo, T., Gjermundsen, A., Debernard, J. B., Gupta, A. K., He, Y., Kirkevåg, A., Schwinger, J., Tjiputra, J., Schancke Aas, K., Bethke, I., Fan, Y., Griesfeller, J., Grini, A., Guo, C., Ilicak, M., Hafsahl Karset, I. H., Landgren, O., Liakka, J., Onsum Moseid, K., Nummelin, A., Spensberger, C., Tang, H., Zhang, Z., Heinze, C., Iverson, T., and Schulz, M.: The Norwegian Earth System Model, NorESM2 - Evaluation of theCMIP6 DECK and historical simulations, *Geosci. Model Dev. Discuss.*, 2020, 1-68, 10.5194/gmd-2019-378, 2020.
- 655 Sellar, A. A., Jones, C. G., Mulcahy, J. P., Tang, Y., Yool, A., Wiltshire, A., O'Connor, F. M., Stringer, M., Hill, R., Palmieri, J., Woodward, S., de Mora, L., Kuhlbrodt, T., Rumbold, S. T., Kelley, D. I., Ellis, R., Johnson, C. E., Walton, J., Abraham, N. L., Andrews, M. B., Andrews, T., Archibald, A. T., Berthou, S., Burke, E., Blockley, E., Carslaw, K., Dalvi, M., Edwards, J., Folberth, G. A., Gedney, N., Griffiths, P. T., Harper, A. B., Hendry, M. A., Hewitt, A. J., Johnson, B., Jones, A., Jones, C. D., Keeble, J., Liddicoat, S., Morgenstern, O., Parker, R. J., Predoi, V., Robertson, E., Siahahan, A., Smith, R. S., Swaminathan, R., Woodhouse, M. T., Zeng, G., and Zerroukat, M.: UKESM1: Description and Evaluation of the U.K. Earth System Model, *J Adv Model Earth Sy*, 11, 4513-4558, 10.1029/2019MS001739, 2019.
- 665 Shindell, D. T., Lamarque, J. F., Schulz, M., Flanner, M., Jiao, C., Chin, M., Young, P. J., Lee, Y. H., Rotstayn, L., Mahowald, N., Milly, G., Faluvegi, G., Balkanski, Y., Collins, W. J., Conley, A. J., Dalsoren, S., Easter, R., Ghan, S., Horowitz, L., Liu, X., Myhre, G., Nagashima, T., Naik, V., Rumbold, S. T., Skeie, R., Sudo, K., Szopa, S., Takemura, T., Voulgarakis, A., Yoon, J. H., and Lo, F.: Radiative forcing in the ACCMIP historical and future climate simulations, *Atmos Chem Phys*, 13, 2939-2974, 10.5194/acp-13-2939-2013, 2013.
- Swart, N. C., Cole, J. N. S., Kharin, V. V., Lazare, M., Scinocca, J. F., Gillett, N. P., Anstey, J., Arora, V., Christian, J. R., Hanna, S., Jiao, Y., Lee, W. G., Majaess, F., Saenko, O. A., Seiler, C., Seinen, C., Shao, A., Sigmond, M., Solheim, L., von



- 670 Salzen, K., Yang, D., and Winter, B.: The Canadian Earth System Model version 5 (CanESM5.0.3), *Geosci. Model Dev.*, 12, 4823-4873, 10.5194/gmd-12-4823-2019, 2019.
- Taylor, K. E., Stouffer, R. J., and Meehl, G. A.: An Overview of CMIP5 and the Experiment Design, *Bulletin of the American Meteorological Society*, 93, 485-498, 10.1175/bams-d-11-00094.1, 2012.
- 675 Teixeira, J., Waliser, D., Ferraro, R., Gleckler, P., Lee, T., and Potter, G.: Satellite Observations for CMIP5: The Genesis of Obs4MIPs, *Bulletin of the American Meteorological Society*, 95, 1329-1334, 10.1175/bams-d-12-00204.1, 2014.
- Tjiputra, J. F., Roelandt, C., Bentsen, M., Lawrence, D. M., Lorentzen, T., Schwinger, J., Seland, O., and Heinze, C.: Evaluation of the carbon cycle components in the Norwegian Earth System Model (NorESM), *Geosci Model Dev*, 6, 301-325, 10.5194/gmd-6-301-2013, 2013.
- 680 Waliser, D., Gleckler, P. J., Ferraro, R., Taylor, K. E., Ames, S., Biard, J., Bosilovich, M. G., Brown, O., Chepfer, H., Cinquini, L., Durack, P., Eyring, V., Mathieu, P. P., Lee, T., Pinnock, S., Potter, G. L., Rixen, M., Saunders, R., Shulz, J., Thepaut, J. N., and Tuma, M.: Observations for Model Intercomparison Project (Obs4MIPs): Status for CMIP6, *Geosci. Model Dev. Discuss.*, 2019, 1-30, 10.5194/gmd-2019-268, 2019.
- 685 Watanabe, S., Hajima, T., Sudo, K., Nagashima, T., Takemura, T., Okajima, H., Nozawa, T., Kawase, H., Abe, M., Yokohata, T., Ise, T., Sato, H., Kato, E., Takata, K., Emori, S., and Kawamiya, M.: MIROC-ESM 2010: model description and basic results of CMIP5-20c3m experiments, *Geosci Model Dev*, 4, 845-872, 10.5194/gmd-4-845-2011, 2011.
- Weigel, K., Eyring, V., Gier, B. K., Lauer, A., Righi, M., Schlund, M., Adeniyi, K., Arnone, E., Berg, P., Bock, L., Corti, S., Caron, L. P., Cionni, I., Drost, N., Hunter, A., Liedô, L., Mohr, C.-M., Pérez-Zanón, N., Predoi, V., Sandstad, M., Sillmann, J., Sterl, A., Vegas-Regidor, J., and von Hardenberg, J.: Earth System Model Evaluation Tool (ESMValTool) v2.0 - diagnostics for Earth system models in CMIP: hydroclimatic intensity, extreme events, regional and impact evaluation, and ensemble member selection, *Geosci. Model Dev.*, 2020, in preparation.
- 690 Wenzel, S., Cox, P. M., Eyring, V., and Friedlingstein, P.: Emergent constraints on climate-carbon cycle feedbacks in the CMIP5 Earth system models, *J Geophys Res-Bioge*, 119, 794-807, 10.1002/2013jg002591, 2014.
- Wenzel, S., Cox, P. M., Eyring, V., and Friedlingstein, P.: Projected land photosynthesis constrained by changes in the seasonal cycle of atmospheric CO₂, *Nature*, 538, 499-501, 10.1038/nature19772, 2016.
- 695 Williams, D. N., Ananthakrishnan, R., Bernholdt, D. E., Bharathi, S., Brown, D., Chen, M., Chervenak, A. L., Cinquini, L., Drach, R., Foster, I. T., Fox, P., Fraser, D., Garcia, J., Hankin, S., Jones, P., Middleton, D. E., Schwidder, J., Schweitzer, R., Schuler, R., Shoshani, A., Siebenlist, F., Sim, A., Strand, W. G., Su, M., and Wilhelmi, N.: The Earth System Grid: Enabling Access to Multimodel Climate Simulation Data, *Bulletin of the American Meteorological Society*, 90, 195-206, 10.1175/2008bams2459.1, 2009.
- 700 Wunch, D., Toon, G. C., Blavier, J. F., Washenfelder, R. A., Notholt, J., Connor, B. J., Griffith, D. W., Sherlock, V., and Wennberg, P. O.: The total carbon column observing network, *Philos Trans A Math Phys Eng Sci*, 369, 2087-2112, 10.1098/rsta.2010.0240, 2011.
- Yin, Y., Ciais, P., Chevallier, F., Li, W., Bastos, A., Piao, S. L., Wang, T., and Liu, H. Y.: Changes in the Response of the Northern Hemisphere Carbon Uptake to Temperature Over the Last Three Decades, *Geophys Res Lett*, 45, 4371-4380, 10.1029/2018gl077316, 2018.
- 705



- Yukimoto, S., Yoshimura, H., Hosaka, M., Sakami, T., Tsujino, H., Hirabara, M., Tanaka, T. Y., Deushi, M., Obata, A., Nakano, H., Adachi, Y., Shindo, E., Yabu, S., Ose, T., and Kitoh, A.: Meteorological Research Institute-Earth System Model Version 1 (MRI-ESM1), Technical Reports, 64, 88, 10.11483/mritechrepo.64, 2011.
- 710 Yukimoto, S., Adachi, Y., Hosaka, M., Sakami, T., Yoshimura, H., Hirabara, M., Tanaka, T. Y., Shindo, E., Tsujino, H., Deushi, M., Mizuta, R., Yabu, S., Obata, A., Nakano, H., Koshiro, T., Ose, T., and Kitoh, A.: A New Global Climate Model of the Meteorological Research Institute: MRI-CGCM3 - Model Description and Basic Performance -, Journal of the Meteorological Society of Japan, 90A, 23-64, 10.2151/jmsj.2012-A02, 2012.
- 715 Yukimoto, S., Kawai, H., Koshiro, T., Oshima, N., Yoshida, K., Urakawa, S., Tsujino, H., Deushi, M., Tanaka, T., Hosaka, M., Yabu, S., Yoshimura, H., Shindo, E., Mizuta, R., Obata, A., Adachi, Y., and Ishii, M.: The Meteorological Research Institute Earth System Model Version 2.0, MRI-ESM2.0: Description and Basic Evaluation of the Physical Component, Journal of the Meteorological Society of Japan. Ser. II, 97, 931-965, 10.2151/jmsj.2019-051, 2019.
- 720 Zhao, F., Zeng, N., Asrar, G., Friedlingstein, P., Ito, A., Jain, A., Kalnay, E., Kato, E., Koven, C. D., Poulter, B., Rafique, R., Sitch, S., Shu, S. J., Stocker, B., Viovy, N., Wiltshire, A., and Zaehle, S.: Role of CO₂, climate and land use in regulating the seasonal amplitude increase of carbon fluxes in terrestrial ecosystems: a multimodel analysis, Biogeosciences, 13, 5121-5137, 10.5194/bg-13-5121-2016, 2016.
- Ziehn, T., Lenton, A., Law, R. M., Matear, R. J., and Chamberlain, M. A.: The carbon cycle in the Australian Community Climate and Earth System Simulator (ACCESS-ESM1) – Part 2: Historical simulations, Geosci. Model Dev., 10, 2591-2614, 10.5194/gmd-10-2591-2017, 2017.

725



Tables

Table 1: List of active NOAA surface flask measurement sites used in this study.

Code	Location	Latitude [°]	Longitude [°]	Altitude [m]	Start Year
ASK	Assekrem, Algeria	23.2625	5.6322	2710.00	1995
CGO	Cape Grim, Australia	-40.6800	144.6800	94.00	1984
LEF	Park Falls, United States	45-945	269.7300	868.00	1994
HUN	Hegyhatsal, Hungary	46.950	16.650	248.00	1993
WIS	Ketura, Israel	30.8595	34.7809	482.00	1995



730 **Table 2: CMIP5 models analysed in this study. Under Comments, D stands for models including dynamic vegetation, and N for models including Nitrogen cycles.**

Model	Institute	Atmosphere Model	Land Model	Ocean Model	Comment	Main Reference
BNU-ESM	College of Global Change and Earth System Science,	CAM3.5	CoLM + BNU-DGVM	MOM4p1 + IBGC	N, D	Ji et al. (2014)
CanESM2	Canadian Center for Climate Modeling and Analysis, BC, Canada	CanAM4	CLASS2.7 + CTEM1	CMOC		Arora et al. (2011)
CESM1-BGC	National Center for Atmospheric Research Boulder, CO, USA	CAM4	CLM4	POP2 + BEC	N	Gent et al. (2011); Lindsay et al. (2014)
FIO-ESM	The First Institute of Oceanography, SOA, China	CAM3.0	CLM3.5 + CASA	POP2.0 + OCMIP-2		Bao et al. (2012); Qiao et al. (2013)
GFDL-ESM2G	Geophysical Fluid Dynamics Laboratory, United States	AM2	LM3.0	GOLD + TOPAZ2	D	Dunne et al. (2012); Dunne et al. (2013)
GFDL-ESM2M	Geophysical Fluid Dynamics Laboratory, United States	AM2	LM3.0	MOM4.1 + TOPAZ2	D	Dunne et al. (2012); Dunne et al. (2013)
MIROC-ESM	Japan Agency for Marine-Earth Science and Technology, Japan; Atmosphere and Ocean Research Institute, Japan	MIROC-AGCM + SPRINTARS	MATSIRO + SEIB-DGVM	COCO3.4 + NPZD	D	Watanabe et al. (2011)
MPI-ESM-LR	Max Planck Institute for Meteorology, Hamburg, Germany	ECHAM6	JSBACH + BETHY	MPIOM + HAMOCC5	D	Giorgetta et al. (2013)



MRI-ESM1	Meteorological Research Institute, Japan	MRI-AGCM3.3 + MASINGAR mk-2 + MRI- CCM2	HAL	MRI.COM3	D	Adachi et al. (2013); Yukimoto et al. (2012); Yukimoto et al. (2011)
NorESM1-ME	Norwegian Climate Center, Norway	CAM4-Oslo	CLM4	HAMOCC5	N	Tjiputra et al. (2013)



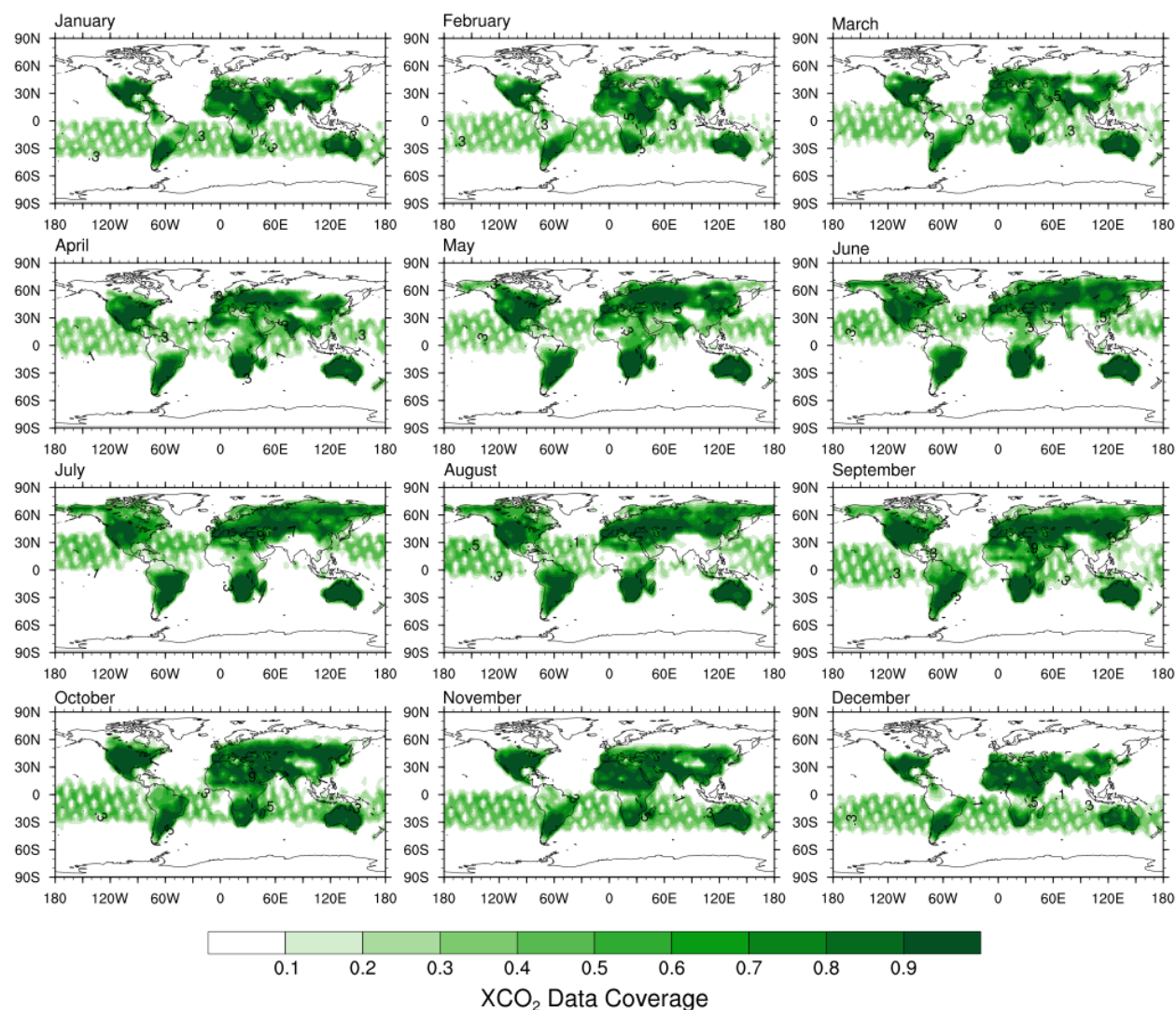
Table 3: CMIP6 models analysed in this study. Under Comments, D stands for models including dynamic vegetation, and N for models including Nitrogen cycles.

735

Model	Institute	Atmosphere Model	Land Model	Ocean Model	Comment	Main Reference
ACCESS-ESM1-5	Commonwealth Scientific and Industrial Research Organisation, Australia	UM7.3	CABLE2.4 with CASA-CNP	MOM5 + WOMBAT	N	Law et al. (2017); Ziehn et al. (2017)
CanESM5	Canadian Center for Climate Modeling and Analysis, BC, Canada	CanAM5	CLASS-CTEM	NEMO 3.4.1. + CMOC		Swart et al. (2019)
GFDL-ESM4	Geophysical Fluid Dynamics Laboratory, United States	AM4.1	LM4.1	OM4 MOM6 + COBALTv2	D	Dunne et al. (2019) (submitted)
MPI-ESM1-2-LR	Max Planck Institute for Meteorology, Hamburg, Germany	ECHAM6.3	JSBACH3.2	MPIOM1.6 + HAMOCC6	N, D	Mauritsen et al. (2019)
MRI-ESM2-0	Meteorological Research Institute, Japan	MRI-AGCM3.5 + MASINGAR mk-2r4c + MRI-CCM2.1	HAL	MRI.COMv4		Yukimoto et al. (2019)
NorESM2-LM	Norwegian Climate Center, Norway	Modified CAM6	CLM5	HAMOCC	N	Seland et al. (2020)
UKESM1-0-LL	Met Office Hadley Centre, United Kingdom	Unified Model	JULES-ES-1.0	NEMO + MEDUSA-2	N, D	Sellar et al. (2019)



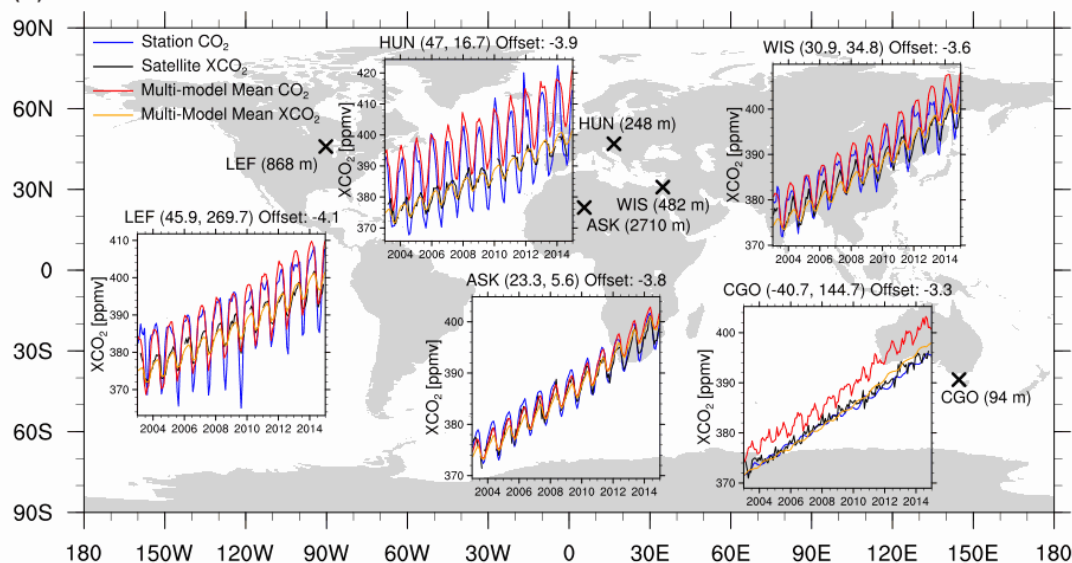
Figures



740 **Figure 1: Mean fractional coverage of monthly satellite data for 2003–2014. A value of 0 (white) signifies no available data, while a value of 1 (dark green) means that this gridcell contains data for all years of this month.**



(a) CMIP6



(b) CMIP5

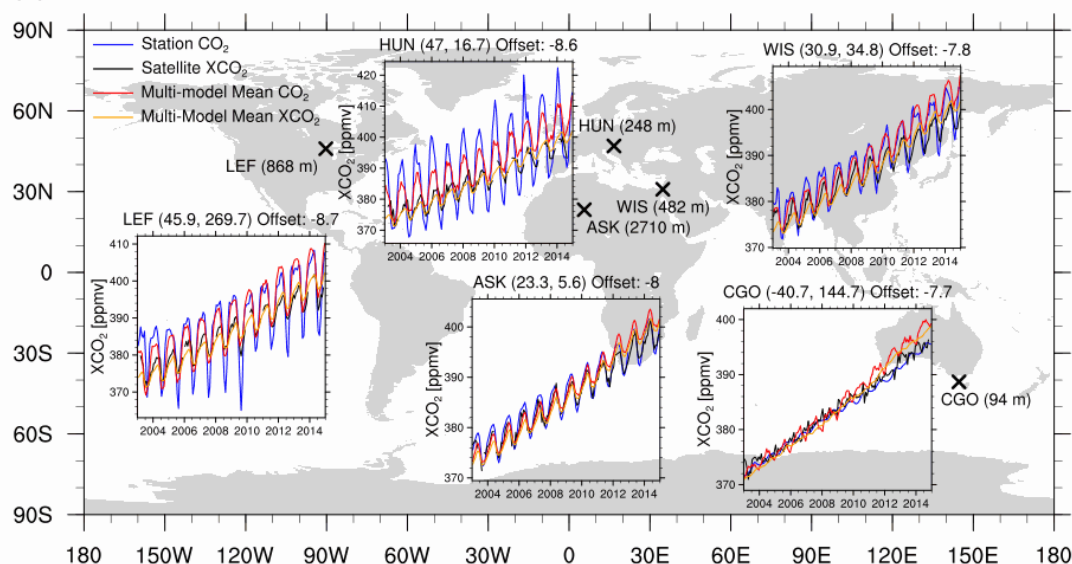
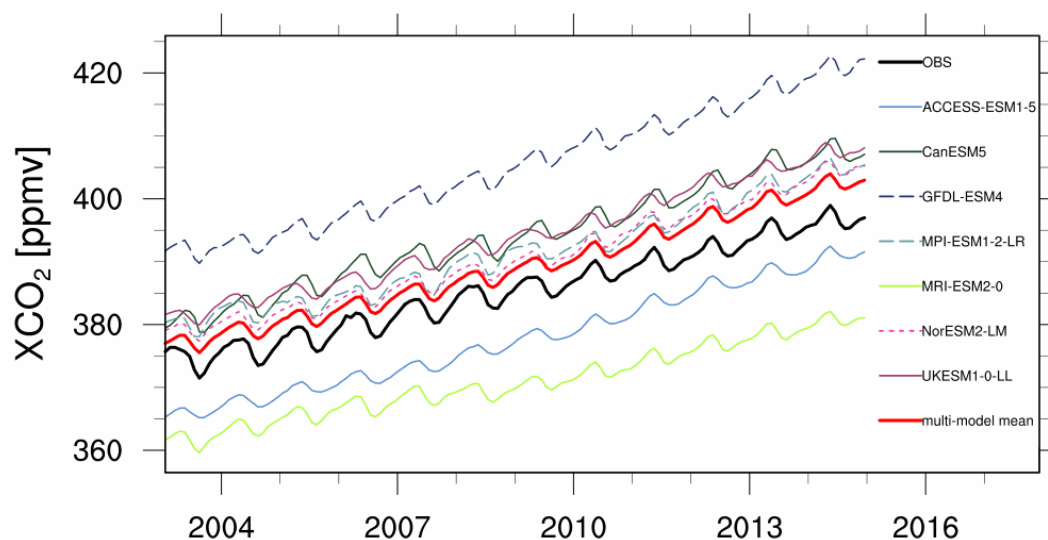


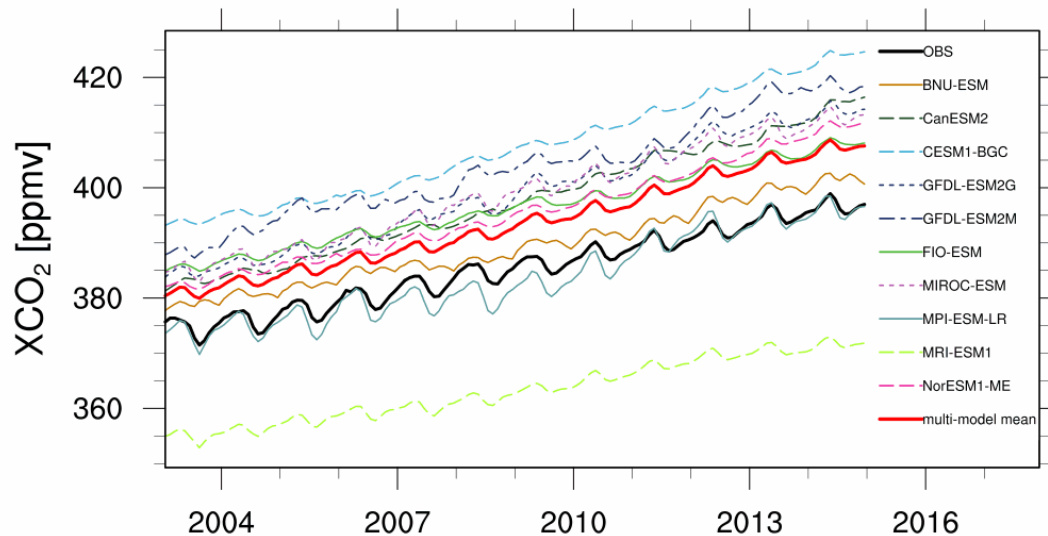
Figure 2: Comparison of time series from satellite XCO₂ (black), multi-model mean XCO₂ (orange) and surface CO₂ (red), and NOAA surface CO₂ station data (blue) at selected sites, with the coordinates noted in brackets above the time series and the altitudes shown in the map plot (see table 1). The multi-model mean for both XCO₂ and surface CO₂ was offset to have the same average value as the satellite XCO₂ for better comparison, and this offset is noted above each time series. CMIP6 and CMIP5 multi-model mean are shown on the top (a) and bottom (b) panels respectively.



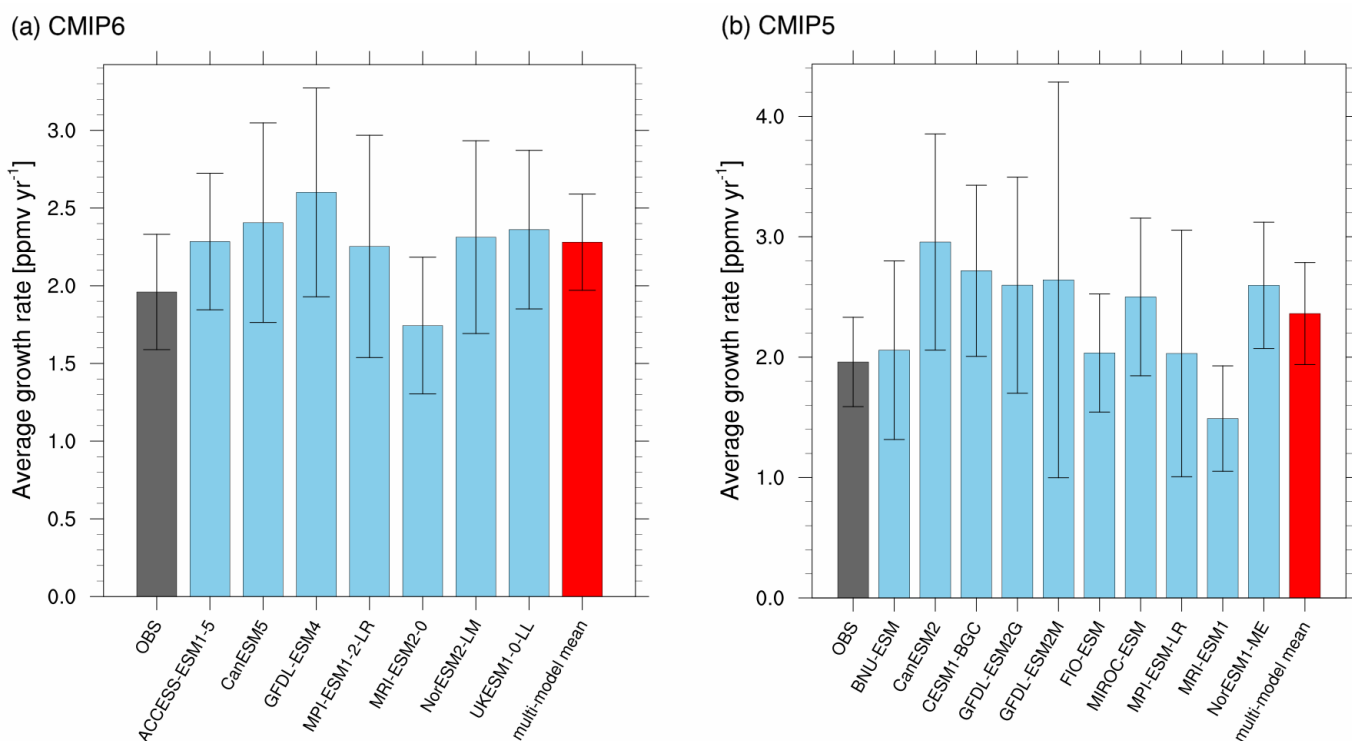
(a) CMIP6



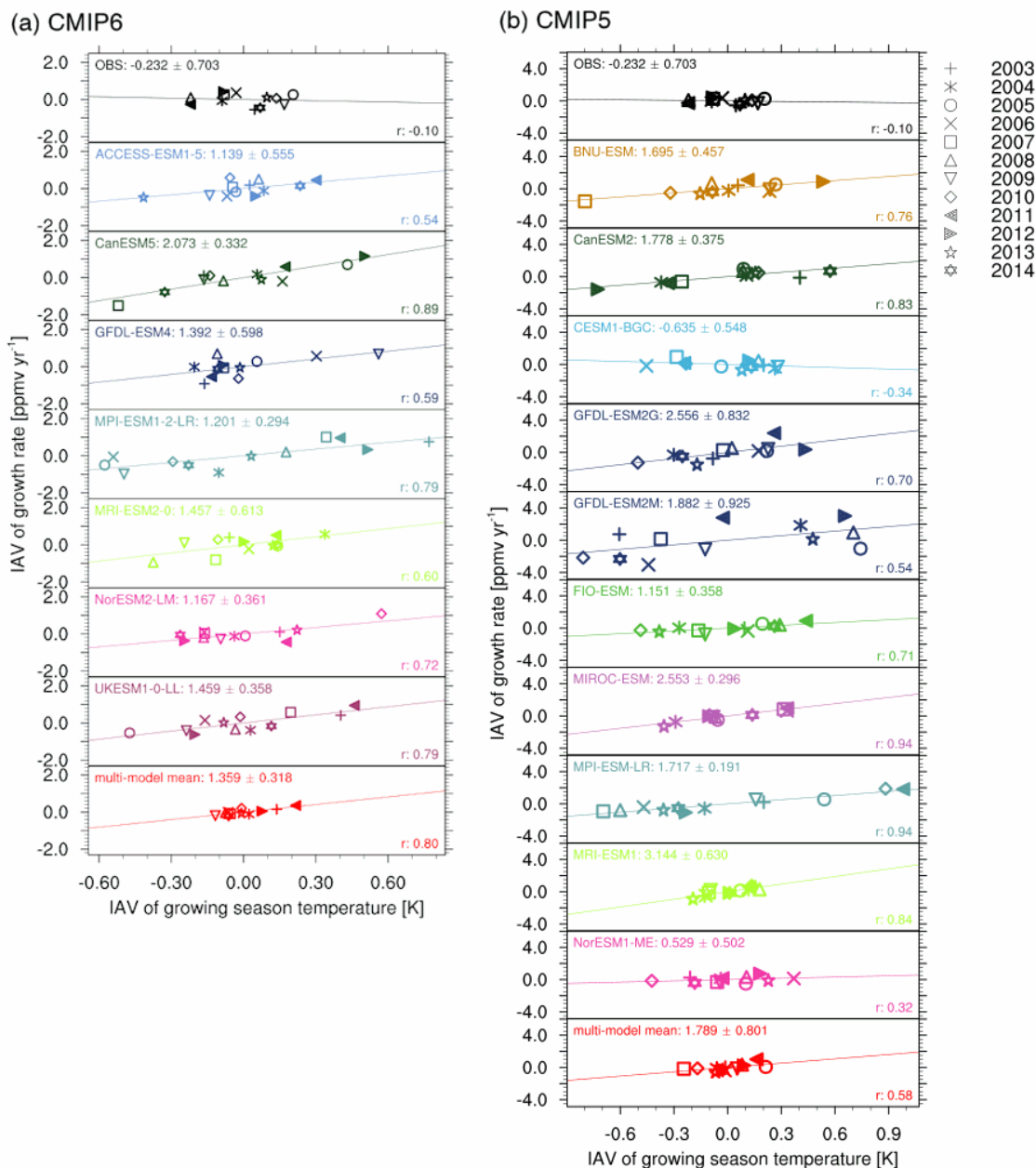
(b) CMIP5



750 **Figure 3: Global time series of monthly mean column averaged carbon dioxide (XCO_2) from 2003 to 2014, for the emission driven CMIP6 (a) and CMIP5 (b) model simulations in comparison to satellite XCO_2 data (bold black line). The model output is sampled as the satellite data.**



755 **Figure 4. Global mean and standard deviation over all years of annual growth rate of XCO₂ during 2003–2014, for CMIP6 models (a) and CMIP5 models (b). The leftmost bar (black) represents the satellite observations, while the rightmost bar (red) depicts the multi-model mean.**



760 **Figure 5: Sensitivity of the interannual variability of the XCO₂ growth rate in the tropics (30° S–30° N) to the interannual variability**
 765 **of tropical growing season temperature for CMIP6 models (a) and CMIP5 models (b). The observational temperature taken from**
the GISTemp temperature anomaly map, while the models use their own simulated temperature. A linear regression is performed
on the data for each dataset. Model colors are the same as in Figure 3, and symbols denote the years. In the top left of each panel the
regression coefficient and its uncertainty is shown, while the bottom right states the correlation coefficient.



(a) CMIP6

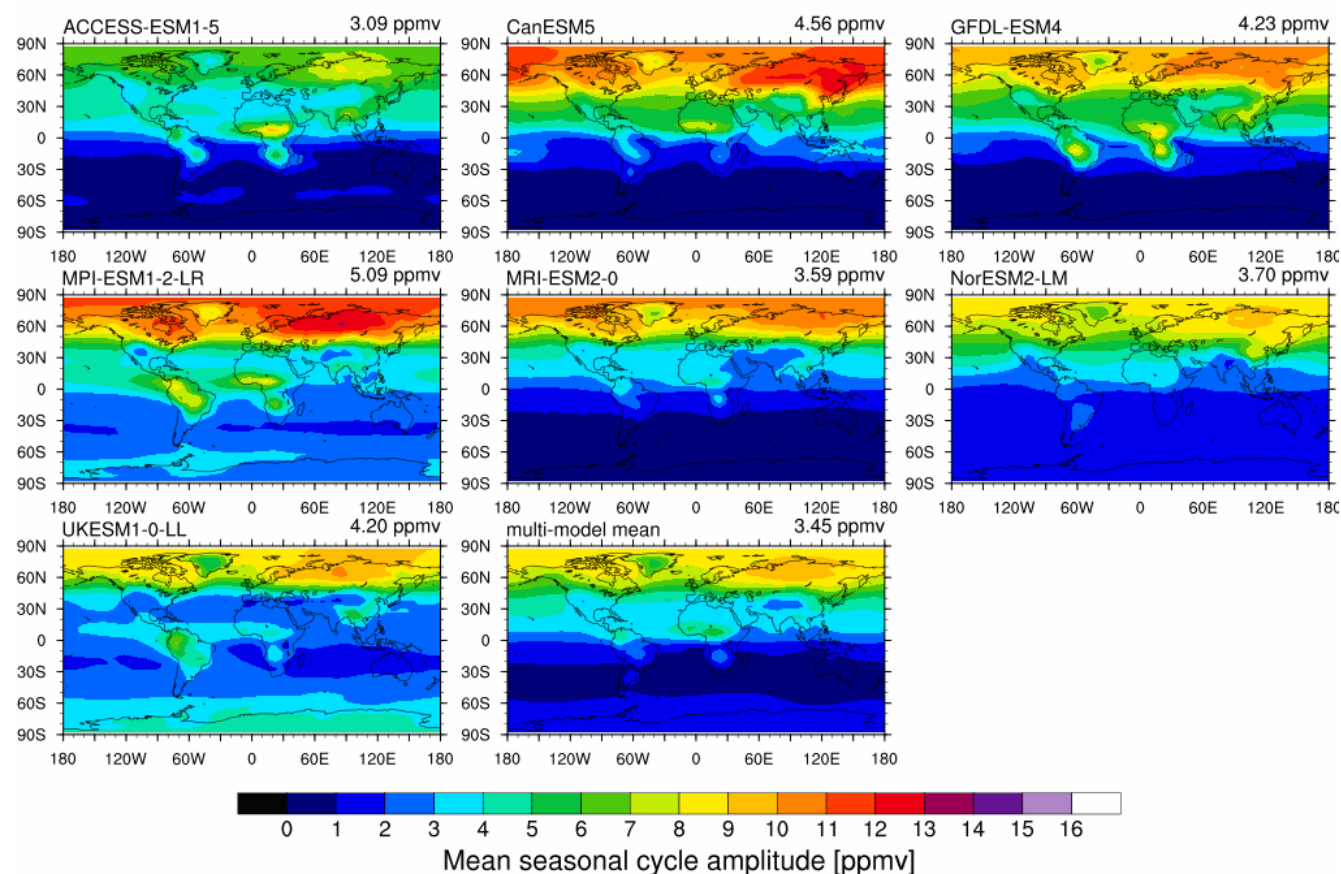
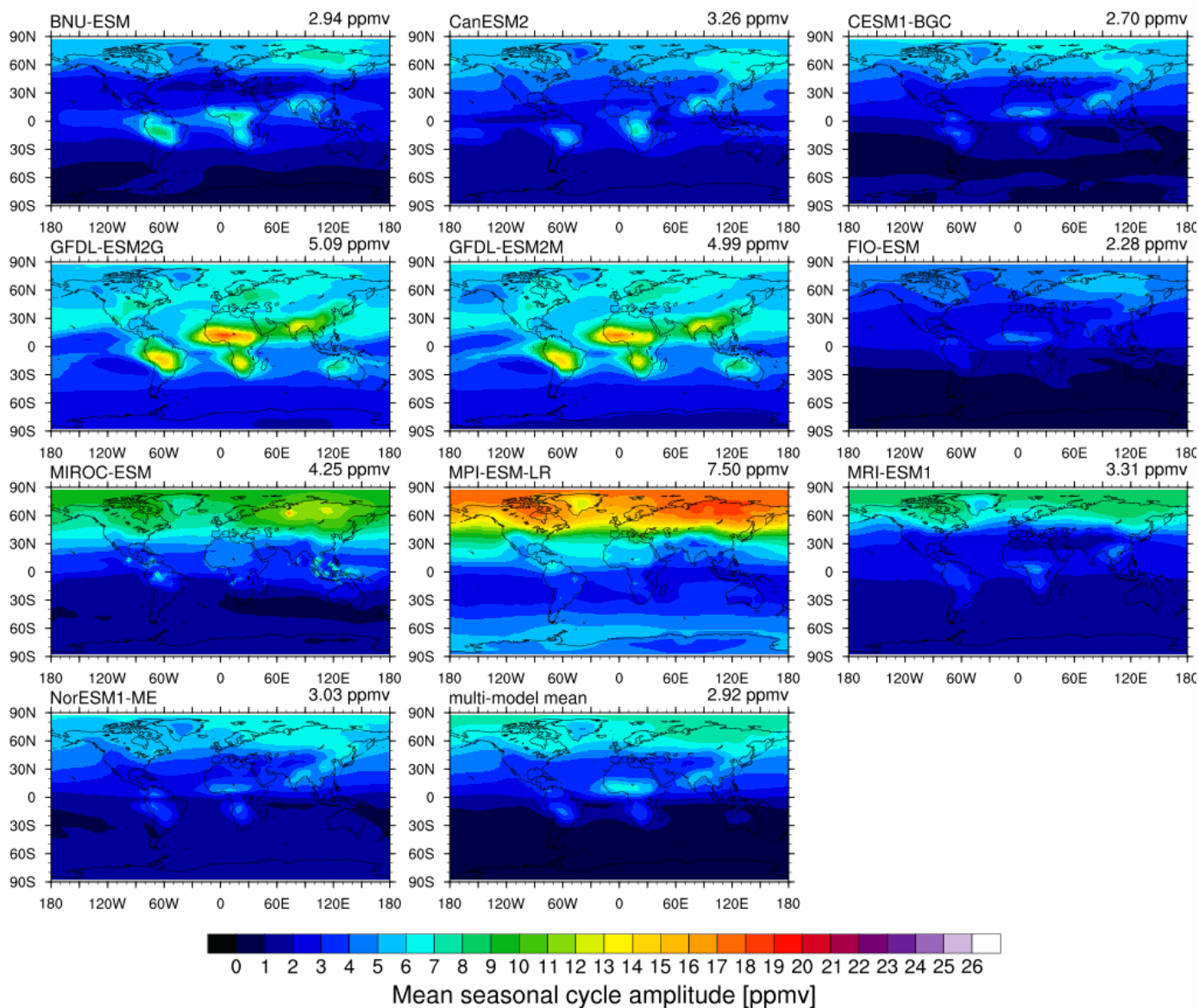


Figure 6a: Maps of mean annual seasonal cycle amplitude for 2003–2014 for CMIP6 models. The model name is given at the top left of each panel, while the top right shows the global average of the mean annual seasonal cycle.



(b) CMIP5



770 **Figure 6b:** Same as Figure 6a but for CMIP5 models. Note the different scale for the CMIP5 models to fit the increased SCA of the MPI-ESM-LR model.



(a) CMIP6

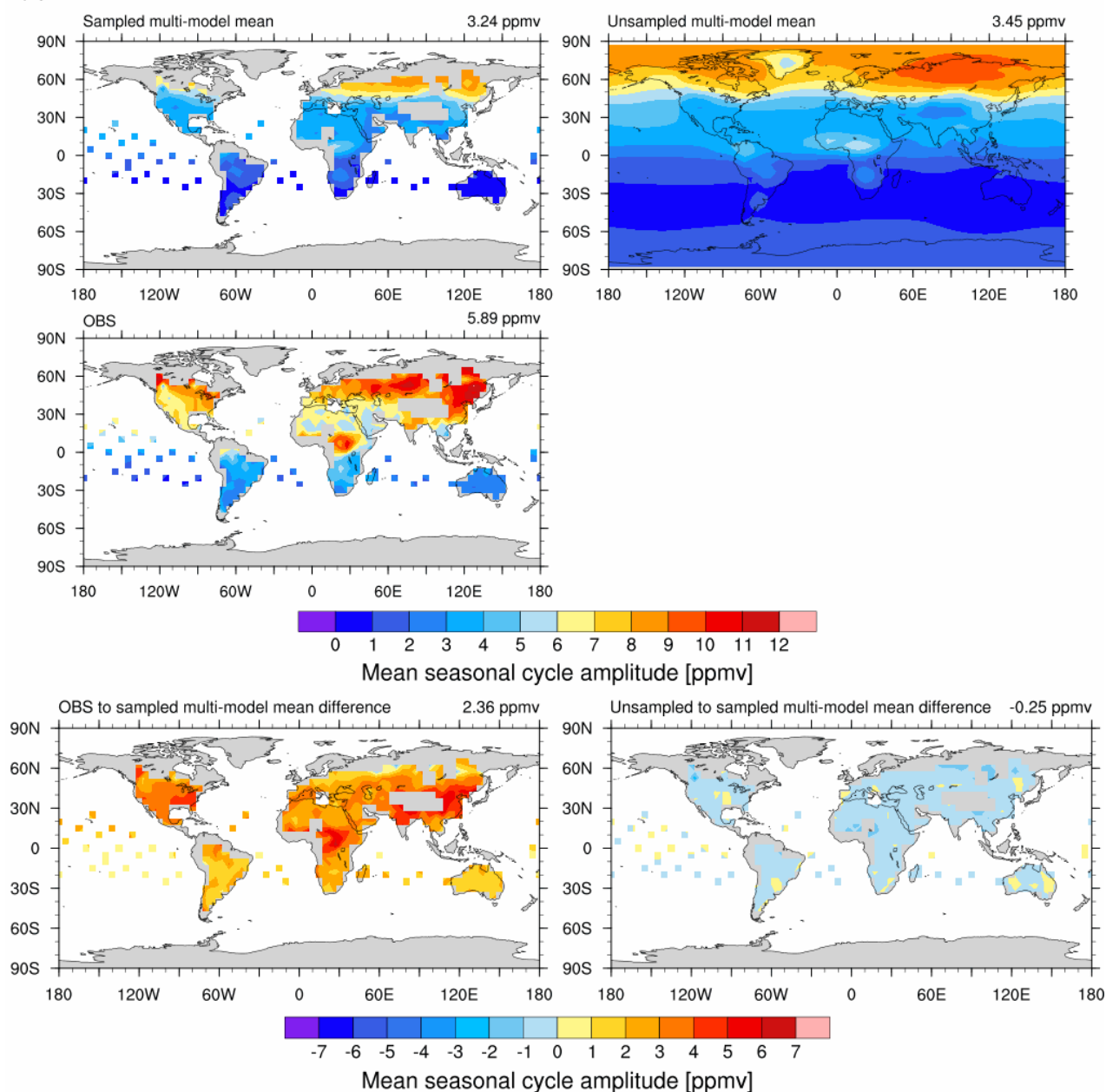


Figure 7a: Maps of mean SCA of the CMIP6 multi-model mean for 2003–2014. Top: SCA of multi-model mean with observational sampling (left) and without sampling (right). Middle: SCA of the satellite observations. Bottom: Difference between observations and sampled model data (left) and sampled and unsampled model (right).



(b) CMIP5

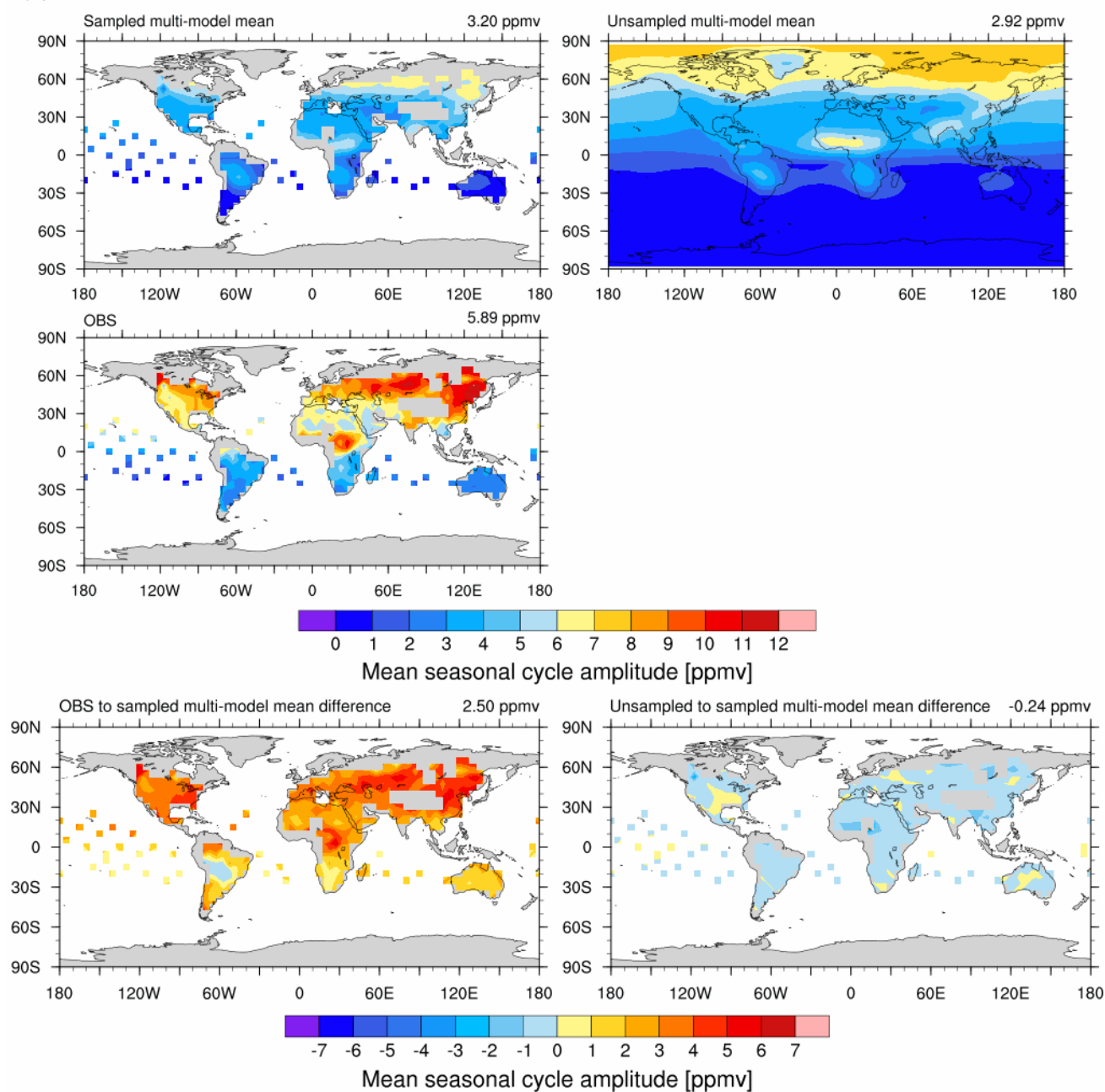


Figure 7b: Same as Figure 7a but for CMIP5 multi-model mean.

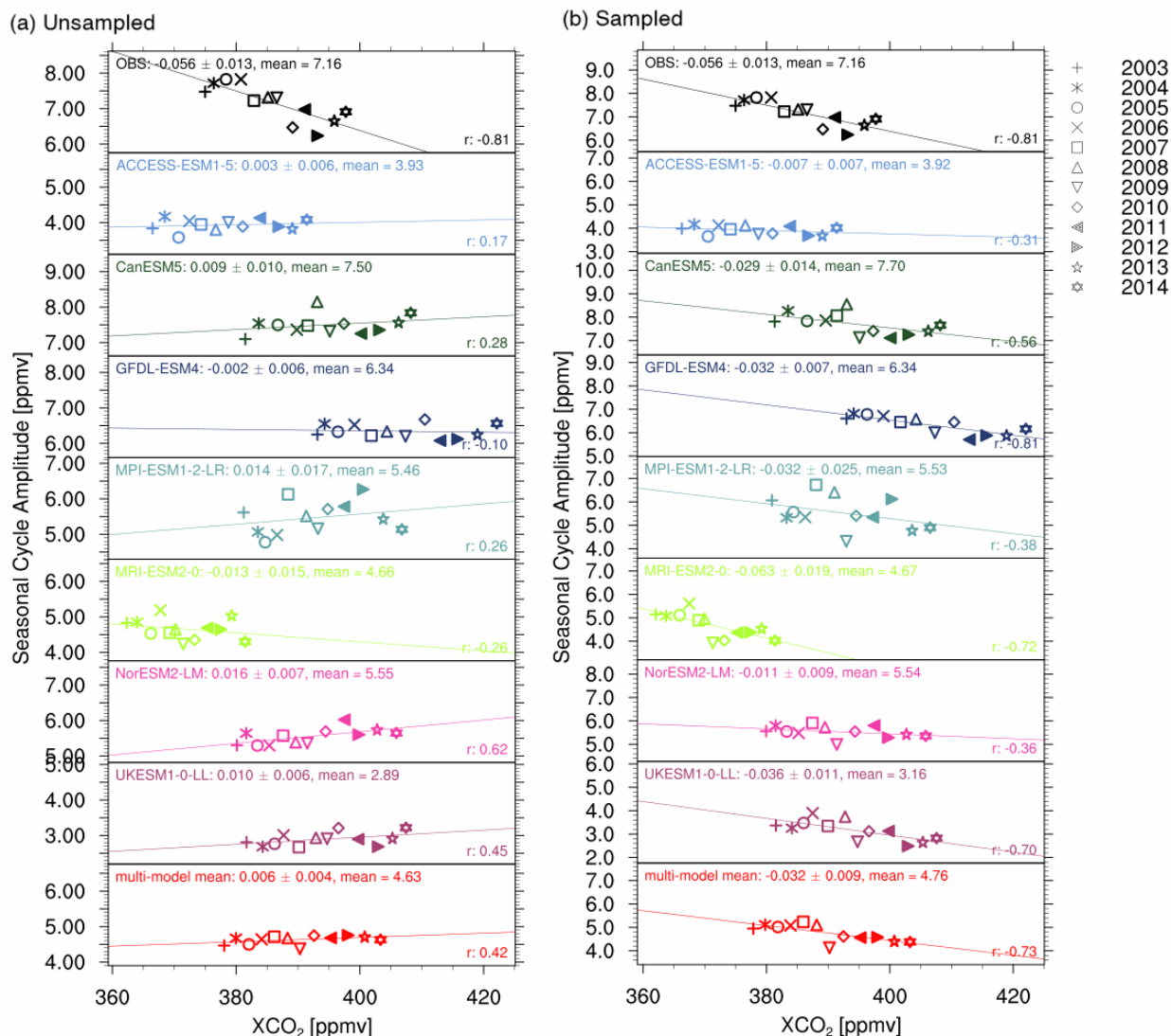


Figure 8: Trend of SCA with XCO₂ 2003–2014 for the northern mid-latitudes (30–60° N), including a linear regression with slope and mean SCA given on the top left of each panel and the correlation on the bottom right. Symbols denote the different years and model colors are consistent with previous figures. The left panels (a) show unsampled CMIP6 models, while CMIP6 models sampled according to the satellite data are shown on the right (b). Note that the y-axis range for each plot is the same and only differs by a shift.

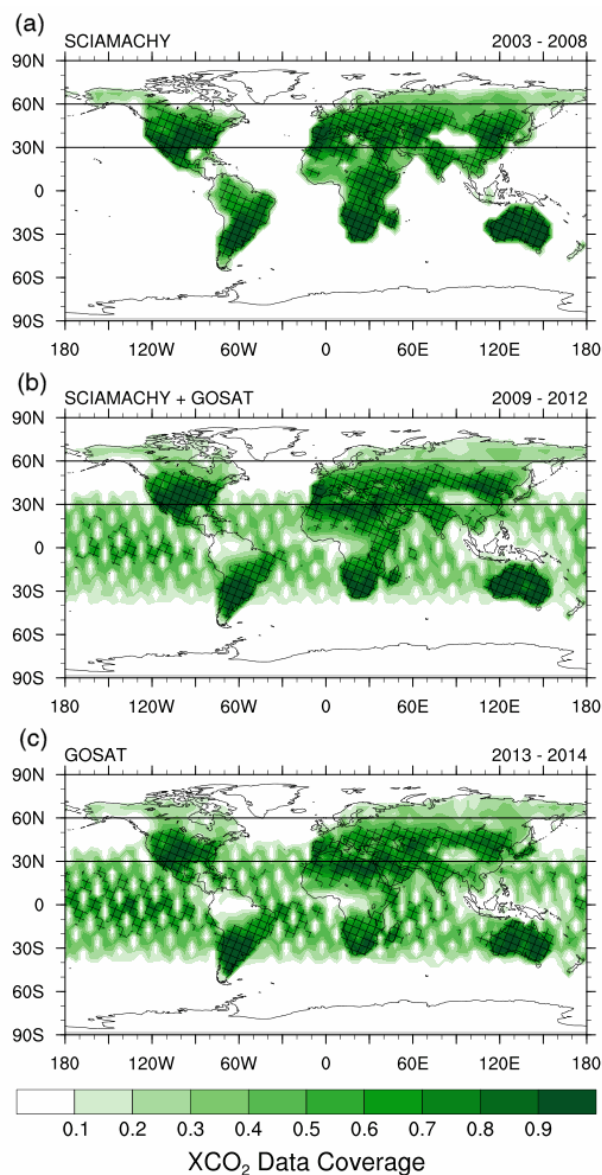
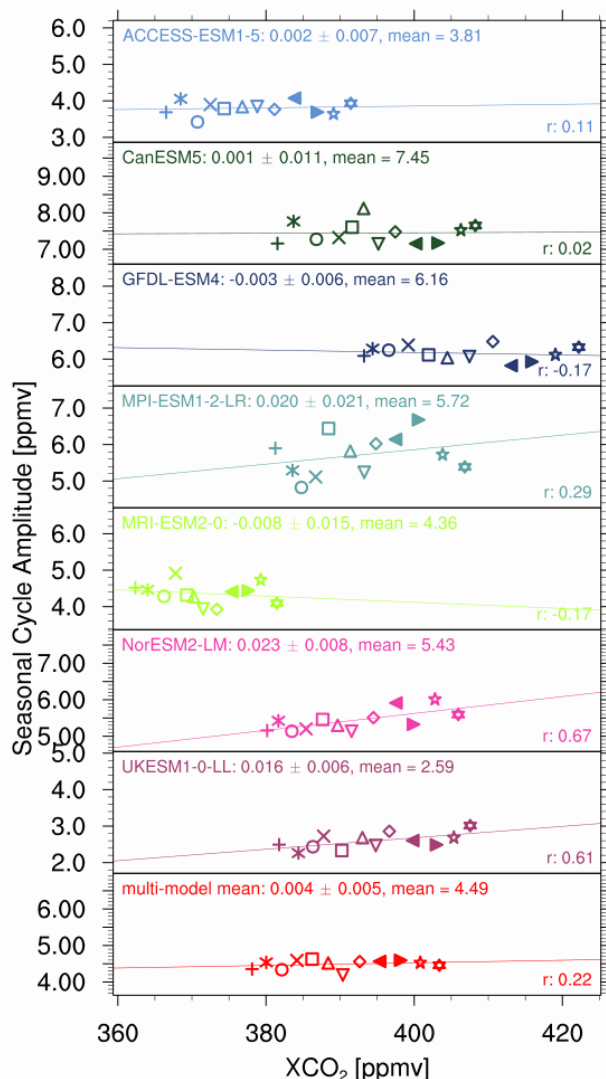


Figure 9: Data Coverage of the satellite observations for (a) 2003–2008, containing only SCIAMACHY data, (b) 2009–2012 containing the overlap of SCIAMACHY and GOSAT data, and (c) 2013–2014 containing only GOSAT data. The patterned area highlights values above 0.5.



(a) SCIAMACHY mask



(b) GOSAT mask

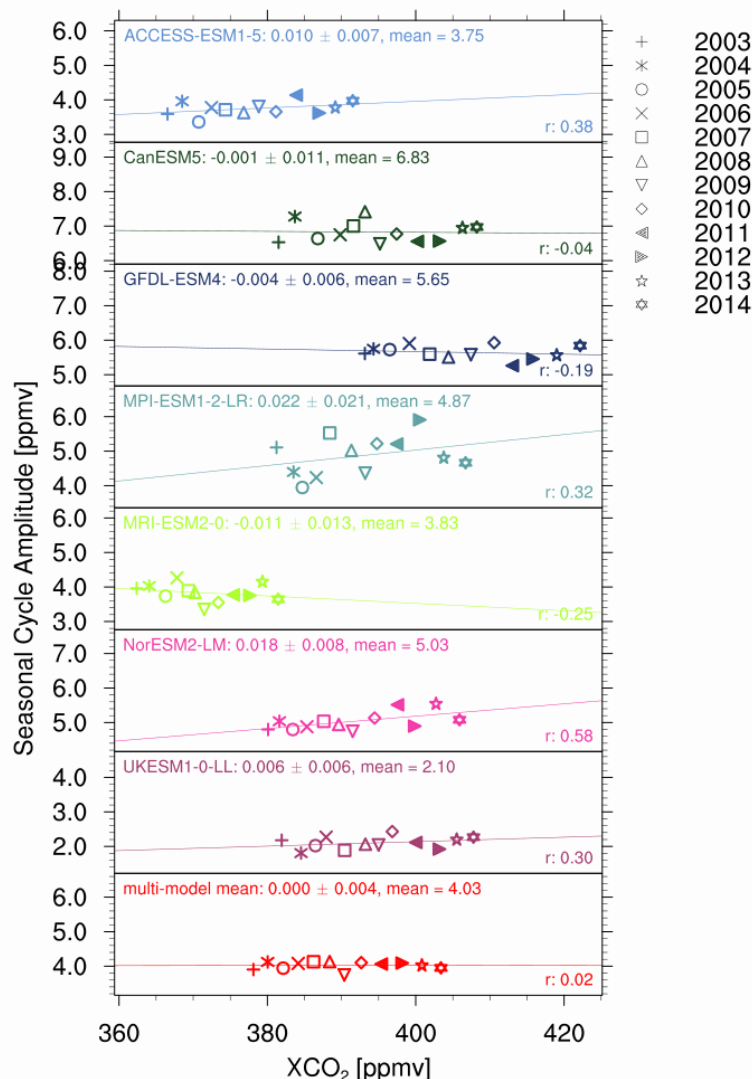


Figure 10: Same as Figure 8, but with CMIP6 models masked using (a) the SCIAMACHY mask and (b) the GOSAT mask, with the masks derived from Figure 9, masking out points with less than 50% coverage in those time periods.

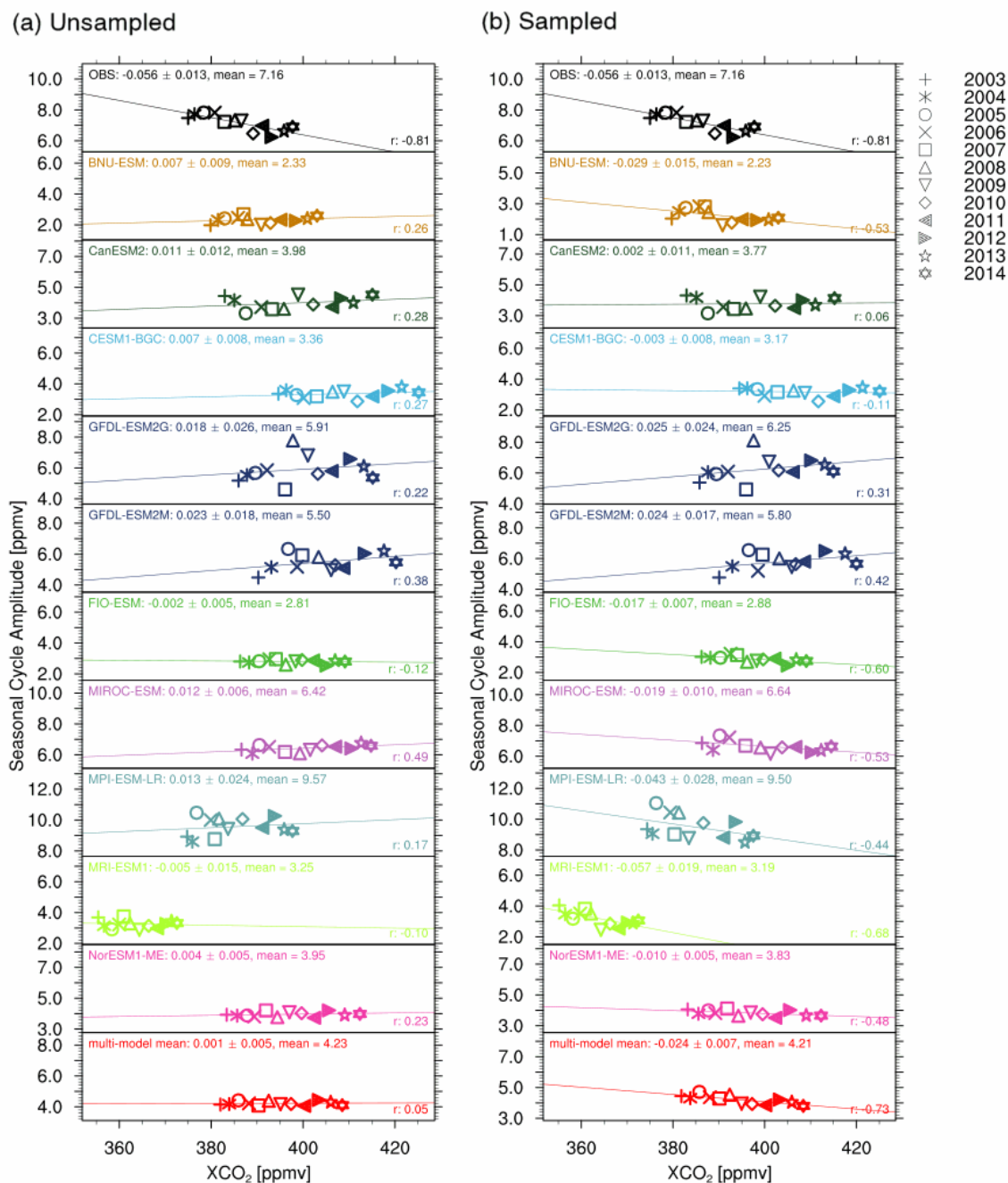


Figure B1: Same as Figure 8 but for CMIP5 models. The left panels (a) show unsampled models, while models sampled according to the satellite data are shown on the right (b). Note that the y-axis range is the same and only differs by a shift.

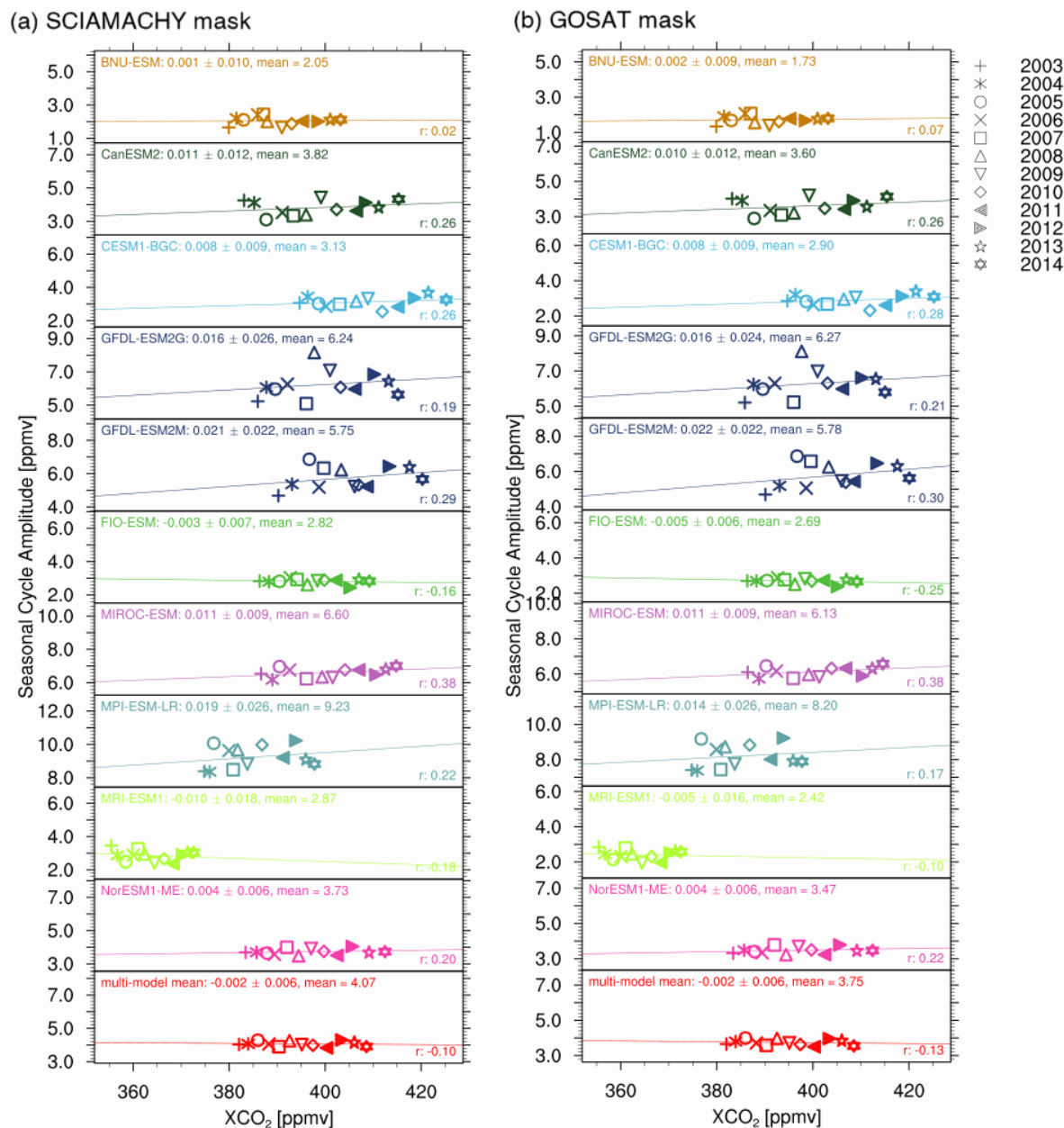


Figure B2: Same as Figure B1, but with CMIP5 models masked using (a) the SCIAMACHY mask and (b) the GOSAT mask, with the masks derived from Figure 9, masking out points with less than 50% coverage in those time periods.