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**

1.1. Motivation . . . 1
1.2. Research Questions . . . 1
1.3. Structure of the Thesis . . . 2

**2. Scientific Background** **3**

2.1. Earth System Models: Simulations and Analysis . . . 3
2.1.1. Numerical Climate Modeling . . . 3
2.1.2. CMIP . . . 5
2.1.3. Sources of Uncertainties in Climate Model Projections . . . 7
2.2. Climate Sensitivity . . . 10
2.2.1. Climate Feedbacks . . . 10
2.2.2. Mathematical Framework for Feedbacks Analysis . . . 12
2.2.3. Equilibrium and Effective Climate Sensitivity . . . 14
2.2.4. Cloud-Related Feedback Parameters . . . 17
2.2.5. Transient Climate Response . . . 17
2.3. The Global Carbon Cycle . . . 19
2.3.1. Overview . . . 19
2.3.2. Anthropogenic Perturbations . . . 21
2.4. Reducing Uncertainties in Multi-Model Climate Projections with Observations . . . 24
2.4.1. Emergent Constraints . . . 25
2.4.2. Performance- and Interdependence-based Weighting of Climate Models . . . 29
2.4.3. Multiple Diagnostic Ensemble Regression . . . 31

**3. Improving Routine Climate Model Evaluation** **35**

3.1. The Earth System Model Evaluation Tool (ESMValTool) . . . 35
3.2. Contributions to ESMValCore . . . 38
3.3. Contributions to ESMValTool . . . 41

**4. Assessment of Policy-Relevant Climate Metrics in CMIP6** **45**

**5. Evaluation of Emergent Constraints on the Effective Climate Sensitivity in CMIP6** **47**
5.1. Comparison of Emergent Constraints on ECS for CMIP5 and CMIP6 47
5.1.1. COX . . . . . 47

**6. Constraining Uncertainties in Future Gross Primary Productivity with Machine Learning** **49**
6.1. Step 1: XXXXXXXXXXXXXXXXX . . . . . 49

**7. Summary and Outlook** **51**

**Appendix** **53**
A. TBA . . . . . 53
A.1. test . . . . . 53
B. TBA . . . . . 53

**List of Acronyms** **55**

**Integrated Author's References** **57**

**References** **59**

**Declaration of Authorship** **73**

# 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 (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 Coupled Model Intercomparison Project (CMIP)?
  - One sentence on Machine Learning (ML).
  - Mention IPCC assessments, older ones and upcoming AR6

## 1.2. Research Questions

This section will present the **AWESOME** research questions!!!

1. Did climate sensitivity change in the latest generation of climate models from CMIP6 and can we see progress compared to older climate model generations?

2. Can we use emergent constraints established in CMIP3 and CMIP5 to reduce uncertainties in climate sensitivity in the CMIP6 ensemble?

3. Can we use ML-based approaches to constrain gridded projections of the future climate with 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). Unless stated otherwise, results, discussions and conclusions presented from these studies originate from the author of this thesis. A detailed list of corresponding references is given at the beginning of the relevant 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 Earth System Models (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 policy-relevant climate metrics like the ECS and the TCR in the latest generation of ESMs. 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, Lauer, et al., 2020). 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 21$^{st}$ century, which is already published in the *Journal of Geophysical Research: Biogeosciences* (Schlund, Eyring, et al., 2020). 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. Next, important metrics describing climate sensitivity and fundamental biogeochemical processes of the global carbon cycle 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. 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.

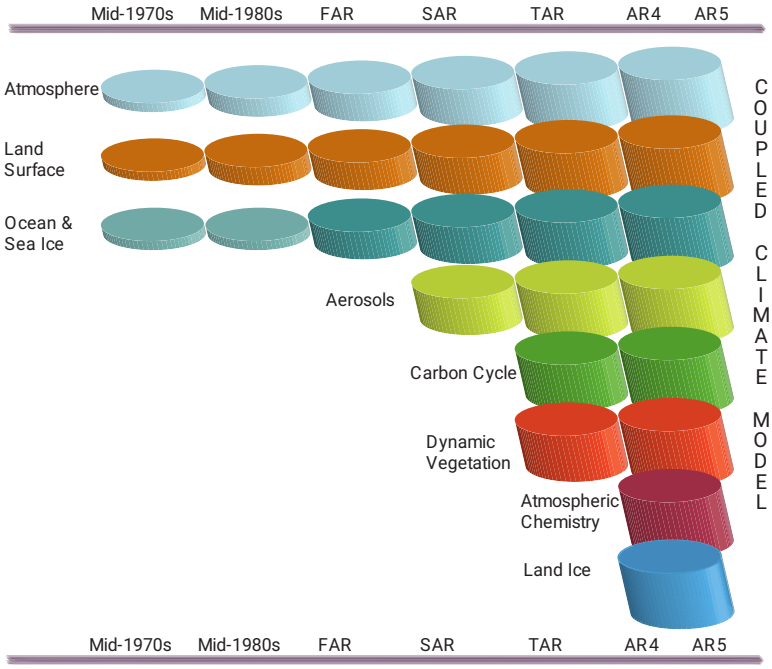



Figure 2.1.: Historical evolution of coupled climate models over the last 45 years. In early days, these models were so-called Atmosphere-Ocean General Circulation Models (AOGCMs) and only included three components: the atmosphere, the land surface and the ocean. Over the time, the individual components 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). Taken from Cubasch et al. (2013).

Over the course of the years, climate models became more and more complex by including a wider range of processes within the components, but also by introducing new components to the coupled system. Examples of these are aerosols, the carbon cycle, 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 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., 2019). 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. 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). 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, Bony, et al., 2016; 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

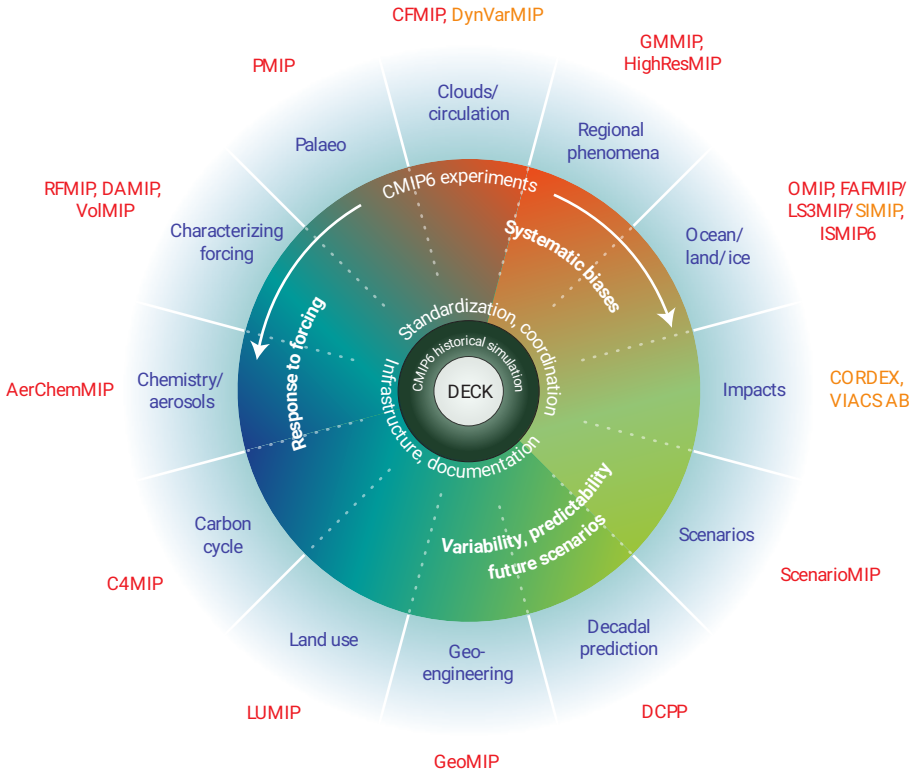



Figure 2.2.: Schematic of the experiment design of Phase 6 of the Coupled Model Intercomparison Project (CMIP6). The center of the circle illustrates the four DECK (Diagnostic, Evaluation, and Characterisation of Klima) experiments and the CMIP6 historical simulation. The circular sectors show additional science themes that can be explored through the 21 CMIP6-Endorsed Model Intercomparison Projects (MIPs). Taken from Simpkins (2017).

(*1pctCO2*) (Eyring, Bony, et al., 2016). 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., 2017; O'Neill et al., 2013). 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, 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 can distinguish: natural variability, climate response uncertainty and emission uncertainty (Hawkins & Sutton, 2009, 2010). Figure 2.3 shows these three sources for the projected global mean surface temperature anomaly over the 21$^{\mathrm{st}}$ 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

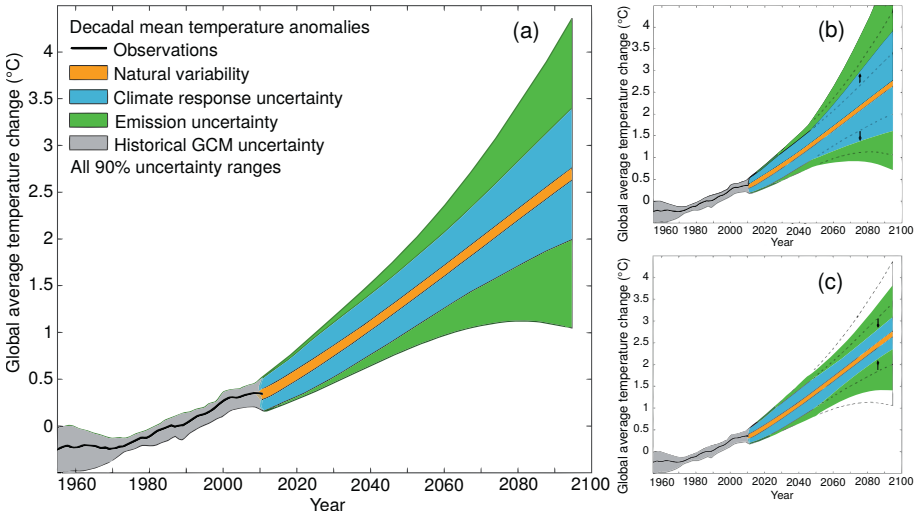



Figure 2.3.: Schematic illustrating the importance of different sources of uncertainties in climate model projections and their evolution in time. (a) Time series of the anomaly of the decadal and global mean surface temperature relative to the period 1961–1980. The black line shows the historical observations 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 & 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. Taken from Cubasch et al. (2013).

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 un-

certainty 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 & 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

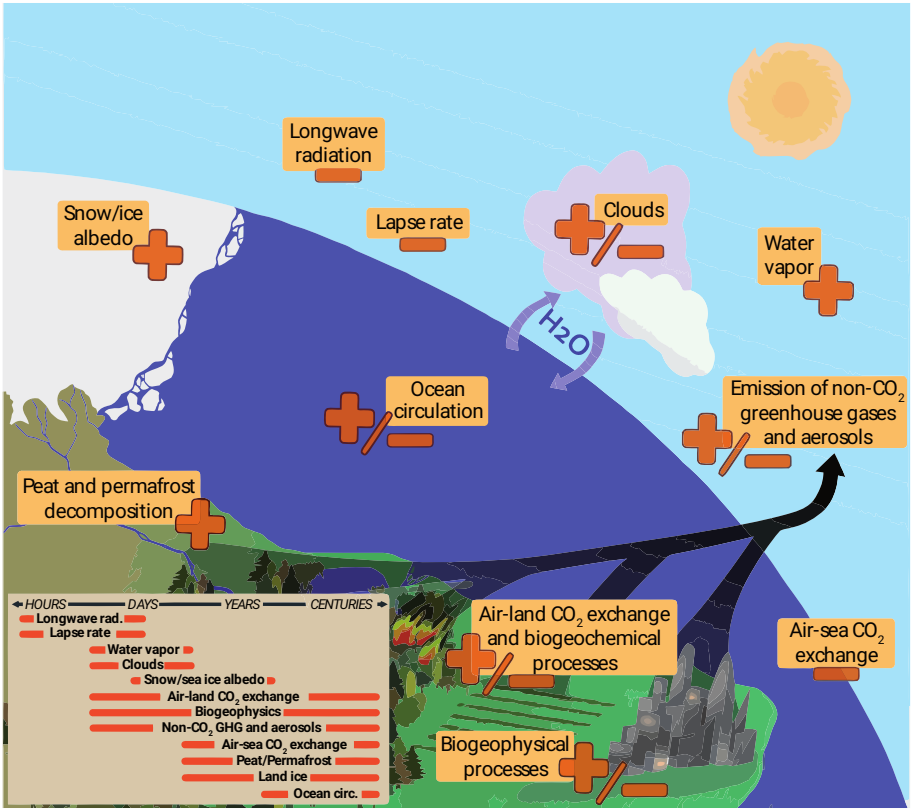



Figure 2.4.: Climate feedbacks and corresponding time scales. "+" refers to positive feedbacks, which amplify the effect of the external forcing (e.g. the water vapor feedback). "-" refers to negative feedbacks, which diminish the effect of the external forcing (e.g. the longwave radiation feedback). "+/-" 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. Taken from Cubasch et al. (2013).

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 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 & Bony, 2008; Soden & 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

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

$\sigma \approx 5.67\,\mathrm{Wm^{-2}K^{-4}}$ 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 $\Delta T$:

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

Since the temperature change caused by an anthropogenic forcing is much smaller than the equilibrium temperature $\Delta T \ll T_0 \approx 255\,\mathrm{K}$, 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

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

with the blackbody feedback parameter $\lambda_{\mathrm{BB}} \approx -3.8\,\mathrm{Wm^{-2}K^{-1}}$. Results from climate models and analyses of observations confirm this linear relationship between $R$ and $\Delta T$ (Gregory et al., 2004). However, the value of this linear constant $\lambda$, the *climate feedback parameter*, is found to be considerably larger than the blackbody response ($\lambda \approx -1.0\,\mathrm{Wm^{-2}K^{-1}}$), indicating that additional processes affect the Earth's radiative balance: the climate feedbacks (Flato et al., 2013; Gregory et al., 2009). Since climate models suggest that the radiative effects of these additional feedbacks are also proportional to $\Delta T$ (Gregory & Webb, 2008), equation (2.5) can be adjusted to

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

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

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

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

### 2.2.3. Equilibrium and Effective Climate Sensitivity

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 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 $\lambda$. The steady state values for $F$ and $\lambda$ 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 $\Delta T$

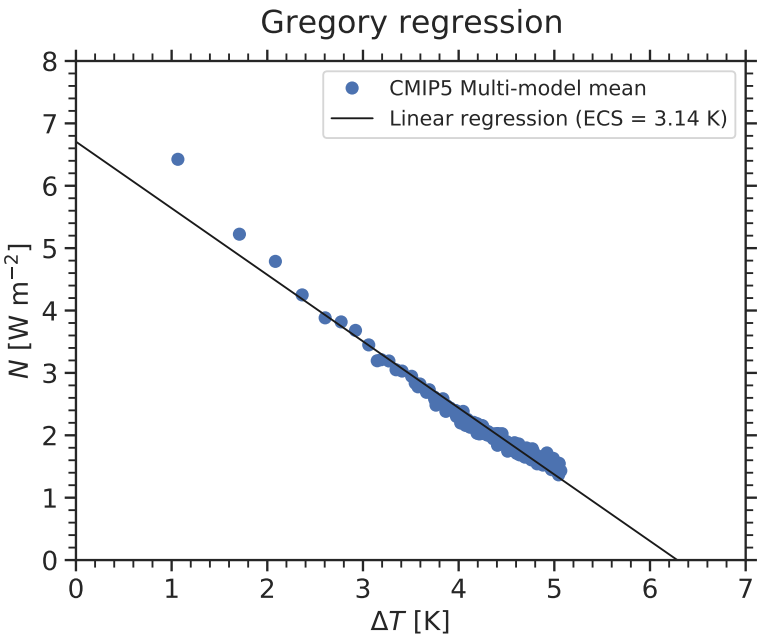



Figure 2.5.: Global and annual mean net top of the atmosphere (TOA) radiation $N$ versus the change in global and annual mean near-surface air temperature $\Delta T$ for 150 years of a simulation with an abrupt quadrupling of the atmospheric $CO_2$ concentration ($4xCO_2$) for the CMIP5 multi-model mean (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 $\lambda$ and the $y$-intercept corresponds to the radiative forcing $F_{4x}$. These can be used to calculate the Effective Climate Sensitivity (ECS) with the Gregory regression method according to equation (2.10) (Gregory et al., 2004). ECS is equivalently given by the $x$-intercept of the linear regression line divided by 2. Here, ECS = 3.14 K.

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

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

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

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

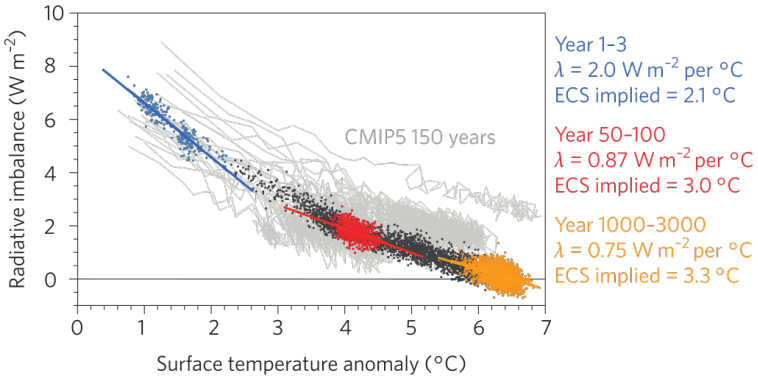



Figure 2.6.: As figure 2.5, but for different time periods considered in the Gregory regression resulting in different values for the Effective Climate Sensitivity (ECS) and the climate feedback parameter $\lambda$. Dark gray, blue, red and orange colors correspond to the Community Earth System Model (CESM) of the National Center for Atmospheric Research (NCAR) for different time periods of a long running simulation (see legend). The initial years are simulated many times for different initial conditions. Light gray colors correspond to the CMIP5 ensemble (150 years each). Taken from Knutti, Rugenstein, and Hegerl (2017).

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, Rugenstein, & Hegerl, 2017; Knutti & Rugenstein, 2015). As figure 2.6 shows, the slope in the Gregory regression is not constant, but rather decreases over time when a long running 4x$CO_2$ simulation with more than 1000 model years is considered. As a result, the climate feedback parameter $\lambda$ decreases over time, resulting in a higher ECS. Major reasons for this are temperature dependencies of the feedbacks, atmospheric and oceanic adjustments over time, changing warming patterns over time, non-additive feedbacks and dependencies on the type and magnitude of the external forcings (Knutti, Rugenstein, & Hegerl, 2017). 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 4x$CO_2$ instead of a 2x$CO_2$ 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 & 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 $\lambda$ and the ECS, the Gregory regression can also be used to estimate cloud-related feedback parameters. For this, the net TOA radiation $N$ on the $y$-axis in figure 2.5 is replaced with the cloud radiative effect (CRE), which is defined as the difference between the all-sky (i.e. with clouds if present) net TOA radiation and the clear-sky (i.e. clouds artificially removed) net TOA radiation (Andrews et al., 2012). This can be done for the shortwave $N_{\text{SWCRE}}$ and longwave $N_{\text{LWCRE}}$ components separately, but also for the combined effect $N_{\text{CRE}} = N_{\text{SWCRE}} + N_{\text{LWCRE}}$. The slopes in the corresponding Gregory regressions are the so-called *CRE feedback parameters* $\lambda_{\text{SWCRE}}$, $\lambda_{\text{LWCRE}}$ and $\lambda_{\text{CRE}}$, which quantify the change in CRE as a response to increasing GSATs.

### 2.2.5. Transient Climate Response

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.

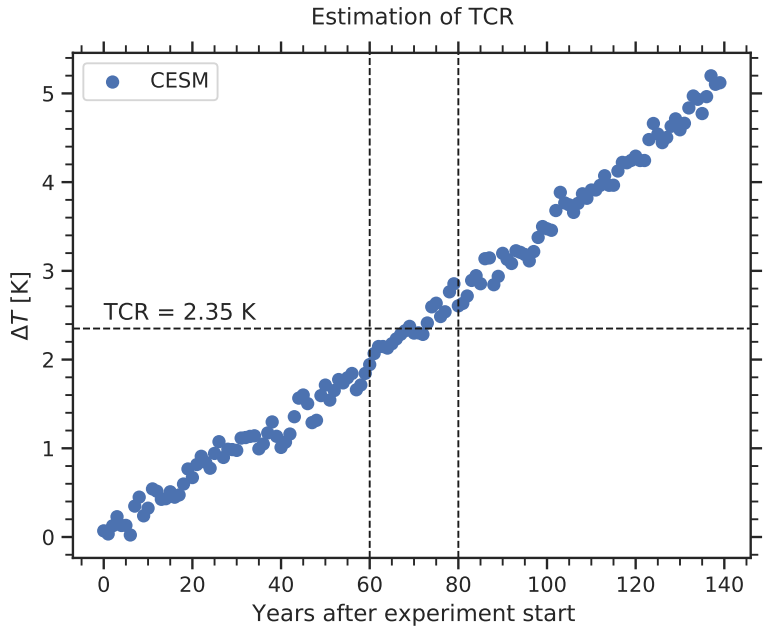



Figure 2.7.: Global and annual mean near-surface air temperature change $\Delta T$ for 140 years of a simulation with a 1 % per year increase of the atmospheric $CO_2$ concentration (1%$CO_2$) for the Community Earth System Model (CESM) of the National Center for Atmospheric Research (NCAR) (blue circles). 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 $\Delta T$ 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).

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, Rugenstein, & Hegerl, 2017). 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 $\kappa$ is the ocean heat uptake efficiency (Gregory & Forster, 2008). Since there is a net energy flux into the climate system ($N > 0$) due to the external forcing $F$, $\kappa$ is positive. This approximation becomes less accurate as the deeper ocean warms up and cannot be applied to simulations with a steady state climate change ($N \to 0$) where the external forcing tends stabilize (Gregory et al., 2009).

By applying the definition of TCR (transient GSAT change $\Delta T$ at the time of $CO_2$ doubling) and combining equations (2.9) and (2.11), TCR can be estimated as

$$\mathrm{TCR} = \frac{F_{2x}}{\kappa - \lambda}, \tag{2.12}$$

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 $\mathrm{ECS} = -F_{2x}/\lambda$ and assuming path independence of the forcing (i.e. the resulting radiative forcing from the 1%$CO_2$ and 2x$CO_2$ runs are comparable) gives

$$\mathrm{TCR} = \frac{F_{2x} \cdot \mathrm{ECS}}{F_{2x} + \kappa \cdot \mathrm{ECS}}, \tag{2.13}$$

which demonstrates the non-linear connection between TCR and ECS (Gregory & 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, Eyring, et al., 2020; 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 $\mathrm{GtC\,yr^{-1}}$ 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

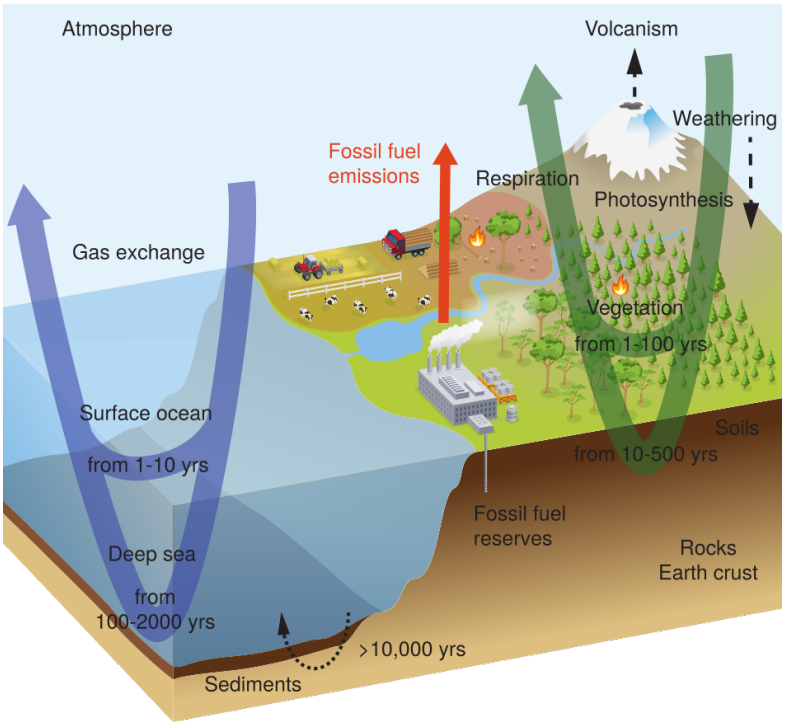



Figure 2.8.: Simplified schematic of the global carbon cycle including the typical turnover time scales for carbon transfers through the major reservoirs (atmosphere, land surface and ocean). Taken from Ciais et al. (2013).

are comparatively small ($< 0.3\,\mathrm{GtC\,yr^{-1}}$) 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., 2019). 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., 2019).

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 Productivity (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 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 $19^{th}$ 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 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

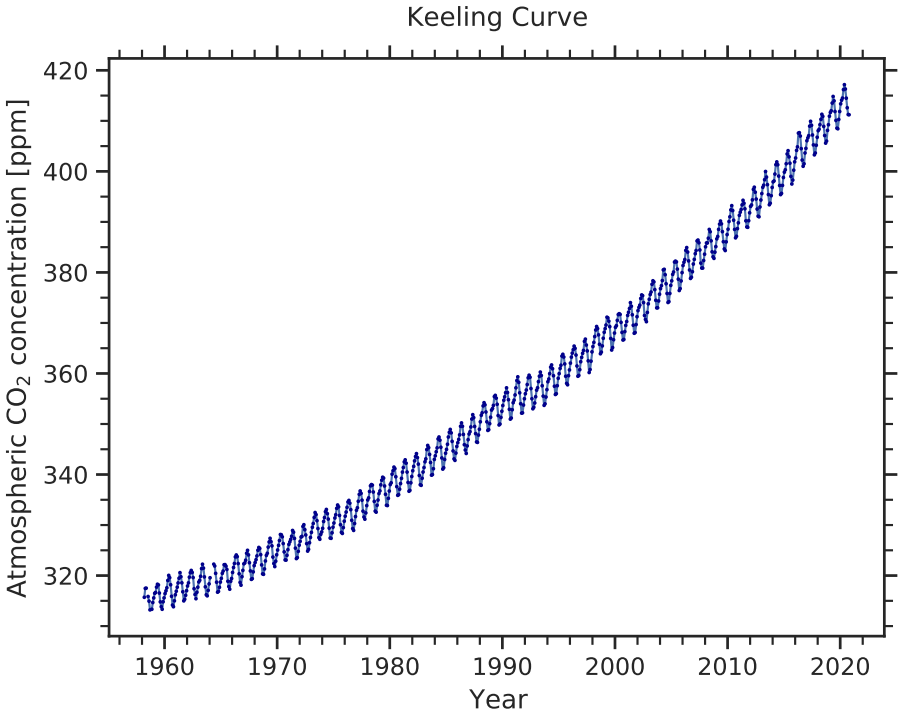



Figure 2.9.: The Keeling Curve: monthly-mean atmospheric $CO_2$ concentration at the Mauna Loa Observatory, Hawaii (19.5 °N, 155.6 °W; elevation: 3397 m) from 1958 to 2019 (Keeling et al., 2005). The steady increase of the atmospheric $CO_2$ concentration is superimposed with a seasonal oscillation caused by the seasonal $CO_2$ cycle (see section 2.3.1).

$CO_2$ 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 2009–2018, anthropogenic activities caused net carbon fluxes of 3.2 GtC $yr^{-1}$ from the atmosphere into the terrestrial biosphere due to increased plant photosynthesis and 2.5 GtC $yr^{-1}$ from the atmosphere into the ocean due to increase dissolution of $CO_2$ into the sea (Friedlingstein et al., 2019). In the

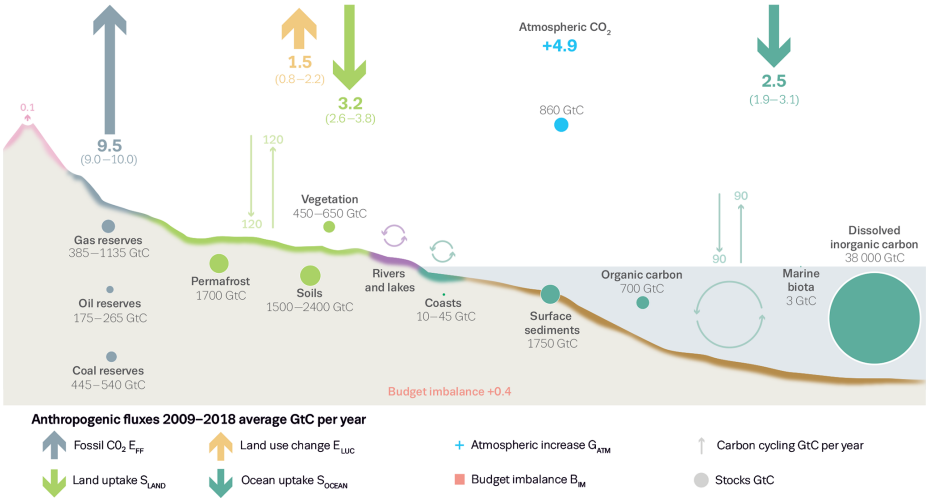



Figure 2.10.: Schematic representation of the overall perturbation of the global carbon cycle caused by anthropogenic activities, averaged globally for the decade 2009–2018. Arrows represent carbon exchange fluxes; circles carbon reservoirs. More details are given in the legend of this figure. Taken from Friedlingstein et al. (2019).

same time, the amount of carbon in the atmosphere reservoir increased with a rate of 4.9 GtC $yr^{-1}$, indicating that only about half of the anthropogenic $CO_2$ emissions in the last decade remained in the atmosphere (Friedlingstein et al., 2019) 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, Sedlacek, et al., 2017). 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 & 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, Sedlacek, et al., 2017).

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 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., 2018; Cox et al., 2013). Let $x_m$ be the observable predictor variable for climate model $m$ and $y_m$ the corresponding target variable. To find the linear emergent relationship for a climate model ensemble with $M$ climate models and data $\{(x_m, y_m) \mid m \in I_M\}$ with index set $I_M = \{1, 2, \ldots, M\}$, 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

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

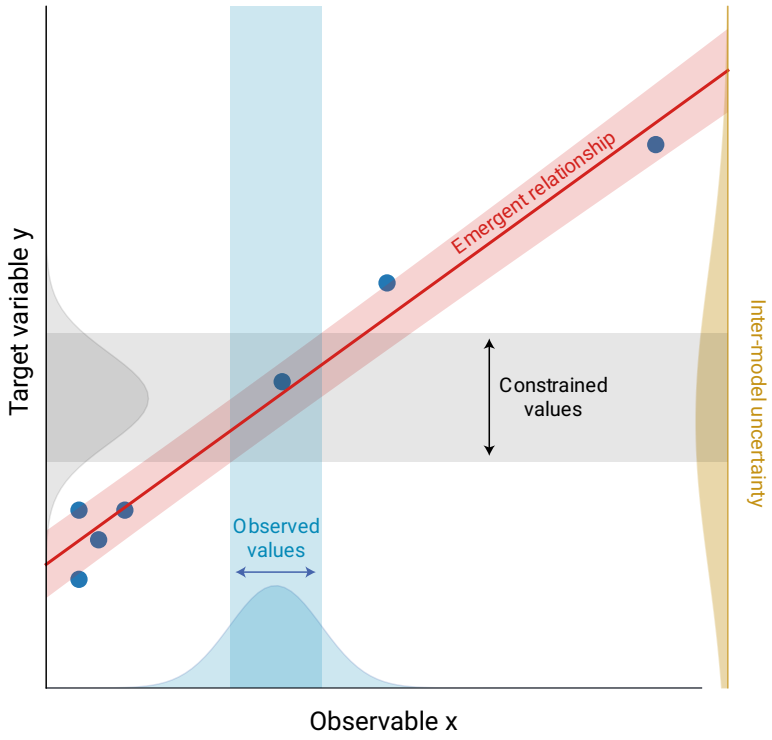



Figure 2.11.: Schematic illustration of the emergent constraint approach. Basis of every emergent constraint is a robust and physically interpretable emergent relationship (red line) between a target variable $y$ (e.g. Earth system sensitivity or projection of future climate change) and an observable $x$ (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). Taken and modified from Eyring et al. (2019).

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

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

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 $x$:

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

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 $\sigma_x$ is the standard deviation of the observation around the true value. Assuming an imperfect uniform prior $P(x) \propto 1$ with cut-offs at $-\infty$ and $+\infty$ and using Bayes' theorem implies $P(x_0|x) = P(x|x_0)$. In a final step, this can be used to calculate the posterior PDF for the constrained prediction of the target variable $y$ given the observation $x_0$ (gray PDF in figure 2.11) 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 & 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

$$t = \frac{r\sqrt{M-2}}{\sqrt{1-r^2}} \tag{2.20}$$

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, Sedlacek, et al., 2017; Sanderson et al., 2015a; Sanderson et al., 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 & Mauritsen, 2019). Moreover, using only a single observational dataset to estimate $x_0$ and $\sigma_x$ when different datasets are available might lead to an underestimation of the observational uncertainty, as different observational datasets might lead to different emergent constraints.

A crucial aspect for every emergent constraint is a verifiable physical process explaining the correlation between $x$ and $y$ (Hall et al., 2019). Only if the underlying emergent relationship can be derived from a robust and plausible physical mechanism, an emergent constraint can be considered credible. The reason for this 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 & Ingram, 2002) and the snow-albedo feedback (Hall & 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 & Schneider, 2016; Brient et al., 2015; Fasullo & 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 5.1. 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 & 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, 2016; Winkler et al., 2019). An extensive discussion on the emergent constraint by Wenzel, Cox, et al. (2016), 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

$$y = \sum_{m=1}^{M} w_m y_m \tag{2.21}$$

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, Sedlacek, et al. (2017) propose a weighting scheme based on climate model performance and interdependence with the following weights:

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

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). $\sigma_D$ and $\sigma_S$ are constants that determine the individual strength of the performance and interdependence weighting, respectively.

A commonly used distance metric to measure model performance $D_m$ and model interdependence $S_{mn}$ is the root-mean-square error (RMSE), but others are possible (Knutti, Sedlacek, et al., 2017). 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, Sedlacek, et al., 2017). In practice, this choice is either based on expert judgment about relevant processes, on emergent relationships (see section 2.4.1) or on multivariate regression models (Karpechko et al., 2013; Sanderson et al., 2015b; see section 2.4.3). It might also be beneficial to use different diagnostics for the calculations of the performance and interdependence metrics (Merrifield et al., 2020) and/or to remove selected diagnostics based on their mutual correlation (Lorenz et al., 2018).

The constants $\sigma_D$ and $\sigma_S$ determine how strongly the climate models' performance and interdependence are weighted (Knutti, Sedlacek, et al., 2017). Small values of the performance parameter $\sigma_D$ lead to an aggressive weighting with only a few climate models receiving a majority of the weight, while large values of $\sigma_D$ result in an equal weighting. For the interdependence parameter $\sigma_S$, this is slightly different: Here, small (all climate models are independent) and large (all climate models are dependent) values lead to an almost equal weighting. Thus, an optimal choice for $\sigma_D$ and $\sigma_S$ is crucial. A useful tool to

estimate these optimal parameters is the *leave-one-model-out cross-validation (CV)* approach, which is also known as *pseudo-reality*, *model-as-truth* or *perfect model* setup (de Elia et al., 2002; Karpechko et al., 2013). For this, a single climate model is removed from the multi-model ensemble and treated as observation (*pseudo-observation*). Then, a weighted MMM with weights computed from the updated model ensemble is calculated, which gives a prediction for the "true" climate model. This allows a simple quantitative assessment of the weighting scheme by calculating the RMSE between the prediction and the known ground truth of the pseudo-observation. 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 $\sigma_D$ and $\sigma_S$ can be assessed using specific criteria to find optimal values for $\sigma_D$ and $\sigma_S$ (Knutti, Sedlacek, et al., 2017). 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 $\sigma_D$ and $\sigma_S$ adds another level of subjectivity to the weighting scheme, which can partly be addressed with the introduced leave-one-model-out CV setup. Nevertheless, due to its flexibility, the climate model weighting scheme of Knutti, Sedlacek, et al. (2017) has already been used for various target variables: Arctic sea ice (Knutti, Sedlacek, et al., 2017), Antarctic ozone concentrations (Amos et al., 2020), North American maximum temperature (Lorenz et al., 2018), European temperature and precipitation (Brunner et al., 2019; Merrifield et al., 2020) and global warming over the 21$^{\text{st}}$ 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 $\left\{x^{(1)}, x^{(2)}, \ldots, x^{(K)}\right\}$ 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

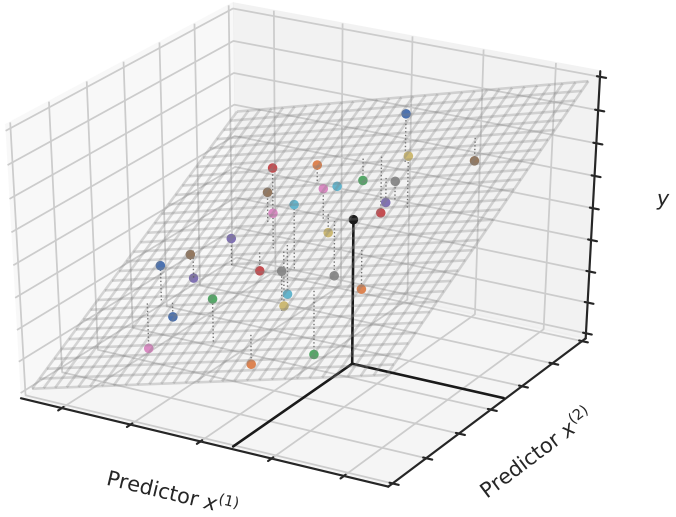



Figure 2.12.: Schematic illustration of the Multiple Diagnostic Ensemble Regression (MDER) approach (Karpechko et al., 2013). First, inter-model relationships between a target variable $y$ and multiple process-based predictors $x^{(k)}$ (here: two predictors $x^{(1)}$ and $x^{(2)}$) are used to fit a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

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 $\boldsymbol{y} = \left(y_1, y_2, \ldots, y_M\right)^T \in \mathbb{R}^M$ be the vector of target variables of the $M$ climate models ($T$ denotes the transpose) and $\boldsymbol{X} \in \mathbb{R}^{M\times(K+1)}$ the design matrix representing the predictors:

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

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

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{b} + \boldsymbol{\varepsilon}, \tag{2.24}$$

where $\boldsymbol{b} = (b_0, b_1, \ldots, b_K)^T \in \mathbb{R}^{(K+1)}$ is the vector of linear coefficients (with intercept $b_0$) and $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_M)^T \in \mathbb{R}^M$ 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 $\hat{\boldsymbol{b}}$ are given by

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

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)).

In a second step, observed data of the process-based diagnostics $\boldsymbol{x}_0 = \left(1, x_0^{(1)}, x_0^{(2)}, \ldots, x_0^{(K)}\right)^T \in \mathbb{R}^{(K+1)}$ 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):

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

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 $\boldsymbol{x}_0$ 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 $\boldsymbol{w} = (w_1, w_2, \ldots, w_M)^T \in \mathbb{R}^M$ can be defined by

$$\boldsymbol{w} = \left[\boldsymbol{x}_0^T\left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T\right]^T, \tag{2.27}$$

which can be used to calculate the weighted target variable by

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

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, Eyring, et al., 2016) and in projections of the Arctic sea ice extent (Senftleben et al., 2020).

# 3. Improving Routine Climate Model Evaluation

In order to answer the research 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 over the ESMValTool's structure, this chapter presents all these contributions that are not documented in other chapters of this thesis.

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

As shown 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

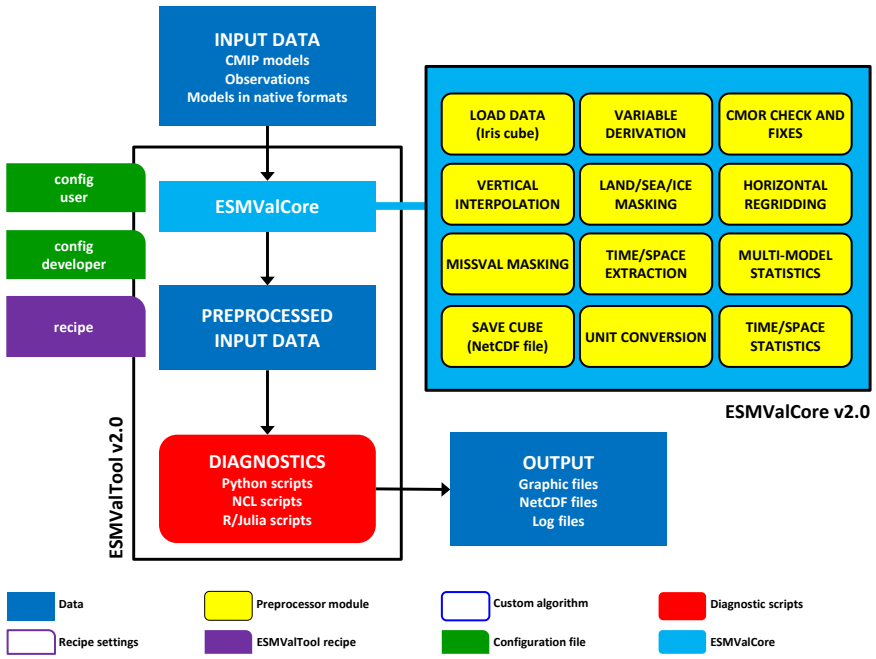



Figure 3.1.: Schematic representation of the Earth System Model Evaluation Tool (ESMValTool). First, input data is preprocessed by the core functionalities of the ESMValTool (ESMValCore). Second, the preprocessed data is read by diagnostic scripts, which create the final output (e.g. plots). The user 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. Taken and modified from Righi et al. (2020).

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, Righi, et al., 2016), the ESMValTool has been greatly extended and improved. A schematic of the current structure of the ESMValTool is illustrated in figure 3.1. The first major element of this workflow is an extensive preprocessing of the input data. For this, the Python-based core functionalities of the ESMValTool (ESMValCore) provide a set of operators (*preprocessors*) that are applied to the input data. There are two classes of preprocessors: non-optional and optional 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.

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, Gleckler, et al., 2016). 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 (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, Broetz, de Mora, Drost, Eyring, Koldunov, Lauer, Predoi, et al., 2020). 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, Broetz, de Mora, Drost, Eyring, Koldunov, Lauer, Mueller, et al., 2020).

## 3.2. Contributions to ESMValCore

As of December 2020, 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).

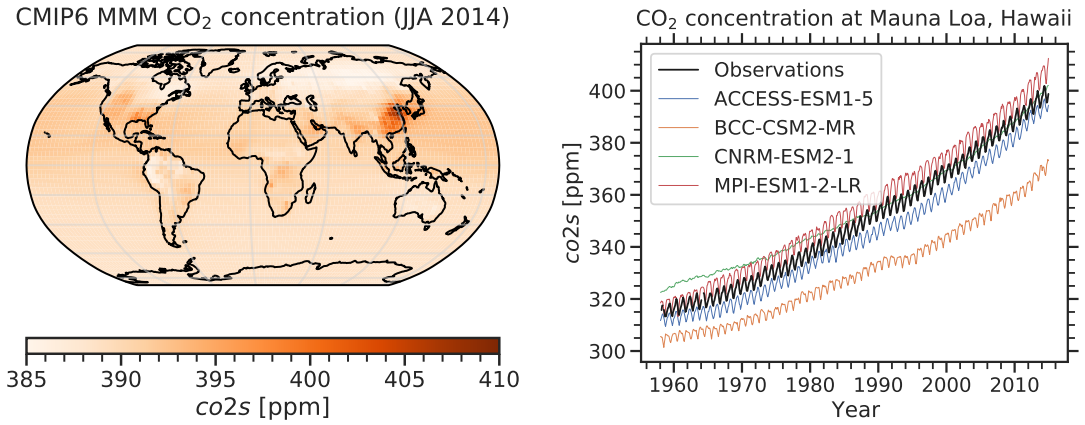
In total, three new preprocessor functions have been added: *amplitude*, *land/sea fraction weighting* and *trend*. The *amplitude* preprocessor calculates the peak-to-peak amplitude of periodic phenomena along an arbitrary coordinate. Usually, this coordinate is a temporal dimension. A common application 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

| Type | Name | Description |
|---|---|---|
| Preprocessor | *amplitude* | Amplitude of periodic phenomena (e.g. diurnal or seasonal 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 ordinary least squares 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.

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 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




(a) (b)

Figure 3.2.: (a) CMIP6 multi-model mean (MMM) of the global 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; elevation: 3397 m) from 1958 to 2014. The thick black line shows observations from the Mauna Loa Observatory (Keeling et al., 2005); the remaining colored lines show emission-driven historical simulations from individual CMIP6 models.

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 climate models 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. 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 December 2020). 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).

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.

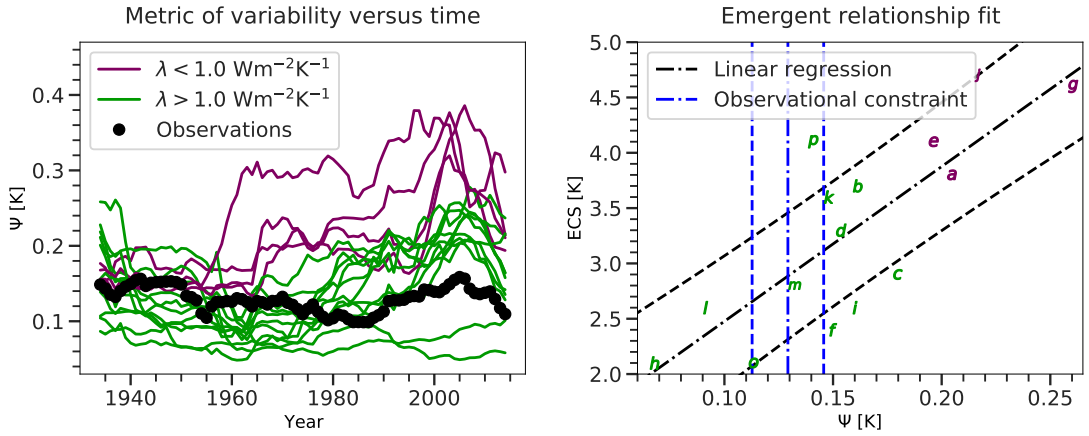
Due to an exhaustive overhaul of the ESMValTool version 2, older recipes from version 1 (Eyring, Righi, et al., 2016) cannot be used directly in the latest release. Porting recipes from the old version to the current version requires substantial changes on the recipes themselves, but also on the corresponding diagnostic scripts. As of December 2020, the author of this thesis has ported two recipes originally written by other authors to the new version of ESMValTool: *recipe_anav13jclim.yml* and *recipe_wenzel16jclim.yml*. *recipe_anav13jclim.yml* reproduces the analysis of Anav et al. (2013), who evaluate carbon cycle-related variables for climate models 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, Eyring, et al. (2016) to constrain future changes in the position of the austral jet stream. Due to its flexible code, the corresponding diagnostic

| Type | Name | Main reference |
|---|---|---|
| CMORizer | CRU | (Harris et al., 2014) |
| | CT2019 | (Jacobson et al., 2020) |
| | GCP | (Friedlingstein et al., 2019) |
| | 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 & Livesey, 2015) |
| | MTE | (Jung et al., 2011) |
| | NDP | (Gibbs, 2006) |
| | Scripps-$CO_2$ | (Keeling et al., 2005) |
| 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) |
| | *schlund20esd* | (Schlund, Lauer, et al., 2020) |
| | *schlund20jgr_gpp_abs_rcp85* | (Schlund, Eyring, et al., 2020) |
| | *schlund20jgr_gpp_change_1pct* | (Schlund, Eyring, et al., 2020) |
| | *schlund20jgr_gpp_change_rcp85* | (Schlund, Eyring, et al., 2020) |
| | *tcr* | (Gregory & Forster, 2008) |

Table 3.2.: Summary of new CMORization scripts for observational data and recipes 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*.

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_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 $\Psi$. Figure 3.3a shows $\Psi$ 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 $\Psi$ than low sensitivity models, an emergent relationship between ECS and $\Psi$ can be established (figure 3.3b).

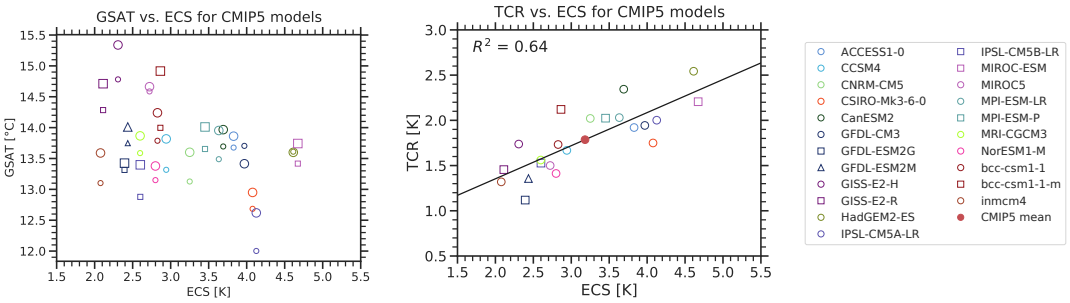



(a) (b)

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 $\Psi$ 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 $\Psi$. 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).

With observations of $\Psi$, ECS can eventually be constrained to $(2.8 \pm 0.6)$ K (66 % confidence range). More details on this including the calculation of $\Psi$ are given in section 5.1.1.

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.




(a) (b)

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.

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.

# 4. Assessment of Policy-Relevant Climate Metrics in CMIP6

TBA.

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

TBA.

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

### 5.1.1. COX

TBA.

# 6. Constraining Uncertainties in Future Gross Primary Productivity with Machine Learning

TBA.

## 6.1. Step 1: XXXXXXXXXXXXXXXXX

# 7. Summary and Outlook

TBA.

# Appendix

## A. TBA

### A.1. test

**test**

**hi** The ECS is really cool. I like it very much!

This is e.g. without an "at" and this is it with an "at" e.g. difference? Test space. Real dot!

E.g.blaa. E.g. blaaaa. i.e.blaaaa, i.e. blaa.

These are really cool papers: (Schlund, Eyring, et al., 2020; Schlund, Lauer, et al., 2020)

autocite: (Lauer et al., 2018)

cite: Lauer et al., 2010 (Anav et al., 2015) (Anav et al., 2013) (Allen & Ingram, 2002)

textcite: Lauer et al. (2010)

And this one, too: (Lauer et al., 2020)

Three authors: (Bao et al., 2020)

Many many authors: (Eyring et al., 2020)

```
input <iostream>
```

## B. TBA

TBA.

# List of Acronyms

**AerChemMIP** Aerosol Chemistry Model Intercomparison Project . . . . . 6

**AOGCM** Atmosphere-Ocean General Circulation Model . . . . . 3

**AR5** Fifth Assessment Report . . . . . 43

**AR6** Sixth Assessment Report . . . . . 38

**C4MIP** Coupled Climate-Carbon Cycle Model Intercomparison Project . . 6

**CESM** Community Earth System Model . . . . . 16

**CMIP** Coupled Model Intercomparison Project . . . . . 5

**$CH_4$** methane . . . . . 20

**CMOR** Climate Model Output Rewriter . . . . . 37

**CO** carbon monoxide . . . . . 20

**$CO_2$** carbon dioxide . . . . . 4

**CRE** cloud radiative effect . . . . . 17

**CV** cross-validation . . . . . 31

**DECK** Diagnostic, Evaluation, and Characterisation of Klima . . . . . 5

**ECS** Effective Climate Sensitivity . . . . . 15

**ENSO** El Niño-Southern Oscillation . . . . . 8

**ESGF** Earth System Grid Federation . . . . . 37

**ESM** Earth System Model . . . . . 4

**ESMValCore** core functionalities of the ESMValTool . . . . . 36

**ESMValTool** Earth System Model Evaluation Tool . . . . . 35

**GHG** Greenhouse Gas . . . . . 4

**GPP** Gross Primary Productivity . . . . . 21

**GSAT** global mean near-surface air temperature . . . . . 9

**GtC** gigatonnes of carbon . . . . . 19

**I/O** Input/Output . . . . . 36

**IAM** Integrated Assessment Models . . . . . 7

**IPCC** Intergovernmental Panel on Climate Change . . . . . 38

**JJA** June-July-August . . . . . 40

**LongRunMIP** Long Run Model Intercomparison Project . . . . . 17

**MDER** Multiple Diagnostic Ensemble Regression . . . . . 31

**MIP** Model Intercomparison Project . . . . . 5

**ML** Machine Learning . . . . . 1

**MMM** multi-model mean . . . . . 24

**NAO** North Atlantic Oscillation . . . . . 8

**NCAR** National Center for Atmospheric Research . . . . . 16

**netCDF** Network Common Data Form . . . . . 37

**PDF** probability density function . . . . . 26

**ppm** parts per million . . . . . 19

**RCP** Representative Concentration Pathway . . . . . 7

**RMSE** root-mean-square error . . . . . 30

**ScenarioMIP** Scenario Model Intercomparison Project . . . . . 6

**SSP** Shared Socioeconomic Pathway . . . . . 6

**TCR** Transient Climate Response . . . . . 17

**TOA** top of the atmosphere . . . . . 13

**WCRP** World Climate Research Programme . . . . . 5

**WGCM** Working Group on Coupled Modelling . . . . . 5

# 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. (2020). 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. (2020). 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

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

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

Anav, A., Friedlingstein, P., Beer, C., Ciais, P., Harper, A., Jones, C., Murray-Tortarolo, G., Papale, D., Parazoo, N. C., Peylin, P., Piao, S., Sitch, S., Viovy, N., Wiltshire, A., & Zhao, M. (2015). Spatiotemporal patterns of terrestrial gross primary production: A review. *Reviews of Geophysics*, *53*(3), 785–818. https://doi.org/10.1002/2015rg000483

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. (2020). 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. (2020). 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

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

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

Bao, Y., Song, Z. Y., & Qiao, F. L. (2020). FIO-ESM Version 2.0: Model Description and Evaluation. *Journal of Geophysical Research-Oceans*, *125*(6). https://doi.org/10.1029/2019JC016036

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

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

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

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., 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

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

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*.

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, 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

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

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

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

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

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

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., & Taylor, K. E. (2016). 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., 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., 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. (2016). 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., 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

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. (2016). 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

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. (2011). Earth system models: an overview. *Wiley Interdisciplinary Reviews: Climate Change*, *2*(6), 783–800. https://doi.org/10.1002/wcc.148

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

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., Jones, M. W., O'Sullivan, M., Andrew, R. M., Hauck, J., Peters, G. P., Peters, W., Pongratz, J., Sitch, S., Le Quere, C., Bakker, D. C. E., Canadell, J. G., Ciais, P., Jackson, R. B., Anthoni, P., Barbero, L., Bastos, A., Bastrikov, V., Becker, M., ... Zaehle, S. (2019). Global Carbon Budget

2019. *Earth System Science Data*, *11*(4), 1783–1838. https://doi.org/10.5194/essd-11-1783-2019

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

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

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., 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., 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

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

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

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

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

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

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.

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., 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

Knutti, R., Rugenstein, M. A. A., & Hegerl, G. C. (2017). 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. (2017). A climate model projection weighting scheme accounting for performance and interdependence. *Geophysical Research Letters*, *44*(4), 1909–1918. https://doi.org/10.1002/2016gl072012

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

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., Hamilton, K., Wang, Y. Q., Phillips, V. T. J., & Bennartz, R. (2010). The Impact of Global Warming on Marine Boundary Layer Clouds over the Eastern Pacific-A Regional Model Study. *Journal of Climate*, *23*(21), 5844–5863. https://doi.org/10.1175/2010jcli3666.1

Lauer, A., Jones, C., Eyring, V., Evaldsson, M., Stefan, H. A., Makela, J., Martin, G., Roehrig, R., & Wang, S. Y. (2018). Process-level improvements in CMIP5 models and their impact on tropical variability, the Southern Ocean, and monsoons. *Earth System Dynamics*, *9*(1), 33–67. https://doi.org/10.5194/esd-9-33-2018

Lauer, A., Eyring, V., Bellprat, O., Bock, L., Gier, B. K., Hunter, A., Lorenz, R., Pérez-Zanón, N., Righi, M., **Schlund**, **M.**, Senftleben, D., Weigel, K., &

Zechlau, S. (2020). Earth System Model Evaluation Tool (ESMValTool) v2.0 – diagnostics for emergent constraints and future projections from Earth system models in CMIP. *Geoscientific Model Development*, *13*(9), 4205–4228. https://doi.org/10.5194/gmd-13-4205-2020

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

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

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 CO2 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

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

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

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., 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

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

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

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., 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

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

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., 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

Sanderson, B. M., Knutti, R., & Caldwell, P. (2015b). Addressing Interdependency in a Multimodel Ensemble by Interpolation of Model Properties. *Journal of Climate*, *28*(13), https://journals.ametsoc.org/jcli/article-pdf/28/13/5150/4043271/jcli-d-14-00361_1.pdf, 5150–5170. https://doi.org/10.1175/JCLI-D-14-00361.1

**Schlund**, **M.**, Eyring, V., Camps-Valls, G., Friedlingstein, P., Gentine, P., & Reichstein, M. (2020). 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. (2020). 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

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

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

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

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

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

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. (2016). 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. (2016). 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

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

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

# 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