



Universität  
Bremen

---

**Constraining Uncertainties in Multi-Model  
Projections of the Future Climate with  
Observations**

---

DOCTORAL DISSERTATION OF  
**Manuel SCHLUND**

April 2021



UNIVERSITY OF BREMEN  
INSTITUTE OF ENVIRONMENTAL PHYSICS (IUP)

---

# Constraining Uncertainties in Multi-Model Projections of the Future Climate with Observations

---

DOCTORAL DISSERTATION OF  
**Manuel SCHLUND**

*A thesis submitted in fulfillment of the requirements for the degree  
Doktor der Naturwissenschaften (Dr. rer. nat.)*

Primary Examiner: Prof. Dr. Veronika EYRING  
Secondary Examiner: Prof. Dr. Pierre GENTINE

Submission Date: 13 April 2021



This work is licensed under a [Creative Commons Attribution 4.0 International License](#).

# Abstract

A precise quantification of climate change is crucial to assess optimal mitigation and adaptation strategies. Earth system models (ESMs), which are state-of-the-art climate models that allow numerical simulations of the complex physical, biological, and chemical processes of the Earth system, are common tools to understand and project climate change. Due to the chaotic nature of the climate system, unknowns in future emission pathways, and uncertainties in the climate models, projections of the future climate are associated with large uncertainties. The main focus of this thesis is the analysis of future climate projections from ESMs participating in the Coupled Model Intercomparison Project (CMIP) with the aim to reduce uncertainties in climate projections with observations.

In a first step, climate sensitivity (i.e., the temperature response of the climate system to an external forcing) is evaluated in the latest generation of ESMs from CMIP6. For the effective climate sensitivity (ECS), which is an estimation of the equilibrium temperature response that follows a doubling of the atmospheric carbon dioxide ( $\text{CO}_2$ ) concentration, a multi-model mean (MMM) of 3.74 K and a multi-model range of 1.8–5.6 K are found. These values are higher than in any previous CMIP ensemble before. Moreover, a third of the analyzed CMIP6 models exceed the upper bound of the likely ECS range of 1.5–4.5 K assessed by the Intergovernmental Panel on Climate Change (IPCC) Fifth Assessment Report (AR5) from 2013. Similarly, the transient response of the climate system to a doubling of  $\text{CO}_2$ , also known as transient climate response (TCR), shows an inter-model range of 1.3–3.0 K with an upper bound again higher than the likely range assessed in AR5 of 1.0–2.5 K. Possible reasons for the increased climate sensitivity in many CMIP6 models are the addition of new prognostic aerosol schemes that include aerosol-cloud interactions and changes in the microphysical representation of mixed-phase clouds. The changes in the mixed-cloud representation reduce the strong negative shortwave cloud phase change feedback over the Southern Ocean that is present in climate models from previous CMIP generations.

To reduce uncertainties in ECS projected by the CMIP6 models, 11 published emergent constraints on ECS (mostly derived from models participating in CMIP5, the predecessor generation of CMIP6) are systematically analyzed. Emergent constraints are potentially promising approaches to reduce uncertainties in climate model projections by combining observations and output from ESMs. The focus of this analysis is on testing if these emergent constraints hold for ESMs participating in CMIP6. Since none of the emergent constraints considered here have been derived from the CMIP6 ensemble, the CMIP6 models can be used for cross-checking the emergent constraints on a new model ensemble. The application of the emergent constraints to CMIP6 data shows a decrease in skill and statistical significance of the emergent

relationship for nearly all constraints, with this decrease being large in many cases. Consequently, the size of the constrained ECS ranges (66 % confidence intervals) widens by 51 % on average in CMIP6 compared to CMIP5. This is likely related to the increased multi-model spread of ECS in CMIP6, but may in some cases also be due to spurious statistical relationships or a too-small number of models in the ensemble that the emergent constraint was originally derived from. The corresponding best estimates of ECS given by the emergent constraints also increase from CMIP5 to CMIP6 by 12 % on average. This can at least be partly explained by the increased number of high-ECS models in CMIP6 without a corresponding change in the constraint predictors, suggesting the emergence of new feedback processes rather than changes in the strength of those previously dominant. The results support previous studies concluding that emergent constraints should be based on an independently verifiable physical mechanism and that process-based emergent constraints on ECS should rather be thought of as constraints for the process or feedback they are actually targeting.

To overcome these issues of single-process-oriented emergent constraints, an alternative approach based on machine learning (ML) is introduced. Since this new technique relies on a large number of data points to train the ML algorithm, the scalar climate sensitivity expressed as ECS or TCR is not an appropriate target variable. Therefore, gross primary production (GPP) as a process that contributes to climate sensitivity is studied as an alternative. GPP is the largest flux of the terrestrial carbon uptake and slows down global warming by removing CO<sub>2</sub> from the atmosphere. In this analysis, an existing emergent constraint on CO<sub>2</sub> fertilization is combined with an ML approach to constrain the spatial variations of multi-model GPP projections. In the first step of the two-step approach, observed changes in the CO<sub>2</sub> seasonal cycle at Cape Kumukahi, Hawaii are used to constrain the global mean GPP at the end of the 21<sup>st</sup> century (2091–2100) in Representative Concentration Pathway (RCP) 8.5 simulations with ESMs participating in CMIP5 to  $(171 \pm 12)$  GtC yr<sup>-1</sup>, compared to the unconstrained model range of 156–247 GtC yr<sup>-1</sup>. In a second step, an ML model is used to constrain gridded future absolute GPP and gridded fractional GPP change in two independent approaches. For this, observational data is fed into the ML algorithm that has been trained on CMIP5 data to learn relationships between present-day physically relevant diagnostics and the target variable. In a leave-one-model-out cross-validation approach, the ML model shows superior performance to the CMIP5 MMM. The new approach predicts a higher GPP increase in high latitudes and a lower GPP increase in regions closer to the equator.

# Integrated Author's References

Parts of this thesis (text, figures, and tables) are already published in the following peer-reviewed publications. More details on this are given in Section 1.3 and at the beginning of the corresponding chapters.

- Bock, L., Lauer, A., **Schlund, M.**, Barreiro, M., Bellouin, N., Jones, C., Meehl, G. A., Predoi, V., Roberts, M. J., & Eyring, V. (2020). Quantifying Progress Across Different CMIP Phases With the ESMValTool. *Journal of Geophysical Research: Atmospheres*, 125(21), e2019JD032321. <https://doi.org/10.1029/2019jd032321>
- 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.,..., Zimmermann, K. (2020). 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(7), 3383–3438. <https://doi.org/10.5194/gmd-13-3383-2020>
- 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., & Zechlau, S. (2020). 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(9), 4205–4228. <https://doi.org/10.5194/gmd-13-4205-2020>
- Meehl, G. A., Senior, C. A., Eyring, V., Flato, G., Lamarque, J.-F., Stouffer, R. J., Taylor, K. E., & **Schlund, M.** (2020). Context for interpreting equilibrium climate sensitivity and transient climate response from the CMIP6 Earth system models. *Science Advances*, 6(26), eaba1981. <https://doi.org/10.1126/sciadv.aba1981>
- Righi, M., Andela, B., Eyring, V., Lauer, A., Predoi, V., **Schlund, M.**, Vegas-Regidor, J., Bock, L., Brotz, B., de Mora, L., Diblen, F., Dreyer, L., Drost, N., Earnshaw, P., Hassler, B., Koldunov, N., Little, B., Tomas, S. L., & Zimmermann, K. (2020). Earth System Model Evaluation Tool (ESMValTool) v2.0-technical overview. *Geoscientific Model Development*, 13(3), 1179–1199. <https://doi.org/10.5194/gmd-13-1179-2020>
- Schlund, M.**, Eyring, V., Camps-Valls, G., Friedlingstein, P., Gentine, P., & Reichstein, M. (2020a). Constraining Uncertainty in Projected Gross Primary Production With Machine Learning. *Journal of Geophysical Research: Biogeosciences*, 125(11), e2019JG005619. <https://doi.org/10.1029/2019jg005619>

- Schlund, M.**, Lauer, A., Gentine, P., Sherwood, S. C., & Eyring, V. (2020b). Emergent constraints on equilibrium climate sensitivity in CMIP5: do they hold for CMIP6? *Earth System Dynamics*, 11(4), 1233–1258. <https://doi.org/10.5194/esd-11-1233-2020>
- 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., . . . Eyring, V. (2020). 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 Discussions*, accepted, 1–43. <https://doi.org/10.5194/gmd-2020-244>

# Contents

<b>Abstract</b>	<b>v</b>
<b>Integrated Author's References</b>	<b>vii</b>
<b>1. Introduction</b>	<b>1</b>
1.1. Motivation . . . . .	1
1.2. Key Science Questions . . . . .	3
1.3. Structure of the Thesis . . . . .	3
<b>2. Scientific Background</b>	<b>5</b>
2.1. Earth System Models: Simulations and Analysis . . . . .	5
2.1.1. Numerical Climate Modeling . . . . .	5
2.1.2. The Coupled Model Intercomparison Project (CMIP) . . . . .	7
2.1.3. Sources of Uncertainties in Climate Model Projections . . . . .	9
2.2. Climate Sensitivity . . . . .	11
2.2.1. Climate Feedbacks . . . . .	11
2.2.2. Mathematical Framework for Feedbacks Analysis . . . . .	14
2.2.3. Equilibrium and Effective Climate Sensitivity (ECS) . . . . .	15
2.2.4. Cloud-Related Feedback Parameters . . . . .	17
2.2.5. Transient Climate Response (TCR) . . . . .	18
2.3. The Global Carbon Cycle . . . . .	19
2.3.1. Overview . . . . .	20
2.3.2. Anthropogenic Perturbations . . . . .	21
2.3.3. Representation in Earth System Models . . . . .	24
2.4. Reducing Uncertainties in Multi-Model Climate Projections with Observations	25
2.4.1. Emergent Constraints . . . . .	26
2.4.2. Performance- and Interdependence-based Weighting of Climate Models	30
2.4.3. Multiple Diagnostic Ensemble Regression . . . . .	31
<b>3. Improving Routine Climate Model Evaluation</b>	<b>35</b>
3.1. The Earth System Model Evaluation Tool (ESMValTool) . . . . .	35
3.2. Contributions to ESMValCore . . . . .	38
3.3. Contributions to ESMValTool . . . . .	40
<b>4. Assessment of Climate Sensitivity in the CMIP6 Ensemble</b>	<b>45</b>
4.1. Evaluation of ECS and TCR in CMIP5 and CMIP6 . . . . .	45

4.2. Comparison to Previous CMIP Generations and International Climate Assessments . . . . .	49
4.3. Possible Reasons for High Climate Sensitivity in CMIP6 . . . . .	50
<b>5. Evaluation of Emergent Constraints on the Effective Climate Sensitivity (ECS) in CMIP6</b>	<b>57</b>
5.1. Data and Methods . . . . .	59
5.2. Comparison of Emergent Constraints on ECS for CMIP5 and CMIP6 . . . . .	59
5.2.1. Sensitivity of Shortwave Cloud Albedo to Changes in Sea Surface Temperature (BRI) . . . . .	60
5.2.2. Temperature Variability Metric (COX) . . . . .	60
5.2.3. Extent of Southern Hemisphere Hadley Cell (LIP) . . . . .	62
5.2.4. Large-Scale Lower-Tropospheric Mixing (SHD) . . . . .	62
5.2.5. Small-Scale Lower-Tropospheric Mixing (SHS) . . . . .	62
5.2.6. Lower-Tropospheric Mixing Index (SHL) . . . . .	64
5.2.7. Error in Vertical Profile of Relative Humidity (SU) . . . . .	64
5.2.8. Tropical Mid-Tropospheric Humidity Asymmetry Index (TIH) . . . . .	65
5.2.9. Southern ITCZ Index (TII) . . . . .	66
5.2.10. Difference Between Tropical and Southern Midlatitudinal Cloud Fraction (VOL) . . . . .	66
5.2.11. Response of Marine Boundary Layer Cloud Fraction to Changes in Sea Surface Temperature (ZHA) . . . . .	67
5.2.12. Constrained ECS Ranges and Statistical Significance of the 11 Emergent Constraints . . . . .	69
5.3. Discussion . . . . .	71
5.4. Summary . . . . .	72
<b>6. Constraining Future Gross Primary Production (GPP) with Machine Learning</b>	<b>75</b>
6.1. Data and Methods . . . . .	76
6.1.1. Constraining the CO <sub>2</sub> Fertilization Effect (Step 1) . . . . .	76
6.1.2. Gradient Boosted Regression Tree (GBRT) Algorithm (Step 2) . . . . .	80
6.1.3. Assessment of Uncertainty and Predictive Skill . . . . .	83
6.2. Results . . . . .	84
6.2.1. Constraining Global Mean GPP Projections (Step 1) . . . . .	84
6.2.2. Constraining Gridded Absolute GPP Projections (Step 2a) . . . . .	86
6.2.3. Constraining Gridded Fractional GPP Change Projections (Step 2b) . . . . .	91
6.2.4. Comparison of the Absolute and Relative GBRT Prediction . . . . .	96
6.3. Summary and Discussion . . . . .	97
<b>7. Conclusion</b>	<b>101</b>
7.1. Overall Summary . . . . .	101
7.2. Outlook . . . . .	104

<b>Appendix</b>	<b>107</b>
A. Supplementary Materials for Chapter 5	107
B. Supplementary Materials for Chapter 6	114
B.1. Data Preprocessing	114
B.2. Gradient Boosted Regression Tree (GBRT) Algorithm	114
B.3. Evaluation of Prediction Uncertainty	115
B.4. Evaluation of Residuals	116
<b>List of Abbreviations</b>	<b>123</b>
<b>List of Figures</b>	<b>127</b>
<b>List of Tables</b>	<b>131</b>
<b>References</b>	<b>133</b>
<b>Acknowledgments</b>	<b>161</b>



# 1. Introduction

## 1.1. Motivation

Climate change is one of the greatest challenges for humankind today. The warming of the climate system is “unequivocal”, and “many of the observed changes are unprecedented over decades to millennia” (IPCC 2014). The changing climate increases the “likelihood of severe, pervasive and irreversible impacts for people, species and ecosystems” with “mostly negative impacts for biodiversity, ecosystem services and economic development”, and amplifies “risks for livelihoods and for food and human security” (IPCC 2014). Potential drivers for climate change are all natural and anthropogenic substances and processes that may alter the Earth’s energy budget. Human influence on the climate system is clear: Ever-increasing emissions of greenhouse gases (GHGs) since the end of the pre-industrial era largely driven by economic and population growth led to atmospheric concentrations of carbon dioxide ( $\text{CO}_2$ ), methane ( $\text{CH}_4$ ), and nitrous oxide ( $\text{N}_2\text{O}$ ) that are “unprecedented in at least the last 800,000 years” (IPCC 2014). The effects of these GHG emissions and other anthropogenic drivers have been “detected throughout the climate system and are extremely likely to have been the dominant cause of the observed warming since the mid-20<sup>th</sup> century” (IPCC 2014). The greenhouse effect is based on the optical properties of the GHGs: While the atmosphere is mostly transparent for the incoming solar (shortwave) radiation, the outgoing infrared (longwave) radiation that is reflected from the planet’s surface is partly absorbed by the GHG molecules through excitation of their corresponding vibrational modes and re-emitted isotropically. This process traps energy near the surface and leads to a warming of the Earth’s surface and the lower atmosphere. The individual impact of the different drivers of climate change is measured with the so-called *radiative forcing*, which quantifies the change in energy fluxes caused by changes in these drivers relative to pre-industrial conditions (IPCC 2013). The total radiative forcing is positive (which corresponds to a warming) and its largest contribution is caused by the increase in the atmospheric concentration of  $\text{CO}_2$  due to anthropogenic fossil fuel emissions since the mid-18<sup>th</sup> century (IPCC 2014). Apart from their physical warming effect on the climate, carbon-based GHGs like  $\text{CO}_2$  also directly influence the global carbon cycle, an important biogeochemical cycle of the Earth. A crucial flux of the carbon cycle is the *gross primary production (GPP)*, which describes the carbon uptake of the terrestrial biosphere due to photosynthesis. Since this land carbon uptake absorbs about 30 % of the anthropogenic  $\text{CO}_2$  emissions in today’s climate (Friedlingstein et al. 2020), this process substantially slows down global warming. Other important anthropogenic drivers of climate change are the emission of aerosols, and land-use/land-cover changes. Apart from anthropogenic drivers, there are

also natural processes that impact the climate system like changes in solar activity or volcanic eruptions. However, there is clear evidence that these natural drivers alone cannot explain the observed climate change (Haustein et al. 2017).

To successfully mitigate the massive impacts of climate change, a first important step is the understanding of climate change and its accurate quantification. Extremely valuable tools in this context are climate models, which allow us to simulate the behavior of the climate system under arbitrary conditions without having to perform (ethnically questionable) experiments in the real world. All around the world climate research institutes develop a variety of different climate models. Many of them participate in the Coupled Model Intercomparison Project (CMIP), which was initiated in 1995 by the Working Group on Coupled Modelling (WGCM) of the World Climate Research Programme (WCRP) to “better understand past, present and future climate changes arising from natural, unforced variability or in response to changes in radiative forcing in a multi-model context” (WCRP 2020). The CMIP models provide crucial input for the international climate assessments given by the Assessment Reports (ARs) of the Intergovernmental Panel on Climate Change (IPCC). For example, the latest generation of climate models from the most recent (sixth) phase of CMIP (Eyring et al. 2016a; known as CMIP6) support the assessment of the upcoming IPCC Sixth Assessment Report (AR6), and their predecessor models from CMIP5 (Taylor et al. 2012) have been assessed as part of the Fifth Assessment Report (AR5) in 2013 (Flato et al. 2013). Modern-day climate models, which allow the simulation of biological and chemical processes in addition to the dynamics of the physical components of the Earth system, are also known as *Earth system models* (ESMs) and provide the most sophisticated simulations of the Earth’s climate. In this thesis, the terms “climate model” and “Earth system model” are used interchangeably since most modern models participating in CMIP are ESMs or at least have ESM versions.

Simulations that extrapolate the state of the climate system into the future are called *climate model projections*. These include idealized simulations with only a prescribed change in the atmospheric CO<sub>2</sub> concentration (e.g., an instantaneous doubling of CO<sub>2</sub> or a CO<sub>2</sub> increase of 1 % per year) as well as more realistic projections that consider different future scenarios (e.g., a fossil fuel-based future or a scenario that is based on sustainable development). Multi-model projections from CMIP show a large inter-model range in many variables that are relevant for climate change (M. Collins et al. 2013; Flato et al. 2013). A crucial and policy-relevant example of this is climate sensitivity, which refers to the change in the global mean near-surface air temperature (GSAT) that results from a change in the radiative forcing. Common metrics for this are the *effective climate sensitivity* (ECS), which describes the equilibrium response of the climate system after a doubling of the atmospheric CO<sub>2</sub> concentration, and the *transient climate response* (TCR), which describes the transient response of the system to a CO<sub>2</sub> doubling. In AR5, the assessed range of both quantities is large with 1.5–4.5 K for ECS and 1.0–2.5 K for TCR (Stocker et al. 2013). The corresponding inter-model ranges from the CMIP5 models show similar results (Flato et al. 2013). For this reason, a careful statistical evaluation and further refinement of the output of multi-model climate projections are necessary to reduce associated uncertainties. A state-of-the-art technique for this is the *emergent constraints* method,

which uses a physically based inter-model relationship between an observable quantity of the Earth system and a target variable to reduce uncertainties in the target variable with observations (Allen and Ingram 2002). An alternative approach is the weighting of climate models based on their performance (i.e., the distance of one model to observational products) and interdependence (i.e., the distance of one model to other climate models) (Knutti et al. 2017b). These techniques form the baseline for the new analyses and results presented in this thesis, which partly utilize methods from a new emerging research field in climate sciences: artificial intelligence (AI) and machine learning (ML).

## 1.2. Key Science Questions

This thesis aims to reduce uncertainties in multi-model climate projections with observations by addressing the following three key science questions:

1. What is the range of climate sensitivity in the latest generation of ESMs from CMIP6 compared to previous multi-model ensembles, and do we understand the processes that determine this uncertainty range?
2. Can uncertainties in climate sensitivity be reduced with observations using the emergent constraint approach?
3. Can uncertainties in multi-dimensional (gridded) climate projections be reduced with ML techniques and observations?

## 1.3. Structure of the Thesis

Parts of this thesis are already published in multiple peer-reviewed publications (two first-author studies and six co-author studies). A complete list of these publications is given on page vii. Wherever material from these studies is presented in this thesis, the pronoun “we” is used to increase readability by avoiding the passive voice and to acknowledge all involved contributors. However, unless stated otherwise, all content from these publications (text, figures, and tables) shown in this thesis originates from the author of this thesis. A detailed list of contributions to these studies is given in the corresponding chapters.

This thesis is structured as follows: Chapter 2 introduces the scientific background. This includes relevant literature that is used as a baseline for this thesis. Chapter 3 gives an overview of the contributions made to the Earth System Model Evaluation Tool (ESMValTool), an open-source software package for the analysis of ESMs. These contributions helped to improve the routine evaluation of ESMs which is useful for the entire scientific community and led to co-authorship in four peer-reviewed studies (Eyring et al. 2020; Lauer et al. 2020; Righi et al. 2020; Weigel et al. 2020). Chapter 4 covers the assessment of climate sensitivity metrics like the ECS or TCR in the latest generation of ESMs from CMIP6. This work is already published in two scientific publications (Bock et al. 2020; Meehl et al. 2020). Since the ECS and

## 1. Introduction

---

TCR are considerably higher in this new climate model generation, Chapter 5 describes the assessment of emergent constraints on the ECS for these CMIP6 models and compares these to results derived from CMIP5 models. The content of this chapter is published in *Earth System Dynamics* (Schlund et al. 2020b). Chapter 6 focuses on a new method to reduce uncertainties in multi-dimensional (gridded) multi-model projections of the future climate with observations based on ML. As an example, the method is applied to GPP at the end of the 21<sup>st</sup> century, which is already published in the *Journal of Geophysical Research: Biogeosciences* (Schlund et al. 2020a). Finally, Chapter 7 provides a summary of the results of this thesis and gives an outlook of possible future work.

## 2. Scientific Background

This chapter introduces the scientific background of this thesis. First, basic concepts of climate model simulations and associated uncertainties are introduced (Section 2.1). Next, important metrics describing climate sensitivity (Section 2.2) and fundamental processes of the global carbon cycle (Section 2.1) are presented. Finally, state-of-the-art techniques that can be used to reduce uncertainties in projections of the future climate with observations are shown (Section 2.4). These methods form the basis for new techniques developed in this thesis.

### 2.1. Earth System Models: Simulations and Analysis

#### 2.1.1. Numerical Climate Modeling

In contrast to many other fields of science, researching the future evolution of the Earth's climate cannot be purely done by performing experiments in a laboratory. Due to the immense complexity of the Earth system (including physical, biological, and chemical processes on various temporal and spatial scales and their mutual interactions), we do not have access to a tiny replica of the Earth that we can analyze when exposed to different external conditions (Flato 2011). While observing the current state of the Earth System is (relatively) straightforward, gaining evidence about the future evolution of the climate by only considering present-day observations is rather difficult.

A possible way out is given by numerical climate models, which offer the possibility to simulate the Earth's climate on a computer. To efficiently replicate the vastly complex Earth system with finite computational resources, climate models divide the Earth into a set of grid cells. The typical size of an atmospheric grid cell in modern-day global climate models is about 100 km along the horizontal dimensions (latitude and longitude) and 1 km along the vertical dimension (pressure level or height). In addition, also the temporal evolution of the Earth system is discretized using time steps that are typically about 30 min long in modern-day global climate models. For each grid cell and time step, a single value per model variable is usually given. Examples for such variables provided by a climate model are the atmospheric prognostic variables velocity (horizontal and vertical), temperature, specific humidity, pressure, and density. These prognostic variables are related to each other via the *primitive equations*, a set of partial differential equations that can be derived from the conservation of momentum, mass, energy, and moisture (Holton 2004). To progress further in time, these primitive equations are solved numerically by the dynamical core of the climate models, which eventually describes the Earth's large-scale atmospheric motions. Similar to

## 2. Scientific Background

---



Figure 2.1.: Historical evolution of coupled climate models over the last 45 years. In the early days, these models were so-called atmosphere-ocean general circulation models (AOGCMs), which only included three components: the atmosphere, the land surface, and the ocean. Over time, the individual components grew in complexity and included a wider range of processes (illustrated by the growing cylinders). Eventually, more and more components (aerosols, carbon cycle, etc.) were added to the coupled system, forming the modern Earth system models (ESMs). Reproduced with permission from Cubasch et al. (2013) (their Figure 1.13).

this, many other processes of the Earth system are simulated by the climate models using other fundamental physical laws and principles. As opposed to these calculated variables that form the output of climate models, the corresponding input is mainly given by the radiative forcing.

Many processes in the Earth system occur at spatial scales much smaller than the size of a typical grid cell. Illustrative examples for this are clouds, which are usually smaller than a  $100\text{ km} \times 100\text{ km}$  grid cell, but still play an important role in the overall climate system by reflecting incoming and outgoing radiation (Boucher et al. 2013). To reasonably approximate these subgrid-scale processes, a concept called *parameterization* is used. Instead of simulating a process exactly, parameterizations aim to represent the effect of that process at the grid scale of the climate model by generating the appropriate forcing terms for the rest of the system and the rest of the processes (Gettelman and Rood 2016). Parameterizations are necessary to simulate many processes of the Earth system, for example, surface heat and moisture fluxes, moist convection, turbulent mixing, and radiation (Holton 2004).

The first numerical climate models came up in the 1960s and were based on weather prediction models (Flato 2011). Early models from the 1970s simulated only the physical components of the climate system: atmosphere, land surface, ocean, and sea ice (see Figure 2.1).

These models are called *atmosphere-ocean general circulation models* (AOGCMs) (Flato et al. 2013). Over the course of the years, climate models became more and more complex by including a wider range of processes within the components, but also by introducing new components to the coupled system. Examples of these are aerosols, the carbon cycle, dynamic vegetation, atmospheric chemistry, and land ice (see Figure 2.1). AOGCMs coupled to these additional components are called *Earth system models* (ESMs), which are the current state-of-the-art models that allow the most sophisticated simulations of the Earth's climate. In contrast to AOGCMs, ESMs enable the simulation of biological and chemical processes in addition to the dynamics of the physical components of the Earth system. Especially in the context of anthropogenic climate change, these additional processes are of uttermost importance for realistic climate model simulations, since the anthropogenic interference with the Earth system directly influences the various biogeochemical cycles of the Earth. For example, the emission of the most prominent greenhouse gas (GHG), carbon dioxide ( $\text{CO}_2$ ), immediately impacts the global carbon cycle by inserting additional carbon into the system (see Section 2.3 for details). Further examples include land-use changes like the deforestation of tropical rainforests, which also directly influence several biogeochemical cycles (e.g., carbon cycle, nitrogen cycle, phosphorus cycle, etc.) by altering respective sinks and sources.

Due to the complex interactions between the different components of the Earth system, these changes in the biogeochemical processes also affect the physical properties of the climate system. For example, due to the global carbon cycle, only about 50 % of the emitted  $\text{CO}_2$  by humankind remains in the atmosphere (Friedlingstein et al. 2020). The residual part is absorbed by the two other main carbon sinks of the planet: the terrestrial biosphere and the ocean. Since only atmospheric  $\text{CO}_2$  can act as GHG by introducing an additional radiative forcing to the Earth System leading to increasing surface temperatures, this uptake of  $\text{CO}_2$  by the carbon cycle slows down global warming.

### 2.1.2. The Coupled Model Intercomparison Project (CMIP)

Due to the complex nature of the Earth system itself, numerical models of it consist of hundreds of thousands of lines of computer code. Thus, a standardization of the experimental setup and model output to a certain degree is crucial for the various research groups developing ESMs all around the world to obtain comparable output and to facilitate model analysis. For this reason, the Working Group on Coupled Modelling (WGCM) of the World Climate Research Programme (WCRP) initiated the Coupled Model Intercomparison Project (CMIP) in 1995 to "better understand past, present and future climate changes arising from natural, unforced variability or in response to changes in radiative forcing in a multi-model context" (WCRP 2020). One major element of CMIP is to establish common standards, coordination, infrastructure, and documentation to facilitate the distribution of climate model output (Eyring et al. 2016a; Juckes et al. 2020).

A further main aspect is to provide a set of standardized experiments for global climate model simulations. To participate in the latest phase of CMIP, CMIP6, climate models need

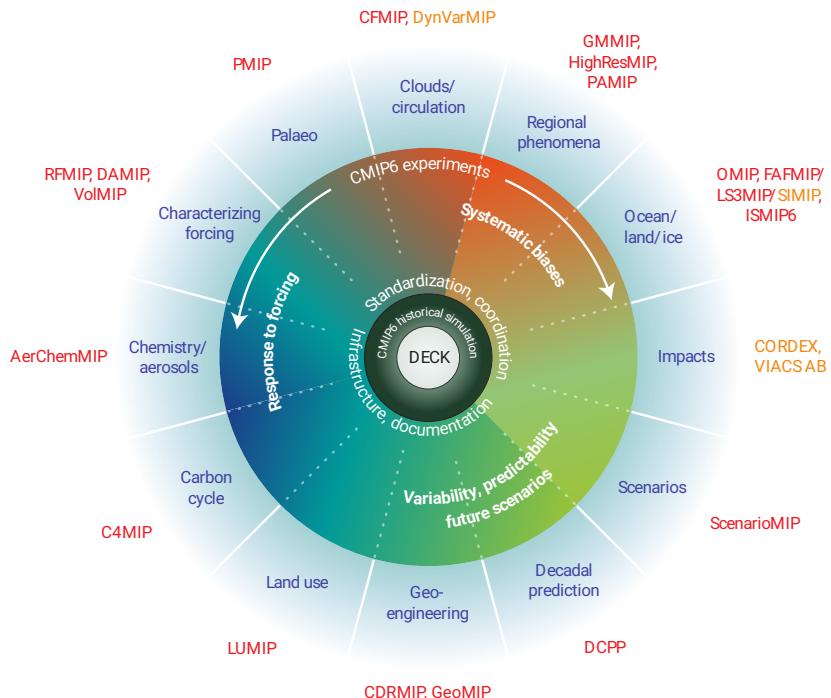


Figure 2.2.: Schematic of the experiment design of Phase 6 of the Coupled Model Intercomparison Project (CMIP6). The center of the circle illustrates the four DECK (Diagnostic, Evaluation, and Characterization of Klima) experiments and the CMIP6 historical simulation. The circular sectors show additional science themes that can be explored through the 23 CMIP6-Endorsed Model Intercomparison Projects (MIPs). Adapted by permission from Springer Nature Customer Service Centre GmbH: Simpkins (2017).

to run a *historical* simulation of the period 1850–2014 and the so-called *Diagnostic, Evaluation, and Characterization of Klima* (DECK) experiments, which include a pre-industrial control run (*piControl*), a historical Atmospheric Model Intercomparison Project (MIP) simulation (*amip*), a simulation forced with an abrupt quadrupling of CO<sub>2</sub> (*abrupt-4xCO<sub>2</sub>*), and a simulation forced with a 1 % per year increase of the atmospheric CO<sub>2</sub> concentration (*1pctCO<sub>2</sub>*) (Eyring et al. 2016a). This is shown in the center of Figure 2.2, which illustrates the experimental design of CMIP6.

To increase diversity and answer more scientific questions, CMIP6 models can participate in the so-called CMIP6-Endorsed MIPs, of which CMIP6 offers 23 (see circular sectors in Figure 2.2). Some MIPs offer additional experiments to explore specific aspects of the Earth system, like the Coupled Climate-Carbon Cycle Model Intercomparison Project (C4MIP) which focuses on the carbon cycle (Jones et al. 2016) or the Aerosol Chemistry Model Intercomparison Project (AerChemMIP) which focuses on aerosol chemistry (W. J. Collins et al. 2017). Other MIPs allow the assessment of future climate change. An example is the Scenario Model Intercomparison Project (ScenarioMIP), which provides common experiments that simulate different possible futures (O'Neill et al. 2016). These experiments are based on the so-called *Shared Socioeconomic Pathways* (SSPs), a set of alternative pathways of future societal development (O'Neill et al. 2013, 2017). For each experiment, specific emissions and land-use

changes are calculated from the SSPs (Riahi et al. 2017) which are then used to force the global climate models. For ScenarioMIP, five different SSPs are considered, ranging from SSP1 (sustainability) to SSP5 (fossil fuel-based development). Each SSP is combined with a climate outcome (measured as radiative forcing in the year 2100) based on a particular forcing pathway that integrated assessment models (IAMs) have shown to be feasible. For example, SSP5-8.5 represents a scenario based on a fossil fuel-based development with a radiative forcing of  $8.5 \text{ Wm}^{-2}$  in 2100 while SSP1-2.6 represents a sustainable future with a radiative forcing of  $2.6 \text{ Wm}^{-2}$  in the year 2100. The four scenarios in the main category of ScenarioMIP, the *Tier 1* experiments, are the SSP1-2.6, SSP2-4.5, SSP3-7.0, and SSP5-8.5 scenarios. In contrast to the ScenarioMIP experiments, the corresponding CMIP5 counterparts (Taylor et al. 2012), the so-called *Representative Concentration Pathways* (RCPs), only used the radiative forcing in 2100 as single dimension to describe the possible futures (e.g., RCP8.5, RCP4.5, RCP2.6, etc.).

In this thesis, climate model data from the two most recent CMIP generations is used, CMIP5 and CMIP6. More detailed information about the specific variables and experiments analyzed is given in the corresponding chapters.

### 2.1.3. Sources of Uncertainties in Climate Model Projections

Simulations from climate model ensembles of CMIP allow us to assess future climate change consistently and transparently. Especially the ScenarioMIP experiments can give valuable insights into possible developments of the Earth system by providing future *climate projections*. In contrast to climate predictions, climate projections run over multiple decades and depend upon the future scenario considered, which are based on assumptions that may or may not turn out to be correct. On the contrary, climate predictions are attempts to predict the actual evolution of the climate on much shorter time scales from seasons to years. Similar to any other scientific experiment, climate model projections suffer from associated uncertainties. There are three major sources of uncertainties in climate model projections we can distinguish: natural variability, climate response uncertainty, and emission uncertainty (Hawkins and Sutton 2009, 2010). Figure 2.3 shows these three sources for the projected global mean surface temperature anomaly over the 21<sup>st</sup> century.

*Natural variability* is connected to the chaotic nature of the Earth system that arises from complex interactions between the ocean, atmosphere, land, biosphere, and cryosphere (Cubasch et al. 2013). It constitutes a fundamental limit of how precisely we can project the future climate since it is inherent in the Earth system and cannot be eliminated by more knowledge and more advanced climate models. Natural variability is more relevant on regional and local scales than on continental or global scales. Further contributions to natural variability on longer time scales come from phenomena like the El Niño–Southern Oscillation (ENSO) or the North Atlantic Oscillation (NAO) and from other events like volcanic eruptions and variations in the solar activity. Natural variability can be seen as the *noise* in the climate record as opposed to the anthropogenic *signal* (Cubasch et al. 2013).

## 2. Scientific Background

---

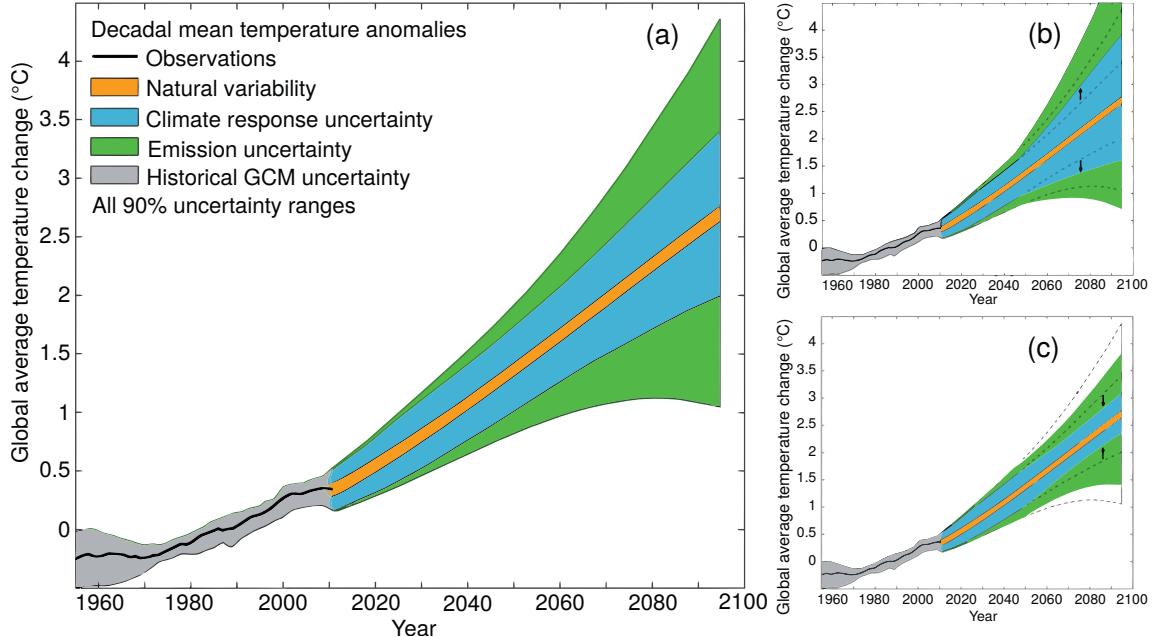


Figure 2.3.: Schematic illustration of the importance of different sources of uncertainties in climate model projections and their evolution in time. (a) Time series of the anomaly of the decadal and global mean surface temperature relative to the period 1961–1980. The black line shows the historical observations with estimates of uncertainty from climate models (gray). The remaining colors show different sources of uncertainty in future climate projections: Natural variability (orange), climate response uncertainty (blue), and emission uncertainty (green) (Hawkins and Sutton 2009, 2010). Climate response uncertainty can increase (b) in newer generations of climate models when a new process is discovered to be relevant or decrease (c) with additional model improvements and observational constraints. Reproduced with permission from Cubasch et al. (2013) (their FAQ 1.1, Figure 1).

The second source of uncertainty in climate model projections is *emission uncertainty*. This arises from the different possible trajectories in terms of future forcing (GHGs, aerosols, land-use changes, etc.) humankind might take. Examples for these are the SSP-based experiments given by ScenarioMIP that include a variety of different scenarios from a sustainable future to a full fossil fuel-based development (see Section 2.1.2). A possible approach to quantify emission uncertainty is to assess the climate impact of these different trajectories. Since the emission uncertainty strongly depends on the future development of human society, it cannot be reduced by improving climate models. In contrast to natural variability, the emission uncertainty increases over time in climate projections, since estimating forcings for the near future is simpler than for the far future.

Finally, the third source of uncertainty in climate model projections is the *climate response uncertainty*, which comes from our imperfect knowledge of how the climate system will respond to anthropogenic forcings. Due to the complexity of the Earth system, the future climate could develop in many different ways that are all consistent with our current knowledge and models (Cubasch et al. 2013). In the context of climate model ensembles, the climate response uncertainty is often also called *model uncertainty* and reflects the different responses of the

different climate models to a given forcing. Even though all climate models are built on the same physical principles, they differ in terms of spatial resolution, processes included, and parameterizations of unresolved processes (see Section 2.1.1).

These differences in the climate models also give rise to different intensities of *climate feedbacks* (or even their presence/absence) in the models. A climate feedback is a mechanism that either amplifies (*positive feedback*) or diminishes (*negative feedback*) the effect of an external forcing. An example of a strong positive feedback is the water vapor feedback, in which the increased surface temperature (caused by anthropogenic forcing) leads to enhanced evaporation of water which increases the amount of water vapor in the atmosphere. Since water vapor itself is a powerful GHG, this amplifies the effect of the anthropogenic forcing by further increasing the surface temperatures (Cubasch et al. 2013). Further examples and a mathematical framework for the analysis of feedbacks are given in Section 2.2.1.

As science evolves, representations of already included processes can be improved in climate models. Moreover, new geophysical and biogeochemical processes can be added to them. On the one hand, this can increase the climate response uncertainty when a new process is discovered to be relevant (Cubasch et al. 2013; see Figure 2.3b). However, such an increase corresponds to a previously unmeasured uncertainty. An example of this has recently happened in CMIP6: most likely due to changes in the cloud representation of the models the spread in the projected global mean near-surface air temperature (GSAT) caused by a doubling of the atmospheric CO<sub>2</sub> concentration has substantially increased in CMIP6 compared to older CMIP generations (Zelinka et al. 2020). On the other hand, the climate response uncertainty can decrease with additional model improvements and a better understanding of the Earth system (see Figure 2.3c). Moreover, it can also be reduced by observational constraints, which is the main topic of this thesis.

## 2.2. Climate Sensitivity

An important policy-relevant metric of the Earth system that can be assessed with numerical climate model simulations is climate sensitivity. Climate sensitivity refers to the change in GSAT that results from a change in the radiative forcing. In other words, it describes how sensitive the climate system is to an external forcing. The source of this forcing might either be natural (changes in the solar activity, volcanic eruptions, etc.) or anthropogenic (emissions of GHGs, land-use changes, etc.). Thus, the assessment of climate sensitivity is essential for a precise quantification of human-made climate change to determine optimal mitigation and adaptation strategies.

### 2.2.1. Climate Feedbacks

As already described in Section 2.1.3, the effects of an external forcing acting on the climate system can additionally be amplified or diminished by climate feedbacks. Thus, feedback processes play a crucial role in determining the magnitude of climate sensitivity. Figure 2.4

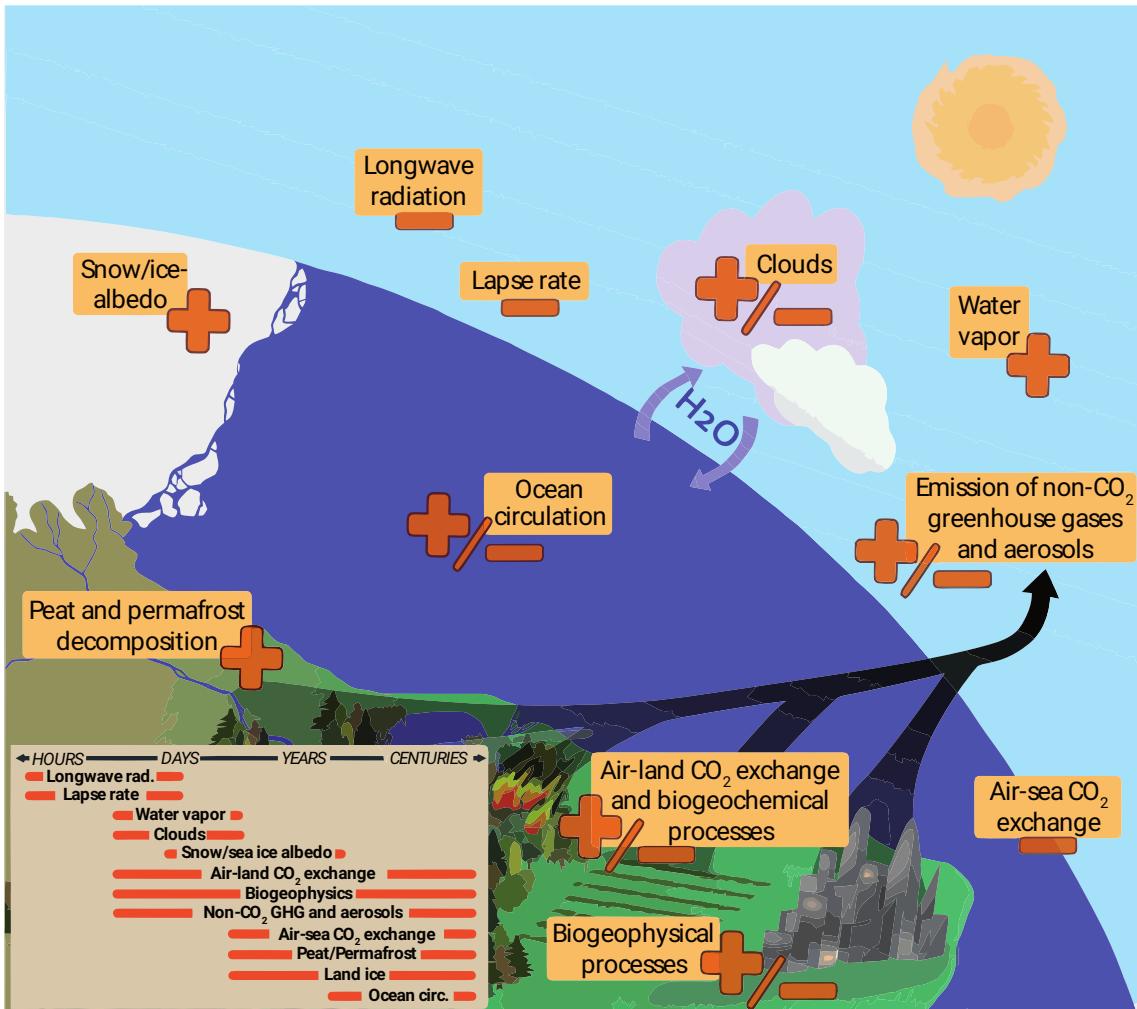


Figure 2.4.: Climate feedbacks and corresponding time scales. The different signs refer to the signs of the different feedbacks. Positive feedbacks amplify the effect of the external forcing (e.g., the water vapor feedback) and negative feedbacks diminish the effect of the external forcing (e.g., the longwave radiation feedback). An example of a feedback that might be either positive or negative is the cloud feedback. The smaller box highlights the large differences in time scales for the various feedbacks. Reproduced with permission from Cubasch et al. (2013) (their Figure 1.2).

shows an overview of important feedbacks in the Earth system with their corresponding time scales on which they operate.

An example of a positive feedback is the already mentioned *water vapor feedback*. Being the primary GHG in the Earth's atmosphere, water vapor is the largest contributor to the natural greenhouse effect. Since its amount in the atmosphere is mainly controlled by the air temperature, and anthropogenic emissions of water vapor are negligible, the influence of water vapor on the climate system is described as a feedback mechanism and not as an external forcing (Myhre et al. 2013). The basis of this feedback is the enhanced evaporation of water with increasing air temperatures. Each degree of warming allows the atmosphere to retain about 7 % more water vapor (Myhre et al. 2013), which closes the positive feedback

loop by further increasing air temperatures through the greenhouse effect. With a typical residence time of water vapor in the atmosphere of several days, the water vapor feedback operates on relatively short time scales. As the largest positive feedback in the Earth system (Soden and Held 2006), the water vapor feedback amplifies any initial forcing (e.g., caused by anthropogenic CO<sub>2</sub> emissions) by a typical factor between 2 and 3, rendering water vapor a fundamental agent of climate change (Myhre et al. 2013). An example of a positive feedback that operates on longer time scales (several years) is the *snow/ice-albedo feedback*, in which the surface albedo decreases as highly reflective ice and snow surfaces melt with global warming, exposing the darker and more absorbing surfaces below (Cubasch et al. 2013).

In contrast to positive feedbacks, negative feedbacks diminish the effect of an external forcing. An example of this is the *blackbody feedback* (also known as *Planck feedback* or *longwave radiation feedback*), which is the strongest negative feedback in the Earth system (Cubasch et al. 2013). It is based on the thermal electromagnetic radiation that any object with a non-zero temperature emits (the so-called *blackbody radiation*). Since the power of this radiation strongly depends on the temperature of the object, higher surface temperatures of the Earth increase the outgoing longwave radiation flux from the surface which reduces the effect of the external forcing and cools the planet.

For some domains of the Earth system, feedbacks can be positive and/or negative, since a variety of different mechanisms are involved. An example of this is the cloud feedback. Changes in the clouds induced by climate change can cause changes in their longwave (greenhouse warming) and shortwave (reflective cooling) effects on the Earth's radiation budget, which both need to be considered for the overall cloud feedback (Boucher et al. 2013). Relevant cloud properties that may change as a response to an external forcing and that may alter the Earth's radiative budget are cloud cover, cloud optical thickness, cloud top and cloud base height, vertical extent, and the geographical distribution of clouds. Examples for robust cloud feedback processes are the increase in cloud top height of high-level clouds in a warming climate which traps longwave radiation and enhances global warming, and the reduction in mid- and low-level cloud cover which diminishes the reflection of incoming solar radiation and also increases the surface warming (Boucher et al. 2013). In global climate model ensembles, the overall cloud feedback shows a large range with positive and negative values, but tends to be slightly positive on average (Dufresne and Bony 2008; Soden and Held 2006; Vial et al. 2013; Zelinka et al. 2020). This large uncertainty in the cloud feedback is a major reason for uncertainties in the climate sensitivity of climate models (Boucher et al. 2013; Flato et al. 2013).

Further examples of feedbacks with positive and negative contributions are biogeochemical feedbacks. Negative contributions come from increased CO<sub>2</sub> fluxes into the land and ocean carbon reservoirs due to increased photosynthesis rates and CO<sub>2</sub> dissolution in the sea, respectively, which decrease the atmospheric CO<sub>2</sub> content and diminish global warming. An example of a positive contribution is the decreased solubility of CO<sub>2</sub> in water in a warmer climate, which reduces the atmosphere-ocean CO<sub>2</sub> flux and enhances climate change. More details on this are given in Section 2.3.2.

### 2.2.2. Mathematical Framework for Feedbacks Analysis

For a precise quantification of climate feedbacks and thus climate sensitivity, a mathematical framework for climate feedback analysis is necessary. One possible approach for this is based on a simple energy balance model (Gregory et al. 2009; Roe 2009). Anthropogenic activities in the Earth system like the emissions of GHGs or aerosols introduce an external forcing to the climate system, which is quantified with a radiative forcing  $F$  measured in  $\text{Wm}^{-2}$ . To restore a stable state, the climate system opposes this forcing with a climate response  $R$ , leading to a net energy flux of

$$N = F + R \quad (2.1)$$

into the system. Positive values of  $N$ ,  $F$ , and  $R$  indicate incoming fluxes; usually  $F > 0$  and  $R < 0$ . On long time scales (multiple years), the net incoming radiative flux at the top of the atmosphere (TOA) and the net heat flux into the ocean are equal definitions of  $N$  since nearly all of the Earth's heat capacity resides in the ocean (Gregory et al. 2009). While  $N \neq 0$ , the climate system evolves; when  $N = 0$  a new steady state has been reached.

To quantify the effects of different feedbacks, a reference system with a basic response needs to be defined, which is a crucial aspect of feedback analysis (Roe 2009). Usually, the idealization of a blackbody Earth without an atmosphere is used for that: In equilibrium, the incoming solar irradiance is balanced with an outgoing thermal irradiance  $J_0$  that solely depends on the global mean surface temperature  $T_0$  following the Stefan-Boltzmann law

$$J_0 = -\sigma T_0^4. \quad (2.2)$$

$\sigma \approx 5.67 \text{ Wm}^{-2}\text{K}^{-4}$  is the Stefan-Boltzmann constant. To answer an external forcing  $F$ , the climate system reacts with a response  $R$  expressed by a change in the global mean surface temperature  $\Delta T$ :

$$J_0 + R = -\sigma (T_0 + \Delta T)^4. \quad (2.3)$$

Since the temperature change caused by an anthropogenic forcing is much smaller than the equilibrium temperature  $\Delta T \ll T_0 \approx 255 \text{ K}$ , a simple first-order Taylor expansion can be used to linearize the blackbody response:

$$-\sigma (T_0 + \Delta T)^4 \approx J_0 - 4\sigma T_0^3 \cdot \Delta T. \quad (2.4)$$

Thus, by comparing Equations (2.3) and (2.4) the climate response  $R$  can be expressed as

$$R = -4\sigma T_0^3 \cdot \Delta T := \lambda_{\text{BB}} \cdot \Delta T \quad (2.5)$$

with the blackbody feedback parameter  $\lambda_{\text{BB}} \approx -3.8 \text{ Wm}^{-2}\text{K}^{-1}$ . Results from climate models and analyses of observations confirm this linear relationship between  $R$  and  $\Delta T$  (Gregory et al. 2004). However, the value of this linear constant  $\lambda$ , the *climate feedback parameter*, is found to be considerably larger than the blackbody response ( $\lambda \approx -1.0 \text{ Wm}^{-2}\text{K}^{-1}$ ), indicating that additional processes affect the Earth's radiative balance: the climate feedbacks (Flato et al. 2013; Gregory et al. 2009). Since climate models suggest that the radiative effects of these

additional feedbacks are also proportional to  $\Delta T$  (Gregory and M. Webb 2008), Equation (2.5) can be adjusted to

$$R = \lambda \cdot \Delta T = (\lambda_{\text{BB}} + \lambda_{\text{WV}} + \lambda_{\text{Albedo}} + \lambda_{\text{Cloud}} + \dots) \cdot \Delta T. \quad (2.6)$$

$\lambda_{\text{WV}}$  refers to the water vapor feedback,  $\lambda_{\text{Albedo}}$  to the snow/ice-albedo feedback, and  $\lambda_{\text{Cloud}}$  to the cloud feedback. Thus, the overall climate feedback parameter  $\lambda$  can be written as the sum of the individual feedback parameters  $\lambda_i$ :

$$\lambda = \sum_i \lambda_i. \quad (2.7)$$

Positive values of  $\lambda_i$  indicate positive feedbacks (e.g., the water vapor feedback) and negative values indicate negative feedbacks (e.g., the blackbody feedback). This equation assumes that the individual radiative responses from the different feedbacks are independent, which is a reasonable first-order approximation but not entirely true (Soden et al. 2008).

### 2.2.3. Equilibrium and Effective Climate Sensitivity (ECS)

An important metric describing climate sensitivity is the *equilibrium climate sensitivity*. It is defined as the change in GSAT after an instantaneous doubling of the atmospheric CO<sub>2</sub> concentration from pre-industrial conditions once the climate system reaches radiative equilibrium (Bindoff et al. 2013). Being already used in one of the first assessments of the anthropogenic climate change, the *Charney Report* from 1979 (Charney et al. 1979), the equilibrium climate sensitivity is one of the oldest metrics describing climate change. However, in practice, this traditional definition is not always useful. Running fully coupled ESMs into equilibrium is computationally expensive as it would require thousands of model years (Rugensteiner et al. 2020).

For this reason, the equilibrium climate sensitivity is commonly approximated with the *effective climate sensitivity (ECS)*, which can be derived from only 150 model years of a simulation with an abrupt quadrupling of the atmospheric CO<sub>2</sub> concentration (4xCO<sub>2</sub>) (Gregory et al. 2004). The basis of the definition of ECS is the simple energy balance model introduced in Section 2.2.2. Assuming radiative equilibrium ( $N = 0$ ), Equations (2.1) and (2.6) imply

$$\Delta T = -\frac{F}{\lambda}. \quad (2.8)$$

Thus, the change in GSAT in radiative equilibrium can be easily calculated from the external forcing  $F$  and the climate feedback parameter  $\lambda$ . The steady-state values for  $F$  and  $\lambda$  can be estimated from a 4xCO<sub>2</sub> simulation that is not in equilibrium by extrapolation with a linear regression (Gregory et al. 2004). For this so-called *Gregory regression* (see Figure 2.5), the global and annual mean net TOA radiation  $N$  versus the change in the annual mean GSAT  $\Delta T$  for all 150 years of the 4xCO<sub>2</sub> run are plotted. To account for energy leakage and remove any model drift present in the control climate, a linear fit of the corresponding pre-industrial control



Figure 2.5.: Global and annual mean net top of the atmosphere (TOA) radiation  $N$  versus the change in global and annual mean near-surface air temperature  $\Delta T$  for 150 years of a simulation with an abrupt quadrupling of the atmospheric  $\text{CO}_2$  concentration ( $4\text{xCO}_2$ ) for the CMIP5 multi-model mean (blue points). To account for energy leakage and model drift, a linear fit of the corresponding pre-industrial control run is subtracted from the  $4\text{xCO}_2$  simulation. As given by Equation (2.9), the slope of the linear regression (black line) corresponds to the climate feedback parameter  $\lambda$ , and the  $y$ -intercept corresponds to the radiative forcing  $F_{4x}$ . These can be used to calculate the effective climate sensitivity (ECS) with the Gregory regression method according to Equation (2.10) (Gregory et al. 2004). ECS is equivalently given by the  $x$ -intercept of the linear regression line divided by 2. Here, ECS = 3.14 K.

run is subtracted from the  $4\text{xCO}_2$  simulation beforehand (Andrews et al. 2012). Since the combination of Equations (2.1) and (2.6) yields

$$N = F + \lambda \cdot \Delta T, \quad (2.9)$$

$F$  is now given by the  $y$ -intercept of this linear regression ( $F_{4x}$ ) and  $\lambda$  by its slope. Thus, the ECS is given by

$$\text{ECS} = -\frac{F_{4x}}{2\lambda}. \quad (2.10)$$

The factor of 2 in the denominator accounts for the fact that the traditional equilibrium climate sensitivity is defined for an abrupt  $\text{CO}_2$  doubling, whereas here a simulation with an abrupt quadrupling is considered.

Although commonly used in the literature, the ECS is known to be only an approximation of the true equilibrium climate sensitivity. One major reason for this is the state and time dependence of the global feedbacks (Knutti and Rugenstein 2015; Knutti et al. 2017a). As Figure 2.6 shows, the slope in the Gregory regression is not constant but rather decreases over time when a long-running  $4\text{xCO}_2$  simulation with more than 1000 model years is considered.



Figure 2.6.: As in Figure 2.5 but for different time periods considered in the Gregory regression resulting in different values for the effective climate sensitivity (ECS) and the climate feedback parameter  $\lambda$ . Dark gray, blue, red, and orange colors correspond to the Community Earth System Model (CESM) of the National Center for Atmospheric Research (NCAR) for different time periods of a long-running simulation (see legend). The initial years are simulated many times for different initial conditions. Light gray colors correspond to the CMIP5 ensemble (150 years each). Adapted by permission from Springer Nature Customer Service Centre GmbH: Knutti et al. (2017a).

As a result, the climate feedback parameter  $\lambda$  decreases over time, resulting in a higher ECS. Major reasons for this are temperature dependencies of the feedbacks, atmospheric and oceanic adjustments over time, changing warming patterns over time, non-additive feedbacks, and dependencies on the type and magnitude of the external forcings (Knutti et al. 2017a). All in all, this demonstrates the limits of the linear feedback framework introduced in Section 2.2.2 that is not capable of describing non-linear effects. A second major reason for the discrepancies between the equilibrium climate sensitivity and ECS is the use of a  $4\text{xCO}_2$  instead of a  $2\text{xCO}_2$  simulation. The factor of 2 in the denominator of Equation (2.10) only partly compensates for this difference since the radiative forcing logarithmically depends on the atmospheric  $\text{CO}_2$  concentration (Huang and Shahabadi 2014).

However, despite these deficiencies, the ECS is still a practical estimate of the equilibrium climate sensitivity. With the help of climate models, Sherwood et al. (2020) showed that the ECS is only about 6 % lower than the best estimate of the true equilibrium warming obtained from integrating climate models until a new steady state is reached. Nevertheless, for CMIP6 long-running simulations from the Long Run Model Intercomparison Project (LongRunMIP) (Rugenstein et al. 2019) can be a promising way forward to estimate the true equilibrium climate sensitivity for ESMs.

#### 2.2.4. Cloud-Related Feedback Parameters

In addition to the calculation of the external forcing  $F$ , the overall climate feedback parameter  $\lambda$ , and the ECS, the Gregory regression can also be used to estimate cloud-related feedback parameters. For this, the net TOA radiation  $N$  on the  $y$ -axis in Figure 2.5 is replaced with the cloud radiative effect (CRE), which is defined as the difference between the all-sky (i.e., with

clouds if present) net TOA radiation and the clear-sky (i.e., clouds artificially removed) net TOA radiation (Andrews et al. 2012). This can be done for the shortwave  $N_{\text{SWCRE}}$  and longwave  $N_{\text{LWCRE}}$  components separately, but also for the combined effect  $N_{\text{CRE}} = N_{\text{SWCRE}} + N_{\text{LWCRE}}$ . The slopes in the corresponding Gregory regressions are the so-called *CRE feedback parameters*  $\lambda_{\text{SWCRE}}$ ,  $\lambda_{\text{LWCRE}}$ , and  $\lambda_{\text{CRE}}$ , which quantify the change in CRE as a response to increasing GSATs.

### 2.2.5. Transient Climate Response (TCR)

A further metric describing climate sensitivity is the *transient climate response (TCR)*. In contrast to the ECS, this metric does not assume radiative equilibrium of the Earth system but describes the transient response of an evolving climate. Following Bindoff et al. (2013), TCR is defined as the change in the GSAT at the time of CO<sub>2</sub> doubling in a simulation with a 1% per year increase of the atmospheric CO<sub>2</sub> concentration (1%CO<sub>2</sub>). For this, the annual mean GSATs are averaged over a 20-year period centered at the time of the CO<sub>2</sub> doubling (years 61–80 when the first year is indexed with 1). To account for model drift, the annual mean changes in GSAT are calculated relative to a corresponding pre-industrial control simulation smoothed with a linear fit that considers 140 model years (length of the 1%CO<sub>2</sub> simulation). An illustration of that calculation is shown in Figure 2.7.

Similar to ECS, TCR can also be defined in terms of an external forcing and climate feedbacks. For this, the energy balance Equation (2.9) can be slightly adjusted. Since over 90 % of the excess energy introduced into the climate system by the radiative forcing  $F$  is taken up by the ocean due to its large heat capacity,  $N$  can be taken equal to the global ocean heat uptake (Knutti et al. 2017a). In experiments with a steadily increasing radiative forcing, which is the case for the 1%CO<sub>2</sub> simulation, this ocean heat uptake can be approximated with

$$N = \kappa \cdot \Delta T, \quad (2.11)$$

where  $\kappa$  is the ocean heat uptake efficiency (Gregory and Forster 2008). Since there is a net energy flux into the climate system ( $N > 0$ ) due to the external forcing  $F$ ,  $\kappa$  is positive. This approximation becomes less accurate as the deeper ocean warms up and cannot be applied to simulations with a steady-state climate change in which  $N \rightarrow 0$  (Gregory et al. 2009). By applying the definition of TCR (transient GSAT change  $\Delta T$  at the time of CO<sub>2</sub> doubling) and combining Equations (2.9) and (2.11), TCR can be estimated as

$$\text{TCR} = \frac{F_{2x}}{\kappa - \lambda}, \quad (2.12)$$

where  $F_{2x}$  is the radiative forcing induced by a doubling of the atmospheric CO<sub>2</sub> concentration.

This equation can be used to derive a relationship between TCR and ECS. Writing  $\text{ECS} = -F_{2x}/\lambda$  and assuming path independence of the forcing (i.e., the resulting radiative forcing from the 1%CO<sub>2</sub> and 2xCO<sub>2</sub> runs are comparable) gives

$$\text{TCR} = \frac{F_{2x} \cdot \text{ECS}}{F_{2x} + \kappa \cdot \text{ECS}}, \quad (2.13)$$



Figure 2.7.: Global and annual mean near-surface air temperature change  $\Delta T$  for 140 years of a simulation with a 1 % per year increase of the atmospheric CO<sub>2</sub> concentration (1%CO<sub>2</sub>) for the Community Earth System Model (CESM) of the National Center for Atmospheric Research (NCAR) (blue points). The temperature change is calculated relative to a corresponding pre-industrial control simulation smoothed with a linear fit over all 140 years. The transient climate response (TCR) is defined as the temperature change  $\Delta T$  at the time of CO<sub>2</sub> doubling averaged over a 20-year period (illustrated by the vertical dashed lines). Here, TCR = 2.35 K (horizontal dashed line).

which demonstrates the non-linear connection between TCR and ECS (Gregory and Forster 2008; Nijssse et al. 2020). Since  $\lambda < 0$  and  $\kappa > 0$ , Equation (2.12) implies that the equilibrium response ECS is (as expected) larger than transient response TCR, i.e., ECS > TCR.

### 2.3. The Global Carbon Cycle

Currently, only about 50 % of the emitted CO<sub>2</sub> by humankind remains in the atmosphere (Friedlingstein et al. 2020). The residual part is absorbed by the two other main carbon sinks of the planet: the terrestrial biosphere and the ocean. These exchange fluxes of carbon are part of the global carbon cycle of planet Earth. Since only atmospheric CO<sub>2</sub> can act as GHG, this removal of CO<sub>2</sub> from the atmosphere substantially slows down global warming. Moreover, since the emission of carbon-based GHGs introduces additional carbon into the Earth system, the carbon exchange fluxes might change under global warming, which in turn directly influences the magnitude of climate sensitivity. However, the two idealized climate sensitivity metrics ECS and TCR introduced in the previous section are by definition independent of carbon cycle-related effects, since both of them are defined in terms of a doubling of the atmospheric CO<sub>2</sub> concentration, not in terms of CO<sub>2</sub> emissions. Nevertheless,

to assess policy-relevant metrics like the allowable fossil fuel emissions to meet particular warming targets, for example, the 1.5 °C of the Paris Agreement (UNFCCC 2015), it is crucial to take the global carbon cycle into account. This section introduces the scientific background of the global carbon cycle and its current anthropogenic perturbations.

### 2.3.1. Overview

A schematic overview of the global carbon cycle is shown in Figure 2.8. To quantify the carbon cycle, common units are parts per million (ppm) for the atmospheric trace gas concentrations (dry-air mole fraction) and gigatonnes of carbon (GtC) or GtC yr<sup>-1</sup> for the reservoir masses or exchange fluxes, respectively. The carbon exchange processes between the different carbon reservoirs run on a wide range of time scales. Conceptually, one can distinguish between two domains of the global carbon cycle: a slow and a fast domain. The slow domain with turnover times (reservoir mass of carbon divided by exchange flux) of more than 10,000 years consists of the large carbon stores in rocks and sediments which are connected to the rapid domain of the carbon cycle through volcanic emissions of CO<sub>2</sub>, chemical weathering, erosion, and sediment formation on the seafloor. These natural exchange fluxes between the slow and the fast domain are comparatively small (< 0.3 GtC yr<sup>-1</sup>) and can be assumed as approximately constant in time over the last few centuries (Ciais et al. 2013).

The fast domain of the global carbon cycle consists of three main carbon reservoirs: the atmosphere, the terrestrial biosphere, and the ocean. In the atmosphere, carbon is mainly stored in trace gases, with CO<sub>2</sub> as the major component with a current (2019) concentration of about 410 ppm (Friedlingstein et al. 2020). Additional contributors to the atmospheric carbon content are the trace gas methane (CH<sub>4</sub>), the trace gas carbon monoxide (CO), hydrocarbons, black carbon aerosols, and organic compounds (Ciais et al. 2013). Carbon in the terrestrial biosphere is mainly stored as organic compounds, with about 450–650 GtC in the living vegetation biomass, 1500–2400 GtC in dead organic matter in litter and soils, and about 1700 GtC in permafrost soils (Ciais et al. 2013). The main component of the oceanic carbon reservoir is dissolved inorganic carbon (carbonic acid, bicarbonate ions, and carbonate ions) with about 38,000 GtC. Further carbon is stored as dissolved organic carbon (about 700 GtC), in surface sediments (about 1750 GtC), and in marine biota (about 3 GtC, predominantly phytoplankton and other microorganisms) (Ciais et al. 2013; Friedlingstein et al. 2020).

In the fast domain of the global carbon cycle, reservoir turnover times range from seconds to millennia. In contrast to the slow domain, the carbon exchange fluxes within the fast domain of the carbon cycle are much higher. One major group of exchange processes in the fast domain connects the atmosphere and the terrestrial biosphere. CO<sub>2</sub> is removed from the atmosphere by plant photosynthesis with about 120 GtC yr<sup>-1</sup> (Ciais et al. 2013). This process is also known as *gross primary production* (GPP). The carbon fixed into plants can be released back into the atmosphere by autotrophic (plant) and heterotrophic (soil microbial and animal) respiration and additional disturbance processes like fires (Ciais et al. 2013). Since the land CO<sub>2</sub> uptake by photosynthesis occurs only during the growing season, whereas respiration

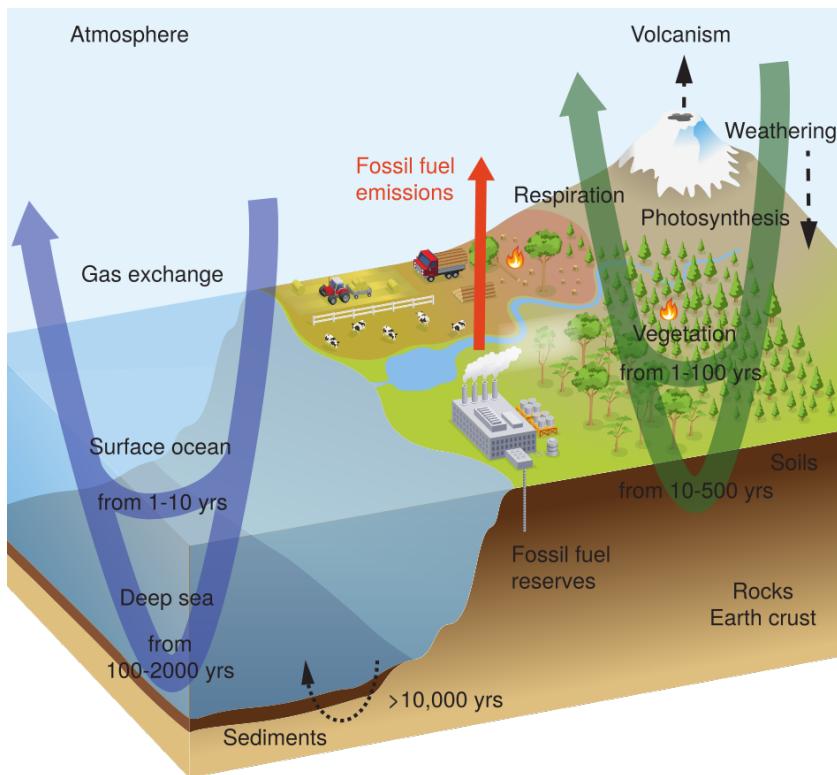


Figure 2.8.: Simplified schematic of the global carbon cycle including the typical turnover time scales for carbon transfers through the major reservoirs (atmosphere, land surface, and ocean). Reproduced with permission from Ciais et al. (2013) (their FAQ 6.2, Figure 1).

occurs nearly all year, the larger amount of vegetation in the Northern Hemisphere (due to the larger landmass) gives rise to a seasonal cycle of the atmospheric CO<sub>2</sub> concentration (Keeling et al. 1995). This seasonal cycle reflects the phase of the global carbon cycle and shows a maximum of the atmospheric CO<sub>2</sub> concentration in the Northern Hemisphere winter (net CO<sub>2</sub> flux into the atmosphere due to respiration) and a minimum during the Northern Hemisphere summer (net CO<sub>2</sub> flux into the land due to photosynthesis). Another major carbon exchange process connects the atmosphere and the ocean. Atmospheric CO<sub>2</sub> is exchanged with the surface ocean through a gas exchange, which is driven by the partial CO<sub>2</sub> pressure difference between the air and the sea (Ciais et al. 2013).

### 2.3.2. Anthropogenic Perturbations

Before the Industrial Era, the global carbon cycle was roughly in a dynamic equilibrium, which means that exchange fluxes balanced each other and the amount of carbon in the different reservoirs did neither increase nor decrease. This can be inferred from ice core measurements, which show an almost constant atmospheric CO<sub>2</sub> concentration over the last several thousand years before the Industrial Revolution in the 19<sup>th</sup> century (Ciais et al. 2013). Since the beginning of the Industrial Era, humanity is constantly emitting carbon-based GHGs (e.g., CO<sub>2</sub> and CH<sub>4</sub>) into the atmosphere. Especially the atmospheric CO<sub>2</sub> concentration has substantially increased, which has already been shown by Charles D. Keeling in 1976 by his



Figure 2.9.: The Keeling Curve: monthly mean atmospheric CO<sub>2</sub> concentration at the Mauna Loa Observatory, Hawaii (19.5 °N, 155.6 °W; elevation: 3397 m) from 1958 to 2019 (Keeling et al. 2005). The steady increase of the atmospheric CO<sub>2</sub> concentration is superimposed with a pronounced seasonal oscillation caused by the seasonal CO<sub>2</sub> cycle (see Section 2.3.1).

continuous CO<sub>2</sub> measurements at Mauna Loa, Hawaii that started in 1958 (Keeling et al. 1976; see Figure 2.9). From 1958, the atmospheric CO<sub>2</sub> concentration at Mauna Loa has steadily increased by about 100 ppm to 410 ppm in the year 2019 (Keeling et al. 2005). In addition to the steady increase, the so-called *Keeling Curve* is further superimposed with the seasonal CO<sub>2</sub> cycle, which gives rise to local maxima of the atmospheric CO<sub>2</sub> concentration in the Northern Hemisphere winter and local minima in the Northern Hemisphere summer (Keeling et al. 1995; see Section 2.3.1). Due to its location in the middle of the Pacific Ocean, the Mauna Loa Observatory offers perfect conditions for CO<sub>2</sub> measurements by being far away from big population centers. Moreover, its elevation of more than 3000 m provides access to the free troposphere where CO<sub>2</sub> is well mixed, which prevents any interference from the vegetation present on the Hawaiian Islands.

Apart from warming the Earth by altering its radiation budget, the anthropogenically emitted CO<sub>2</sub> directly influences the carbon exchange fluxes of the global carbon cycle. Due to the excess carbon in the atmosphere, there is now a net carbon flux from the atmosphere into the land and ocean reservoirs (see Figure 2.10). Thus, the carbon cycle is not in a

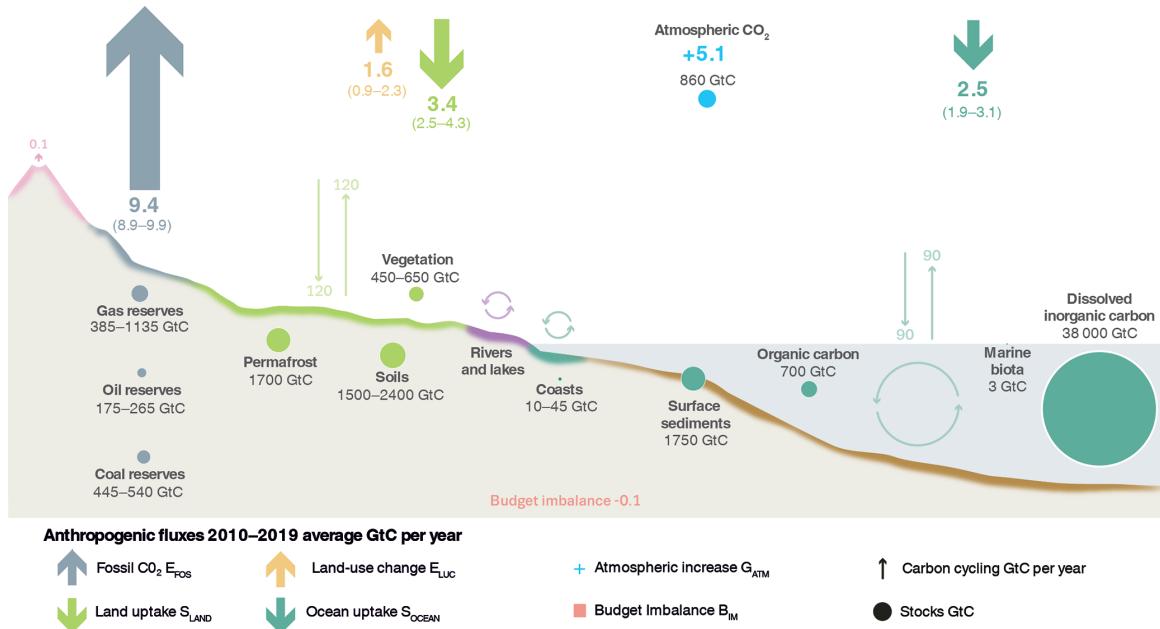


Figure 2.10.: Schematic representation of the overall perturbation of the global carbon cycle caused by anthropogenic activities, averaged globally for the decade 2010–2019. Arrows represent carbon exchange fluxes; circles carbon reservoirs. More details are given in the legend of this figure. Adapted with permission from Friedlingstein et al. (2020).

steady state anymore. In the decade 2010–2019, anthropogenic activities caused net carbon fluxes of  $3.4 \text{ GtC yr}^{-1}$  from the atmosphere into the terrestrial biosphere due to increased plant photosynthesis and  $2.5 \text{ GtC yr}^{-1}$  from the atmosphere into the ocean due to increased dissolution of CO<sub>2</sub> into the sea (Friedlingstein et al. 2020). At the same time, the amount of carbon in the atmosphere reservoir increased with a rate of  $5.1 \text{ GtC yr}^{-1}$ , indicating that only about half of the anthropogenic CO<sub>2</sub> emissions in the last decade remained in the atmosphere (Friedlingstein et al. 2020) where they can act as GHG.

Thus, this removal of CO<sub>2</sub> from the atmosphere actively slows down global warming. However, whether this benefit will persist in the future remains unclear, which is primarily linked to two feedback processes connecting the physical climate system and the global carbon cycle: the *concentration-carbon feedback* and the *climate-carbon feedback* (M. Collins et al. 2013; Friedlingstein et al. 2006; Gregory et al. 2009). For the terrestrial biosphere, the concentration-carbon feedback is connected to the *CO<sub>2</sub> fertilization effect* (Walker et al. 2020), which causes an increase in photosynthesis rates when the atmospheric CO<sub>2</sub> concentration increases, which in turn removes CO<sub>2</sub> from the atmosphere, forming a negative feedback. For the ocean, the concentration-carbon feedback is negative as well. In this case, an elevated atmospheric CO<sub>2</sub> concentration causes an increased dissolution of CO<sub>2</sub> into the sea, which increases the ocean carbon uptake. On the other hand, the climate-carbon feedback is thought to be positive for both the terrestrial biosphere and the ocean (Gregory et al. 2009). In the first case, temperature and precipitation changes due to anthropogenic activities decrease the land carbon uptake because of increased temperature and water stress on photosynthesis and higher ecosystem

respiration costs, which accelerates global warming due to more CO<sub>2</sub> that remains in the atmosphere. For the ocean, increased temperatures lead to a reduction of vertical transport in the ocean resulting from increased stability and reduced solubility of CO<sub>2</sub> in the sea, which reduces the ocean carbon uptake and enhances climate change (Gregory et al. 2009).

### 2.3.3. Representation in Earth System Models

In modern ESMs, the carbon cycle is usually represented by a land carbon cycle model and an ocean carbon cycle model that are both coupled to the other components of the ESM. An overarching principle is the conservation of the total carbon mass in the Earth system (Gregory et al. 2009), i.e.,

$$C_E = C_A + C_L + C_O. \quad (2.14)$$

Here,  $C_E$  represents the cumulative anthropogenic carbon emissions that are distributed among the three main carbon reservoirs of the Earth system.  $C_A$ ,  $C_L$ , and  $C_O$  describe the corresponding changes in these carbon stores, namely the atmosphere, the terrestrial biosphere (land), and the ocean, respectively.

Similar to many cloud processes, most carbon cycle-related processes need to be parameterized in ESMs since they occur on spatial scales much smaller than the typical grid-cell sizes (see Section 2.1.1). For the terrestrial carbon cycle, the major carbon exchange processes that need to be considered are photosynthesis, respiration, and disturbances, which connect the atmospheric carbon reservoir and the land carbon reservoir. The latter mainly consists of the vegetation biomass, litter, and soil. The terrestrial carbon exchange processes are commonly described with GPP, net primary production (NPP), net ecosystem production (NEP), and net biome production (NBP), which are all illustrated in Figure 2.11. GPP describes the total carbon uptake of the vegetation by photosynthesis, while NPP refers to the net carbon gain of the plants after accounting for autotrophic respiration ( $R_a$ ). NEP additionally considers heterotrophic respiration ( $R_h$ ) and describes the carbon uptake/release of the entire ecosystem. Finally, NBP describes changes in the long-term carbon storage of the terrestrial biosphere by including disturbances like fires ( $L_d$ ).

The different land carbon cycle models integrated into modern ESMs use a variety of different parameterizations for the different processes (Arora et al. 2020). Examples for photosynthesis (GPP) are given by Farquhar et al. (1980) for C<sub>3</sub> plants and Collatz et al. (1992) for C<sub>4</sub> plants. The terms “C<sub>3</sub>” and “C<sub>4</sub>” refer to the different metabolic pathways used in carbon fixation during photosynthesis. In both cases, the enzyme ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO) is an important catalyst for the corresponding chemical reactions. Both parameterizations utilize variables that are known to affect photosynthesis, such as the availability of light, the CO<sub>2</sub> concentration, the soil moisture, and the temperature.



Figure 2.11.: Schematic illustration of the processes of the terrestrial carbon uptake. The gross primary production (GPP) describes the total carbon uptake of the vegetation by photosynthesis, while the net primary production (NPP) refers to the net carbon gain of the plants after accounting for autotrophic respiration ( $R_a$ ). The net ecosystem production (NEP) additionally considers heterotrophic respiration ( $R_h$ ) and describes the carbon uptake/release of the entire ecosystem. Finally, the net biome production (NBP) describes changes in the long-term carbon storage of the terrestrial biosphere by including disturbance losses ( $L_d$ ).

## 2.4. Reducing Uncertainties in Multi-Model Climate Projections with Observations

As shown in Section 2.1.3, projections of the future climate are always associated with uncertainties. In the context of this thesis, the most relevant source of uncertainty is the climate response uncertainty. It originates from necessary simplifications that have to be implemented into the climate models due to limited computational resources and from our imperfect knowledge on how the climate system will respond to external forcing. In a multi-model ensemble (e.g., from CMIP), the climate response uncertainty is expressed in the different responses of the different climate models to a given forcing. A common approach to distill information about a projected quantity from multi-model ensembles is to treat the arithmetic multi-model mean (MMM) as best estimate and the multi-model range as uncertainty measure (M. Collins et al. 2013). This *model democracy* approach assumes that all climate models are independent, equally plausible, distributed around reality, and that the projected multi-model range is representative of the uncertainty in the projected quantity (Knutti et al. 2017b). However, since the CMIP ensembles, sometimes referred to as *ensembles of opportunity*, have not been designed to represent a true statistical sample of the reality composed of independent climate models (Tebaldi and Knutti 2007), these assumptions do not hold in practice. The main reasons for this are that different climate models (even for different modeling institutions) share parts of their code (Abramowitz et al. 2019; Knutti et al. 2013), that models do not equally well represent the observed past and present-day climate (Gleckler et al. 2008; Knutti et al. 2013), and that

models might suffer from common structural limitations like missing processes (Knutti et al. 2017b).

Thus, more sophisticated techniques are necessary to evaluate multi-model climate projections. This section introduces three state-of-the-art methods to assess multi-model projections and reduce associated uncertainties with observations. These techniques form the baseline of the work presented in the following chapters.

#### 2.4.1. Emergent Constraints

As indicated at the beginning of this section, one main issue of multi-model ensembles is that not all participating climate models are equally plausible. Usually, this is quantified with some kind of measure of the models' *performance*, i.e., their agreement with observations of the real climate system. However, this model performance can only be evaluated against observations of the past and present-day climate, which does not necessarily provide insights into the quality of model projections of the future climate.

The *emergent constraint* approach tackles this problem by "identifying robust, physically interpretable relationships between Earth system feedback behaviors on short, well-observed time scales and on time scales that span the twenty-first century and beyond" (Eyring et al. 2019). An illustration of the concept of emergent constraints is shown in Figure 2.12. Each emergent constraint requires two key components: an *emergent relationship* and a corresponding observation of the real world (Eyring et al. 2019). The emergent relationship (red line in Figure 2.12) is a robust and physically interpretable inter-model relationship between a target variable  $y$  related to the future climate and an observable  $x$  of the past or present-day climate. The basis for the relationship is output from the different climate models of a multi-model ensemble (blue points in Figure 2.12). Using an observation of  $x$ , this emergent relationship can then be used to derive an emergent constraint on  $y$  (gray shaded area in Figure 2.12) that considers uncertainties in the emergent relationship itself (red shaded area in Figure 2.12) and uncertainties in the observation (blue shaded area in Figure 2.12).

One possible mathematical framework for the evaluation of emergent constraints is based on linear regression and Gaussian probability densities (Cox et al. 2013, 2018). Let  $x_m$  be the observable predictor variable for climate model  $m$  and  $y_m$  the corresponding target variable. To find the linear emergent relationship for a climate model ensemble with  $M$  climate models and data  $\{(x_m, y_m) \mid m \in I_M\}$  with index set  $I_M = \{1, 2, \dots, M\}$ , a linear regression model

$$\hat{y}(x) = \hat{b}_0 + \hat{b}_1 x \quad (2.15)$$

for the predicted target variable  $\hat{y}$  with estimated intercept  $\hat{b}_0$  and slope  $\hat{b}_1$  is used (see Equation (2.26) for details). Fitting this regression line with ordinary least squares includes minimizing the standard error  $s$  of the estimate

$$s^2 = \frac{1}{M-2} \sum_{m=1}^M (y_m - \hat{y}_m)^2, \quad (2.16)$$



Figure 2.12.: Schematic illustration of the emergent constraint approach. Basis of every emergent constraint is a robust and physically interpretable emergent relationship (red line) between a target variable  $y$  (e.g., Earth system sensitivity or projection of future climate change) and an observable  $x$  (e.g., past or present-day trend or variation) for the climate models of a multi-model ensemble (blue points). With an observation of  $x$ , uncertainties in  $y$  in the multi-model ensemble, illustrated by the yellow probability density function (PDF), can be reduced (gray PDF). Uncertainties in the target variable  $y$  arise from two sources: uncertainties in the observation (blue shaded area) and uncertainties in the emergent relationship (red shaded area). Adapted by permission from Springer Nature Customer Service Centre GmbH: Eyring et al. (2019).

where  $\hat{y}_m := \hat{y}(x_m)$  is the predicted target variable for climate model  $m$  and  $M$  is the total number of climate models. The uncertainty of the emergent relationship for a value  $x$  that has not been used to fit the regression line is given by the standard prediction error (SPE)  $\sigma_{\hat{y}}(x)$ :

$$\sigma_{\hat{y}}^2(x) = s^2 \left[ 1 + \frac{1}{M} + \frac{(x - \bar{x})^2}{\sum_{m=1}^M (x_m - \bar{x})^2} \right]. \quad (2.17)$$

Here,  $\bar{x}$  indicates the arithmetic mean of  $x$  calculated over all climate models. Assuming Gaussian errors and a mean of  $\hat{y}(x)$  (i.e., the best estimate of the target variable  $y$  is given by the regression line), Equation (2.17) can be used to define a conditional probability density function (PDF) for predicting a target variable  $y$  given  $x$ :

$$P(y|x) = \frac{1}{\sqrt{2\pi\sigma_{\hat{y}}^2(x)}} \exp \left[ -\frac{(y - \hat{y}(x))^2}{2\sigma_{\hat{y}}^2(x)} \right]. \quad (2.18)$$

This distribution describes the uncertainty in the emergent relationship itself introduced by the imperfect alignment of the climate model data (red shaded area in Figure 2.12). Its maximum is given by the emergent relationship itself (red line in Figure 2.12). The conditional PDF can be interpreted as the posterior distribution of the regression model based on the climate model output but constrained on the observable  $x$ . However, the observed value of  $x$ , called  $x_0$ , also has uncertainties associated with it (blue shaded area in Figure 2.12). Assuming again a Gaussian distribution, the observational PDF for observing  $x_0$  given the true value  $x$  can be written as

$$P(x_0|x) = \frac{1}{\sqrt{2\pi\sigma_x^2}} \exp\left[-\frac{(x_0 - x)^2}{2\sigma_x^2}\right], \quad (2.19)$$

where  $\sigma_x$  is the standard deviation of the observation around the true value. Assuming an imperfect uniform prior  $P(x) \propto 1$  with cut-offs at  $-\infty$  and  $+\infty$  and using Bayes' theorem implies  $P(x_0|x) = P(x|x_0)$ . In a final step, this can be used to calculate the posterior PDF for the constrained prediction of the target variable  $y$  given the observation  $x_0$  (gray PDF in Figure 2.12) with numerical integration:

$$P(y|x_0) = \int_{-\infty}^{+\infty} P(y|x) P(x|x_0) dx. \quad (2.20)$$

Posterior estimates of the target variable are influenced by the way the statistical inference has been performed. Alternative methods that can be used include Bayesian frameworks (Renoult et al. 2020), information-theoretic approaches based on the Kullback-Leibler divergence between the models' PDFs of  $x$  and the observational PDF (Brient and Schneider 2016), and linear regression models based on hierarchical Bayesian models (Nijssse et al. 2020). However, no consensus has yet been found for this statistical inference (Brient 2020).

A convenient metric to quantify the skill of an emergent relationship is the coefficient of determination  $R^2$  of its underlying statistical model. In the presented framework which is based on univariate ordinary least squares regression,  $R^2$  is given by the squared Pearson correlation coefficient  $r$  evaluated on the climate model ensemble data  $\{(x_m, y_m) \mid m \in I_M\}$ , i.e.,  $R^2 = r^2$ . A further quantity describing the skill of an emergent relationship is its statistical significance. In the introduced framework, a two-sided  $t$ -test can be used to determine how likely the correlation found between the target variable  $y$  and the predictor  $x$  would be to appear by chance. The null hypothesis for this test is that the predictor and the target variable are not linearly correlated, i.e., that the true underlying Pearson correlation coefficient of the population is zero. If this null hypothesis is true, the probability distribution of the variable

$$t = \frac{r\sqrt{M-2}}{\sqrt{1-r^2}} \quad (2.21)$$

is a Student's  $t$ -distribution with  $M - 2$  degrees of freedom. The statistical significance can then be measured with the  $p$ -value of this two-sided  $t$ -test, which describes the probability of obtaining an absolute sample Pearson correlation coefficient greater than  $|r|$  if the null hypothesis is true. Smaller  $p$ -values indicate a higher statistical significance and vice versa.

One limitation of the presented framework is the assumption that the individual data points from the different climate models are independent. As already noted at the beginning of Section 2.4, this is not the case for typical climate model ensembles, as some modeling groups provide output for multiple climate models and some climate models from different modeling institutions share components and code (Knutti et al. 2013). The duplicated code in the different climate models leads to an overestimation of the sample size and may result in spurious correlations (Sanderson et al. 2015a). Possible approaches to tackle this problem are presented in Section 2.4.2 and include a weighting of the climate models based on their degree of interdependence (Knutti et al. 2017b; Sanderson et al. 2015a, 2017). A further limitation of this approach is the use of an ordinary least squares linear regression model. Sometimes, this is not appropriate; for example, when a non-linear emergent relationship is expected (Nijssse et al. 2020) or when additional physical considerations further constrain the regression model, e.g., by demanding a zero intercept ( $\hat{b}_0 = 0$ ) (Annan et al. 2020; Jimenez-de-la-Cuesta and Mauritsen 2019). Moreover, using only a single observational dataset to estimate  $x_0$  and  $\sigma_x$  when different datasets are available might lead to an underestimation of the observational uncertainty, as different observational datasets might lead to different emergent constraints.

A crucial aspect for every emergent constraint is a verifiable physical process explaining the correlation between  $x$  and  $y$  (Hall et al. 2019). Only if the underlying emergent relationship can be derived from a robust and plausible physical mechanism, an emergent constraint can be considered credible. The reason for this is the occurrence of spurious relationships: Due to the large number of possible observables provided by modern ESMs and the comparatively small number of climate models, spurious relationships are possible just by chance (Caldwell et al. 2014). Furthermore, out-of-sample tests offer an important tool to evaluate the credibility of emergent constraints (Hall et al. 2019). These ensure that the existence of an emergent relationship is not limited to a certain climate model ensemble and might indicate that the relationship is also valid for the true climate system. Testing emergent constraints in different CMIP generations offers a straightforward setup for out-of-sample testing (Caldwell et al. 2018), which is discussed in more detail in Chapter 5, where 11 emergent constraints on ECS are evaluated on the new CMIP6 ensemble.

In the last two decades, many emergent constraints on various aspects of the Earth system have been published. Early studies tackled the hydrological cycle (Allen and Ingram 2002) and the snow/ice-albedo feedback (Hall and Qu 2006). Over the years, climate sensitivity expressed by the ECS has been a prominent target variable. Since cloud feedbacks are a major source of uncertainty for climate sensitivity, a variety of scientific papers focus on constraining ECS with cloud-related processes (Brient et al. 2015; Brient and Schneider 2016; Fasullo and Trenberth 2012; Lipat et al. 2017; Qu et al. 2013; Sherwood et al. 2014; Su et al. 2014; Tian 2015; Volodin 2008; Zhai et al. 2015), which are discussed in detail in Section 5.2. More recent studies aim to constrain ECS with the historical temperature variability (Cox et al. 2018) or the historical warming trend (Jimenez-de-la-Cuesta and Mauritsen 2019; Nijssse et al. 2020; Tokarska et al. 2020). Emergent constraints are not only limited to physical processes but can also be applied to other domains, like the global carbon cycle (Cox et al. 2013; Kwiatkowski

et al. 2017; Wenzel et al. 2014, 2016a; Winkler et al. 2019a). An extensive discussion on the emergent constraint by Wenzel et al. (2016a), which focuses on the concentration-carbon feedback, is given in Section 6.1.1.

#### 2.4.2. Performance- and Interdependence-based Weighting of Climate Models

Further techniques to reduce uncertainties in climate model projections with observations are model weighting schemes. Their basic idea is to abandon model democracy by replacing the arithmetic mean used to calculate the MMMs by a weighted mean of the form

$$y = \sum_{m=1}^M w_m y_m \quad (2.22)$$

with normalized weights  $w_m$ . Similar to the notation introduced in the previous section,  $y$  is a target variable (e.g., a projection of the future climate) and  $m$  indexes the  $M$  different climate models. To address two major issues of model democracy (different climate models are not equally plausible and not independent), Knutti et al. (2017b) propose a weighting scheme based on climate model performance and interdependence with the following weights:

$$w_m \propto \frac{\exp\left(-\frac{D_m^2}{\sigma_D^2}\right)}{1 + \sum_{n \neq m}^M \exp\left(-\frac{S_{mn}^2}{\sigma_S^2}\right)}. \quad (2.23)$$

The metric  $D_m$  describes the distance between climate model  $m$  and observations (= model performance) and the metric  $S_{mn}$  describes the distance between climate model  $m$  and  $n$  (= model interdependence).  $\sigma_D$  and  $\sigma_S$  are constants that determine the individual strength of the performance and interdependence weighting, respectively.

A commonly used distance metric to measure model performance  $D_m$  and model interdependence  $S_{mn}$  is the root mean square error (RMSE), but others are possible (Knutti et al. 2017b). The metrics are evaluated on a set of past or present-day diagnostics and variables, whose choice is crucial for the weighting scheme. A helpful strategy for this is to focus on addressing the question “which climate model is adequate for predicting the target variable  $y$ ?” instead of trying to answer the question “which climate model is the best?” (Parker 2009). Thus, diagnostics and variables are chosen that are relevant for the projection of the target variable (Knutti et al. 2017b). In practice, this choice is either based on expert judgment about relevant processes, on emergent relationships (see Section 2.4.1), or on multivariate regression models (Karpechko et al. 2013; Sanderson et al. 2015b; see Section 2.4.3). It might also be beneficial to use different diagnostics for the calculations of the performance and interdependence metrics (Merrifield et al. 2020) and/or to remove selected diagnostics based on their mutual correlation (Lorenz et al. 2018).

The constants  $\sigma_D$  and  $\sigma_S$  determine how strongly the climate models’ performance and interdependence are weighted (Knutti et al. 2017b). Small values of the performance parameter  $\sigma_D$  lead to an aggressive weighting with only a few climate models receiving a majority of

the weight, while large values of  $\sigma_D$  result in an equal weighting. For the interdependence parameter  $\sigma_S$ , this is slightly different: Here, small (all climate models are independent) and large (all climate models are dependent) values lead to an almost equal weighting. Thus, an optimal choice for  $\sigma_D$  and  $\sigma_S$  is crucial. A useful tool to estimate these optimal parameters is the *leave-one-model-out cross-validation (CV)* approach, which is also known as *pseudo-reality, model-as-truth, or perfect model* setup (de Elia et al. 2002; Karpechko et al. 2013). For this, a single climate model is removed from the multi-model ensemble and treated as observation (*pseudo-observation*). Then, a weighted MMM with weights computed from the updated model ensemble is calculated, which gives a prediction for the “true” climate model. This allows a simple quantitative assessment of the weighting scheme by calculating the RMSE between the prediction and the known ground truth of the pseudo-observation, which is also known as *root mean square error of prediction (RMSEP)*. The whole process is repeated for every climate model of the ensemble to get a statistical distribution of RMSEPs. Finally, different RMSEP distributions calculated from different parameters  $\sigma_D$  and  $\sigma_S$  can be assessed using specific criteria to find optimal values for  $\sigma_D$  and  $\sigma_S$  (Knutti et al. 2017b). Furthermore, the leave-one-model-out CV approach can be used to evaluate different climate model weightings schemes (including the unweighted MMM) and compare them against each other.

The definition of the weights according to Equation (2.23) is based on reasonable and comprehensible principles. However, the exact form of the equation is purely subjective. Moreover, the additional freedom in choosing a suitable metric and optimal values for the parameters  $\sigma_D$  and  $\sigma_S$  adds another level of subjectivity to the weighting scheme, which can partly be addressed with the introduced leave-one-model-out CV setup. Nevertheless, due to its flexibility, the climate model weighting scheme of Knutti et al. (2017b) has already been used for various target variables: Arctic sea ice (Knutti et al. 2017b), Antarctic ozone concentrations (Amos et al. 2020), North American maximum temperature (Lorenz et al. 2018), European temperature and precipitation (Brunner et al. 2019; Merrifield et al. 2020), and global warming over the 21<sup>st</sup> century (Brunner et al. 2020; Liang et al. 2020).

### 2.4.3. Multiple Diagnostic Ensemble Regression

An alternative climate model weighting scheme is the *multiple diagnostic ensemble regression (MDER)* approach (Karpechko et al. 2013). Similar to all methods presented in Section 2.4, it can be used to reduce uncertainties in climate model projections with observations. The basis of MDER is a set of  $K$  predictor diagnostics  $\{x^{(1)}, x^{(2)}, \dots, x^{(K)}\}$  which are relevant for the projection of the target variable  $y$ . The reasoning for this choice of diagnostics is similar to the one presented in the previous section: Weighting schemes should address the question “which climate model is adequate for predicting the target variable?” and not “which climate model is the best?”.

The concept of MDER is mathematically similar to the concept of emergent constraints. In a first step, an inter-model relationship between the target variable and the process-relevant diagnostics is used to fit a multivariate linear regression model. Let  $\mathbf{y} = (y_1, y_2, \dots, y_M)^T \in \mathbb{R}^M$

be the vector of target variables of the  $M$  climate models ( $T$  denotes the transpose) and  $\mathbf{X} \in \mathbb{R}^{M \times (K+1)}$  the design matrix representing the predictors:

$$\mathbf{X} = \begin{pmatrix} 1 & x_1^{(1)} & x_1^{(2)} & \dots & x_1^{(K)} \\ 1 & x_2^{(1)} & x_2^{(2)} & \dots & x_2^{(K)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_M^{(1)} & x_M^{(2)} & \dots & x_M^{(K)} \end{pmatrix}. \quad (2.24)$$

The entry  $x_m^{(k)}$  of this matrix refers to the diagnostic variable of diagnostic  $k$  and climate model  $m$ . With this notation, the linear inter-model relationship can be written as

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \boldsymbol{\varepsilon}, \quad (2.25)$$

where  $\mathbf{b} = (b_0, b_1, \dots, b_K)^T \in \mathbb{R}^{(K+1)}$  is the vector of linear coefficients (with intercept  $b_0$ ) and  $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_M)^T \in \mathbb{R}^M$  a vector of independent random variables representing the noise in the target variable. Figure 2.13 shows a schematic that illustrates this linear relationship (gray surface) for the different climate models (colored points) for two diagnostics ( $K = 2$ ). Using ordinary least squares regression, the estimated linear coefficients  $\hat{\mathbf{b}}$  are given by

$$\hat{\mathbf{b}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}, \quad (2.26)$$

where the exponent  $-1$  denotes the inverse matrix. Since this definition works for any number of diagnostics  $K$ , it can also be used to calculate the linear coefficients  $\hat{b}_0$  (intercept) and  $\hat{b}_1$  (slope) that define emergent relationships (see Equation (2.15)).

In the second step of the algorithm, observed data of the process-based diagnostics  $\mathbf{x}_0 = (1, x_0^{(1)}, x_0^{(2)}, \dots, x_0^{(K)})^T \in \mathbb{R}^{(K+1)}$  is fed into the multivariate linear regression model to get an observation-based prediction of the target variable  $\hat{y}_0$  (Karpechko et al. 2013):

$$\hat{y}_0 = \mathbf{x}_0^T \hat{\mathbf{b}} \quad (2.27)$$

This is mathematically similar to the calculation of the best estimate target variable  $y$  for emergent constraints. In Figure 2.13, the observations of the predictors  $\mathbf{x}_0$  are illustrated with horizontal black lines, and the best estimate  $\hat{y}_0$  is shown as a black point. By combining Equations (2.26) and (2.27) and comparing this to the definition of weighted means in Equation (2.22), climate model weights  $\mathbf{w} = (w_1, w_2, \dots, w_M)^T \in \mathbb{R}^M$  can be defined by

$$\mathbf{w} = \left[ \mathbf{x}_0^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \right]^T, \quad (2.28)$$

which can be used to calculate the weighted target variable by

$$\hat{y}_0 = \hat{\mathbf{w}}^T \mathbf{y}. \quad (2.29)$$

A crucial aspect for the success of the MDER approach is the choice of the process-relevant diagnostics. In addition to the pre-selection based on expert judgment, an additional selection based on statistical criteria is necessary for two reasons: First, predictors which only show



Figure 2.13.: Schematic illustration of the multiple diagnostic ensemble regression (MDER) approach (Karpechko et al. 2013). First, inter-model relationships between a target variable  $y$  and multiple process-based predictors  $x^{(k)}$  (here: two predictors  $x^{(1)}$  and  $x^{(2)}$ ) are used to fit a multivariate linear regression model (gray surface). Second, observations of the predictors (horizontal black lines) are fed into the regression model to calculate an observation-based best estimate of the target variable  $\hat{y}_0$  following Equation (2.27). The black point indicates the best estimate for the target variable  $y$  given by the observed values of  $x^{(1)}$  and  $x^{(2)}$ . Each of the remaining colored points represents a single climate model of the multi-model ensemble. The vertical dashed lines visualize the distance between the climate model data and the linear regression surface, which represents the noise term  $\epsilon$  in Equation (2.25).

a weak correlation with the target variable should not be included in the regression model since they introduce additional noise and might lead to overconfident results. Second, multicollinearity (i.e., mutually correlated predictors) should be avoided since this reduces the robustness of the linear regression. A common technique to deal with these problems is a stepwise feature selection algorithm based on statistical tests of the correlations between the involved variables (Karpechko et al. 2013).

The basic assumption of the MDER algorithm is that the inter-model relationship between the process-based predictors and the target variable also holds for the true climate. This may seem weak at first glance, especially since it explicitly requires climate models that deviate from the observed climate to span the desired relationship (similar to emergent constraints). However, a much weaker assumption is made traditional in weighting approaches which assume that climate models that are better in simulating the past or present-day climate are necessarily better in simulating the future climate. In contrast to these other weighting approaches, MDER explicitly establishes the relationship between past/present and future within the climate model ensemble (Karpechko et al. 2013).

Drawbacks of the MDER approach are the missing consideration of errors in the observational data, the limitation to linear relationships between the process-relevant diagnostics

## *2. Scientific Background*

---

and the target variable, and the limitation to a single data point per climate model. Despite these, MDER has been successfully used to constrain uncertainties in Antarctic total ozone projections (Karpechko et al. 2013), in the projected change of the austral jet position (Wenzel et al. 2016b), and in projections of the Arctic sea ice extent (Senftleben et al. 2020).

### 3. Improving Routine Climate Model Evaluation

To address the key science questions of this thesis posed in Section 1.2, a reliable and efficient tool to read, process, and evaluate climate model output and observational data is necessary. A valuable software package that fits these criteria is the Earth System Model Evaluation Tool (ESMValTool). The ESMValTool is an open-source community diagnostics and performance metrics tool for the routine evaluation of ESM output, which notably facilitates the analysis of CMIP models (<https://www.esmvaltool.org>). For this reason, all analyses presented in Chapters 4–6 of this thesis have been implemented into the ESMValTool. Apart from that, further substantial changes and additions to the code base of the ESMValTool have been contributed that improve the routine evaluation of climate models, which is beneficial for the entire scientific community. This led to co-authorship in the scientific documentation of the ESMValTool, which is published in four peer-reviewed studies (Eyring et al. 2020; Lauer et al. 2020; Righi et al. 2020; Weigel et al. 2020). After a brief overview of the ESMValTool’s structure (Section 3.1), this chapter presents all these contributions that are not documented in other chapters of this thesis and puts them into context (Sections 3.2 and 3.3).

#### 3.1. The Earth System Model Evaluation Tool (ESMValTool)

As outlined in Section 2.1.1, climate models have been continuously improved and extended over the last decades from the relatively simple AOGCMs to the complex state-of-the-art ESMs that include an immense number of variables and processes. In CMIP6, more modeling institutes provide data for more versions of these complex models running simulations for more experiments. Consequently, the data volume of the entire CMIP6 archive is expected to reach up to 80 PB (Balaji et al. 2018), which is a vast increase in comparison to the 2 PB of CMIP5. The increasing complexity and data volume of the climate models pose a major challenge for the evaluation and analysis of the model output. To address this big data challenge and support the scientific community, the ESMValTool has been developed to provide an “open-source, standardized, community-based software package for the systematic, efficient, and well-documented analysis of ESM results” (Righi et al. 2020). Moreover, the ESMValTool allows a routine comparison of single or multiple climate models against predecessor versions and/or observations.

Since its first release in 2016 (Eyring et al. 2016b), the ESMValTool has been greatly extended and improved. A schematic representation of the current structure of the ESMValTool



Figure 3.1.: Schematic representation of the structure of the Earth System Model Evaluation Tool (ESMValTool). First, input data is preprocessed by the core functionalities of the ESMValTool (ESMValCore). Second, the preprocessed data is read by diagnostic scripts, which create the final output (e.g., plots). The user can specify the desired input datasets, preprocessing operations, and diagnostic scripts in the recipe, which is a configuration file that controls the main workflow of the ESMValTool. Adapted with permission from Righi et al. (2020).

is illustrated in Figure 3.1. The first major element of this workflow is an extensive preprocessing of the input data. For this, the Python-based core functionalities of the ESMValTool (ESMValCore) provide a set of operators (*preprocessors*) that are applied to the input data. There are two classes of preprocessors: non-optional and optional preprocessors. Examples for non-optional preprocessors are the input/output operations *load* (loading input data from climate models and observations) and *save* (saving the preprocessed data), which utilize the Network Common Data Form (NetCDF) format (a self-describing, machine-independent binary file format used for the storage of array-based scientific data). Further non-optional preprocessors include *checks* and *fixes* of the input data, which test whether the input data adheres to standards given by the Climate Model Output Rewriter (CMOR) format and fixes the data if necessary. The CMOR format ensures that the output from the many different modeling institutes within a generation of CMIP follows common standards. The optional preprocessors include commonly used operations on spatiotemporal datasets, like spatial and temporal *subsetting*, the calculation of *spatial and temporal statistics*, horizontal and vertical *in-*

terpolation, land/sea/ice masking, unit conversion, and the calculation of *multi-model statistics*. A further example is *variable derivation*, which can be used to derive non-CMOR variables from the input data. The preprocessing functions of ESMValCore aim to facilitate the routine evaluation of climate models by providing a set of commonly used data operations. To reduce computation times, ESMValCore allows parallel preprocessing of different datasets. However, since this leads to higher memory usage, a compromise has to be found in practice to optimize the use of time and memory resources.

The second main element of the ESMValTool workflow is the calculation of *diagnostics*. This task is performed by the diagnostic scripts, which can be written in multiple programming languages. Currently, the languages Python, NCL, R, and Julia are supported. The diagnostic scripts, which contain the code that runs the actual scientific evaluation, read the preprocessed datasets from ESMValCore and use these to create the final output of the tool. Apart from NetCDF files and plots, this includes log files with provenance information that ensure the reproducibility and transparency of the results. The whole workflow of the ESMValTool is controlled with the *recipe*, which is a configuration file in which the user can specify the desired input datasets, preprocessing operations, and diagnostic scripts. Implementing a new scientific assessment into the ESMValTool usually includes writing a new recipe and one or more diagnostic scripts. A third major element of the ESMValTool is the *CMORization* (i.e., the reformatting of data so it adheres to CMOR standards) of raw observational datasets (not shown in Figure 3.1). This ensures that the ESMValTool can process arbitrary observational datasets, which can be used, for example, to assess the skill of climate model simulations. More details on the structure of the ESMValTool are given in its extensive documentation (<https://docs.esmvaltool.org>).

To ensure a rapid evaluation of the CMIP6 models, the ESMValTool is fully integrated into the infrastructure of the Earth System Grid Federation (ESGF), which provides the CMIP model output for the general public (Eyring et al. 2016c). As soon as new model data is published on the ESGF servers, it can be accessed with the ESMValTool and analyzed. This instantaneous evaluation of the CMIP models is urgently needed since there is a growing dependency on CMIP products by a broad research community and by national and international climate assessments. For this reason, the ESMValTool is used to evaluate climate model output and observational data in several chapters of the upcoming Sixth Assessment Report (AR6) of the Intergovernmental Panel on Climate Change (IPCC).

The ESMValTool is developed open-source on GitHub, a web service that provides hosting for version control with *git* (<https://github.com/ESMValGroup>). It is released under the Apache License, Version 2.0. The source code of the latest released version of the ESMValCore package, which includes the core functionalities of the ESMValTool, is publicly available at Zenodo (Andela et al. 2020b). Similarly, the source code of the latest released version of the ESMValTool package, which includes all publicly available recipes and diagnostic scripts, is also available at Zenodo (Andela et al. 2020a).

## 3.2. Contributions to ESMValCore

As of April 2021, 41,780 lines of code have been added and 21,713 lines of code have been removed from the GitHub repository of ESMValCore by the author of this thesis. Apart from general improvements of the code base, these changes and additions mainly include new preprocessor functions that can be applied to the input data and derivation scripts for new non-CMOR variables. Table 3.1 shows a summary of these main contributions, which are partly published in the scientific documentation of ESMValCore (Righi et al. 2020).

In total, three new preprocessor functions have been added: *amplitude*, *land/sea fraction weighting*, and *trend*. The *amplitude* preprocessor calculates the peak-to-peak amplitude of periodic phenomena along an arbitrary coordinate, which is usually time. A common application of this preprocessor is the computation of a variable's diurnal or seasonal cycle amplitude. For instance, this is necessary to calculate the predictor variable (CO<sub>2</sub> seasonal cycle amplitude) of the emergent constraint on CO<sub>2</sub> fertilization by Wenzel et al. (2016a) that is also used in Chapter 6 to constrain the global mean GPP at the end of the 21<sup>st</sup> century. The *land/sea fraction weighting* preprocessor weights fields with the land or sea fraction of the respective grid cells. For example, this is necessary for the spatial integration of flux-related variables which are reported in units of "per square meter of land/sea" and not in "per square meter of grid cell". After the weighting, the grid-cell areas can be used to integrate the flux-related variable over any desired region. In this thesis, *land/sea fraction weighting* is applied to GPP fields given by observational products and models to compute the corresponding global mean results (see Chapter 6). The *trend* preprocessor calculates the linear trend of a variable along an arbitrary coordinate. The linear trend is defined as the slope of an ordinary least squares linear regression of the variable against the selected coordinate. For example, this can

Type	Name	Description
Preprocessor	<i>amplitude</i>	Amplitude of periodic phenomena (e.g., cycles)
	<i>land/sea fraction weighting</i>	Weighting of fields based on the land or sea fraction of the respective grid cells
	<i>trend</i>	Linear trend (slope of linear regression)
Derived variable	<i>asr</i>	Absorbed shortwave radiation
	<i>co2s</i>	Atmospheric CO <sub>2</sub> concentration at the surface
	<i>et</i>	Evapotranspiration
	<i>rlntcs</i>	Clear-sky net top of the atmosphere (TOA) longwave radiation
	<i>rsntcs</i>	Clear-sky net TOA shortwave radiation
	<i>uajet</i>	Position of austral jet stream

Table 3.1.: Summary of new preprocessor functions and variable derivation scripts contributed to the core functionalities of the ESMValTool (ESMValCore) by the author of this thesis.



Figure 3.2.: (a) CMIP6 multi-model mean (MMM) of the atmospheric CO<sub>2</sub> concentration at the surface (*co2s*) averaged over the months June-July-August (JJA) of the year 2014 in the emission-driven historical simulation. The CMIP6 MMM includes the four climate models that provide all data needed for the calculation of *co2s* (see legend in (b)). (b) Monthly mean *co2s* at Mauna Loa, Hawaii (19.5°N, 155.6°W) from 1958 to 2014. The black line shows observations from the Mauna Loa Observatory (Keeling et al. 2005); the remaining lines show emission-driven historical simulations from individual CMIP6 models.

be used to calculate the historical trend of the GSAT, which is used as a predictor for ECS in several recently published emergent constraints (Jimenez-de-la-Cuesta and Mauritsen 2019; Nijssse et al. 2020; Tokarska et al. 2020).

In addition to new preprocessor functions, six derivation scripts for the non-CMOR variables *asr*, *co2s*, *et*, *rlnacs*, *rsntcs*, and *uajet* have been implemented into the ESMValCore. Apart from the atmospheric CO<sub>2</sub> concentration at the surface (*co2s*), which is used in Chapter 6 to derive the predictor variable of the emergent constraint by Wenzel et al. (2016a), these variables are not directly used in this thesis but have been added to reproduce the MDER analysis of Wenzel et al. (2016b) (*asr* and *uajet*) or have been used in earlier versions of the studies presented in this thesis (*et* for the study presented in Chapter 6; *rlnacs* and *rsntcs* for the studies presented in Chapter 4).

The absorbed shortwave radiation (*asr*) is defined as the difference between the incoming TOA shortwave radiation and the outgoing TOA shortwave radiation. The atmospheric CO<sub>2</sub> concentration at the surface (*co2s*) can be calculated from the pressure level-dependent atmospheric CO<sub>2</sub> concentration and the surface air pressure using interpolation. An example illustrating *co2s* is given in Figure 3.2. Figure 3.2a shows the CMIP6 MMM of the global *co2s* for the months June-July-August (JJA) of the year 2014 in the emission-driven historical simulation. As expected, the map shows high atmospheric CO<sub>2</sub> concentrations over large metropolitan areas with high CO<sub>2</sub> emissions (e.g., over North America, the Arabian Peninsula, and East Asia). Moreover, regions with high photosynthetic activity (e.g., the boreal forests in the Northern Hemisphere summer and the tropical rainforests in South America) exhibit

smaller CO<sub>2</sub> concentrations. Figure 3.2b shows the monthly mean *co2s* at Mauna Loa, Hawaii from 1958 to 2014 for observations performed at the Mauna Loa Observatory (Keeling et al. 2005) (thick black line) and emission-driven historical simulations from four CMIP6 models (remaining colored lines). Apart from the model CNRM-ESM2-1, the simulated atmospheric CO<sub>2</sub> concentrations show the expected seasonal cycle present in the Keeling Curve (see Figure 2.9). All ESMs correctly simulate the increase in *co2s* over the years. However, there are some differences in the absolute values with the model BCC-CSM2-MR showing the largest deviations.

The evapotranspiration (*et*) defined as the sum of evaporation and plant transpiration can be calculated from the surface latent heat flux and the latent heat vaporization constant. The longwave/shortwave clear-sky net TOA radiation (*rlntcs/rsntcs*) is defined as the difference between the longwave/shortwave incoming TOA radiation assuming clear-sky and the longwave/shortwave outgoing TOA radiation assuming clear-sky. Finally, the position of the austral jet stream (*uajet*) is given by the latitude with maximum zonal mean eastward wind speed at 850 hPa on the Southern Hemisphere between 80 °S and 30 °S.

### 3.3. Contributions to ESMValTool

In addition to the analyses presented in Chapters 4–6, several more recipes and CMORization scripts for observational data have been added to the main ESMValTool repository on GitHub. A summary of these is given in Table 3.2. Moreover, multiple minor changes and additions have been implemented to improve the code base and/or to fix bugs. In total, the author of this study has added 98,662 lines of code and removed 49,823 lines of code from the ESMValTool repository (as of April 2021). Parts of these implementations are already published in the scientific documentation of the ESMValTool, which covers large-scale diagnostics (Eyring et al. 2020), diagnostics for emergent constraints and future projections (Lauer et al. 2020), and diagnostics for extreme events and regional evaluation (Weigel et al. 2020).

Due to an exhaustive overhaul of the ESMValTool Version 2, older recipes from Version 1 (Eyring et al. 2016b) cannot be used directly in the latest release. Porting recipes from the old version to the current version requires substantial changes on the recipes themselves, but also on the corresponding diagnostic scripts. As of April 2021, the author of this thesis has ported two recipes originally written by other authors to the new version of ESMValTool: *recipe\_anav13jclim.yml* and *recipe\_wenzel16jclim.yml*. *recipe\_anav13jclim.yml* reproduces the analysis of Anav et al. (2013), who evaluate carbon cycle-related variables for ESMs of the CMIP5 ensemble. This includes climatologies, trends, and variabilities of important quantities of the terrestrial and oceanic carbon cycle like carbon fluxes, carbon reservoir sizes, and vegetation distributions. *recipe\_wenzel16jclim.yml* includes the MDER analysis performed by Wenzel et al. (2016b) to constrain future changes in the position of the austral jet stream. Due to its flexible code, the corresponding diagnostic scripts can be used to apply the MDER method to arbitrary target variables and predictors. Both recipes include analyses that contributed ideas and methods for the study presented in Chapter 6, in which the

Type	Name	Main reference
Recipe	<i>bock20jgr</i> (Figures 8 and 10)	(Bock et al. 2020)
	<i>cox18nature</i>	(Cox et al. 2018)
	<i>ecs</i>	(Gregory et al. 2004)
	<i>flato13ipcc</i> (Figure 9.42)	(Flato et al. 2013)
	<i>meehl20sciadv</i>	(Meehl et al. 2020)
	<i>schlund20esd</i>	(Schlund et al. 2020b)
	<i>schlund20jgr_gpp_abs_rcp85</i>	(Schlund et al. 2020a)
	<i>schlund20jgr_gpp_change_1pct</i>	(Schlund et al. 2020a)
	<i>schlund20jgr_gpp_change_rcp85</i>	(Schlund et al. 2020a)
	<i>tcr</i>	(Gregory and Forster 2008)
CMORizer	CRU	(Harris et al. 2014)
	CT2019	(Jacobson et al. 2020)
	GCP	(Friedlingstein et al. 2020)
	HWSD	(Wieder et al. 2014)
	JMA-TRANSCOM	(Maki et al. 2010)
	LAI3g	(Zhu et al. 2013)
	LandFlux-EVAL	(Mueller et al. 2013)
	MLS-Aura	(Read and Livesey 2015)
	FLUXNET-MTE	(Jung et al. 2011)
	NDP	(Gibbs 2006)
	Scripps-CO <sub>2</sub>	(Keeling et al. 2005)

Table 3.2.: Summary of new recipes and CMORization scripts for observational data contributed to the Earth System Model Evaluation Tool (ESMValTool) by the author of this thesis. The corresponding file names of the recipes in the ESMValTool repository are given by *recipe\_(name\_in\_table).yml*.

MDER method is expanded with a machine learning (ML) approach and applied to a carbon cycle-related target variable (GPP).

In addition to the ported recipes, several new recipes and diagnostic scripts have been contributed to the public ESMValTool repository (see Table 3.2 for a summary). Some of these include the analyses performed in other chapters of this thesis, namely *recipe\_bock20jgr.yml* (Chapter 4), *recipe\_meehl20sciadv.yml* (Chapter 4), *recipe\_schlund20esd.yml* (Chapter 5), and *recipe\_schlund20jgr\_\*.yml* (Chapter 6). *recipe\_ecs.yml* and *recipe\_tcr.yml* contain diagnostic scripts that allow the calculation of the climate metrics ECS and TCR for arbitrary model output (see Section 2.2), which is also used in Chapters 4 and 5. Corresponding plots are shown in Figures 2.5 and 2.7. *recipe\_cox18nature.yml* reproduces the analysis of Cox et al. (2018), who introduce an emergent constraint in ECS based on a global temperature variability metric  $\Psi$ . Figure 3.3a shows  $\Psi$  over the past 85 years for CMIP5 models and observations from HadCRUT4 (Morice et al. 2012). Since high-sensitivity climate models exhibit higher values of  $\Psi$  than low-sensitivity models, an emergent relationship between ECS and  $\Psi$  can be established (Figure 3.3b). With observations of  $\Psi$ , ECS can eventually be constrained to  $(2.8 \pm 0.6)$  K (66 % confidence range). The emergent constraint by Cox et al. (2018) is, together



Figure 3.3.: Emergent constraint on the effective climate sensitivity (ECS) from global temperature variability (Cox et al. 2018). Magenta colors indicate high-sensitivity climate models ( $|\lambda| < 1 \text{ Wm}^{-2}\text{K}^{-1}$ ), and green colors indicate low-sensitivity climate models ( $|\lambda| > 1 \text{ Wm}^{-2}\text{K}^{-1}$ ). Both panels have been created with the ESMValTool using *recipe\_cox18nature.yml* and are similar to Cox et al. (2018). (a) Temperature variability metric  $\Psi$  versus time for historical simulations from CMIP5 models (colored lines) and observations from the HadCRUT4 dataset (black points) (Morice et al. 2012). (b) Emergent relationship between ECS and the temperature variability metric  $\Psi$ . The dot-dashed black line shows the linear regression across the climate model ensemble with its associated uncertainties indicated by the dashed black lines (standard prediction error; see Equation (2.17)). Blue lines show the observational constraint from HadCRUT4 with its best estimate (dot-dashed line) and standard deviation (dashed lines). The letters represent individual CMIP5 models (see Cox et al. (2018) for details).

with 10 other emergent constraints on ECS, analyzed in great detail for the CMIP5 and CMIP6 ensemble in Chapter 5, which also includes a description of the calculation of  $\Psi$  in Section 5.2.2.

An important class of recipes in the ESMValTool covers international climate assessments. One such example is *recipe\_flato13ipcc.yml*, which reproduces large parts of the climate model evaluation performed in Chapter 9 of the IPCC's Fifth Assessment Report (AR5) (Flato et al. 2013). For this recipe, the Figure 9.42 of Flato et al. (2013) has been added, which shows the relationship between the historical and pre-industrial GSAT and ECS (Figure 3.4a) and the relationship between TCR and ECS (Figure 3.4b) within the CMIP5 ensemble. Figure 3.4a illustrates that there is no clear connection between ECS and the historical or pre-industrial GSAT, i.e., models with high temperatures in the historical or pre-industrial period do not necessarily exhibit a high equilibrium warming in the future or vice versa. On the other hand, TCR and ECS are well correlated within the CMIP5 ensemble as theoretically expected from Equation (2.13). However, this figure only shows a linear regression fit between the two variables and not the expected reciprocal relation. The output of *recipe\_flato13ipcc.yml* provided useful input for the analysis of climate sensitivity in CMIP6 performed in Meehl et al. (2020), which is also presented in Chapter 4.

The implemented CMORization scripts (also referred to as CMORizers) allow the processing of arbitrary observational data with the ESMValTool. For this, the raw observational



Figure 3.4.: Relationships between climate metrics within the CMIP5 climate model ensemble.

Both panels have been created with the ESMValTool using `recipe_f1ato13ipcc.yml` and are similar to Flato et al. (2013). (a) Global mean near-surface air temperature (GSAT) versus the effective climate sensitivity (ECS) for CMIP5 models for the period 1961–1990 (larger symbols) and for pre-industrial control runs (smaller symbols). (b) Transient climate response (TCR) versus ECS for CMIP5 models. The black line shows a linear fit between TCR and ECS.

datasets are reformatted in such a way that they adhere to the CMOR standards. Since the CMORization scripts are publicly available like any other code in the ESMValTool and ESMValCore repositories, the entire scientific community can use them to CMORize the observational data. A complete list of all 11 newly implemented CMORizers including the main reference for the corresponding observational datasets is shown in Table 3.2. The majority of these datasets are used for analyses presented in this thesis (e.g., see Tables 5.1 and 6.1).



## 4. Assessment of Climate Sensitivity in the CMIP6 Ensemble

To reduce uncertainties in multi-model climate projections, a first important step is the assessment of the desired target variables in the corresponding climate model ensemble. This is particularly relevant when a new generation of ESMs is published that includes considerable modifications compared to the respective predecessor model versions. Currently, we are in a situation like this with the CMIP6 ensemble, for which new data is still released every day (as of April 2021). In light of the upcoming international climate assessment of the IPCC's AR6, the evaluation of policy-relevant metrics like ECS or TCR for these new CMIP6 models is crucial since they provide vital information about the future climate of the planet Earth. As the ESMValTool is an excellent tool that allows for a quick and robust evaluation of new CMIP data as soon as it gets published on the ESGF servers (see Section 3.1), it is used here to assess and analyze ECS and TCR of the CMIP6 models and compare it to corresponding results of predecessor model generations and international climate assessments. This work is already published in two scientific papers: Bock et al. (2020) and Meehl et al. (2020). For Bock et al. (2020), the author of this thesis contributed their Figure 8 (a bar chart showing the ECS for several CMIP3, CMIP5, and CMIP6 models), their Figure 10 (map plots showing cloud-related feedback parameters for the CMIP5 and CMIP6 MMMs; see Figure 4.4), code to create these figures with the ESMValTool, and text for the manuscript (in particular a section about the evaluation of ECS). For Meehl et al. (2020), the author of this thesis contributed all figures and tables, code to produce these with the ESMValTool, and text for the manuscript. This chapter first presents the evaluation of climate sensitivity for the CMIP5 and CMIP6 ensembles (Section 4.1) and puts the corresponding ECS and TCR values into historical context (Section 4.2). Finally, possible reasons for the apparent increase of climate sensitivity in the CMIP6 models are discussed in detail (Section 4.3).

### 4.1. Evaluation of ECS and TCR in CMIP5 and CMIP6

Following Section 2.2.3, ECS is calculated with the Gregory regression method using GSAT and TOA net radiation data for 150 years of a 4xCO<sub>2</sub> simulation (Gregory et al. 2004). This calculation is illustrated in Figure 4.1 for the CMIP6 MMM, which yields an ECS of 3.74 K when all 150 years of the run are used. However, similar to the CMIP5 ensemble, the exact value of ECS depends on the years considered in the Gregory regression (see Figure 2.6). Using only the first 20 years of the simulation gives a significantly lower ECS of 3.31 K than



Figure 4.1.: Gregory regression for the CMIP6 multi-model mean (MMM) following Gregory et al. (2004): Global and annual mean net top of the atmosphere radiation  $N$  versus the change in global and annual mean near-surface air temperature  $\Delta T$  for 150 years of a simulation with an abrupt quadrupling of the atmospheric CO<sub>2</sub> concentration (4xCO<sub>2</sub>) for the CMIP6 MMM (points). To account for energy leakage and model drift, a linear fit of the corresponding pre-industrial control run is subtracted from the 4xCO<sub>2</sub> simulation. A linear regression can be used to calculate the effective climate sensitivity (ECS) according to Equation (2.10) with the radiative forcing  $F_{4x}$  ( $y$ -intercept) and the climate feedback parameter  $\lambda$  (slope) for all 150 years (black line), which results in ECS = 3.74 K. Due to non-linear effects, the exact value of ECS depends on the years considered for the Gregory regression (see Section 2.2.3 and Figure 2.6): Using only the first 20 years of the simulation (blue points and line) yields a considerably smaller ECS than using only the last 130 years of the simulation (orange points and line). Adapted with permission from Meehl et al. (2020).

using only the last 130 years of the simulation, which gives an ECS of 4.05 K. As thoroughly described in Section 2.2.3, the reason for this is the state and time dependence of the climate feedback parameter, which is given by the slope of the Gregory regression line. Due to non-linear effects in the feedbacks, this slope changes over time, resulting in lower (higher) values of ECS when only early (late) years of the simulation are considered. In contrast to that, the transient response of the climate system TCR is calculated from GSAT data at the time of CO<sub>2</sub> doubling in a 1%CO<sub>2</sub> run (Bindoff et al. 2013; see Section 2.2.5). Using these two definitions and the ESMValTool, ECS and TCR are assessed for all CMIP5 and CMIP6 models where the necessary temperature and radiation data are available. Tables 4.1 and 4.2 give an overview of the results for the CMIP5 and CMIP6 models, respectively.

The first striking feature of these two tables is the increased MMM of ECS and TCR in CMIP6. For ECS, the CMIP6 MMM is about 16 % (0.51 K) higher than the corresponding CMIP5 MMM. For TCR, the relative difference between the two model ensembles is notably

Model	Index used in plots	ECS [K]	TCR [K]
ACCESS1-0	1	3.83	1.92
ACCESS1-3	2	3.53	1.63
BNU-ESM	3	3.92	2.49
CCSM4	4	2.94	1.67
CNRM-CM5	5	3.25	2.02
CNRM-CM5-2	6	3.44	1.79
CSIRO-Mk3-6-0	7	4.08	1.75
CanESM2	8	3.69	2.34
FGOALS-g2	9	3.38	1.38
GFDL-CM3	10	3.97	1.94
GFDL-ESM2G	11	2.39	1.12
GFDL-ESM2M	12	2.44	1.36
GISS-E2-H	13	2.31	1.74
GISS-E2-R	14	2.11	1.46
HadGEM2-ES	15	4.61	2.54
IPSL-CM5A-LR	16	4.13	2.00
IPSL-CM5A-MR	17	4.12	1.99
IPSL-CM5B-LR	18	2.60	1.53
MIROC-ESM	19	4.67	2.21
MIROC5	20	2.72	1.43
MPI-ESM-LR	21	3.63	2.03
MPI-ESM-MR	22	3.46	2.03
MPI-ESM-P	23	3.45	2.02
MRI-CGCM3	24	2.60	1.56
NorESM1-M	25	2.80	1.41
bcc-csm1-1	26	2.83	1.73
bcc-csm1-1-m	27	2.86	2.12
inmcm4	28	2.08	1.32
Multi-model mean		3.23	1.81
Multi-model standard deviation		0.72	0.36

Table 4.1.: Effective climate sensitivity (ECS) and transient climate response (TCR) calculated for the CMIP5 models. Details on this are given in Section 2.2.3 and Section 2.2.5, respectively. The multi-model mean (MMM) is calculated from the Gregory regression method using the MMM net top of the atmosphere (TOA) radiation and the MMM change in global mean near-surface air temperature (GSAT) similar to Figure 4.1. The multi-model standard deviation is given by the sample standard deviation of ECS calculated over all climate models (using the normalization  $1/M$ , where  $M$  is the number of models). Corresponding references for each model are given in Table A.3. Adapted with permission from Meehl et al. (2020).

smaller with about 10 % (0.19 K). The multi-model spread (expressed as the multi-model standard deviation) shows an even larger increase in the CMIP6 ensemble: For ECS, the relative difference is about 49 % (0.72 K in CMIP5 to 1.07 K in CMIP6), and for TCR, the relative difference is about 17 % (0.36 K in CMIP5 to 0.42 K in CMIP6). The main reason for the increased MMM and spread in the CMIP6 ensemble is the existence of several models

Model	Index used in plots	ECS [K]	TCR [K]
ACCESS-CM2	29	4.72	2.10
ACCESS-ESM1-5	30	3.87	1.95
AWI-CM-1-1-MR	31	3.16	2.06
BCC-CSM2-MR	32	3.04	1.72
BCC-ESM1	33	3.26	1.77
CAMS-CSM1-0	34	2.29	1.73
CAS-ESM2-0	35	3.51	2.04
CESM2	36	5.16	2.06
CESM2-FV2	37	5.14	2.05
CESM2-WACCM	38	4.75	1.98
CESM2-WACCM-FV2	39	4.79	2.01
CMCC-CM2-SR5	40	3.52	2.09
CNRM-CM6-1	41	4.83	2.14
CNRM-CM6-1-HR	42	4.28	2.48
CNRM-ESM2-1	43	4.76	1.86
CanESM5	44	5.62	2.74
E3SM-1-0	45	5.32	2.99
EC-Earth3-Veg	46	4.31	2.62
FGOALS-f3-L	47	3.00	1.94
FGOALS-g3	48	2.88	
GISS-E2-1-G	49	2.72	1.79
GISS-E2-1-H	50	3.11	1.93
HadGEM3-GC31-LL	51	5.55	2.55
HadGEM3-GC31-MM	52	5.42	2.58
INM-CM4-8	53	1.83	1.33
INM-CM5-0	54	1.92	1.37
IPSL-CM6A-LR	55	4.56	2.32
KACE-1-0-G	56	4.48	1.41
MCM-UA-1-0	57	3.65	1.94
MIROC-ES2L	58	2.68	1.55
MIROC6	59	2.61	1.55
MPI-ESM-1-2-HAM	60	2.96	1.80
MPI-ESM1-2-HR	61	2.98	1.66
MPI-ESM1-2-LR	62	3.00	1.84
MRI-ESM2-0	63	3.15	1.64
NESM3	64	4.72	2.72
NorCPM1	65	3.05	1.56
NorESM2-LM	66	2.54	1.48
NorESM2-MM	67	2.50	1.33
SAM0-UNICON	68	3.72	2.27
TaiESM1	69	4.31	2.34
UKESM1-0-LL	70	5.34	2.79
Multi-model mean		3.74	2.00
Multi-model standard deviation		1.07	0.42

Table 4.2.: As in Table 4.1 but for the CMIP6 models. Corresponding references for each model are given in Table A.4. Adapted with permission from Meehl et al. (2020).

with very high values of ECS and TCR. In addition to that, there also exist some models with very low values of ECS. Therefore, the CMIP6 model ranges of ECS and TCR are well outside the corresponding assessed ranges given by the latest published Assessment Report (AR) of the IPCC from 2013 (Stocker et al. 2013) with 1.8–5.6 K for ECS (AR5: 1.5–4.5 K) and 1.3–3.0 K (AR5: 1.0–2.5 K) for TCR.

## 4.2. Comparison to Previous CMIP Generations and International Climate Assessments

To illustrate the results of the previous section and to put them into historical context, Figure 4.2 shows the assessed ECS and TCR ranges over the years from the Charney report (Charney et al. 1979) and the different ARs of the IPCC in combination with the corresponding modeled ranges from the different CMIP generations. Since the Charney report in 1979, the assessed range of ECS of 1.5–4.5 K has almost remained unchanged during the last 40 years (Albritton et al. 2001; Charney et al. 1979; Kattenberg et al. 1996; Mitchell et al. 1990; Stocker et al. 2013), with the exception of AR4, where the lower bound was temporarily increased to 2.0 K (Solomon et al. 2007). For AR2, AR4, and AR5, the corresponding climate model generations from CMIP1, CMIP3, and CMIP5 more or less agree with this assessed range (Flato et al. 2013; Kattenberg et al. 1996; Randall et al. 2007). In contrast to that, some of the early climate models used in AR1 and some CMIP2 models used in AR3 exhibit values of ECS well above 4.5 K, resulting in an upper model range of about 0.6 K higher than the assessed range (Cubasch et al. 2001; Mitchell et al. 1990). However, these deviations from the assessed range are small compared to the ones present in the CMIP6 ensemble, which have become evident over the course of 2020. As shown in Figure 4.2, both the lower and upper bound are more extreme than in any previous climate model generation, resulting in a CMIP6 ECS model range of 1.8–5.6 K. This increase in range is particularly relevant for the upper bound, which is about 1.1 K higher than the assessed upper bound of 4.5 K. Moreover, it is not only a small fraction of models that exceed this upper bound: out of the 42 analyzed CMIP6 models, a third (14 models) have an ECS above 4.5 K. In addition to the analysis presented here, several modeling centers independently confirmed high ECS values in their models (Andrews et al. 2019; Bodas-Salcedo et al. 2019; Gettelman et al. 2019a; Wyser et al. 2020). A list of possible reasons for the increased climate sensitivity in the CMIP6 ensemble including aerosol-cloud interactions and changes in the shortwave cloud feedback over the Southern Ocean is given in Section 4.3.

In contrast to ECS, TCR has only been evaluated since AR1 in 1990, in which it is estimated with 2.3 K using only a single climate model (F. P. Bretherton et al. 1990; see single green point in Figure 4.2). Over the years, the modeled range of TCR has decreased from about 1.3–3.8 K in the CMIP1 models (Kattenberg et al. 1996) to about 1.1–2.6 K in the CMIP5 ensemble (Flato et al. 2013). Similarly, the assessed range of TCR, which was first quoted in AR4 with 1.0–3.0 K (Solomon et al. 2007), has been reduced to 1.0–2.5 K in the subsequent AR5 (Stocker et al. 2013). While the corresponding climate model ranges from CMIP3 and CMIP5 agree with these



Figure 4.2.: Assessed values of the effective climate sensitivity (ECS) (blue bars) and the transient climate response (TCR) (red bars) over the years in the Charney report from 1979 (Charney et al. 1979) and the subsequent Assessment Reports (ARs) of the Intergovernmental Panel on Climate Change (IPCC) (Albritton et al. 2001; Kattenberg et al. 1996; Mitchell et al. 1990; Solomon et al. 2007; Stocker et al. 2013). Orange and green bars represent the modeled ranges of ECS and TCR, respectively, from the different climate model ensembles of the Coupled Model Intercomparison Project (CMIP) with the corresponding multi-model mean (MMM) illustrated with the horizontal black lines (the green point represents the output from a single climate model). The numbers correspond to individual CMIP5 and CMIP6 models (see Tables 4.1 and 4.2). Adapted with permission from Meehl et al. (2020).

assessed ranges (Flato et al. 2013; Randall et al. 2007), the upper bound of the CMIP6 model range (1.3–3.0 K) is well above the assessed upper bound of 2.5 K given by AR5. However, unlike ECS, also the lower bound of the CMIP6 model range increased compared to the CMIP5 ensemble, leading to a similar TCR model spread in both climate model generations. Due to the correlation of TCR and ECS given by Equation (2.13), the increase of TCR in many CMIP6 models is not surprising in light of the many high ECS models that are present in the CMIP6 ensemble.

### 4.3. Possible Reasons for High Climate Sensitivity in CMIP6

Because of the massive political and societal relevance of ECS, its apparent increase in the CMIP6 ensemble is currently one of the most important questions for the entire climate modeling community. For this reason, this section discusses possible reasons for the increased climate sensitivity in CMIP6.

Various improvements of the underlying physical, biological, and chemical processes have been introduced to the CMIP6 ESMs to represent the coupled Earth system in more detail. Since many of these processes directly influence the models' ECS (Forster et al. 2020), determining reasons for the increased ECS in the CMIP6 ensemble is highly non-trivial. As discussed in detail by Meehl et al. (2020), one possible reason is the consideration of aerosol-cloud interactions. New prognostic aerosol schemes added to some CMIP6 models that include aerosol-cloud interactions could have produced overly large negative radiative forcing, which then required a stronger model response to GHG forcing to reproduce the observed historical temperature trend. In fact, Meehl et al. (2020) find a weak inter-model relationship between the aerosol forcing and ECS for some CMIP6 models, with larger negative present-day aerosol forcing associated with higher ECS. Such relationships have also been found for previous climate model generations (Forster et al. 2013; Kiehl 2007). However, due to the varying aerosol forcing over time, these relations cannot be used to tune ECS based on the models' responses to aerosols over the different periods of the 20<sup>th</sup> century (Dittus et al. 2020). Further aerosol-cloud interactions that are relevant in the context of high climate sensitivity in CMIP6 can be found, for example, in the Hadley Centre Global Environmental Model (HadGEM) Version 3. The new aerosol scheme in this model suppresses a strong negative cloud feedback over the Southern Ocean that is found in the model's predecessor versions (Bodas-Salcedo et al. 2019). The Southern Ocean is one of the most pristine aerosol regions in the world (Hamilton et al. 2014), where aerosols predominantly originate from natural sources like marine emissions of gaseous dimethyl sulfide. Earlier HadGEM versions show a large reduction in cloud droplet size with warming over the Southern Ocean, which increases the cloud albedo and thus produces a negative shortwave cloud feedback. On the contrary, in HadGEM3, this negative feedback is completely suppressed or even positive due to a small increase of the cloud droplet size with warming, which eventually increases ECS due to an increased net cloud feedback (Bodas-Salcedo et al. 2019).

Another possible reason for the high climate sensitivity in the CMIP6 ensemble is related to cloud feedbacks (Bock et al. 2020). As already discussed in Section 2.2.1, uncertainties in cloud feedbacks are a major source of uncertainty of ECS in modern ESMs (Boucher et al. 2013; Flato et al. 2013). Thus, changes in processes that are related to cloud feedbacks immediately impact the climate models' climate sensitivities. The particular feedback mechanism that is thought to be relevant for the high ECS in CMIP6 is connected to cloud phase changes over the Southern Ocean that are present in earlier CMIP generations (e.g., CMIP5). This so-called *cloud phase change feedback* is illustrated in Figure 4.3 (green arrows). As the climate warms, the predominantly ice clouds over the Southern Ocean in these climate models become liquid clouds (McCoy et al. 2015). As a result, the cloud will get more reflective to the incoming solar radiation since a cloud consisting of smaller liquid droplets reflects more sunlight than a cloud consisting of larger ice crystals (for fixed water content), which has a cooling effect on the climate system. In addition, predominantly liquid clouds tend to precipitate less than mixed-phase clouds (composed of liquid water and ice). This leads to a higher liquid water content of the cloud and a further amplification of the cooling effect. Overall, this forms a



Figure 4.3.: Schematic illustration of the strong negative shortwave cloud phase change feedback over the Southern Ocean, which is present in earlier CMIP model generations like CMIP5 (green arrows). Due to improvements in the microphysical representation of mixed-phase clouds, some CMIP6 models simulate a higher fraction of (supercooled) liquid water in the clouds over the present-day Southern Ocean than their corresponding predecessor models (Bodas-Salcedo et al. 2019; Gettelman et al. 2019a). Thus, in the affected CMIP6 models, the strong negative cloud feedback that results from a cloud phase change from ice clouds in the present day to liquid clouds in the future is substantially reduced (orange arrows). This leads to an increase in the global net cloud feedback (Bodas-Salcedo et al. 2019; Tan et al. 2016) and the effective climate sensitivity (ECS) (Andrews et al. 2019; Gettelman et al. 2019a), which is a possible explanation for the high ECS values found in some CMIP6 models.

strong negative shortwave cloud feedback over the Southern Ocean in climate models from previous CMIP generations.

However, it is well known that the low-level mixed-phase clouds over the Southern Ocean in these earlier climate models are biased towards too large amounts of ice crystals and too little amounts of supercooled liquid water when compared to satellite observations (Bodas-Salcedo et al. 2016). Thus, the size of the resulting strong negative cloud phase change feedback in these models has long been questioned (McCoy et al. 2015; Tan et al. 2016). Improvements in the representation of mixed-phase clouds are known to reduce this long-standing cloud phase bias over the Southern Ocean (Bodas-Salcedo et al. 2016; McCoy et al. 2016) and to improve the representation of both cloud microphysical structure and cloud radiative impacts in this region (Hyder et al. 2018; Kay et al. 2016). For this reason, the microphysical representation of mixed-phase clouds has been updated in some CMIP6 models. For example, in CESM2, a new mixed-phase ice nucleation scheme by Hoose et al. (2010) replaces the Meyers et al. (1992) empirical scheme used in predecessor models (Gettelman et al. 2019a). Simulations with the Meyers et al. (1992) scheme show almost no supercooled liquid water in high-latitude clouds and a strong negative cloud phase change feedback when ice turns to liquid. On the contrary, simulations with the new ice nucleation scheme by Hoose et al. (2010) show a strong reduction

in this feedback. Since the new scheme also reduces significant cloud phase biases over the Southern Ocean that are present in the old scheme, the reduced negative cloud phase change feedback is thought to be more consistent with observations of the current climate (Gettelman et al. 2019a). Another example is the HadGEM3 model, for which a new cloud scheme that accounts for the turbulent production of liquid water in mixed-phase clouds has been implemented (Bodas-Salcedo et al. 2019). The new scheme parameterizes the role of subgrid-scale turbulence in the production and maintenance of supercooled liquid water (Furtado et al. 2016). As a consequence, the cloud liquid water path over the Southern Ocean is increased, which is more in line with observations from microwave satellite retrievals. Moreover, the increase in the cloud liquid water path with warming is reduced in HadGEM3, leading to a reduction in the negative cloud phase change feedback over the Southern Ocean (Bodas-Salcedo et al. 2019). In addition, Williams et al. (2020) show that short-term forecasts created with HadGEM3 are more accurate than comparable simulations from predecessor model versions (using traditional metrics from numerical weather prediction), which indicates an improvement of the physical processes in the new climate model.

Overall, such changes in the CMIP6 models contribute to a better agreement of the CMIP6 MMM of the TOA shortwave CRE with observations resulting in higher average pattern correlations of the model output with satellite observations (Bock et al. 2020). Moreover, the strong negative cloud feedback described above that results from a cloud phase change from ice clouds in the present day to liquid clouds in the future is substantially reduced in those CMIP6 models that simulate a higher fraction of (supercooled) liquid water in the clouds over the present-day Southern Ocean than their corresponding CMIP5 predecessors. This is illustrated in Figure 4.3 (orange arrows). The reduction of the negative shortwave cloud phase change feedback over the Southern Ocean increases the global net cloud feedback (Bodas-Salcedo et al. 2019; Tan et al. 2016) and eventually the ECS of the affected CMIP6 models (Andrews et al. 2019; Gettelman et al. 2019a), offering a possible explanation for the high ECS values found in some CMIP6 models.

To further confirm this hypothesis, the global distributions of relevant cloud feedback parameters (expressed by the net, shortwave, and longwave CRE feedback parameters; see Section 2.2.4) are shown for the CMIP6 MMM and its difference to the CMIP5 MMM in Figure 4.4. First of all, the comparison of Figures 4.4a, 4.4c and 4.4e demonstrates that the shortwave component dominates the net CRE feedback parameter over large swaths of the globe in the CMIP6 MMM. The sign change at approximately 60 °S in the shortwave CRE feedback parameter (see Figure 4.4c) roughly shows where the climate models simulate a high fraction of ice clouds (south of 60 °S) for which possible phase changes under global warming are potentially important. Such possible phase changes (from ice to liquid) contribute to a negative shortwave cloud feedback in this region as liquid clouds tend to be more reflective than ice clouds with a similar cloud water content. Figures 4.4b and 4.4d show a higher shortwave and net CRE feedback parameter over the Southern Ocean in CMIP6 compared to CMIP5. This is consistent with the hypothesis presented in the paragraph above, which claims that the reduced negative cloud phase change feedback over the Southern Ocean in



Figure 4.4.: Geographical distributions of the net (a), shortwave (c), and longwave (e) cloud radiative effect (CRE) feedback parameters for the CMIP6 multi-model mean (MMM) and the corresponding differences to the CMIP5 MMM, again for the net (b), shortwave (d), and longwave (f) components. The CRE feedback parameters for each grid cell are calculated with a Gregory regression of the grid cell's CRE versus the change in the global mean near-surface air temperature (see Section 2.2.4). Adapted with permission from Bock et al. (2020).

CMIP6 leads to an increased net cloud feedback parameter in comparison to older CMIP generations. This result is further supported by Zelinka et al. (2020), who find a larger positive zonal mean net cloud feedback parameter in CMIP6 compared to CMIP5 due to a more positive (reduced negative) low-level cloud feedback, primarily in the extratropics. In comparison to earlier CMIP generations, the CMIP6 models exhibit weaker increases in the extratropical low-level cloud cover and the corresponding liquid water content as a result of global warming. Moreover, Zelinka et al. (2020) show that this mainly results from an increase in the liquid condensate fraction in the CMIP6 clouds for the pre-industrial and present-day periods, leading to the aforementioned reduction in the negative cloud phase change feedback on warming and eventually to higher ECS values.



## 5. Evaluation of Emergent Constraints on the Effective Climate Sensitivity (ECS) in CMIP6

The massive spread in ECS within the CMIP6 ensemble discussed in the previous chapter begs the question about the interpretation and credibility of these results. Since the CMIP6 models provide crucial input for the upcoming international climate assessment of the IPCC (AR6), this is not only important for the scientific community but also directly affects policymakers and society. Due to the many new and improved processes included in the models, the CMIP6 ensemble is arguably more realistic than its predecessor generations. However, does this automatically make results derived from CMIP6 models more trustworthy than those derived from earlier CMIP models? In general, the answer to this question is “not necessarily”. Even the CMIP6 models are far from fully simulating all aspects and processes of the vastly complex Earth system. Therefore, the addition of new processes to these models could introduce new biases which have not been present in earlier CMIP models. For example, the apparent increase in climate sensitivity in the CMIP6 models might be wrongly caused by missing processes that cancel out the effects of the newly included processes. Thus, it is vital to take other sources of information into account to assess the credibility of ESM results.

One of the most important sources of information for this purpose is undoubtedly observational data. As presented in Section 2.4.1, appropriate observations of the Earth system can be utilized within the so-called *emergent constraints* framework to assess the accuracy of climate model projections and reduce associated uncertainties. For ECS, a large variety of different emergent constraints have been developed over the years, which we now evaluate on the CMIP6 ensemble to assess whether they still hold for these new climate models, i.e., whether they have skill to reduce uncertainties in ECS in the CMIP6 ensemble. In total, we assess 11 emergent constraints on ECS (see Table 5.1), out of which the most are related to cloud feedbacks since these constitute the most important source of uncertainty for ECS (Boucher et al. 2013; Flato et al. 2013; see also Section 2.2.1). Since all of the evaluated emergent constraints have been derived from the CMIP3 or CMIP5 ensemble, the CMIP6 models offer an exciting possibility to perform out-of-sample testing using the previously unseen data.

This work is already published in Schlund et al. (2020b). For this paper, the author of this thesis led the writing and the analysis and implemented the code to reproduce this study with all figures and tables using the ESMValTool. In this chapter, we first introduce the data (climate model output and observations) and methods used in this study (Section 5.1). After that, we present the 11 emergent constraints on ECS and evaluate them on the CMIP5 and

## 5. Evaluation of Emergent Constraints on the Effective Climate Sensitivity (ECS) in CMIP6

Label	Reference	Description of $x$ -axis	Variables and corresponding observations
BRI	Brient and Schneider (2016)	Sensitivity of shortwave cloud albedo to changes in sea surface temperature (SST) [% K $^{-1}$ ]	<ul style="list-style-type: none"> <li>• <math>hur</math> (ERA-Interim)</li> <li>• <math>rsdt</math> (CERES-EBAF)</li> <li>• <math>rsut</math> (CERES-EBAF)</li> <li>• <math>rsutcs</math> (CERES-EBAF)</li> <li>• <math>ts</math> (HadISST)</li> </ul>
COX	Cox et al. (2018)	$\Psi$ (temperature variability metric) [K]	<ul style="list-style-type: none"> <li>• <math>tas</math> (HadCRUT4)</li> </ul>
LIP	Lipat et al. (2017)	Extent of Southern Hemisphere Hadley cell [ $^{\circ}$ ]	<ul style="list-style-type: none"> <li>• <math>va</math> (ERA-Interim)</li> </ul>
SHD	Sherwood et al. (2014)	$D$ -index (large-scale lower-tropospheric mixing) [1]	<ul style="list-style-type: none"> <li>• <math>wap</math> (ERA-Interim)</li> </ul>
SHL	Sherwood et al. (2014)	LTMI (lower-tropospheric mixing index) [1]	<ul style="list-style-type: none"> <li>• <math>hur</math> (ERA-Interim)</li> <li>• <math>ta</math> (ERA-Interim)</li> <li>• <math>wap</math> (ERA-Interim)</li> </ul>
SHS	Sherwood et al. (2014)	$S$ -index (small-scale lower-tropospheric mixing) [1]	<ul style="list-style-type: none"> <li>• <math>hur</math> (ERA-Interim)</li> <li>• <math>ta</math> (ERA-Interim)</li> <li>• <math>wap</math> (ERA-Interim)</li> </ul>
SU	Su et al. (2014)	Error in vertical profile of relative humidity [1]	<ul style="list-style-type: none"> <li>• <math>hur</math> (AIRS, MLS-Aura)</li> </ul>
TIH	Tian (2015)	Tropical mid-tropospheric humidity asymmetry index from AIRS [%]	<ul style="list-style-type: none"> <li>• <math>hus</math> (AIRS)</li> </ul>
TII	Tian (2015)	Southern ITCZ index from GPCP [mm day $^{-1}$ ]	<ul style="list-style-type: none"> <li>• <math>pr</math> (GPCP)</li> </ul>
VOL*	Volodin (2008)	Difference between tropical and southern midlatitudinal cloud fraction [%]	<ul style="list-style-type: none"> <li>• <math>clt</math> (ISCCP D-2)</li> </ul>
ZHA*	Zhai et al. (2015)	Seasonal response of marine boundary layer cloud fraction to changes in SST [% K $^{-1}$ ]	<ul style="list-style-type: none"> <li>• <math>cl</math> (CloudSat/CALIPSO)</li> <li>• <math>tos</math> (AMSRE SST)</li> <li>• <math>wap</math> (ERA-Interim)</li> </ul>

Table 5.1.: Overview of the 11 emergent constraints on the effective climate sensitivity (ECS) evaluated in this chapter. Detailed descriptions of the variables used to calculate the  $x$ -axis values of the emergent constraints and the references for the corresponding observations (fourth column) are given in Tables A.1 and A.2, respectively. For emergent constraints marked with an asterisk (\*), the observational value of the original publication is used to calculate the observational constraint. For all others, the observational value is computed with the ESMValTool. Adapted with permission from Schlund et al. (2020b).

CMIP6 ensemble (Section 5.2). Finally, we provide a discussion (Section 5.3) and a summary of the results (Section 5.4).

## 5.1. Data and Methods

In this study, we use the output from climate models participating in CMIP5 and CMIP6 as shown in Tables A.3 and A.4, respectively. In addition, for each emergent constraint, at least one observational dataset is used to calculate the observational constraint. A complete list of these datasets is given in Table 5.1 and Table A.2. Following other similar studies, we calculate ECS with the Gregory regression method (Gregory et al. 2004), which is described in detail in Section 2.2.3. An overview of the 11 emergent constraints on ECS analyzed in this study including the variables required for their calculations is given in Table 5.1 and Section 5.2. We chose these particular emergent constraints since these had already been implemented in the ESMValTool at the time of writing this study, which greatly facilitated this analysis. For all emergent constraints, we use the historical simulations of CMIP5 and CMIP6 to ensure maximum agreement with the observational data. If necessary, the historical simulation of CMIP5 is extended after its final year 2005 with data from the RCP8.5 scenario (Riahi et al. 2011). Note that we only use data through 2014, during which time all RCP scenarios behave similarly and the choice of the scenario is not expected to affect the results considerably. Such an extension is not needed for CMIP6 models as their historical simulations cover a longer time period until 2014. We quantify the skill of the emergent constraints with three metrics: the coefficient of determination  $R^2$  of the linear relationship, the statistical significance using a two-sided  $t$ -test based on the Pearson correlation coefficient  $r$  (see Equation (2.21)), and the constrained range of ECS based on the constrained PDF (see Equation (2.20)).

## 5.2. Comparison of Emergent Constraints on ECS for CMIP5 and CMIP6

In this section, we describe and discuss the 11 emergent constraints on ECS summarized in Table 5.1 using CMIP5 and CMIP6 data (Sections 5.2.1–5.2.11) and provide a best estimate for ECS and statistical significance of the 11 emergent constraints in Section 5.2.12. While most of these emergent constraints have been derived using data from the CMIP5 and/or CMIP3 ensembles, to our knowledge none of them has been evaluated on the CMIP6 ensemble so far. The results for the individual emergent constraints described in the following are shown in Figures 5.1–5.4. The left columns in these figures show the emergent relationships, including the uncertainty of the linear regressions (blue and orange shaded areas; see Equation (2.17)) and the uncertainty in the observations (gray shaded area; see Equation (2.19)). The right columns show the PDFs of ECS in the original model ensemble (histogram) and the constrained distribution given by the emergent constraints (blue and orange line; see Equation (2.20)). Table 5.2 shows the corresponding 66 % confidence intervals (i.e., the 17–83 % intervals) of ECS derived from the PDFs given by Equation (2.20) and the  $p$ -values used to assess the significance of the emergent relationships.

### 5.2.1. Sensitivity of Shortwave Cloud Albedo to Changes in Sea Surface Temperature (BRI)

In this emergent constraint proposed by Brient and Schneider (2016), ECS is correlated with the tropical low-level cloud (TLC) albedo, i.e., using the covariance of clouds with changes in SSTs. Differences in the TLC albedo account for more than half of the variance of ECS in the CMIP5 ensemble. Following Brient and Schneider (2016), TLC regions are defined as grid points that are in the driest quartile of 500 hPa relative humidity of all grid cells over the ocean between 30°S and 30°N. The TLC albedo is obtained by calculating the ratio of TOA shortwave cloud radiative forcing and solar insolation averaged over the TLC region. The regression coefficients of deseasonalized variations of TLC shortwave albedo and SST (in % K<sup>-1</sup>) are then used as an emergent constraint for ECS. Here, we use observational data from HadISST for SST (Rayner et al. 2003), ERA-Interim for 500 hPa relative humidity (Dee et al. 2011), and CERES-EBAF (Loeb et al. 2018) for the TOA radiative fluxes over the period 2001–2005. In the original publication, Brient and Schneider (2016) use similar observation-based datasets with the exception of SST, where they take ERSST data (Smith and Reynolds 2003) as reference instead. Our analysis yields a 66 % confidence range for ECS of  $(3.72 \pm 0.59)$  K for CMIP5 ( $R^2 = 0.38$ ) and  $(4.32 \pm 1.07)$  K for CMIP6, with much lower  $R^2 = 0.12$ . The original publication states a best estimate of 4.0 K with a very low likelihood of values below 2.3 K (90 % confidence). The statistical significance of the emergent relationship drops from  $p = 0.0005$  for CMIP5 to  $p = 0.0355$  for CMIP6.

### 5.2.2. Temperature Variability Metric (COX)

The emergent constraint on ECS proposed by Cox et al. (2018) uses a temperature variability metric  $\Psi$  that is based on the interannual variation of GSAT calculated from its variance (in time) and 1-year lag autocorrelation. In contrast to the majority of emergent constraints that focus on cloud-related processes, this constraint is based on the fluctuation-dissipation theorem, which relates the long-term response of the climate system to an external forcing (ECS) and short-term variations of the climate system (climate variability). This arguably places the constraint on a more solid theoretical foundation, although several questions have been raised on the robustness of the results to choices made in the analysis (Brown et al. 2018; Po-Chedley et al. 2018; Rypdal et al. 2018). For example, Annan et al. (2020) has shown that the assumed linear relationship between  $\Psi$  and ECS does not hold when adding a deep ocean to the model. As observational data, here we use the HadCRUT4 dataset (Morice et al. 2012) over the period 1880–2014. Under the COX constraint we assess a 66 % ECS range of  $(3.03 \pm 0.73)$  K for CMIP5 ( $R^2 = 0.31$ ) and  $(3.71 \pm 1.09)$  K for CMIP6 ( $R^2 = 0.01$ ). Cox et al. (2018) derive a 66 % range of  $(2.8 \pm 0.6)$  K from a different subset of CMIP5 models but with the same observations. When moving from CMIP5 to CMIP6, the significance of the emergent relation drops massively from  $p = 0.0032$  to  $p = 0.5415$ , respectively.



Figure 5.1.: Emergent constraints BRI, COX, and LIP applied to the CMIP5 (blue) and CMIP6 ensemble (orange). (a), (c), (e) Emergent relationships (solid lines) for the CMIP models (numbers; see Tables A.5 and A.7 for details) with their standard prediction errors (blue and orange shaded areas; see Equation (2.17)). The vertical dashed line corresponds to the observational reference with its standard error (gray shaded area). The horizontal dashed lines show the best estimates of the constrained effective climate sensitivity (ECS), and the colored points mark the multi-model means for CMIP5 (blue) and CMIP6 (orange). (b), (d), (f) Probability density functions for ECS of the emergent constraints (solid lines; see Equation (2.20)) and the unconstrained model ensembles (histograms). Due to the availability of data, the histograms may differ for the different emergent constraints. Adapted with permission from Schlund et al. (2020b).

### 5.2.3. Extent of Southern Hemisphere Hadley Cell (LIP)

The results of Lipat et al. (2017) show that the multi-year average extent of the Hadley cell correlates with ECS in CMIP5 models. The Hadley cell edge is defined as the latitude of the first two grid cells from the Equator going south where the zonal average 500 hPa mass stream function calculated from December-January-February (DJF) means of the meridional wind field changes sign from negative to positive. Lipat et al. (2017) explain this correlation by tying it to the observed correlation of the interannual variability in midlatitudinal clouds and their radiative effects with the poleward extent of the Hadley cell. For the calculation of the emergent constraint, we use reanalysis data from ERA-Interim (Dee et al. 2011) for the meridional wind speed over the period 1980–2005. Our application of this emergent constraint gives ECS 66 % ranges of  $(2.97 \pm 0.75)$  K for CMIP5 ( $R^2 = 0.18$ ) and  $(3.75 \pm 1.11)$  K for CMIP6 ( $R^2 < 0.01$ ). The original publication does not specify an ECS range. For CMIP6, the emergent constraint shows a much lower statistical significance ( $p = 0.6791$ ) than for CMIP5 ( $p = 0.0228$ ).

### 5.2.4. Large-Scale Lower-Tropospheric Mixing (SHD)

Sherwood et al. (2014) propose that the degree of mixing in the lower troposphere determines the response of boundary layer clouds and humidity to climate warming, as the associated moisture transport would increase rapidly in a warmer atmosphere due to the Clausius-Clapeyron relationship. The large-scale component  $D$  of this mixing is defined as the ratio of shallow to deep overturning.  $D$  is calculated from the vertical velocities averaged over two height regions: 850 and 700 hPa for shallow overturning and 600, 500, and 400 hPa for deep overturning. Both quantities are averaged over parts of the tropical ocean region away from the regions of highest SST and strongest mid-level ascent, specifically the region  $30^\circ\text{S}$ – $30^\circ\text{N}$ ,  $160^\circ\text{W}$ – $30^\circ\text{E}$ , wherever air is ascending at low levels. As observation-based data, we use vertical velocities from ERA-Interim (Dee et al. 2011) over the period 1989–1998 similar to the original publication. We derive ECS 66 % confidence ranges of  $(3.65 \pm 0.64)$  K for CMIP5 ( $R^2 = 0.28$ ) and  $(3.77 \pm 1.06)$  K for CMIP6 ( $R^2 = 0.03$ ). Sherwood et al. (2014) do not give a best estimate for ECS based on the large-scale component of mixing  $D$  or its small-scale counterpart  $S$  (see Section 5.2.5) but instead for the sum of  $D + S$  only (see Section 5.2.6). The regression shows a much lower significance for CMIP6 ( $p = 0.2805$ ) than for CMIP5 ( $p = 0.0037$ ).

### 5.2.5. Small-Scale Lower-Tropospheric Mixing (SHS)

The small-scale mixing  $S$  (Sherwood et al. 2014) is calculated from the differences in relative humidity and temperature between 700 and 850 hPa. The differences are averaged over all grid cells within the upper quartile of the annual mean 500 hPa ascent rate (within ascending regions) in the tropics. The tropics are defined as the region between  $30^\circ\text{S}$  and  $30^\circ\text{N}$ . In the Cloud Feedback Model Intercomparison Project (CFMIP) models (M. J. Webb et al. 2017), for which convective tendencies are available, upward moisture transport by parameterized convection is shown to increase more rapidly with warming for higher values of  $S$ . We use

## 5.2. Comparison of Emergent Constraints on ECS for CMIP5 and CMIP6



Figure 5.2.: As in Figure 5.1 but for the emergent constraints SHD, SHL, and SHS (see Tables A.5 and A.7 for details). Adapted with permission from Schlund et al. (2020b).

reanalysis data from ERA-Interim (Dee et al. 2011) for temperature and relative humidity to calculate the observational constraint (1989–1998). Our analysis shows a 66 % range of ECS of  $(3.07 \pm 0.73)$  K for CMIP5 ( $R^2 = 0.13$ ) and  $(3.48 \pm 1.07)$  K for CMIP6 ( $R^2 = 0.12$ ). The correlation of  $S$  and ECS shows a slightly higher significance in the CMIP6 ensemble ( $p = 0.0396$ ) than in the CMIP5 ensemble ( $p = 0.0647$ ). The SHS constraint is one of the two

analyzed emergent constraints (ZHA being the other exception) that shows a higher statistical significance for the CMIP6 than for the CMIP5 ensemble.

### 5.2.6. Lower-Tropospheric Mixing Index (SHL)

The lower-tropospheric mixing index (LTMI) formulated by Sherwood et al. (2014) is defined as the sum of the small-scale mixing  $S$  (see Section 5.2.5) and the large-scale mixing  $D$  (see Section 5.2.4), which are supposed to capture complementary components of the total mixing phenomenon. Sherwood et al. (2014) argue that the increase in dehydration depends on initial mixing linking it to cloud feedbacks and thus also to ECS. For this constraint, we derive an ECS 66 % confidence range of  $(3.42 \pm 0.65)$  K for CMIP5 ( $R^2 = 0.41$ ) and  $(3.67 \pm 1.06)$  K for CMIP6 ( $R^2 = 0.16$ ). Sherwood et al. (2014) give a best estimate of about 4 K with a lower limit of 3 K. Similar to both other constraints by Sherwood et al. (2014), SHD and SHS, the statistical significance of the SHL emergent relation decreases in CMIP6 ( $p = 0.0138$ ) compared to CMIP5 ( $p = 0.0002$ ).

### 5.2.7. Error in Vertical Profile of Relative Humidity (SU)

Another emergent constraint on ECS that targets uncertainties in cloud feedbacks is proposed by Su et al. (2014). They show that changes in the Hadley circulation are physically connected to changes in tropical clouds and thus ECS. Consequently, the inter-model spread in the change of the Hadley circulation in an ensemble of climate models is well correlated with the corresponding changes in the TOA CRE. Moreover, Su et al. (2014) find a correlation between a model's ECS and its ability to represent the present-day Hadley circulation. The latter is calculated from the tropical ( $45^\circ\text{S}$ – $40^\circ\text{N}$ ) zonal mean vertical profiles of relative humidity from the surface to 100 hPa. These profiles are then used to define the  $x$ -axis of the SU constraint by calculating a performance metric based on the slope of the linear regression between a climate model's relative humidity profile and the corresponding observational reference. Similar to the original publication, we use humidity observations from AIRS (Aumann et al. 2003) for pressure levels greater than 300 hPa and MLS-Aura data (Beer 2006) for pressure levels of less than 300 hPa. Our analysis yields a constrained 66 % range of ECS of  $(3.30 \pm 0.88)$  K for CMIP5 ( $R^2 = 0.08$ ) and  $(3.77 \pm 1.35)$  K for CMIP6 ( $R^2 = 0.05$ ). The original publication gives a best estimate of 4 K with a lower limit of 3 K. Figure 5.3a shows that in addition to the low  $R^2$  values, the emergent relationship shows different slopes for CMIP5 and CMIP6. For CMIP5, the expected positive correlation is found, while for CMIP6, a negative correlation is found. This suggests that the constraint is not working (anymore) when applied to the CMIP6 data. Consequently, the SU constraint shows a weaker statistical significance in CMIP6 ( $p = 0.1935$ ) than in CMIP5 ( $p = 0.1676$ ). The SU constraint is related to an emergent constraint on ECS proposed by Fasullo and Trenberth (2012), who correlate May–August zonal mean relative humidity against ECS. In contrast to Su et al. (2014), they do not use the entire tropics but identify two distinct regions with the largest correlation.



Figure 5.3.: As in Figure 5.1 but for the emergent constraints SU, TIH, and TII (see Tables A.6 and A.8 for details). Adapted with permission from Schlund et al. (2020b).

### 5.2.8. Tropical Mid-Tropospheric Humidity Asymmetry Index (TIH)

Tian (2015) finds a link between mid-tropospheric humidity over the tropical Pacific and simulated moisture, precipitation, clouds, large-scale circulation, and thus ECS in CMIP3 and CMIP5 models. The study explains this link with the similarity of mid-tropospheric

humidity and precipitation patterns as both are related to the Intertropical Convergence Zone (ITCZ). The proposed tropical mid-tropospheric humidity asymmetry index to constrain ECS is defined as relative bias (in %) in simulated annual mean 500 hPa specific humidity averaged over the Southern Hemisphere tropical Pacific ( $30^{\circ}\text{S}$ – $0^{\circ}\text{N}$ ,  $80^{\circ}\text{W}$ – $120^{\circ}\text{E}$ ) minus the bias averaged over the Northern Hemisphere tropical Pacific ( $0$ – $20^{\circ}\text{N}$ ,  $80^{\circ}\text{W}$ – $120^{\circ}\text{E}$ ) when compared with observations. Similar to the SU constraint, the index proposed by Tian (2015) seems to be related to the emergent constraint by Fasullo and Trenberth (2012), who find correlations between relative humidity of the middle and upper troposphere and ECS. Here, we use humidity observations from AIRS (Aumann et al. 2003) over the period 2003–2005 as the reference dataset. We assess a 66 % ECS range of  $(3.88 \pm 0.75)$  K for CMIP5 ( $R^2 = 0.24$ ) and  $(4.15 \pm 1.10)$  K for CMIP6 ( $R^2 = 0.06$ ). Tian (2015) specifies a best estimate of 4.0 K. The significance of the emergent relationship drops massively from  $p = 0.0089$  in CMIP5 to  $p = 0.1348$  in CMIP6.

### 5.2.9. Southern ITCZ Index (TII)

In addition to the humidity index, Tian (2015) proposes an emergent constraint on ECS based on the southern ITCZ index (Bellucci et al. 2010; Hirota et al. 2011). This index is defined as the climatological annual mean precipitation bias averaged over the southeastern Pacific ( $30^{\circ}\text{S}$ – $0^{\circ}\text{N}$ ,  $150$ – $100^{\circ}\text{W}$ ). The southern ITCZ index is calculated in  $\text{mm day}^{-1}$  and dominated by the so-called double-ITCZ, a common problem in many CMIP5 climate models. Tian (2015) finds a link between double-ITCZ bias and simulated moisture, precipitation, clouds, and large-scale circulation in CMIP3 and CMIP5 models. He argues that this could explain the link between the double-ITCZ bias and ECS. As reference data, we use observed precipitation data for the years 1986–2005 from GPCP (Adler et al. 2003). We calculate an ECS 66 % confidence range of  $(3.87 \pm 0.67)$  K for CMIP5 ( $R^2 = 0.33$ ) and  $(3.84 \pm 1.09)$  K for CMIP6 ( $R^2 < 0.01$ ). Tian (2015) specifies a best estimate of 4.0 K. The emergent relationship shows a much lower statistical significance in CMIP6 ( $p = 0.8236$ ) than in CMIP5 ( $p = 0.0013$ ).

### 5.2.10. Difference Between Tropical and Southern Midlatitudinal Cloud Fraction (VOL)

The study by Volodin (2008) aims at constraining ECS based on the geographical distribution of clouds in climate models. Since this early emergent constraint has originally been trained on CMIP3 models, both CMIP5 and CMIP6 are out-of-sample tests for it. Volodin (2008) shows that high ECS models tend to simulate a higher total cloud cover over the southern midlatitudes and a lower total cloud cover over the tropics (relative to the MMM). This can be used to establish an emergent relationship between the ECS and the difference in tropical total cloud cover ( $28^{\circ}\text{S}$ – $28^{\circ}\text{N}$ ) and the southern midlatitudinal total cloud cover ( $56$ – $36^{\circ}\text{S}$ ). Analogous to the original study, we use the ISCCP-D2 data (Rossow and Schiffer 1991) as observational reference. For the VOL constraint, we calculate a constrained 66 % range of ECS of  $(3.74 \pm 0.64)$  K for CMIP5 ( $R^2 = 0.38$ ) and  $(4.21 \pm 1.04)$  K for CMIP6 ( $R^2 = 0.18$ ),



Figure 5.4.: As in Figure 5.1 but for the emergent constraints VOL and ZHA (see Tables A.6 and A.8 for details). Adapted with permission from Schlund et al. (2020b).

whereas the original publication gives a range of  $(3.6 \pm 0.4)$  K (standard deviation) for a climate model ensemble of CMIP3 models. The emergent constraint by Volodin (2008) shows a lower significance in the CMIP6 ensemble ( $p = 0.0056$ ) than in the CMIP5 ensemble ( $p = 0.0004$ ).

### 5.2.11. Response of Marine Boundary Layer Cloud Fraction to Changes in Sea Surface Temperature (ZHA)

Zhai et al. (2015) focus on the variations of marine boundary layer clouds (MBLCs), which largely contribute to the shortwave cloud feedback and thus to the uncertainty in modeled ECS. Their central quantity is the response of the MBLC fraction to changes in the SST in subtropical oceanic subsidence regions for both hemispheres ( $20\text{--}40^\circ$ ). On short (seasonal) and long (centennial under a forcing) time scales, this quantity is well correlated with ECS among an ensemble of CMIP3 and CMIP5 models. Together with observations of cloud fraction from CloudSat/CALIPSO (Mace et al. 2009), SST from AMSRE SST (AMSR-E 2011), and vertical velocity from ERA-Interim (Dee et al. 2011), the seasonal response of MBLC fraction to changes in SST forms an emergent constraint on ECS. We assess a 66 % ECS range



Figure 5.5.: Emergent relationship ZHA (Zhai et al. 2015) for different subsets of CMIP5 models. Blue points show the 15 CMIP5 models used in the original publication (except for CESM1-CAM5). The solid blue line and blue shaded area show the emergent relationships evaluated on these models including the uncertainty range. In our study, we added 11 more CMIP5 models (red points). The corresponding emergent relationship that considers all available CMIP5 models is shown in red colors. This relationship shows a considerably lower coefficient of determination ( $R^2$ ) and higher  $p$ -value than the relationship using the original subset of CMIP5 models. The vertical dashed line and shaded area correspond to the observational reference, and the horizontal dashed lines show the corresponding constraints on the effective climate sensitivity (ECS) using this observation. Adapted with permission from Schlund et al. (2020b).

of  $(3.35 \pm 0.74)$  K for CMIP5 ( $R^2 = 0.05$ ) and  $(3.79 \pm 0.67)$  K for CMIP6 ( $R^2 = 0.62$ ). In their original publication, Zhai et al. (2015) find an ECS range of  $(3.90 \pm 0.45)$  K (standard deviation) for a combination of CMIP3 and CMIP5 models. In terms of statistical significance, the results of the ZHA constraints are somewhat surprising: although CMIP5 data (in combination with CMIP3 data) has successfully been used in the original publication, our approach finds that the statistical significance of the emergent relationship is much higher in the unseen CMIP6 ensemble ( $p < 0.0001$ ) than in the previously available CMIP5 ensemble ( $p = 0.2567$ ). The ZHA constraint is the only emergent constraint analyzed here that shows this extreme behavior (only one other constraint, SHS, shows a slightly higher significance in CMIP6; all other constraints show lower significances in CMIP6). The reason for the erratic skill in CMIP5 is the set of climate models used. For our analysis, we use 11 additional CMIP5 models that have not been used in the original publication (ACCESS1-0, ACCESS1-3, bcc-csm1-1, bcc-csm1-1-m, CCSM4, GFDL-ESM2G, GFDL-ESM2M, IPSL-CM5A-MR, IPSL-CM5B-LR, MPI-ESM-MR, and MPI-ESM-P). Due to a lack of publicly available data, the model CESM1-CAM5 that is used in the original publication is not included in our analysis. The effect of choosing different subsets of CMIP5 models on the emergent relationship is illustrated in Figure 5.5. Using the original CMIP5 models from the original publication gives a considerably higher correlation

( $R^2 = 0.38$ ) than using all available CMIP5 models ( $R^2 = 0.05$ ). This result shows a strong dependency of this emergent constraint on the subset of climate models used. Nonetheless, the performance on CMIP6 models is, surprisingly, the best of all the constraints and much better than on either subset of CMIP5 models.

### 5.2.12. Constrained ECS Ranges and Statistical Significance of the 11 Emergent Constraints

In most cases, the emergent relationships (left columns of Figures 5.1–5.4) show the same sign of the slope (as expected from the theory) for CMIP5 and CMIP6, with the SU constraint being the only exception. However, the coefficient of determination ( $R^2$ ) is lower for CMIP6 compared to CMIP5 for all one constraint: ZHA. The probability distributions of the constrained ECS that we obtain (right columns of Figures 5.1–5.4) give similar results: except for the ZHA constraint, the constraint on the CMIP6 ensemble is weaker, i.e., the constrained PDFs derived from the CMIP6 ensemble are broader than their respective CMIP5 counterparts. As shown in Table 5.2 and Figure 5.6, for CMIP5, the range of the best (maximum likelihood) estimates for ECS is 2.97 K to 3.88 K, while the corresponding CMIP6 best estimates are higher for almost

Label	ECS (original publication)	ECS [K] (CMIP5)	ECS [K] (CMIP6)	$p$ (CMIP5)	$p$ (CMIP6)
BRI	most likely 4.0 K, < 2.30 K very unlikely (90 % confidence)	$3.72 \pm 0.59$	$4.32 \pm 1.07$	0.0005	0.0355
COX	$(2.8 \pm 0.6)$ K	$3.03 \pm 0.73$	$3.71 \pm 1.09$	0.0032	0.5415
LIP	no best estimate given	$2.97 \pm 0.75$	$3.75 \pm 1.11$	0.0228	0.6791
SHD	none — see SHL	$3.65 \pm 0.64$	$3.77 \pm 1.06$	0.0037	0.2805
SHL	most likely 4 K, lower limit 3 K	$3.42 \pm 0.65$	$3.67 \pm 1.06$	0.0002	0.0138
SHS	none — see SHL	$3.07 \pm 0.73$	$3.48 \pm 1.07$	0.0647	0.0396
SU	most likely 4 K, lower limit 3 K	$3.30 \pm 0.88$	$3.77 \pm 1.35$	0.1676	0.1935
TIH	most likely 4.0 K	$3.88 \pm 0.75$	$4.15 \pm 1.10$	0.0089	0.1348
TII	most likely 4.0 K	$3.87 \pm 0.67$	$3.84 \pm 1.09$	0.0013	0.8236
VOL	$(3.6 \pm 0.4)$ K (std. dev.)	$3.74 \pm 0.64$	$4.21 \pm 1.04$	0.0004	0.0056
ZHA	$(3.90 \pm 0.45)$ K (std. dev.)	$3.35 \pm 0.74$	$3.79 \pm 0.67$	0.2567	< 0.0001

Table 5.2.: Overview of the constrained effective climate sensitivity (ECS) ranges and  $p$ -values for all 11 analyzed emergent constraints. If not further specified, the uncertainty ranges correspond to the 66 % confidence intervals (17–83 %). For CMIP5 and CMIP6, these are evaluated from the probability density function given by Equation (2.20) (see also the right columns of Figures 5.1–5.4). Note that even though CMIP5 models have been used for some constraints in the original publications, the constrained ranges in the second and third column might differ due to the use of a different subset of climate models (in this study, we use output from all CMIP models that are publicly available; see Tables A.5–A.8 for details). The  $p$ -values describing the significances of the emergent relationships are defined as the probability to obtain an absolute correlation coefficient  $|r|$  or higher under the null hypothesis that the true underlying correlation coefficient between the predictor and ECS is zero. Smaller  $p$ -values point to a higher significance and vice versa (see Section 2.4.1 for details). Adapted with permission from Schlund et al. (2020b).



Figure 5.6.: Overview of the constrained effective climate sensitivity (ECS) ranges given by the 11 analyzed emergent constraints using the CMIP5 (blue) and CMIP6 (orange) ensemble. The bars correspond to the 66 % confidence intervals (17–83 %) calculated from the constrained probability density function of ECS following Equation (2.20). The vertical black lines illustrate the corresponding best estimates of the distributions. For almost all emergent constraints, the constrained best estimate and range of ECS are higher in the CMIP6 ensemble.

every tested emergent constraint (TII being the only exception), resulting in a range of best estimates of 3.48 K to 4.32 K. Using the arithmetic mean of all analyzed emergent constraints, this results in a mean increase of the ECS best estimate of 12 % in CMIP6 compared to CMIP5. Similarly, the size of the 66 % ECS ranges (17–83 % confidence) shows values of 1.16 K to 1.75 K in CMIP5 and 1.32 K to 2.70 K in CMIP6, resulting in an increase of 51 % averaged over all emergent constraints.

In summary, the  $R^2$  of the emergent relationships and the constrained range of ECS each depend strongly on the climate model ensemble used, even though a physical explanation is given for each emergent constraint that is thought to be valid for every climate model ensemble. The same behavior is found for the statistical significance of the emergent relationships using the null hypothesis that there is no correlation between the predictor and ECS. Except for the ZHA and SHS constraints, every emergent relationship investigated shows a lower statistical significance (i.e., a higher  $p$ -value) in the CMIP6 ensemble than in the CMIP5 ensemble. If a conventional significance test of  $p < 0.05$  was applied, eight of the 11 constraints would pass this test on CMIP5 model data but only five (BRI, SHL, SHS, VOL, and ZHA) would pass this test on CMIP6. This is still much better than would be expected purely by chance. Hence, there is still skill in at least a few of the constraints, but it is much lower than suggested in nearly all of the initial studies.

### 5.3. Discussion

As shown in the previous sections, most emergent relationships show smaller coefficients of determination when evaluated on the new CMIP6 ensemble compared to the CMIP5 ensemble. In this section, we discuss possible reasons for, and implications of, these differences. As reported by Caldwell et al. (2014), the large amount of data provided by modern ESMs can generate spurious correlations of variables between past climate and ECS just by chance, especially when only a few climate models are considered. This would cause the performance of the emergent constraint to be reduced on out-of-sample data (like the new CMIP6 ensemble) since the emergent relationship appeared just by chance and not because of a physically based mechanism.

A further reason for the weaker emergent relationships in CMIP6 may be the increased complexity of the participating ESMs. Each emergent constraint approach is based on the assumption that a single observable process or physical aspect in the current climate dominates the uncertainty in ECS. Some emergent constraints such as ZHA and BRI relate changes in cloud properties (ZHA: low-level cloud fraction; BRI: cloud reflectivity) on seasonal or interannual time scales driven by changes in SST to ECS. This means that it has to be implicitly assumed that the observable changes in these properties on seasonal or interannual time scales are related to those occurring as a result of climate forcing in a way successfully captured by the ESMs. While this assumption seems to make sense, we do not know whether the ESMs cover all relevant processes of the real Earth system. For example, it may be possible that there exist processes that are unimportant in the ESMs (and hence are not captured by the emergent constraints) but are actually important in reality. This lack of relevant processes may lead to an overconfident constraint. Thus, the more complex ESMs of the CMIP6 ensemble are more likely to capture relevant processes of the true climate, which leads to weaker emergent relationships. On the other hand, emergent constraints on the less complex CMIP3 and CMIP5 ensembles may be overconfident.

For CMIP6 models, Zelinka et al. (2020) show that cloud feedbacks and thus ECS in high-sensitivity models are to some extent associated with changes in clouds over the Southern Ocean, while in CMIP3 and CMIP5 the uncertainty in cloud feedbacks is dominated by clouds in the subtropical subsidence regions. One might speculate that a possible reason for this might be an improved simulation of clouds over the Southern Ocean in some models (Bodas-Salcedo et al. 2019; Gettelman et al. 2019a), as shown for some pre-CMIP6 model versions evaluated by Lauer et al. (2018). The findings of Zelinka et al. (2020) can also at least partly explain the larger inter-model spread in climate sensitivity due to a greater diversity of cloud feedbacks, which also results in a weaker emergent constraint compared with CMIP5 models, as most of them constrain low-cloud feedbacks. They find that on average, the shortwave low-cloud feedback is larger in CMIP6 than in CMIP5, which they primarily relate to changes in the representation of clouds. As a possible explanation, Zelinka et al. (2020) give an increase in mean-state supercooled liquid water (i.e., increase in the cloud water liquid fraction) in

mixed-phase clouds resulting in less pronounced increases in low-level cloud cover and water content with warmer SSTs, particularly in midlatitudes.

Our findings suggest that the process-oriented emergent constraints (i.e., all of the emergent constraints investigated here except COX) are only successful in constraining ECS as long as the uncertainty in ECS is dominated by the same process or feedback. In the CMIP5 ensemble, cloud feedback is the main contributor to the spread in ECS with low-level clouds in tropical subsidence regions dominating the spread in cloud feedback (Ceppi et al. 2017). If any other process or feedback is biased (or missing) in the ensemble as a whole, then these process-oriented emergent constraints will be biased in their estimates of ECS. The appearance of diverse new feedback processes in CMIP6 could explain the reduced skill when applied to CMIP6 data, and a tendency for these to be positive would explain the upward shift in the model ECS distribution that is not captured by the CMIP5-trained constraints. Process-oriented emergent constraints are therefore perhaps best thought of as constraints on the processes that they target, rather than constraints on ECS.

Emergent constraints that use global temperature change as a way of constraining ECS could in principle overcome this problem. If one feedback is biased in an ensemble the constraint might still work as both global temperature change and ECS might similarly reflect the sum of all feedbacks. Emergent constraints of this kind include the tropical temperature during the Last Glacial Maximum (Hargreaves et al. 2012), tropical temperature anomalies during the mid-Pliocene Warm Period (Hargreaves and Annan 2016), and post-1970s warming (Jimenez-de-la-Cuesta and Mauritsen 2019). This seems to be supported by the findings of Tokarska et al. (2020), who test an emergent constraint for the TCR based on recent global warming trends on the CMIP5 and CMIP6 ensembles with similar results for both model ensembles. Nevertheless, these temperature-based estimates are sensitive to assumptions about forcings and unforced decadal temperature variations, which could also be incorrect, as could model-predicted relationships between feedbacks on short and long time scales that are implicit in most such measures.

## 5.4. Summary

In this chapter, 11 different emergent constraints on ECS, which are mostly directly or indirectly related to cloud feedbacks, are assessed by applying them to results from ESMs contributing to CMIP5 and CMIP6. Of particular interest are the results from CMIP6, since all analyzed emergent constraints have been published prior to the availability of CMIP6 data. In summary, the best estimate of ECS averaged over all emergent constraints increases by 12 % when moving from relationships trained on CMIP5 to those trained on CMIP6. Some increase is predicted by almost every constraint we analyze (with TII as the only exception) and can at least be partly explained by the increased multi-model mean ECS of CMIP6, which has not been accompanied by systematic changes in the constraint variables that could explain this increase, leading to regression fits with higher intercept values at observed constraint values. This is also illustrated by the CMIP5 and CMIP6 MMMs in the left columns of Figures 5.1–5.4 (colored

points), in which the connecting line between the CMIP5 and CMIP6 MMM is not parallel to the CMIP5 emergent relationships for all emergent constraints.

Our results also show that, except for ZHA, all emergent relationships are weaker (in terms of the coefficient of determination  $R^2$ ) in CMIP6 compared to CMIP5, which means that the corresponding emergent relationships are able to explain less of the ECS variation simulated by the newer CMIP6 models than by those of CMIP5. This is also demonstrated by the statistical significance, which is lower in CMIP6 than in CMIP5 for all emergent constraints except for ZHA and SHS. As described in Section 2.4.1, our test for statistical significance uses the null hypothesis that there is no correlation between the predictor variable of the emergent constraint and ECS. Further evidence of the decreased performance of the emergent constraints in CMIP6 is given by the size of the constrained ECS ranges, which widens by 51 % in CMIP6 compared to CMIP5 on average. Moreover, our more detailed analysis of the ZHA constraint (see Figure 5.5) shows that this emergent constraint is very sensitive to outliers and the subset of the climate model ensemble used to fit the emergent relationship. Such behavior might not be unique to the ZHA constraint but could apply to other emergent constraints as well. This in turn suggests that the number of climate models commonly used for emergent constraints might be too low, leading to non-robust relationships.

Our analysis makes several simplifying assumptions common to other studies, such as model independence, discussed in detail in Sections 2.2.3 and 2.4.1. These assumptions affect the significance of emergent relationships and the PDFs of ECS based on a constraint. However, they do not affect our main conclusions here, which concern the change in performance on CMIP6 relative to CMIP5 and the implications for robustness and future use of emergent constraints.

ECS is the product of the complex interactions of the many components and feedbacks. Thus, constraining ECS with a single physical process might overly simplify this problem. Such single-process-oriented emergent constraints therefore do not seem to be helpful in constraining ECS but should probably rather be thought of as constraints for the process or feedback they are actually targeting (if that can be clearly identified). With increasing computational resources available to climate science, more and more detailed process interactions can be taken into account in a modern ESM. In contrast, the predecessor versions CMIP3 and CMIP5 are less complex with simpler atmospheric process representations, so constraining uncertainties of a single dominant process may allow for an apparently more successful constraint of ECS than would be achieved in more complex models. As a conclusion, we argue that to constrain ECS more robustly, it might be beneficial to apply multivariate approaches that are able to consider multiple (different) relevant physical processes and feedbacks at once and thus are able to get a broader picture of the complex reality. A possible approach for this is given by C. S. Bretherton and Caldwell (2020), who combine the information from multiple emergent constraints on ECS using a multivariate Gaussian and multivariate linear regression. For the CMIP3 and CMIP5 ensembles, they find an increased best estimate relative to the unweighted ensemble mean similar to the participating individual emergent constraints, but with a lower uncertainty range. Moreover, new ML techniques are a promising avenue

forward for such multivariate approaches and for constraining uncertainties in multi-model projections (Schlund et al. 2020a; see Chapter 6) with the aim of further improving climate modeling and analysis (Reichstein et al. 2019).

## 6. Constraining Future Gross Primary Production (GPP) with Machine Learning

As argued in detail in the previous chapter, single-process-oriented emergent constraints do not seem to be beneficial in constraining climate projections from the vastly complex modern-day ESMs since they might overly simplify the problem. To overcome this issue, we explore an alternative approach that expands the multiple diagnostic ensemble regression (MDER) method (see Section 2.4.3) with a supervised machine learning (ML) algorithm. In contrast to emergent constraints, this approach considers (1) multiple process-based predictors, (2) multi-dimensional (gridded) target variables, and (3) non-linear relationships between the predictors and the target variable. Since the new technique relies on a large number of data points to train the ML algorithm, ECS as a scalar climate sensitivity metric is not an appropriate target variable. Thus, this analysis does not focus on reducing uncertainties in climate sensitivity itself, but rather on a selected process that contributes to it: GPP. As described in detail in Sections 2.3.1 and 2.3.2, GPP constitutes the largest flux of the terrestrial carbon uptake and actively slows down global warming by removing CO<sub>2</sub> from the atmosphere. In particular, the terrestrial biosphere absorbs about 30 % of the total anthropogenic CO<sub>2</sub> emission (Friedlingstein et al. 2020). Thus, accurately quantifying the future evolution of the terrestrial carbon sink is crucial for reliable climate change projections.

Due to complex feedback processes connected to the global carbon cycle (concentration–carbon feedback and climate–carbon feedback; see Section 2.3.2), the response of the terrestrial carbon cycle to changes in the atmospheric CO<sub>2</sub> concentration and climate is a major source of uncertainty in climate projections (Bodman et al. 2013; Booth et al. 2012; M. Collins et al. 2013). For example, this applies to multi-model projections of future GPP of the CMIP5 ensemble. Thus, this study aims to reduce uncertainties in multi-model projections of future GPP in the CMIP5 RCP8.5 scenario at the end of the 21<sup>st</sup> century with a new ML-based weighting technique and appropriate observations. The new two-step approach utilizes aspects of emergent constraints and the MDER technique in combination with an ML algorithm. In the first step, we apply an existing emergent constraint on CO<sub>2</sub> fertilization (Wenzel et al. 2016a) to constrain the ESMs' responses to rising atmospheric CO<sub>2</sub> concentration using observations of the increase of the CO<sub>2</sub> seasonal cycle amplitude at Cape Kumukahi, Hawaii (KUM) (Keeling et al. 2005). In a second step, we introduce a supervised ML algorithm based on boosting trees (Friedman 2001) to learn an empirical spatial relationship that links grid-wise future GPP to historical processes relevant to its simulation under present-day conditions. In combination with observational products of the predictors, that relationship can be used to further constrain

uncertainties in the projected spatial maps of GPP at the end of the 21<sup>st</sup> century in the RCP8.5 scenario. We examine both constraining the absolute GPP and the fractional change in GPP as two independent approaches and target variables. Unlike univariate linear regression used in the MDER algorithm, the proposed *gradient boosted regression tree (GBRT)* algorithm is able to handle multiple predictors and copes with non-linearities in the data. GBRT is a well-known and successful tool used for interpolation, classification, and prediction in other fields of data science and engineering (De'ath 2007; Elith et al. 2008). In the context of climate science, GBRT was recently applied to identify the key drivers of spatial variations of the ratio of plant transpiration to total terrestrial evapotranspiration in ESMs (Lian et al. 2018).

This work is already published in Schlund et al. (2020a). For this paper, the author of this thesis led the writing and the analysis and implemented the code to reproduce this study with all figures and tables using the ESMValTool. Section 6.1 of this chapter provides an overview of the data and methods used in this paper. The results are presented in Section 6.2 and Section 6.3 closes with a summary and discussion.

## 6.1. Data and Methods

Here we briefly describe the data and methods used in this study. Section 6.1.1 reviews the method of a previously published emergent constraint on CO<sub>2</sub> fertilization from Wenzel et al. (2016a) that we apply here to our study to rescale the climate model output (Step 1). This includes a discussion of the caveats that are associated with this emergent constraint that attributes changes in the observed increase in the CO<sub>2</sub> seasonal cycle amplitude at KUM entirely to the increase in atmospheric CO<sub>2</sub> concentration, yet other studies provide indications that part of this increase is due to climate and land-use changes (Bastos et al. 2019; Forkel et al. 2016; Piao et al. 2018; Zhao et al. 2016). In Section 6.1.2 the ML technique used in Step 2 to constrain gridded GPP projections is presented. Figure 6.1 shows an overview of our two-step approach. Section 6.1.3 describes how uncertainty and predictive skill are quantified. A complete list of all CMIP5 models used in this study is given in Table B.1. To apply the emergent constraint in Step 1, we need emission-driven CMIP5 simulations. The number of ESMs included is therefore seven, as in Wenzel et al. (2016a). More details on the methods and the experimental setup are given in Appendix B.

### 6.1.1. Constraining the CO<sub>2</sub> Fertilization Effect (Step 1)

In this section, we present the observational emergent constraint on the CO<sub>2</sub> fertilization effect by Wenzel et al. (2016a) which we use to globally rescale the models for their individual bias in the sensitivity of GPP to an increase in atmospheric CO<sub>2</sub> concentration. Our objective here is to present a general methodology and suggest further improvements upon this constraint that could be used in the future. Wenzel et al. (2016a) show a correlation between the increase in the CO<sub>2</sub> seasonal cycle amplitude simulated by models and the projected CO<sub>2</sub> fertilization. When combined with the observed trends in the CO<sub>2</sub> seasonal cycle amplitude,



Figure 6.1.: Schematic illustration of our two-step approach. In Step 1 an emergent constraint by Wenzel et al. (2016a) is used to constrain the global mean fractional change in gross primary production (GPP) over the 21<sup>st</sup> century. Moreover, this constraint is used to rescale two different gridded target variables: absolute GPP at the end of the 21<sup>st</sup> century and fractional GPP change over the 21<sup>st</sup> century. In Step 2, a machine learning model is used to constrain these two target variables (Step 2a: absolute GPP; Step 2b: fractional GPP change) in two independent approaches. Adapted with permission from Schlund et al. (2020a).

the fractional GPP change at the time of CO<sub>2</sub> doubling relative to pre-industrial conditions in a biogeochemically coupled simulation where the atmospheric CO<sub>2</sub> concentration increases by 1 % per year (1%BC) is constrained. The predictor in this study is the sensitivity of the CO<sub>2</sub> seasonal cycle amplitude to rising atmospheric CO<sub>2</sub> concentrations, defined as the slope of the linear regression between the CO<sub>2</sub> seasonal cycle amplitude and the annual mean atmospheric CO<sub>2</sub> concentrations (see Figure B.1 for details). The emergent constraint is physically motivated by the hypothesis that increasing terrestrial GPP is the main driver for the observed changes in the CO<sub>2</sub> seasonal cycle amplitude (Gray et al. 2014; Keeling et al. 1996; Zhao and Zeng 2014). This hypothesis is based on the fact that the seasonal cycle in atmospheric CO<sub>2</sub> concentration originates from photosynthesis and decomposition processes: in summer, photosynthesis removes CO<sub>2</sub> from the atmosphere; in winter, additional CO<sub>2</sub> is added to the atmosphere due to the decomposition of organic matter (Keeling et al. 1995). Thus, CO<sub>2</sub> fertilization leads to an increase of the CO<sub>2</sub> seasonal cycle amplitude due to increased CO<sub>2</sub> removal from the atmosphere. We note that this emergent constraint implicitly assumes that there is no temperature-driven increase in respiration over time. Therefore, this approach most likely overestimates the true sensitivity of GPP to an increase in the atmospheric CO<sub>2</sub> concentration.

However, we note that this emergent constraint is not undisputed. In other studies, the observed change in the CO<sub>2</sub> seasonal cycle amplitude is attributed to other factors. Using a dynamic global vegetation model, Forkel et al. (2016) find that the increase in the CO<sub>2</sub> seasonal cycle amplitude above 40 °N is mainly driven by the response of plants to global warming instead of CO<sub>2</sub>. Similarly, Piao et al. (2018) show that elevated atmospheric CO<sub>2</sub> concentration and climate change equally contribute to the increase in the CO<sub>2</sub> seasonal cycle in northern high latitudes with additional smaller contributions from air-sea carbon fluxes and land-use changes. For their analysis, Piao et al. (2018) use CO<sub>2</sub> records from 26 stations and nine terrestrial ecosystem models. An evaluation of two atmospheric inversions and 11 land surface models shows contrasting effects of CO<sub>2</sub> fertilization (positive) and global warming (negative) on the CO<sub>2</sub> seasonal cycle amplitude increase in northern high latitudes with negligible contributions from land-use changes (Bastos et al. 2019). A further study by Zhao et al. (2016) finds that for seven out of nine dynamic vegetation models, the CO<sub>2</sub> fertilization effect is the strongest driver for the observed changes in the CO<sub>2</sub> seasonal cycle amplitude, while the two remaining models equally attribute this to CO<sub>2</sub> fertilization, climate change, and land-use changes. Moreover, Winkler et al. (2019b) discuss limitations of carbon cycle emergent constraints related to the calculation of the predictor from observation-based data and the inter-model relationship. As a case study, they adopt a further emergent constraint by Winkler et al. (2019a) that uses the response of the leaf area index (LAI) to ambient CO<sub>2</sub> instead of the CO<sub>2</sub> seasonal cycle amplitude sensitivity to constrain the GPP increase in the northern high latitudes. Furthermore, it has been shown that the CMIP5 models (Graven et al. 2013) and the TRENDY dynamic vegetation models (Thomas et al. 2016) greatly underestimate long-term changes in the CO<sub>2</sub> seasonal cycle which questions their ability to attribute changes in the CO<sub>2</sub> seasonal cycle to different contributors. Nevertheless, we emphasize that we here present a generic methodology and use the results of an existing emergent constraint. Our proposed framework is universal and could be based on other emergent constraints in the future.

In contrast to Wenzel et al. (2016a), instead of idealized simulations we apply the presented emergent constraint on the emission-driven RCP8.5 simulation (Riahi et al. 2011) and examine whether it still holds for this simulation where forcings other than CO<sub>2</sub> also change. Moreover, we consider the global mean GPP change instead of the GPP change in the northern extratropics as in Wenzel et al. (2016a). Due to the strong annual mixing of CO<sub>2</sub> in the atmosphere, the change in the CO<sub>2</sub> seasonal cycle amplitude at KUM is representative of the whole globe. Thus, we apply this emergent constraint to the global mean GPP change. However, we note that this approach might introduce errors since the CO<sub>2</sub> fertilization effect could be different for tropical and extratropical ecosystems due to nutrient limitations and the temperature dependence of RuBisCO kinetics (Crafts-Brandner and Salvucci 2000) in the tropics. Nevertheless, we argue that we can use the constraint if the emergent relationship still holds within our considered framework (the RCP8.5 scenario simulated by the CMIP5 models).

Therefore, our target variable in Step 1 is the global mean fractional GPP change over the 21<sup>st</sup> century  $f$  in the RCP8.5 scenario, which is calculated from the 10-year mean GPP at the

end of the 21<sup>st</sup> century in the emission-driven fully coupled RCP8.5 simulations (2091–2100) and the 10-year mean GPP at the end of the 20<sup>th</sup> century (1991–2000) in the emission-driven historical simulation:

$$f = \frac{\text{GPP (2091–2100)}}{\text{GPP (1991–2000)}} - 1. \quad (6.1)$$

This variable is given as a percentage, where negative values correspond to a decrease in GPP, positive values to an increase in GPP, and a value of 0 % to no GPP change. The CO<sub>2</sub> amplitude sensitivity is calculated for the years 1979–2019, which is the full available time range for the observed CO<sub>2</sub> record at KUM (Keeling et al. 2005). Similar to Wenzel et al. (2016a), we extract the observed CO<sub>2</sub> amplitude from smoothed atmospheric CO<sub>2</sub> concentrations (using a stiff cubic spline function plus a four-harmonics function with linear gain) as provided by Keeling et al. (2005). For the CMIP5 models, the respective grid cell closest to KUM (19.5 °N, 154.8 °W) is extracted. Additionally, the emission-driven historical simulations are extended with the emission-driven RCP8.5 simulations for the years 2006 to 2019. More details including an illustration of this calculation are shown in Figure B.1.

Since the global constraint above cannot be used directly as a predictor in the ML approach which operates on gridded values (see Section 6.1.2), we use it as a first step to globally rescale the ESM output. This important first step allows us to constrain the large uncertainty in the global CO<sub>2</sub> fertilization effect of the CMIP5 models. In this study, two different gridded target variables are examined (see Figure 6.1): the monthly climatology of future absolute GPP at the end of the 21<sup>st</sup> century (2091–2100) in the RCP8.5 scenario (see Step 2a) and the fractional GPP change over the 21<sup>st</sup> century as defined in Equation (6.1) (see Step 2b). The global rescaling of the gridded target variables works as follows: Let  $f_m$  be the global mean fractional GPP change over the 21<sup>st</sup> century of ESM  $m$ ,  $f'$  the constrained value of the global mean fractional GPP change, and  $y_{mi}$  the target variable of climate model  $m$ . Then, we define the rescaled target variable  $y'_{mi}$  by

$$y'_{mi} = y_{mi} \cdot \frac{f'}{f_m}, \quad (6.2)$$

where the index  $i$  counts over all grid cells (and months when absolute GPP is used as target variable in Step 2a) and  $m$  refers to the ESM ( $m \in \{1, 2, \dots, 7\}$ ). Thus, values for the individual grid cells/months are rescaled by a factor that is determined by the ratio between the models' global mean GPP change and the constrained global mean GPP change derived from the Wenzel et al. (2016a) emergent constraint using the observed change in the CO<sub>2</sub> cycle amplitude. Models that simulate a too high global mean GPP change are nudged toward smaller values and vice versa. The rescaled target variables are then fed into the ML model (Step 2) presented in the following section. We note that this global rescaling assumes that the relative biases in the fractional GPP change of the individual ESMs are constant over the whole globe.

### 6.1.2. Gradient Boosted Regression Tree (GBRT) Algorithm (Step 2)

For Step 2 we use the GBRT algorithm to find an optimal regression function  $F$  relating the target variable  $y$  to a set of  $K$  predictors  $\mathbf{x} = (x^{(1)}, x^{(2)}, \dots, x^{(K)})$  (also called *features* or *covariates*) such that the predicted target variable is given by  $\hat{y} = F(\mathbf{x})$  (De'ath 2007; Elith et al. 2008; Friedman 2001, 2002). Its basic elements are decision trees, which use binary splits of the input data to create decision rules. Due to their simple nature, decision trees are easy to interpret but cannot be used to create satisfying predictions for very complex datasets. This issue can be overcome by a technique called *boosting* (Freund and Schapire 1996). Boosting iteratively improves the performance of *weak learners* (in our case decision trees) in the following way: At first, a single decision tree is fitted to a randomly selected subsample of the training data  $\{(x_i, y_i) \mid i \in I\}$  with an arbitrary index set  $I$  (using a predefined loss function, e.g., the mean square error). Then, further decision trees are linearly added to the regression function  $F$  (scaled by a regularization parameter  $v \ll 1$  called *learning rate*) and the model is fitted again in the same manner. This ensures that the prediction for training points that are poorly modeled in the beginning will gradually improve during the training. The whole procedure is repeated until the loss function does not improve anymore, as evaluated on an independent validation dataset. Relevant hyperparameters of this algorithm (maximum number of decision trees, maximum depth of the decision trees, learning rate, and the subsample ratio) are optimized by 5-fold cross-validation (CV) using the training data. More details on the GBRT algorithm are given in Appendix B.2.

A convenient way to evaluate and interpret GBRT models (or other statistical models in general) is to identify the most relevant predictors. Globally, this so-called *feature importance* is closely related to the number of appearances of a predictor in the different decision trees that set up the GBRT model (Friedman 2001). To evaluate the feature importance regionally (for the different grid cells), we use a model distillation technique called *local interpretable model-agnostic explanations (LIME)*, which approximates the GBRT model with a local linear regression model for each grid cell (Ribeiro et al. 2016). The corresponding linear coefficients  $b_k$  for each predictor  $x^{(k)}$  constitute a good proxy to the model's Jacobian

$$b_k \approx \frac{\partial y}{\partial x^{(k)}}, \quad (6.3)$$

which is difficult to compute explicitly for this particular ML model. Thus, features with high (low) absolute  $b_k$  have a high (low) impact on the local prediction of the target variable  $y$ .

As already described in the previous section, we analyze two different target variables in this study: the monthly climatology of absolute GPP at the end of the 21<sup>st</sup> century (Step 2a) and the fractional GPP change over the 21<sup>st</sup> century (Step 2b). A schematic representation of Step 2a is shown in Figure 6.2. The target variable  $y$  for the ML model is the rescaled monthly climatology of future GPP at the end of the 21<sup>st</sup> century (2091–2100) in the RCP8.5 scenario given by Equation (6.2). We relate this target variable to several process-oriented diagnostics  $\mathbf{x}$  of the past climate (historical simulation) listed in Table 6.1. A single training point  $(\mathbf{x}_i, y_i)$  for the supervised ML algorithm corresponds to the set of the aforementioned diagnostics



Figure 6.2.: Schematic illustration of our machine learning (ML) approach to constrain projected absolute gross primary production (GPP) in Step 2a. (a) In the training phase of the algorithm, the model is fitted to the training data interpolating the empirical (non-linear) relationship between two process-oriented diagnostics of the past climate  $\{x^{(1)}, x^{(2)}\}$  and (rescaled) future GPP (gray surface). The points show the training points for the supervised ML algorithm, each of them representing a single grid cell/month of a single climate model (the different colors correspond to different climate models). (b) In the prediction phase, observation-based values of the diagnostic (black points) are fed into the trained ML model to constrain GPP for every grid cell/month to a value that best agrees with the observations. (c) For independent validation of our method, we use an out-of-sample testing setup based on a leave-one-model-out cross-validation approach (see Section 2.4.2 for details). The schematic illustration of Step 2b differs only in the target variable used (fractional GPP change instead of absolute GPP). Adapted with permission from Schlund et al. (2020a).

Predictor	Observation-driven data	Used time range	Physical connection to GPP
Gross primary production (GPP)	FLUXNET-MTE (Jung et al. 2011)	1991–2000	–
Leaf area index (LAI)	LAI3g (Zhu et al. 2013)	1982–2005	LAI 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).
Precipitation (PR)	CRU (Harris et al. 2014)	1901–2005	Water is essential for the chemical processes of photosynthesis.
Incoming solar radiation at the surface (RSDS)	ERA-Interim (Dee et al. 2011)	1979–2005	Solar radiation is essential for the chemical processes of photosynthesis.
Near-surface air temperature (T)	CRU (Harris et al. 2014)	1901–2005	T and photosynthesis rate have a common driver (incoming solar radiation).

Table 6.1.: Process-oriented diagnostics (also called *predictors* or *features*) used in the gradient boosted regression tree (GBRT) model to predict the target variables. For Step 2a (target variable: absolute GPP), all listed variables are monthly climatologies of the specified time ranges in the historical climate. For Step 2b (target variable: fractional GPP change), the temporal mean over the specified time ranges is calculated for all variables. Adapted with permission from Schlund et al. (2020a).

evaluated on a single grid cell/month of a single climate model (see colored points in Figures 6.2a and 6.2b). Step 2b differs from Step 2a only in the target variable used. Both steps are independent approaches and are evaluated separately.

The main part of steps 2a and 2b each consist of two phases. In the first phase (training of the algorithm), we build a statistical model based on the empirical relationship between observable process-oriented variables of the present-day climate and future GPP in the CMIP5 ensemble (see gray surface in Figures 6.2a and 6.2b). This has a flavor similar to the MDER algorithm (Karpechko et al. 2013; Senftleben et al. 2020; Wenzel et al. 2016b; see Section 2.4.3). However, in contrast to the original MDER algorithm, our multivariate GBRT approach is able to handle multiple predictors with non-linear dependencies. Moreover, instead of considering a single globally or regionally averaged value per climate model, our approach uses spatiotemporally gridded data for every ESM in the ensemble. This dramatically increases the available number of training data (even if they are not statistically independent) and uniquely enables the statistical model to identify and exploit regional processes. Furthermore, the large training database allows us to perform extensive out-of-sample testing, which is crucial to ensure the ML model’s robustness.

In the second phase (predicting with the trained GBRT model), we use observational data products to constrain the RCP8.5 CMIP5 projection of GPP. To do this, observational data for all predictors  $x$  of the historical climate (for every grid cell/month) are fed into the trained GBRT model to get a spatially constrained projection of GPP (see black points in Figure 6.2b). Similar to the MDER algorithm, our approach assumes that there exists an inter-model relationship between the predictors and the target variable which also holds for the true climate. This assumption may seem weak at first glance, but compared to the assumption of traditional weighting approaches (models that are better in simulating the present-day climate are also better in simulating the future climate and vice versa), the MDER assumption is more reasonable and allows correcting models instead of just weighing (Karpechko et al. 2013; see Section 2.4.3). Following other MDER studies (Karpechko et al. 2013; Senftleben et al. 2020; Wenzel et al. 2016b), we also select diagnostics as predictor variables  $x$  that are related to GPP-relevant processes (see Table 6.1 for detailed physical explanations of the links between the predictors and GPP).

The raw CMIP5 and observation-driven datasets are preprocessed before entering the ML algorithm (see Appendix B.1). In particular, increases greater than 300 % in the fractional GPP change that is used as the target variable in Step 2b are removed. These unrealistically high values (making up about 5.9 % of the data) occur in places with very small absolute GPP values in the historical simulation that are used in the denominator of the derivation of the fractional GPP change (see Equation (6.1)). In addition, we randomly select 25 % of the training data and use it as hold-out test data. This dataset is excluded from the training and can therefore be used for independent validation of our statistical model.

### 6.1.3. Assessment of Uncertainty and Predictive Skill

The estimation of the standard prediction error (SPE) for the GBRT models needs to consider three sources of uncertainty: the uncertainty of the ML model itself, the uncertainty in the rescaling of the target variable, and the uncertainty in the input data, that is, errors in the observation-based products. First, the SPE of the GBRT model itself (i.e., uncertainty in the position of the gray surface in Figures 6.2a and 6.2b) is estimated as the *root mean square error of prediction (RMSEP)*, that is, the RMSE between the predicted and true values of the independent hold-out test dataset (Bishop 2006). This approach is justified when the distribution of residuals is unbiased (see Appendix B.4), in which case the RMSEP equals the standard deviation of the residuals (which we interpret as error). Figure B.2 shows that this is approximately correct.

The SPE of the GBRT model can be understood as combined error due to the internal variability of the climate system, the different model responses within the climate model ensemble to forcing, and the incomplete description of the climate system with the limited number of variables used as input features. Second, the uncertainty in the rescaling is simply assessed as the uncertainty given by the emergent constraint from Step 1. Finally, the uncertainty in the observation-based data (i.e., uncertainty in the position of the black points in Figure 6.2b) is

assessed by error propagation. The necessary sensitivities of the target variable to the various predictors for each grid cell/month are given by the linear coefficients  $b_k$  given by the LIME approach introduced in Section 6.1.2. All sources of uncertainty are assumed independent and are combined by adding the squared errors, yielding a single estimate of the SPE for each grid cell/month. Further details on the calculation of errors are given in Appendix B.3.

The predictive power of the GBRT algorithm and its robustness is assessed using a *leave-one-model-out* CV approach (de Elia et al. 2002; Karpechko et al. 2013; see Figure 6.2c and Section 2.4.2). For this, a single climate model is removed from the multi-model ensemble and considered as being the true climate. Our statistical model is then fitted on all remaining climate models and a prediction for the “true” model is created. This allows a simple evaluation of the predictive power by calculating the RMSEP. The whole process is repeated for every climate model of the ensemble to get a distribution of RMSEPs. Since this approach is not limited to GBRT, it allows a simple comparison of the predictive skill of different statistical models. We compare the GBRT model to the simple CMIP5 multi-model mean (MMM) and a least absolute shrinkage and selection operator (LASSO) model (Tibshirani 1996). The LASSO model is an extended linear regression model that regularizes the least squares solution with an L1 penalty term. In contrast to the ordinary least squares regression, the LASSO model promotes sparse solutions (i.e., only a few non-zero coefficients/weights), which acts as an intrinsic feature selection and provides an effective way to combat overfitting induced by collinearity.

## 6.2. Results

### 6.2.1. Constraining Global Mean GPP Projections (Step 1)

In Wenzel et al. (2016a), the observed atmospheric CO<sub>2</sub> concentration at KUM (Keeling et al. 2005) is used to constrain the GPP change in the 1%BGC run resulting from a doubling of atmospheric CO<sub>2</sub> concentrations to (32 ± 9) % for extratropical ecosystems (30–90 °N). As mentioned in Section 6.1.1, we argue that due to the strong annual mixing of CO<sub>2</sub> in the atmosphere, this emergent constraint can also be applied to the global mean GPP change if the emergent relationship holds. To test this, we first use the idealized CMIP5 simulations that were analyzed by Wenzel et al. (2016a) to calculate the emergent constraint (Figure 6.3a). The relationship is statistically significant at a 5 % significance level ( $R^2 = 0.79$ ,  $p = 0.007$ ) and predicts a constrained global mean GPP change of (30 ± 9) %, which is consistent with the findings of Wenzel et al. (2016a). This result gives us confidence that the emergent relationship holds for the global mean GPP. Details on the mathematical derivation of the constrained range are given in Section 2.4.1.

In addition to applying it to the idealized simulations, we apply the emergent constraint to the RCP8.5 scenario. In the 1%BGC simulation, the CO<sub>2</sub> fertilization effect is the only driver of future GPP change since the carbon cycle components of the ESMs only see the increasing CO<sub>2</sub> concentration. However, since the magnitude of the concentration-carbon feedback is



Figure 6.3.: Emergent relationship between the global mean fractional change in gross primary production (GPP) and the sensitivity of the CO<sub>2</sub> seasonal cycle amplitude to rising atmospheric CO<sub>2</sub> concentrations observed at Cape Kumukahi, Hawaii (KUM). Colored points refer to CMIP5 models, the orange line and shaded area to the linear regression fit and its corresponding standard prediction error, and the dashed black lines to the observational constraint. (a) Similar to Wenzel et al. (2016a): The global mean fractional GPP change after CO<sub>2</sub> doubling in a biogeochemically coupled simulation where the atmospheric CO<sub>2</sub> concentration increases by 1 % per year (1% BGC) defined as the 10-year mean of the 1% BGC run centered at the time of CO<sub>2</sub> doubling relative to the 10-year mean of pre-industrial control conditions at the beginning of the 1% BGC run. The sensitivity of the CO<sub>2</sub> amplitude for the CMIP5 models is calculated from the years 1860–2005, the corresponding observational value from the years 1979–2019. The constrained global mean GPP change after CO<sub>2</sub> doubling in the 1% BGC run is (30 ± 9) %. (b) Fractional global mean GPP change over the 21<sup>st</sup> century calculated from the 10-year mean GPP at the end of the 21<sup>st</sup> century (2091–2100) in the emission-driven fully coupled RCP8.5 simulations and the 10-year mean GPP at the end of the 20<sup>th</sup> century (1991–2000) in the emission-driven fully coupled historical run (see Equation (6.1)). In contrast to Wenzel et al. (2016a), the sensitivity of the CO<sub>2</sub> amplitude is calculated from the years 1979–2019 for climate models and observations for better comparability (the historical CMIP5 simulations are extended with the RCP8.5 simulations for the years 2006–2019). The constrained global mean GPP change over the 21<sup>st</sup> century is (39 ± 7) %. Adapted with permission from Schlund et al. (2020a).

believed to be four times the size of the climate-carbon feedback (Gregory et al. 2009), the CO<sub>2</sub> fertilization effect is still the dominant driver of GPP change in the fully coupled simulations (Huntzinger et al. 2017). This is also supported by Wenzel et al. (2016a), who show that the GPP increase due to rising atmospheric CO<sub>2</sub> concentrations in the fully coupled historical simulations is proportional to the same quantity in the 1% BGC run. Additional lines of evidence suggest that the modeled CO<sub>2</sub> sensitivity of photosynthesis is the main source of uncertainty of future GPP (Arora et al. 2013; Haverd et al. 2020; Rogers et al. 2017). Thus, the sensitivity of the CO<sub>2</sub> amplitude can be used to approximately constrain global mean GPP change over the 21<sup>st</sup> century in RCP8.5, which is shown in Figure 6.3b. Assuming a significance level of 5 %, the emergent relationship is statistically significant ( $R^2 = 0.65, p = 0.029$ ).

Using the emergent relationship, the global mean GPP change over the 21<sup>st</sup> century can be constrained to (39 ± 7) % (standard error), which is at the lower end of the original CMIP5 range (31–57 %). Moreover, the best estimate of 39 % is also slightly lower than the CMIP5 MMM of 43 %. The observed GPP averaged over the years 1991–2000 from the FLUXNET-

MTE product is  $(123 \pm 6)$  GtC yr $^{-1}$  (Jung et al. 2011). Assuming independent errors and using Gaussian error propagation, the constrained fractional change corresponds to a global mean GPP of  $(171 \pm 12)$  GtC yr $^{-1}$  (standard error) at the end of the 21<sup>st</sup> century (2091–2100) in the RCP8.5 scenario. In contrast to that, the unconstrained CMIP5 ensemble has a model range of 156–247 GtC yr $^{-1}$  for future GPP in the RCP8.5 scenario (2091–2100). The resulting global mean GPP change is used to rescale the gridded output of the different climate models with Equation (6.2). The following two sections show the results for the rescaled absolute future GPP (Step 2a) and the rescaled fractional GPP change (Step 2b) further constrained by the GBRT model.

### 6.2.2. Constraining Gridded Absolute GPP Projections (Step 2a)

In the second step, the objective is to constrain the spatial distribution of the projected GPP. We use the grid-wise monthly mean climatology of absolute GPP at the end of the 21<sup>st</sup> century (2091–2100) as the target variable, rescaled using Equation (6.2) to reduce the large uncertainties in the CO<sub>2</sub> fertilization effect in the models. In the first part, we evaluate the GBRT model using the leave-one-model-out CV approach and compare it to other statistical approaches and the unweighted MMM. In the second part, we use observation-based products to predict the target variable.

#### Prediction Error in Leave-One-Model-Out CV Approach and Feature Importance

To get a detailed insight into the performance of our GBRT model, we compare it to five other statistical models: the unweighted CMIP5 MMM of future GPP, its rescaled version using Equation (6.2) (rMMM), a linear LASSO regression model using all features as defined in Table 6.1, a single-predictor LASSO model (LASSO-1D), and a single-predictor GBRT model (GBRT-1D). Both single-predictor models use the historical GPP as the only feature. The predictive power (in terms of RMSEP) of each statistical model is assessed using the leave-one-model-out CV approach (see Section 6.1.3). This allows us to create an RMSEP distribution for each statistical approach, where the different points of the distributions refer to different training/prediction dataset combinations generated by the leave-one-model-out CV approach. Thus, each distribution consists of seven points, one for each climate model. Figure 6.4a shows the RMSEP distributions for the six different statistical models.

In terms of raw predictive power, the simple MMM is outperformed by every other model. Its prediction uncertainty expressed as mean RMSEP can be reduced by more than 48 % by using other statistical models. However, this is not surprising: In contrast to the other statistical models, the simple RCP8.5 MMM does not take further evidence in form of observations of the historical climate into account. Step 1 of our algorithm (rMMM) can reduce this mean RMSEP by 15 % due to the simple rescaling of the gridded climate model output with the global mean GPP constraint on the CO<sub>2</sub> fertilization effect. However, since there is a considerably large GPP range in the individual climate models themselves, this reduction is rather small for the gridded values. A far larger reduction of the RMSEP can be achieved by using the



Figure 6.4.: (a) Box plot of the root mean square error of prediction (RMSEP) distributions for six different statistical models used to predict future absolute gross primary production (GPP) in Step 2a using a leave-one-model-out cross-validation approach. The distribution for each statistical model contains seven elements (black points; one for each climate model used as truth) and is represented in the following way: The lower and upper limit of the blue boxes correspond to the 25 % and 75 % quantiles, respectively. The central line in the box shows the median and the black “x” shows the mean of the distribution. The whiskers outside the box represent the range of the distribution. Compared to the CMIP5 multi-model mean (MMM) and its corresponding rescaled version (rMMM), the prediction uncertainty measured by the mean RMSEP is significantly reduced by up to 48 % and 39 %, respectively, when using other statistical models. Moreover, the non-linear gradient boosted regression tree (GBRT) models can slightly reduce the mean RMSEP compared to the linear least absolute shrinkage and selection operator (LASSO) models by about 2 % for GBRT-1D (using historical GPP as single predictor) and 3 % for the full GBRT model (using all predictors). (b) Relative global feature importance for the different predictors used in the GBRT model to predict future absolute GPP (Step 2a). The red bars correspond to positive Pearson correlation coefficients between all predictors and the target variable. Due to its strong positive linear relationship with the future GPP, the historical GPP is by far the most important predictor in the model. A local feature importance map using local interpretable model-agnostic explanations (LIME) is not shown here because GPP is the dominant predictor for all grid cells. Adapted with permission from Schlund et al. (2020a).

regression models LASSO-1D, LASSO, GBRT-1D, and GBRT. All of them share similar RMSEP distributions, which can be explained as follows: The historical GPP is strongly (near linearly) correlated to the rescaled end-of-century GPP (with pattern correlation of  $R^2 = 0.83$  for the whole multi-model ensemble). Because of that, all regression models are able to considerably reduce the RMSEPs compared to the MMM and rMMM. This also shows that the GBRT models successfully learn the linear connection between past and future GPP. Moreover, the non-linear GBRT models can slightly reduce the mean RMSEP even further compared to the linear LASSO models by about 2 % for GBRT-1D and 3 % for the full GBRT model. This leads to the conclusion that in addition to the strong linear relationship between past and future GPP, there are small non-linear relations between the predictors and the target variable

that can be used to further reduce the RMSEP. Since the full GBRT model with access to all predictors shows the minimal mean RMSEP, we argue that it is beneficial for our approach to use the multivariate non-linear GBRT model instead of a linear model because only the GBRT algorithm is able to make use of all predictors and exploit more complex relationships.

The influence of the different predictors can be further analyzed by evaluating the global feature importance by using the whole training dataset and the already trained model. This is shown in Figure 6.4b. The past GPP with its strong positive linear correlation to future GPP is the most relevant predictor of the trained model with a relative importance of approximately 95 %. All other features show values of less than 2 %. The global feature importance determines the expected impact of a feature in predictions of the target variable. In our GBRT model, the prediction input data of historical GPP mainly determines the ML model's prediction of the future GPP in the RCP8.5 scenario at the end of the 21<sup>st</sup> century. However, we emphasize that this is only valid for our specific setup (i.e., which variables are used, which algorithm is used, which climate models are considered, etc.). For another problem setup, the global feature importance could change significantly. Since the historical GPP is the most important predictor for all grid cells, we do not show the plot of the local feature importance using LIME. We can get further insights into our GBRT model by analyzing the residuals of the prediction using an independent test dataset. The resulting plots show that the algorithm is not overfitting the training data and that the prediction errors are approximately unbiased (see Figure B.2a).

### Observation-Based GBRT Prediction of Absolute GPP

In this section, we use the different statistical models to predict absolute GPP at the end of the 21<sup>st</sup> century. For this, we feed observation-based data (see Table 6.1) into the regression models GBRT, GBRT-1D, LASSO, and LASSO-1D.

The result of Step 1 on the gridded data is illustrated in Figure 6.5a, which shows the ratio between rMMM and MMM. This ratio is almost constant over the whole globe, which is not surprising due to the global nature of the rescaling (see Equation (6.2)). The use of the global emergent constraint predicts a slightly lower GPP increase over the 21<sup>st</sup> century (39 %) than the unweighted CMIP5 ensemble mean (43 %). The corresponding ratio of 39 % / 43 %  $\approx$  0.91 is approximately equal to the mean value of rMMM/MMM (0.92). Consequently, the global estimate of rMMM gives a lower GPP at the end of the century ( $179 \text{ GtC yr}^{-1}$ ) as MMM ( $198 \text{ GtC yr}^{-1}$ ), which can be interpreted as a correction of the ESMs' overestimation of future GPP to changes in the atmospheric CO<sub>2</sub> concentration.

The spatial result of our GBRT model in Step 2a can be visualized by comparing rMMM and the output of the GBRT model. Figure 6.5b shows the bias of the rescaled CMIP5 RCP8.5 MMM (rMMM) compared to our GBRT-predicted GPP at the end of the 21<sup>st</sup> century (2091–2100), while Figure 6.5c shows the bias of the historical CMIP5 MMM compared to the observational FLUXNET-MTE product (Jung et al. 2011) averaged for the period 1991–2000. As reported by Anav et al. (2013), the historical CMIP5 ensemble mean overestimates GPP in most regions,



Figure 6.5.: (a) Ratio of the rescaled CMIP5 ensemble mean of the absolute gross primary production (GPP) at the end of the 21<sup>st</sup> century (2091–2100) in the RCP8.5 scenario using Equation (6.2) (rMMM) and its unweighted version (MMM). The plot shows an almost constant value over the whole globe with a mean of 0.92, which corresponds to the ratio of the constrained global mean GPP change over the 21<sup>st</sup> century from Step 1 (39 %) and the CMIP5 ensemble mean global mean GPP change (43 %). All values close to 0 for the dataset in the denominator have been masked to avoid divisions by 0. (b) Bias between rMMM and our gradient boosted regression tree (GBRT) prediction for the end of the 21<sup>st</sup> century. This corresponds to Step 2a of our approach. (c) Bias between the modeled GPP in the CMIP5 MMM of the historical simulation and the FLUXNET-MTE observation-based estimate of GPP (Jung et al. 2011) averaged between 1991 and 2000. Over large swaths of the globe, the CMIP5 ensemble overestimates GPP (red color). Panels (b) and (c) show similar bias patterns (pattern correlation of  $R^2 = 0.88$ ). Thus, the GBRT prediction in Step 2a is able to correct the historical bias of the CMIP5 ensemble relative to the FLUXNET-MTE product. Adapted with permission from Schlund et al. (2020a).

leading to a global mean of  $138 \text{ GtC yr}^{-1}$ , which is larger than the  $123 \text{ GtC yr}^{-1}$  estimate for the FLUXNET-MTE product. Regions where GPP is largely overestimated are the western parts of South America, central and southern Africa, and East Asia. On the contrary, GPP is underestimated in small areas of Central America and northern parts of South America. The bias patterns in Figures 6.5b and 6.5c are very similar (pattern correlation of  $R^2 = 0.88$ ),



Figure 6.6.: Difference (a) and ratio (b) of the biases shown in Figures 6.5b and 6.5c. For panel (b), all values close to 0 for the dataset in the denominator have been masked to avoid divisions by 0. Both panels show that our approach is only to first order a bias correction, in which case both plots would only show constant values. (c) Comparison of the gradient boosted regression tree (GBRT) versus GBRT-1D projections of future gross primary production (GPP). This panel indicates a clear difference between the full GBRT model using all predictors and the GBRT model using only historical GPP as a single predictor. (d) Comparison of the GBRT versus the least absolute shrinkage and selection operator (LASSO) projections of future GPP. This panel shows that there is a clear difference between using the non-linear GBRT model and the linear LASSO model. The results of the LASSO-1D model are not shown here because they are very similar to the results of the LASSO model. Adapted with permission from Schlund et al. (2020a).

which means that the GBRT algorithm detects regional biases in the historical simulation and corrects them in its future predictions. This is also illustrated in Figure 6.6b, in which many values are close to 1, while only a few negative values exist. However, we emphasize that our approach is only to first order a bias correction (i.e., subtracting the historical bias from the RCP8.5 MMM). There are clear differences between Figures 6.5b and 6.5c, which are also illustrated in Figures 6.6a and 6.6b. A simple bias correction would only show constant values for the whole globe, whereas our approach applies different corrections (in sign and magnitude) for different regions. On the global domain, our approach predicts a GPP at the

end of the century of  $169 \text{ GtC yr}^{-1}$ , which is consistent with the global constraint from Step 1 of  $(171 \pm 12) \text{ GtC yr}^{-1}$ . In summary, our approach first corrects the CMIP5 models' response to CO<sub>2</sub> (Step 1) and second corrects the historical bias of the ESMs relative to observations (Step 2a).

Further illustrations of these results including the uncertainties can be found in Figures B.3–B.5. Analogous to the results from the leave-one-model-out CV approach, the uncertainties on grid-cell level are significantly reduced for the GBRT prediction compared to the MMM results. These errors consider all three sources of uncertainty that we present in Section 6.1.3 (error in the GBRT model, error in the rescaling, and error in the observational-driven products). In contrast to that, the error derived from the leave-one-model-out CV approach illustrated in Figure 6.4a only shows the error in the statistical models themselves. We note that local relative errors can be very large (especially in regions with low absolute GPP) due to the uncertainty calculation method: Regardless of the absolute value, the error of each grid cell is at least  $535 \text{ gC m}^{-2} \text{ yr}^{-1}$  (estimated SPE of the GBRT model using the RMSEP), which is added to the propagated errors of the observation-driven datasets and the error derived from the rescaling using Equation (B.5). The GBRT model error is the dominant source of uncertainty. Since the true covariance structure of these global fields is unknown, a global (or at least regional) aggregation of these uncertainties is not possible.

To further compare our GBRT approach to other statistical models, we also fed the observation-driven data into the GBRT-1D, LASSO, and LASSO-1D models to get a constrained projection of the RCP8.5 GPP. Figures 6.6c and 6.6d show the difference of the other models to the full GBRT approach. Both panels show values close to 0 for most of the globe (i.e., the differences between the GBRT and GBRT-1D/LASSO models are small), indicating that the bias-correcting property of our approach is also present in the other regression models that use only the historical GPP as single predictors. Thus, we conclude that the bias correction of our GBRT model originates from the observation of the historical GPP, which is also supported by Figure 6.4b, which shows historical GPP as the by far most important feature. On the contrary, second-order corrections originate from the observations in the remaining predictors. These second-order corrections are also visible as the non-zero values in Figures 6.6c and 6.6d. This indicates that using the non-linear GBRT model with all predictors improves the final result. Moreover, the similar patterns in both panels demonstrate that the GBRT-1D model emulates the strong linear relation of the historical GPP to future GPP and performs equally well as the linear models. The global estimates of  $167 \text{ GtC yr}^{-1}$  (GBRT-1D),  $163 \text{ GtC yr}^{-1}$  (LASSO), and  $162 \text{ GtC yr}^{-1}$  (LASSO-1D) are all consistent with the global constraint of  $(171 \pm 12) \text{ GtC yr}^{-1}$ .

### 6.2.3. Constraining Gridded Fractional GPP Change Projections (Step 2b)

In Step 2b, we constrain the gridded fractional GPP change over the 21<sup>st</sup> century (2100 versus 2000) in the emission-driven RCP8.5 scenario (rescaled using Equation (6.2)). This step is independent of Step 2a and only differs in the target variable (GPP change instead of absolute

GPP). The remaining setup including the predictors, the GBRT model itself, and the datasets are similar.

### Prediction Error in Leave-One-Model-Out CV Approach and Feature Importance

Figure 6.7a shows the RMSEP distributions for four different statistical models: the simple CMIP5 MMM of the fractional GPP change, its rescaled version using Equation (6.2) (rMMM), a linear LASSO regression model, and the GBRT model. Similar to Step 2a, the GBRT approach shows the smallest mean RMSEP values of all statistical models. Compared to MMM and rMMM the mean is reduced by 16 % and 9 %, respectively. Moreover, the non-linear GBRT model also outperforms the linear LASSO model (mean reduced by 3 %). In contrast to Step 2a, there is not a single predictor which heavily dominates the feature importance of the GBRT model. Thus, using single-predictor models (GBRT-1D and LASSO-1D) is not necessary for Step 2b. Figures 6.7b and 6.7d show the global feature importance for the LASSO and GBRT model, respectively. Both models agree that the near-surface air temperature (T) and LAI are the two dominant predictors with a relative importance of more than 65 %. The LASSO model shows a relative importance of less than 10 % for each of the remaining features, while the GBRT model exhibits a considerably high relative importance for GPP with almost 20 %. All predictors are negatively correlated to the target variable (as given by the sign of the linear Pearson correlation coefficient). That means cold areas with a small LAI in the historical climate show a larger GPP change over the 21<sup>st</sup> century and vice versa. Indeed, the fertilization effect is expected to be stronger in lower LAI regions. To get more detailed insights on this, observation-driven data is fed into the GBRT model to obtain a constrained fractional GPP change over the 21<sup>st</sup> century.

### Observation-Based GBRT Prediction of Relative GPP

The global distribution of the fractional GPP change for the different statistical models is shown in Figure 6.8. All panels show a GPP increase over the 21<sup>st</sup> century in the RCP8.5 scenario for almost all regions of the globe. Compared to the unweighted CMIP5 MMM, its rescaled version (rMMM) shows a slightly lower GPP increase, which is consistent with the emergent constraint approach from Step 1 that shows a lower global mean GPP increase over the 21<sup>st</sup> century (39 %) than the global CMIP5 ensemble mean (43 %). This is also illustrated in Figure 6.9a, which shows an almost constant pattern over the whole globe with a mean of 0.91 that is consistent with the ratio 39 % / 43 %  $\approx$  0.91. Similar to Step 2a, this can be interpreted as a global correction of the ESMs' response of future GPP to changes in the atmospheric CO<sub>2</sub> concentration.

All four patterns in Figure 6.8 show very similar geographical patterns. Obvious exceptions are the Sahara Desert and the Arabian Peninsula, which show a noisy behavior for the multi-model means (MMM and rMMM) and a small GPP increase for the two regression models (LASSO and GBRT). This noise-like pattern occurs due to numerical inconsistencies produced by very small values in the absolute historical and future GPP in the climate models in this



Figure 6.7.: (a) Box plot of the root mean square error of prediction (RMSEP) distributions for four different statistical models used to predict the fractional change in gross primary production (GPP) over the 21<sup>st</sup> century (Step 2b) using the leave-one-model-out cross-validation approach. A detailed description of the representation of these distributions is given in Figure 6.4a. The gradient boosted regression tree (GBRT) algorithm shows the minimal mean and median RMSEP. Compared to the CMIP5 multi-model mean (MMM), its corresponding rescaled version (rMMM), and the linear least absolute shrinkage and selection operator (LASSO) model, the mean RMSEP of the GBRT model is reduced by more than 16 %, 9 %, and 3 %, respectively. (b) Relative global feature importance for the different features used in the LASSO model to predict the fractional GPP change (Step 2b). The feature importance for the LASSO model is given by the normalized linear coefficients of the model. (c) Local feature importance for the GBRT model used to predict fractional GPP change (Step 2b) calculated using the local interpretable model-agnostic explanations (LIME) approach (Ribeiro et al. 2016) for the three dominant features near-surface air temperature (T), leaf area index (LAI), and GPP. The relative influence of these three features (ignoring all other features) is color-coded according to the triangle in the lower-left corner. Over large parts of the globe, T is the dominant feature. (d) Relative global feature importance for the different features used in the GBRT model to predict the fractional GPP change (Step 2b). The blue bars for the two global feature importance plots (b) and (d) correspond to negative Pearson correlation coefficients between all predictors and the target variable. Adapted with permission from Schlund et al. (2020a).



Figure 6.8.: Fractional change in gross primary production (GPP) over the 21<sup>st</sup> century (2100 versus 2000) in the RCP8.5 scenario for different statistical models (Step 2b): (a) CMIP5 multi-model mean (MMM) of the fractional GPP change, (b) rescaled CMIP5 multi-model mean (rMMM) using Equation (6.2), (c) linear least absolute shrinkage and selection operator (LASSO) model, and (d) gradient boosted regression tree (GBRT) model. The geographical patterns from the different statistical models are very similar and show a higher GPP increase in high latitudes and a lower GPP increase in regions closer to the equator. Adapted with permission from Schlund et al. (2020a).

region that are used to derive the fractional GPP change following Equation (6.1). However, due to the small impact on the global mean GPP, this region is negligible in our analysis. The fractional GPP change is not uniformly distributed over the globe. Instead, there is a pronounced latitudinal dependency in its geographical pattern: In high latitudes, the projected GPP change is larger than in regions closer to the equator. This effect is particularly strong for the northern high latitudes and consistent with the results of Wenzel et al. (2016a), who find an increased GPP change in northern high-latitude ecosystems compared to the northern extratropics. A possible reason for this is the extension of the growing season in high latitudes caused by climate change connected to a *greening*, which has already been observed in the past climate through satellite measurements (Lucht et al. 2002; Myneni et al. 1997; Zhang et al. 2020) and associated with the increase in the CO<sub>2</sub> seasonal cycle amplitude (Forkel et al.



Figure 6.9.: (a) Ratio of the rescaled CMIP5 ensemble mean of the fractional change in gross primary production (GPP) over the 21<sup>st</sup> century using Equation (6.2) (rMMM; see Figure 6.8b) and its unweighted version (MMM; see Figure 6.8a). The plot shows an almost constant value over the whole globe with a mean of 0.91, which corresponds to the ratio of the constrained global mean GPP change over the 21<sup>st</sup> century from Step 1 (39 %) and the CMIP5 ensemble mean global mean GPP change (43 %). All values close to 0 for the dataset in the denominator have been masked to avoid divisions by 0. Absolute (b) and relative (c) differences between the fractional GPP change over the 21<sup>st</sup> century predicted by the gradient boosted regression tree (GBRT) model (see Figure 6.8d) and rMMM (see Figure 6.8b). For panel (c), all values close to 0 for the dataset in the denominator have been masked to avoid divisions by 0. Both plots show a good agreement over large parts of the globe (corresponding to values of 0). There are large absolute differences in the Sahara region and central Asia. The largest relative differences (except for the Sahara and Arabian Peninsula region) appear over South America, South Africa, the west coast of Africa, the Middle East, parts of Australia, and western parts of the United States. Adapted with permission from Schlund et al. (2020a).

2016). This is consistent with the feature map in Figure 6.7c, which shows an increased relative importance of the LAI (a quantity directly related to the greening) in the high-latitude regions with high GPP changes. On the contrary, in tropical ecosystems, the growing season already covers the whole year. Thus, a further climate change-induced extension is not possible, leading to a smaller fractional GPP change in regions closer to the equator. For most parts

of the remaining globe, T is the dominant predictor, with some areas in northern Africa, the Middle East, India, and Australia showing an increased relative importance of the historical GPP. Overall, the negative correlation of the three most important features (T, LAI, and historical GPP) to the target variable is well reflected in the global pattern of the fractional GPP change: The high latitudes with lower T, lower LAI, and lower historical GPP show higher GPP changes, whereas regions closer to the equator with higher T, higher LAI, and higher historical GPP show smaller GPP changes. At this point we want to emphasize that it is not possible to infer global or other large-scale averages of the fractional GPP change from the geographical distributions in Figure 6.8 since division operations and averaging operations do not commute: in general,  $\sum_i (a_i/b_i) \neq (\sum_i a_i) / (\sum_i b_i)$ . Thus, a direct comparison to other studies that give locally or globally aggregated results is not possible.

To get further insights into the GBRT algorithm, we compare it directly to the rescaled CMIP5 MMM (rMMM), which is illustrated in Figures 6.9b and 6.9c. The mismatch between GBRT and rMMM can be interpreted as additional information that is added to the target variable by the observations in the predictors in Step 2b. Overall, the absolute difference (Figure 6.9b) is small over the whole globe, indicating that the rMMM and GBRT agree in the general pattern of the fractional GPP change. The most striking deviation in this panel is again the Sahara and Arabian Peninsula region, which shows the already discussed large absolute differences. Apart from that, there is a large patch in central Asia that shows high negative differences, that is, the GBRT model predicts a smaller GPP change in this area. The relative difference (Figure 6.9c) indicates many regions where the GBRT model predicts an increased GPP change compared to rMMM, in particular South America, South Africa, the west coast of Africa, the Middle East, parts of Australia, and western parts of the United States. However, all these correspond to low absolute GPP changes. Figure B.6 shows the corresponding gridded uncertainties for all four statistical models that also cover the uncertainty in the emergent constraint and the observational uncertainty (in contrast to Figure 6.7a, which only covers the uncertainty in the statistical models). Similar to Step 2a, the local standard errors are large, even for the GBRT model (that performs better than any other statistical model) which shows errors of at least 43.6 % for every grid cell caused by the uncertainty in the statistical model itself. Figure B.2b illustrates the residuals for the training and independent test datasets and shows that the GBRT model has symmetrical residuals and is not overfitting.

#### 6.2.4. Comparison of the Absolute and Relative GBRT Prediction

In a last step, we use the FLUXNET-MTE observational product to constrain the absolute GPP increase at the end of the 21<sup>st</sup> century (2091–2100) from the fractional GPP change given by the GBRT model in Step 2b. This can be directly compared to the output of Step 2a and serves as a sanity check between steps 2a and 2b. As shown by Figure 6.10, both approaches show similar geographical patterns (pattern correlation:  $R^2 = 0.97$ ). The difference between the two approaches is in the same order of magnitude as the difference between the different statistical models used in Step 2a (see Figures 6.6c and 6.6d) and considerably smaller than the



Figure 6.10.: Top row: Absolute future gross primary production (GPP) at the end of the 21<sup>st</sup> century (2091–2100) calculated using Step 2a (a) and Step 2b (b). Both approaches give similar results with global averages of  $169 \text{ GtC yr}^{-1}$  (a) and  $175 \text{ GtC yr}^{-1}$  (b), which are both consistent with the global result of Step 1 of  $(171 \pm 12) \text{ GtC yr}^{-1}$ . The pattern correlation between both approaches is  $R^2 = 0.97$ . Bottom row: Absolute (c) and relative (d) differences between panels (a) and (b). Adapted with permission from Schlund et al. (2020a).

difference between the GBRT and rMMM approaches (see Figure 6.5b). Moreover, the globally aggregated results ( $169 \text{ GtC yr}^{-1}$  for Step 2a and  $175 \text{ GtC yr}^{-1}$  for Step 2b) are both compatible with the global value of  $(171 \pm 12) \text{ GtC yr}^{-1}$  given by the global emergent constraint in Step 1.

### 6.3. Summary and Discussion

In this chapter, a two-step approach to constrain the projected GPP at the end of the 21<sup>st</sup> century (2091–2100) in the RCP8.5 scenario is developed. In the first step, we constrain the global mean GPP to  $(171 \pm 12) \text{ GtC yr}^{-1}$  using a published emergent constraint approach which we assume can also be applied to the global mean GPP at the end of the 21<sup>st</sup> century in the RCP8.5 scenario (see discussion in Section 6.1.1). This step corresponds to the correction of the ESMs' biases in the response of future GPP to rising atmospheric CO<sub>2</sub> concentration, which is the main

source of uncertainty in future GPP projections (Arora et al. 2013; Haverd et al. 2020; Rogers et al. 2017). As already discussed in Section 6.1.1, due to other than CO<sub>2</sub> fertilization drivers of the increase in the CO<sub>2</sub> seasonal cycle amplitude (Bastos et al. 2019; Forkel et al. 2016; Piao et al. 2018; Zhao et al. 2016) this emergent constraint is not undisputed and can be replaced with another emergent constraint in future studies. In the second step, an ML approach is used to further constrain the gridded GPP based on process-based present-day predictors. We consider two target variables: first (Step 2a) the gridded monthly climatologies of absolute GPP (2091–2100) and second (Step 2b) the gridded fractional GPP change over the 21<sup>st</sup> century (years 2091–2100 versus years 1991–2000; see Equation (6.1)). Both approaches give consistent results (see Figure 6.10). The latter quantity shows a higher GPP increase in high latitudes and a lower GPP increase closer to the equator, which might be attributed to an additional greening trend in the high latitudes as supported by the feature map in Figure 6.7c that shows LAI as the dominant predictor in this region. However, in the absence of a robust physical mechanism, this connection cannot be proven. For both target variables, the GBRT algorithm is superior to all other considered statistical models in terms of prediction uncertainty evaluated in a leave-one-model-out CV approach among different statistical models. Compared to the unweighted CMIP5 MMM the mean prediction error is reduced by 48 % (Step 2a) and 16 % (Step 2b); compared to a linear LASSO regression model the mean prediction error is reduced by approximately 3 % for both cases. However, local standard errors are still large for both target variables, even for the GBRT model. Due to the unknown covariance structure, a global aggregation of these errors is not possible. Step 2a mainly corrects the bias of the simulated absolute historical GPP relative to observations (Anav et al. 2013). Consequently, the historical GPP is by far the most important predictor for future absolute GPP (regions with high GPP in the past are likely to have high GPP in the future and vice versa). However, as we show in Figure 6.6, the GBRT model expands this bias correction by taking more predictors than the historical GPP into account. A similar result is provided by Step 2b: Figures 6.9b and 6.9c show the impact of the additional predictors as difference between the CMIP5 MMM and the GBRT prediction of the GPP change. In this case, the target variable (fractional GPP change) is already normalized with the historical GPP. Therefore, T (the second most important feature in Step 2a) is now the most important predictor instead of historical GPP (see Figure 6.7d). As shown in Figure 6.7c, the different features have also different dominant regional impacts; but again, since there is no distinct physical mechanism relating the predictors to the target variable this map requires further analysis and needs to be treated with care.

Our GBRT approach (Step 2) is mathematically similar to MDER (Karpechko et al. 2013; Seneftleben et al. 2020; Wenzel et al. 2016b): We establish a relationship between process-oriented, physically relevant diagnostics and future projections and then utilize this to project today's observed conditions into the future. We emphasize that the exact nature of this relationship is strongly dependent on the climate model ensemble considered. Although derived empirically, we argue that this relationship can still be used to extract further information from the climate model ensemble for two reasons: First, we consider process-oriented variables which are physically linked to GPP, and second, we ensure its statistical robustness. The latter is achieved

by considering gridded climatological data instead of global or regional means. The number of points establishing the relation is dramatically increased from about 50 points for classical emergent constraints to 237,852 points (Step 2a) and 16,503 (Step 2b) in our approach. We validate the relationship by extensive out-of-sample testing using a leave-one-model-out CV approach and randomly excluding parts of the data prior to training. Furthermore, the large number of points in combination with the non-linear GBRT model allows us to exploit more information than classical emergent constraints and include non-linear multivariate relations in the statistical model, including the information encoded in spatial variation. The prediction phase of our approach (feeding observation-based data into the trained GBRT model) can be interpreted as an implicit performance weighting: The GBRT model creates predictions based on today's conditions (in form of variables that are physically connected to GPP). However, unlike other performance weighting techniques (Knutti et al. 2017b; Sanderson et al. 2017), we do not assign constant weights to the different climate models. On the contrary, since the algorithm works on grid-cell level, it is able to adapt to regional climate characteristics and thus applies an implicit local and climate-dependent weighting. Moreover, the algorithm adds another level of weighting by assigning a relative relevance score to the different predictors. We emphasize that this is only an implicit weighting since it is not possible to extract specific values for the individual weights due to the complex structure of the GBRT model.

Yet, like emergent constraints and performance weighting techniques, our GBRT approach in Step 2 hinges on the assumption that the climate models reflect the real world. This is certainly only partially true since there are processes like CO<sub>2</sub> fertilization, land-use and land-cover changes, nutrient cycles and limitations (Du et al. 2020; Fleischer et al. 2019), disturbances, and induced vegetation dynamics that may strongly alter the future trajectories of GPP without this being fully encoded in today's observational products. Moreover, all CO<sub>2</sub>-related effects are only implicitly incorporated in our approach by the target period we are using and the global rescaling using the emergent constraint in Step 1. Thus, our reduction of uncertainty only applies to the specific setup we are considering here: the future GPP in the CMIP5 RCP8.5 scenario at the end of the 21<sup>st</sup> century. It has to be viewed as a nominal reduction, while the real uncertainty remains unknown and is likely larger than the nominal one. These issues could be tested in the future by including other (even offline) models, which account for the aforementioned processes. A true validation will only be possible once we experience the changed condition or could be based on paleo data. In summary, our approach does not address the coupled system directly but hinges on the climate predictions of the individual ESMs and hence is not able to implicitly correct respective biases of the CMIP5 ensemble.

The presented approach based on ML is not limited to projections of the future carbon cycle, which is used as an illustrative example. Indeed, since its only prerequisites are the availability of gridded climate model data and gridded observation-based data, the GBRT algorithm can be applied to any variable of interest if physically relevant diagnostics that influence the target variable are known. This opens a wide range of possibilities for constraining uncertainties in projected variable scenarios. In addition, maps of the local relative feature importance

as shown in Figure 6.7c can be used to reveal connections between different Earth system variables that are currently unknown. With the proposed data-driven approach, a possibly extended application could be to constrain projections of not just one but several variables simultaneously. Moreover, the concept of our method is not limited to GBRT but can be used with any other (ML) regression algorithm.

# 7. Conclusion

## 7.1. Overall Summary

The analysis of future climate projections from numerical climate model simulations is of paramount importance to assess future climate change under different forcing scenarios. Since this involves metrics of public interest like the allowable fossil fuel emissions to meet particular warming targets, for example, the 1.5 °C of the Paris Agreement (UNFCCC 2015), research in this field of science is not only relevant to climate scientists but also to policymakers and human society as a whole. In the light of the large spread of climate sensitivity in the most recent generation of climate models from CMIP6, which has been demonstrated in Chapter 4, a careful statistical evaluation and refinement of the output of multi-model climate projections is as relevant as ever. This thesis quantifies uncertainties in climate model projections, presents the evaluation of established methods to reduce these uncertainties in a new climate model ensemble, and finally introduces a novel alternative technique based on supervised ML.

To answer the key science questions posed in Section 1.2 and to ensure a consistent evaluation of the participating climate model ensembles in this thesis, the ESMValTool is used, an open-source community diagnostics and performance metrics tool for the routine evaluation of ESMs. All analyses shown in this thesis are performed with the ESMValTool. Apart from that, further substantial changes and additions to the code base of the tool have been implemented as part of this thesis (see Chapter 3), which led to co-authorship in the technical and scientific documentation of the ESMValTool (Eyring et al. 2020; Lauer et al. 2020; Righi et al. 2020; Weigel et al. 2020). Since the open-source tool is freely available, the code that has been implemented as part of this thesis is beneficial for the entire scientific community.

In the first study of this thesis, the climate sensitivity metrics ECS and TCR are evaluated for the latest generation of ESMs from CMIP6. This work, which is presented in Chapter 4 and already published in Bock et al. (2020) and Meehl et al. (2020), directly addresses Key Science Question 1 (“What is the range of climate sensitivity in the latest generation of ESMs from CMIP6 compared to previous multi-model ensembles, and do we understand the processes that determine this uncertainty range?”). For ECS, a CMIP6 model range of 1.8–5.6 K is found, which is higher on the upper and lower bound than any model range from previous CMIP generations. In comparison to CMIP5, the CMIP6 MMM of ECS is about 16 % higher (3.74 K in CMIP6 versus 3.23 K in CMIP5). Moreover, especially the upper bound of the CMIP6 model range is considerably larger than in the assessed range of 1.5–4.5 K given by the latest published AR of the IPCC from 2013 (Stocker et al. 2013). The assessed upper bound of 4.5 K is exceeded by a third of the CMIP6 models, with many models showing ECS values

above 5 K. For TCR, the model range of CMIP6 is 1.3–3.0 K, which also exceeds the CMIP5 range of 1.1–2.5 K and the assessed range from AR5 of 1.0–2.5 K. One possible reason for the increased climate sensitivity in some CMIP6 models is the addition of prognostic aerosol schemes that include aerosol-cloud interactions, which might result in overly large negative radiative forcing. In this case, a stronger response to GHG forcing is required to correctly reproduce the observed historical temperature trend, resulting in higher climate sensitivity in the corresponding models. A further reason for the high ECS in CMIP6 is a change in the microphysical representation of mixed-phase clouds in some CMIP6 models. This change was implemented to improve the representation of the fractioning of cloud ice and cloud liquid in those models, e.g., by allowing for (more) supercooled cloud liquid water. In addition, the CMIP6 MMM shows an improved simulation of the shortwave CRE when compared to satellite observations. The change in the representation of mixed-phase clouds in some CMIP6 models reduces the strong negative shortwave cloud feedback over the Southern Ocean that is present in predecessor versions of these models and that results from a cloud phase change from ice clouds in the present day to liquid clouds in the future. In the affected CMIP6 models, this cloud phase change due to warming is reduced since these models simulate less cloud ice over the present-day Southern Ocean than their predecessor versions with no supercooled cloud liquid.

To reduce this large range of ECS in the latest generation of climate models, already-published emergent constraints that have been derived from models from the previous CMIP generations CMIP3 and CMIP5 are evaluated for their skill in the CMIP6 ensemble. Emergent constraints use a physically based inter-model relationship between an observable quantity of the Earth system and a target variable to reduce uncertainties in the target variable with observations (Allen and Ingram 2002). In total, 11 emergent constraints on ECS are assessed, which are mostly related to cloud feedbacks since these constitute the most important source of uncertainty for ECS (Boucher et al. 2013; Flato et al. 2013). Since all of the evaluated emergent constraints have been derived from the CMIP3 or CMIP5 ensemble, out-of-sample tests on the emergent constraints can be performed by assessing whether they still hold for the CMIP6 models. In this study, which is shown in Chapter 5 and already published in Schlund et al. (2020b), a substantial reduction of skill for the majority of emergent constraints is found when applied to the CMIP6 ensemble in comparison to the CMIP5 ensemble. This drop in skill is expressed as a decrease of the coefficient of determination  $R^2$  of the emergent relationship and a decrease in the statistical significance using the null hypothesis that there is no correlation between the predictor and ECS. Moreover, the corresponding PDFs for the emergent constraints show higher 66 % ECS ranges (17–83 % confidence) for almost all emergent constraints, resulting in values of 1.32–2.70 K for CMIP6 (CMIP5: 1.16–1.75 K). Averaged over all emergent constraints, this is an increase in the 66 % ECS range of 51 %. Similarly, the best estimates for ECS show values of 2.97–3.88 K in CMIP5 and 3.48–4.32 K in CMIP6, resulting in an increase of about 12 % averaged over all emergent constraints. Thus, Key Science Question 2 (“Can uncertainties in climate sensitivity be reduced with observations using the emergent constraint approach?”) needs to be answered with a “not very well” for

the CMIP6 ensemble. The increased best estimates and spreads resulting from the emergent constraints in CMIP6 are likely related to the increased MMM and multi-model spread of ECS in CMIP6. A possible reason for the reduced skill of the emergent constraints when applied to the CMIP6 ensemble is the increased complexity of the CMIP6 models: A basic assumption for these single-process-oriented emergent constraints is that a single observable process dominates the uncertainty in ECS, which might not be valid anymore for the latest generation of ESMs from CMIP6 due to an increased number of processes that are included in these models.

To overcome these issues of single-process-oriented emergent constraints, Chapter 6 introduces an alternative approach based on ML. This work is already published in Schlund et al. (2020a). Since the new technique relies on a large number of data points to train the ML algorithm, the scalar climate sensitivity expressed as ECS or TCR is not an appropriate target variable. Therefore, this analysis does not focus on reducing uncertainties in climate sensitivity itself but rather on a selected process that contributes to it: GPP. GPP is the largest flux of the terrestrial carbon uptake and slows down global warming by removing CO<sub>2</sub> from the atmosphere. In the first step of the new two-step approach, a published emergent constraint by Wenzel et al. (2016a) is used to constrain the global mean GPP at the end of the 21<sup>st</sup> century in the CMIP5 RCP8.5 scenario to  $(171 \pm 12)$  GtC yr<sup>-1</sup>, compared to unconstrained CMIP5 model range of 156–247 GtC yr<sup>-1</sup>. This first step corrects the CMIP5 models' GPP response to CO<sub>2</sub>, which is primarily driven by CO<sub>2</sub> fertilization. In the second step, an ML-based climate model weighting approach is used to further constrain the gridded GPP based on present-day predictors that are relevant for the simulation of GPP in the ESMs. The ML approach is mathematically similar to the MDER approach (Karpechko et al. 2013; Seneftleben et al. 2020; Wenzel et al. 2016b), but additionally considers multi-dimensional (gridded) target variables and non-linear relationships between the predictors and the target variables. A relationship between process-oriented predictors and future projections of GPP is established and then utilized to project today's observed conditions into the future. The prediction phase of the new method can be interpreted as an implicit performance weighting. However, due to the complex structure of the used ML algorithm (GBRT), it is not possible to extract specific values for the individual weights. Two target variables are considered: the gridded monthly climatologies of absolute GPP (2091–2100) and the gridded fractional GPP change over the 21<sup>st</sup> century (2100 versus 2000). The latter quantity shows an increased GPP change in the high latitudes compared to regions closer to the equator. The results of both approaches are consistent with each other and with the global constraint of the first step. The new approach is validated by comparing it to other statistical models (the CMIP5 MMM and a linear LASSO model) in a leave-one-model-out CV setup. Compared to MMM (LASSO), a reduction of the resulting mean RMSEP of up to 48 % (3 %) is found when using the ML approach. Moreover, the evaluation of the global and local feature importance allows further insights into the ML model. For the first target variable (absolute GPP), historical GPP is by far the most important predictor, which can be explained with a correction of the historical bias in GPP by the new approach. For the second variable (fractional change in GPP), T and LAI are the

most important features. A corresponding feature map shows that T dominates in the tropics and LAI dominates in the northern high latitudes, which might be attributed to an additional greening trend in the high latitudes that enhances the GPP increase in this region compared to regions closer to the equator through an extension of the growing season. This study directly addresses Key Science Question 3 (“Can uncertainties in multi-dimensional (gridded) climate projections be reduced with ML techniques and observations?”), which can be confidently answered with “yes” based on the results found.

## 7.2. Outlook

Climate sensitivity is a policy-relevant and easy-to-use metric to assess the strength of climate change. The range of model results for this important climate metric, however, remains large and could not be narrowed down over the last decades. In the latest generation of climate models contributing to CMIP6, the range has even increased (Meehl et al. 2020). Chapter 4 presents two possible reasons for this: changes in the aerosol-cloud interaction and changes in the shortwave cloud phase change feedback over the Southern Ocean. While these are likely explanations, there might be additional relevant aspects that have not been analyzed in detail yet. This could include additional feedback processes that are not present in older CMIP model generations. Obvious candidates for these feedback processes are cloud-related feedbacks, which pose a major source of uncertainty in climate sensitivity in modern-day ESMs (Boucher et al. 2013). Identifying and quantifying such feedback mechanisms can further help to gain a better understanding of the CMIP6 models and potentially to reduce associated uncertainties in the entire multi-model ensemble. A potentially interesting process relevant in this context is the influence of the midlatitude jet position on clouds and CREs (Grise and Medeiros 2016).

A promising way forward to reduce uncertainties in multi-model climate projections is to apply the new flexible ML-based climate model weighting approach introduced in Chapter 6 to other target variables. While the scalar metrics ECS or TCR with one value per climate model do not offer enough training points for this technique, a possible alternative could be the gridded near-surface air temperature over the 21<sup>st</sup> century in the SSP scenarios. A relevant study in this context is given by Brunner et al. (2020), who use performance- and interdependence-based climate model weighting to constrain the near-surface air temperature over the 21<sup>st</sup> century based on observable process-based predictors of today’s climate and which could serve as a possible baseline for the new ML-based weighting approach. An exciting and valuable addition to the method could be causal inference (Nowack et al. 2020; Runge et al. 2019). In the current approach, the relationships between the different predictors and the target variable are based on statistical correlation alone. Moreover, the local feature importance maps do not reveal true causal relationships between the predictors and the target variable, but only show the relative weight that is given to a specific predictor at a specific location for the prediction of the target variable by the ML model. By integrating causal networks into the new approach, it might be possible to explore and utilize true causal connections instead.

The massive progress in artificial intelligence (AI) and ML in recent years in combination with ever-increasing computational power and resources provided by modern supercomputers has the potential for enormous improvements in climate modeling and analysis (Reichstein et al. 2019). Apart from the presented ML-based weighting approach, further examples of applications of AI and ML in climate science involve ML-based parameterizations that are learned from high-resolution climate models (Gentine et al. 2018; Rasp et al. 2018), ML-based analysis and prediction of forcing patterns (Barnes et al. 2019; Mansfield et al. 2020), or learning of entire ESMs from observational products (Geer 2021). Therefore, the foundation for innovative and groundbreaking research in climate science for the near future has been laid, which will help to further understand the Earth system and fight one of the greatest challenges for humankind today: climate change.



# Appendix

## A. Supplementary Materials for Chapter 5

Variable short name	Description
<i>cl</i>	Cloud area fraction
<i>clt</i>	Total cloud area fraction
<i>hur</i>	Relative humidity
<i>hus</i>	Specific humidity
<i>pr</i>	Precipitation
<i>rsdt</i>	Top of the atmosphere (TOA) incoming shortwave radiation
<i>rsut</i>	TOA outgoing shortwave radiation
<i>rsutcs</i>	Clear-sky TOA outgoing shortwave radiation
<i>ta</i>	Air temperature
<i>tas</i>	Surface air temperature
<i>tos</i>	Sea surface temperature
<i>ts</i>	Surface temperature
<i>va</i>	Northward wind speed
<i>wap</i>	Vertical velocity

Table A.1.: Overview of the variables used in Chapter 5. More details are given in Table 5.1, which lists the individual variables used for each emergent constraint.

Observational dataset	Corresponding variables	Reference
AIRS	<i>hur, hus</i>	Aumann et al. (2003)
AMSR-E SST	<i>tos</i>	AMSR-E (2011)
CERES-EBAF	<i>rsdt, rsut, rsutcs</i>	Loeb et al. (2018)
CloudSat/CALIPSO	<i>cl</i>	Mace et al. (2009)
ERA-Interim	<i>hur, ta, va, wap</i>	Dee et al. (2011)
GPCP	<i>pr</i>	Adler et al. (2003)
HadCRUT4	<i>tas</i>	Morice et al. (2012)
HadISST	<i>ts</i>	Rayner et al. (2003)
ISCCP D-2	<i>clt</i>	Rossow and Schiffer (1991)
MLS-Aura	<i>hur</i>	Beer (2006)

Table A.2.: References for all observational datasets used in Chapter 5. More details are given in Table 5.1, which lists the observational datasets used for each emergent constraint.

Model	Index	Main reference
ACCESS1-0	1	Dix et al. (2013)
ACCESS1-3	2	Dix et al. (2013)
BNU-ESM	3	Ji et al. (2014)
CCSM4	4	Gent et al. (2011) and Meehl et al. (2012)
CNRM-CM5	5	Voldoire et al. (2013)
CNRM-CM5-2	6	Voldoire et al. (2013)
CSIRO-Mk3-6-0	7	Rotstayn et al. (2012)
CanESM2	8	Arora et al. (2011)
FGOALS-g2	9	L. J. Li et al. (2013)
GFDL-CM3	10	Donner et al. (2011)
GFDL-ESM2G	11	Dunne et al. (2012)
GFDL-ESM2M	12	Dunne et al. (2012)
GISS-E2-H	13	Schmidt et al. (2006)
GISS-E2-R	14	Schmidt et al. (2006)
HadGEM2-ES	15	W. J. Collins et al. (2011)
IPSL-CM5A-LR	16	Dufresne et al. (2013)
IPSL-CM5A-MR	17	Dufresne et al. (2013)
IPSL-CM5B-LR	18	Dufresne et al. (2013)
MIROC-ESM	19	S. Watanabe et al. (2011)
MIROC5	20	M. Watanabe et al. (2010)
MPI-ESM-LR	21	Giorgetta et al. (2013)
MPI-ESM-MR	22	Giorgetta et al. (2013)
MPI-ESM-P	23	Giorgetta et al. (2013)
MRI-CGCM3	24	Yukimoto et al. (2012)
NorESM1-M	25	Bentsen et al. (2013) and Iversen et al. (2013)
bcc-csm1-1	26	Wu et al. (2014)
bcc-csm1-1-m	27	Wu et al. (2014)
inmcm4	28	Volodin et al. (2010)

Table A.3.: List of CMIP5 models used in Chapter 5 alongside the main reference and the index used in the corresponding figures. Adapted with permission from Schlund et al. (2020b).

Model	Index	Main reference
ACCESS-CM2	29	Bi et al. (2013)
ACCESS-ESM1-5	30	Law et al. (2017) and Ziehn et al. (2017)
AWI-CM-1-1-MR	31	Rackow et al. (2018) and Sidorenko et al. (2015)
BCC-CSM2-MR	32	Wu et al. (2019)
BCC-ESM1	33	Wu et al. (2019)
CAMS-CSM1-0	34	Rong et al. (2018)
CAS-ESM2-0	35	Wang et al. (2020)
CESM2	36	Danabasoglu et al. (2020)
CESM2-FV2	37	Danabasoglu et al. (2020)
CESM2-WACCM	38	Danabasoglu et al. (2020) and Gettelman et al. (2019b)
CESM2-WACCM-FV2	39	Danabasoglu et al. (2020) and Gettelman et al. (2019b)
CMCC-CM2-SR5	40	Cherchi et al. (2019)
CNRM-CM6-1	41	Voldoire et al. (2019)
CNRM-CM6-1-HR	42	Voldoire et al. (2019)
CNRM-ESM2-1	43	Séfériant et al. (2019)
CanESM5	44	Swart et al. (2019)
E3SM-1-0	45	Golaz et al. (2019)
EC-Earth3-Veg	46	Wyser et al. (2020)
FGOALS-f3-L	47	Guo et al. (2020) and He et al. (2019, 2020)
FGOALS-g3	48	L. Li et al. (2020)
GISS-E2-1-G	49	Rind et al. (2020)
GISS-E2-1-H	50	Rind et al. (2020)
HadGEM3-GC31-LL	51	Kuhlbrodt et al. (2018)
HadGEM3-GC31-MM	52	Williams et al. (2018)
INM-CM4-8	53	Volodin et al. (2017a,b)
INM-CM5-0	54	Volodin et al. (2017a,b)
IPSL-CM6A-LR	55	Boucher et al. (2020)
KACE-1-0-G	56	J. Lee et al. (2020)
MCM-UA-1-0	57	Delworth et al. (2002)
MIROC-ES2L	58	Hajima et al. (2020)
MIROC6	59	Tatebe et al. (2019)
MPI-ESM-1-2-HAM	60	Mauritsen et al. (2019)
MPI-ESM1-2-HR	61	Muller et al. (2018)
MPI-ESM1-2-LR	62	Mauritsen et al. (2019)
MRI-ESM2-0	63	Yukimoto et al. (2019)
NESM3	64	Cao et al. (2018)
NorCPM1	65	Counillon et al. (2016)
NorESM2-LM	66	Seland et al. (2020)
NorESM2-MM	67	Seland et al. (2020)
SAM0-UNICON	68	Park et al. (2019)
TaiESM1	69	W. L. Lee et al. (2020)
UKESM1-0-LL	70	Sellar et al. (2019)

Table A.4.: As in Table A.3 but for the CMIP6 models. Adapted with permission from Schlund et al. (2020b).

Model	Index	ECS [K]	BRI [% K <sup>-1</sup> ]	COX [K]	LIP [°]	SHD [1]	SHL [1]	SHS [1]
ACCESS1-0	1	3.83	-1.59	0.20	-33.70	0.45	0.84	0.39
ACCESS1-3	2	3.53	-1.59	0.15	-34.32	0.54	0.90	0.36
BNU-ESM	3	3.92	-1.39	0.14	-33.38	0.35	0.75	0.41
CCSM4	4	2.94	-0.09	0.18	-37.64	0.39	0.75	0.37
CNRM-CM5	5	3.25	0.29	0.15	-34.53	0.38	0.71	0.33
CNRM-CM5-2	6	3.44	0.56		-34.62	0.40	0.73	0.33
CSIRO-Mk3-6-0	7	4.08	-1.57	0.20	-35.12	0.61	0.97	0.36
CanESM2	8	3.69	-0.72	0.16	-34.28	0.30	0.61	0.31
FGOALS-g2	9	3.38	0.33	0.05	-30.75	0.29	0.79	0.51
GFDL-CM3	10	3.97	-0.26	0.32	-35.03	0.33	0.67	0.34
GFDL-ESM2G	11	2.39	-0.80	0.19	-36.47	0.30	0.76	0.45
GFDL-ESM2M	12	2.44	-0.34	0.15	-35.95	0.25	0.68	0.43
GISS-E2-H	13	2.31	1.18	0.10	-33.87	0.32	0.61	0.29
GISS-E2-R	14	2.11	0.89	0.11	-34.14	0.32	0.60	0.28
HadGEM2-ES	15	4.61	-2.45	0.26	-34.58	0.43	0.81	0.38
IPSL-CM5A-LR	16	4.13	-0.96	0.21	-32.13	0.41	0.86	0.45
IPSL-CM5A-MR	17	4.12	-1.18	0.15	-33.61	0.48	0.92	0.44
IPSL-CM5B-LR	18	2.60	-0.50	0.16	-32.02	0.25	0.67	0.41
MIROC-ESM	19	4.67	-0.99	0.22	-31.87	0.33	0.89	0.56
MIROC5	20	2.72	0.27	0.22	-35.51	0.36	0.78	0.42
MPI-ESM-LR	21	3.63	-0.35	0.15	-34.67	0.41	0.86	0.45
MPI-ESM-MR	22	3.46	-0.49	0.16	-34.33	0.42	0.87	0.45
MPI-ESM-P	23	3.45	-0.68		-34.36	0.41	0.87	0.46
MRI-CGCM3	24	2.60	-1.08	0.09	-35.01	0.37	0.78	0.41
NorESM1-M	25	2.80	-0.74	0.13	-37.44	0.44	0.82	0.38
bcc-csm1-1	26	2.83	-0.11	0.18	-34.25	0.37	0.78	0.41
bcc-csm1-1-m	27	2.86	-0.46	0.13	-36.36	0.34	0.74	0.40
inmcm4	28	2.08	-0.18	0.07	-36.43	0.19	0.52	0.33

Table A.5.: All CMIP5 models used in Chapter 5 including their effective climate sensitivity (ECS) and  $x$ -axis values for the emergent constraints BRI, COX, LIP, SHD, SHL, and SHS. More details on the emergent constraints are given in Table 5.1. The specified index corresponds to the index used in the associated figures. Adapted with permission from Schlund et al. (2020b).

Model	Index	ECS [K]	SU [1]	TIH [%]	TII [mm day <sup>-1</sup> ]	VOL [%]	ZHA [% K <sup>-1</sup> ]
ACCESS1-0	1	3.83	0.94	14.14	0.54	-26.16	-2.32
ACCESS1-3	2	3.53	0.93	21.05	0.64	-20.21	-2.11
BNU-ESM	3	3.92	1.11	22.56	1.47	-11.91	
CCSM4	4	2.94		24.32	1.17	-9.88	0.00
CNRM-CM5	5	3.25	1.04	14.54	1.50	-15.43	-0.40
CNRM-CM5-2	6	3.44		15.83	1.44	-15.08	
CSIRO-Mk3-6-0	7	4.08	1.00	11.19	1.01	-11.53	-0.79
CanESM2	8	3.69	1.00	20.00	0.69	-21.21	-1.18
FGOALS-g2	9	3.38	1.26	24.61	0.93	-12.06	-0.54
GFDL-CM3	10	3.97	1.01	18.05	1.43	-14.67	-2.33
GFDL-ESM2G	11	2.39	0.93	18.60	0.74	-17.63	-2.43
GFDL-ESM2M	12	2.44	0.94	17.43	1.36	-17.75	-2.64
GISS-E2-H	13	2.31	0.71	21.70	2.58	3.57	-0.27
GISS-E2-R	14	2.11	0.71	12.88	2.13	2.52	-0.21
HadGEM2-ES	15	4.61	0.95	11.54	0.87	-24.69	-1.54
IPSL-CM5A-LR	16	4.13	0.95	11.58	0.53	-34.04	-1.08
IPSL-CM5A-MR	17	4.12	0.98	5.73	0.36	-32.90	-0.95
IPSL-CM5B-LR	18	2.60	0.89	21.25	0.92	-27.02	-1.90
MIROC-ESM	19	4.67	0.95	0.94	-0.54	-27.09	0.07
MIROC5	20	2.72	1.01	4.90	0.40	-10.99	0.41
MPI-ESM-LR	21	3.63	1.07	3.16	0.24	-17.96	0.07
MPI-ESM-MR	22	3.46	1.04	6.10	0.30	-18.92	-0.04
MPI-ESM-P	23	3.45		-0.23	0.13	-17.97	0.06
MRI-CGCM3	24	2.60	1.04	34.02	2.04	-11.08	0.40
NorESM1-M	25	2.80	1.07	26.23	0.65	-7.13	0.51
bcc-csm1-1	26	2.83	1.15	30.84	1.34	-8.77	0.41
bcc-csm1-1-m	27	2.86	1.19	43.64	2.69	-7.56	0.66
inmcm4	28	2.08	0.84	28.16	1.89	-14.52	0.75

Table A.6.: As in Table A.5 but for the emergent constraints SU, TIH, TII, VOL, and ZHA. Adapted with permission from Schlund et al. (2020b).



Model	Index	ECS [K]	SU [1]	TIH [%]	TII [mm day <sup>-1</sup> ]	VOL [%]	ZHA [% K <sup>-1</sup> ]
ACCESS-CM2	29	4.72	0.99	21.63	1.62	-24.48	
ACCESS-ESM1-5	30	3.87	0.94	15.62	1.00	-20.15	
AWI-CM-1-1-MR	31	3.16	1.05	2.15	-0.12	-19.44	
BCC-CSM2-MR	32	3.04	1.10	32.92	1.50	-3.93	-0.10
BCC-ESM1	33	3.26	1.18	38.93	0.94	2.69	-0.22
CAMS-CSM1-0	34	2.29	1.09	27.52	1.40	-13.84	-0.36
CAS-ESM2-0	35	3.51	0.95	6.98	1.58	-14.42	-0.06
CESM2	36	5.16	0.93	-7.25	0.45	-24.99	-2.39
CESM2-FV2	37	5.14	0.93	12.09	0.30	-25.68	-2.44
CESM2-WACCM	38	4.75	0.93	-5.37	-0.13	-23.81	-2.14
CESM2-WACCM-FV2	39	4.79		26.98	0.96	-25.34	
CMCC-CM2-SR5	40	3.52	1.03	23.28	1.12	-22.46	-1.00
CNRM-CM6-1	41	4.83	1.03	16.59	0.92	-9.16	
CNRM-CM6-1-HR	42	4.28	1.04	13.31	1.04	-6.63	
CNRM-ESM2-1	43	4.76	1.04	16.17	0.95	-8.60	
CanESM5	44	5.62	0.97	15.39	0.04	-20.02	
E3SM-1-0	45	5.32	0.96	16.73	1.26	-25.38	-3.11
EC-Earth3-Veg	46	4.31			0.82		
FGOALS-f3-L	47	3.00	1.06	11.75	1.00	-0.19	
FGOALS-g3	48	2.88	1.18	19.06	0.32	-15.13	1.06
GISS-E2-1-G	49	2.72	0.98	28.46	2.39	-7.75	-0.07
GISS-E2-1-H	50	3.11	1.00	24.55	1.84	-6.82	0.00
HadGEM3-GC31-LL	51	5.55	1.02	9.56	0.99	-24.87	-2.66
HadGEM3-GC31-MM	52	5.42	1.01	4.73	1.19	-22.48	
INM-CM4-8	53	1.83	0.99	12.22	0.62	-16.71	
INM-CM5-0	54	1.92	0.99	11.22	0.62	-14.38	
IPSL-CM6A-LR	55	4.56	0.76	11.90	0.41	-34.90	
KACE-1-0-G	56	4.48	1.03	22.16	1.66	-24.51	
MCM-UA-1-0	57	3.65	0.91	14.50	-0.56		
MIROC-ES2L	58	2.68	0.91	18.53	0.68	-13.03	-0.14
MIROC6	59	2.61	0.99	5.79	0.29	-13.43	-1.08
MPI-ESM-1-2-HAM	60	2.96	0.96	14.47	0.70	-23.06	-0.73
MPI-ESM1-2-HR	61	2.98	1.04	4.33	0.53	-20.21	-0.16
MPI-ESM1-2-LR	62	3.00	1.02	5.92	0.40	-19.40	-0.42
MRI-ESM2-0	63	3.15	0.95	3.33	0.75	-22.83	-1.35
NESM3	64	4.72	1.05	28.78	0.55	-18.81	-0.62
NorCPM1	65	3.05		30.50	0.54	-8.09	
NorESM2-LM	66	2.54	1.01	34.40	1.30	-16.22	-1.25
NorESM2-MM	67	2.50	1.00	11.07	0.27	-25.33	-1.03
SAM0-UNICON	68	3.72	1.01	11.99	1.13	-25.25	-1.85
TaiESM1	69	4.31	1.02	11.20	1.17	-16.53	-1.89
UKESM1-0-LL	70	5.34	1.03	2.22	0.86	-26.83	-2.72

Table A.8.: As in Table A.6 but for the CMIP6 models. Adapted with permission from Schlund et al. (2020b).

## B. Supplementary Materials for Chapter 6

### B.1. Data Preprocessing

The raw monthly mean output of every participating climate model and observation-driven dataset is regridded to a  $2^\circ \times 2^\circ$  grid and masked with a common mask to remove all oceans and Antarctica. For Step 2a (target variable: absolute gross primary production (GPP) at the end of the 21<sup>st</sup> century), monthly climatologies are calculated for every dataset by averaging over all available years for every month. For Step 2b (target variable: fractional GPP change over the 21<sup>st</sup> century), temporal means are calculated for every dataset by averaging over the full time dimension. In addition, values greater than 300 % in the target variable in Step 2b (fractional GPP change over the 21<sup>st</sup> century) are masked to avoid numerical inconsistencies caused by the division of small numbers in the derivation of the target variable. In the next step, the multi-dimensional data is flattened and all the training data from the different climate models is stacked into a single large training array. Finally, to account for the varying magnitudes of the different predictors, all of them are linearly scaled by their respective means and standard deviations so that they have a mean of zero and unit variance. In total, 237,852 (16,503) training data points, 79,284 (5501) hold-out test data points, and 46,344 (3727) points for the prediction are used in the machine learning model for Step 2a (Step 2b).

### B.2. Gradient Boosted Regression Tree (GBRT) Algorithm

The basic elements of the GBRT algorithm are decision trees. These models create decision rules based on binary splits to predict a target variable  $y$  from a set of predictors  $\mathbf{x} = (x^{(1)}, x^{(2)}, \dots, x^{(K)})$  ( $K$  is the number of predictors). These predictors do not need to be of the same type: GBRT allows the simultaneous input of numerical and categorical variables, which is a great advantage for our use case. There is no need to encode the categorical predictors in any way. Due to their simple nature, machine learning (ML) models based on decision trees are easy to interpret and explain but cannot be used to create satisfying predictions for complex datasets. This issue can be overcome by a technique called *boosting*. Boosting improves the performance of *weak learners* (in our case decision trees) by combining a large number of them (Freund and Schapire 1996). The regression function used to predict  $\hat{y} = F(\mathbf{x})$  can be written as a linear combination of simple decision trees  $h(\mathbf{x})$

$$F(\mathbf{x}) = \sum_{j=0}^H \beta_j h(\mathbf{x}, \boldsymbol{\alpha}_j), \quad (\text{B.1})$$

where  $H$  is the total number of decision trees,  $\beta_j$  the expansion coefficient for tree  $j$ , and  $\boldsymbol{\alpha}_j$  the hyperparameters for tree  $j$ . Using all  $N$  training data points  $\{(x_i, y_i) \mid i \in I_N\}$  with index set  $I_N = \{1, 2, \dots, N\}$  (*classic gradient boosting*), the expansion coefficients and tree parameters are jointly fitted by minimizing a loss function  $L(y, F(\mathbf{x}))$  in a forward iteration:

$$(\beta_j, \boldsymbol{\alpha}_j) = \arg \min_{\beta, \boldsymbol{\alpha}} \sum_{i=1}^N L(y_i, F_{j-1}(\mathbf{x}_i) + \beta h(\mathbf{x}_i, \boldsymbol{\alpha})). \quad (\text{B.2})$$

In practice, this iteration step only uses a randomly selected subsample of the training data (drawn without replacement), i.e., the sum does not cover all  $N$  training points. Starting with an initial guess  $F_0(\mathbf{x})$ , the model is recursively built by

$$F_j(\mathbf{x}) = F_{j-1}(\mathbf{x}) + \beta_j h(\mathbf{x}, \alpha_j). \quad (\text{B.3})$$

The minimization procedure of the loss function (mean square error with additional sample weights determined by the grid-cell areas) is called *stochastic gradient boosting* and is explained in detail by Friedman (2001, 2002). Fitting the GBRT model involves building the decision trees by splitting the data at points with maximum information gain. Boosting those simple trees greatly improves the overall predictive power of the ML algorithm: Poorly modeled training points in the early stages of the algorithm will gradually improve throughout the training process.

A crucial criterion for the successful application of any GBRT algorithm is the choice of several hyperparameters. The three main control parameters of the learning procedure are the total number of decision trees  $H$ , the complexity of the individual trees (for example, measured by the maximum tree depth), and the learning rate  $\nu \ll 1$ . The latter parameter is used for regularization and dramatically reduces the risk of overfitting by scaling down the contribution of each added weak learner (De'ath 2007; Elith et al. 2008; Friedman 2001). A common way to optimize the algorithm is  $n$ -fold cross-validation (CV) (Bishop 2006): The data is randomly divided into a training and a validation dataset and the GBRT model is fitted on the training data only. After that, the performance of this model can be evaluated on the validation dataset by a suitable metric (e.g., the mean square error). This process is repeated  $n$  times so that every input point is part of the validation set at least once. The optimal hyperparameters are the hyperparameters that show optimal performance on the validation datasets (Elith et al. 2008).

### B.3. Evaluation of Prediction Uncertainty

We estimate the standard prediction error (SPE) of the GBRT model itself as the RMSE of the predicted target variable  $\hat{y}$  and the corresponding true values  $y$  of a hold-out test dataset, the so-called *root mean square error of prediction (RMSEP)* (Bishop 2006), which is assumed to be constant for all prediction input points:

$$\sigma_{\text{GBRT}} = \text{RMSE}(\hat{y}, y). \quad (\text{B.4})$$

For this, we randomly select 25 % of the input data prior to training, so that this part of the data neither enters the training of the GBRT model nor the hyperparameter optimization process. Moreover, the test dataset allows an assessment of the prediction residuals, which is useful to detect overfitting (see Figure B.2).

A second source of uncertainty is the error in the rescaling of the target variable in Step 1 of our approach. Analogous to Equation (6.2), we estimate this error as

$$\sigma_{i,\text{RESC}} = \bar{y}_i \cdot \frac{\sigma_{f'}}{\bar{f}} \quad (\text{B.5})$$

for each prediction input point  $i$  ( $i$  runs over all grid cells and months).  $\bar{y}_i$  is the CMIP5 multi-model mean (MMM) of the target variable at prediction input point  $i$  (Step 2a: absolute GPP at the end of the 21<sup>st</sup> century; Step 2b: fractional GPP change over the 21<sup>st</sup> century),  $\sigma_f$  the standard error in the global mean fractional GPP change given by the emergent constraint from Step 1, and  $\bar{f}$  the CMIP5 MMM global mean fractional GPP change over the 21<sup>st</sup> century.

The final source of uncertainty is the error of the prediction input data  $\sigma_{ik}$  ( $i$  corresponds to the prediction input point and  $k$  to the predictor). These are only available for the FLUXNET-MTE product (Jung et al. 2011). To account for this, we use the local interpretable model-agnostic explanations (LIME) technique (Ribeiro et al. 2016; see Section 6.1.2) to build a local linear model for every sample point, which yields the linear coefficients  $b_{ik}$ . Using error propagation for all predictors and assuming independence of all individual errors, the SPE due to observational uncertainty  $\sigma_{i,\text{OBS}}$  then be calculated as

$$\sigma_{i,\text{OBS}} = \sum_{k=1}^K b_{ik}^2 \sigma_{ik}^2. \quad (\text{B.6})$$

The total SPE at a prediction input point  $i$  is the sum of the squared errors presented above (assuming all of them are independent):

$$\sigma_i^2 = \sigma_{\text{GBRT}}^2 + \sigma_{i,\text{RESC}}^2 + \sigma_{i,\text{OBS}}^2. \quad (\text{B.7})$$

The specified error ranges for the MMM approaches are calculated similarly: The constant SPE per grid cell is estimated by the mean RMSEP given in the leave-one-model-out CV experiment (see Section 2.4.2). For the plain MMM, this is the only source of uncertainty. For the rescaled MMM, the total error can be calculated similarly to Equation (B.7) without the last term (observational uncertainty).

#### B.4. Evaluation of Residuals

A convenient way to gain information about statistical models is to analyze the residuals  $\varepsilon_i$ , which are defined as the difference between the true value of the target variable  $y_i$  and the predicted value  $\hat{y}_i$  at a sample point  $i$  with known ground truth:

$$\varepsilon_i = y_i - \hat{y}_i. \quad (\text{B.8})$$

A common way to visualize the residuals is to plot their probability density functions (PDFs) (see Figure B.2). The two panels (for steps 2a and 2b) show that the ML model is not overfitting in both cases: The distributions of the training data and the independent hold-out data are very similar. Moreover, the PDFs do not show significant biases, as the residuals are approximately unbiased (zero mean) for the training and the test dataset. This justifies the use of the RMSEP to estimate the SPE since for unbiased residuals the RMSEP is equal to the standard deviation of the residuals.

Climate model	Land model	Main reference
CESM1-BGC	CLM4	Gent et al. (2011)
CanESM2	CLASS2.7 + CTEM1	Arora et al. (2011)
GFDL-ESM2M	LM3	Dunne et al. (2012)
HadGEM2-ES	JULES + TRIFFID	W. J. Collins et al. (2011)
MIROC-ESM	MATSIRO + SEIB-DGVM	S. Watanabe et al. (2011)
MPI-ESM-LR	JSBACH + BETHY	Giorgetta et al. (2013)
NorESM1-ME	CLM4	Iversen et al. (2013)

Table B.1.: List of the seven CMIP5 models used in Chapter 6 alongside the main references. More details are given by Anav et al. (2013). We chose all CMIP5 models which provide all necessary variables ( $co2$ ,  $gpp$ ,  $lai$ ,  $pr$ ,  $rsds$ , and  $tas$ ) for all used experiments ( $esmHistorical$ ,  $esmrcp85$ , and  $esmFixClim1$ ). For all models, we only use the first ensemble member available. Adapted with permission from Schlund et al. (2020a).



Figure B.1.: (a) Monthly mean atmospheric CO<sub>2</sub> concentration at Cape Kumukahi, Hawaii (KUM) from 1979 to 2019. The thin colored lines show the individual CMIP5 models (emission-driven historical simulations for the years 1979–2005 and emission-driven RCP8.5 simulations for the years 2006–2019; the latter is not available for HadGEM2-ES). The thick black line shows the observations. For the CMIP5 models, the grid cell closest to KUM is considered. The curves show an increase of the atmospheric CO<sub>2</sub> concentration superimposed by a pronounced seasonal cycle (see Section 2.3.2). (b) Annual amplitude of the seasonal cycle of CO<sub>2</sub> (defined as the difference between the maximum and the minimum monthly mean atmospheric CO<sub>2</sub> concentration for each year) against the annual mean atmospheric CO<sub>2</sub> concentration at KUM. Colored points show the CMIP5 models (similar time ranges as in (a)), and thick black points show the observations. The lines show the corresponding linear regression fits for each dataset. The slopes of these linear fits define the sensitivity of the seasonal CO<sub>2</sub> cycle amplitude to atmospheric CO<sub>2</sub> concentrations, which is used as predictor for the emergent constraint step of our approach (Step 1). Adapted with permission from Schlund et al. (2020a).



Figure B.2.: Distribution of the residuals of the gradient boosted regression tree (GBRT) model for the two different target variables used in Step 2a (absolute gross primary production (GPP) at the end of the 21<sup>st</sup> century) and Step 2b (fractional GPP change over the 21<sup>st</sup> century). The distributions are derived by kernel density estimation using training (blue) and test (green) data. The plots show approximately unbiased distributions for the training and the test datasets, which are very similar to each other. This indicates that the machine learning model does not overfit the data. Adapted with permission from Schlund et al. (2020a).

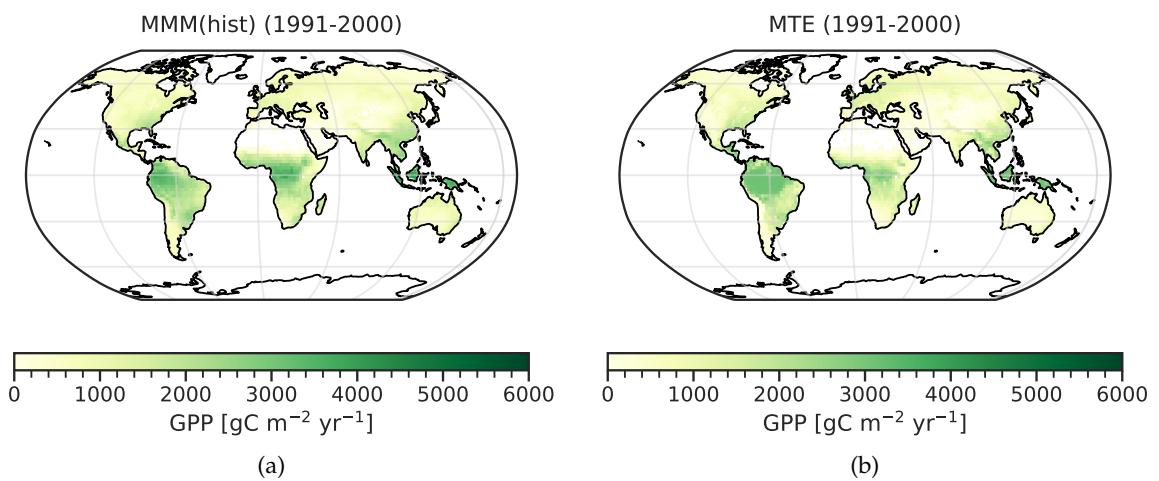


Figure B.3.: Geographical distributions of the historical gross primary production (GPP) averaged between 1991 and 2000. (a) CMIP5 multi-model mean (MMM). (b) FLUXNET-MTE product (Jung et al. 2011). Adapted with permission from Schlund et al. (2020a).



Figure B.4.: Geographical distributions of the absolute gross primary production (GPP) at the end of the 21<sup>st</sup> century in the RCP8.5 scenario (Step 2a) for different statistical models. (a) CMIP5 multi-model mean (MMM). (b) Rescaled CMIP5 MMM using Equation (6.2). (c) Least absolute shrinkage and selection operator (LASSO) model using only the historical GPP as a single predictor. (d) Gradient boosted regression tree (GBRT) model using only the historical GPP as a single predictor. (e) LASSO model using all predictors. (f) GBRT model using all predictors. Adapted with permission from Schlund et al. (2020a).

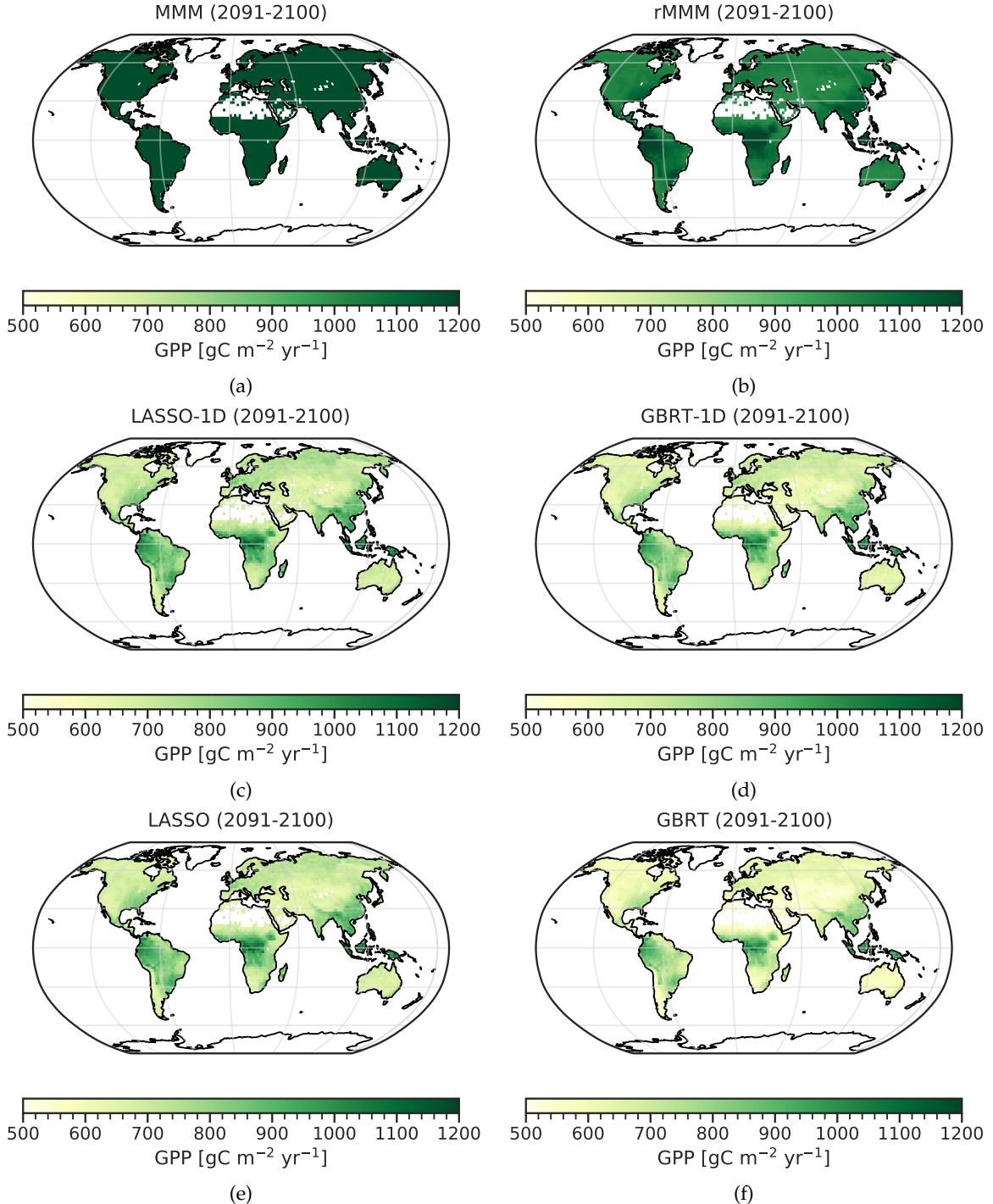


Figure B.5.: Geographical distributions of the standard prediction errors (SPEs) of the absolute gross primary production (GPP) at the end of the 21<sup>st</sup> century in the RCP8.5 scenario (Step 2a) for different statistical models. Details on the calculation of the SPE are given in Appendix B.3. (a) CMIP5 multi-model mean (MMM). (b) Rescaled CMIP5 MMM using Equation (6.2). (c) Least absolute shrinkage and selection operator (LASSO) model using only the historical GPP as a single predictor. (d) Gradient boosted regression tree (GBRT) model using only the historical GPP as a single predictor. (e) LASSO model using all predictors. (f) GBRT model using all predictors. The SPE is minimal for the GBRT model using all predictors. Adapted with permission from Schlund et al. (2020a).



Figure B.6.: Geographical distributions of the standard prediction errors (SPEs) of the fractional change in gross primary production (GPP) over the 21<sup>st</sup> century in the RCP8.5 scenario (Step 2b) for different statistical models. Details on the calculation of the SPE are given in Appendix B.3. (a) CMIP5 multi-model mean (MMM). (b) Rescaled CMIP5 MMM using Equation (6.2). (c) Least absolute shrinkage and selection operator (LASSO) model. (d) Gradient boosted regression tree (GBRT) model. The SPE is minimal for the GBRT model. Adapted with permission from Schlund et al. (2020a).



# List of Abbreviations

<b>1%BGC</b>	biogeochemically coupled simulation where the atmospheric CO <sub>2</sub> concentration increases by 1 % per year . . . . .	77
<b>AerChemMIP</b>	Aerosol Chemistry Model Intercomparison Project . . . . .	8
<b>AI</b>	artificial intelligence . . . . .	105
<b>AOGCM</b>	atmosphere-ocean general circulation model . . . . .	6
<b>AR</b>	Assessment Report . . . . .	49
<b>C4MIP</b>	Coupled Climate-Carbon Cycle Model Intercomparison Project . . . . .	8
<b>CESM</b>	Community Earth System Model . . . . .	17
<b>CFMIP</b>	Cloud Feedback Model Intercomparison Project . . . . .	62
<b>CH<sub>4</sub></b>	methane . . . . .	20
<b>CMIP</b>	Coupled Model Intercomparison Project . . . . .	7
<b>CMOR</b>	Climate Model Output Rewriter . . . . .	36
<b>CO</b>	carbon monoxide . . . . .	20
<b>CO<sub>2</sub></b>	carbon dioxide . . . . .	7
<b>CRE</b>	cloud radiative effect . . . . .	17
<b>CV</b>	cross-validation . . . . .	31
<b>DECK</b>	Diagnostic, Evaluation, and Characterization of Klima . . . . .	8
<b>DJF</b>	December-January-February . . . . .	62
<b>ECS</b>	effective climate sensitivity . . . . .	16
<b>ENSO</b>	El Niño–Southern Oscillation . . . . .	9
<b>ESGF</b>	Earth System Grid Federation . . . . .	37
<b>ESM</b>	Earth system model . . . . .	6
<b>ESMValCore</b>	core functionalities of the ESMValTool . . . . .	36
<b>ESMValTool</b>	Earth System Model Evaluation Tool . . . . .	35
<b>GBRT</b>	gradient boosted regression tree . . . . .	76
<b>GHG</b>	greenhouse gas . . . . .	7
<b>GPP</b>	gross primary production . . . . .	20

<b>GSAT</b>	global mean near-surface air temperature . . . . .	11
<b>GtC</b>	gigatonnes of carbon . . . . .	20
<b>HadGEM</b>	Hadley Centre Global Environmental Model . . . . .	51
<b>IAM</b>	integrated assessment model . . . . .	9
<b>IPCC</b>	Intergovernmental Panel on Climate Change . . . . .	37
<b>ITCZ</b>	Intertropical Convergence Zone . . . . .	66
<b>JJA</b>	June-July-August . . . . .	39
<b>KUM</b>	Cape Kumukahi, Hawaii . . . . .	75
<b>LAI</b>	leaf area index . . . . .	78
<b>LASSO</b>	least absolute shrinkage and selection operator . . . . .	84
<b>LIME</b>	local interpretable model-agnostic explanations . . . . .	80
<b>LongRunMIP</b>	Long Run Model Intercomparison Project . . . . .	17
<b>LTMI</b>	lower-tropospheric mixing index . . . . .	64
<b>MBLC</b>	marine boundary layer cloud . . . . .	67
<b>MDER</b>	multiple diagnostic ensemble regression . . . . .	31
<b>MIP</b>	Model Intercomparison Project . . . . .	8
<b>ML</b>	machine learning . . . . .	41
<b>MMM</b>	multi-model mean . . . . .	25
<b>N<sub>2</sub>O</b>	nitrous oxide . . . . .	1
<b>NAO</b>	North Atlantic Oscillation . . . . .	9
<b>NBP</b>	net biome production . . . . .	24
<b>NCAR</b>	National Center for Atmospheric Research . . . . .	17
<b>NEP</b>	net ecosystem production . . . . .	24
<b>NetCDF</b>	Network Common Data Form . . . . .	36
<b>NPP</b>	net primary production . . . . .	24
<b>PDF</b>	probability density function . . . . .	27
<b>ppm</b>	parts per million . . . . .	20
<b>PR</b>	precipitation . . . . .	82
<b>RCP</b>	Representative Concentration Pathway . . . . .	9
<b>RMSE</b>	root mean square error . . . . .	30
<b>RMSEP</b>	root mean square error of prediction . . . . .	31
<b>RSDS</b>	incoming solar radiation at the surface . . . . .	82

---

<b>RuBisCO</b>	ribulose-1,5-bisphosphate carboxylase/oxygenase . . . . .	24
<b>ScenarioMIP</b>	Scenario Model Intercomparison Project . . . . .	8
<b>SPE</b>	standard prediction error . . . . .	27
<b>SSP</b>	Shared Socioeconomic Pathway . . . . .	8
<b>SST</b>	sea surface temperature . . . . .	58
<b>T</b>	near-surface air temperature . . . . .	82
<b>TCR</b>	transient climate response . . . . .	18
<b>TLC</b>	tropical low-level cloud . . . . .	60
<b>TOA</b>	top of the atmosphere . . . . .	14
<b>WCRP</b>	World Climate Research Programme . . . . .	7
<b>WGCM</b>	Working Group on Coupled Modelling . . . . .	7



# List of Figures

2.1. Historical evolution of coupled climate models over the last 45 years. . . . .	6
2.2. Schematic of the experiment design of Phase 6 of the Coupled Model Intercomparison Project (CMIP6). . . . .	8
2.3. Schematic illustration of the importance of different sources of uncertainties in climate model projections and their evolution in time. . . . .	10
2.4. Climate feedbacks and corresponding time scales. . . . .	12
2.5. Gregory regression for the CMIP5 multi-model mean illustrating the definition of the effective climate sensitivity (ECS). . . . .	16
2.6. Gregory regression for different time periods. . . . .	17
2.7. Illustration of the definition of the transient climate response (TCR). . . . .	19
2.8. Simplified schematic of the global carbon cycle. . . . .	21
2.9. The Keeling Curve. . . . .	22
2.10. Schematic representation of the overall perturbation of the global carbon cycle caused by anthropogenic activities. . . . .	23
2.11. Schematic illustration of the processes of the terrestrial carbon uptake. . . . .	25
2.12. Schematic illustration of the emergent constraint approach. . . . .	27
2.13. Schematic illustration of the multiple diagnostic ensemble regression (MDER) approach. . . . .	33
3.1. Schematic representation of the structure of the Earth System Model Evaluation Tool (ESMValTool). . . . .	36
3.2. Illustration of the atmospheric CO <sub>2</sub> concentration at the surface ( <i>co2</i> ) for CMIP6 models. . . . .	39
3.3. Emergent constraint on the effective climate sensitivity (ECS) from global temperature variability. . . . .	42
3.4. Relationships between climate metrics within the CMIP5 climate model ensemble. . . . .	43
4.1. Gregory regression for the CMIP6 multi-model mean. . . . .	46
4.2. Historical values of the effective climate sensitivity (ECS) and the transient climate response (TCR) for different Assessment Reports and climate model ensembles. . . . .	50
4.3. Schematic illustration of the strong negative shortwave cloud phase change feedback over the Southern Ocean. . . . .	52

4.4. Geographical distributions of the cloud radiative effect feedback parameters for the CMIP5 and CMIP6 multi-model means. . . . .	54
5.1. Emergent constraints BRI, COX, and LIP applied to the CMIP5 and CMIP6 ensemble. . . . .	61
5.2. As in Figure 5.1 but for the emergent constraints SHD, SHL, and SHS. . . . .	63
5.3. As in Figure 5.1 but for the emergent constraints SU, TIH, and TII. . . . .	65
5.4. As in Figure 5.1 but for the emergent constraints VOL and ZHA. . . . .	67
5.5. Emergent relationship ZHA for different subsets of CMIP5 models. . . . .	68
5.6. Overview of the constrained effective climate sensitivity (ECS) ranges given by the 11 emergent constraints using the CMIP5 and CMIP6 ensemble. . . . .	70
6.1. Schematic illustration of our two-step approach. . . . .	77
6.2. Schematic illustration of our machine learning approach to constrain projected absolute gross primary production (GPP) at the end of the 21 <sup>st</sup> century in Step 2a. . . . .	81
6.3. Emergent relationship between the global mean fractional change in gross primary production (GPP) and the sensitivity of the CO <sub>2</sub> seasonal cycle amplitude to rising atmospheric CO <sub>2</sub> concentrations observed at Cape Kumukahi, Hawaii. . . . .	85
6.4. Prediction error in a leave-one-model-out cross-validation setup and feature importance of our machine learning approach to constrain projected absolute gross primary production (GPP) at the end of the 21 <sup>st</sup> century in Step 2a. . . . .	87
6.5. Illustration of our machine learning approach to constrain projected absolute gross primary production (GPP) at the end of the 21 <sup>st</sup> century with observations in Step 2a. . . . .	89
6.6. Further illustrations of our machine learning approach to constrain projected absolute gross primary production (GPP) at the end of the 21 <sup>st</sup> century with observations in Step 2a and comparison to other statistical models. . . . .	90
6.7. Prediction error in a leave-one-model-out cross-validation setup and feature importance of our machine learning approach to constrain the fractional change in gross primary production (GPP) over the 21 <sup>st</sup> century in Step 2b. . . . .	93
6.8. Illustration of our machine learning approach to constrain the fractional change in gross primary production (GPP) over the 21 <sup>st</sup> century with observations in Step 2b. . . . .	94
6.9. Further illustrations of our machine learning approach to constrain the fractional change in gross primary production (GPP) over the 21 <sup>st</sup> century with observations in Step 2b. . . . .	95
6.10. Comparison of our two independent approaches from Step 2a and Step 2b. . . . .	97
B.1. Monthly mean atmospheric CO <sub>2</sub> concentration at Cape Kumukahi, Hawaii from 1979 to 2019 and the corresponding annual amplitudes of the seasonal cycle of CO <sub>2</sub> . . . . .	117

B.2. Distribution of the residuals of the gradient boosted regression tree (GBRT) model for the two different target variables used in Step 2a and Step 2b. . . . .	118
B.3. Geographical distributions of the historical gross primary production (GPP) averaged between 1991 and 2000. . . . .	118
B.4. Geographical distributions of the absolute gross primary production (GPP) at the end of the 21 <sup>st</sup> century in the RCP8.5 scenario (Step 2a) for different statistical models. . . . .	119
B.5. Geographical distributions of the standard prediction errors of the absolute gross primary production (GPP) at the end of the 21 <sup>st</sup> century in the RCP8.5 scenario (Step 2a) for different statistical models. . . . .	120
B.6. Geographical distributions of the standard prediction errors of the fractional change in gross primary production (GPP) over the 21 <sup>st</sup> century in the RCP8.5 scenario (Step 2b) for different statistical models. . . . .	121



# List of Tables

3.1. Summary of new preprocessor functions and variable derivation scripts contributed to the core functionalities of the ESMValTool (ESMValCore) by the author of this thesis. . . . .	38
3.2. Summary of new recipes and CMORization scripts for observational data contributed to the Earth System Model Evaluation Tool (ESMValTool) by the author of this thesis. . . . .	41
4.1. Effective climate sensitivity (ECS) and transient climate response (TCR) calculated for the CMIP5 models. . . . .	47
4.2. As in Table 4.1 but for the CMIP6 models. . . . .	48
5.1. Overview of the 11 emergent constraints on the effective climate sensitivity (ECS) evaluated in Chapter 5. . . . .	58
5.2. Overview of the constrained effective climate sensitivity (ECS) ranges and <i>p</i> -values for all 11 analyzed emergent constraints. . . . .	69
6.1. Process-oriented diagnostics used in the gradient boosted regression tree (GBRT) model to predict the target variables. . . . .	82
A.1. Overview of the variables used in Chapter 5. . . . .	107
A.2. References for all observational datasets used in Chapter 5. . . . .	107
A.3. List of CMIP5 models used in Chapter 5. . . . .	108
A.4. As in Table A.3 but for the CMIP6 models. . . . .	109
A.5. All CMIP5 models used in Chapter 5 including their effective climate sensitivity (ECS) and <i>x</i> -axis values for the emergent constraints BRI, COX, LIP, SHD, SHL, and SHS. . . . .	110
A.6. As in Table A.5 but for the emergent constraints SU, TIH, TII, VOL, and ZHA. .	111
A.7. As in Table A.5 but for the CMIP6 models. . . . .	112
A.8. As in Table A.6 but for the CMIP6 models. . . . .	113
B.1. List of the seven CMIP5 models used in Chapter 6 alongside the main references.	117



# References

- Abramowitz, G., Herger, N., Gutmann, E., Hammerling, D., Knutti, R., Leduc, M., Lorenz, R., Pincus, R., & Schmidt, G. A. (2019). ESD Reviews: Model dependence in multi-model climate ensembles: weighting, sub-selection and out-of-sample testing. *Earth System Dynamics*, 10(1), 91–105. <https://doi.org/10.5194/esd-10-91-2019>
- Adler, R. F., Huffman, G. J., Chang, A., Ferraro, R., Xie, P. P., Janowiak, J., Rudolf, B., Schneider, U., Curtis, S., Bolvin, D., Gruber, A., Susskind, J., Arkin, P., & Nelkin, E. (2003). The version-2 global precipitation climatology project (GPCP) monthly precipitation analysis (1979–present). *Journal of Hydrometeorology*, 4(6), 1147–1167. [https://doi.org/10.1175/1525-7541\(2003\)004<1147:TVGCP>2.0.CO;2](https://doi.org/10.1175/1525-7541(2003)004<1147:TVGCP>2.0.CO;2)
- Albritton, D. L., Filho, L. G. M., Cubasch, U., Dai, X., Ding, Y., Griggs, D. J., Hewitson, B., Houghton, J. T., Isaksen, I., Karl, T., McFarland, M., Meleshko, V. P., Mitchell, J. F. B., Noguer, M., Nyenzi, B. S., Oppenheimer, M., Penner, J. E., Pollonais, S., Stocker, T., & Trenberth, K. E. (2001). Technical Summary. Cambridge University Press. [https://archive.ipcc.ch/ipccreports/tar/wg1/pdf/WG1\\_TAR-FRONT.PDF](https://archive.ipcc.ch/ipccreports/tar/wg1/pdf/WG1_TAR-FRONT.PDF)
- Allen, M. R., & Ingram, W. J. (2002). Constraints on future changes in climate and the hydrologic cycle. *Nature*, 419(6903), 224–232. <https://doi.org/10.1038/nature01092>
- Amos, M., Young, P. J., Hosking, J. S., Lamarque, J.-F., Abraham, N. L., Akiyoshi, H., Archibald, A. T., Bekki, S., Deushi, M., Jöckel, P., Kinnison, D., Kirner, O., Kunze, M., Marchand, M., Plummer, D. A., Saint-Martin, D., Sudo, K., Tilmes, S., & Yamashita, Y. (2020). Projecting ozone hole recovery using an ensemble of chemistry–climate models weighted by model performance and independence. *Atmospheric Chemistry and Physics*, 20(16), 9961–9977. <https://doi.org/10.5194/acp-20-9961-2020>
- AMSR-E. (2011). AMSR-E Level 3 Sea Surface Temperature for Climate Model Comparison. Ver. 1. PO.DAAC, CA, USA. *Remote Sensing Systems*. <https://doi.org/10.5067/SST00-1D1M1>
- Anav, A., Friedlingstein, P., Kidston, M., Bopp, L., Ciais, P., Cox, P., Jones, C., Jung, M., Myneni, R., & Zhu, Z. (2013). 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>
- Andela, B., Broetz, B., de Mora, L., Drost, N., Eyring, V., Koldunov, N., Lauer, A., Mueller, B., Predoi, V., Righi, M., Schlund, M., Vegas-Regidor, J., Zimmermann, K., Adeniyi, K., Arnone, E., Bellprat, O., Berg, P., Bock, L., Caron, L.-P., … Weigel, K. (2020a). ESMValTool. <https://doi.org/10.5281/ZENODO.3401363>

- Andela, B., Broetz, B., de Mora, L., Drost, N., Eyring, V., Koldunov, N., Lauer, A., Predoi, V., Righi, M., **Schlund, M.**, Vegas-Regidor, J., Zimmermann, K., Bock, L., Diblen, F., Dreyer, L., Earnshaw, P., Hassler, B., Little, B., Loosveldt-Tomas, S., ... Jury, M. (2020b). ESMValCore. <https://doi.org/10.5281/ZENODO.3387139>
- Andrews, T., Gregory, J. M., Webb, M. J., & Taylor, K. E. (2012). Forcing, feedbacks and climate sensitivity in CMIP5 coupled atmosphere-ocean climate models. *Geophysical Research Letters*, 39(9), L09712. <https://doi.org/10.1029/2012gl051607>
- Andrews, T., Andrews, M. B., Bodas-Salcedo, A., Jones, G. S., Kuhlbrodt, T., Manners, J., Menary, M. B., Ridley, J., Ringer, M. A., Sellar, A. A., Senior, C. A., & Tang, Y. (2019). Forcings, Feedbacks, and Climate Sensitivity in HadGEM3-GC3.1 and UKESM1. *Journal of Advances in Modeling Earth Systems*, 11(12), 4377–4394. <https://doi.org/10.1029/2019ms001866>
- Annan, J. D., Hargreaves, J. C., Mauritsen, T., & Stevens, B. (2020). What could we learn about climate sensitivity from variability in the surface temperature record? *Earth System Dynamics*, 11(3), 709–719. <https://doi.org/10.5194/esd-11-709-2020>
- Arora, V. K., Scinocca, J. F., Boer, G. J., Christian, J. R., Denman, K. L., Flato, G. M., Kharin, V. V., Lee, W. G., & Merryfield, W. J. (2011). Carbon emission limits required to satisfy future representative concentration pathways of greenhouse gases. *Geophysical Research Letters*, 38(5), L05805. <https://doi.org/10.1029/2010gl046270>
- 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., & Wu, T. (2013). Carbon-Concentration and Carbon-Climate Feedbacks in CMIP5 Earth System Models. *Journal of Climate*, 26(15), 5289–5314. <https://doi.org/10.1175/Jcli-D-12-00494.1>
- 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. D., ... Ziehn, T. (2020). 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>
- Aumann, H. H., Chahine, M. T., Gautier, C., Goldberg, M. D., Kalnay, E., McMillin, L. M., Revercomb, H., Rosenkranz, P. W., Smith, W. L., Staelin, D. H., Strow, L. L., & Susskind, J. (2003). AIRS/AMSU/HSB on the aqua mission: Design, science objectives, data products, and processing systems. *Ieee Transactions on Geoscience and Remote Sensing*, 41(2), 253–264. <https://doi.org/10.1109/Tgrs.2002.808356>
- Balaji, V., Taylor, K. E., Juckes, M., Lawrence, B. N., Durack, P. J., Lautenschlager, M., Blanton, C., Cinquini, L., Denvil, S., Elkington, M., Guglielmo, F., Guilyardi, E., Hassell, D., Kharin, S., Kindermann, S., Nikonorov, S., Radhakrishnan, A., Stockhouse, M., Weigel, T., & Williams, D. (2018). Requirements for a global data infrastructure in support of CMIP6. *Geoscientific Model Development*, 11(9), 3659–3680. <https://doi.org/10.5194/gmd-11-3659-2018>

- Barnes, E. A., Hurrell, J. W., Ebert-Uphoff, I., Anderson, C., & Anderson, D. (2019). Viewing Forced Climate Patterns Through an AI Lens. *Geophysical Research Letters*, 46(22), 13389–13398. <https://doi.org/10.1029/2019gl084944>
- 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., Liennert, S., ... Zhu, D. (2019). Contrasting effects of CO<sub>2</sub> fertilization, land-use change and warming on seasonal amplitude of Northern Hemisphere CO<sub>2</sub> exchange. *Atmospheric Chemistry and Physics*, 19(19), 12361–12375. <https://doi.org/10.5194/acp-19-12361-2019>
- Beer, R. (2006). TES on the Aura mission: Scientific objectives, measurements, and analysis overview. *Ieee Transactions on Geoscience and Remote Sensing*, 44(5), 1102–1105. <https://doi.org/10.1109/Tgrs.2005.863716>
- Bellucci, A., Gualdi, S., & Navarra, A. (2010). The Double-ITCZ Syndrome in Coupled General Circulation Models: The Role of Large-Scale Vertical Circulation Regimes. *Journal of Climate*, 23(5), 1127–1145. <https://doi.org/10.1175/2009jcli3002.1>
- Bentsen, M., Bethke, I., Debernard, J. B., Iversen, T., Kirkevag, A., Seland, O., Drange, H., Roelandt, C., Seierstad, I. A., Hoose, C., & Kristjansson, J. E. (2013). The Norwegian Earth System Model, NorESM1-M - Part 1: Description and basic evaluation of the physical climate. *Geoscientific Model Development*, 6(3), 687–720. <https://doi.org/10.5194/gmd-6-687-2013>
- Bi, D. H., Dix, M., Marsland, S. J., O'Farrell, S., Rashid, H. A., Uotila, P., Hirst, A. C., Kowalczyk, E., Golebiewski, M., Sullivan, A., Yan, H. L., Hannah, N., Franklin, C., Sun, Z. A., Vohralik, P., Watterson, I., Zhou, X. B., Fiedler, R., Collier, M., ... Puri, K. (2013). The ACCESS coupled model: description, control climate and evaluation. *Australian Meteorological and Oceanographic Journal*, 63(1), 41–64. <https://doi.org/10.22499/2.6301.004>
- Bindoff, N. L., Stott, P. A., AchutaRao, K. M., Allen, M. R., Gillett, N., Gutzler, D., Hansingo, K., Hegerl, G., Hu, Y., Jain, S., Mokhov, I. I., Overland, J., Perlitz, J., Sebbari, R., & Zhang, X. (2013). Detection and Attribution of Climate Change: from Global to Regional. Cambridge University Press. [https://www.ipcc.ch/site/assets/uploads/2018/02/WG1AR5\\_Chapter10\\_FINAL.pdf](https://www.ipcc.ch/site/assets/uploads/2018/02/WG1AR5_Chapter10_FINAL.pdf)
- Bishop, C. M. (2006). *Pattern recognition and machine learning*.
- Bock, L., Lauer, A., Schlund, M., Barreiro, M., Bellouin, N., Jones, C., Meehl, G. A., Predoi, V., Roberts, M. J., & Eyring, V. (2020). Quantifying Progress Across Different CMIP Phases With the ESMValTool. *Journal of Geophysical Research: Atmospheres*, 125(21), e2019JD032321. <https://doi.org/10.1029/2019jd032321>
- Bodas-Salcedo, A., Hill, P. G., Furtado, K., Williams, K. D., Field, P. R., Manners, J. C., Hyder, P., & Kato, S. (2016). Large Contribution of Supercooled Liquid Clouds to the Solar Radiation Budget of the Southern Ocean. *Journal of Climate*, 29(11), 4213–4228. <https://doi.org/10.1175/jcli-d-15-0564.1>

- Bodas-Salcedo, A., Mulcahy, J. P., Andrews, T., Williams, K. D., Ringer, M. A., Field, P. R., & Elsaesser, G. S. (2019). Strong Dependence of Atmospheric Feedbacks on Mixed-Phase Microphysics and Aerosol-Cloud Interactions in HadGEM3. *Journal of Advances in Modeling Earth Systems*, 11(6), 1735–1758. <https://doi.org/10.1029/2019ms001688>
- Bodman, R., Rayner, P. J., & Karoly, D. J. (2013). Uncertainty in temperature projections reduced using carbon cycle and climate observations. *Nature Climate Change*, 3(8), 725–729. <https://doi.org/10.1038/Nclimate1903>
- Booth, B. B. B., Jones, C. D., Collins, M., Totterdell, I. J., Cox, P. M., Sitch, S., Huntingford, C., Betts, R. A., Harris, G. R., & Lloyd, J. (2012). High sensitivity of future global warming to land carbon cycle processes. *Environmental Research Letters*, 7(2), 024002. <https://doi.org/10.1088/1748-9326/7/2/024002>
- Boucher, O., Randall, D., Artaxo, P., Bretherton, C., Feingold, G., Forster, P., Kerminen, V.-M., Kondo, Y., Liao, H., & Lohmann, U. (2013). Clouds and Aerosols. Cambridge University Press. [https://www.ipcc.ch/site/assets/uploads/2018/02/WG1AR5\\_Chapter07\\_FINAL-1.pdf](https://www.ipcc.ch/site/assets/uploads/2018/02/WG1AR5_Chapter07_FINAL-1.pdf)
- Boucher, O., Servonnat, J., Albright, A. L., Aumont, O., Balkanski, Y., Bastrikov, V., Bekki, S., Bonnet, R., Bony, S., Bopp, L., Braconnot, P., Brockmann, P., Cadule, P., Caubel, A., Cheruy, F., Codron, F., Cozic, A., Cugnet, D., D'Andrea, F., ... Vuichard, N. (2020). Presentation and Evaluation of the IPSL-CM6A-LR Climate Model. *Journal of Advances in Modeling Earth Systems*, 12(7), e2019MS002010. <https://doi.org/10.1029/2019MS002010>
- Bretherton, F. P., Bryan, K., & Woods, J. D. (1990). Time-Dependent Greenhouse-Gas-Induced Climate Change. Cambridge University Press. [https://archive.ipcc.ch/ipccreports/far/wg\\_I/ipcc\\_far\\_wg\\_I\\_chapter\\_06.pdf](https://archive.ipcc.ch/ipccreports/far/wg_I/ipcc_far_wg_I_chapter_06.pdf)
- Bretherton, C. S., & Caldwell, P. M. (2020). Combining Emergent Constraints for Climate Sensitivity. *Journal of Climate*, 33(17), 7413–7430. <https://doi.org/10.1175/JCLI-D-19-0911.1>
- Brient, F., Schneider, T., Tan, Z., Bony, S., Qu, X., & Hall, A. (2015). Shallowness of tropical low clouds as a predictor of climate models' response to warming. *Climate Dynamics*, 47(1-2), 433–449. <https://doi.org/10.1007/s00382-015-2846-0>
- Brient, F., & Schneider, T. (2016). Constraints on Climate Sensitivity from Space-Based Measurements of Low-Cloud Reflection. *Journal of Climate*, 29(16), 5821–5835. <https://doi.org/10.1175/Jcli-D-15-0897.1>
- Brient, F. (2020). Reducing Uncertainties in Climate Projections with Emergent Constraints: Concepts, Examples and Prospects. *Advances in Atmospheric Sciences*, 37(1), 1–15. <https://doi.org/10.1007/s00376-019-9140-8>
- Brown, P. T., Stolpe, M. B., & Caldeira, K. (2018). Assumptions for emergent constraints. *Nature*, 563(7729), E1–E3. <https://doi.org/10.1038/s41586-018-0638-5>
- Brunner, L., Lorenz, R., Zumwald, M., & Knutti, R. (2019). Quantifying uncertainty in European climate projections using combined performance-independence weighting. *Environmental Research Letters*, 14(12), 124010. <https://doi.org/10.1088/1748-9326/ab492f>

- Brunner, L., Pendergrass, A. G., Lehner, F., Merrifield, A. L., Lorenz, R., & Knutti, R. (2020). Reduced global warming from CMIP6 projections when weighting models by performance and independence. *Earth System Dynamics*, 11(4), 995–1012. <https://doi.org/10.5194/esd-11-995-2020>
- Caldwell, P. M., Bretherton, C. S., Zelinka, M. D., Klein, S. A., Santer, B. D., & Sanderson, B. M. (2014). Statistical significance of climate sensitivity predictors obtained by data mining. *Geophysical Research Letters*, 41(5), 1803–1808. <https://doi.org/10.1002/2014gl059205>
- Caldwell, P. M., Zelinka, M. D., & Klein, S. A. (2018). Evaluating Emergent Constraints on Equilibrium Climate Sensitivity. *Journal of Climate*, 31(10), 3921–3942. <https://doi.org/10.1175/Jcli-D-17-0631.1>
- Cao, J., Wang, B., Yang, Y. M., Ma, L. B., Li, J., Sun, B., Bao, Y., He, J., Zhou, X., & Wu, L. G. (2018). The NUIST Earth System Model (NESM) version 3: description and preliminary evaluation. *Geoscientific Model Development*, 11(7), 2975–2993. <https://doi.org/10.5194/gmd-11-2975-2018>
- Ceppi, P., Brient, F., Zelinka, M. D., & Hartmann, D. L. (2017). Cloud feedback mechanisms and their representation in global climate models. *Wiley Interdisciplinary Reviews-Climate Change*, 8(4), e465. <https://doi.org/10.1002/wcc.465>
- Charney, J. G., Arakawa, A., Baker, D. J., Bolin, B., Dickinson, R. E., Goody, R. M., Leith, C. E., Stommel, H. M., & Wunsch, C. I. (1979). *Carbon dioxide and climate: a scientific assessment*.
- Cherchi, A., Fogli, P. G., Lovato, T., Peano, D., Iovino, D., Gualdi, S., Masina, S., Scoccimarro, E., Materia, S., Bellucci, A., & Navarra, A. (2019). Global Mean Climate and Main Patterns of Variability in the CMCC-CM2 Coupled Model. *Journal of Advances in Modeling Earth Systems*, 11(1), 185–209. <https://doi.org/10.1029/2018ms001369>
- Ciais, P., Sabine, C., Bala, G., Bopp, L., Brovkin, V., Canadell, J., Chhabra, A., DeFries, R., Galloay, J., & Heimann, M. (2013). Carbon and Other Biogeochemical Cycles. Cambridge University Press. [https://www.ipcc.ch/site/assets/uploads/2018/02/WG1AR5\\_Chapter06\\_FINAL.pdf](https://www.ipcc.ch/site/assets/uploads/2018/02/WG1AR5_Chapter06_FINAL.pdf)
- Collatz, G. J., Ribas-Carbo, M., & Berry, J. A. (1992). Coupled Photosynthesis-Stomatal Conductance Model for Leaves of C4 Plants. *Functional Plant Biology*, 19(5), 519. <https://doi.org/10.1071/pp9920519>
- Collins, W. J., Bellouin, N., Doutriaux-Boucher, M., Gedney, N., Halloran, P., Hinton, T., Hughes, J., Jones, C. D., Joshi, M., Liddicoat, S., Martin, G., O'Connor, F., Rae, J., Senior, C., Sitch, S., Totterdell, I., Wiltshire, A., & Woodward, S. (2011). Development and evaluation of an Earth-System model-HadGEM2. *Geoscientific Model Development*, 4(4), 1051–1075. <https://doi.org/10.5194/gmd-4-1051-2011>
- Collins, M., Knutti, R., Arblaster, J., Dufresne, J.-L., Fichefet, T., Friedlingstein, P., Gao, X., Gutowski, W. J., Johns, T., & Krinner, G. (2013). Long-term Climate Change: Projections, Commitments and Irreversibility. Cambridge University Press. [https://www.ipcc.ch/site/assets/uploads/2018/02/WG1AR5\\_Chapter12\\_FINAL.pdf](https://www.ipcc.ch/site/assets/uploads/2018/02/WG1AR5_Chapter12_FINAL.pdf)
- Collins, W. J., Lamarque, J.-F., Schulz, M., Boucher, O., Eyring, V., Hegglin, M. I., Maycock, A., Myhre, G., Prather, M., Shindell, D., & Smith, S. J. (2017). AerChemMIP: quantifying

- the effects of chemistry and aerosols in CMIP6. *Geoscientific Model Development*, 10(2), 585–607. <https://doi.org/10.5194/gmd-10-585-2017>
- Counillon, F., Keenlyside, N., Bethke, I., Wang, Y. G., Billeau, S., Shen, M. L., & Bentsen, M. (2016). Flow-dependent assimilation of sea surface temperature in isopycnal coordinates with the Norwegian Climate Prediction Model. *Tellus Series a-Dynamic Meteorology and Oceanography*, 68(1), 32437. <https://doi.org/10.3402/tellusa.v68.32437>
- Cox, P. M., Pearson, D., Booth, B. B., Friedlingstein, P., Huntingford, C., Jones, C. D., & Luke, C. M. (2013). Sensitivity of tropical carbon to climate change constrained by carbon dioxide variability. *Nature*, 494(7437), 341–344. <https://doi.org/10.1038/nature11882>
- Cox, P. M., Huntingford, C., & Williamson, M. S. (2018). Emergent constraint on equilibrium climate sensitivity from global temperature variability. *Nature*, 553(7688), 319–322. <https://doi.org/10.1038/nature25450>
- Crafts-Brandner, S. J., & Salvucci, M. E. (2000). Rubisco activase constrains the photosynthetic potential of leaves at high temperature and CO<sub>2</sub>. *Proceedings of the National Academy of Sciences of the United States of America*, 97(24), 13430–13435. <https://doi.org/10.1073/pnas.230451497>
- Cubasch, U., Meehl, G. A., Boer, G. J., Stouffer, R. J., Dix, M., Noda, A., Senior, C. A., Raper, S., & Yap, K. S. (2001). Projections of Future Climate Change. Cambridge University Press. <https://archive.ipcc.ch/ipccreports/tar/wg1/pdf/TAR-09.PDF>
- Cubasch, U., Wuebbles, D., Chen, D., Facchini, M. C., Frame, D., Mahowald, N., & Winther, J.-G. (2013). Introduction. Cambridge University Press. [https://www.ipcc.ch/site/assets/uploads/2017/09/WG1AR5\\_Chapter01\\_FINAL.pdf](https://www.ipcc.ch/site/assets/uploads/2017/09/WG1AR5_Chapter01_FINAL.pdf)
- Danabasoglu, G., Lamarque, J. F., Bacmeister, J., Bailey, D. A., DuVivier, A. K., Edwards, J., Emmons, L. K., Fasullo, J., Garcia, R., Gettelman, A., Hannay, C., Holland, M. M., Large, W. G., Lauritzen, P. H., Lawrence, D. M., Lenaerts, J. T. M., Lindsay, K., Lipscomb, W. H., Mills, M. J., ... Strand, W. G. (2020). The Community Earth System Model Version 2 (CESM2). *Journal of Advances in Modeling Earth Systems*, 12(2), e2019MS001916. <https://doi.org/10.1029/2019MS001916>
- De'ath, G. (2007). Boosted trees for ecological modeling and prediction. *Ecology*, 88(1), 243–251. [https://doi.org/10.1890/0012-9658\(2007\)88\[243:Btfema\]2.0.Co;2](https://doi.org/10.1890/0012-9658(2007)88[243:Btfema]2.0.Co;2)
- Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., ... Vitart, F. (2011). 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>
- de Elia, R., Laprise, R., & Denis, B. (2002). Forecasting skill limits of nested, limited-area models: A perfect-model approach. *Monthly Weather Review*, 130(8), 2006–2023. [https://doi.org/10.1175/1520-0493\(2002\)130<2006:Fslonl>2.0.Co;2](https://doi.org/10.1175/1520-0493(2002)130<2006:Fslonl>2.0.Co;2)
- Delworth, T. L., Stouffer, R. J., Dixon, K. W., Spelman, M. J., Knutson, T. R., Broccoli, A. J., Kushner, P. J., & Wetherald, R. T. (2002). Review of simulations of climate variability

- and change with the GFDL R30 coupled climate model. *Climate Dynamics*, 19(7), 555–574. <https://doi.org/10.1007/s00382-002-0249-5>
- Dittus, A. J., Hawkins, E., Wilcox, L. J., Sutton, R. T., Smith, C. J., Andrews, M. B., & Forster, P. M. (2020). Sensitivity of Historical Climate Simulations to Uncertain Aerosol Forcing. *Geophysical Research Letters*, 47(13), e2019GL085806. <https://doi.org/10.1029/2019gl085806>
- Dix, M., Vohralik, P., Bi, D. H., Rashid, H., Marsland, S., O'Farrell, S., Uotila, P., Hirst, T., Kowalczyk, E., Sullivan, A., Yan, H. L., Franklin, C., Sun, Z. A., Watterson, I., Collier, M., Noonan, J., Rotstayn, L., Stevens, L., Uhe, P., & Puri, K. (2013). The ACCESS coupled model: documentation of core CMIP5 simulations and initial results. *Australian Meteorological and Oceanographic Journal*, 63(1), 83–99. <https://doi.org/10.22499/2.6301.006>
- Donner, L. J., Wyman, B. L., Hemler, R. S., Horowitz, L. W., Ming, Y., Zhao, M., Golaz, J. C., Ginoux, P., Lin, S. J., Schwarzkopf, M. D., Austin, J., Alaka, G., Cooke, W. F., Delworth, T. L., Freidenreich, S. M., Gordon, C. T., Griffies, S. M., Held, I. M., Hurlin, W. J., ... Zeng, F. R. (2011). The Dynamical Core, Physical Parameterizations, and Basic Simulation Characteristics of the Atmospheric Component AM3 of the GFDL Global Coupled Model CM3. *Journal of Climate*, 24(13), 3484–3519. <https://doi.org/10.1175/2011jcli3955.1>
- Du, E. Z., Terrer, C., Pellegrini, A. F. A., Ahlstrom, A., van Lissa, C. J., Zhao, X., Xia, N., Wu, X. H., & Jackson, R. B. (2020). Global patterns of terrestrial nitrogen and phosphorus limitation. *Nature Geoscience*, 13(3), 221–226. <https://doi.org/10.1038/s41561-019-0530-4>
- Dufresne, J.-L., & Bony, S. (2008). An Assessment of the Primary Sources of Spread of Global Warming Estimates from Coupled Atmosphere–Ocean Models. *Journal of Climate*, 21(19), 5135–5144. <https://doi.org/10.1175/2008jcli2239.1>
- Dufresne, J.-L., Foujols, M. A., Denvil, S., Caubel, A., Marti, O., Aumont, O., Balkanski, Y., Bekki, S., Bellenger, H., Benshila, R., Bony, S., Bopp, L., Braconnot, P., Brockmann, P., Cadule, P., Cheruy, F., Codron, F., Cozic, A., Cugnet, D., ... Vuichard, N. (2013). Climate change projections using the IPSL-CM5 Earth System Model: from CMIP3 to CMIP5. *Climate Dynamics*, 40(9-10), 2123–2165. <https://doi.org/10.1007/s00382-012-1636-1>
- 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., Phillipps, P. J., Sentman, L. T., Samuels, B. L., Spelman, M. J., Winton, M., Wittenberg, A. T., & Zadeh, N. (2012). GFDL's ESM2 Global Coupled Climate–Carbon Earth System Models. Part I: Physical Formulation and Baseline Simulation Characteristics. *Journal of Climate*, 25(19), 6646–6665. <https://doi.org/10.1175/Jcli-D-11-00560.1>
- Elith, J., Leathwick, J. R., & Hastie, T. (2008). A working guide to boosted regression trees. *Journal of Animal Ecology*, 77(4), 802–813. <https://doi.org/10.1111/j.1365-2656.2008.01390.x>
- Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., & Taylor, K. E. (2016a). Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6)

- experimental design and organization. *Geoscientific Model Development*, 9(5), 1937–1958. <https://doi.org/10.5194/gmd-9-1937-2016>
- 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., . . . Williams, K. D. (2016b). ESMValTool (v1.0) - a community diagnostic and performance metrics tool for routine evaluation of Earth system models in CMIP. *Geoscientific Model Development*, 9(5), 1747–1802. <https://doi.org/10.5194/gmd-9-1747-2016>
- Eyring, V., Gleckler, P. J., Heinze, C., Stouffer, R. J., Taylor, K. E., Balaji, V., Guilyardi, E., Joussaume, S., Kindermann, S., Lawrence, B. N., Meehl, G. A., Righi, M., & Williams, D. N. (2016c). Towards improved and more routine Earth system model evaluation in CMIP. *Earth System Dynamics*, 7(4), 813–830. <https://doi.org/10.5194/esd-7-813-2016>
- 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., Hurt, G. C., Jahn, A., Jones, C. D., Klein, S. A., Krasting, J. P., Kwiatkowski, L., Lorenz, R., Maloney, E., Meehl, G. A., . . . Williamson, M. S. (2019). Taking climate model evaluation to the next level. *Nature Climate Change*, 9(2), 102–110. <https://doi.org/10.1038/s41558-018-0355-y>
- 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., . . . Zimmermann, K. (2020). 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(7), 3383–3438. <https://doi.org/10.5194/gmd-13-3383-2020>
- Farquhar, G. D., von Caemmerer, S., & Berry, J. A. (1980). A biochemical model of photosynthetic CO<sub>2</sub> assimilation in leaves of C<sub>3</sub> species. *Planta*, 149(1), 78–90. <https://doi.org/10.1007/bf00386231>
- Fasullo, J. T., & Trenberth, K. E. (2012). A Less Cloudy Future: The Role of Subtropical Subsidence in Climate Sensitivity. *Science*, 338(6108), 792–794. <https://doi.org/10.1126/science.1227465>
- Flato, G. M., Marotzke, J., Abiodun, B., Braconnot, P., Chou, S. C., Collins, W., Cox, P., Driouech, F., Emori, S., Eyring, V., Forest, C., Gleckler, P., Guilyardi, E., Jakob, C., Kattsov, V., Reason, C., & Rummukainen, M. (2013). Evaluation of Climate Models. Cambridge University Press. [https://www.ipcc.ch/site/assets/uploads/2018/02/WG1AR5\\_Chapter09\\_FINAL.pdf](https://www.ipcc.ch/site/assets/uploads/2018/02/WG1AR5_Chapter09_FINAL.pdf)
- Flato, G. M. (2011). Earth system models: an overview. *Wiley Interdisciplinary Reviews: Climate Change*, 2(6), 783–800. <https://doi.org/10.1002/wcc.148>
- Fleischer, K., Rammig, A., De Kauwe, M. G., Walker, A. P., Domingues, T. F., Fuchslueger, L., Garcia, S., Goll, D. S., Grandis, A., Jiang, M. K., Haverd, V., Hofhansl, F., Holm, J. A., Kruijt, B., Leung, F., Medlyn, B. E., Mercado, L. M., Norby, R. J., Pak, B., . . . Lapola, D. M. (2019). Amazon forest response to CO<sub>2</sub> fertilization dependent on plant phosphorus

- acquisition. *Nature Geoscience*, 12(9), 736–741. <https://doi.org/10.1038/s41561-019-0404-9>
- Forkel, M., Carvalhais, N., Rodenbeck, C., Keeling, R., Heimann, M., Thonicke, K., Zaehle, S., & Reichstein, M. (2016). Enhanced seasonal CO<sub>2</sub> exchange caused by amplified plant productivity in northern ecosystems. *Science*, 351(6274), 696–699. <https://doi.org/10.1126/science.aac4971>
- Forster, P. M., Andrews, T., Good, P., Gregory, J. M., Jackson, L. S., & Zelinka, M. (2013). Evaluating adjusted forcing and model spread for historical and future scenarios in the CMIP5 generation of climate models. *Journal of Geophysical Research: Atmospheres*, 118(3), 1139–1150. <https://doi.org/10.1002/jgrd.50174>
- Forster, P. M., Maycock, A. C., McKenna, C. M., & Smith, C. J. (2020). Latest climate models confirm need for urgent mitigation. *Nature Climate Change*, 10(1), 7–10. <https://doi.org/10.1038/s41558-019-0660-0>
- Freund, Y., & Schapire, R. E. (1996). Experiments with a New Boosting Algorithm. In *Proceedings of the Thirteenth International Conference on International Conference on Machine Learning*. <https://cseweb.ucsd.edu/~yfreund/papers/boostingexperiments.pdf>
- 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., ... Zeng, N. (2006). 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>
- Friedlingstein, P., O'Sullivan, M., Jones, M. W., Andrew, R. M., Hauck, J., Olsen, A., Peters, G. P., Peters, W., Pongratz, J., Sitch, S., Le Quéré, C., Canadell, J. G., Ciais, P., Jackson, R. B., Alin, S., Aragão, L. E. O. C., Arneth, A., Arora, V., Bates, N. R., ... Zaehle, S. (2020). Global Carbon Budget 2020. *Earth System Science Data*, 12(4), 3269–3340. <https://doi.org/10.5194/essd-12-3269-2020>
- Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, 29(5), 1189–1232. <https://doi.org/10.1214/aos/1013203451>
- Friedman, J. H. (2002). Stochastic gradient boosting. *Computational Statistics & Data Analysis*, 38(4), 367–378. [https://doi.org/10.1016/S0167-9473\(01\)00065-2](https://doi.org/10.1016/S0167-9473(01)00065-2)
- Furtado, K., Field, P. R., Boutle, I. A., Morcrette, C. J., & Wilkinson, J. M. (2016). A Physically Based Subgrid Parameterization for the Production and Maintenance of Mixed-Phase Clouds in a General Circulation Model. *Journal of the Atmospheric Sciences*, 73(1), 279–291. <https://doi.org/10.1175/jas-d-15-0021.1>
- Geer, A. J. (2021). Learning earth system models from observations: machine learning or data assimilation? *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 379(2194), 20200089. <https://doi.org/10.1098/rsta.2020.0089>
- 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., & Zhang, M. H. (2011). The Community Climate System Model Version 4. *Journal of Climate*, 24(19), 4973–4991. <https://doi.org/10.1175/2011jcli4083.1>

- Gentine, P., Pritchard, M., Rasp, S., Reinaudi, G., & Yacalis, G. (2018). Could Machine Learning Break the Convection Parameterization Deadlock? *Geophysical Research Letters*, 45(11), 5742–5751. <https://doi.org/10.1029/2018gl078202>
- Gettelman, A., & Rood, R. B. (2016). *Demystifying Climate Models: A Users Guide to Earth System Models*. Springer Nature.
- Gettelman, A., Hannay, C., Bacmeister, J. T., Neale, R. B., Pendergrass, A. G., Danabasoglu, G., Lamarque, J. F., Fasullo, J. T., Bailey, D. A., Lawrence, D. M., & Mills, M. J. (2019a). High Climate Sensitivity in the Community Earth System Model Version 2 (CESM2). *Geophysical Research Letters*, 46(14), 8329–8337. <https://doi.org/10.1029/2019gl083978>
- Gettelman, A., Mills, M. J., Kinnison, D. E., Garcia, R. R., Smith, A. K., Marsh, D. R., Tilmes, S., Vitt, F., Bardeen, C. G., McInerny, J., Liu, H. L., Solomon, S. C., Polvani, L. M., Emmons, L. K., Lamarque, J. F., Richter, J. H., Glanville, A. S., Bacmeister, J. T., Phillips, A. S., . . . Randel, W. J. (2019b). The Whole Atmosphere Community Climate Model Version 6 (WACCM6). *Journal of Geophysical Research-Atmospheres*, 124(23), 12380–12403. <https://doi.org/10.1029/2019jd030943>
- Gibbs, H. (2006). Olson's Major World Ecosystem Complexes Ranked by Carbon in Live Vegetation: An Updated Database Using the GLC2000 Land Cover Product (NDP-017b, a 2006 update of the original 1985 and 2001 data file)). <https://doi.org/10.3334/CDIAC/LUE.NDP017.2006>
- 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., . . . Stevens, B. (2013). Climate and carbon cycle changes from 1850 to 2100 in MPI-ESM simulations for the Coupled Model Intercomparison Project phase 5. *Journal of Advances in Modeling Earth Systems*, 5(3), 572–597. <https://doi.org/10.1002/jame.20038>
- Gleckler, P. J., Taylor, K. E., & Doutriaux, C. (2008). Performance metrics for climate models. *Journal of Geophysical Research*, 113(D6), D06104. <https://doi.org/10.1029/2007jd008972>
- Golaz, J. C., Caldwell, P. M., Van Roekel, L. P., Petersen, M. R., Tang, Q., Wolfe, J. D., Abeshu, G., Anantharaj, V., Asay-Davis, X. S., Bader, D. C., Baldwin, S. A., Bisht, G., Bogenschutz, P. A., Branstetter, M., Brunke, M. A., Brus, S. R., Burrows, S. M., Cameron-Smith, P. J., Donahue, A. S., . . . Zhu, Q. (2019). The DOE E3SM Coupled Model Version 1: Overview and Evaluation at Standard Resolution. *Journal of Advances in Modeling Earth Systems*, 11(7), 2089–2129. <https://doi.org/10.1029/2018ms001603>
- 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., Kelley, J. J., Daube, B. C., Kort, E. A., Santoni, G. W., & Bent, J. D. (2013). Enhanced Seasonal Exchange of CO<sub>2</sub> by Northern Ecosystems Since 1960. *Science*, 341(6150), 1085–1089. <https://doi.org/10.1126/science.1239207>
- Gray, J. M., Frolking, S., Kort, E. A., Ray, D. K., Kucharik, C. J., Ramankutty, N., & Friedl, M. A. (2014). Direct human influence on atmospheric CO<sub>2</sub> seasonality from increased crop-land productivity. *Nature*, 515(7527), 398–401. <https://doi.org/10.1038/nature13957>

- Gregory, J. M., Ingram, W. J., Palmer, M. A., Jones, G. S., Stott, P. A., Thorpe, R. B., Lowe, J. A., Johns, T. C., & Williams, K. D. (2004). A new method for diagnosing radiative forcing and climate sensitivity. *Geophysical Research Letters*, 31(3), L03205. <https://doi.org/10.1029/2003gl018747>
- Gregory, J. M., & Forster, P. M. (2008). Transient climate response estimated from radiative forcing and observed temperature change. *Journal of Geophysical Research*, 113(D23), D23105. <https://doi.org/10.1029/2008jd010405>
- Gregory, J. M., & Webb, M. (2008). Tropospheric Adjustment Induces a Cloud Component in CO<sub>2</sub> Forcing. *Journal of Climate*, 21(1), 58–71. <https://doi.org/10.1175/2007jcli1834.1>
- Gregory, J. M., Jones, C. D., Cadule, P., & Friedlingstein, P. (2009). Quantifying Carbon Cycle Feedbacks. *Journal of Climate*, 22(19), 5232–5250. <https://doi.org/10.1175/2009jcli2949.1>
- Grise, K. M., & Medeiros, B. (2016). Understanding the Varied Influence of Midlatitude Jet Position on Clouds and Cloud Radiative Effects in Observations and Global Climate Models. *Journal of Climate*, 29(24), 9005–9025. <https://doi.org/10.1175/jcli-d-16-0295.1>
- Guo, Y., Yu, Y., Lin, P., Liu, H., He, B., Bao, Q., Zhao, S., & Wang, X. (2020). Overview of the CMIP6 Historical Experiment Datasets with the Climate System Model CAS FGOALS-f3-L. *Advances in Atmospheric Sciences*, 37(10), 1057–1066. <https://doi.org/10.1007/s00376-020-2004-4>
- Hajima, T., Watanabe, M., Yamamoto, A., Tatebe, H., Noguchi, M. A., Abe, M., Ohgaito, R., Ito, A., Yamazaki, D., Okajima, H., Ito, A., Takata, K., Ogochi, K., Watanabe, S., & Kawamiya, M. (2020). Development of the MIROC-ES2L Earth system model and the evaluation of biogeochemical processes and feedbacks. *Geoscientific Model Development*, 13(5), 2197–2244. <https://doi.org/10.5194/gmd-13-2197-2020>
- Hall, A., & Qu, X. (2006). Using the current seasonal cycle to constrain snow albedo feedback in future climate change. *Geophysical Research Letters*, 33(3), L03502. <https://doi.org/10.1029/2005gl025127>
- Hall, A., Cox, P., Huntingford, C., & Klein, S. (2019). Progressing emergent constraints on future climate change. *Nature Climate Change*, 9(4), 269–278. <https://doi.org/10.1038/s41558-019-0436-6>
- Hamilton, D. S., Lee, L. A., Pringle, K. J., Reddington, C. L., Spracklen, D. V., & Carslaw, K. S. (2014). Occurrence of pristine aerosol environments on a polluted planet. *Proceedings of the National Academy of Sciences*, 111(52), 18466–18471. <https://doi.org/10.1073/pnas.1415440111>
- Hargreaves, J. C., Annan, J. D., Yoshimori, M., & Abe-Ouchi, A. (2012). Can the Last Glacial Maximum constrain climate sensitivity? *Geophysical Research Letters*, 39(24), L24702. <https://doi.org/10.1029/2012gl053872>
- Hargreaves, J. C., & Annan, J. D. (2016). Could the Pliocene constrain the equilibrium climate sensitivity? *Climate of the Past*, 12(8), 1591–1599. <https://doi.org/10.5194/cp-12-1591-2016>

- Harris, I., Jones, P. D., Osborn, T. J., & Lister, D. H. (2014). 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>
- Haustein, K., Allen, M. R., Forster, P. M., Otto, F. E. L., Mitchell, D. M., Matthews, H. D., & Frame, D. J. (2017). A real-time Global Warming Index. *Scientific Reports*, 7(1), 1–6. <https://doi.org/10.1038/s41598-017-14828-5>
- Haverd, V., Smith, B., Canadell, J. G., Cuntz, M., Mikaloff-Fletcher, S., Farquhar, G., Woodgate, W., Briggs, P. R., & Trudinger, C. M. (2020). Higher than expected CO<sub>2</sub> fertilization inferred from leaf to global observations. *Global Change Biology*, 2390–2402. <https://doi.org/10.1111/gcb.14950>
- Hawkins, E., & Sutton, R. (2009). The Potential to Narrow Uncertainty in Regional Climate Predictions. *Bulletin of the American Meteorological Society*, 90(8), 1095–1108. <https://doi.org/10.1175/2009bams2607.1>
- Hawkins, E., & Sutton, R. (2010). The potential to narrow uncertainty in projections of regional precipitation change. *Climate Dynamics*, 37(1-2), 407–418. <https://doi.org/10.1007/s00382-010-0810-6>
- He, B., Bao, Q., Wang, X. C., Zhou, L. J., Wu, X. F., Liu, Y. M., Wu, G. X., Chen, K. J., He, S. C., Hu, W. T., Li, J. D., Li, J. X., Nian, G. K., Wang, L., Yang, J., Zhang, M. H., & Zhang, X. Q. (2019). CAS FGOALS-f3-L Model Datasets for CMIP6 Historical Atmospheric Model Intercomparison Project Simulation. *Advances in Atmospheric Sciences*, 36(8), 771–778. <https://doi.org/10.1007/s00376-019-9027-8>
- He, B., Liu, Y. M., Wu, G. X., Bao, Q., Zhou, T. J., Wu, X. F., Wang, L., Li, J. D., Wang, X. C., Li, J. X., Hu, W. T., Zhang, X. Q., Sheng, C., & Tang, Y. Q. (2020). CAS FGOALS-f3-L Model Datasets for CMIP6 GMMIP Tier-1 and Tier-3 Experiments. *Advances in Atmospheric Sciences*, 37(1), 18–28. <https://doi.org/10.1007/s00376-019-9085-y>
- Hirota, N., Takayabu, Y. N., Watanabe, M., & Kimoto, M. (2011). Precipitation Reproducibility over Tropical Oceans and Its Relationship to the Double ITCZ Problem in CMIP3 and MIROC5 Climate Models. *Journal of Climate*, 24(18), 4859–4873. <https://doi.org/10.1175/2011jcli4156.1>
- Holton, J. R. (2004). *An Introduction to Dynamic Meteorology* (4th ed.). Elsevier Academic Press.
- Hoose, C., Kristjánsson, J. E., Chen, J.-P., & Hazra, A. (2010). A Classical-Theory-Based Parameterization of Heterogeneous Ice Nucleation by Mineral Dust, Soot, and Biological Particles in a Global Climate Model. *Journal of the Atmospheric Sciences*, 67(8), 2483–2503. <https://doi.org/10.1175/2010jas3425.1>
- Huang, Y., & Shahabadi, M. B. (2014). Why logarithmic? A note on the dependence of radiative forcing on gas concentration. *Journal of Geophysical Research: Atmospheres*, 119(24), 13683–13689. <https://doi.org/10.1002/2014jd022466>
- Huntzinger, D. N., Michalak, A. M., Schwalm, C., Ciais, P., King, A. W., Fang, Y., Schaefer, K., Wei, Y., Cook, R. B., Fisher, J. B., Hayes, D., Huang, M., Ito, A., Jain, A. K., Lei, H., Lu, C., Maignan, F., Mao, J., Parazoo, N., ... Zhao, F. (2017). Uncertainty in the response of

- terrestrial carbon sink to environmental drivers undermines carbon-climate feedback predictions. *Scientific Reports*, 7, 1–8. <https://doi.org/10.1038/s41598-017-03818-2>
- Hyder, P., Edwards, J. M., Allan, R. P., Hewitt, H. T., Bracegirdle, T. J., Gregory, J. M., Wood, R. A., Meijers, A. J. S., Mulcahy, J., Field, P., Furtado, K., Bodas-Salcedo, A., Williams, K. D., Copsey, D., Josey, S. A., Liu, C., Roberts, C. D., Sanchez, C., Ridley, J., . . . Belcher, S. E. (2018). Critical Southern Ocean climate model biases traced to atmospheric model cloud errors. *Nature Communications*, 9(1), 1–17. <https://doi.org/10.1038/s41467-018-05634-2>
- IPCC. (2013). Summary for Policymakers. Cambridge University Press. [https://www.ipcc.ch/site/assets/uploads/2018/02/WG1AR5\\_SPM\\_FINAL.pdf](https://www.ipcc.ch/site/assets/uploads/2018/02/WG1AR5_SPM_FINAL.pdf)
- IPCC. (2014). *Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. IPCC. [https://www.ipcc.ch/site/assets/uploads/2018/02/SYR\\_AR5\\_FINAL\\_full.pdf](https://www.ipcc.ch/site/assets/uploads/2018/02/SYR_AR5_FINAL_full.pdf)
- Iversen, T., Bentsen, M., Bethke, I., Debernard, J. B., Kirkevag, A., Selander, O., Drange, H., Kristjansson, J. E., Medhaug, I., Sand, M., & Seierstad, I. A. (2013). The Norwegian Earth System Model, NorESM1-M - Part 2: Climate response and scenario projections. *Geoscientific Model Development*, 6(2), 389–415. <https://doi.org/10.5194/gmd-6-389-2013>
- Jacobson, A. R., Schuldt, K. N., Miller, J. B., Oda, T., Tans, P., Andrews, A., Mund, J., Ott, L., Collatz, G. J., Aalto, T., Afshar, S., Aikin, K., Aoki, S., Apadula, F., Baier, B., Bergamaschi, P., Beyersdorf, A., Biraud, S. C., Bollenbacher, A., . . . Zimnoch, M. (2020). CarbonTracker CT2019. <https://doi.org/10.25925/39M3-6069>
- 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., & Zhou, M. (2014). Description and basic evaluation of Beijing Normal University Earth System Model (BNU-ESM) version 1. *Geoscientific Model Development*, 7(5), 2039–2064. <https://doi.org/10.5194/gmd-7-2039-2014>
- Jimenez-de-la-Cuesta, D., & Mauritsen, T. (2019). Emergent constraints on Earth's transient and equilibrium response to doubled CO<sub>2</sub> from post-1970s global warming. *Nature Geoscience*, 12(11), 902–905. <https://doi.org/10.1038/s41561-019-0463-y>
- Jones, C. D., Arora, V., Friedlingstein, P., Bopp, L., Brovkin, V., Dunne, J., Graven, H., Hoffman, F., Ilyina, T., John, J. G., Jung, M., Kawamiya, M., Koven, C., Pongratz, J., Raddatz, T., Randerson, J. T., & Zaehle, S. (2016). C4MIP – The Coupled Climate–Carbon Cycle Model Intercomparison Project: experimental protocol for CMIP6. *Geoscientific Model Development*, 9(8), 2853–2880. <https://doi.org/10.5194/gmd-9-2853-2016>
- Juckes, M., Taylor, K. E., Durack, P. J., Lawrence, B., Mizielinski, M. S., Pamment, A., Peterschmitt, J.-Y., Rixen, M., & Sénési, S. (2020). The CMIP6 Data Request (DREQ, version 01.00.31). *Geoscientific Model Development*, 13(1), 201–224. <https://doi.org/10.5194/gmd-13-201-2020>
- Jung, M., Reichstein, M., Margolis, H. A., Cescatti, A., Richardson, A. D., Arain, M. A., Arneth, A., Bernhofer, C., Bonal, D., Chen, J. Q., Gianelle, D., Gobron, N., Kiely, G., Kutsch, W.,

- Lasslop, G., Law, B. E., Lindroth, A., Merbold, L., Montagnani, L., ... Williams, C. (2011). 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-Biogeosciences*, 116, G00J07. <https://doi.org/10.1029/2010jg001566>
- Karpechko, A. Y., Maraun, D., & Eyring, V. (2013). 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>
- Kattenberg, A., Giorgi, F., Grassl, H., Meehl, G. A., Mitchell, J. F. B., Stouffer, R. J., Tokioka, T., Weaver, A. J., & Wigley, T. M. L. (1996). Climate models – projections of future climate. Cambridge University Press. [https://archive.ipcc.ch/ipccreports/sar/wg\\_I/ipcc\\_sar\\_wg\\_I\\_full\\_report.pdf](https://archive.ipcc.ch/ipccreports/sar/wg_I/ipcc_sar_wg_I_full_report.pdf)
- Kay, J. E., Bourdages, L., Miller, N. B., Morrison, A., Yettella, V., Chepfer, H., & Eaton, B. (2016). Evaluating and improving cloud phase in the Community Atmosphere Model version 5 using spaceborne lidar observations. *Journal of Geophysical Research: Atmospheres*, 121(8), 4162–4176. <https://doi.org/10.1002/2015jd024699>
- Keeling, C. D., Bacastow, R. B., Bainbridge, A. E., Jr., C. A. E., Guenther, P. R., Waterman, L. S., & Chin, J. F. S. (1976). Atmospheric carbon dioxide variations at Mauna Loa Observatory, Hawaii. *Tellus*, 28(6), 538–551. <https://doi.org/10.3402/tellusa.v28i6.11322>
- Keeling, C. D., Whorf, T. P., Wahlen, M., & Vanderplicht, J. (1995). Interannual Extremes in the Rate of Rise of Atmospheric Carbon-Dioxide since 1980. *Nature*, 375(6533), 666–670. <https://doi.org/10.1038/375666a0>
- Keeling, C. D., Chin, J. F. S., & Whorf, T. P. (1996). Increased activity of northern vegetation inferred from atmospheric CO<sub>2</sub> measurements. *Nature*, 382(6587), 146–149. <https://doi.org/10.1038/382146a0>
- Keeling, C. D., Piper, S. C., Bacastow, R. B., Wahlen, M., Whorf, T. P., Heimann, M., & Meijer, H. A. (2005). Atmospheric CO<sub>2</sub> and <sup>13</sup>CO<sub>2</sub> exchange with the terrestrial biosphere and oceans from 1978 to 2000: Observations and carbon cycle implications.
- Kiehl, J. T. (2007). Twentieth century climate model response and climate sensitivity. *Geophysical Research Letters*, 34(22), L22710. <https://doi.org/10.1029/2007gl031383>
- Knutti, R., Masson, D., & Gettelman, A. (2013). Climate model genealogy: Generation CMIP5 and how we got there. *Geophysical Research Letters*, 40(6), 1194–1199. <https://doi.org/10.1002/grl.50256>
- Knutti, R., & Rugenstein, M. A. A. (2015). Feedbacks, climate sensitivity and the limits of linear models. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 373(2054), 20150146. <https://doi.org/10.1098/rsta.2015.0146>
- Knutti, R., Rugenstein, M. A. A., & Hegerl, G. C. (2017a). Beyond equilibrium climate sensitivity. *Nature Geoscience*, 10(10), 727–736. <https://doi.org/10.1038/Ngeo3017>
- Knutti, R., Sedlacek, J., Sanderson, B. M., Lorenz, R., Fischer, E. M., & Eyring, V. (2017b). A climate model projection weighting scheme accounting for performance and interde-

- pendence. *Geophysical Research Letters*, 44(4), 1909–1918. <https://doi.org/10.1002/2016gl072012>
- Kuhlbrodt, T., Jones, C. G., Sellar, A., Storkey, D., Blockley, E., Stringer, M., Hill, R., Graham, T., Ridley, J., Blaker, A., Calvert, D., Copsey, D., Ellis, R., Hewitt, H., Hyder, P., Ineson, S., Mulcahy, J., Sahaan, A., & Walton, J. (2018). The Low-Resolution Version of HadGEM3 GC3.1: Development and Evaluation for Global Climate. *Journal of Advances in Modeling Earth Systems*, 10(11), 2865–2888. <https://doi.org/10.1029/2018ms001370>
- Kwiatkowski, L., Bopp, L., Aumont, O., Caias, P., Cox, P. M., Laufkötter, C., Li, Y., & Séférian, R. (2017). Emergent constraints on projections of declining primary production in the tropical oceans. *Nature Climate Change*, 7(5), 355–358. <https://doi.org/10.1038/nclimate3265>
- Lauer, A., Jones, C., Eyring, V., Evaldsson, M., Stefan, H. A., Makela, J., Martin, G., Roehrig, R., & Wang, S. Y. (2018). Process-level improvements in CMIP5 models and their impact on tropical variability, the Southern Ocean, and monsoons. *Earth System Dynamics*, 9(1), 33–67. <https://doi.org/10.5194/esd-9-33-2018>
- 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., & Zechlau, S. (2020). 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(9), 4205–4228. <https://doi.org/10.5194/gmd-13-4205-2020>
- Law, R. M., Ziehn, T., Matear, R. J., Lenton, A., Chamberlain, M. A., Stevens, L. E., Wang, Y. P., Srbinovsky, J., Bi, D. H., Yan, H. L., & Vohralík, P. F. (2017). The carbon cycle in the Australian Community Climate and Earth System Simulator (ACCESS-ESM1) - Part 1: Model description and pre-industrial simulation. *Geoscientific Model Development*, 10(7), 2567–2590. <https://doi.org/10.5194/gmd-10-2567-2017>
- Lee, J., Kim, J., Sun, M. A., Kim, B. H., Moon, H., Sung, H. M., Kim, J., & Byun, Y. H. (2020). Evaluation of the Korea Meteorological Administration Advanced Community Earth-System model (K-ACE). *Asia-Pacific Journal of Atmospheric Sciences*, 56(3), 381–395. <https://doi.org/10.1007/s13143-019-00144-7>
- Lee, W. L., Wang, Y. C., Shiu, C. J., Tsai, I. C., Tu, C. Y., Lan, Y. Y., Chen, J. P., Pan, H. L., & Hsu, H. H. (2020). Taiwan Earth System Model Version 1: description and evaluation of mean state. *Geoscientific Model Development*, 13(9), 3887–3904. <https://doi.org/10.5194/gmd-13-3887-2020>
- Li, L. J., Lin, P. F., Yu, Y. Q., Wang, B., Zhou, T. J., Liu, L., Liu, J. P., Bao, Q., Xu, S. M., Huang, W. Y., Xia, K., Pu, Y., Dong, L., Shen, S., Liu, Y. M., Hu, N., Liu, M. M., Sun, W. Q., Shi, X. J., . . . Qiao, F. L. (2013). The flexible global ocean-atmosphere-land system model, Grid-point Version 2: FGOALS-g2. *Advances in Atmospheric Sciences*, 30(3), 543–560. <https://doi.org/10.1007/s00376-012-2140-6>
- Li, L., Yu, Y., Tang, Y., Lin, P., Xie, J., Song, M., Dong, L., Zhou, T., Liu, L., Wang, L., Pu, Y., Chen, X., Chen, L., Xie, Z., Liu, H., Zhang, L., Huang, X., Feng, T., Zheng, W., . . . Wei, J. (2020). The Flexible Global Ocean-Atmosphere-Land System Model Grid-Point

- Version 3 (FGOALS-g3): Description and Evaluation. *Journal of Advances in Modeling Earth Systems*, 12(9), e2019MS002012. <https://doi.org/10.1029/2019ms002012>
- Lian, X., Piao, S. L., Huntingford, C., Li, Y., Zeng, Z. Z., Wang, X. H., Ciais, P., McVicar, T. R., Peng, S. S., Ottle, C., Yang, H., Yang, Y. T., Zhang, Y. Q., & Wang, T. (2018). Partitioning global land evapotranspiration using CMIP5 models constrained by observations. *Nature Climate Change*, 8(7), 640–646. <https://doi.org/10.1038/s41558-018-0207-9>
- Liang, Y., Gillett, N. P., & Monahan, A. H. (2020). Climate Model Projections of 21st Century Global Warming Constrained Using the Observed Warming Trend. *Geophysical Research Letters*, 47(12), e2019GL086757. <https://doi.org/10.1029/2019gl086757>
- Lipat, B. R., Tselioudis, G., Grise, K. M., & Polvani, L. M. (2017). CMIP5 models' shortwave cloud radiative response and climate sensitivity linked to the climatological Hadley cell extent. *Geophysical Research Letters*, 44(11), 5739–5748. <https://doi.org/10.1002/2017gl073151>
- Loeb, N. G., Doelling, D. R., Wang, H. L., Su, W. Y., Nguyen, C., Corbett, J. G., Liang, L. S., Mitrescu, C., Rose, F. G., & Kato, S. (2018). Clouds and the Earth's Radiant Energy System (CERES) Energy Balanced and Filled (EBAF) Top-of-Atmosphere (TOA) Edition-4.0 Data Product. *Journal of Climate*, 31(2), 895–918. <https://doi.org/10.1175/Jcli-D-17-0208.1>
- Lorenz, R., Herger, N., Sedláček, J., Eyring, V., Fischer, E. M., & Knutti, R. (2018). Prospects and Caveats of Weighting Climate Models for Summer Maximum Temperature Projections Over North America. *Journal of Geophysical Research: Atmospheres*, 123(9), 4509–4526. <https://doi.org/10.1029/2017jd027992>
- Lucht, W., Prentice, I. C., Myneni, R. B., Sitch, S., Friedlingstein, P., Cramer, W., Bousquet, P., Buermann, W., & Smith, B. (2002). Climatic control of the high-latitude vegetation greening trend and Pinatubo effect. *Science*, 296(5573), 1687–1689. <https://doi.org/10.1126/science.1071828>
- Mace, G. G., Zhang, Q. Q., Vaughan, M., Marchand, R., Stephens, G., Trepte, C., & Winker, D. (2009). A description of hydrometeor layer occurrence statistics derived from the first year of merged Cloudsat and CALIPSO data. *Journal of Geophysical Research-Atmospheres*, 114(D8), D00A26. <https://doi.org/10.1029/2007jd009755>
- Maki, T., Ikegami, M., Fujita, T., Hirahara, T., Yamada, K., Mori, K., Takeuchi, A., Tsutsumi, Y., Suda, K., & Conway, T. J. (2010). New technique to analyse global distributions of CO<sub>2</sub> concentrations and fluxes from non-processed observational data. *Tellus B: Chemical and Physical Meteorology*, 62(5), 797–809. <https://doi.org/10.1111/j.1600-0889.2010.00488.x>
- Mansfield, L. A., Nowack, P. J., Kasoar, M., Everitt, R. G., Collins, W. J., & Voulgarakis, A. (2020). Predicting global patterns of long-term climate change from short-term simulations using machine learning. *npj Climate and Atmospheric Science*, 3(1), 1–9. <https://doi.org/10.1038/s41612-020-00148-5>
- Mauritsen, T., Bader, J., Becker, T., Behrens, J., Bittner, M., Brokopf, R., Brovkin, V., Claussen, M., Crueger, T., Esch, M., Fast, I., Fiedler, S., Flaeschner, D., Gayler, V., Giorgetta, M., Goll, D. S., Haak, H., Hagemann, S., Hedemann, C., ... Roeckner, E. (2019). Developments

- in the MPI-M Earth System Model version 1.2 (MPI-ESM1.2) and Its Response to Increasing CO<sub>2</sub>. *Journal of Advances in Modeling Earth Systems*, 11(4), 998–1038. <https://doi.org/10.1029/2018ms001400>
- McCoy, D. T., Hartmann, D. L., Zelinka, M. D., Ceppi, P., & Grosvenor, D. P. (2015). Mixed-phase cloud physics and Southern Ocean cloud feedback in climate models. *Journal of Geophysical Research: Atmospheres*, 120(18), 9539–9554. <https://doi.org/10.1002/2015jd023603>
- McCoy, D. T., Tan, I., Hartmann, D. L., Zelinka, M. D., & Storelvmo, T. (2016). On the relationships among cloud cover, mixed-phase partitioning, and planetary albedo in GCMs. *Journal of Advances in Modeling Earth Systems*, 8(2), 650–668. <https://doi.org/10.1002/2015ms000589>
- Meehl, G. A., Washington, W. M., Arblaster, J. M., Hu, A. X., Teng, H. Y., Tebaldi, C., Sanderson, B. N., Lamarque, J. F., Conley, A., Strand, W. G., & White, J. B. (2012). Climate System Response to External Forcings and Climate Change Projections in CCSM4. *Journal of Climate*, 25(11), 3661–3683. <https://doi.org/10.1175/Jcli-D-11-00240.1>
- Meehl, G. A., Senior, C. A., Eyring, V., Flato, G., Lamarque, J.-F., Stouffer, R. J., Taylor, K. E., & Schlund, M. (2020). Context for interpreting equilibrium climate sensitivity and transient climate response from the CMIP6 Earth system models. *Science Advances*, 6(26), eaba1981. <https://doi.org/10.1126/sciadv.aba1981>
- Merrifield, A. L., Brunner, L., Lorenz, R., Medhaug, I., & Knutti, R. (2020). An investigation of weighting schemes suitable for incorporating large ensembles into multi-model ensembles. *Earth System Dynamics*, 11(3), 807–834. <https://doi.org/10.5194/esd-11-807-2020>
- Meyers, M. P., DeMott, P. J., & Cotton, W. R. (1992). New Primary Ice-Nucleation Parameterizations in an Explicit Cloud Model. *Journal of Applied Meteorology*, 31(7), 708–721. [https://doi.org/10.1175/1520-0450\(1992\)031<0708:npinpi>2.0.co;2](https://doi.org/10.1175/1520-0450(1992)031<0708:npinpi>2.0.co;2)
- Mitchell, J. F. B., Manabe, S., Meleshko, V., & Tokioka, T. (1990). Equilibrium Climate Change and its Implications for the Future. [https://archive.ipcc.ch/ipccreports/far/wg\\_I/ipcc\\_far\\_wg\\_I\\_chapter\\_05.pdf](https://archive.ipcc.ch/ipccreports/far/wg_I/ipcc_far_wg_I_chapter_05.pdf)
- Morice, C. P., Kennedy, J. J., Rayner, N. A., & Jones, P. D. (2012). Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: The HadCRUT4 data set. *Journal of Geophysical Research-Atmospheres*, 117(D8), D08101. <https://doi.org/10.1029/2011jd017187>
- 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., & Seneviratne, S. I. (2013). Benchmark products for land evapotranspiration: LandFlux-EVAL multi-data set synthesis. *Hydrology and Earth System Sciences*, 17(10), 3707–3720. <https://doi.org/10.5194/hess-17-3707-2013>
- Muller, W. A., Jungclaus, J. H., Mauritzen, T., Baehr, J., Bittner, M., Budich, R., Bunzel, F., Esch, M., Ghosh, R., Haak, H., Ilyina, T., Kleine, T., Kornblueh, L., Li, H., Modali, K., Notz, D., Pohlmann, H., Roeckner, E., Stemmler, I., ... Marotzke, J. (2018). A

- Higher-resolution Version of the Max Planck Institute Earth System Model (MPI-ESM1.2-HR). *Journal of Advances in Modeling Earth Systems*, 10(7), 1383–1413. <https://doi.org/10.1029/2017ms001217>
- Myhre, G., Shindell, D., Bréon, F.-M., Collins, W., Fuglestvedt, J., Huang, J., Koch, D., Lamarque, J.-F., Lee, D., Mendoza, B., Nakajima, T., Robock, A., Stephens, G., Takemura, T., & Zhang, H. (2013). Anthropogenic and Natural Radiative Forcing. Cambridge University Press. [https://www.ipcc.ch/site/assets/uploads/2018/02/WG1AR5\\_Chapter08\\_FINAL.pdf](https://www.ipcc.ch/site/assets/uploads/2018/02/WG1AR5_Chapter08_FINAL.pdf)
- Myneni, R. B., Keeling, C. D., Tucker, C. J., Asrar, G., & Nemani, R. R. (1997). Increased plant growth in the northern high latitudes from 1981 to 1991. *Nature*, 386(6626), 698–702. <https://doi.org/10.1038/386698a0>
- Nijssse, F. J. M. M., Cox, P. M., & Williamson, M. S. (2020). Emergent constraints on transient climate response (TCR) and equilibrium climate sensitivity (ECS) from historical warming in CMIP5 and CMIP6 models. *Earth System Dynamics*, 11(3), 737–750. <https://doi.org/10.5194/esd-11-737-2020>
- Nowack, P., Runge, J., Eyring, V., & Haigh, J. D. (2020). Causal networks for climate model evaluation and constrained projections. *Nature Communications*, 11(1), 1–11. <https://doi.org/10.1038/s41467-020-15195-y>
- O'Neill, B. C., Kriegler, E., Riahi, K., Ebi, K. L., Hallegatte, S., Carter, T. R., Mathur, R., & van Vuuren, D. P. (2013). A new scenario framework for climate change research: the concept of shared socioeconomic pathways. *Climatic Change*, 122(3), 387–400. <https://doi.org/10.1007/s10584-013-0905-2>
- O'Neill, B. C., Tebaldi, C., van Vuuren, D. P., Eyring, V., Friedlingstein, P., Hurt, G., Knutti, R., Kriegler, E., Lamarque, J.-F., Lowe, J., Meehl, G. A., Moss, R., Riahi, K., & Sanderson, B. M. (2016). The Scenario Model Intercomparison Project (ScenarioMIP) for CMIP6. *Geoscientific Model Development*, 9(9), 3461–3482. <https://doi.org/10.5194/gmd-9-3461-2016>
- O'Neill, B. C., Kriegler, E., Ebi, K. L., Kemp-Benedict, E., Riahi, K., Rothman, D. S., van Ruijven, B. J., van Vuuren, D. P., Birkmann, J., Kok, K., Levy, M., & Solecki, W. (2017). The roads ahead: Narratives for shared socioeconomic pathways describing world futures in the 21st century. *Global Environmental Change*, 42, 169–180. <https://doi.org/10.1016/j.gloenvcha.2015.01.004>
- Park, S., Shin, J., Kim, S., Oh, E., & Kim, Y. (2019). Global Climate Simulated by the Seoul National University Atmosphere Model Version 0 with a Unified Convection Scheme (SAM0-UNICON). *Journal of Climate*, 32(10), 2917–2949. <https://doi.org/10.1175/Jcli-D-18-0796.1>
- Parker, W. S. (2009). Confirmation and Adequacy-for-Purpose in Climate Modelling. *Aristotelian Society Supplementary Volume*, 83(1), 233–249. <https://doi.org/10.1111/j.1467-8349.2009.00180.x>
- Piao, S. L., Liu, Z., Wang, Y. L., Ciais, P., Yao, Y. T., Peng, S., Chevallier, F., Friedlingstein, P., Janssens, I. A., Penuelas, J., Sitch, S., & Wang, T. (2018). On the causes of trends

- in the seasonal amplitude of atmospheric CO<sub>2</sub>. *Global Change Biology*, 24(2), 608–616. <https://doi.org/10.1111/gcb.13909>
- Po-Chedley, S., Proistosescu, C., Armour, K. C., & Santer, B. D. (2018). Climate constraint reflects forced signal. *Nature*, 563(7729), E6–E9. <https://doi.org/10.1038/s41586-018-0640-y>
- Qu, X., Hall, A., Klein, S. A., & Caldwell, P. M. (2013). On the spread of changes in marine low cloud cover in climate model simulations of the 21st century. *Climate Dynamics*, 42(9-10), 2603–2626. <https://doi.org/10.1007/s00382-013-1945-z>
- Rackow, T., Goessling, H. F., Jung, T., Sidorenko, D., Semmler, T., Barbi, D., & Handorf, D. (2018). Towards multi-resolution global climate modeling with ECHAM6-FESOM. Part II: climate variability. *Climate Dynamics*, 50(7-8), 2369–2394. <https://doi.org/10.1007/s00382-016-3192-6>
- Randall, D. A., Wood, R. A., Bony, S., Colman, R., Fichefet, T., Fyfe, J., Kattsov, V., Pitman, A., Shukla, J., Srinivasan, J., Stouffer, R. J., Sumi, A., & Taylor, K. E. (2007). Climate Models and Their Evaluation. Cambridge University Press. <https://www.ipcc.ch/site/assets/uploads/2018/02/ar4-wg1-chapter8-1.pdf>
- Rasp, S., Pritchard, M. S., & Gentine, P. (2018). Deep learning to represent subgrid processes in climate models. *Proceedings of the National Academy of Sciences of the United States of America*, 115(39), 9684–9689. <https://doi.org/10.1073/pnas.1810286115>
- Rayner, N. A., Parker, D. E., Horton, E. B., Folland, C. K., Alexander, L. V., Rowell, D. P., Kent, E. C., & Kaplan, A. (2003). Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *Journal of Geophysical Research-Atmospheres*, 108(D14), 4407. <https://doi.org/10.1029/2002jd002670>
- Read, W., & Livesey, N. (2015). MLS/Aura L2 Relative Humidity With Respect To Ice - Version 4. <https://doi.org/10.5067/AURA/MLS/DATA2019>
- Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., & Prabhat. (2019). Deep learning and process understanding for data-driven Earth system science. *Nature*, 566(7743), 195–204. <https://doi.org/10.1038/s41586-019-0912-1>
- Renoult, M., Annan, J. D., Hargreaves, J. C., Sagoo, N., Flynn, C., Kapsch, M.-L., Li, Q., Lohmann, G., Mikolajewicz, U., Ohgaito, R., Shi, X., Zhang, Q., & Mauritsen, T. (2020). A Bayesian framework for emergent constraints: case studies of climate sensitivity with PMIP. *Climate of the Past*, 16(5), 1715–1735. <https://doi.org/10.5194/cp-16-1715-2020>
- Riahi, K., Rao, S., Krey, V., Cho, C. H., Chirkov, V., Fischer, G., Kindermann, G., Nakicenovic, N., & Rafaj, P. (2011). RCP 8.5-A scenario of comparatively high greenhouse gas emissions. *Climatic Change*, 109(1-2), 33–57. <https://doi.org/10.1007/s10584-011-0149-y>
- Riahi, K., van Vuuren, D. P., Kriegler, E., Edmonds, J., O'Neill, B. C., Fujimori, S., Bauer, N., Calvin, K., Dellink, R., Fricko, O., Lutz, W., Popp, A., Cuaresma, J. C., KC, S., Leimbach, M., Jiang, L., Kram, T., Rao, S., Emmerling, J., . . . Tavoni, M. (2017). The Shared Socioeconomic Pathways and their energy, land use, and greenhouse gas emissions implications: An overview. *Global Environmental Change*, 42, 153–168. <https://doi.org/10.1016/j.gloenvcha.2016.05.009>

- Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why Should I Trust You?": Explaining the Predictions of Any Classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. <https://doi.org/10.1145/2939672.2939778>
- Righi, M., Andela, B., Eyring, V., Lauer, A., Predoi, V., **Schlund, M.**, Vegas-Regidor, J., Bock, L., Brotz, B., de Mora, L., Diblen, F., Dreyer, L., Drost, N., Earnshaw, P., Hassler, B., Koldunov, N., Little, B., Tomas, S. L., & Zimmermann, K. (2020). Earth System Model Evaluation Tool (ESMValTool) v2.0-technical overview. *Geoscientific Model Development*, 13(3), 1179–1199. <https://doi.org/10.5194/gmd-13-1179-2020>
- Rind, D., Orbe, C., Jonas, J., Nazarenko, L., Zhou, T., Kelley, M., Lacis, A., Shindell, D., Faluvegi, G., Romanou, A., Russell, G., Tausnev, N., Bauer, M., & Schmidt, G. (2020). GISS Model E2.2: A Climate Model Optimized for the Middle Atmosphere—Model Structure, Climatology, Variability, and Climate Sensitivity. *Journal of Geophysical Research: Atmospheres*, 125(10), e2019JD032204. <https://doi.org/10.1029/2019jd032204>
- Roe, G. (2009). Feedbacks, Timescales, and Seeing Red. *Annual Review of Earth and Planetary Sciences*, 37(1), 93–115. <https://doi.org/10.1146/annurev.earth.061008.134734>
- Rogers, A., Medlyn, B. E., Dukes, J. S., Bonan, G., von Caemmerer, S., Dietze, M. C., Kattge, J., Leakey, A. D. B., Mercado, L. M., Niinemets, U., Prentice, I. C., Serbin, S. P., Sitch, S., Way, D. A., & Zaehle, S. (2017). A roadmap for improving the representation of photosynthesis in Earth system models. *New Phytologist*, 213(1), 22–42. <https://doi.org/10.1111/nph.14283>
- Rong, X. Y., Li, J., Chen, H. M., Xin, Y. F., Su, J. Z., Hua, L. J., Zhou, T. J., Qi, Y. J., Zhang, Z. Q., Zhang, G., & Li, J. D. (2018). The CAMS Climate System Model and a Basic Evaluation of Its Climatology and Climate Variability Simulation. *Journal of Meteorological Research*, 32(6), 839–861. <https://doi.org/10.1007/s13351-018-8058-x>
- Rossow, W. B., & Schiffer, R. A. (1991). Isccp Cloud Data Products. *Bulletin of the American Meteorological Society*, 72(1), 2–20. [https://doi.org/10.1175/1520-0477\(1991\)072<0002:Icdp>2.0.Co;2](https://doi.org/10.1175/1520-0477(1991)072<0002:Icdp>2.0.Co;2)
- Rotstayn, L. D., Jeffrey, S. J., Collier, M. A., Dravitzki, S. M., Hirst, A. C., Syktus, J. I., & Wong, K. K. (2012). Aerosol- and greenhouse gas-induced changes in summer rainfall and circulation in the Australasian region: a study using single-forcing climate simulations. *Atmospheric Chemistry and Physics*, 12(14), 6377–6404. <https://doi.org/10.5194/acp-12-6377-2012>
- Rugenstein, M. A. A., Bloch-Johnson, J., Abe-Ouchi, A., Andrews, T., Beyerle, U., Cao, L., Chadha, T., Danabasoglu, G., Dufresne, J. L., Duan, L., Foujols, M. A., Frolicher, T., Geoffroy, O., Gregory, J., Knutti, R., Li, C., Marzocchi, A., Mauritzen, T., Menary, M., ... Yang, S. T. (2019). LongRunMIP: Motivation and Design for a Large Collection of Millennial-Length AOGCM Simulations. *Bulletin of the American Meteorological Society*, 100(12), 2551–2570. <https://doi.org/10.1175/Bams-D-19-0068.1>
- Rugenstein, M. A. A., Bloch-Johnson, J., Gregory, J., Andrews, T., Mauritzen, T., Li, C., Frolicher, T. L., Paynter, D., Danabasoglu, G., Yang, S. T., Dufresne, J. L., Cao, L., Schmidt,

- G. A., Abe-Ouchi, A., Geoffroy, O., & Knutti, R. (2020). Equilibrium Climate Sensitivity Estimated by Equilibrating Climate Models. *Geophysical Research Letters*, 47(4), e2019GL083898. <https://doi.org/10.1029/2019GL083898>
- Runge, J., Nowack, P., Kretschmer, M., Flaxman, S., & Sejdinovic, D. (2019). Detecting and quantifying causal associations in large nonlinear time series datasets. *Science Advances*, 5(11), eaau4996. <https://doi.org/10.1126/sciadv.aau4996>
- Rypdal, M., Fredriksen, H. B., Rypdal, K., & Steene, R. J. (2018). Emergent constraints on climate sensitivity. *Nature*, 563(7729), E4–E5. <https://doi.org/10.1038/s41586-018-0639-4>
- Sanderson, B. M., Knutti, R., & Caldwell, P. (2015a). A Representative Democracy to Reduce Interdependency in a Multimodel Ensemble. *Journal of Climate*, 28(13), 5171–5194. <https://doi.org/10.1175/Jcli-D-14-00362.1>
- Sanderson, B. M., Knutti, R., & Caldwell, P. (2015b). Addressing Interdependency in a Multi-model Ensemble by Interpolation of Model Properties. *Journal of Climate*, 28(13), 5150–5170. <https://doi.org/10.1175/JCLI-D-14-00361.1>
- Sanderson, B. M., Wehner, M., & Knutti, R. (2017). Skill and independence weighting for multi-model assessments. *Geoscientific Model Development*, 10(6), 2379–2395. <https://doi.org/10.5194/gmd-10-2379-2017>
- Schlund, M.**, Eyring, V., Camps-Valls, G., Friedlingstein, P., Gentine, P., & Reichstein, M. (2020a). Constraining Uncertainty in Projected Gross Primary Production With Machine Learning. *Journal of Geophysical Research: Biogeosciences*, 125(11), e2019JG005619. <https://doi.org/10.1029/2019jg005619>
- Schlund, M.**, Lauer, A., Gentine, P., Sherwood, S. C., & Eyring, V. (2020b). Emergent constraints on equilibrium climate sensitivity in CMIP5: do they hold for CMIP6? *Earth System Dynamics*, 11(4), 1233–1258. <https://doi.org/10.5194/esd-11-1233-2020>
- Schmidt, G. A., Ruedy, R., Hansen, J. E., Aleinov, I., Bell, N., Bauer, M., Bauer, S., Cairns, B., Canuto, V., Cheng, Y., Del Genio, A., Faluvegi, G., Friend, A. D., Hall, T. M., Hu, Y. Y., Kelley, M., Kiang, N. Y., Koch, D., Lacis, A. A., … Yao, M. S. (2006). Present-day atmospheric simulations using GISS ModelE: Comparison to in situ, satellite, and reanalysis data. *Journal of Climate*, 19(2), 153–192. <https://doi.org/10.1175/Jcli3612.1>
- Séférian, R., Nabat, P., Michou, M., Saint-Martin, D., Voldoire, A., Colin, J., Decharme, B., Delire, C., Berthet, S., Chevallier, M., Sénési, S., Franchisteguy, L., Vial, J., Mallet, M., Joetzjer, E., Geoffroy, O., Guérémy, J.-F., Moine, M.-P., Msadek, R., … Madec, G. (2019). Evaluation of CNRM Earth System Model, CNRM-ESM2-1: Role of Earth System Processes in Present-Day and Future Climate. *Journal of Advances in Modeling Earth Systems*, 11(12), 4182–4227. <https://doi.org/10.1029/2019MS001791>
- Seland, Ø., Bentsen, M., Seland Graff, L., Olivie, D., Tonizzo, 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., … Schulz, M. (2020). The Norwegian Earth System Model, NorESM2 - Evaluation of the CMIP6 DECK and historical simulations. *Geoscientific Model Development Discussions*, 2020, 1–68. <https://doi.org/10.5194/gmd-2019-378>

- Sellar, A. A., Jones, C. G., Mulcahy, J. P., Tang, Y. M., 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., . . . Zerroukat, M. (2019). UKESM1: Description and Evaluation of the UK Earth System Model. *Journal of Advances in Modeling Earth Systems*, 11(12), 4513–4558. <https://doi.org/10.1029/2019ms001739>
- Senftleben, D., Lauer, A., & Karpechko, A. (2020). Constraining Uncertainties in CMIP5 Projections of September Arctic Sea Ice Extent with Observations. *Journal of Climate*, 33(4), 1487–1503. <https://doi.org/10.1175/jcli-d-19-0075.1>
- Sherwood, S. C., Bony, S., & Dufresne, J. L. (2014). Spread in model climate sensitivity traced to atmospheric convective mixing. *Nature*, 505(7481), 37–42. <https://doi.org/10.1038/nature12829>
- Sherwood, S. C., Webb, M. J., Annan, J. D., Armour, K. C., Forster, P. M., Hargreaves, J. C., Hegerl, G., Klein, S. A., Marvel, K. D., Rohling, E. J., Watanabe, M., Andrews, T., Braconnot, P., Bretherton, C. S., Foster, G. L., Hausfather, Z., Heydt, A. S., Knutti, R., Mauritzen, T., . . . Zelinka, M. D. (2020). An Assessment of Earth's Climate Sensitivity Using Multiple Lines of Evidence. *Reviews of Geophysics*, 58(4), e2019RG000678. <https://doi.org/10.1029/2019rg000678>
- Sidorenko, D., Rackow, T., Jung, T., Semmler, T., Barbi, D., Danilov, S., Dethloff, K., Dorn, W., Fieg, K., Goessling, H., Handorf, D., Harig, S., Hiller, W., Juricke, S., Losch, M., Schroter, J., Sein, D. V., & Wang, Q. (2015). Towards multi-resolution global climate modeling with ECHAM6-FESOM. Part I: model formulation and mean climate. *Climate Dynamics*, 44(3-4), 757–780. <https://doi.org/10.1007/s00382-014-2290-6>
- Simpkins, G. (2017). Progress in climate modelling. *Nature Climate Change*, 7(10), 684–685. <https://doi.org/10.1038/nclimate3398>
- Smith, T. M., & Reynolds, R. W. (2003). Extended reconstruction of global sea surface temperatures based on COADS data (1854–1997). *Journal of Climate*, 16(10), 1495–1510. <https://doi.org/10.1175/1520-0442-16.10.1495>
- Soden, B. J., & Held, I. M. (2006). An Assessment of Climate Feedbacks in Coupled Ocean–Atmosphere Models. *Journal of Climate*, 19(14), 3354–3360. <https://doi.org/10.1175/jcli3799.1>
- Soden, B. J., Held, I. M., Colman, R., Shell, K. M., Kiehl, J. T., & Shields, C. A. (2008). Quantifying Climate Feedbacks Using Radiative Kernels. *Journal of Climate*, 21(14), 3504–3520. <https://doi.org/10.1175/2007jcli2110.1>
- Solomon, S., Qin, D., Manning, M., Alley, R. B., Berntsen, T., Bindoff, N. L., Chen, Z., Chidthaisong, A., Gregory, J. M., Hegerl, G. C., Heimann, M., Hewitson, B., Hoskins, B. J., Joos, F., Jouzel, J., Kattsov, V., Lohmann, U., Matsuno, T., Molina, M., . . . Wratt, D. (2007). Technical Summary. Cambridge University Press. <https://www.ipcc.ch/site/assets/uploads/2018/02/ar4-wg1-ts-1.pdf>
- Stocker, T. F., Qin, D., Plattner, G.-K., Alexander, L. V., Allen, S. K., Bindoff, N. L., Bréon, F.-M., Church, J. A., Cubasch, U., Emori, S., Forster, P., Friedlingstein, P., Gillett, N., Gregory,

- J. M., Hartmann, D. L., Jansen, E., Kirtman, B., Knutti, R., Kumar, K. K., . . . Xie, S.-P. (2013). Technical Summary. Cambridge University Press. [https://www.ipcc.ch/site/assets/uploads/2018/02/WG1AR5\\_TS\\_FINAL.pdf](https://www.ipcc.ch/site/assets/uploads/2018/02/WG1AR5_TS_FINAL.pdf)
- Su, H., Jiang, J. H., Zhai, C. X., Shen, T. J., Neelin, J. D., Stephens, G. L., & Yung, Y. L. (2014). Weakening and strengthening structures in the Hadley Circulation change under global warming and implications for cloud response and climate sensitivity. *Journal of Geophysical Research-Atmospheres*, 119(10), 5787–5805. <https://doi.org/10.1002/2014jd021642>
- 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. J., Lee, W. G., Majaess, F., Saenko, O. A., Seiler, C., Seinen, C., Shao, A., Sigmond, M., Solheim, L., . . . Winter, B. (2019). The Canadian Earth System Model version 5 (CanESM5.0.3). *Geoscientific Model Development*, 12(11), 4823–4873. <https://doi.org/10.5194/gmd-12-4823-2019>
- Tan, I., Storelvmo, T., & Zelinka, M. D. (2016). Observational constraints on mixed-phase clouds imply higher climate sensitivity. *Science*, 352(6282), 224–227. <https://doi.org/10.1126/science.aad5300>
- Tatebe, H., Ogura, T., Nitta, T., Komuro, Y., Ogochi, K., Takemura, T., Sudo, K., Sekiguchi, M., Abe, M., Saito, F., Chikira, M., Watanabe, S., Mori, M., Hirota, N., Kawatani, Y., Mochizuki, T., Yoshimura, K., Takata, K., O’ishi, R., . . . Kimoto, M. (2019). Description and basic evaluation of simulated mean state, internal variability, and climate sensitivity in MIROC6. *Geoscientific Model Development*, 12(7), 2727–2765. <https://doi.org/10.5194/gmd-12-2727-2019>
- Taylor, K. E., Stouffer, R. J., & Meehl, G. A. (2012). An Overview of Cmip5 and the Experiment Design. *Bulletin of the American Meteorological Society*, 93(4), 485–498. <https://doi.org/10.1175/Bams-D-11-00094.1>
- Tebaldi, C., & Knutti, R. (2007). The use of the multi-model ensemble in probabilistic climate projections. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 365(1857), 2053–2075. <https://doi.org/10.1098/rsta.2007.2076>
- Thomas, R. T., Prentice, L. C., Graven, H., Ciais, P., Fisher, J. B., Hayes, D. J., Huang, M. Y., Huntzinger, D. N., Ito, A., Jain, A., Mao, J. F., Michalak, A. M., Peng, S. S., Poulter, B., Ricciuto, D. M., Shi, X. Y., Schwalm, C., Tian, H. Q., & Zeng, N. (2016). Increased light-use efficiency in northern terrestrial ecosystems indicated by CO<sub>2</sub> and greening observations. *Geophysical Research Letters*, 43(21), 11339–11349. <https://doi.org/10.1002/2016gl070710>
- Tian, B. J. (2015). Spread of model climate sensitivity linked to double-Intertropical Convergence Zone bias. *Geophysical Research Letters*, 42(10), 4133–4141. <https://doi.org/10.1002/2015gl064119>
- Tibshirani, R. (1996). Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society Series B-Methodological*, 58(1), 267–288. <https://doi.org/10.1111/j.2517-6161.1996.tb02080.x>

- Tokarska, K. B., Stolpe, M. B., Sippel, S., Fischer, E. M., Smith, C. J., Lehner, F., & Knutti, R. (2020). Past warming trend constrains future warming in CMIP6 models. *Science Advances*, 6(12), eaaz9549. <https://doi.org/10.1126/sciadv.aaz9549>
- UNFCCC. (2015). Adoption of the Paris Agreement, 1–32. <http://unfccc.int/resource/docs/2015/cop21/eng/109r01.pdf>
- Vial, J., Dufresne, J. L., & Bony, S. (2013). On the interpretation of inter-model spread in CMIP5 climate sensitivity estimates. *Climate Dynamics*, 41(11-12), 3339–3362. <https://doi.org/10.1007/s00382-013-1725-9>
- Voldoire, A., Sanchez-Gomez, E., Melia, D. S. Y., Decharme, B., Cassou, C., Senesi, S., Valcke, S., Beau, I., Alias, A., Chevallier, M., Deque, M., Deshayes, J., Douville, H., Fernandez, E., Madec, G., Maisonnave, E., Moine, M. P., Planton, S., Saint-Martin, D., . . . Chauvin, F. (2013). The CNRM-CM5.1 global climate model: description and basic evaluation. *Climate Dynamics*, 40(9-10), 2091–2121. <https://doi.org/10.1007/s00382-011-1259-y>
- Voldoire, A., Saint-Martin, D., Sénési, S., Decharme, B., Alias, A., Chevallier, M., Colin, J., Guérémy, J.-F., Michou, M., Moine, M.-P., Nabat, P., Roehrig, R., y Méliá, D. S., Séférian, R., Valcke, S., Beau, I., Belamari, S., Berthet, S., Cassou, C., . . . Waldman, R. (2019). Evaluation of CMIP6 DECK Experiments With CNRM-CM6-1. *Journal of Advances in Modeling Earth Systems*, 11(7), 2177–2213. <https://doi.org/10.1029/2019ms001683>
- Volodin, E. M., Dianskii, N. A., & Gusev, A. V. (2010). Simulating present-day climate with the INMCM4.0 coupled model of the atmospheric and oceanic general circulations. *Izvestiya Atmospheric and Oceanic Physics*, 46(4), 414–431. <https://doi.org/10.1134/S000143381004002x>
- Volodin, E. M., Mortikov, E. V., Kostrykin, S. V., Galin, V. Y., Lykosov, V. N., Gritsun, A. S., Diansky, N. A., Gusev, A. V., & Yakovlev, N. G. (2017a). Simulation of modern climate with the new version of the INM RAS climate model. *Izvestiya Atmospheric and Oceanic Physics*, 53(2), 142–155. <https://doi.org/10.1134/S0001433817020128>
- Volodin, E. M., Mortikov, E. V., Kostrykin, S. V., Galin, V. Y., Lykossov, V. N., Gritsun, A. S., Diansky, N. A., Gusev, A. V., & Iakovlev, N. G. (2017b). Simulation of the present-day climate with the climate model INMCM5. *Climate Dynamics*, 49(11-12), 3715–3734. <https://doi.org/10.1007/s00382-017-3539-7>
- Volodin, E. M. (2008). Relation between temperature sensitivity to doubled carbon dioxide and the distribution of clouds in current climate models. *Izvestiya Atmospheric and Oceanic Physics*, 44(3), 288–299. <https://doi.org/10.1134/S0001433808030043>
- Walker, A. P., Kauwe, M. G. D., Bastos, A., Belmecheri, S., Georgiou, K., Keeling, R. F., McMahon, S. M., Medlyn, B. E., Moore, D. J. P., Norby, R. J., Zaehle, S., Anderson-Teixeira, K. J., Battipaglia, G., Brienen, R. J. W., Cabugao, K. G., Cailleret, M., Campbell, E., Canadell, J. G., Ciais, P., . . . Zuidema, P. A. (2020). Integrating the evidence for a terrestrial carbon sink caused by increasing atmospheric CO<sub>2</sub>. *New Phytologist*, 229(5), 2413–2445. <https://doi.org/10.1111/nph.16866>

- Wang, T., Jiang, J., Zhang, M., Zhang, H., He, J., Hao, H., & Chi, X. (2020). Design and Research of CAS-CIG for Earth System Models. *Earth and Space Science*, 7(7), e2019EA000965. <https://doi.org/10.1029/2019EA000965>
- Watanabe, M., Suzuki, T., O'ishi, R., Komuro, Y., Watanabe, S., Emori, S., Takemura, T., Chikira, M., Ogura, T., Sekiguchi, M., Takata, K., Yamazaki, D., Yokohata, T., Nozawa, T., Hasumi, H., Tatebe, H., & Kimoto, M. (2010). Improved Climate Simulation by MIROC5. Mean States, Variability, and Climate Sensitivity. *Journal of Climate*, 23(23), 6312–6335. <https://doi.org/10.1175/2010jcli3679.1>
- 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., & Kawamiya, M. (2011). MIROC-ESM 2010: model description and basic results of CMIP5-20c3m experiments. *Geoscientific Model Development*, 4(4), 845–872. <https://doi.org/10.5194/gmd-4-845-2011>
- WCRP. (2020). WCRP Coupled Model Intercomparison Project (CMIP). <https://www.wcrp-climate.org/wgcm-cmip>
- Webb, M. J., Andrews, T., Bodas-Salcedo, A., Bony, S., Bretherton, C. S., Chadwick, R., Chepfer, H., Douville, H., Good, P., Kay, J. E., Klein, S. A., Marchand, R., Medeiros, B., Siebesma, A. P., Skinner, C. B., Stevens, B., Tselioudis, G., Tsushima, Y., & Watanabe, M. (2017). The Cloud Feedback Model Intercomparison Project (CFMIP) contribution to CMIP6. *Geoscientific Model Development*, 10(1), 359–384. <https://doi.org/10.5194/gmd-10-359-2017>
- 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., ... Eyring, V. (2020). 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 Discussions*, accepted, 1–43. <https://doi.org/10.5194/gmd-2020-244>
- Wenzel, S., Cox, P. M., Eyring, V., & Friedlingstein, P. (2014). Emergent constraints on climate-carbon cycle feedbacks in the CMIP5 Earth system models. *Journal of Geophysical Research-Biogeosciences*, 119(5), 794–807. <https://doi.org/10.1002/2013jg002591>
- Wenzel, S., Cox, P. M., Eyring, V., & Friedlingstein, P. (2016a). Projected land photosynthesis constrained by changes in the seasonal cycle of atmospheric CO<sub>2</sub>. *Nature*, 538(7626), 499–501. <https://doi.org/10.1038/nature19772>
- Wenzel, S., Eyring, V., Gerber, E. P., & Karpechko, A. Y. (2016b). Constraining Future Summer Austral Jet Stream Positions in the CMIP5 Ensemble by Process-Oriented Multiple Diagnostic Regression. *Journal of Climate*, 29(2), 673–687. <https://doi.org/10.1175/Jcli-D-15-0412.1>
- Wieder, W. R., Boehnert, J., Bonan, G. B., & Langseth, M. (2014). Regridded Harmonized World Soil Database v1.2. <https://doi.org/10.3334/ORNLDAA/1247>
- Williams, K. D., Copsey, D., Blockley, E. W., Bodas-Salcedo, A., Calvert, D., Comer, R., Davis, P., Graham, T., Hewitt, H. T., Hill, R., Hyder, P., Ineson, S., Johns, T. C., Keen, A. B.,

- Lee, R. W., Megann, A., Milton, S. F., Rae, J. G. L., Roberts, M. J., . . . Xavier, P. K. (2018). The Met Office Global Coupled Model 3.0 and 3.1 (GC3.0 and GC3.1) Configurations. *Journal of Advances in Modeling Earth Systems*, 10(2), 357–380. <https://doi.org/10.1002/2017ms001115>
- Williams, K. D., Hewitt, A. J., & Bodas-Salcedo, A. (2020). Use of Short-Range Forecasts to Evaluate Fast Physics Processes Relevant for Climate Sensitivity. *Journal of Advances in Modeling Earth Systems*, 12(4), e2019MS001986. <https://doi.org/10.1029/2019ms001986>
- Winkler, A. J., Myneni, R. B., Alexandrov, G. A., & Brovkin, V. (2019a). Earth system models underestimate carbon fixation by plants in the high latitudes. *Nature Communications*, 10, 1–8. <https://doi.org/10.1038/s41467-019-10863-z>
- Winkler, A. J., Myneni, R. B., & Brovkin, V. (2019b). Investigating the applicability of emergent constraints. *Earth System Dynamics*, 10(3), 501–523. <https://doi.org/10.5194/esd-10-501-2019>
- Wu, T. W., Song, L. C., Li, W. P., Wang, Z. Z., Zhang, H., Xin, X. G., Zhang, Y. W., Zhang, L., Li, J. L., Wu, F. H., Liu, Y. M., Zhang, F., Shi, X. L., Chu, M., Zhang, J., Fang, Y. J., Wang, F., Lu, Y. X., Liu, X. W., . . . Zhou, M. Y. (2014). An Overview of BCC Climate System Model Development and Application for Climate Change Studies. *Journal of Meteorological Research*, 28(1), 34–56. <https://doi.org/10.1007/s13351-014-3041-7>
- Wu, T. W., Lu, Y. X., Fang, Y. J., Xin, X. G., Li, L., Li, W. P., Jie, W. H., Zhang, J., Liu, Y. M., Zhang, L., Zhang, F., Zhang, Y. W., Wu, F. H., Li, J. L., Chu, M., Wang, Z. Z., Shi, X. L., Liu, X. W., Wei, M., . . . Liu, X. H. (2019). The Beijing Climate Center Climate System Model (BCC-CSM): the main progress from CMIP5 to CMIP6. *Geoscientific Model Development*, 12(4), 1573–1600. <https://doi.org/10.5194/gmd-12-1573-2019>
- Wyser, K., van Noije, T., Yang, S. T., von Hardenberg, J., O'Donnell, D., & Doscher, R. (2020). On the increased climate sensitivity in the EC-Earth model from CMIP5 to CMIP6. *Geoscientific Model Development*, 13(8), 3465–3474. <https://doi.org/10.5194/gmd-13-3465-2020>
- 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., & Kitoh, A. (2012). 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. <https://doi.org/10.2151/jmsj.2012-A02>
- 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., & Ishii, M. (2019). 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*, 97(5), 931–965. <https://doi.org/10.2151/jmsj.2019-051>
- Zelinka, M. D., Myers, T. A., Mccoy, D. T., Po-Chedley, S., Caldwell, P. M., Ceppi, P., Klein, S. A., & Taylor, K. E. (2020). Causes of Higher Climate Sensitivity in CMIP6 Models. *Geophysical Research Letters*, 47(1), e2019GL085782. <https://doi.org/10.1029/2019GL085782>

- Zhai, C. X., Jiang, J. H., & Su, H. (2015). Long-term cloud change imprinted in seasonal cloud variation: More evidence of high climate sensitivity. *Geophysical Research Letters*, 42(20), 8729–8737. <https://doi.org/10.1002/2015gl065911>
- Zhang, Y., Parazoo, N. C., Williams, A. P., Zhou, S., & Gentine, P. (2020). Large and projected strengthening moisture limitation on end-of-season photosynthesis. *Proceedings of the National Academy of Sciences of the United States of America*, 117(17), 9216–9222. <https://doi.org/10.1073/pnas.1914436117>
- Zhao, F., & Zeng, N. (2014). Continued increase in atmospheric CO<sub>2</sub> seasonal amplitude in the 21st century projected by the CMIP5 Earth system models. *Earth System Dynamics*, 5(2), 423–439. <https://doi.org/10.5194/esd-5-423-2014>
- 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., & Zaehle, S. (2016). Role of CO<sub>2</sub>, climate and land use in regulating the seasonal amplitude increase of carbon fluxes in terrestrial ecosystems: a multimodel analysis. *Biogeosciences*, 13(17), 5121–5137. <https://doi.org/10.5194/bg-13-5121-2016>
- Zhu, Z. C., Bi, J., Pan, Y. Z., Ganguly, S., Anav, A., Xu, L., Samanta, A., Piao, S. L., Nemani, R. R., & Myneni, R. B. (2013). 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>
- Ziehn, T., Lenton, A., Law, R. M., Matear, R. J., & Chamberlain, M. A. (2017). The carbon cycle in the Australian Community Climate and Earth System Simulator (ACCESS-ESM1) - Part 2: Historical simulations. *Geoscientific Model Development*, 10(7), 2591–2614. <https://doi.org/10.5194/gmd-10-2591-2017>



# Acknowledgments

This work has been supported by the European Union's Horizon 2020 project *Coordinated Research in Earth Systems and Climate: Experiments, kNowledge, Dissemination and Outreach* (CRESCKENDO) under Grant Agreement Number 641816, the European Union's Horizon 2020 project *Climate-Carbon Interactions in the Coming Century (4C)* under Grant Agreement Number 821003, the European Research Council Synergy Grant *Understanding and Modelling the Earth System with Machine Learning (USMILE)* under Grant Agreement Number 855187, and the *Advanced Earth System Model Evaluation for CMIP (EVal4CMIP)* project funded by the Helmholtz Society. I acknowledge the WCRP's WGCM, which promoted CMIP5 and CMIP6. I thank the climate modeling groups for producing and making available their model output, the ESGF for archiving the data and providing access, and the multiple funding agencies who support CMIP and the ESGF. The computational resources of the German Climate Computing Center (DKRZ; Hamburg, Germany) where the ESMValTool is fully integrated into the ESGF infrastructure are kindly acknowledged.

First of all, I would like to express my sincere gratitude to my supervisor and primary examiner, Prof. Dr. Veronika Eyring. Without her continuous support, advice, enthusiasm, expertise, and patience, this Ph.D. project would not have been possible. I further thank her for providing me the opportunity to visit many international conferences, which allowed me to meet amazing people and collaborate with colleagues from all around the world.

I am grateful to my second examiner, Prof. Dr. Pierre Gentine, for his great advice and feedback on my work.

My sincere thanks go to all my co-authors, namely Prof. Dr. Gustau Camps-Valls, Prof. Dr. Veronika Eyring, Prof. Dr. Pierre Friedlingstein, Prof. Dr. Pierre Gentine, Dr. Axel Lauer, Prof. Dr. Markus Reichstein, and Prof. Dr. Steven Sherwood, for sharing their expertise, their advice, and helpful comments on my studies.

I would like to thank Prof. Dr. Gustau Camps-Valls and his Image and Signal Processing (ISP) group for hosting me during my research stay at the University of Valencia. I really enjoyed the time there, where I had the opportunity to meet great people and expand my knowledge about machine learning.

Many thanks to Dr. Axel Lauer for being an amazing second author and for the endless patience he showed whenever I had questions for him. I am also very grateful to Dr. Birgit Hassler for her advice in all circumstances and for providing the urgently-needed chocolate and cookies during long hours in the office. A warm thank you to all my colleagues at DLR-EVA and IUP for their professional advice, helpful discussions, and simply the great time we spent together inside and outside of the office.

## *Acknowledgments*

---

A special thank you to the ESMValTool core development team for providing me useful insights into modern software development and for the great workshops and beers we shared.

I am deeply grateful to my friends for all the amazing time we spent together in the many years of our friendship, which allowed me to stay focused on my work.

A very special thanks to the (extended) Langer family for their endless support and advice in all situations of life.

I am forever grateful to my parents and my sister for their everlasting support and their unconditional love, without which this work would not have been possible.

Finally, I would like to thank Hannah with all my heart, for everything.