
# Report on the robustness of different weighting schemes for $CO_2$ projections

## *Deliverable 3.5*

**Authors: Manuel Schlund, Lisa Bock, Pauline Bonnet, Birgit Hassler, Fernando Iglesias-Suarez, Rémi Kazeroni, Axel Lauer, Talytha Pereira Barbosa, Breixo Solino-Fernandez and Veronika Eyring**

This project received funding from the Horizon 2020 programme under the grant agreement No. 821003.

# Document Information

| | |
|---|---|
| **GRANT AGREEMENT** | 821003 |
| **PROJECT TITLE** | Climate Carbon Interactions in the Current Century |
| **PROJECT ACRONYM** | 4C |
| **PROJECT START** | 1/6/2019 |
| **RELATED WORK PACKAGE** | WP3 |
| **RELATED TASK(S)** | T3.3 |
| **LEAD ORGANIZATION** | DLR |
| **AUTHORS** | Manuel Schlund, Lisa Bock, Pauline Bonnet, Birgit Hassler, Fernando Iglesias-Suarez, Rémi Kazeroni, Axel Lauer, Talytha Pereira Barbosa, Breixo Solino-Fernandez and Veronika Eyring |
| **SUBMISSION DATE** | 24.03.2023 |
| **DISSEMINATION LEVEL** | PU |

# History

| DATE | SUBMITTED BY | REVIEWED BY | VISION (NOTES) |
|---|---|---|---|
| 24.03.2023 | Manuel Schlund | | |
| 02.05.2023 | | Kerry Hope | |
| | | | |

**Please cite this report as:** Schlund, M., Bock, L., Bonnet, P., Hassler, B., Iglesias-Suarez, F., Kazeroni, R., Lauer, A., Pereira Barbosa, T., Solino-Fernandez, B., & Eyring, V. (2023), Report on the robustness of different weighting schemes for $CO_2$ projections, D3.5 of the 4C project.

# Table of Contents



# List of tables



# List of figures

# About 4C

**Climate-Carbon Interactions in the Coming Century** (4C) is an EU-funded H2020 project that addresses the crucial knowledge gap in the climate sensitivity to carbon dioxide emissions, by reducing the uncertainty in our quantitative understanding of carbon-climate interactions and feedbacks. This will be achieved through innovative integration of models and observations, providing new constraints on modelled carbon-climate interactions and climate projections, and supporting Intergovernmental Panel on Climate Change (IPCC) assessments and policy objectives.

# Executive Summary

This report summarizes the analysis of three different climate model weighting methods used for reducing uncertainties in projected $CO_2$ fluxes in Earth System Models (ESMs) within 4C. Here, we specifically focus on gross primary production (GPP), which is the largest flux of the terrestrial carbon uptake and mainly determined by photosynthesis. The weighting methods analyzed here include a performance- and interdependence-based scheme (Knutti et al. 2017), the multiple diagnostics ensemble regression (MDER; Karpechko et al. 2013), and a machine learning (ML)-based weighting approach (Schlund et al. 2020). All three methods have been successfully used to constrain a variety of target variables in the past and have been made publicly available as open-source-developed code in the Earth System Model Evaluation Tool (ESMValTool) as part of 4C. We find that the resulting weighted GPP is robust across all three weighting methods in terms of the globally aggregated result, as well as in the spatial distribution. This corroborates the credibility and plausibility of the different weighting approaches.

## Keywords

Climate projections, climate model weighting, carbon cycle, land photosynthesis, gross primary production, ESMValTool.

# 1 Introduction

A major source of uncertainty in future climate projections of Earth System Model (ESM) ensembles like the Coupled Model Intercomparison Project Phase 6 (CMIP6; Eyring et al. 2016) are uncertainties in carbon cycle feedbacks and $CO_2$ fluxes between the atmosphere, land and ocean (Arora et al. 2020). The $CO_2$ fluxes play a crucial role in determining the amount of $CO_2$ emissions that remain in the atmosphere as greenhouse gases, hence precise projections of these fluxes are necessary for accurate assessments of policy-relevant metrics like the TCRE (transient climate response to cumulative carbon emissions) and remaining carbon budgets for achieving specific warming targets. The largest flux of the terrestrial carbon uptake is gross primary production (GPP), which is defined as the production of carbohydrates by photosynthesis. In the future, elevated atmospheric $CO_2$ concentration is expected to increase GPP through the $CO_2$ fertilization effect, which forms a negative feedback with global warming as additional $CO_2$ is removed from the atmosphere (e.g., Friedlingstein et al. 2006, Walker et al. 2020).

Task 3.3 aims to increase the accuracy of $CO_2$ flux projections of large ESM ensembles by developing and applying a new climate model weighting method and evaluating its robustness and performance against unweighted ensembles and other established weighting schemes. In this report, we compare a machine learning (ML)–based weighting method developed within the 4C project to two other weighting schemes, all of them applied to 21$^{st}$ century GPP projections of CMIP5 simulations.

# 2 Methods

## 2.1 Target variable and diagnostics

The target variable used for all weighting schemes is the same as used by Schlund et al. (2020) in the ML-based weighting approach: the rescaled future GPP in the CMIP5 Representative Concentration Pathway 8.5 (RCP 8.5) scenario (emission-driven simulation). Here, “rescaled” refers to a correction of the global mean GPP by an emergent constraint on the $CO_2$ fertilization effect that accounts for the ESMs’ biases in the response of future GPP to rising atmospheric $CO_2$ concentrations (Wenzel et al. 2016).

As diagnostics for the weighting schemes, we use process-oriented diagnostics which are known to be physically relevant for the simulation of GPP. These diagnostics are listed in Table 1. All diagnostics are inferred from emission-driven CMIP5 simulations of the recent past (*esmHistorical*), but also from observations of the real Earth system. These observation-based products are ultimately used to constrain the target variable.

**Table 1. Process-oriented diagnostics used in the weighting schemes to constrain the target variable.**

| NAME | DESCRIPTION | OBSERVATION-BASED DATASET | USED TIME RANGE | PHYSICAL CONNECTION TO GPP |
|---|---|---|---|---|
| GPP | Gross primary production | FLUXNET-MTE (Jung et al. 2011) | 1991–2000 | - |
| LAI | Leaf are index | LAI3g (Zhu et al. 2013) | 1982–2005 | The leaf area index is a measure for the number of leaves in a grid cell. The photosynthesis rate is highly dependent on the number of leaves (and vegetation in general). |
| PR | Precipitation | CRU (Harris et al. 2014) | 1901–2005 | Water is essential for the chemical processes of photosynthesis. |
| RSDS | Downwelling solar radiation at surface | ERA-Interim (Dee et al. 2011) | 1979–2005 | Solar radiation is essential for the chemical processes of photosynthesis |
| T | Near-surface air temperature | CRU (Harris et al. 2014) | 1901–2005 | Near-surface air temperature and photosynthesis rate have a common driver (incoming solar radiation). |

## 2.2 Climate model weighting schemes

An overview of the weighting schemes analyzed in this report is given in Table 2. All of them are publicly available in the most recent version of the Earth System Model Evaluation Tool (ESMValTool; Righi et al. 2020, Eyring et al. 2020, Lauer et al. 2020, Weigel et al. 2021, Schlund et al. 2023).

**Table 2. Overview of the weighting schemes used in this report.**

| WEIGHTING SCHEME | REFERENCE |
|---|---|
| Performance- and interdependence-based weighting | Knutti et al. 2017 |
| Multiple diagnostic ensemble regression (MDER) | Karpechko et al. 2013 |
| ML-based weighting scheme | Schlund et al. 2020 |

### 2.2.1 Climate model weighting based on performance and interdependence

To improve the reliability of projections from large ESM ensembles, one approach is to weight the models based on their skill and interdependence (Knutti et al. 2017). This method assumes that the accuracy of multi-model projections can be improved by assigning larger weights to models that perform better compared to observational data. Similarly, models that are more similar to each other (i.e., are more dependent) receive lower weights, as these models provide redundant information. All in all, this weighting approach aims to reduce the spread of multi-model projections and produce a more robust ensemble average, which can provide more reliable and robust projections of future climate change.

### 2.2.2 Multiple diagnostic ensemble regression (MDER)

Multiple diagnostic ensemble regression (MDER) is a regression algorithm that uses iterative steps to select a subset of pre-selected diagnostic variables as input predictors to construct a multivariate linear regression model that optimally predicts a future target variable (Karpechko et al. 2013). In contrast to the other weighting schemes presented in this report, this method only works with scalar predictors, meaning each climate model is represented by a single value. MDER uses observational data for the selected diagnostics to create a multi-diagnostic constraint and calculates weights for each model based on this constraint, which are then used to get a weighted projection of the target variable with lower uncertainties.

### 2.2.3 Machine learning (ML)–based weighting scheme

Within the 4C project, Schlund et al. (2020) expanded on the MDER approach by incorporating gridded data and utilizing a non-linear machine learning regression model. This involves inputting gridded observational data for the diagnostic/predictor variables into a machine learning algorithm that has been trained on climate model data to understand the relationship between current physically relevant predictors and a future climate target variable (as shown in Figures 1a and 1b). The approach was thoroughly tested through a leave-one-model-out cross-validation setup (as shown in Figure 1c). The resulting weighting scheme can be viewed as an implicit performance weighting that takes into account local characteristics and different significance levels for the various process-based predictors.

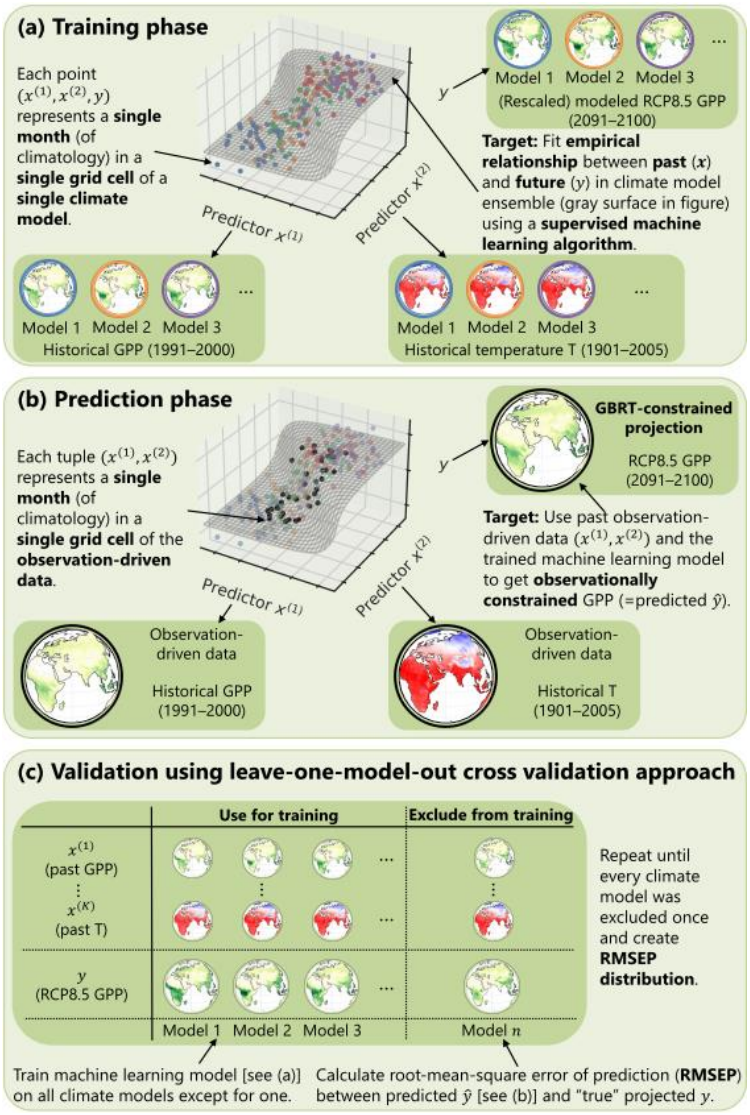


**Figure 1. Graphical representation of the ML-based climate model weighting approach by Schlund et al. (2020). Panel (a) shows the training phase of the algorithm, where the ML model is fitted to the training data by finding an empirical (non-linear) relationship between two (in general, an arbitrary number of predictors is supported) process-oriented predictors ($x_1$, $x_2$) of the past climate and a target variable y (gray surface). The points on the graph represent grid cells from different climate models and are used as training data by the supervised ML algorithm (different colors correspond to different climate models). Panel (b) depicts the prediction phase, where the ML model is applied to observation-based values of the predictors (black points), resulting in ML-based predictions of the target variable for each grid cell. Panel (c) shows how the method is independently validated using a leave-one-model-out cross-validation approach. This approach involves removing one climate model at a time from the training data and evaluating the ML model's performance on the remaining data. From Schlund et al. (2020).**

# 3 Results

## 3.1 Climate model weighting based on performance and interdependence

The performance- and interdependence-based weighting scheme by Knutti et al. (2017) takes into account the performance of the climate models (measured relative to observation-based products) and their interdependence (measured relative to other climate models). We use all diagnostics shown in Table 1 to calculate the performance and intercedence weights for the target variable. Figure 2 shows the weighted multi-model mean and standard prediction error of GPP averaged over 2091–2100 in the RCP 8.5 scenario. This corresponds to a globally aggregated GPP of 180 GtC $yr^{-1}$.

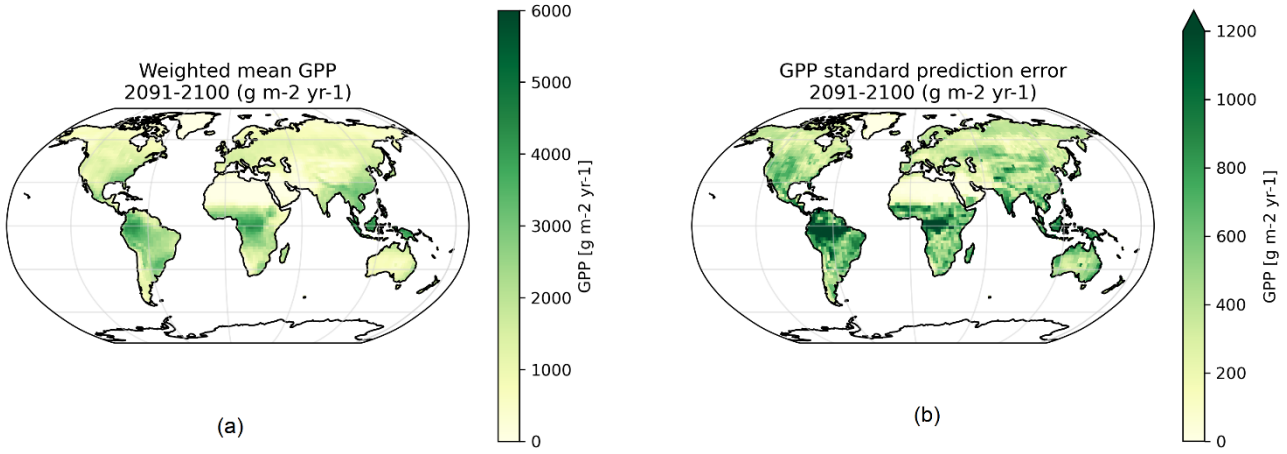


**Figure 2. Weighted multi-model mean (a) and standard prediction error (b) of the target variable GPP (2091–2100) using the performance- and interdependence-based weighting approach.**

## 3.2 Multiple diagnostic ensemble regression (MDER)

In contrast to the other two weighting approaches presented in this report, MDER only considers globally aggregated values. Thus, as a first step, all diagnostics from Table 1 and the target variable are globally summed (GPP) or averaged (other variables) before entering the MDER algorithm. In the first step of this weighting scheme, diagnostics are selected by their correlation relative to the target variable. Due to the strong relationship between past and future GPP, the historical GPP remains the only predictor for the MDER model. In the second step, a linear model is built that relates the predictor and the target variable, and eventually used to create predictions for the target variables using observations of the predictor (see Figure 3a). MDER predicts a global GPP of (170 ± 26) GtC $yr^{-1}$ in the period 2091–2100 (see Figure 3b).

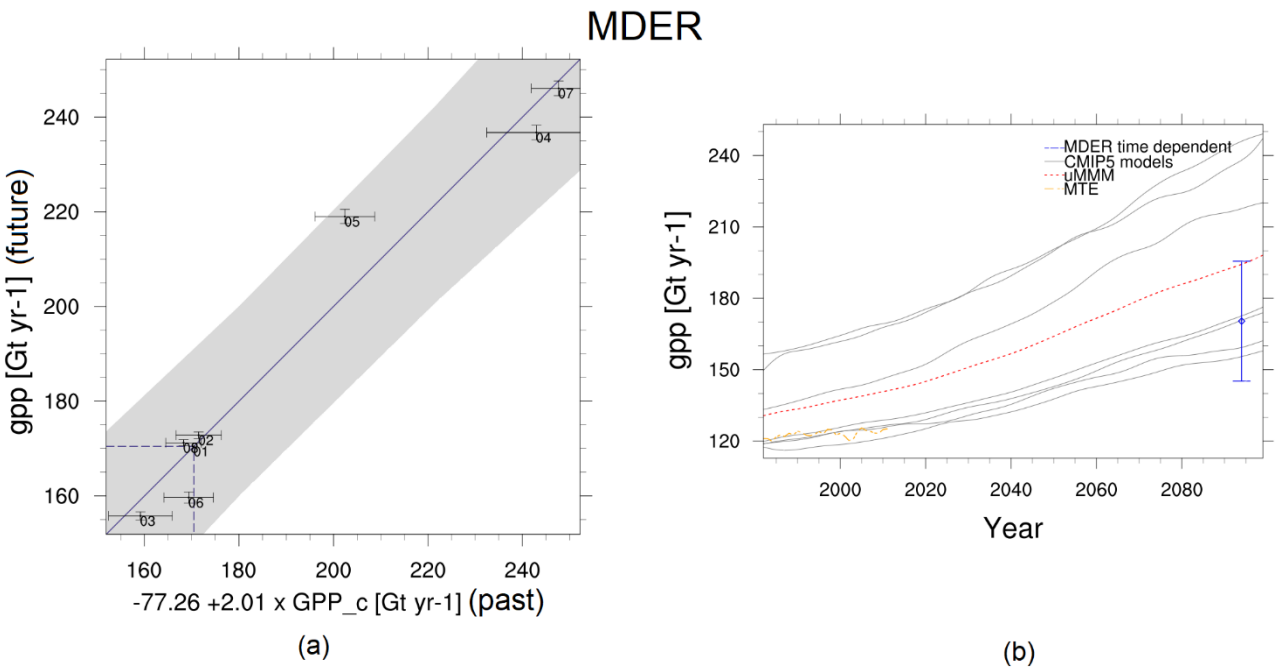


**Figure 3. Result of the MDER weighting. (a) Linear regression to predict future GPP from past GPP using a relationship (blue line) of the two. Numbers correspond to different CMIP5 models, the gray shaded area show the uncertainty of the linear regression, and the dashed lines the observation and the corresponding prediction. (b) Time evolution of unweighted GPP of the different models (gray lines), multi-model mean (red line), and observations (yellow line). The blue marker shows the weighted mean with uncertainties.**

## 3.3 Machine learning (ML)–based weighting scheme

Similar to the MDER approach, the ML-based weighting approach also builds a statistical model between the diagnostics and the target variable. However, in contrast to MDER it uses all diagnostics (the importance of the individual predictors is determined implicitly) and applies a non-linear regression (here: Gradient Boosted Regression Tree algorithm; GBRT). The resulting mean prediction field and its corresponding standard prediction error are shown in Figure 4. This corresponds to a globally aggregated GPP of 169 GtC $yr^{-1}$.

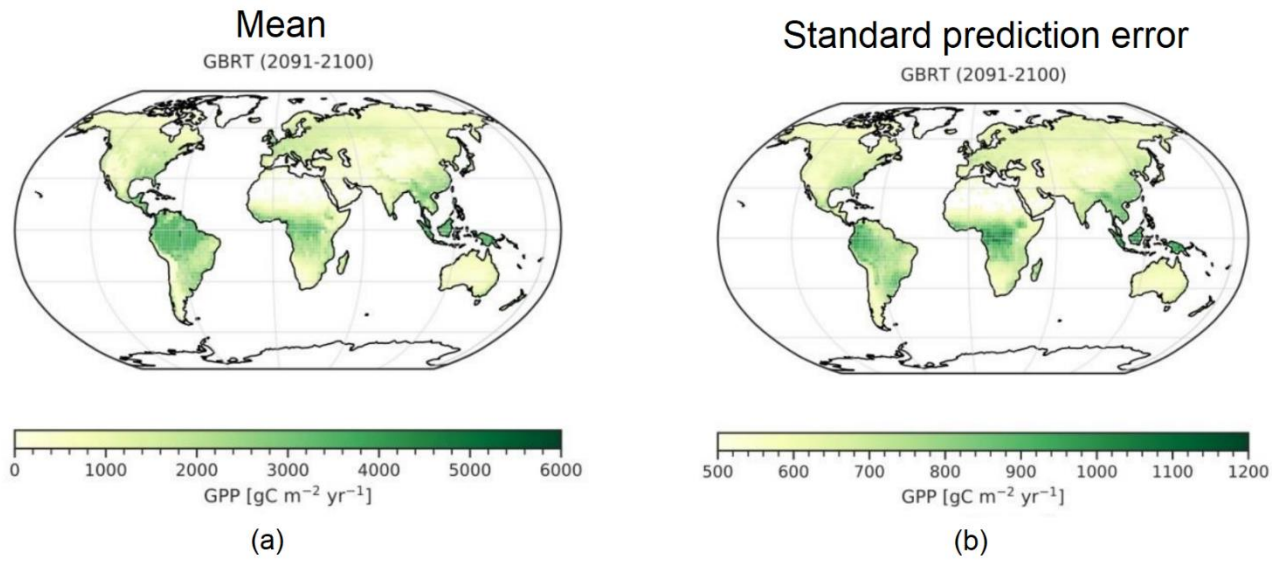


**Figure 4. ML-based prediction of the mean (a) and standard prediction error of the target variable GPP (2091–2100) using a Gradient Boosted Regression Tree (GBRT) algorithm.**

# 4 Discussion and conclusions

All weighting approaches presented in this report could be successfully applied to multi-model projections of the rescaled future GPP using ESMValTool. In terms of globally aggregated GPP, all approaches give yield similar results for the period 2091–2100 in the RCP 8.5 scenario: performance/interdependence weighting: 180 GtC $yr^{-1}$, MDER: 170 GtC $yr^{-1}$, and ML weighting: 169 GtC $yr^{-1}$. These results are all consistent with the emergent constraint on $CO_2$ fertilization that has been used to rescale the target variable in the first place (see Schlund et al. 2020 for details), which gives a globally aggregated GPP of (171 ± 12) GtC $yr^{-1}$ for the period 2091–2100 in the RCP 8.5 scenario. Thus, the globally aggregated result is robust across all investigated weighting methods, which further corroborates the credibility and plausibility of the different approaches.

In addition, in the performance/interdependence and ML-based weighting schemes, global patterns of future GPP can be analyzed (see Figure 2 and Figure 4). For both approaches, the mean predictions and corresponding standard prediction errors are very similar, with pattern correlations of r=0.81 (mean prediction) and r=0.56 (standard prediction error). Thus, the exact results of the two weighting schemes are slightly different, but the general patterns across the globe are mostly the same, especially for the mean prediction, which further underlines the robustness of the results in terms of the different weighting methods.

# 5 References


Arora, V. K., Katavouta, A., Williams, R. G., Jones, C. D., Brovkin, V., Friedlingstein, P., ... & Ziehn, T.: Carbon–concentration and carbon–climate feedbacks in CMIP6 models and their comparison to CMIP5 models. Biogeosciences, 17(16), 4173-4222, https://doi.org/10.5194/bg-17-4173-2020, 2020.

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., et al.: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. Quarterly Journal of the Royal Meteorological Society, 137(656), 553–597. https://doi.org/10.1002/qj.828, 2011.

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, Geoscientific Model Development, 9, 1937–1958, https://doi.org/10.5194/gmd-9-1937-2016, 2016.

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. L., 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., Hunter, A., Kadow, C., Kindermann, S., Koirala, S., Koldunov, N., 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.: Earth System Model Evaluation Tool (ESMValTool) v2.0 – an extended set of large-scale diagnostics for quasi-operational and comprehensive evaluation of Earth system models in CMIP, Geoscientific Model Development, 13, 3383–3438, https://doi.org/10.5194/gmd-13-3383-2020, 2020.

Friedlingstein, P., Cox, P., Betts, R., Bopp, L., Von Bloh, W., Brovkin, V., et al.: Climate-carbon cycle feedback analysis: Results from the (CMIP)-M-4 model intercomparison. Journal of Climate, 19(14), 3337–3353. https://doi.org/10.1175/jcli3800.1, 2006.

Harris, I., Jones, P. D., Osborn, T. J., & Lister, D. H.: Updated high-resolution grids of monthly climatic observations—The CRU TS3.10 dataset. International Journal of Climatology, 34(3), 623–642. https://doi.org/10.1002/joc.3711, 2014.

Jung, M., Reichstein, M., Margolis, H. A., Cescatti, A., Richardson, A. D., Arain, M. A., et al.: Global patterns of land-atmosphere fluxes of carbon dioxide, latent heat, and sensible heat derived from eddy covariance, satellite, and meteorological observations. Journal of Geophysical Research, 116, G00J07. https://doi.org/10.1029/2010JG001566, 2011.

Karpechko, A. Y., Maraun, D., & Eyring, V.: Improving Antarctic total ozone projections by a process-oriented multiple diagnostic ensemble regression. Journal of the atmospheric sciences, 70(12), 3959-3976, https://doi.org/10.1175/JAS-D-13-071.1, 2013.

Knutti, R., Sedláček, J., Sanderson, B. M., Lorenz, R., Fischer, E. M., & Eyring, V.: A climate model projection weighting scheme accounting for performance and interdependence. Geophysical Research Letters, 44(4), 1909-1918, https://doi.org/10.1002/2016GL072012, 2017.

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, Geoscientific Model Development, 13, 4205–4228, https://doi.org/10.5194/gmd-13-4205-2020, 2020.

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, Geoscientific Model Development, 13, 1179–1199, https://doi.org/10.5194/gmd-13-1179-2020, 2020.

Schlund, M., Eyring, V., Camps-Valls, G., Friedlingstein, P., Gentine, P., & Reichstein, M.: Constraining uncertainty in projected gross primary production with machine learning. Journal of Geophysical Research: Biogeosciences, 125(11), e2019JG005619, https://doi.org/10.1029/2019JG005619, 2020.

Schlund, M., Hassler, B., Lauer, A., Andela, B., Jöckel, P., Kazeroni, R., Loosveldt Tomas, S., Medeiros, B., Predoi, V., Sénési, S., Servonnat, J., Stacke, T., Vegas-Regidor, J., Zimmermann, K., and Eyring, V.: Evaluation of native Earth system model output with ESMValTool v2.6.0, Geosci. Model Dev., 16, 315–333, https://doi.org/10.5194/gmd-16-315-2023, 2023.

Walker, A. P., De Kauwe, M. G., Bastos, A., Belmecheri, S., Georgiou, K., Keeling, R., et al.: Integrating the evidence for a terrestrial carbon sink caused by increasing atmospheric CO2. New Phytologist, 1. https://doi.org/10.1111/nph.16866, 2020.

Weigel, K., Bock, L., Gier, B. K., Lauer, A., Righi, M., Schlund, M., Adeniyi, K., Andela, B., Arnone, E., Berg, P., Caron, L.-P., Cionni, I., Corti, S., Drost, N., Hunter, A., Lledó, L., Mohr, C. W., Paçal, A., Pérez-Zanón, N., Predoi, V., Sandstad, M., Sillmann, J., Sterl, A., Vegas-Regidor, J., von Hardenberg, J., and Eyring, V.: Earth System Model Evaluation Tool (ESMValTool) v2.0 – diagnostics for extreme events, regional and impact evaluation, and analysis of Earth system models in CMIP, Geoscientific Model Development, 14, 3159–3184, https://doi.org/10.5194/gmd-14-3159-2021, 2021.

Wenzel, S., Cox, P. M., Eyring, V., & Friedlingstein, P.: Projected land photosynthesis constrained by changes in the seasonal cycle of atmospheric CO2. Nature, 538(7626), 499. https://doi.org/10.1038/nature19772, 2016.

Zhu, Z. C., Bi, J., Pan, Y. Z., Ganguly, S., Anav, A., Xu, L., et al.: Global data sets of vegetation leaf area index (LAI)3g and fraction of photosynthetically active radiation (FPAR)3g derived from Global Inventory Modeling and Mapping Studies (GIMMS) normalized difference vegetation index (NDVI3g) for the period 1981 to 2011. Remote Sensing, 5(2), 927–948. https://doi.org/10.3390/rs5020927, 2013.