UNIVERSITY OF BREMEN

INSTITUTE OF ENVIRONMENTAL PHYSICS (IUP)

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

DISSERTATION

Author:
Manuel SCHLUND

Supervisors:
Prof. Dr. Veronika EYRING
Prof. Dr. Pierre GENTINE

This thesis is submitted for the degree
Doktor der Naturwissenschaften (Dr. rer. nat.)

March 2021

Abstract (English version)

TBA.

Abstract (German version)

TBA.

Contents

1. Introduction

1.1. Motivation

- Why is climate change relevant for the society?

 - Cite latest warming reports, also other variables (droughts, extreme events, etc.)

 - basic physics of greenhouse gases (vibration modes, etc.)

 - definition of radiative forcing

 - well-mixed greenhouse gases (water vapor, CO2, CH4) + sources

 - long-lived vs. short-lived GHGs?

 - aerosols + cooling effect

 - natural vs. anthropogenic effects (D&A) -> clear evidence that climate change over 21st century is caused by human influence. Website: (https://globalwarmingindex.org/, Haustein et al., 2017)

 - remove description of emergent constraints from section 1.3, since this will probably be included in this section.

 - answer question: what is climate sensitivity? What is Effective Climate Sensitivity (ECS)? What is Transient Climate Response (TCR)? What is CMIP?

 - One sentence on Machine Learning (ML).

 - Mention IPCC assessments, older ones and upcoming AR6

 - "climate model" is used as generic term which include Atmosphere-Ocean General Circulation Models (AOGCMs) and Earth System Models (ESMs) (see section 2.1.1) (subcategories). Since most modern climate models participating in CMIP are ESMs (or at least have ESM versions), the terms "climate model" and "ESM" are used interchangeably in this thesis.

1.2. Key Science Questions

This section will present the **AWESOME** key science questions!!!

1. What is the range of climate sensitivity in the latest generation of ESMs 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). 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 contents from these publications (text, figures and tables) shown in this thesis originate from the author of this thesis. A detailed list of contributions to these studies (including their full reference) 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 over the contributions made to the ESMValTool, an open-source software for the analysis of ESMs. These contributions helped improving the routine evaluation of ESMs which is useful for the entire scientific community and lead 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 (CMIP6). This work is already published in two scientific publications (Bock et al. 2020; Meehl et al. 2020). Since the ECS and TCR are considerably higher in this new climate model generation, chapter 5 describes the assessment of emergent constraints (a technique to reduce uncertainties in multi-model climate projections, see section 2.4.1) on the ECS for these ESMs. The contents of this chapter are published in *Earth System Dynamics* (Schlund et al. 2020b). Chapter 6 focuses on a new method to reduce uncertainties in gridded multi-model projections of the future climate with observations based on ML. As an example, the method is applied to the photosynthesis rate at the end of the 21st 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 works.

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 biogeochemical 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 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. 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). The basis of these so-called Atmosphere-Ocean General Circulation Models (AOGCMs) (Flato et al. 2013) is the numerical solving of the differential equations describing the exchange of energy and matter between these physical components.

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, a 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
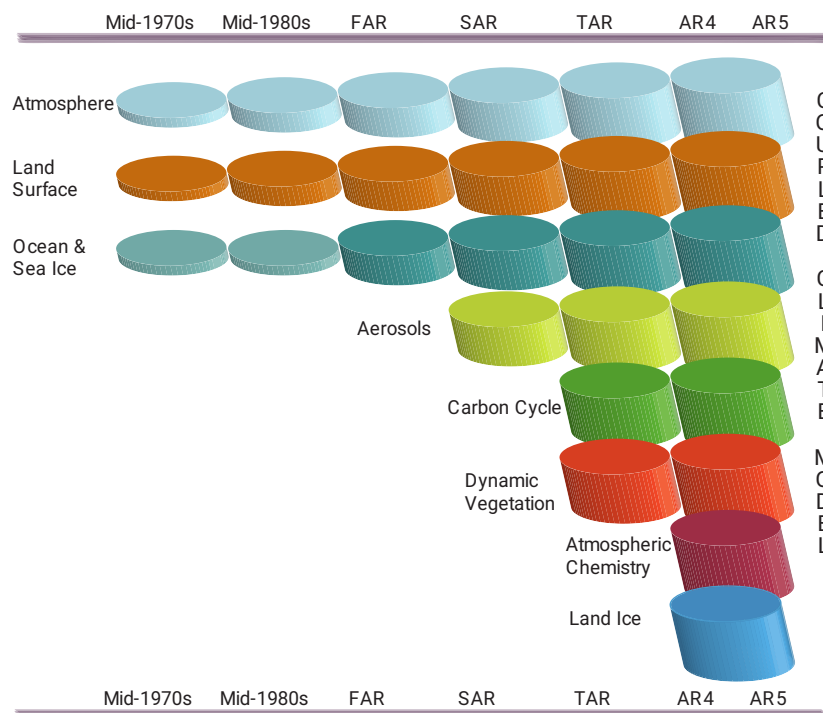
Figure 2.1.: Historical evolution of coupled climate models over the last 45 years. In early days, these models were so-called Atmosphere-Ocean General Circulation Models (AOGCMs) and only included three components: the atmosphere, the land surface and the ocean. Over the time, the individual components 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).

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 (CO_2), immediately impacts the global carbon cycle by inserting additional carbon into the system (for details see section 2.3). Further examples include land use changes like the deforestation of tropical rainforests, which also directly influences 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 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 CO_2 can act as GHG by introducing an additional radiative forcing to the Earth System leading to increasing surface temperatures, this uptake of 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 to a certain degree is crucial for the various research groups developing ESMs all around the world in order 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, with the objective 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, November). One major element of CMIP is to establish common standards, coordination, infrastructure and documentation in order 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 to run a *historical* simulation of the period 1850–2014 and the so-called Diagnostic, Evaluation, and Characterisation 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_2 (*abrupt-4xCO2*) and a simulation forced with a 1 % per year increase of the atmospheric CO_2 concentration (*1pctCO2*) (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 21 (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, a set of emissions and land use changes is 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 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 Modelss (IAMs) have shown to be feasible. For example, SSP5-8.5 represents a scenario based on a fossil-fuel development with a radiative forcing of $8.5\,\mathrm{Wm}^{-2}$ in 2100 while SSP1-2.6 represents a sustainable future with a radiative forcing of $2.6\,\mathrm{Wm}^{-2}$ in the year 2100. The two scenarios in the main category of ScenarioMIP, the *Tier 1* experiments,
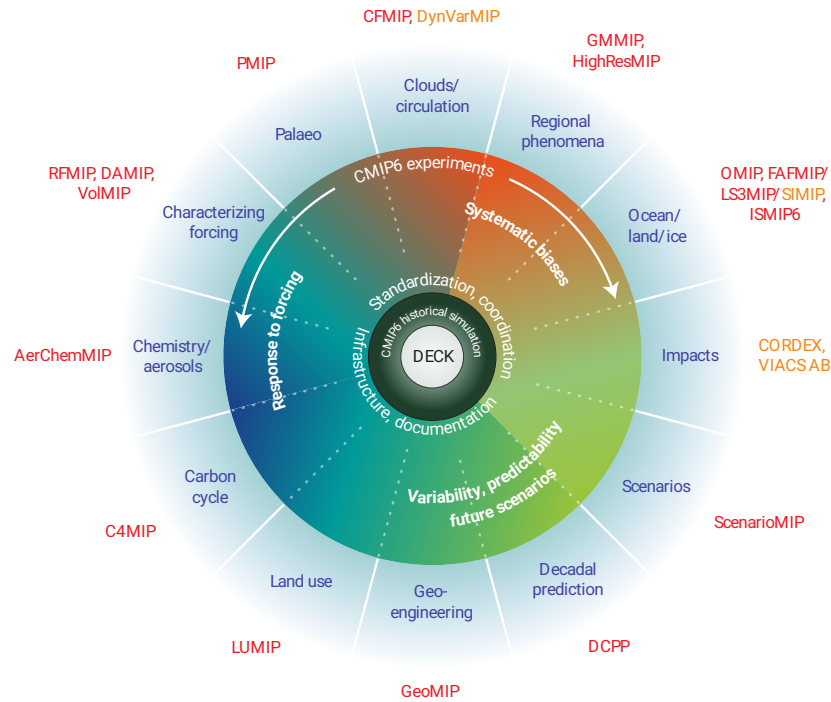
Figure 2.2.: Schematic of the experiment design of Phase 6 of the Coupled Model Intercomparison Project (CMIP6). The center of the circle illustrates the four DECK (Diagnostic, Evaluation, and Characterisation of Klima) experiments and the CMIP6 historical simulation. The circular sectors show additional science themes that can be explored through the 21 CMIP6-Endorsed Model Intercomparison Projects (MIPs). Adapted by permission from Springer Nature Customer Service Centre GmbH: Simpkins (2017).

are the SSP2-4.5 and SSP3-7.0 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 only 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 the impact of future climate change in a consistent and transparent way. Especially the ScenarioMIP experiments can give valuable insights into the upcoming development of the Earth system by providing *projections* of the future climate. 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 climate model projections we

Figure 2.3.: Schematic illustrating 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 (b) increase in newer generations of climate models when a new process is discovered to be relevant or (c) decrease with additional model improvements and observational constraints. Reproduced with permission from Cubasch et al. (2013) (their FAQ 1.1, figure 1).

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 21st 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 externally (and thus explainable) 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). As illustrated by figure 2.3, the uncertainty associated with natural variability is constant over time.

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 the 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 parametrizations of unresolved processes. The latter can be thought of as different approximations that are necessary to represent processes that take place on scales smaller than the common size of grid boxes in modern global climate models (about $1° \times 1°$ horizontally), which is for example the case for cloud processes and vegetation processes.

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 an 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 sciences 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 for 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_2 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 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

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.).

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 the evaluation of climate sensitivity. Figure 2.4 shows an overview of important feedbacks in the Earth system with their corresponding time scales on which they operate.

An example for 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). 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_2 emissions) by a typical factor between 2 and 3, rendering water vapor a fundamental agent of climate change (Myhre et al. 2013). An example for 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 for 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

Figure 2.4.: Climate feedbacks and corresponding time scales. "+" refers to positive feedbacks, which amplify the effect of the external forcing (e.g. the water vapor feedback). "-" refers to negative feedbacks, which diminish the effect of the external forcing (e.g. the longwave radiation feedback). "+/-" refers to feedbacks which might be either positive or negative (e.g. 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).

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 is involved. An example for this is the cloud feedback. Changes in clouds induced by climate change can cause both longwave (greenhouse warming) and shortwave (reflective cooling) effects, 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 altitude and the geographical distribution of clouds. Examples for robust cloud feedback processes are the rise of high 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_2 fluxes into the land and ocean carbon reservoirs due to increased photosynthesis rates and CO_2 dissolution in the sea, respectively, which decrease the atmospheric CO_2 content and diminish global warming. An example for a positive contribution is the decreased solubility of CO_2 in water in a warmer climate, which reduces the atmosphere-ocean CO_2 flux and enhances climate change. More details on this are given in section 2.3.2.

2.2.2. Mathematical Framework for Feedbacks Analysis

The foundation of a basic mathematical framework for climate feedback analysis is 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 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

$$N = F + R \tag{2.1}$$

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 basically 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

$$J_0 = -\sigma T_0^4. \tag{2.2}$$

$\sigma \approx 5.67\,Wm^{-2}K^{-4}$ 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 ΔT:

$$J_0 + R = -\sigma \left(T_0 + \Delta T\right)^4. \tag{2.3}$$

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:

$$-\sigma \left(T_0 + \Delta T\right)^4 \approx J_0 - 4\sigma T_0^3 \cdot \Delta T. \tag{2.4}$$

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

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

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 ΔT (Gregory et al. 2004). However, the value of this linear constant λ, 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 ΔT (Gregory and Webb 2008), equation (2.5) can be adjusted to

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

λ_{WV} refers to the water vapor feedback, λ_{Albedo} to the snow/ice albedo feedback and λ_{Cloud} to the cloud feedback. Thus, the overall climate feedback parameter λ can be written as the sum of the individual feedback parameters λ_i:

$$\lambda = \sum_i \lambda_i. \tag{2.7}$$

Positive values of λ_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

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_2 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 of is not useful. Running fully-coupled ESMs into equilibrium is computationally expensive as it would require thousands of model years (M. Rugenstein 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_2 concentration (4xCO$_2$) (Gregory
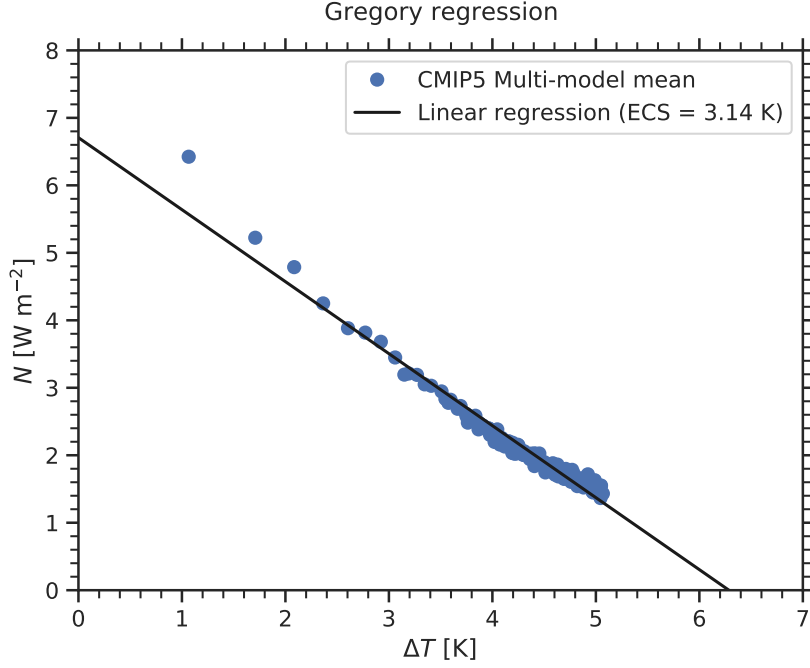
Figure 2.5.: Global and annual mean net top of the atmosphere (TOA) radiation N versus the change in global and annual mean near-surface air temperature ΔT for 150 years of a simulation with an abrupt quadrupling of the atmospheric CO_2 concentration (4xCO_2) for the CMIP5 multi-model mean (blue circles). To account for energy leakage and model drift, a linear fit of the corresponding pre-industrial control run is subtracted from the 4xCO_2 simulation. As given by equation (2.9), the slope of the linear regression (black line) corresponds to the climate feedback parameter λ 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.

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

$$\Delta T = -\frac{F}{\lambda}. \tag{2.8}$$

Thus, the change in GSAT in radiative equilibrium can be easily calculated from the external forcing F and the climate feedback parameter λ. The steady state values for F and λ can be estimated from a 4xCO_2 simulation that is not in equilibrium by extrapolation with a linear regression (Gregory et al. 2004). For this method, the so-called *Gregory regression*, the global and annual mean net TOA radiation N versus the change in the annual mean GSAT ΔT for all 150 years of the 4xCO_2 run are plotted (see figure 2.5). To account for energy leakage and remove any model drift present in the control climate, a linear fit of the corresponding pre-industrial control run is subtracted from the 4xCO_2 simulation beforehand (Andrews et al. 2012). Since the combination of equations (2.1) and (2.6) yields

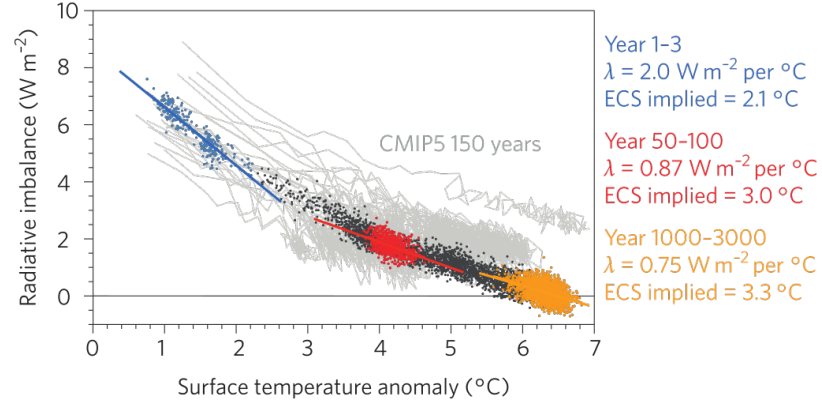$$N = F + \lambda \cdot \Delta T, \tag{2.9}$$

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 λ. 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).

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

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

The factor of 2 in the denominator accounts for the fact the traditional Equilibrium Climate Sensitivity is defined for an abrupt 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 M. A. A. 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 4xCO2 simulation with more than 1000 model years is considered. As a result, the climate feedback parameter λ 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 4xCO2 instead of a 2xCO2 simulation. The factor of 2 in the denominator of equation (2.10) only partly compensates this difference since the radiative forcing logarithmically depends on the atmospheric 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) (M. 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 λ 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_{SWCRE} and longwave N_{LWCRE} components separately, but also for the combined effect $N_{CRE} = N_{SWCRE} + N_{LWCRE}$. The slopes in the corresponding Gregory regressions are the so-called *CRE feedback parameters* λ_{SWCRE}, λ_{LWCRE} and λ_{CRE}, which quantify the change in CRE as a response to increasing GSATs.

2.2.5. Transient Climate Response

A further metric describing the 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_2 doubling in a simulation with a 1 % per year increase of the atmospheric CO_2 concentration (1%CO_2). For this, the annual mean GSATs are averaged over a 20 year period centered at the time of the CO_2 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_2 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_2 simulation, this ocean heat uptake can be approximated with

$$N = \kappa \cdot \Delta T, \tag{2.11}$$

where κ 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, κ is positive. This approximation becomes less accurate as the deeper ocean warms up and cannot be applied

Estimation of TCR

Figure 2.7.: Global and annual mean near-surface air temperature change ΔT for 140 years of a simulation with a 1 % per year increase of the atmospheric CO_2 concentration (1%CO_2) for the Community Earth System Model (CESM) of the National Center for Atmospheric Research (NCAR) (blue circles). 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 ΔT at the time of CO_2 doubling averaged over a 20 year period (illustrated by the vertical dashed lines). Here, TCR = 2.35 K (horizontal dashed line).

to simulations with a steady state climate change ($N \rightarrow 0$) where the external forcing tends stabilize (Gregory et al. 2009). By applying the definition of TCR (transient GSAT change ΔT at the time of CO_2 doubling) and combining equations (2.9) and (2.11), TCR can be estimated as

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

where F_{2x} is the radiative forcing induced by a doubling of the atmospheric CO_2 concentration.

This equation can be used to derive a relationship between TCR and ECS. Writing ECS = $-F_{2x}/\lambda$ and assuming path independence of the forcing (i.e. the resulting radiative forcing from the 1%CO_2 and 2xCO_2 runs are comparable) gives

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

which demonstrates the non-linear connection between TCR and ECS (Gregory and Forster 2008; Nijsse 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

Since one study presented in this thesis aims to reduce uncertainties in carbon cycle-related processes (Schlund et al. 2020a; see chapter 6), this chapter introduces the scientific background of the global carbon cycle.

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 $GtC\,yr^{-1}$ for the reservoirs 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 10000 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_2, chemical weathering, erosion and sediment formation on the sea floor. These natural exchange fluxes between the slow and the fast domain are comparatively small ($< 0.3\,GtC\,yr^{-1}$) 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_2 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_4), 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 38000 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_2 is removed from the atmosphere by plant photosynthesis with about $120\,GtC\,yr^{-1}$ (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_2 uptake

17
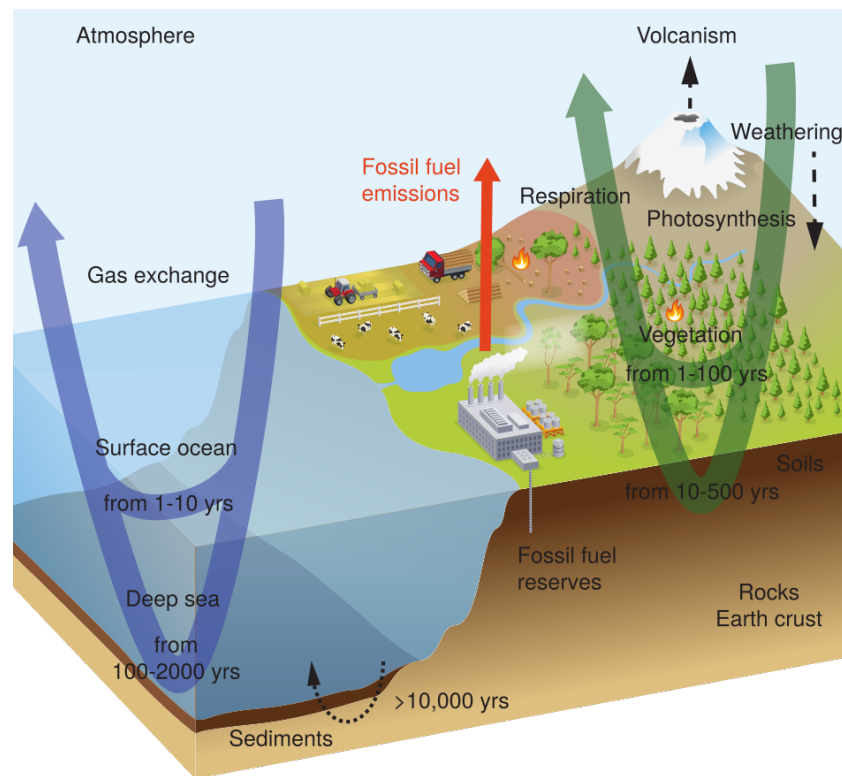
Figure 2.8.: Simplified schematic of the global carbon cycle including the typical turnover time scales for carbon transfers through the major reservoirs (atmosphere, land surface and ocean). Reproduced with permission from Ciais et al. (2013) (their FAQ 6.2, figure 1).

by photosynthesis occurs only during the growing season, whereas respiration occurs nearly all year, the larger amount of vegetation in the Northern Hemisphere (due to the larger land mass) gives rise to a seasonal cycle of the atmospheric CO_2 concentration (Keeling et al. 1995). This seasonal cycle reflects the phase of the global carbon cycle and shows a maximum of the atmospheric CO_2 concentration in the Northern Hemisphere winter (net CO_2 flux into atmosphere due to respiration) and a minimum during the Northern Hemisphere summer (net CO_2 flux into the land due to photosynthesis). Another major carbon exchange process connects the atmosphere and the ocean. Atmospheric CO_2 is exchanged with the surface ocean through gas exchange, which is driven by the partial CO_2 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_2 concentration over the last several thousand years before the Industrial Revolution in the 19th century (Ciais et al. 2013). Since the beginning of the Industrial Era, humanity is constantly emitting carbon-based GHGs (e.g. CO_2 and CH_4) into the atmosphere. Especially the atmospheric CO_2 concentration has

Figure 2.9.: The Keeling Curve: monthly-mean atmospheric CO_2 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_2 concentration is superimposed with a seasonal oscillation caused by the seasonal CO_2 cycle (see section 2.3.1).

substantially increased, which has already been shown by Charles D. Keeling in 1976 by his continuous CO_2 measurements at Mauna Loa, Hawaii that started in 1958 (Keeling et al. 1976; see figure 2.9). From 1958, the atmospheric CO_2 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_2 cycle, which gives rise to local maxima of the atmospheric CO_2 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_2 measurements by being far away from big population centers. Moreover, its elevation of more than 3000 m provides access to the well-mixed air of the Pacific Ocean in high altitudes, 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_2 directly influences the carbon exchange fluxes of the global carbon cycle. Due to the excessive 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 steady state anymore. In the decade 2010–2019, anthropogenic activities caused net carbon fluxes of $3.4\,\mathrm{GtC\,yr^{-1}}$ from the atmosphere into the terrestrial biosphere due to increased plant photosynthesis and $2.5\,\mathrm{GtC\,yr^{-1}}$ from the atmosphere into the ocean due to increase dissolution of CO_2 into the sea (Friedlingstein et al. 2020). In the same time, the amount of carbon in the atmosphere reservoir increased with a rate of $5.1\,\mathrm{GtC\,yr^{-1}}$, indicating that only about half of the
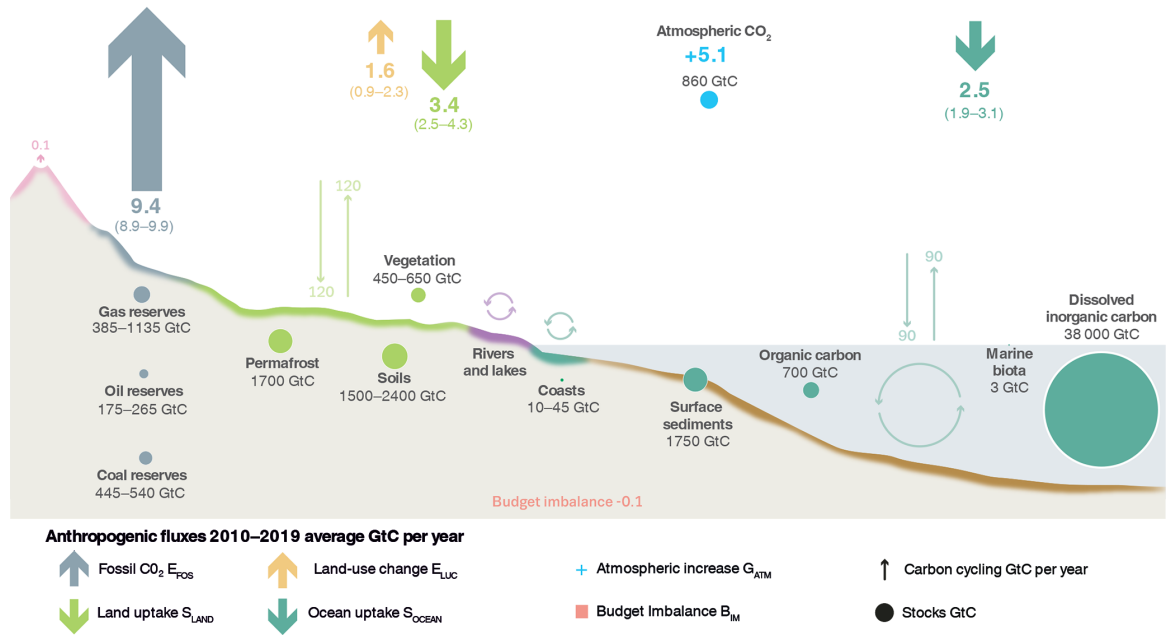
Figure 2.10.: Schematic representation of the overall perturbation of the global carbon cycle caused by anthropogenic activities, averaged globally for the decade 2010–2019. Arrows represent carbon exchange fluxes; circles carbon reservoirs. More details are given in the legend of this figure. Adapted with permission from Friedlingstein et al. (2020).

anthropogenic CO_2 emissions in the last decade remained in the atmosphere (Friedlingstein et al. 2020) where they can act as GHG.

Thus, this removal of CO_2 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_2 fertilization effect* (Walker et al. 2020), that causes an increase of photosynthesis rates when the atmospheric CO_2 concentration increases, which in turns removes CO_2 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_2 concentration causes an increased dissolution of CO_2 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_2 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_2 in the sea, which reduces the ocean carbon uptake and enhances climate change (Gregory et al. 2009).

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 our imperfect knowledge on how the climate system will respond to external forcing. In a multi-model ensemble (e.g. from CMIP), this is expressed as 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) and the multi-model range as best estimate and uncertainty measure of this quantity (M. Collins et al. 2013). This *model democracy* approach basically assumes that all climate models are independent, equally plausible, distributed around reality and that the projected multi-model range is representative for 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 2.11. 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 2.11) is a robust and physically interpretable inter-model relationship between a
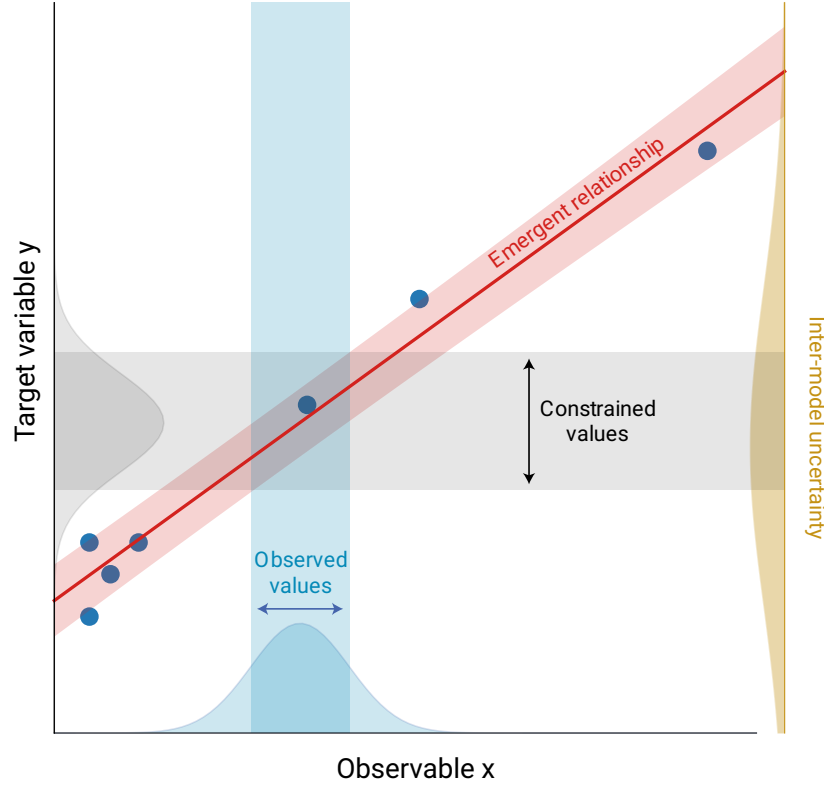
Figure 2.11.: Schematic illustration of the emergent constraint approach. Basis of every emergent constraint is a robust and physically interpretable emergent relationship (red line) between a target variable y (e.g. Earth system sensitivity or projection of future climate change) and an observable x (e.g. past or present-day trend or variation) for the climate models of a multi-model ensemble (blue circles). 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).

target variable y related to the future climate and an observable x of the past or present-day climate. Basis for the relationship is output of the different climate models of a multi-model ensemble (blue circles in figure 2.11). Using an observation of x, this emergent relationship can then be used to derive a emergent constraint on y (gray shaded area in figure 2.11) that considers uncertainties in the emergent relationship itself (red shaded area in figure 2.11) and uncertainties in the observation (blue shaded area in figure 2.11).

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 $\left\{ (x_m, y_m) \mid m \in I_M \right\}$ with index set $I_M = \{1, 2, \dots, M\}$, a linear regression model

$$\hat{y}(x) = \hat{b}_0 + \hat{b}_1 x \tag{2.14}$$

for the predicted target variable $\hat{y}$ with estimated intercept $\hat{b}_0$ and slope $\hat{b}_1$ is used (see equation (2.25) for details on this). Fitting this regression line with ordinary least squares includes minimizing the standard error s of the estimate

$$s^2 = \frac{1}{M-2} \sum_{m=1}^{M} \left(y_m - \hat{y}_m\right)^2 , \tag{2.15}$$

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 $\sigma_{\hat{y}}(x)$:

$$\sigma_{\hat{y}}^2(x) = s^2 \left[1 + \frac{1}{M} + \frac{(x - \bar{x})^2}{\sum_{m=1}^{M} (x_m - \bar{x})^2}\right]. \tag{2.16}$$

Here, $\bar{x}$ indicates the arithmetic mean of x 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 (2.16) can be used to define a conditional probability density function (PDF) for predicting a target variable of y given x:

$$P(y|x) = \frac{1}{\sqrt{2\pi\sigma_{\hat{y}}^2(x)}} \exp\left[-\frac{\left(y - \hat{y}(x)\right)^2}{2\sigma_{\hat{y}}^2(x)}\right]. \tag{2.17}$$

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 2.11). Its maximum is given by the emergent relationship itself (red line in figure 2.11). 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 are in figure 2.11). Assuming again a Gaussian distribution, the observational PDF for observing x_0 given the true value x can be written as

$$P(x_0|x) = \frac{1}{\sqrt{2\pi\sigma_x^2}} \exp\left[-\frac{(x_0 - x)^2}{2\sigma_x^2}\right], \tag{2.18}$$

where σ_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 2.11) with numerical integration:

$$P(y|x_0) = \int_{-\infty}^{+\infty} P(y|x)\, P(x|x_0)\, dx. \tag{2.19}$$

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 (Nijsse 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

$$t = \frac{r\sqrt{M-2}}{\sqrt{1-r^2}} \qquad (2.20)$$

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 higher 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 in the beginning of section 2.4, this is not the case for typical climate models 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. This is not always appropriate, for example when a non-linear emergent relationship is expected (Nijsse 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 σ_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 are 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 eleven 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-albedo feedback (Hall and Qu 2006). Over the years, the 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 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; Nijsse 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. 2019). 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.

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

A further technique 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

$$y = \sum_{m=1}^{M} w_m y_m \qquad (2.21)$$

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; see beginning of section 2.4), Knutti et al. (2017b) propose a weighting scheme based on climate model performance and interdependence with the following weights:

$$w_m \propto \frac{\exp\left(-\frac{D_m^2}{\sigma_D^2}\right)}{1 + \sum_{n \neq m}^{M} \exp\left(-\frac{S_{mn}^2}{\sigma_S^2}\right)}. \qquad (2.22)$$

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). σ_D and σ_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 σ_D and σ_S determine how strongly the climate models' performance and interdependence are weighted (Knutti et al. 2017b). Small values of the performance parameter σ_D lead to an aggressive weighting with only a few climate models receiving a majority of the weight, while large values of σ_D result in an equal weighting. For the interdependence parameter σ_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 σ_D and σ_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. The whole process is repeated for every climate model of the ensemble to get a statistical distribution of RMSEs. Finally, different RMSE distributions calculated from different parameters σ_D and σ_S can be assessed using specific criteria to find optimal values for σ_D and σ_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 (2.22) 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 σ_D and σ_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 21st 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 $\{x^{(1)}, x^{(2)}, \ldots, x^{(K)}\}$ which are relevant for the projection of the target variable y. The reasoning for this choice of the 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 $y = (y_1, y_2, \ldots, y_M)^T \in \mathbb{R}^M$ be the vector of target variables of the M climate models (T denotes the transpose) and $X \in \mathbb{R}^{M \times (K+1)}$ the design matrix representing the predictors:

$$X = \begin{pmatrix} 1 & x_1^{(1)} & x_1^{(2)} & \cdots & x_1^{(K)} \\ 1 & x_2^{(1)} & x_2^{(2)} & \cdots & x_2^{(K)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_M^{(1)} & x_M^{(2)} & \cdots & x_M^{(K)} \end{pmatrix}. \tag{2.23}$$

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

$$y = Xb + \varepsilon, \tag{2.24}$$

where $b = (b_0, b_1, \ldots, b_K)^T \in \mathbb{R}^{(K+1)}$ is the vector of linear coefficients (with intercept b_0) and $\varepsilon = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_M)^T \in \mathbb{R}^M$ a vector of independent random variables representing the noise in the target variable. Figure 2.12 shows a schematic that illustrates this linear relationship (gray surface) for the different climate models (colored circles) for two diagnostics ($K = 2$). Using ordinary least squares regression, the estimated linear coefficients $\hat{b}$ are given by

$$\hat{b} = \left(X^T X\right)^{-1} X^T y, \tag{2.25}$$

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 (2.14)).
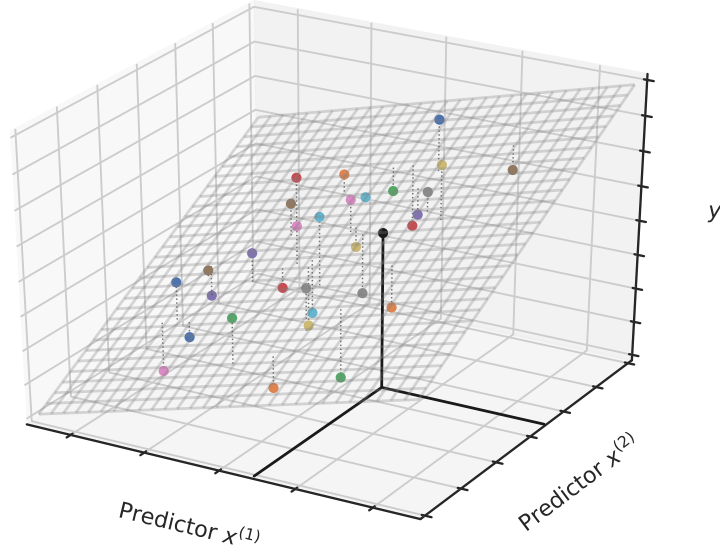
Figure 2.12.: Schematic illustration of the Multiple Diagnostic Ensemble Regression (MDER) approach (Karpechko et al. 2013). First, inter-model relationships between a target variable y and multiple process-based predictors $x^{(k)}$ (here: two predictors $x^{(1)}$ and $x^{(2)}$) are used to fit a multivariate 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 (2.26). The black circle 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 circles 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 and represent the noise terms ε in equation (2.24).

In the second step of the algorithm, observed data of the process-based diagnostics $x_0 = \left(1, x_0^{(1)}, x_0^{(2)}, \ldots, x_0^{(K)}\right)^T \in \mathbb{R}^{(K+1)}$ 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):

$$\hat{y}_0 = x_0^T \hat{b} \tag{2.26}$$

This is mathematically similar to the calculation of the best estimate target variable y for emergent constraints. In figure 2.12, the observations of the predictors x_0 are illustrated with horizontal black lines and the best estimate $\hat{y}_0$ is shown as a black circle. By combining equations (2.25) and (2.26) and comparing this to the definition of weighted means in equation (2.21), climate model weights $w = (w_1, w_2, \ldots, w_M)^T \in \mathbb{R}^M$ can be defined by

$$w = \left[x_0^T \left(X^T X \right)^{-1} X^T \right]^T, \tag{2.27}$$

which can be used to calculate the weighted target variable by

$$\hat{y}_0 = \hat{w}^T y. \tag{2.28}$$

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 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 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

In order to answer 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 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 were 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 lead 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 (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 is illus-
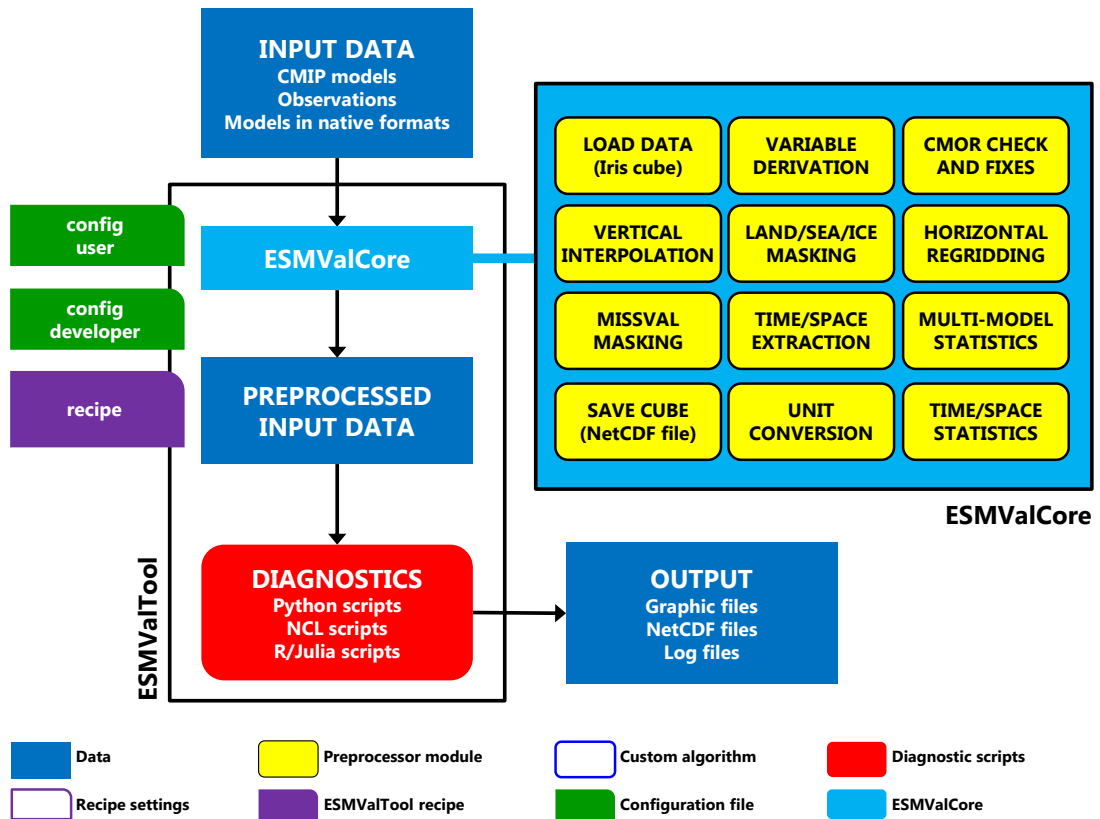
Figure 3.1.: Schematic representation of the Earth System Model Evaluation Tool (ESMValTool). First, input data is preprocessed by the core functionalities of the ESMValTool (ESMValCore). Second, the preprocessed data is read by diagnostic scripts, which create the final output (e.g. plots). The user 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).

trated 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 ones. Examples for non-optional preprocessors are the Input/Output (I/O) 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 *interpolation*, land/sea/ice

masking, *unit conversion*, or the calculation of *multi-model statistics*. A further example is the *variable derivation*, which can be used to derive non-CMOR variables from the input data. The aim of the preprocessor is 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 a 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 for example be used to assess the skill of climate model simulations. More details on the structure of the ESMValTool are given in its extensive documentation (https://esmvaltool.readthedocs.io).

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 for example 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 the recipe and diagnostic scripts that can for example be used to reproduce scientific assessments, is also available at Zenodo (Andela et al. 2020a).

3.2. Contributions to ESMValCore

As of January 2021, 40408 lines of code have been added and 21050 lines of code have been removed from to 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 calculation of a variable's diurnal or seasonal cycle amplitude. 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 a desired region. 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 be used to calculate the temporal trend of the GSAT over the 20$^{\text{th}}$ century.

Apart from new preprocessor functions, six derivation scripts for the non-CMOR variables *asr*, *co2s*, *et*, *rlntcs*, *rsntcs* and *uajet* have been added. The absorbed shortwave radiation (*asr*) is defined as the difference in the incoming TOA shortwave radiation and the outgoing TOA shortwave radiation. The atmospheric CO_2 concentration at surface (*co2s*) can be calculated

Type	Name	Description
Preprocessor	*amplitude*	Amplitude of periodic phenomena (e.g. cycles)
	land/sea fraction weighting	Weighting of fields based on the land or sea fraction of the respective grid cells
	trend	Linear trend (slope of linear regression)
Derived variable	*asr*	Absorbed shortwave radiation
	co2s	Atmospheric CO_2 concentration at surface
	et	Evapotranspiration
	rlntcs	Clear-sky net top of the atmosphere (TOA) longwave radiation
	rsntcs	Clear-sky net TOA shortwave radiation
	uajet	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.
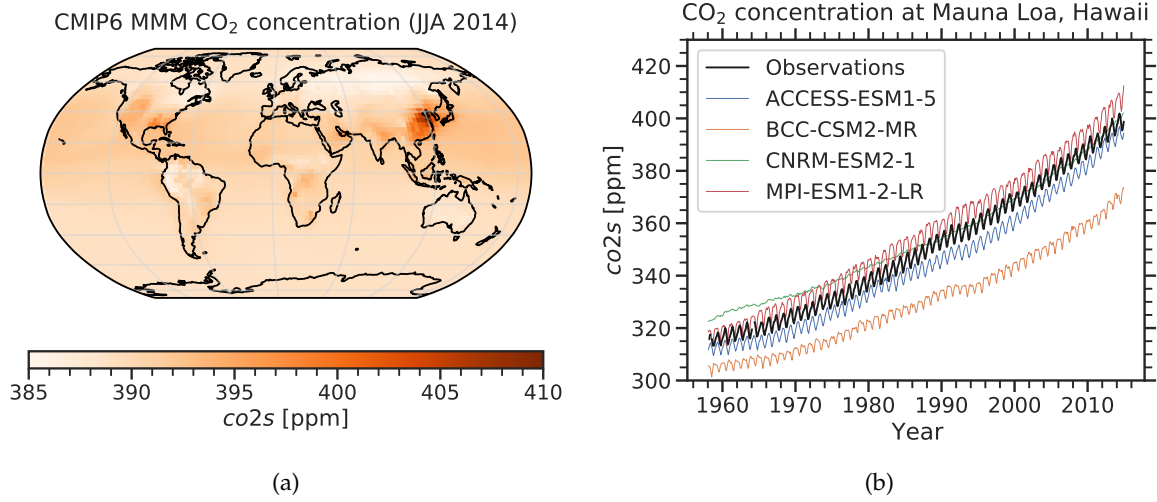
CMIP6 MMM CO$_2$ concentration (JJA 2014)

co2s [ppm]

CO$_2$ concentration at Mauna Loa, Hawaii

(a)   (b)

Figure 3.2.: (a) CMIP6 multi-model mean (MMM) of the atmospheric CO$_2$ concentration at 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.

from the pressure level-dependent atmospheric CO$_2$ 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$_2$ concentrations over large metropolitan areas with high CO$_2$ 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$_2$ 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$_2$ 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 top of the atmosphere (TOA) radiations (*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 88668 lines of code and removed 47816 lines of code from the ESMValTool repository (as of January 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 January 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* re-

Type	Name	Main reference
Recipe	*bock20jgr* (figures 8 and 10)	(Bock et al. 2020)
	cox18nature	(Cox et al. 2018)
	ecs	(Gregory et al. 2004)
	flato13ipcc (figure 9.42)	(Flato et al. 2013)
	meehl20sciadv	(Meehl et al. 2020)
	schlund20esd	(Schlund et al. 2020b)
	schlund20jgr_gpp_abs_rcp85	(Schlund et al. 2020a)
	schlund20jgr_gpp_change_1pct	(Schlund et al. 2020a)
	schlund20jgr_gpp_change_rcp85	(Schlund et al. 2020a)
	tcr	(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)
	MTE	(Jung et al. 2011)
	NDP	(Gibbs 2006)
	Scripps-CO$_2$	(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*.
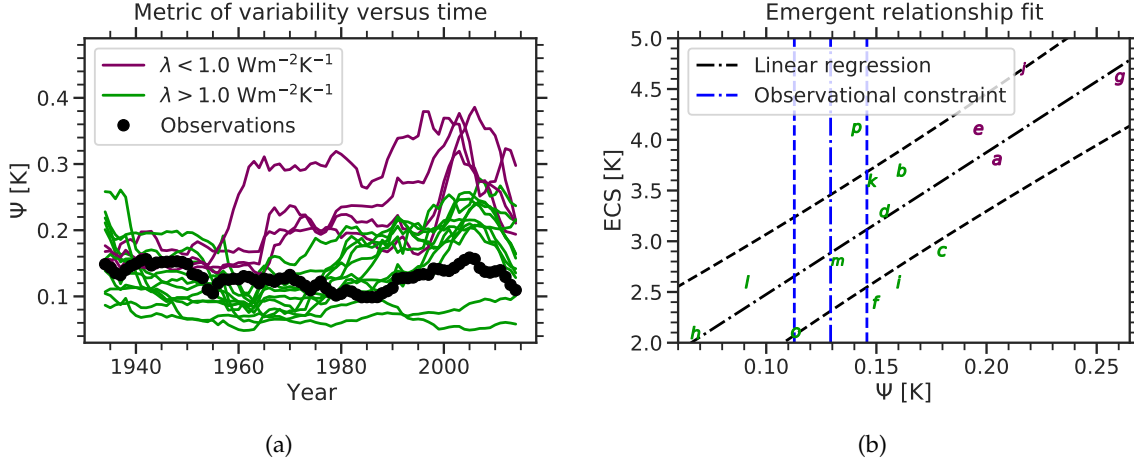
Figure 3.3.: Emergent constraint on the Effective Climate Sensitivity (ECS) from global temperature variability (Cox et al. 2018). Magenta colors indicate high sensitivity climate models ($\lambda < 1\,\mathrm{Wm^{-2}K^{-1}}$) and green colors indicate low sensitivity climate models ($\lambda > 1\,\mathrm{Wm^{-2}K^{-1}}$). Both panels were created with the ESMValTool using *recipe_cox18nature.yml* and are similar to Cox et al. (2018). (a) Temperature variability metric Ψ versus time for historical simulations from CMIP5 models (colored lines) and observations from the HadCRUT4 dataset (black circles) (Morice et al. 2012). (b) Emergent relationship between ECS and the temperature variability metric Ψ. The black dot-dashed line shows the linear regression across the climate model ensemble with its associated uncertainties indicated by the black dashed lines (standard prediction error; see equation (2.16)). 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).

produces 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.

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 are able to calculate the climate metrics ECS and TCR for arbitrary model output (see section 2.2). 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 Ψ. Figure 3.3a shows Ψ over the past 85 years for CMIP5 models and observations form HadCRUT4 (Morice et al. 2012). Since high sensitivity climate models exhibit higher values of Ψ than low sensitivity models, an emergent
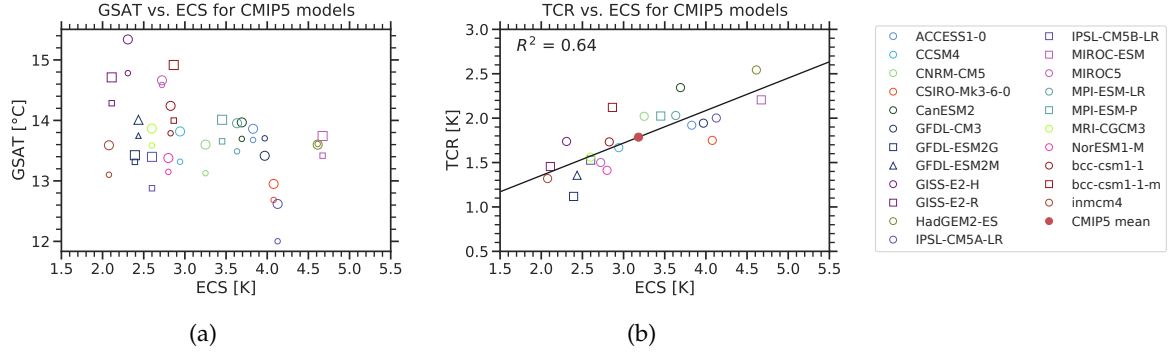
Figure 3.4.: Relationships between climate metrics within the CMIP5 climate model ensemble. Both panels were created with the ESMValTool using *recipe_flato13ipcc.yml* and are similar to Flato et al. (2013). (a) Global mean near-surface air temperature (GSAT) versus the Effective Climate Sensitivity (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.

relationship between ECS and Ψ can be established (figure 3.3b). With observations of Ψ, ECS can eventually be constrained to (2.8 ± 0.6) K (66 % confidence range). More details on this including the calculation of Ψ are given in section 5.2.1.

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 two-panel figure 9.42 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 no necessarily exhibit a high equilibrium warming in the future or vice versa. On the other hand, TCR and ECS are in fact 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 implemented CMORization scripts (also referred to as *CMORizers*) allow the processing of arbitrary observational data with the ESMValTool. For this, the raw observational datasets are reformatted in such 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 eleven newly implemented CMORizers including the main reference for the corresponding observational datasets is shown in table 3.2.

4. Assessment of Climate Sensitivity in the CMIP6 Ensemble

In order 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 January 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 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), we use it 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_2 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 using only
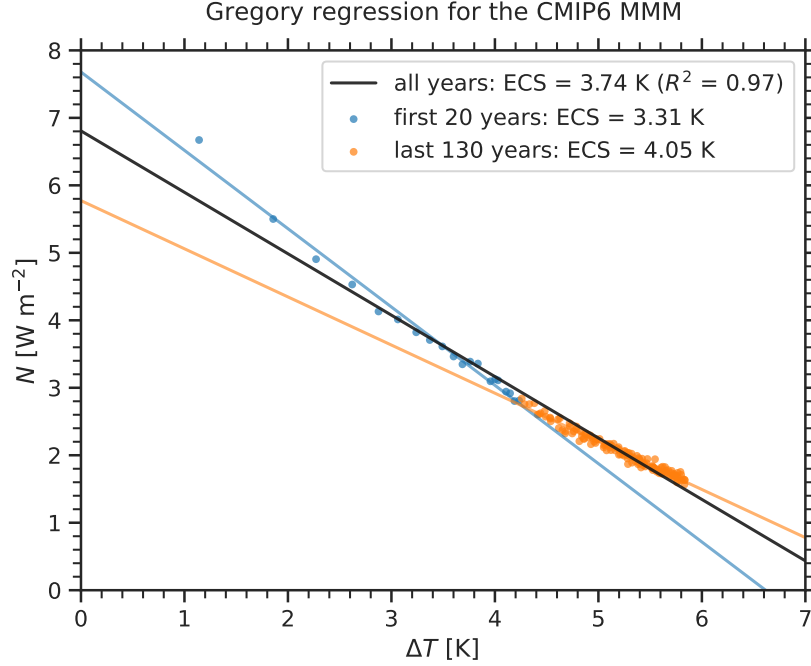
Figure 4.1.: Gregory regression for the CMIP6 multi-model mean (MMM) following Gregory et al. (2004): Global and annual mean net top of the atmosphere radiation N versus the change in global and annual mean near-surface air temperature ΔT for 150 years of a simulation with an abrupt quadrupling of the atmospheric CO_2 concentration (4xCO_2) for the CMIP6 MMM (circles). To account for energy leakage and model drift, a linear fit of the corresponding pre-industrial control run is subtracted from the 4xCO_2 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 λ (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 circles and line) yields a considerably smaller ECS than using only the last 130 years of the simulation (orange circles and line). Adapted with permission from Meehl et al. (2020).

the last 130 years of the simulation, which gives an ECS of 4.05 K. As thoroughly describes 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_2 doubling in a 1%CO_2 run (Bindoff et al. 2013; see section 2.2.5). Using these two definitions and the ESMValTool, we assess ECS and TCR for all CMIP5 and CMIP6 models where the necessary temperature and radiation data is available (as of January 2021). Tables 4.1 and 4.2 give an overview over 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) evaluated for the CMIP5 models. Details on the calculation of ECS and TCR 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 evaluated over all climate models (using the normalization $1/M$, where M is the number of models). Adapted with permission from Meehl et al. (2020).

smaller with about 10 % (0.19 K). The spread in the multi-model ensembles (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 with very high values of ECS and TCR. In addition to that, there also exists a number of

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. Adapted with permission from Meehl et al. (2020).

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 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 a 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 insignificant 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 (as of January 2021). 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 exceeds this upper bound: out of the 42 analyzed CMIP6 models, a third (14 models) has an ECS above 4.5 K. In addition to our analysis, several modeling centers independently confirmed high ECS values in their models (Andrews 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 (Bretherton et al. 1990; see single green circle 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 assessed ranges (Flato et al. 2013; Randall et al. 2007), the upper bound of the CMIP6 model

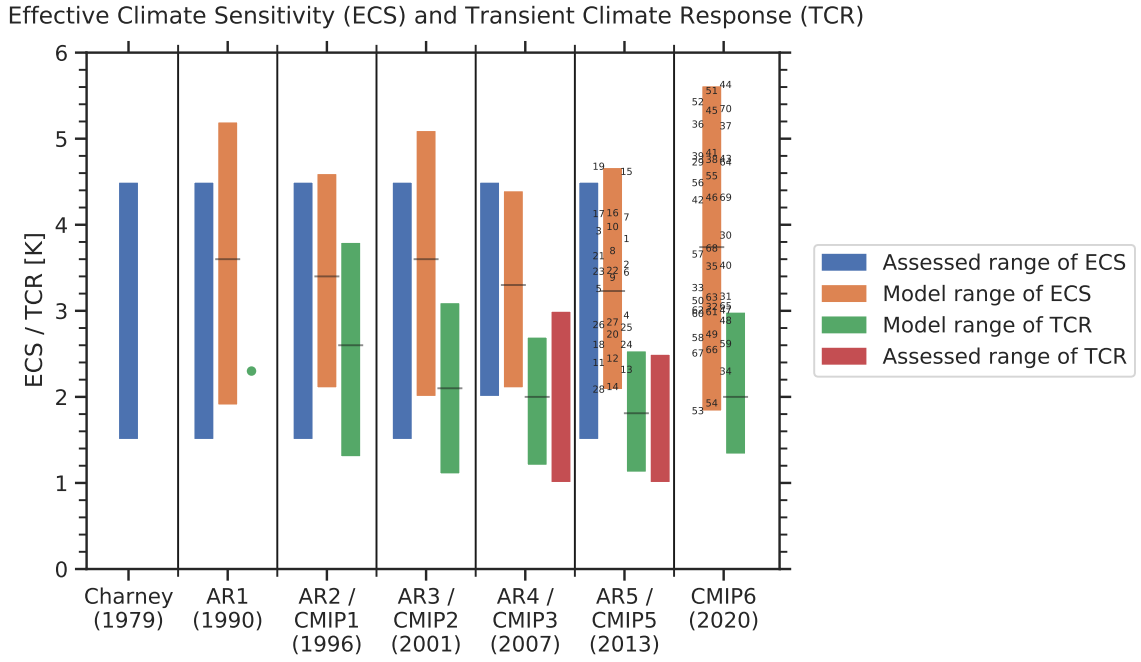Effective Climate Sensitivity (ECS) and Transient Climate Response (TCR)

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 circle represents 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).

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 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, chemical and biological processes have been introduced to the CMIP6 ESMs in order 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 in order 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 20th century (Dittus et al. 2020).

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 sensitivity. 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 a 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 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 (Bodas-Salcedo et al. 2016; McCoy et al. 2015, 2016; Tan et al. 2016). For this reason, the microphysical representation of mixed-phase clouds has been substantially improved in many CMIP6 models (Bodas-Salcedo et al. 2019; Gettelman et al. 2019a; Mulcahy et al. 2020; Williams et al. 2020), leading for example to a better agreement of the simulated TOA shortwave CRE with corresponding observations (Bock et al. 2020). As a consequence, 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 removed (or significantly reduced) in those CMIP6 models that simulate predominantly liquid clouds over the present-day Southern Ocean. This is illustrated in figure 4.3 (orange arrows). Moreover,
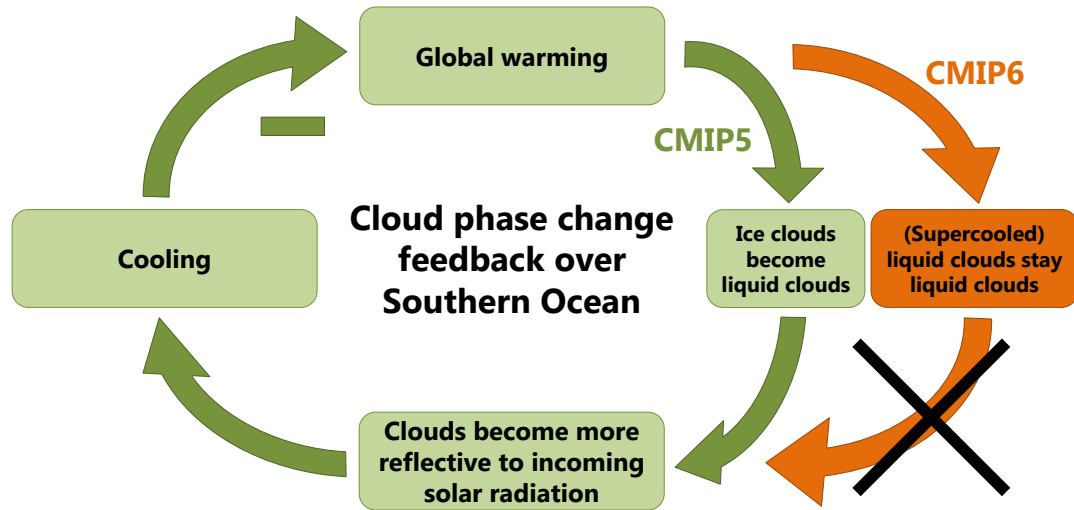
Figure 4.3.: Schematic illustration of the strong negative shortwave cloud phase change feedback over the Southern Ocean, which is present in earlier CMIP model generations like CMIP5 (green arrows). Due to substantial improvements in the microphysical representation of mixed-phase clouds in many CMIP6 models, the predominantly ice clouds over the Southern Ocean in the present-day climate of previous climate model generations have been replaced with clouds that predominantly consist of (supercooled) liquid clouds (Bodas-Salcedo et al. 2019; Gettelman et al. 2019a; Mulcahy et al. 2020; Williams et al. 2020). Thus, in these models a cloud phase change due to global warming is no longer possible, which removes (or significantly reduces) the strong negative cloud phase change feedback (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) in the corresponding CMIP6 models (Andrews et al. 2019; Gettelman et al. 2019a), which is a possible explanation for the high ECS values found in many CMIP6 models.

this reduction of the negative shortwave cloud phase change feedback 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 in some CMIP6 models.

To further confirm this hypothesis, we show the global distributions of relevant cloud feedback parameters (expressed by the net, shortwave and longwave CRE feedback parameters; see section 2.2.4) 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 4.4b) shows where the climate models are switching from simulating almost purely liquid clouds (north of 60 °S) to clouds with an increasing ice component (south of 60 °S) in their pre-industrial and present-day experiments. The increasing fraction of ice in the simulated mixed-phase clouds when moving south supports a negative shortwave cloud feedback on warming in these areas. 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
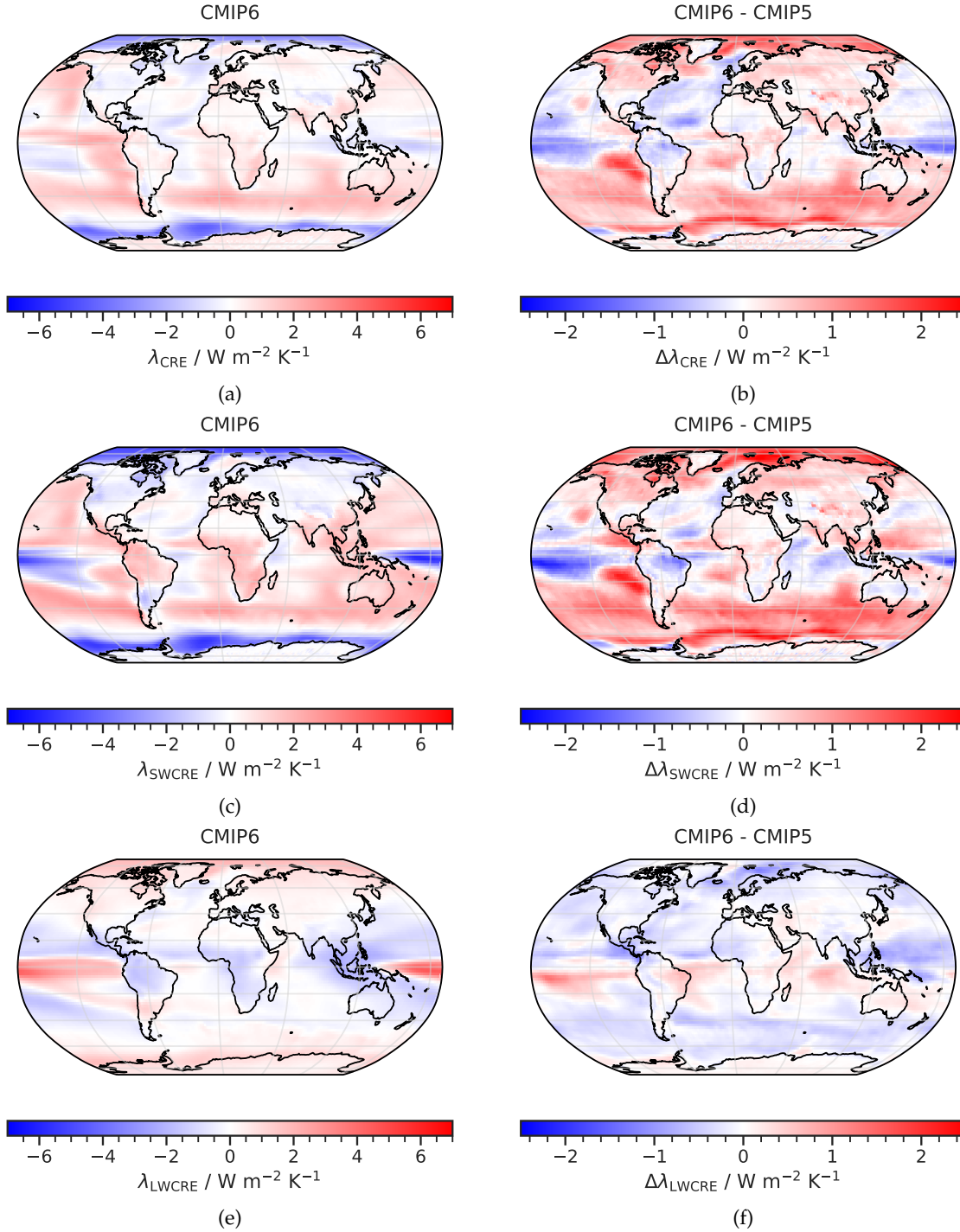
Figure 4.4.: Geographical distributions of the net (a), shortwave (c) and longwave (e) cloud radiative effect (CRE) feedback parameters for the CMIP6 multi-model mean (MMM) and the corresponding differences to the CMIP5 MMM, again for the net (b), shortwave (d) and longwave (f) components. The CRE feedback 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).

consistent to the hypothesis presented in the paragraph above, which claims that the reduced negative cloud phase change feedback over the Southern Ocean in 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 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 society and policymakers. An important question in this context could be the following: Is it reasonable to adapt the assessed ECS range from AR5 (1.5–4.5 K) in AR6 based on the CMIP6 model range of 1.8–5.6 K? An argument for this is that due to the many new and improved processes included to the models, the CMIP6 ensemble is arguably more realistic than its predecessor generations. However, even the CMIP6 models are far from fully simulating all aspects and processes of the full vastly complex Earth system. This 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 before answering the question raised above.

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 emergent constraints framework to assess the accuracy of climate model projections and reduce associated uncertainties. For ECS, a large variety of different emergent constraints has been developed over the years, which we now evaluate on the CMIP6 ensemble in order to assess whether they still hold for these new climate models. In total we assess eleven 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 on the CMIP3 or CMIP5 ensemble, the CMIP6 models offer an exciting possibility to perform out-of-sample testing using the previously unseen data.

This analysis is already published in Schlund et al. (2020b). For this paper, the author of this thesis lead the writing and the analysis and implemented the code to reproduce this analysis 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 eleven emergent constraints on ECS and evaluate them on the CMIP5 and CMIP6 ensemble (section 5.2). Finally, we provide a discussion (section 5.3) and a summary of the results (section 5.4).

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) [$\% \, K^{-1}$]	• *hur* (ERA-Interim) • *rsdt* (CERES-EBAF) • *rsut* (CERES-EBAF) • *rsutcs* (CERES-EBAF) • *ts* (HadISST)
COX	Cox et al. (2018)	Ψ (temperature variability metric) [K]	• *tas* (HadCRUT4)
LIP	Lipat et al. (2017)	Extent of Southern Hemisphere Hadley cell [°]	• *va* (ERA-Interim)
SHD	Sherwood et al. (2014)	D-index (large-scale lower-tropospheric mixing) [1]	• *wap* (ERA-Interim)
SHL	Sherwood et al. (2014)	LTMI (lower-tropospheric mixing index) [1]	• *hur* (ERA-Interim) • *ta* (ERA-Interim) • *wap* (ERA-Interim)
SHS	Sherwood et al. (2014)	S-index (small-scale lower-tropospheric mixing) [1]	• *hur* (ERA-Interim) • *ta* (ERA-Interim) • *wap* (ERA-Interim)
SU	Su et al. (2014)	Error in vertical profile of relative humidity [1]	• *hur* (AIRS, MLS-Aura)
TIH	Tian (2015)	Tropical mid-tropospheric humidity asymmetry index from AIRS [%]	• *hus* (AIRS)
TII	Tian (2015)	Southern ITCZ index from GPCP [$mm \, day^{-1}$]	• *pr* (GPCP)
VOL*	Volodin (2008)	Difference between tropical and southern midlatitudinal cloud fraction [%]	• *clt* (ISCCP D-2)
ZHA*	Zhai et al. (2015)	Seasonal response of marine boundary layer cloud fraction to changes in SST [$\% \, K^{-1}$]	• *cl* (CloudSat/CALIPSO) • *tos* (AMSRE SST) • *wap* (ERA-Interim)

Table 5.1.: Overview of the eleven 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).

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 in 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 eleven 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 were already 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 in order 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 (2.20)) and the constrained range of ECS based on the constrained PDF (see equation (2.19)).

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

5.2.1. COX

5.2.2. SHL

This is the lower-tropospheric mixing index (LTMI)!!

<div align="center">hi YOTHISISATEST!</div>

5.3. Discussion

5.4. Summary
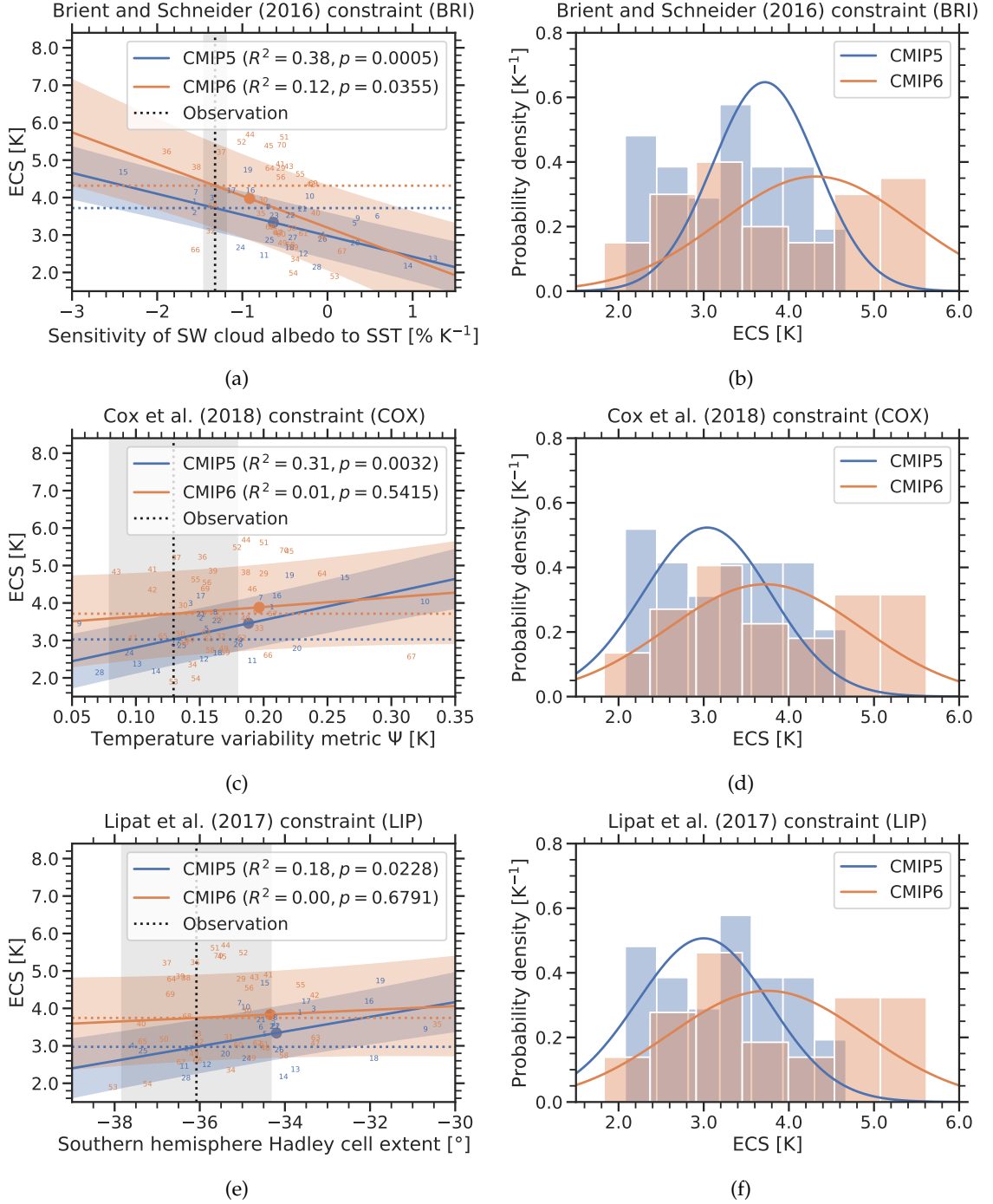
Figure 5.1.: Emergent constraints BRI, COX and LIP applied to the CMIP5 (blue) and CMIP6 ensemble (orange). (a), (c), (e) Emergent relationships (solid lines) for the CMIP models (numbers; see tables A.3 and A.4 for details) with their standard prediction errors (blue and orange shaded areas; see equation (2.16)). 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 (2.19)) 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).

(a)

(b)

(c)

(d)

(e)

(f)
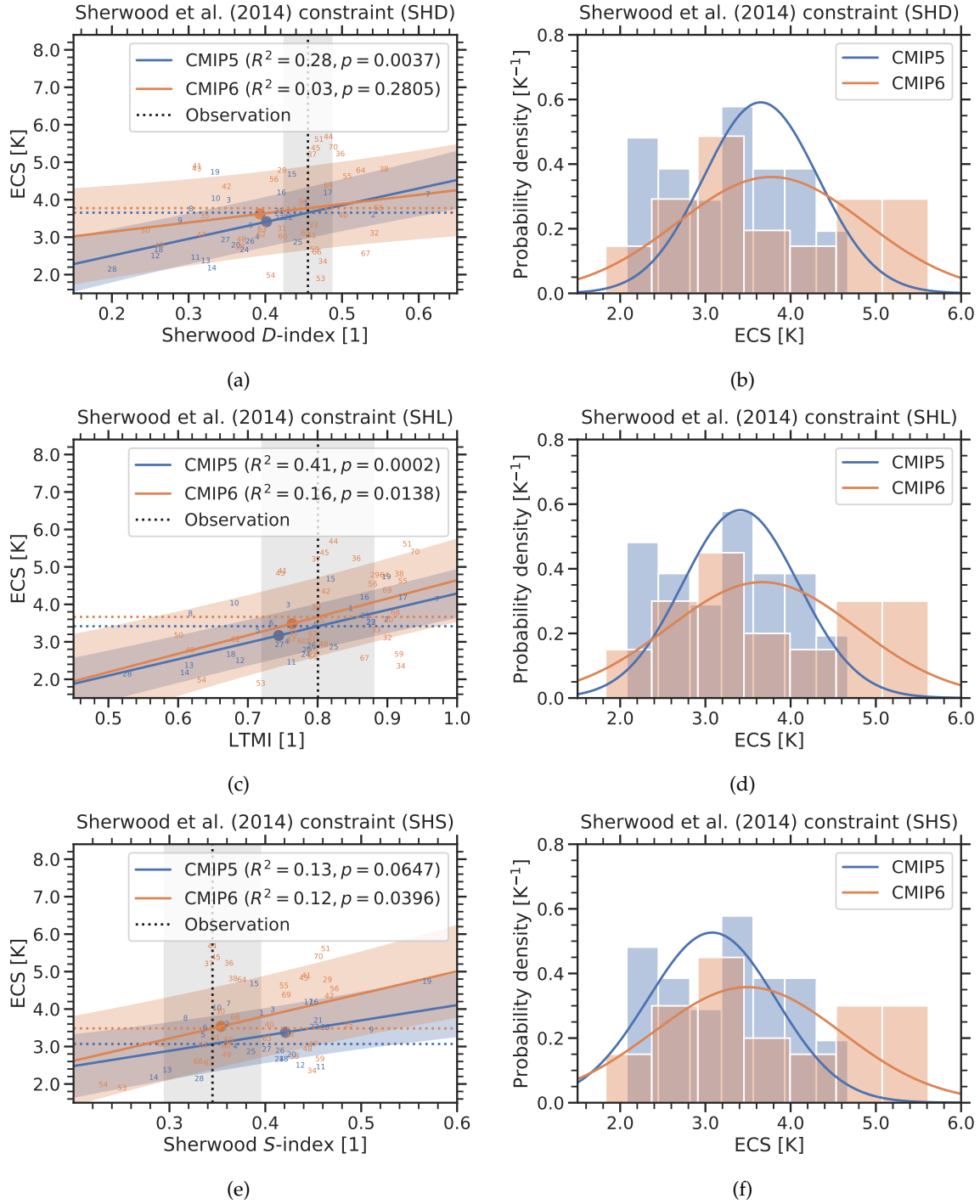
Figure 5.2.: As in figure 5.1 but for the emergent constraints SHD, SHL and SHS. Adapted with permission from Schlund et al. (2020b).
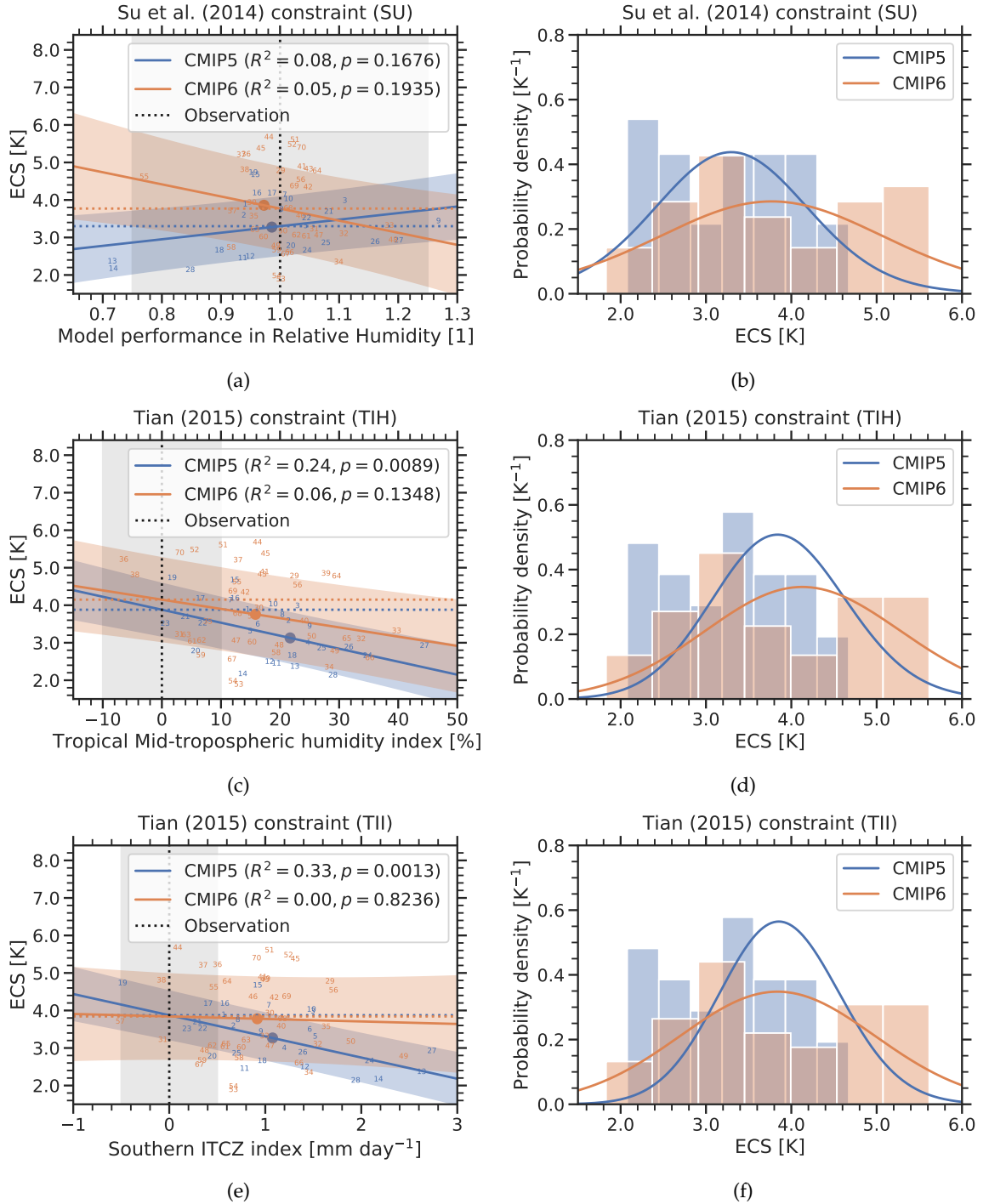
Figure 5.3.: As in figure 5.1 but for the emergent constraints SU, TIH and TII. Adapted with permission from Schlund et al. (2020b).
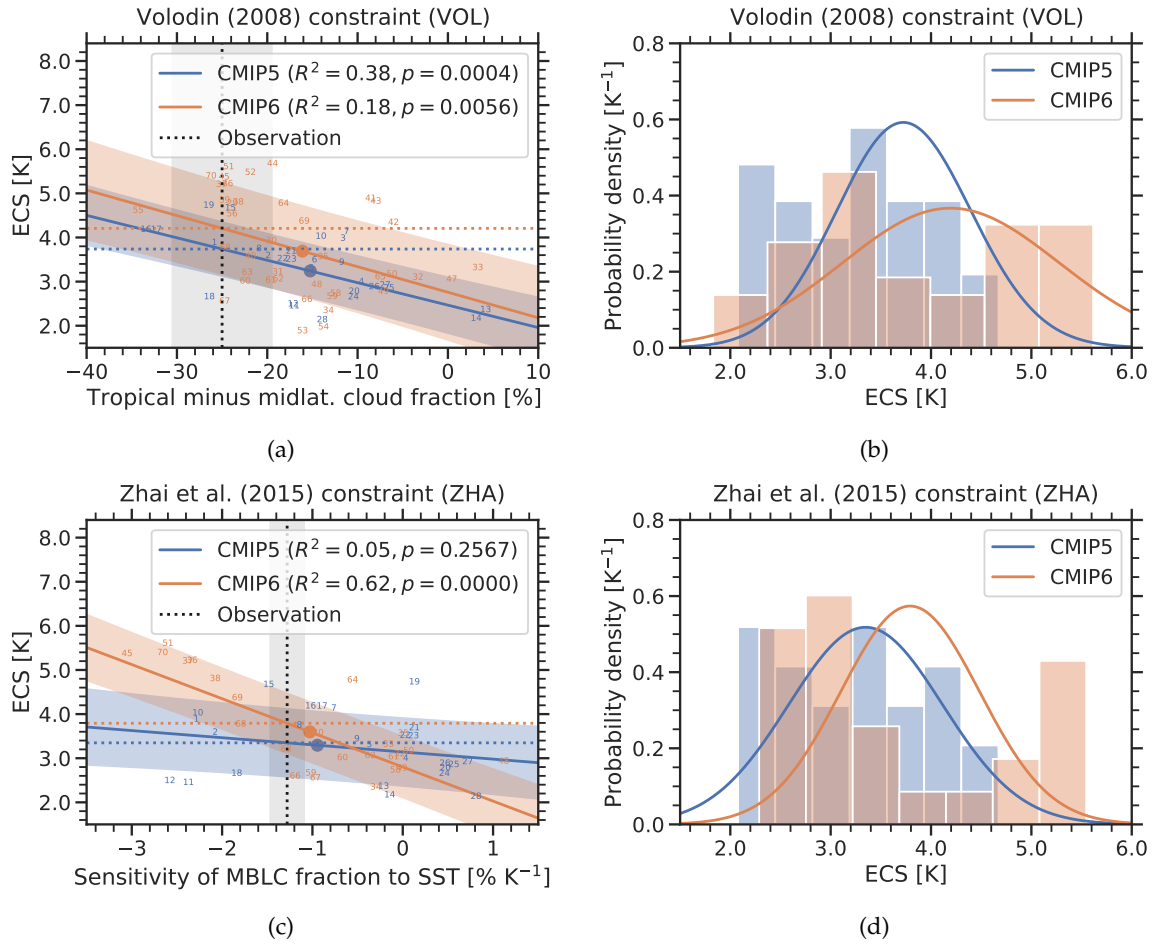
Figure 5.4.: As in figure 5.1 but for the emergent constraints VOL and ZHA. Adapted with permission from Schlund et al. (2020b).

6. Constraining Future Gross Primary Production with Machine Learning

- Single-process emergent constraints have disadvantages

To overcome these, we explore an alternative approach that expands the MDER method with an 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. The target variable analyzed is GPP. As described in detail in section XXX, the carbon cycle slows down global warming by absorbing 30 % CO2, thus accurately quantifying the carbon sink is crucial. GPP is most largest flux of the terrestrial carbon cycle. Due to feedbacks (section XXX, NAME OF FEEDBACKS), multi-model projectionf of future GPP is uncertain, especially in CMIP5 RCP 8.5 at the end of the century. Thus, the aim of the works presented in this chapter is to reduce uncertainties in multi-model projections of future GPP in the RCP 8.5 scenario at the end of the 21st century with a new ML-based weighting technique and observations. The new two-step approach utilizes aspects of emergent constraints and the MDER technique in combination with a supervised machine learning algorithm. In the first step, we apply an existing emergent... (see SCHLUND ET AL., 2020).

6.1. Step 1: XXXXXXXXXXXXXXXXXX

TBA.

7. Summary and Outlook

Outlook:

   - Why GBRT not used for ECS?

   - GBRT used for temperature anomaly based on other studies (emergent constraints, historical T trend, weighting following Brunner et al.)

Appendix

A. Supplementary Materials for Chapter 5

This section contains supplementary tables for chapter 5 (Evaluation of Emergent Constraints on the Effective Climate Sensitivity in CMIP6; see Schlund et al. 2020b).

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

Table A.1.: Overview of the variables used in chapter 5 (Evaluation of Emergent Constraints on the Effective Climate Sensitivity in CMIP6). More details are given in table 5.1, which lists the corresponding variable short names for each emergent constraint.

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

Table A.2.: References for all observational datasets used in chapter 5 (Evaluation of Emergent Constraints on the Effective Climate Sensitivity in CMIP6).

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 (Evaluation of Emergent Constraints on the Effective Climate Sensitivity in CMIP6) 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	Séférian 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⁻¹]	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 (Evaluation of Emergent Constraints on the Effective Climate Sensitivity in CMIP6) 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^{-1}]	VOL [%]	ZHA [% K^{-1}]
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]	BRI [% K^{-1}]	COX [K]	LIP [°]	SHD [1]	SHL [1]	SHS [1]
ACCESS-CM2	29	4.72	−0.60	0.20	−35.14	0.42	0.88	0.46
ACCESS-ESM1-5	30	3.87	−0.81	0.13	−35.00	0.44	0.79	0.35
AWI-CM-1-1-MR	31	3.16	−0.70		−35.43	0.42	0.76	0.34
BCC-CSM2-MR	32	3.04	−1.43	0.16	−36.12	0.54	0.89	0.36
BCC-ESM1	33	3.26		0.19	−36.18	0.46	0.88	0.42
CAMS-CSM1-0	34	2.29	−0.44	0.14	−35.38	0.47	0.91	0.44
CAS-ESM2-0	35	3.51	−0.84	0.16	−30.53	0.31	0.80	0.48
CESM2	36	5.16	−1.94	0.15	−36.22	0.49	0.85	0.36
CESM2-FV2	37	5.14	−1.30	0.13	−36.88	0.45	0.79	0.34
CESM2-WACCM	38	4.75	−1.60	0.18	−36.43	0.55	0.91	0.36
CESM2-WACCM-FV2	39	4.79		0.16	−36.57			
CMCC-CM2-SR5	40	3.52	−0.19	0.18	−37.48	0.50	0.90	0.40
CNRM-CM6-1	41	4.83	−0.61	0.11	−34.50	0.30	0.74	0.44
CNRM-CM6-1-HR	42	4.28	−0.25	0.11	−33.41	0.34	0.81	0.46
CNRM-ESM2-1	43	4.76	−0.51	0.08	−34.83	0.30	0.74	0.44
CanESM5	44	5.62	−0.97	0.18	−35.50	0.48	0.82	0.34
E3SM-1-0	45	5.32	−0.74	0.22	−35.59	0.46	0.80	0.34
EC-Earth3-Veg	46	4.31		0.19				
FGOALS-f3-L	47	3.00	−0.63	0.09	−33.40	0.31	0.76	0.44
FGOALS-g3	48	2.88	−0.12	0.13	−34.57	0.36	0.80	0.44
GISS-E2-1-G	49	2.72	−0.58	0.17	−36.24	0.26	0.61	0.35
GISS-E2-1-H	50	3.11	−0.48	0.13	−36.95	0.24	0.59	0.36
HadGEM3-GC31-LL	51	5.55	−0.56	0.20	−35.75	0.46	0.92	0.46
HadGEM3-GC31-MM	52	5.42	−1.06	0.18	−35.09			
INM-CM4-8	53	1.83	0.03	0.13	−38.15	0.47	0.71	0.25
INM-CM5-0	54	1.92	−0.46	0.14	−37.34	0.40	0.63	0.23
IPSL-CM6A-LR	55	4.56	−0.38	0.14	−33.74	0.50	0.92	0.41
KACE-1-0-G	56	4.48	−0.60	0.15	−34.95	0.41	0.87	0.47
MCM-UA-1-0	57	3.65		0.20				
MIROC-ES2L	58	2.68	−0.49	0.15	−34.13	0.36	0.79	0.43
MIROC6	59	2.61	−0.45	0.17	−34.90	0.46	0.91	0.45
MPI-ESM-1-2-HAM	60	2.96	−0.60	0.14	−35.21	0.42	0.77	0.35
MPI-ESM1-2-HR	61	2.98	−0.34	0.15	−34.60	0.45	0.78	0.33
MPI-ESM1-2-LR	62	3.00	−0.64	0.18	−34.76	0.39	0.68	0.29
MRI-ESM2-0	63	3.15	−0.73	0.15	−33.39	0.39	0.79	0.40
NESM3	64	4.72	−0.73	0.24	−36.77	0.52	0.89	0.37
NorCPM1	65	3.05		0.12	−37.45	0.45		
NorESM2-LM	66	2.54	−1.61	0.20	−36.17	0.46	0.79	0.33
NorESM2-MM	67	2.50	0.12	0.31	−36.55	0.52	0.86	0.34
SAM0-UNICON	68	3.72	−1.40	0.19	−36.41	0.54	0.90	0.36
TaiESM1	69	4.31	−0.22	0.15	−36.80	0.48	0.89	0.42
UKESM1-0-LL	70	5.34	−0.59	0.21	−35.67	0.48	0.93	0.45

Table A.7.: As in table A.5 but for the CMIP6 models. Adapted with permission from Schlund et al. (2020b).

Model	Index	ECS [K]	SU [1]	TIH [%]	TII [mm day^{-1}]	VOL [%]	ZHA [% K^{-1}]
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).

List of Acronyms

Integrated Author's References

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). 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*, *in review*, 1–43. https://doi.org/10.5194/gmd-2020-244

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:TVGPCP>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

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

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., Nikonov, S., Radhakrishnan, A., Stockhause, 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

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

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., Perlwitz, 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

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). 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

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

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

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

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

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., Galloway, 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

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

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

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

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

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

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). 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

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

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., Hurtt, 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

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

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

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

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

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

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). 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

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). https://doi.org/10.1029/2008jd010405

Gregory, J. M., & Webb, M. (2008). Tropospheric Adjustment Induces a Cloud Component in CO2 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

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

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

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

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), 13, 683–13, 689. https://doi.org/10.1002/2014jd022466

Iversen, T., Bentsen, M., Bethke, I., Debernard, J. B., Kirkevag, A., Seland, 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_2 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*. 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

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., Piper, S. C., Bacastow, R. B., Wahlen, M., Whorf, T. P., Heimann, M., & Meijer, H. A. (2005). Atmospheric CO2 and 13CO2 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). 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 interdependence. *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., Siahaan, 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., Ciais, 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., 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., & Vohralik, 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

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). 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

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_2 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

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 CO2. *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

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

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

Mulcahy, J. P., Johnson, C., Jones, C. G., Povey, A. C., Scott, C. E., Sellar, A., Turnock, S. T., Woodhouse, M. T., Abraham, N. L., Andrews, M. B., Bellouin, N., Browse, J., Carslaw, K. S., Dalvi, M., Folberth, G. A., Glover, M., Grosvenor, D. P., Hardacre, C., Hill, R.,

. . . Yool, A. (2020). Description and evaluation of aerosol in UKESM1 and HadGEM3-GC3.1 CMIP6 historical simulations. *Geoscientific Model Development*, *13*(12), 6383–6423. https://doi.org/10.5194/gmd-13-6383-2020

Muller, W. A., Jungclaus, J. H., Mauritsen, 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

Nijsse, 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

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., Hurtt, 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

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

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

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

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

References

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

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

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., 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., Mauritsen, 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., Bloch-Johnson, J., Gregory, J., Andrews, T., Mauritsen, 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

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 Multimodel 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., Olivié, D., Toniazzo, T., Gjermundsen, A., Debernard, J. B., Gupta, A. K., He, Y., Kirkevåg, A., Schwinger, J., Tjiputra, J., Schancke Aas, K., Bethke, I., Fan, Y., Griesfeller, J., Grini, A., Guo, C., Ilicak, M., . . . 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., Mauritsen, 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

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

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

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

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

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 Mélia, 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., De Kauwe, M. G., Bastos, A., Belmecheri, S., Georgiou, K., Keeling, R., 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., Ciais, P., . . . Zuidema, P. A. (2020). Integrating the evidence for a terrestrial carbon sink

caused by increasing atmospheric CO2. *New Phytologist*, *Accepted Author Manuscript*. 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, November). *WCRP Coupled Model Intercomparison Project (CMIP)*. https://www.wcrp-climate.org/wgcm-cmip

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*, *in review*, 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 CO2. *Nature*, *538*(7626), 499–+. 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/ORNLDAAC/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). https://doi.org/10.1029/2019ms001986

Winkler, A. J., Myneni, R. B., Alexandrov, G. A., & Brovkin, V. (2019). Earth system models underestimate carbon fixation by plants in the high latitudes. *Nature Communications*, *10*. https://doi.org/10.1038/s41467-019-08633-z

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

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

I assure that this thesis is a result of my personal work and that no other than the indicated aids have been used for its completion. Furthermore I assure that all quotations and statements that have been inferred literally or in a general manner from published or unpublished writings are marked as such. Beyond this I assure that the work has not been used, neither completely nor in parts, to pass any previous examination.

Declaration of Authorship

I assure that this thesis is a result of my personal work and that no other than the indicated aids have been used for its completion. Furthermore I assure that all quotations and statements that have been inferred literally or in a general manner from published or unpublished writings are marked as such. Beyond this I assure that the work has not been used, neither completely nor in parts, to pass any previous examination.

Oberpfaffenhofen, March 2021

_____

Manuel SCHLUND