Inclusion of new benchmarking in ESMValTool

Deliverable 1.10

Authors: Katja Weigel, Manuel Schlund, Bettina K. Gier, 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 DATE	1/6/2019
RELATED WORK PACKAGE	WP1
RELATED TASK(S)	T1.4.3
LEAD ORGANIZATION	UBREMEN
AUTHORS	Katja Weigel, Manuel Schlund, Bettina K. Gier, and Veronika Eyring
SUBMISSION DATE	21.11.2022
DISSEMINATION LEVEL	PU

History

DATE	SUBMITTED BY	REVIEWED BY	VISION (NOTES)
21.11.2022	Katja Weigel		
22.11.2022		P. Friedlingstein	

Please cite this report as: Weigel, K, Gier, B. K., Schlund, M., & Eyring, V., (2022), Inclusion of new benchmarking in ESMValToo, D1.10 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

The Earth System Model Evaluation Tool (ESMValTool) is a community diagnostics and performance metrics tool for the routine evaluation of Earth system models (ESMs). Here, we present extensions to ESMValTool that improve its benchmarking abilities, i.e., the assessment of the performance of ESMs with observation-based products. These extensions include the addition of new observation-based data sets (partly developed within the 4C project), in particular the CDS-XCO2, LandFlux-EVAL, Landschuetzer2016, Landschuetzer2020, MOBO-DIC_MPIM, and OceanSODA-ETHZ products. Moreover, new diagnostics have been added, which include general-purpose diagnostics useful for a broad evaluation of ESM output, diagnostics for the evaluation of the terrestrial carbon cycle, and diagnostics for the evaluation of the column-averaged mole fraction of CO_2. All these extensions allow for a more detailed and in-depth analysis of ESM output, which can be used to improve the models, and ultimately leads to a better representation of the Earth system and to more accurate projections of the future climate change.

Keywords

Land carbon-cycle, Ocean carbon-cycle, Earth system model evaluation, ESMValTool.

1 Introduction

Earth system models (ESMs) are state-of-the-art tools used to improve our understanding of mechanisms and feedbacks in present-day climate but also to project climate change for different future scenarios. Modern climate models have come a long way starting from simple atmosphere-only models some decades ago to today's complex ESMs participating in the latest (sixth) phase of the Coupled Model Intercomparison Project (CMIP6; Eyring et al. 2016). Constant improvement of the models was and is needed to represent key processes of the Earth system that affect climate change. This increasing complexity, however, is also a possible driver for an increase in the inter-model spread of climate projections within the multi-model ensemble. Thus, more than ever, these developments require innovative and comprehensive model evaluation tools to assess the performance of these increasingly complex models. One of these software tools is the Earth System Model Evaluation Tool (ESMValTool; Righi et al. 2020, Eyring et al. 2020, Lauer et al. 2020, Weigel et al. 2021), a community diagnostics and performance metrics tool for the evaluation of ESMs that allows for routine comparison of single or multiple models, either against predecessor versions or against observations.

Here, we present extensions to ESMValTool that have been implemented as part of the 4C project to improve its benchmarking capabilities (i.e., the evaluation of ESMs with observation-based data). These extensions include the addition of observation-based data sets used and/or developed within the 4C project, new general-purpose diagnostics useful to get a broad overview of ESM output, diagnostics for a general evaluation of the terrestrial carbon cycle, and diagnostics for the evaluation of ESMs using a satellite-based atmospheric CO_2 data set.

2 Inclusion of observation-based products into ESMValTool

To ensure a fair and meaningful comparison of different ESMs, the input data for ESMValTool has to be formatted according to common standards, the so-called CMOR (Climate Model Output Rewriter) tables and definitions. This process is usually referred to as “CMORization”. This does not only apply to ESM output, but also to observation-based products. For this reason, ESMValTool conveniently provides scripts for many observation-based products that can be used to CMORize the data once before it can be used within ESMValTool. To improve the carbon cycle evaluation with ESMValTool, many new CMORizer scripts for observation-based products (and thus the data itself) have been made available as part of the 4C project. Table 1 provides an overview of these new data sets.

Table 1. Observation-based products used and/or developed within the 4C projects that have been made available to ESMValTool through CMORization scripts. Variable names are described in Table 2.

DATA SET NAME	4C TASK	VARIABLES	TIME PERIOD	REFERENCE	FIGURE IN THIS REPORT
CDS-XCO2	T1.1.3	xco2	2003–2017	Reuter et al. (2020)	Figure 5
LandFlux-EVAL	T1.2.3	et, etStderr	1989–2005	Mueller et al. (2013)	-
Landschuetzer2016	T1.2.1	dpco2, fgco2, spco2	1982–2015	Landschützer et al. (2016)	-
Landschuetzer2020	T1.4.1	spco2	1988–2019 (monthly climatology)	Landschützer et al. (2020)	Figure 1
MOBO-DIC_MPIM	T1.2.2	dissic	2004–2017 (monthly climatology)	Keppler et al. (2020)	Figure 2
OceanSODA-ETHZ	T1.2.1	areacello, co3os, dissicos, fgco2, phos, spco2, talkos	1982–2020	Gregor et al. (2021)	Figure 3

Table 2. Variables provided by the observation-based products listed in Table 1.

VARIABLE	DESCRIPTION	UNITS	DIMENSIONS
areacello	Grid-cell area for ocean variables	m^2	latitude, longitude
co3os	Surface carbonate ion concentration	mol m^{-3}	time, latitude, longitude
dissic	Dissolved inorganic carbon concentration	mol m^{-3}	time, depth, latitude, longitude

VARIABLE	DESCRIPTION	UNITS	DIMENSIONS
dissicos	Surface dissolved inorganic carbon concentration	mol m^{-3}	time, latitude, longitude
dpco2	Delta CO_2 partial pressure	Pa	time, latitude, longitude
et, Stderr	Evapotranspiration (error)	mm day^{-1}	time, latitude, longitude
fgco2	Surface downward mass flux of carbon as CO_2	kgC m^{-2} s^{-1}	time, latitude, longitude
phos	Surface pH	1	time, latitude, longitude
spco2	Surface aqueous partial pressure of CO_2	Pa	time, latitude, longitude
talkos	Surface total alkalinity	mol m^{-3}	time, latitude, longitude
xco2	Column-average dry-air mole fraction of atmospheric CO_2	1	time, latitude, longitude

3 General-purpose diagnostics

As part of an initiative to make ESMValTool more user-friendly and versatile, a set of general-purpose diagnostics have been implemented (Schlund et al., in review). These diagnostics are able to handle arbitrary variables from arbitrary data sets, which makes them flexible to use. Moreover, they are highly customizable, and new plot types can be easily added. Example use cases for these diagnostics are the monitoring of running climate simulations (i.e., to get a quick overview of simulation results), comparison of different versions of a climate model, and the assessment of the performance of ESMs with regards to observational data.

In the following, we show example plots created with the monitoring diagnostics where we compare model output with observational products from Table 1. Instead of individual ESMs, we show the multi-model mean (MMM) of the CMIP6 ensemble (one ensemble member [r1i1p1f1] for each model that provides the necessary data).

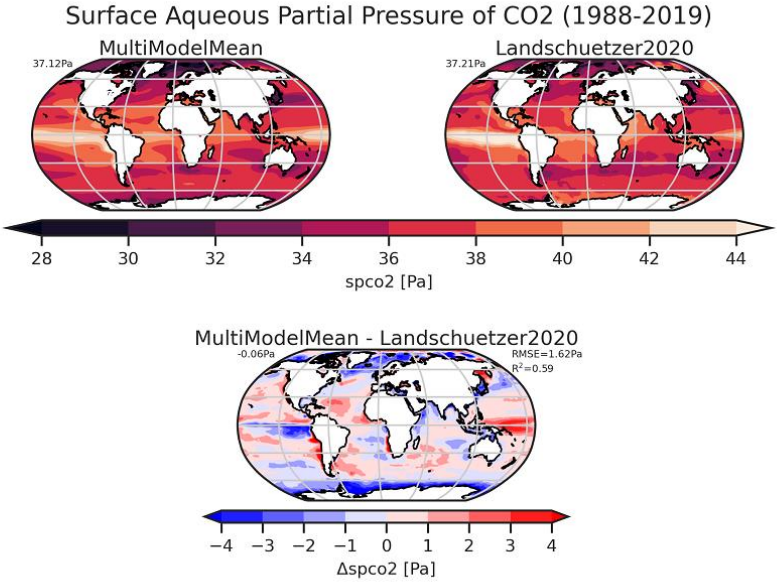

Figure 1. Climatology of surface aqueous partial pressure of CO_2 (spco2) for the CMIP6 multi-model mean (top left) and the Landschuetzer2020 product (top right; Landschützer et al. 2020) averaged over the period 1988–2019. The bottom panel shows the bias between the two. Numbers in the top left corners correspond to the (area-weighted) average of the fields. Numbers in the top right corner of the bias plots correspond to the (area-weighted) root mean square error (RMSE) and the (area-weighted) coefficient of determination (R^2) of the CMIP6 multi-model mean and Landschuetzer2020 fields.

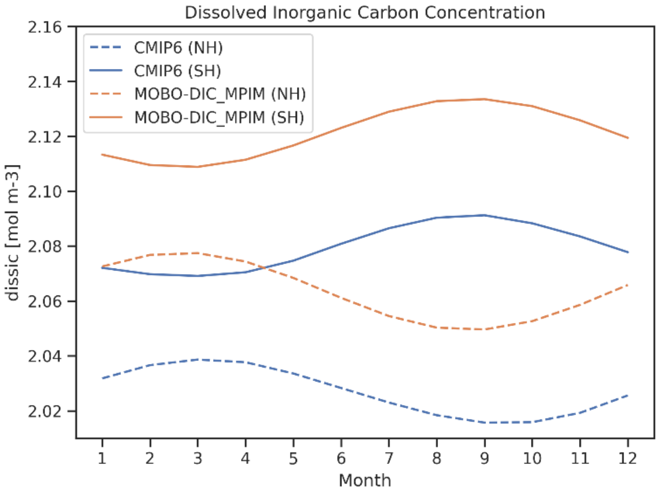

Figure 2. Annual cycle of dissolved inorganic carbon concentration (dissic) measured at a depth of 2.5m for the CMIP6 multi-model mean and the MOBO-DIC_MPIM product (Keppler et al. 2020) averaged over the period 2004–2017. Dashed lines show the northern hemisphere (NH), solid lines the southern hemisphere (SH). The model output has been masked with the MOBO-DIC_MPIM prior to calculating spatial means to get consistent results.

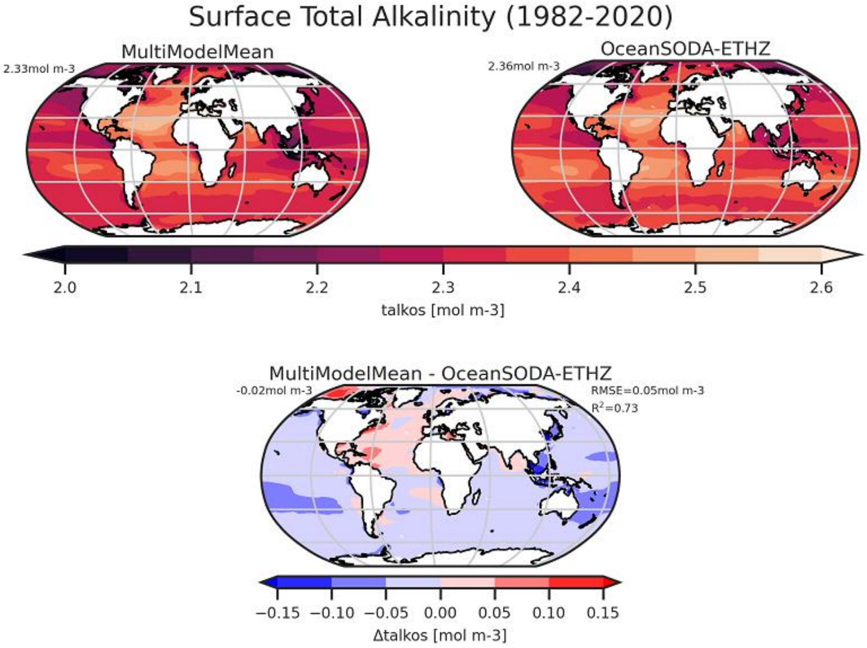

Figure 3. Climatology of surface total alkalinity (talkos) for the CMIP6 multi-model mean (top left) and the OceanSODA-ETHZ product (top right; Gregor et al. 2021) averaged over the period 1982–2020. The bottom panel shows the bias between the two. Numbers in the top left corners correspond to the (area-weighted) average of the fields. Numbers in the top right corner of the bias plots correspond to the (area-weighted) root mean square error (RMSE) and the (area-weighted) coefficient of determination (R^2) of the CMIP6 multi-model mean and OceanSODA-ETHZ fields.

All figures show a good agreement of the MMM with the corresponding observational products. The global mean surface aqueous partial pressure of CO_2 climatology (averaged over 1988–2019; see Figure 1) is slightly underestimated in the models, which is the result of an overestimation in the Atlantic Ocean, Indian Ocean and western Pacific Ocean, and an underestimation in the Arctic Ocean, Southern Ocean and tropical eastern Pacific Ocean. In general, relative errors are smaller than 10% for most parts of the globe. In addition, the geographical patters in the MMM and observational product (Landschuetzer2020) match well, with a root mean square error (RMSE) of 1.62 Pa (< 5% of the observational global mean) and a pattern correlation of $R^2 = 0.59$. Figure 2 shows the annual cycle of the dissolved inorganic carbon concentration averaged over the period 2004–2017 for both hemispheres separately. For both hemispheres, the MMM underestimates the observation-based data set (MOBO-DIC_MPIM) by about 2%. Nevertheless, the phase and the amplitude of the annual cycle match very well. Finally, Figure 3 shows the 1982–2020 climatology of the surface total alkalinity. Similar to both other variables, the global mean surface total alkalinity is also underestimated by the models. This is true for most parts of the oceans, except for the northern Atlantic Ocean and the Arctic Ocean north of Canada. The spatial distribution of the surface total alkalinity matches very well between the MMM and the OceanSODA-

ETHZ observational product, with a low RMSE of 0.05 mol m^{-3} (< 3% of the observed global mean) and a high pattern correlation of $R^2 = 0.73$.

4 Evaluation of the terrestrial carbon cycle

Established carbon cycle benchmarks (Anav et al. 2013) have been implemented into ESMValTool. This includes the evaluation of time series, climatologies, trends, seasonal cycles, and performance metrics of carbon cycle–related variables like gross primary production (GPP), net biome production (NBP), leaf area index (LAI), and atmosphere-ocean CO_2 flux. In a subsequent study, Gier et al. (in prep.) provide an extension of the carbon cycle evaluation on CMIP5 and CMIP6 models that specifically focuses on the differences between ESMs that include an interactive nitrogen cycle with ESMs that do not. Overall, the authors find a slight improvement in the simulation of land carbon cycle parameters when moving from CMIP5 to CMIP6. In addition, the inclusion of nitrogen limitation through an interactive nitrogen cycle in the models leads to a large improvement in the simulation of GPP (see Figure 4). Currently, these diagnostics are under active development, and will be published to the public ESMValTool repository as soon as the study is accepted.

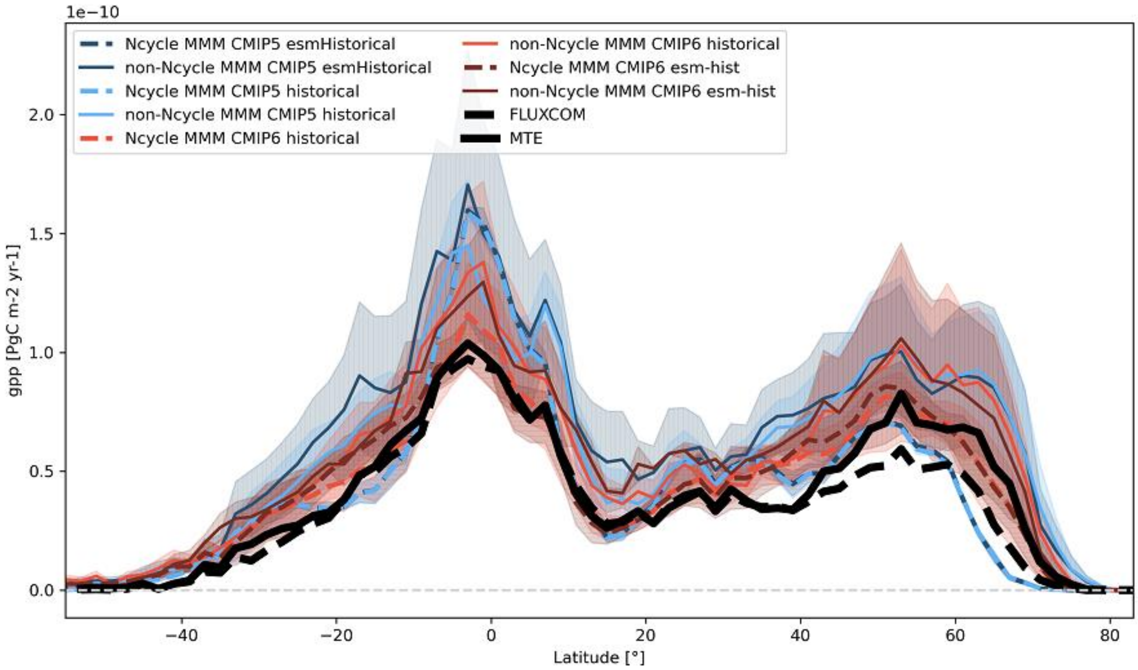

Figure 4. Zonal means of gross primary production (GPP) for the reference data sets FLUXCOM and MTE, and the different multi-model means (MMMs; see legend). Hatching for MMMs shows their standard deviation, with a horizontal hatching for models with and vertical hatching for models without interactive nitrogen cycle.

5 Spatially resolved evaluation of Earth system models with satellite column-averaged CO_2

Gier et al. (2020) provide new diagnostics to evaluate column-averaged dry-air mole fractions of atmospheric CO_2 (XCO2). The evaluation compares the spatially resolved CDS-XCO2 satellite-based data set (Reuter et al. 2020; see Section 2) with CMIP emission-driven simulations, with an emphasis on the impact of the satellite sampling on the results. One of the main findings of the authors is the resolution of the previously believed discrepancy of a strong negative trend in the northern midlatitude (30–60 °N) seasonal cycle amplitude (SCA) of XCO2 with rising XCO2 seen in the satellite observations which is neither seen in the models nor in the in-situ data. The observational data set is composed of a synthesis of data from two different satellites which have different spatial resolutions resulting in different northern midlatitude mean SCAs during their active time. As this difference in mean SCA is larger than the SCA variation in the data, it introduces an artificial negative trend in the combined satellite data timeseries which can be reproduced by the models when they are sampled as the observations (see Figure 5).

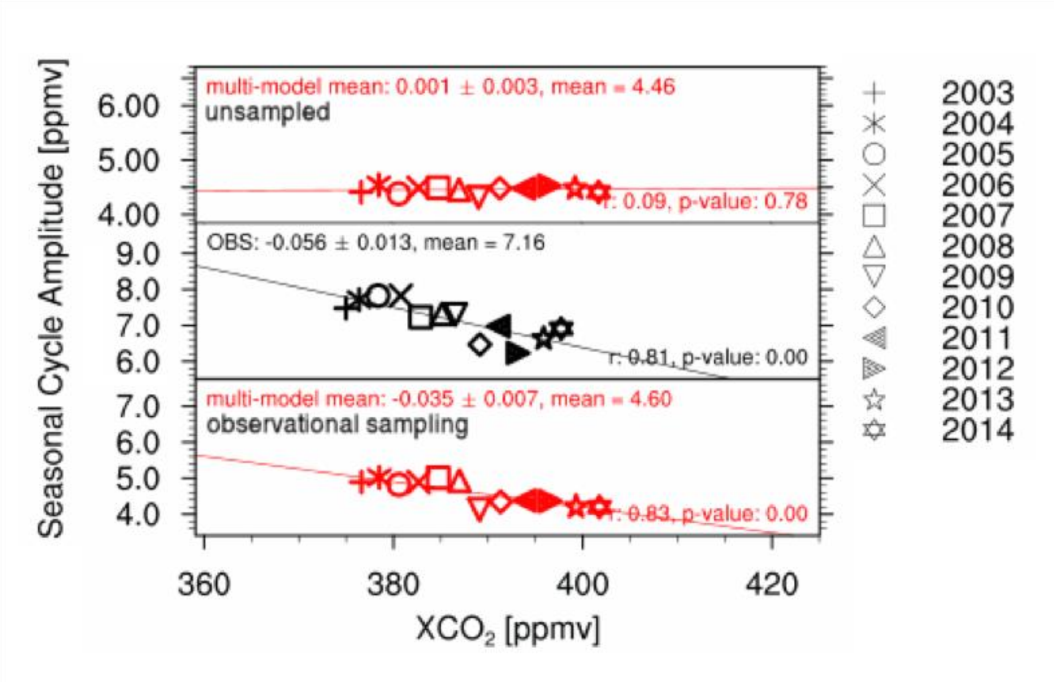

Figure 5. Seasonal cycle amplitude of column-averaged CO_2 (XCO2) with respect to atmospheric XCO2 content. (Top) unsampled models, (middle) observations, (bottom) models sampled as observations. Similar to Figure 7 of Gier et al. (2020).

6 Conclusions and Outlook

ESMValTool is a community-based diagnostic tool for the routine evaluation of ESMs. It is developed open-source by an increasingly large number of developers from many international institutes. Over the last couple of years, more and more diagnostics and functionalities have been implemented, and it has been used as the primary evaluation tool for a large number of publications (e.g., ESMValTool has been used to produce figures for some chapters of the latest Assessment Report 6 of the Intergovernmental Panel on Climate Change IPCC; e.g., Eyring et al. 2021).

In this report, we demonstrate new additions to ESMValTool that have been implemented as part of the 4C project to improve the benchmarking of ESMs. These additions include new observation-based data sets (see Section 2) and new diagnostics (see Sections 3, 4, and 5). The figures and analysis shown in this report only provide examples and by no means represent the full set of available diagnostics. For example, the new observation-based products are now available for every diagnostic in ESMValTool (not only for the general-purpose diagnostics presented in Section 3), which opens up new opportunities for much more complex and in-depth analyses of the ESM output.

7 References

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, Journal of Climate, 26(18), 6801–6843, https://doi.org/10.1175/JCLI-D-12-00417.1, 2013.

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.

Eyring, V., N.P. Gillett, K.M. Achuta Rao, R. Barimalala, M. Barreiro Parrillo, N. Bellouin, C. Cassou, P.J. Durack, Y. Kosaka, S. McGregor, S. Min, O. Morgenstern, and Y. Sun: Human Influence on the Climate System. In Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [Masson-Delmotte, V., P. Zhai, A. Pirani, S.L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M.I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J.B.R. Matthews, T.K. Maycock, T. Waterfield, O. Yelekçi, R. Yu, and B. Zhou (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, pp. 423–552, 2021.

Gier, B. K., Buchwitz, M., Reuter, M., Cox, P. M., Friedlingstein, P., and Eyring, V.: Spatially resolved evaluation of Earth system models with satellite column-averaged CO2, Biogeosciences, 17, 6115–6144, https://doi.org/10.5194/bg-17-6115-2020, 2020.

Gier, B. K., Schlund, M., Arora, V. K., Cox, P. M., Friedlingstein, P., Jones, C. D., Jones, C., Zaehle, S., and Eyring, V.: Evaluation of the Terrestrial Carbon Cycle in CMIP6 historical simulations, in prep., 2022.

Gregor, L. and Gruber, N.: OceanSODA-ETHZ: a global gridded data set of the surface ocean carbonate system for seasonal to decadal studies of ocean acidification, Earth Syst. Sci. Data, 13, 777–808, https://doi.org/10.5194/essd-13-777-2021, 2021.

Keppler, L., Landschützer, P., Gruber, N., Lauvset, S. K., and Stemmler, I.: Seasonal carbon dynamics in the near-global ocean, Global Biogeochemical Cycles, 34, e2020GB006571. https://doi.org/10.1029/2020GB006571, 2020.

Landschützer, P., Gruber, N., and Bakker, D. C. E.: Decadal variations and trends of the global ocean carbon sink, Global Biogeochem. Cycles, 30, 1396–1417, https://doi.org/10.1002/2015GB005359, 2016.

Landschützer, P., Laruelle, G. G., Roobaert, A., and Regnier, P.: A uniform pCO2 climatology combining open and coastal oceans, Earth Syst. Sci. Data, 12, 2537–2553, https://doi.org/10.5194/essd-12-2537-2020, 2020.

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.

Mueller, B., Hirschi, M., Jimenez, C., Ciais, P., Dirmeyer, P. A., Dolman, A. J., Fisher, J. B., Jung, M., Ludwig, F., Maignan, F., Miralles, D. G., McCabe, M. F., Reichstein, M., Sheffield, J., Wang, K., Wood, E. F., Zhang, Y., and Seneviratne, S. I.: Benchmark products for land evapotranspiration: LandFlux-EVAL multi-data set synthesis, Hydrol. Earth Syst. Sci., 17, 3707–3720, https://doi.org/10.5194/hess-17-3707-2013, 2013.

Reuter, M., Buchwitz, M., Schneising, O., Noël, S., Bovensmann, H., Burrows, J. P., Boesch, H., Di Noia, A., Anand, J., Parker, R. J., Somkuti, P., Wu, L., Hasekamp, O. P., Aben, I., Kuze, A., Suto, H., Shiomi, K., Yoshida, Y., Morino, I., Crisp, D., O'Dell, C. W., Notholt, J., Petri, C., Warneke, T., Velazco, V. A., Deutscher, N. M., Griffith, D. W. T., Kivi, R., Pollard, D. F., Hase, F., Sussmann, R., Té, Y. V., Strong, K., Roche, S., Sha, M. K., De Mazière, M., Feist, D. G., Iraci, L. T., Roehl, C. M., Retscher, C., and Schepers, D.: Ensemble-based satellite-derived carbon dioxide and methane column-averaged dry-air mole fraction data sets (2003–2018) for carbon and climate applications, Atmos. Meas. Tech., 13, 789–819, https://doi.org/10.5194/amt-13-789-2020, 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., 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. Discuss. [preprint], https://doi.org/10.5194/gmd-2022-205, in review, 2022.

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.