



# Assessing the uncertainty in climate sensitivity

Daniel Klocke



## Hinweis

Die Berichte zur Erdsystemforschung werden vom Max-Planck-Institut für Meteorologie in Hamburg in unregelmäßiger Abfolge herausgegeben.

Sie enthalten wissenschaftliche und technische Beiträge, inklusive Dissertationen.

Die Beiträge geben nicht notwendigerweise die Auffassung des Instituts wieder.

Die "Berichte zur Erdsystemforschung" führen die vorherigen Reihen "Reports" und "Examensarbeiten" weiter.



## Notice

*The Reports on Earth System Science are published by the Max Planck Institute for Meteorology in Hamburg. They appear in irregular intervals.*

*They contain scientific and technical contributions, including Ph. D. theses.*

*The Reports do not necessarily reflect the opinion of the Institute.*

*The "Reports on Earth System Science" continue the former "Reports" and "Examensarbeiten" of the Max Planck Institute.*

## Anschrift / Address

Max-Planck-Institut für Meteorologie  
Bundesstrasse 53  
20146 Hamburg  
Deutschland

Tel.: +49-(0)40-4 11 73-0  
Fax: +49-(0)40-4 11 73-298  
Web: [www.mpimet.mpg.de](http://www.mpimet.mpg.de)

## Layout:

Bettina Diallo, PR & Grafik

Titelfotos:

vorne:

Christian Klepp - Jochem Marotzke - Christian Klepp

hinten:

Clotilde Dubois - Christian Klepp - Katsumasa Tanaka

Assessing the uncertainty  
in climate sensitivity

Daniel Klocke

aus Aachen

Hamburg 2011

Daniel Klocke  
Max-Planck-Institut für Meteorologie  
Bundesstrasse 53  
20146 Hamburg  
Germany

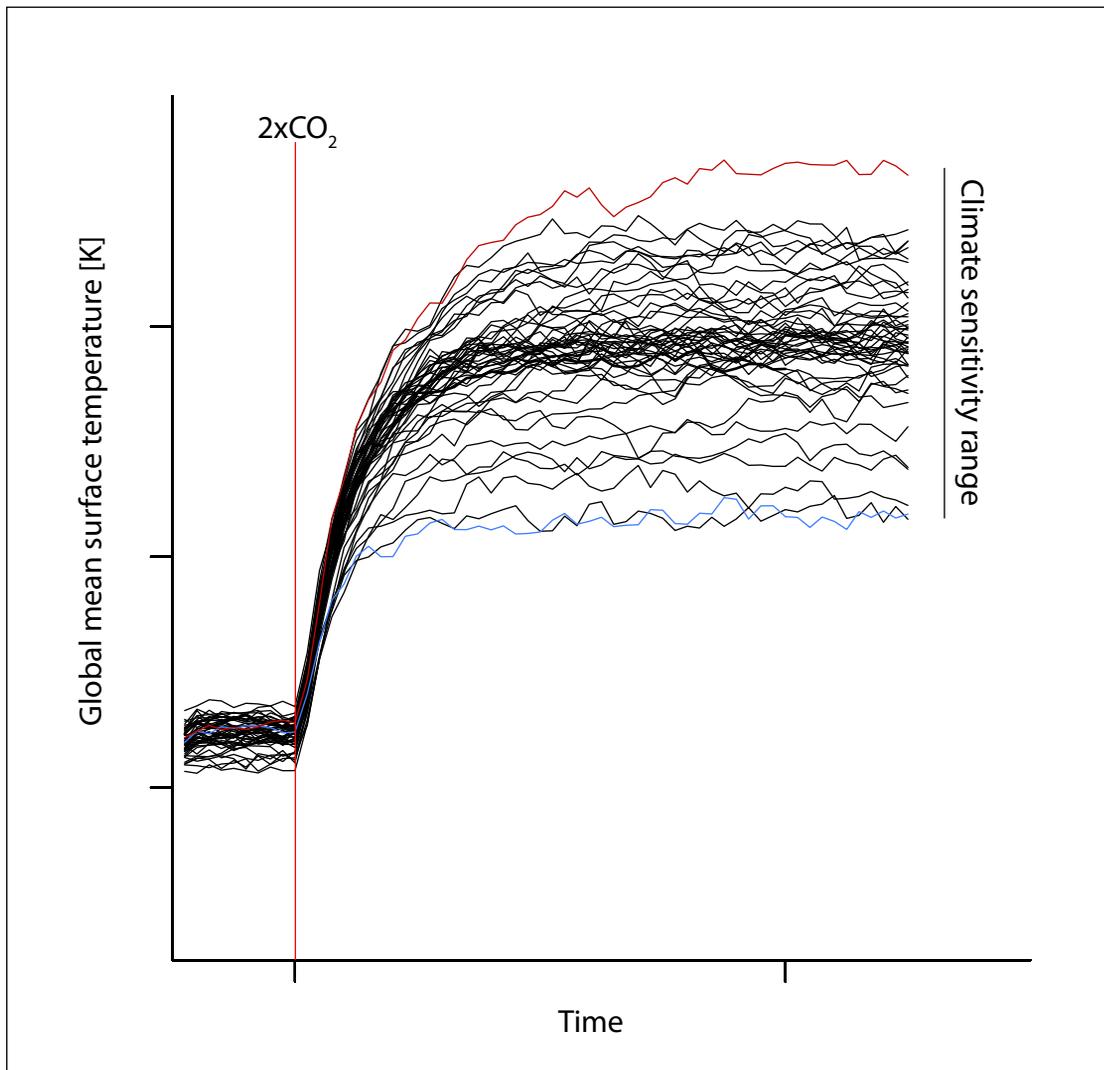
Als Dissertation angenommen  
vom Department Geowissenschaften der Universität Hamburg

auf Grund der Gutachten von  
Prof. Dr. Bjorn Stevens  
und  
Prof. Dr. Johannes Quaas

Hamburg, den 7. Juli 2011  
Prof. Dr. Jürgen Oßenbrügge  
Leiter des Departments für Geowissenschaften

# Assessing the uncertainty in climate sensitivity

---



Daniel Klocke

Hamburg 2011



## Abstract

Large uncertainties in estimates of the Earth's climate sensitivity remain. This dissertation explores and discusses several approaches to narrow the range in climate sensitivity estimates.

The use of volcanic eruptions to constrain climate sensitivity estimates is tested with ensemble simulations of the last millennium, with volcanic radiative forcing based on reconstructions. More than 45 strong eruptions, all larger than Mt. Pinatubo, or more than 10000 years of volcanic activity such as present in the past thousand years, would be necessary to obtain a range of climate sensitivity estimates that is narrower than the most recent IPCC estimates. This conclusion is based on the assumption that the response is directly related to equilibrium climate sensitivity, and in this way circumvents complications of linking the climate response following large volcanic events to climate sensitivity. If this assumption is relaxed, the number of volcanoes needed to obtain accurate estimates of climate sensitivity is likely even higher.

The uncertainty in climate sensitivity estimates may be analyzed in terms of the contributions of feedbacks from individual physical processes in the climate system using feedback quantification methods. Results of a feedback assessment depend on the method chosen and the period over which feedbacks are derived. The sampling errors can be large when averaging periods are too short. This is especially true for the cloud feedback, which is highly variable in space and time. Temporal variability, such as the year-to-year variability, can be as large as the inter-model spread in a given feedback from the CMIP3 multi-model ensemble. Even in our simulations, where many sources of natural variability are neglected, and where the forcing is rather homogeneous as exerted by the carbon dioxide concentration doubling, multi-year averages are necessary to get a reliable estimate of the simulated cloud feedback. Considering the large natural variability and relatively small, and spatially heterogeneous forcing present in the real world, this implies that using observations to constrain feedbacks, in particular the cloud feedback, is a challenging task.

Another option to better constrain climate sensitivity is to weigh model projections according to measures of model fidelity. This has failed so far, largely because climate sensitivity could not be related to aggregate measures of skill in current ensembles of model simulations. Here we show that measures of model fidelity that are effective at narrowing the distribution of future projections still may be a poor measure of the likelihood that a model will provide an accurate estimate of climate sensitivity. Two ensembles of climate model simulations are

considered in this analysis: an ensemble obtained by perturbing parameters in a single model and the CMIP3 multi-model ensemble, containing the majority of the worlds climate models. The single model ensemble reproduces many aspects of the multi-model ensemble, including the distributions of skill in simulating cloud related properties, the distribution of climate sensitivity, and the relation between the simulated climate sensitivities and cloud feedbacks. By restricting error measures to sub-tropical marine cloud regimes tighter relationships between climate sensitivity and model error can be identified, allowing to narrow the distribution of climate sensitivity in the simple ensemble. This relationship, however, does not carry into the multi-model ensemble. This suggests that model weighting based on statistical relationships alone is insufficient, and that structural errors in climate models are still too large to obtain reliable climate sensitivity estimates by attributing weights to individual models.

Finally, a possibility to more directly evaluate climate models with observations is outlined. The use of data assimilation allows to identify which fast processes lead to a drift from the observed state. This is done on short time scales rather than by evaluating long-term statistics of results of those processes. As fast processes are the largest contributor to the uncertainty in model-based estimates of climate sensitivity this is a promising way towards a better understanding of the climate system and a more reliable quantification of its response to perturbations.

# Contents

<b>1</b>	<b>Introduction</b>	<b>5</b>
1.1	Forcing, feedbacks and climate sensitivity . . . . .	5
1.2	Outline of Thesis . . . . .	6
<b>2</b>	<b>How many volcanoes would be necessary to better constrain climate sensitivity?</b>	<b>9</b>
2.1	Introduction . . . . .	10
2.2	Ensemble simulations of the last millennium . . . . .	11
2.2.1	Volcanic radiative forcing . . . . .	15
2.3	Composite forcing and response . . . . .	15
2.4	How many versus how strong volcanoes? . . . . .	17
<b>3</b>	<b>Assessment of different feedback metrics</b>	<b>21</b>
3.1	Introduction . . . . .	21
3.2	Methods to quantify feedbacks in GCMs . . . . .	24
3.3	Model and experimental set-up . . . . .	28
3.4	Feedback factors . . . . .	29
3.5	Feedback variability . . . . .	33
3.5.1	Spatial variability . . . . .	34
3.5.2	Temporal variability . . . . .	40
3.6	Implications and conclusions . . . . .	42
<b>4</b>	<b>On constraining estimates of climate sensitivity with present-day observations through model weighting</b>	<b>45</b>
4.1	Model error and climate sensitivity . . . . .	46
4.2	A simple ensemble spanning a range of errors and climate sensitivities	47
4.3	The simple ensemble as proxy for the multi-model ensemble . . .	50
4.4	Developing measures of model error linked to climate sensitivity .	51
4.5	Implications for weighting projections from multi-model ensembles	54
<b>5</b>	<b>Towards the use of data assimilation for process-oriented climate model evaluation and climate sensitivity constraints</b>	<b>57</b>
5.1	Introduction . . . . .	58

## CONTENTS

5.2	Technical notes on the integration of ECHAM in a data assimilation/forecast cycle . . . . .	60
5.2.1	Modifications to the model restart . . . . .	61
5.2.2	The interface to DART . . . . .	64
5.2.3	Data assimilation/forecast cycle work flow . . . . .	65
5.3	Assimilation increments as a skill measure . . . . .	66
5.4	First results . . . . .	67
5.5	Outlook and research questions . . . . .	71
<b>6</b>	<b>Conclusion and Outlook</b>	<b>73</b>
	<b>Bibliography</b>	<b>77</b>
	<b>Acknowledgements</b>	<b>87</b>

# Chapter 1

## Introduction

**How sensitive is the Earth's climate system to perturbations?** To accurately answer this question we require a detailed understanding of many interacting physical, chemical and biological processes that act on various time and space scales. Despite large advances in all involved fields, crucial parts of the climate system remain poorly understood, and the uncertainty in estimates of climate sensitivity is therefore large.

The complex problem of quantifying the sensitivity of the Earth's climate is generally approached with a measure called *equilibrium climate sensitivity*: the change in global mean surface temperature following a doubling in atmospheric carbon dioxide (CO<sub>2</sub>) concentrations. This simple and useful measure represents a fundamental characteristic of the climate system with which many other aspects of climate change scale. Its value depends crucially on the different feedbacks that take place within the climate system in response to an imposed climate forcing. Obtaining a better estimate of climate sensitivity requires a good understanding of the interplay between forcing, response and feedbacks.

### 1.1 Forcing, feedbacks and climate sensitivity

The concept of forcing, feedbacks and climate sensitivity describes the response of the climate system to changes in boundary conditions which affect the radiation budget of the Earth. Such a change is called radiative forcing ( $\Delta F$ ) and can be initiated naturally, for instance by changes in the solar irradiation or by volcanic eruptions, or due to an anthropogenic impact, such as changes in atmospheric CO<sub>2</sub> concentrations. The radiative forcing leads to a change in net energy flux,  $\Delta R$ , into the climate system, measured as the difference of incoming and outgoing radiation at the top-of-atmosphere (ToA). The mean temperature of the climate system adjusts, as to restore the radiation balance at the ToA. A positive (negative) forcing leads to a warming (cooling), so that the atmosphere emits more (less) energy out to space in the infrared part of the electromagnetic spectrum.

The concept of forcing, feedbacks and climate sensitivity is summarized in the following energy budget equation:

$$\Delta R = \Delta F - \lambda \Delta T_S \quad (1.1)$$

where  $\lambda$ , the feedback parameter, represents all processes that change their radiative impact following a change in the surface temperature ( $\Delta T_S$ ). These feedback processes can amplify or dampen the initial perturbation and comprise the largest uncertainties in estimating the climate system's sensitivity. The concept of *equilibrium climate sensitivity* is introduced when Equation 1.1 is applied to two stationary climate states, where one has doubled atmospheric CO<sub>2</sub> concentrations.

The first estimates of equilibrium climate sensitivity were based on calculations by Arrhenius (1896) (5.5 K) and Callendar (1938) (2 K). More recently, comprehensive numerical models of the Earth system are employed that can be integrated under constant boundary conditions until a new stationary state is reached. However, the range in equilibrium climate sensitivity estimates from ensembles of climate models, often treated as the uncertainty in climate sensitivity, has not changed much from the first climate model results of Charney (1979) to the state-of-the-art climate models: giving a range of 2.1 to 4.4 K (Randall et al. 2007). Although equilibrium climate sensitivity is a useful metric to compare and test climate models, because it is relatively easy to derive, extrapolating this concept to the real world has proven difficult for various reasons (Knutti et al. 2008).

First, the climate system is driven by radiative forcings from different components of the system which change continuously. The different forcing components interfere and it is difficult to quantify them. Second, the inertia of the world's oceans buffer the forcing imposed, leading to long time scales (from decades to centuries) that are necessary to approach a stationary state. Finally, because of observational uncertainties the detection of a response to a comparably small forcing, when compared to the response to a forcing induced by instantaneously doubled CO<sub>2</sub> concentrations, remains a challenging task.

## 1.2 Outline of Thesis

This thesis uses the above concepts to discuss in four self-contained chapters the challenges in obtaining a narrower range in estimates of equilibrium climate sensitivity.

In Chapter 2 ensemble simulations of the last millennium, with boundary conditions derived from reconstructions, are used to test how many volcanoes are

necessary to obtain narrower estimates of climate sensitivity. In this ensemble of model integrations the stratospheric aerosol forcing due to many volcanic eruptions is imposed, with each eruption being different in magnitude and location, and each forcing being imposed at a different climate state. Climate sensitivity is derived by assuming that it relates directly to the response of the climate system to volcanic forcing. It is discussed how many volcanoes, and with what intensity of radiative forcing, are necessary to obtain a range of climate sensitivity estimates, comparable to the one given by the multi-model ensemble of the Coupled Model Inter-comparison Project phase-3 (CMIP3; Meehl et al. 2007). Parts of this chapter are already published in *Climate of the Past*<sup>1</sup> and are reproduced with adjustments to serve the purpose of this chapter. The study presented in this chapter is in preparation for submission to *Geophysical Research Letters*.

Chapter 3 is dedicated to the feedback parameter  $\lambda$  and its different components. A variety of methods have been developed to isolate specific feedback mechanisms in climate models, raising the question as to how sensitive the results of such analyses are to the methods employed, which is explored here. Furthermore, the spatial and temporal variability of each physical feedback process is analyzed to estimate the averaging time necessary to minimize the sampling error. This chapter will be submitted to *Climate Dynamics*<sup>2</sup>.

Chapter 4 constructs a single-model perturbed-physics ensemble where only cloud parameterization parameters are perturbed and demonstrates how this ensemble reproduces many aspects of the CMIP3 multi-model ensemble. For this perturbed-physics ensemble, an observational constraint on climate sensitivity is derived, which, however, is not applicable to the multi-model ensemble. What this implies for the interpretation of results from multi-model ensembles such as used for the IPCC report is discussed in detail. This chapter is re-submitted after revisions to the *Journal of Climate*<sup>3</sup> and is reproduced with editorial adjustments.

Chapter 5 gives a detailed outlook on possible next steps to evaluate models and potentially link processes to climate sensitivity by employing data assimilation. A technical description is given and first results are shown, which show potential

---

<sup>1</sup>Jungclaus, J. H., S. J. Lorenz, C. Timmreck, C. H. Reick, V. Brovkin, K. Six, J. Segschneider, M. A. Giorgetta, T. J. Crowley, J. Pongratz, N. A. Krivova, L. E. Vieira, S. K. Solanki, D. Klocke, M. Botzet, M. Esch, V. Gayler, H. Haak, T. J. Raddatz, E. Roeckner, R. Schnur, H. Widmann, M. Claussen, B. Stevens, and J. Marotzke, 2010: Climate and carbon-cycle variability over the last millennium. *Climate of the Past*, 6, 723–737, doi:10.5194/cp-6-723-2010.

<sup>2</sup>Klocke, D., J. Quaas, M. Giorgetta, B. Stevens, Assessment of different feedback metrics, *to be submitted to Climate Dynamics*, 2011.

<sup>3</sup>Klocke, D., R. Pincus, J. Quaas, On constraining estimates of climate sensitivity with present-day observations through model weighting, *re-submitted after revisions to Journal of Climate*, 2011.

## CHAPTER 1 INTRODUCTION

for future use in the spirit of the problems outlined in the previous chapters. A summary and conclusion of the main findings as well as an outlook are given in Chapter 6.

## Chapter 2

# How many volcanoes would be necessary to better constrain climate sensitivity?<sup>1</sup>

In this chapter the possibility of better constraining climate sensitivity from volcanic forcings is explored by using ensemble simulations of the last millennium with a comprehensive earth system model, including all known forcings. Circumventing the complications associated with deriving climate sensitivity from volcanic eruptions, we assume here that the response to the short pulse forcing is linearly related to climate sensitivity. We use the ensemble mean of many different volcanic forcing realizations in many different states of the climate system to estimate for various thresholds of forcing intensity how many volcanic eruptions would have to be observed to yield an estimate of climate sensitivity that is narrower than the one currently obtained from multi-model ensembles. Few volcanoes with a strong radiative forcing are better suited for deriving a stronger composite normalized response than many weak volcanic events. To obtain a comparable range of climate sensitivities as given by the multi-model ensemble of CMIP3, more than 45 volcanoes, all larger than Mt. Pinatubo (the eruption with the largest radiative impact in the last century) are needed. If the frequency of large volcanic eruptions remains compa-

---

<sup>1</sup>Part of this chapter is already published in: Jungclaus, J. H., S. J. Lorenz, C. Timmreck, C. H. Reick, V. Brovkin, K. Six, J. Segschneider, M. A. Giorgetta, T. J. Crowley, J. Pongratz, N. A. Krivova, L. E. Vieira, S. K. Solanki, D. Klocke, M. Botzet, M. Esch, V. Gayler, H. Haak, T. J. Raddatz, E. Roeckner, R. Schnur, H. Widmann, M. Claussen, B. Stevens, and J. Marotzke, 2010: Climate and carbon-cycle variability over the last millennium. *Climate of the Past*, 6, 723–737, doi:10.5194/cp-6-723-2010. The results most relevant to this chapter - i.e. concerning the climate sensitivity constraint, are still unpublished and are in preparation for submission as Klocke et al. to Geophysical Research Letters.

rable to the last millennium, more than 10000 years are necessary to derive a tighter range of climate sensitivity from responses following volcanic eruptions, that is, assuming that forcing and response relate linearly to equilibrium climate sensitivity.

## 2.1 Introduction

The metric commonly used to measure the response of the earth climate system to an external forcing is equilibrium climate sensitivity, the change of the global mean surface temperature after doubling  $\text{CO}_2$ . This metric is generally used for climate models, which can be integrated until a new equilibrium is reached (Bony et al. 2006; Randall et al. 2007). Over the last decades every new generation of multi-model ensembles produced comparable ranges of climate sensitivity (for the current generation of climate models the range is 2.1 to 4.4 K; Randall et al. 2007). Different approaches have been explored to estimate climate sensitivity from observations of the changing climate over the instrumental period since the beginning of the industrialization (e.g. Forster and Gregory 2006; Gregory et al. 2002), but uncertainties in the forcing and in measuring the global mean surface temperature change make it so far impossible to confine the range of climate sensitivity (Wigley et al. 1997).

One possibility to observe the climate response to an external forcing are volcanoes. Volcanoes can strongly perturb the climate with short (negative) pulse forcings. Large volcanoes inject  $\text{SO}_2$  and other sulphuric gases into the stratosphere. Those gases oxidize and can form sulphate aerosols, which impact the radiation budget by reflecting short-wave radiation back to space and to a lesser extent warm the atmosphere through absorption of long-wave radiation. The resulting net effect is a reduction in top-of-atmosphere (ToA) downward radiative fluxes, leading to a cooling of the climate system. This negative forcing to the climate system can dominate all other forcings on short time-scales so that climate variability over the last millennium was dominated by large volcanic events (Jungclaus et al. 2010, see also Figure 2.1b and 2.2).

Volcanic radiative forcing is distinct from other natural forcings to the climate system. The forcing is short lived (a few years), with a strong and rather heterogeneous distribution in space and time, leading to characteristics different from the climate system's response to greenhouse gas forcing. On short time scales the system never approaches an equilibrium and feedbacks may differ from the ones expected at equilibrium. Further, observations from space are contaminated and uncertainties are larger, when volcanic aerosols are present in the atmosphere, and unforced variability is large on various time-scales leading to difficulties in

## 2.2 ENSEMBLE SIMULATIONS OF THE LAST MILLENNIUM

measuring the response of the climate system to volcanoes alone. Hansen et al. (1997) argue that the Earth system’s sensitivity to volcanic eruptions is not a good measure of equilibrium climate sensitivity and that volcanoes do not provide a good constraint. Important feedbacks that determine equilibrium climate sensitivity do not come into play, as the equilibrium response is not fully developed on the short time scale of the pulse forcing (Hansen et al. 1984). In this study, this problem is circumvented by simply assuming a linear relationship, and its influence on the conclusions will be discussed.

The response to volcanic forcing is relatively weak, because a large portion is quickly buffered by heat release of the upper ocean and the nonlinear relation of the response to equilibrium climate sensitivity hampers constraining climate sensitivity estimates (Wigley et al. 2005; Boer et al. 2007). Nevertheless, the observed response to a volcanic eruption can be usefully compared to climate model simulations, in terms of evaluating fast responses (Soden et al. 2002; Yokohata et al. 2005; Bender et al. 2010).

A further difficulty in analyzing responses to volcanic forcing arises when separating the weak surface temperature signal from the noise due to natural variability. To get a clear response independent of the state of the system one could use a composite volcanic response (e.g. Hansen et al. 1997; Mass and Portman 1989), but volcanic events with significant impact on ToA radiation are seldom and the only major volcanic eruption in the satellite era with reliable global satellite observation was Mt. Pinatubo in June 1991.

Here we make use of the large number of volcanic forcings generated with an ensemble of millennium-timescale simulations with one climate model. We estimate, for different thresholds of eruption strength, how many volcanoes would be sufficient to statistically separate the response from the background variability.

We describe the model simulations of the last millennium in section 2.2, which include hundreds of volcanoes exerting forcings differing in strength, location, and - through the use of a large ensemble - also in the climate state they perturb. In section 2.2.1 we describe the calculation of radiative forcing from effective radius and aerosol optical depth time series. This is followed by creating a composite volcano whose influence on the simulated climate is described in section 2.3. We conclude this chapter by estimating how many volcanoes of a certain strength would lead to tighter estimates of climate sensitivity.

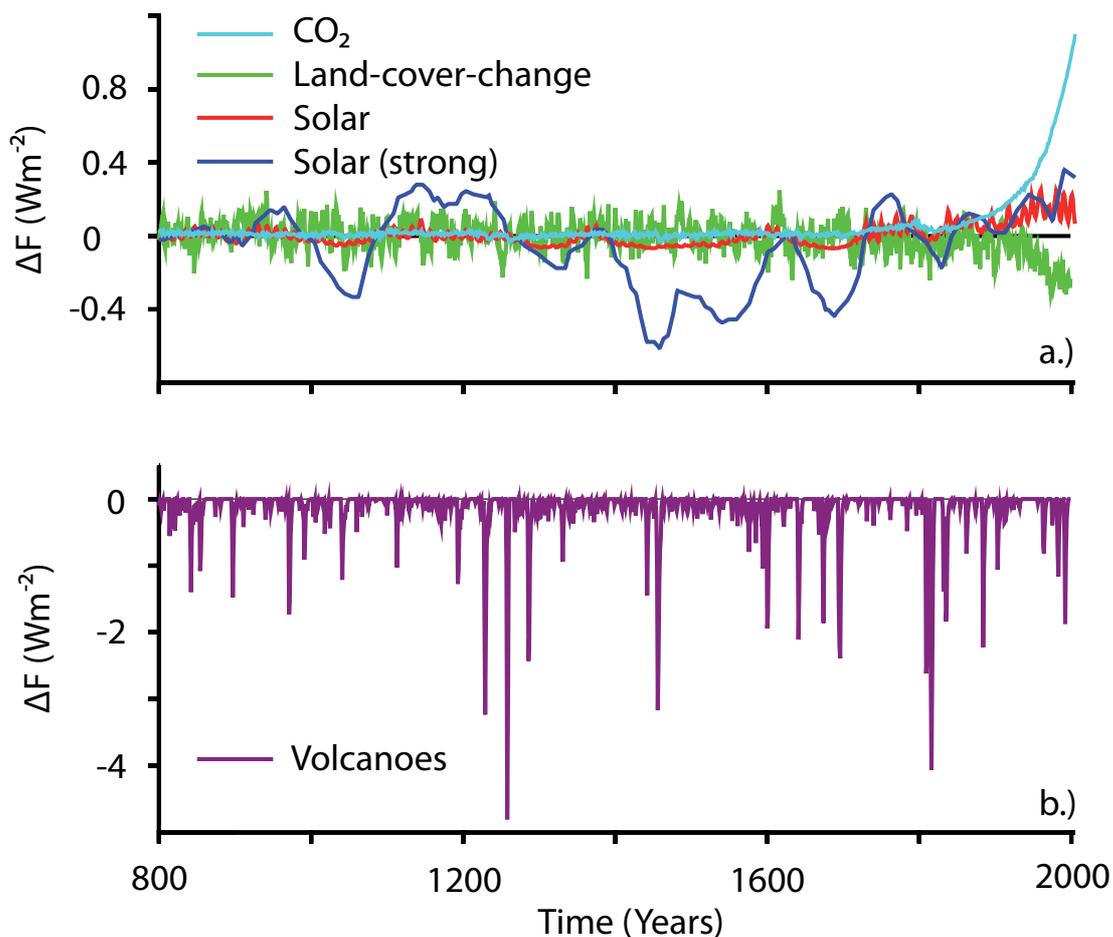
## 2.2 Ensemble simulations of the last millennium

This study is based on ensemble simulations of the last millennium (from 800 – 2000 AD) with an Atmosphere-Surface-Ocean-Biogeochemistry earth system

model as conducted by Jungclaus et al. (2010) at the Max Planck Institute for Meteorology (MPI-M). The earth system model consists of the atmospheric component ECHAM5 (Roeckner et al. 2003), the ocean model MPIOM (Marsland et al. 2003), and modules for land vegetation (JSBACH; Raddatz et al. 2007) and ocean bio-geochemistry (HAMOCC; Wetzel et al. 2006) including the full interactive carbon cycle. The atmosphere model ECHAM5 is run at a spectral resolution of T31 ( $\sim 3.75^\circ \times 3.75^\circ$  spatial resolution), with 19 vertical levels and a model top at 10 hPa. The ocean model MPIOM uses a conformal mapping grid with a horizontal grid spacing of  $3.0^\circ$  and 40 unevenly spaced vertical levels.

The ensemble consists of one 3000-year control simulation with constant 1860 boundary conditions, eight simulations which are forced with reconstructions of the last millennium and one simulation forced with only volcanic aerosol reconstructions. The forced ensemble member simulations are started from different initial conditions derived from the control simulation. The external forcing data sets consist of the total solar irradiation (TSI), volcanic forcing considering aerosol optical depth (AOD) and effective radius ( $R_{\text{eff}}$ ) distributions, land use change, and anthropogenic greenhouse gas and aerosol emissions.. Five ensemble members are forced with a weaker TSI variability of 0.1% of the standard TSI value of  $1367 \text{ Wm}^{-2}$  between the Maunder minimum and present-day TSI, while three ensemble members are forced with a higher TSI variability of 0.25% (Figure 2.1a). All other external forcings are identical between the eight ensemble members. This gives in total nine ensemble members with the identical volcanic forcing, each run for 1200 years.

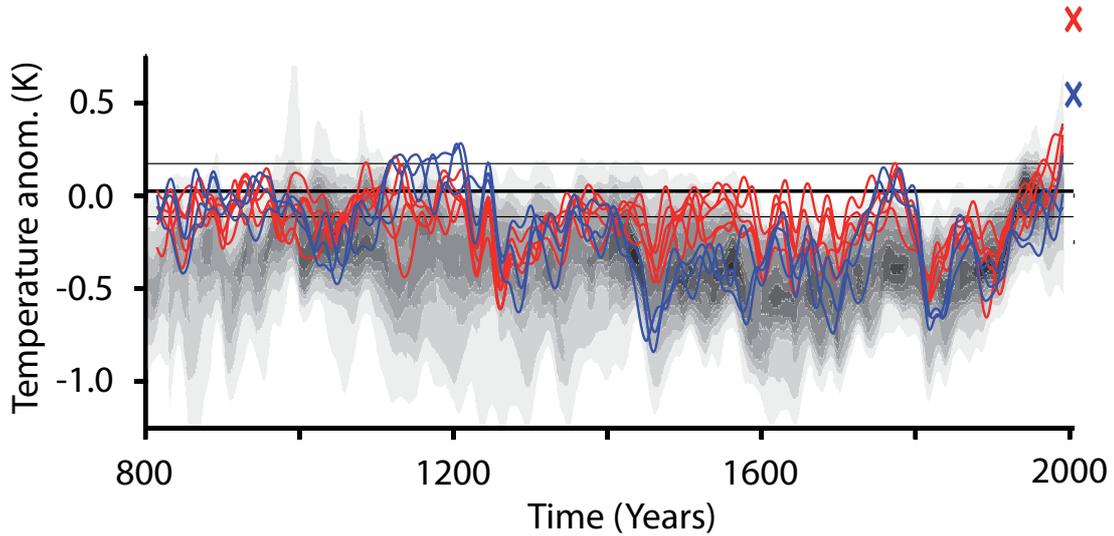
The volcanic radiative effects are taken into account by time series of AOD at  $0.55 \mu\text{m}$  and  $R_{\text{eff}}$ , which are included in the on-line radiative transfer calculations when volcanic aerosols are present. The data is specified on three zonal bands for the northern hemisphere (north pole to  $30^\circ\text{N}$ ), tropics ( $30^\circ\text{N}$  to  $30^\circ\text{S}$ ) and southern hemisphere ( $30^\circ\text{S}$  to south pole) with a time resolution of ten days. AOD estimates (Crowley et al. 2008) are based on a correlation between sulfate concentrations found in Antarctic ice cores and satellite observations after the Mt. Pinatubo eruption (Sato et al. 1993). For each eruption  $R_{\text{eff}}$  growth and decay rates are based on satellite observations of Mt. Pinatubo (Sato et al. 1993). Eruptions with a global-mean AOD exceeding 0.2 (Mt. Pinatubo AOD was about 0.15) are empirically scaled by comparison with the theoretical calculations for very large eruptions (Pinto et al. 1989). In the vertical, AOD is spread over three stratospheric model levels between 20 and 86 hPa, with a maximum AOD at 50 hPa. Sensitivity experiments for the model response to the Mt. Pinatubo eruption in June 1991 yield an average global mean surface temperature response (about 0.4 K for a global monthly mean net forcing of  $-2.33 \text{ Wm}^{-2}$ ) comparable to observations. Timmreck et al. (2009) have shown for the 1258 “unknown”



**Figure 2.1:** Radiative forcing ( $\Delta F$ ,  $\text{Wm}^{-2}$ ) at the top-of-atmosphere displayed as annual means (a) for the greenhouse-gas forcing ( $\text{CO}_2$ , light blue), land-cover change (albedo effect only, green), and solar forcing (red for the standard simulations and dark blue for simulations with enhanced solar variability), and (b) for volcanic forcing displayed with a different axis. Anomalies from solar irradiation and  $\text{CO}_2$  variations are calculated w.r.t. their pre-industrial control mean ( $1367 \text{ Wm}^{-2}$  and  $280.02 \text{ ppm}$ , respectively). The radiative forcing from volcanic aerosol injections and land-cover changes are calculated off-line relative to a control year (after Jungclaus et al. 2010).

volcanic eruption that a shift of the volcanic aerosol size distribution toward larger particles reduces the cooling effect for large eruptions by an increase of the long-wave radiative forcing and improves the consistency with temperature reconstructions. For details of the other forcing data sets, we refer to Jungclaus et al. (2010) and the references therein, as they are not essential to this study.

The comparison of the impact of different external forcings in the “millennium” experiments yields that the variability is dominated by volcanic eruptions



**Figure 2.2:** Evolution of simulated temperature over the last 1200 years: Northern Hemisphere 2m land temperature anomalies w.r.t. the 1961–1990 mean for weak solar forcing (red) and strong solar forcing (blue) in comparison with the range of reconstructions (gray scale, redrawn from Jansen et al. 2007). Black horizontal lines indicate the control experiment mean and its 5th–95th percentile range. Time series are smoothed by a 31-yr running mean. Crosses at the right axis denote the ensemble means (annual average) at the end of the simulation (2005) (after Jungclaus et al. 2010).

decreasing the net radiation at the top-of-atmosphere (Figure 2.2; Jungclaus et al. 2010). The most severe eruption occurred in 1258, with a global monthly mean net radiative forcing of  $-5.13 \text{ Wm}^{-2}$ , and the second most severe was the Tambora eruption in 1815 with a net radiative forcing of  $-4.85 \text{ Wm}^{-2}$ , both leading to a distinct cooling in all ensemble members (Figure 2.2). The 1991 eruption of Mt. Pinatubo was the strongest volcanic eruption in recent decades, but only the 14<sup>th</sup> largest eruption in the simulations used here. In the 1200 simulated years, 66% of the time volcanic aerosols are present in the atmosphere, with a mean volcanic forcing over this period of  $-0.16 \text{ Wm}^{-2}$ , leading to a colder mean climate in the forced simulations in comparison to the control simulation indicated by the black horizontal line in Figure 2.2. Longer cool periods like in the 19<sup>th</sup> century (Figure 2.2) are caused by clusters of larger volcanic events leading to a mean forcing of  $-0.31 \text{ Wm}^{-2}$  for this century, while the preceding 18<sup>th</sup> century with the weakest century-averaged forcing of  $-0.04 \text{ Wm}^{-2}$  was warmer in all ensemble members. The averaged volcanic forcing exceeds all other forcing in the preindustrial period

## 2.3 COMPOSITE FORCING AND RESPONSE

in the ensemble simulations at any time except for the stronger solar forcing in some periods and the anthropogenic greenhouse gas forcing towards the end of the 20<sup>th</sup> century (Figure 2.1a).

### 2.2.1 Volcanic radiative forcing

In the simulations described in section 2.2, AOD and  $R_{\text{eff}}$  are prescribed in the radiative flux calculations, similar to the TSI variability and land cover changes. From the model output only the radiative effect, but not the radiative forcing can be derived. The diagnostic ToA radiative fluxes also include responses by other radiation relevant processes to the presence of volcanic aerosols. The effective radiative forcings (Figure 2.1a, b) are calculated off-line with the isolated ECHAM5 radiative transfer code following the Wetherald and Manabe (1988) approach for calculating radiative feedbacks. The radiative forcing at the top-of-atmosphere is defined as the change in radiative fluxes at the top-of-atmosphere due to the change in one single variable  $x$ . All other variables are taken from one reference year of the control simulation and do not change from year to year. The radiative flux calculation is done once without the presence of volcanic aerosols and once including them. The difference of those two calculations yields the change in radiative fluxes due to the presence of volcanic aerosols alone.

The forcings here are all defined instantaneously and do not allow for any atmospheric adjustment. For  $\text{CO}_2$  concentration, the forcing is calculated in bins of  $\text{CO}_2$  changes with respect to the average  $\text{CO}_2$  concentrations of the control run (280.08 ppm). Land-cover-change related radiative forcing reflects only the effect of changing surface albedo and is calculated relative to the period of 800 to 850 AD from the experiment with land-cover-changes as the only forcing. Volcanic forcing introduces the strongest disturbances in terms of amplitude, but these are short-lived events. If the volcanic forcing time series was smoothed by, for example, a 30-year running mean, the amplitude would be of similar magnitude as the other forcings. However, the volcanic forcing is, in fact, concentrated in individual years rather than spread out over decades to centuries (see Figure 2.1a).

## 2.3 Composite forcing and response

The seasonal cycle of each global monthly mean quantity in every ensemble member time series is removed to isolate the volcano signals. For every individual volcano the peak forcing ( $F_{\text{max}}$ ) is determined and two years before and ten years after  $F_{\text{max}}$  are selected for the composite volcano. To further separate the signal from the long term natural variability, the average over three years preceding the

beginning of each volcanic event is subtracted and to avoid interfering volcanoes, only eruptions are selected without other volcanic events three years before and five years after  $F_{\max}$ . In the last five years of the composite interference of volcanoes with an  $F_{\max}$  of 25% of the selected volcano are allowed in order keep a large sample size. The composite forcing, ToA radiation imbalance, surface temperature response and ocean heat release are shown in Figure 2.3, all normalized by the peak forcing before averaging to make individual volcanoes comparable.

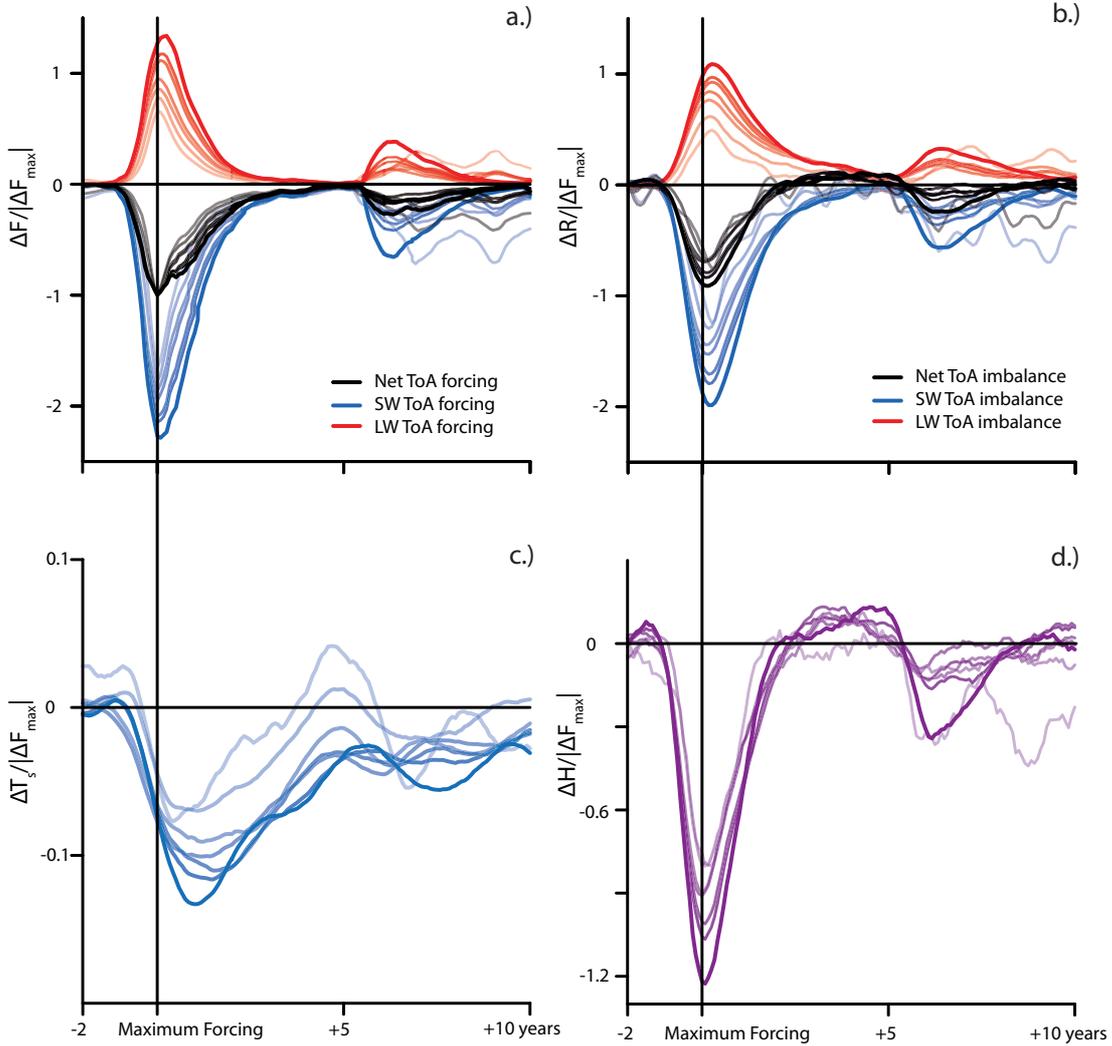
A threshold for the minimum absolute forcing ( $F_{\min}$ ) is defined for volcanoes considered in the composites in order to obtain a clean signal. All volcanoes never exceeding  $-0.1 \text{ Wm}^{-2}$  in monthly global mean net forcing are excluded from the analysis. Weaker eruptions do not, even averaged over a large sample size, separate from the natural variability. If included in the analysis, the weak volcanoes increase the noise, delude the signal and make the interpretation harder. In Figure 2.3 different choices of the minimum threshold are indicated by the color strengths. Table 2.1 gives the corresponding number of volcanoes used to derive the composite response for the different thresholds. The normalized response strength and associated time scale with the response depends on the choice of the threshold.

The net ToA radiative forcing builds up within about one year and decays over about five years back to zero (Figure 2.3a). The short-wave component of the forcing is about twice as large as the long-wave component, which is of opposite (positive) sign. The ToA radiative imbalance in Figure 2.3b is comparable to the forcing. The absolute magnitude of each component is smaller indicating contributions to the radiative fluxes at ToA by the response of the system via physical feedbacks. This is most prominent in the LW component which approaches the stationary state slower than the LW forcing. The volcanic aerosol effect decreases, but the atmospheric state is altered to adjust to the perturbation. The decrease in global mean surface temperature leads to less outgoing LW radiation and results in a positive net ToA radiative flux anomaly three to five years after  $F_{\max}$  of up to  $\sim 10\%$  of  $F_{\max}$ .

The surface temperature response in Figure 2.3c lags the forcing by 12 to 24 months depending on the strength of the forcing and adjusts back to the equilibrium temperature slower than the radiative flux perturbations decays. The forcing is damped by ocean heat release resulting in a maximal global mean surface temperature anomaly of  $0.13 \text{ K (Wm}^{-2}\text{)}^{-1}$  in the composite, when allowing only forcings larger  $-3.0 \text{ Wm}^{-2}$ . This is only a fraction of the equilibrium climate response of  $0.86 \text{ K (Wm}^{-2}\text{)}^{-1}$  for this model version.

The ocean responds immediately to the forcing by releasing heat to the atmosphere that is of the same order of magnitude as the forcing (Figure 2.3d), even exceeding the forcing for the largest eruptions. This delays and weakens the

### 2.3 COMPOSITE FORCING AND RESPONSE



**Figure 2.3:** Composite of the temporal evolution (monthly mean values with the seasonal cycle removed) of the top-of-atmosphere radiative forcing (a), radiative imbalance (b), surface temperature response (c) and ocean heat release  $H$  (d) from ten ensemble simulations of the last millennium. All quantities are normalized by the maximum forcing of each volcano before averaging. Color coding indicates the threshold of the minimum forcing. Lightest color considers all volcanoes with a forcing larger than  $-0.1 \text{ Wm}^{-2}$  and then the threshold increases in steps of  $0.5 \text{ Wm}^{-2}$  from  $-0.5$  to  $-3.0 \text{ Wm}^{-2}$ .

surface temperature response. Following the ToA net radiation imbalance, the ocean heat release turns into an ocean heat uptake from three to five years after  $F_{\text{max}}$  in the composite of up to  $\sim 10\%$  of the maximum heat release.

## 2.4 How many versus how strong volcanoes?

As discussed in section 2.1, estimating equilibrium climate sensitivity from volcanic eruptions has proven difficult. Here, we simply assume that the transient climate response after volcanic eruptions linearly relates to the equilibrium climate sensitivity, and ask whether even in such a simplified framework volcanoes could help constrain climate sensitivity.

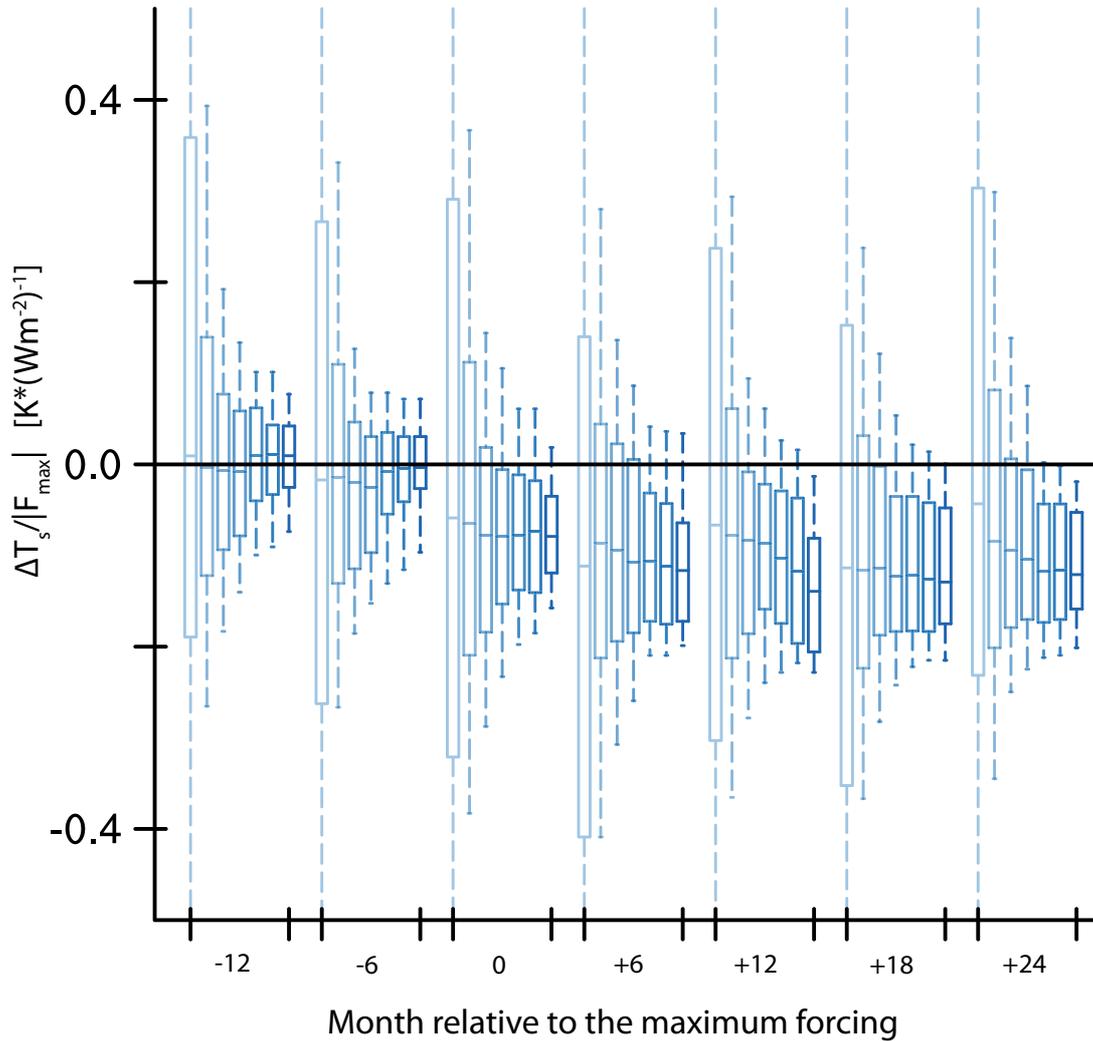
Figure 2.3 shows that few large volcanic events give a stronger normalized average response than if many smaller events are included in the composite response. Furthermore, the strongest surface temperature response is reached later after  $F_{\max}$  if the composite only includes larger eruptions. This behavior is also illustrated in Figure 2.4 which shows statistics for seven minimum forcing thresholds ( $F_{\min}$ ) given in Table 2.1.

Figure 2.4 shows how responses converge, when the forcing is large in contrast to many samples of weaker volcanoes. 18 months after  $F_{\max}$  the medians of different  $F_{\min}$  agree best. We linearly relate this median normalized response to the equilibrium climate sensitivity of the ECHAM5 model (3.27 K). The derived scaling factor is also used to scale the upper and lower quartile of the responses to obtain a range of “climate sensitivity” estimates, shown in Table 2.1. By definition the mean of the range is close to the equilibrium climate sensitivity of the model, but the range depends on the characteristics of the response. This in turn depends on the location of the volcanoes, the strength of the forcing and the state of the climate system.

Considering only volcanoes with an absolute net forcing larger than  $3.0 \text{ Wm}^{-2}$  gives a range of “equilibrium climate sensitivity” closest to, but still larger than the range of estimates from the CMIP3 climate model ensemble (2.1 - 4.4 K; Randall et al. 2007). This estimate is based on a sample of 45 volcanoes with five different characteristics, perturbing the climate in nine different states for each eruption. Allowing for weaker forcing to increase the sample size leads to a larger range of responses, especially if considering volcanoes with an absolute forcing smaller than  $0.5 \text{ Wm}^{-2}$  (see also Table 2.1).

Given the frequency of occurrence of volcanic eruptions with climate impact as in the last 1200 years, more than 10000 years, or alternatively about 45 volcanoes that have an impact larger than Mt. Pinatubo, would be needed to derive a range in estimated climate sensitivity comparable to the current range of climate sensitivities provided by IPCC. Even if such a large sample of volcanoes was to occur, the fact that climate sensitivity does not scale linearly with the following response, and observations carry many uncertainties, it seems unlikely that this range can be further narrowed with the use of volcanoes in the future. We conclude that the climate system’s response following volcanic eruptions is a poor

## 2.4 HOW MANY VERSUS HOW STRONG?



**Figure 2.4:** Statistics of the global monthly mean surface temperature ( $\Delta T_s$ ) response after volcanic eruptions normalized by the maximum forcing ( $F_{\max}$ ) of each volcano for 12 and 6 months before the peak forcing and 6, 12, 18 and 24 months after the peak forcing. The box center line indicates the median, the box boundaries the upper and lower quartile and the whiskers show the upper and lower deciles. Color intensity indicates the minimum global mean volcanic forcing threshold, increasing with color intensity (-0.1, -0.5, -1.0, -1.5, -2.0, -2.5, -3.0  $\text{Wm}^{-2}$ , in accordance with decreasing the sample size (see Table 2.1))

surrogate for climate change.

## CHAPTER 2 HOW MANY VOLCANOES?

**Table 2.1:** Minimum forcing of considered volcanoes ( $F_{\min}$ ), number of volcanoes in the simulation ensemble above this threshold, maximum mean normalized response 18 months after the peak forcing, the range of “climate sensitivity” to which 50% of responses to a volcanic event belong. This is based on the simple assumption that the response would linearly relate to the equilibrium climate sensitivity.

$F_{\min}$ [ $\text{Wm}^{-2}$ ]	#Volcanoes	$\frac{\Delta T_s}{F_{\max}}$ [ $\text{K/Wm}^{-2}$ ]	range of “climate sensitivity” [K]
-0.1	648	0.081	-4.08 – 9.58
-0.5	351	0.087	-0.82 – 6.09
-1.0	216	0.096	0.05 – 5.10
-1.5	153	0.109	0.95 – 4.99
-2.0	90	0.109	0.95 – 4.97
-2.5	72	0.115	1.14 – 4.98
-3.0	45	0.120	1.31 – 4.76

# Chapter 3

## Assessment of different feedback metrics<sup>1</sup>

We quantify physical radiative feedbacks for idealized climate simulations using four different methods. The results differ between the methods and differences are largest for the cloud feedback. The spatial and temporal variability of each feedback are used to estimate the averaging scale necessary to satisfy the feedback concept of one constant global mean value. We find that the year-to-year variability of each feedback process in this single model is comparable to the model-to-model spread in feedback strength of the CMIP3 ensemble. The strongest spatial and temporal variability is in the short-wave component of the cloud feedback. In our simulations, where many sources of natural variability are neglected, long-term averages are necessary to get a reliable feedback estimates. Considering the large natural variability and relatively small forcing present in the real world, as compared to the forcing imposed by doubling CO<sub>2</sub> concentrations in the simulations, implies that using observations to constrain feedbacks is a challenging task and requires reliable long-term measurements.

### 3.1 Introduction

Climate models still give a wide range of surface temperature responses to the same idealized external forcing, for example a doubling the CO<sub>2</sub> concentrations (Solomon et al. 2007). Most of these differences arise from physical processes, which are usefully conceptualized as feedbacks and can be isolated through a feedback analysis (Cess et al. 1990; Colman 2003; Soden and Held 2006). A variety of methods have been developed to isolate specific feedback mechanisms

---

<sup>1</sup>This chapter will be submitted to *Climate Dynamics*: Klocke, D., J. Quaas, M. Giorgetta, B. Stevens, Assessment of different feedback metrics, *Climate Dynamics*, 2011.

in climate models, raising the question as to how sensitive the results of such an analysis are to the methods employed (e.g. Colman 2003; Soden and Held 2006).

In stationarity, the global mean surface temperature,  $T_s$ , does not change ( $\frac{\Delta T_s}{\Delta t} = 0$ ) and the top-of-atmosphere radiation (ToA),  $R$ , is in balance ( $\frac{\Delta R}{\Delta t} = 0$ ), if averaged over a sufficiently long period  $\Delta t$ . If an external forcing,  $F$ , is imposed, for example through a change in greenhouse gas concentrations, the radiative budget at ToA is in imbalance. The climate system responds by changing its mean temperature to radiate excessive energy back into space, in order to return to equilibrium (sometimes called ‘‘Planck’’ feedback). The change in temperature affects other temperature dependent climate processes. If those processes in turn have an effect on the radiation budget (and hence on temperature), they are referred to as climate ‘‘feedbacks’’, analogous to the feedback definition in electronic circuits. Those feedbacks can have amplifying (positive feedback) and dampening (negative feedback) effects on the initial perturbation to the ToA radiation budget. This feedback concept is summarized in Equation 3.1, where the feedback parameter  $\lambda$  (in units of  $\text{Wm}^{-2}\text{K}^{-1}$ ) includes all physical feedback processes plus their mutual interactions.  $\Delta T_s$  is the change in global mean surface temperature. For a forcing from a doubling of the atmospheric  $\text{CO}_2$  concentration, this temperature change is often referred to as the equilibrium climate sensitivity.

$$\Delta R = \Delta F - \lambda \Delta T_s \quad (3.1)$$

Physical feedbacks can be linked to quantities that change in response to a change in global mean surface temperature. The relevant physical quantities we will focus on in this study are the temperature, water vapor, surface albedo in snow and ice regions, and clouds. Other feedbacks due to biogeochemical processes are not considered here (e.g., Friedlingstein et al. 2006).

The feedback factor  $\lambda$  can be formally defined as

$$\lambda = \frac{\partial R}{\partial T_s} = \sum_x \frac{\partial R}{\partial x} \frac{\partial x}{\partial T_s} + \phi(\partial^2) \approx \sum_x \lambda_x \quad (3.2)$$

$$\text{with } \lambda_x = \frac{\partial R}{\partial x} \frac{\partial x}{\partial T_s} \quad (3.3)$$

where  $x$  denotes individual feedback processes. The second-order term and all higher-order terms represent the interactions between different feedbacks. In a linear approximation, which may be considered valid for doubled  $\text{CO}_2$  conditions and the associated temperature changes (Boer and Yu 2003), these interactions are neglected.

### 3.1 INTRODUCTION

Four physical feedback processes have been identified. The total physical feedback factor  $\lambda$  can be separated, under the assumption of linearity, into a temperature ( $\lambda_T$ ), water vapor ( $\lambda_{WV}$ ), surface albedo ( $\lambda_A$ ) and a cloud ( $\lambda_C$ ) component:

$$\lambda = \lambda_{PL} + \lambda_{LR} + \lambda_{WV} + \lambda_A + \lambda_C \quad (3.4)$$

The temperature feedback contribution ( $\lambda_T$ ) to the total feedback can be further separated into a contribution by the Planck response  $\lambda_{PL}$ , or a homogeneous change in temperature, and a contribution by the change in the tropospheric temperature lapse rate  $\lambda_{LR}$ , which measures the rate at which temperature decreases with height ( $\lambda_T = \lambda_{PL} + \lambda_{LR}$ ). The Planck response is the most fundamental feedback, characterized by the temperature dependence of the long-wave (LW) emission, where the emitted energy is proportional to the fourth power of the temperature,  $\sigma T^4$  ( $\sigma$  being the Stefan-Boltzmann constant) and is often referred to as the “no feedback” response.

The linearization in equation 3.2 is useful to disaggregate contributions of individual processes to the overall feedback and to estimate their relative importance. The quantification of individual feedbacks then allows one to compare models to quantify the uncertainty contributions, measured as the model to model difference, of single feedbacks to climate sensitivity estimates (e.g., Bony and Dufresne 2005), and if possible compare them to theoretical expectations or observations of a single component (e.g., Hall and Qu 2006).

All processes in the climate system change in concert when the climate is changing, as measured by the change in global mean surface temperature. Different methods can be utilized to break down  $\lambda$  into the different contributions, all having in common that forcing and response are separated. How parts of the contributions are separated into forcing or response depends on the adopted feedback framework.

Distinctions between feedbacks can also be arbitrary if the strength in the different physical feedbacks is related to the same processes. For example, the water vapor feedback and the tropospheric temperature lapse rate feedback are anti-correlated. If the lapse rate feedback is strongly negative (i.e., a strong reduction in lapse rate, and thus a strong decrease in the greenhouse effect), the water vapor feedback is strongly positive. The reason that both feedbacks are related to the same mechanism, which is a change in deep convection. A weaker temperature lapse rate is generated by a greater warming at high altitudes than at the surface due to heat transport by convection. At the same time, enhanced convection also leads to more upper tropospheric water vapor (e.g., Cess 1975; Held and Soden 2000). For this reason, these two feedbacks are often added together to a single feedback ( $\lambda_{WV+LR}$ ), in which they partly compensate each other. By this the inter-model spread in the strength of this combined feedback is reduced. Huy-

bers (2010) reports further compensations between different feedbacks (especially surface albedo and cloud feedback), but argues that those relations can in fact be an artifact due to, (1) the methods used to estimate the feedbacks, (2) the representation of physical relationships in the models, or (3) conditioning the models upon some combination of observations and expectations.

The concept of feedbacks, forcing and climate sensitivity has proved to be helpful in the idealized model world, but extrapolation to the real world has proven to be complicated. Partial derivatives can hardly be derived from observations, due to many interfering processes, that are difficult to separate and to isolate from the background variability. In the idealized model world, this can be done with different methods. The choice for a certain analyzing method introduces limitations to the results and their interpretation.

Although the feedback parameters defined in Eq. 3.2 are a constant, there is variability in the relevant variables at various space- and timescales. In order to obtain approximately global-mean constant values, the relevant quantities need to be averaged in time. A feedback estimated for a certain year may be very different in other years and the necessary averaging time may be different for different physical processes. The biggest problem arises for clouds, which are highly variable in space and time. This has implications for quantifying feedbacks from climate models and for deriving feedback factors from observations, or finding observational constraints.

The aim of this study is to compare and assess different feedback quantification methods, and to analyze spatiotemporal variability of the different feedbacks. To do so, we use climate model simulations with the atmospheric general circulation model ECHAM5 (Roeckner et al. 2003), coupled to a mixed-layer ocean. This idealized framework neglects factors contributing to natural variability such as volcanic eruptions, El Niño variability and varying modes of ocean circulations as well as less well defined contributions to the forcing (land use change, ocean heat uptake, aerosols).

In section 3.2 we review the different methods to quantify feedbacks, and in section 3.3 we describe the experiment set-up for the idealized climate change simulations. In section 3.4 we analyze the different feedbacks, and discuss their geographical temporal variability in section 3.5 using the different methods. These results have implications for estimating feedback factors in the climate system from observations, which is discussed in the conclusions, section 3.6

## 3.2 Methods to quantify feedbacks in GCMs

Four different methods to estimate climate feedbacks have been proposed in the literature. They are based on two different principles. In this section we briefly describe these four methods. The first two are based on differences based on a recalculation of radiative fluxes perturbed by specific contributions; while the other two depend on differences in the all-sky and clear-sky radiative fluxes. The first principle is less ambiguous, but involves performing radiative transfer computations and special model diagnostics (the kernel method helps to obviate this, by approximating the partial radiative perturbation method, without the need for repeated radiative transfer calculations on ancillary data). The second principle is only applicable to the cloud component of the feedback parameter, but as climate models differ mostly in the cloud feedback component, this approach is often used to estimate the radiative impact of changed clouds in a perturbed climate.

### Partial radiative perturbation

This technique was first introduced by Wetherald and Manabe (1988) and more recently applied to an ensemble of atmosphere-ocean general circulation models by Colman (2003) and Soden and Held (2006). Offline radiative transfer calculations are used to estimate the effect of single variables such as temperature, water vapor, surface albedo or clouds on the ToA radiation. Under the assumption of linearity and separability each variable is substituted separately, one at a time, from a perturbed simulation, while all other radiation relevant variables are taken from a control simulation. This allows one to calculate each feedback factor separately for any variable  $x$ , as follows.

$$\lambda_x = \frac{\Delta_x R}{\Delta x} \frac{\Delta x}{\Delta T_s} \quad (3.5)$$

where  $\Delta x$  and  $\Delta T_s$  are obtained from the difference between a perturbed and a control simulation, and  $\Delta_x R/\Delta x$  from off-line radiation calculations. With this method the partial derivatives are calculated directly and it is closest to the formal definition of the feedback factor as defined in equation 3.2, for a few exceptions aside. These are: interactions between feedbacks are neglected, the climate change signal in any variable is the total derivative of variable  $x$  with temperature instead of the partial derivative, and the difference between perturbed and control simulations might not be small enough to allow for the discrete approximation of the derivative by the differentiation.

Colman and McAvaney (1997), Schneider et al. (1999) and Soden et al. (2004) pointed out that the assumption that all fields are uncorrelated introduces bi-

ases. Unintended perturbations are introduced to the radiation by de-correlating variables, but this can be partially overcome by applying this method twice. Once forward (FW), by substituting a variable from the perturbed climate (ptr) into the control climate (ctl) ( $\partial_{\text{ptr-ctl}}R_x$ ) and once backward (BW) by taking a variable from the control climate and substituting it into the perturbed climate ( $\partial_{\text{ctl-ptr}}R_x$ ). The final radiative perturbation is then better approximated as the average of these two estimates,  $\frac{\partial_{\text{ptr-ctl}}R_x - \partial_{\text{ctl-ptr}}R_x}{2}$ .

Because this method is less easy to implement than the other methods, spurious differences may arise depending on the exact implementation. It is also computationally expensive and needs special instantaneous model output. Most importantly, the radiative transfer part of the climate model needs to be isolated for the off-line radiative transfer computations.

### Radiative kernels

This method is a linearization of the previously described partial radiative perturbation (PRP) method and was introduced by Soden et al. (2008). Instead of perturbing one variable at a time by an increment defined from a perturbed and a control simulation, as described above for the PRP method, the mean climate state is perturbed incrementally in the radiative flux computations level by level for each variable at a time by a pre-defined small increment, and the changes to the ToA radiation balance are computed as a “radiative kernel” for variable  $x$  ( $K_x$ ) as a function of latitude, longitude, model level and time. The kernel for each variable  $x$  represents the first fraction of formula 3.5 and is multiplied with the climate change signal from a forced simulation to calculate  $\lambda_x$ . The advantage is that once those kernels are computed, offline radiation calculations are no longer necessary.

The temperature kernel ( $K_T$ ) is computed by perturbing the temperature at every level at each time by an increment of 1 K, while the specific humidity kernel ( $K_W$ ) is calculated by perturbing the specific humidity by an amount corresponding to about a 1-K warming at fixed relative humidity. The 3-D surface albedo kernel is computed by perturbing the surface albedo fields by a 1% increment (Soden et al. 2008).

A radiative kernel for clouds cannot be computed because an incremental change in “clouds”, which depends on cloud fraction as well as cloud ice and cloud water mixing ratios, is not easy to define. The clear-sky component in the  $\Delta\text{CRE}$  calculation, however, can be corrected for the influence of the other feedbacks in the clear-sky by using the difference of the full-sky kernels for each variable  $x$  ( $K_x$ ) and clear-sky kernels ( $K_x^0$ ) for the temperature, water vapor and surface albedo feedbacks.

### 3.2 METHODS TO QUANTIFY FEEDBACKS IN GCMs

$$\begin{aligned} \Delta_C R = & \Delta \text{CRE} + (K_T^0 - K_T)\Delta T + (K_W^0 - K_W)\Delta W \\ & + (K_A^0 - K_A)\Delta A + (G^0 - G). \end{aligned} \quad (3.6)$$

This compensates for cloud masking effects in the  $\Delta \text{CRE}$  calculations, as described in section 3.2. The stratospheric adjusted radiative forcing ( $G$ ) and clear-sky stratospheric adjusted radiative forcing ( $G^0$ ) are also needed for this correction.

#### Change in cloud radiative effect

This method is most commonly used and easiest to apply, but is only applicable to the cloud feedback contribution to the total climate feedback parameter. It makes use of diagnostic variables that are commonly calculated on-line in climate simulations (Cess and Potter 1987). Clear-sky radiative fluxes are calculated (subscript c), by setting all cloud related variables (cloud water, cloud ice, cloud fraction) to zero for a second diagnostic radiation call. This is done for the short-wave (SW) and long-wave (LW) component separately. The difference between the full-sky radiative flux calculations and the diagnostic clear-sky calculations yields the cloud radiative effect (CRE), where the sum of the SW and LW component is the net cloud radiative effect<sup>1</sup>.

$$\text{CRE} = (F^{\text{SW}} - F_c^{\text{SW}}) + (F^{\text{LW}} - F_c^{\text{LW}}) \quad (3.7)$$

The difference of CRE between a perturbed climate ( $\Delta \text{CRE}_{\text{prt}}$ ) and a control climate ( $\Delta \text{CRE}_{\text{ctl}}$ ) defines the change in cloud radiative effect ( $\Delta \text{CRE}$ ).

$$\Delta \text{CRE} = \text{CRE}_{\text{prt}} - \text{CRE}_{\text{ctl}} \quad (3.8)$$

This quantity is often used as a proxy for the cloud feedback.

The difficulty with this method lies in the components being very large, on the order of hundreds of  $\text{Wm}^{-2}$ , but the resulting  $\Delta \text{CRE}$  is close to zero. Furthermore, the clear-sky components of the perturbed climate include contributions from the temperature lapse rate, water vapor and surface albedo feedback, which does not allow for an accurate separation of the cloud feedback from these other feedbacks. Some part of the change in cloud radiative forcing does thus not result from changes in cloud properties, but from a change in cloud masking,

---

<sup>1</sup>Negative radiative fluxes are defined here as energy loss for the climate system, while positive radiative fluxes are an energy gain for the climate system.

so that the  $\Delta\text{CRE}$  does not accurately reflect the cloud feedback (Zhang et al. 1994; Colman 2003; Soden et al. 2004). It is often negative, even though the actual cloud feedback is generally slightly positive if diagnosed more accurately in climate models. However, when compared across models, the differences in cloud radiative forcing are predominantly a measure of differences due to clouds.

This method is widely used, because it gives an uncomplicated first estimate of the cloud influence on the radiation budget. Also, the cloud radiative effect is comparable with satellite observations (e.g. the Clouds and the Earth's Radiant Energy System, CERES, Wielicki et al. 1996), yet it should be mentioned that such comparisons should be done with caution, as clear-sky radiative fluxes from models and satellites are not directly comparable (Sohn et al. 2006).

### **Linear regression of TOA radiative flux imbalance versus surface temperature change**

This method was proposed by Gregory et al. (2004) for a constant forcing over longer time periods (years). It makes use of the relationship of the change in global-mean surface temperature ( $\Delta T_s$ ) and the forcing ( $\Delta F$ ), which is expressed as the energy balance at the top-of-atmosphere ( $\Delta R$ ).

$$\Delta R = \Delta F - \alpha \Delta T_s \quad (3.9)$$

The variations of  $\Delta R(t)$  and  $\Delta T_s(t)$  with time are regressed against each other as long term averages (e.g. yearly averages). This yields a regression line with a slope,  $-\alpha$ , and an intercept,  $\Delta F$ . The regression can be separated into the short- and long-wave components of  $R$  and  $F$ , and - analogous to the  $\Delta\text{CRE}$  calculations above - also for clear and cloudy skies, respectively. Then  $-\alpha$  is proportional to the cloud feedback estimate through the  $\Delta\text{CRE}$  calculations and if regressed for the net full-sky radiative ToA imbalance, it is an estimate for the total feedback factor. The use of clear-sky fluxes is identical to  $\Delta\text{CRE}$  so this method is facing the same interpretational issues.

The  $\Delta T_s$  intercept is equal to  $\frac{F}{\alpha}$  which is the equilibrium  $\Delta T_s$ , or climate sensitivity.  $\Delta R$  is approximately equal to the stratospheric temperature adjusted radiative forcing, for  $\Delta T_s \rightarrow 0$  (see also Figure 3.2). The advantage of this method is that forcing, cloud feedbacks and climate sensitivity can be estimated with the use of only a few years of model integration, without a need for any further diagnostics.

This method disaggregates forcing and response depending on time scales they act on. For example, clouds instantaneously respond to the increased  $\text{CO}_2$  concentrations, due to changes in heating rates, which is independent from the response to changes in surface temperature (Gregory and Webb 2008). This is referred

to as the “indirect CO<sub>2</sub> effect”, which in the definition used in this method is part of the forcing, while in the other methods it is attributed to the feedback. In this regression method, compared to other methods, the forcing is thus more loosely defined to include all processes acting on “fast” timescales, while the feedbacks are well-defined as only such processes responding to a surface temperature change. In the following, we will refer to this method as the “Gregory-method”.

### 3.3 Model and experimental set-up

All feedback metrics are applied to the same set of simulations, using ECHAM5.4 (Roeckner et al. 2003), with a relatively coarse spectral resolution of T31 (approximately 3.75 degree resolution) and 19 vertical levels. First, a 20 year control integration is conducted with prescribed present-day greenhouse gas concentrations and with prescribed monthly varying sea-surface temperatures and sea-ice cover maps. The heat fluxes from this control simulation are used for the mixed-layer ocean integrations. Coupled to a 50 m mixed-layer ocean, a 20 year control integration and a 50 year integration with doubled CO<sub>2</sub> concentrations are performed, until a new equilibrium is reached. For our analysis the last six years of the control and the perturbed simulations are used. For all four applied methods, the same six hourly model output is used. The radiation code of ECHAM5.4, with 16 long-wave and 6 short-wave bands (Cagnazzo et al. 2007), is isolated from the model and used for the offline calculations for the radiative perturbation method and for computing the radiative kernels. For the calculations of the radiative kernels, incremental perturbations are applied to output fields of temperature, specific humidity and surface albedo from the control simulation.

For quantification of the lapse rate feedback diagnostics of the tropopause height are necessary to exclude the stratospheric temperature change. Here we use the WMO defined tropopause of the control simulation which is saved together with the other instantaneous model output every six hours.

By using the same model output and radiation code throughout this study we strive to be as consistent as possible. Differences in the results will only depend on the method used and its underlying assumptions.

### 3.4 Feedback factors

Figure 3.1 shows the global-, long-term averages (six years) of the physical feedback factors analyzed using the different methods described in section 3.2. The error bars indicate the sampling error over the six years. The boxes indicate  $\pm$  one standard deviation of single year averages, while the whiskers indicate the

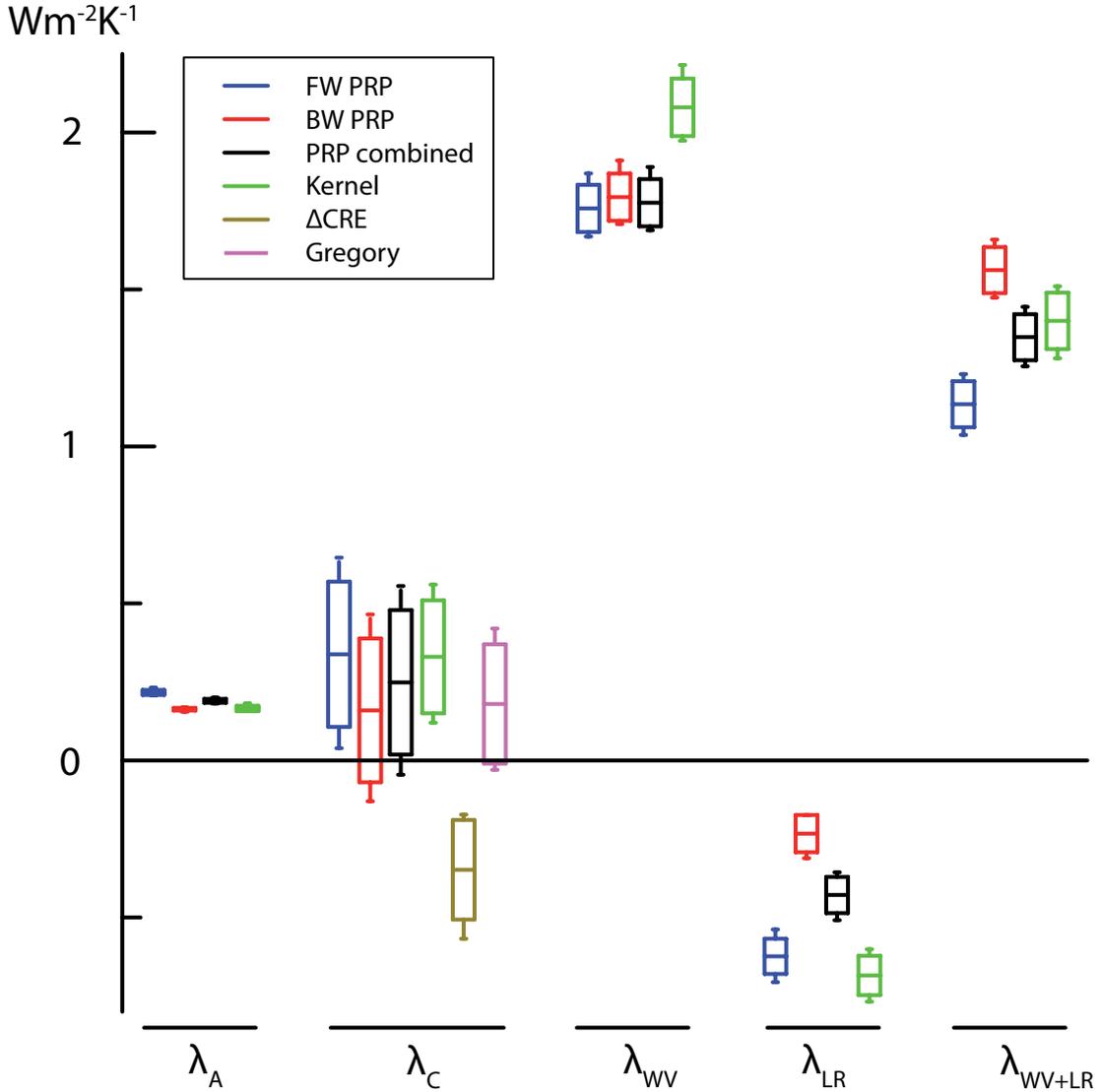
## CHAPTER 3 FEEDBACK METRICS

maximum and minimum value in a single year. Table 3.1 gives further details of the LW and SW contributions separately. The standard deviations apply to feedback estimates for six single-year averages. The cloud feedback factor,  $\lambda_C$ , is calculated with four different methods, while for the tropospheric temperature lapse rate, water vapor and surface albedo feedbacks only the PRP and the radiative kernel methods are applicable. For the PRP method the forward (FW) calculated and the backward (BW) calculated feedback factor along with the average values are provided.

The Planck response is the strongest negative feedback with  $-3.23/-3.17 \text{ Wm}^{-2}\text{K}^{-1}$ , calculated with the PRP method, and  $-3.08 \text{ Wm}^{-2}\text{K}^{-1}$ , if calculated using radiative kernels (not shown). These two estimates of the Planck response overlap within the range of uncertainty as obtained from the year-to-year variability. The tropospheric temperature lapse rate feedback,  $\lambda_{LR}$ , is negative on a global, long-term average. This feedback differs the most depending on whether the FW or BW PRP is used, with  $-0.61$  and  $-0.23 \text{ Wm}^{-2}\text{K}^{-1}$ , respectively, indicating strongest perturbations by de-correlating the different variables in the radiative flux calculations. The radiative kernel yields a  $\lambda_{LR}$  of  $-0.68 \text{ Wm}^{-2}\text{K}^{-1}$ , comparable to the FW PRP, but much larger in absolute terms than the BW PRP. The estimates of  $\lambda_{LR}$  obtained with the two different methods do not overlap within one standard deviation of the year-to-year variability, if the FW and BW calculated PRP are combined to a lapse rate feedback of  $-0.42 \text{ Wm}^{-2}\text{K}^{-1}$ .

The water vapor feedback,  $\lambda_{WV}$  is the strongest positive feedback with  $1.76/1.79 \text{ Wm}^{-2}\text{K}^{-1}$  for the FW/BW PRP and  $2.08 \text{ Wm}^{-2}\text{K}^{-1}$  using the radiative kernel method. For the water vapor feedback, these two methods differ the most in an absolute sense. The water vapor feedback estimates derived with the PRP and kernel method do not agree within one standard deviation of the year-to-year variability. This is mainly due to the large differences in the LW component of this feedback. This feedback acts in both the SW and the LW spectra, but is dominated by the LW contribution that is responsible for  $\sim 75\%$  of the total water vapor feedback. For the LW contribution the water vapor feedback factor differs strongly, depending on the method chosen. Here the PRP LW water vapor feedback is  $1.32/1.42 \text{ Wm}^{-2}\text{K}^{-1}$ , while the LW component is larger with  $1.71 \text{ Wm}^{-2}\text{K}^{-1}$ , if calculated using the linearization through the LW water vapor kernel. For the SW the PRP and kernel method indicate a feedback strength of  $0.43/0.38$  and  $0.37 \text{ Wm}^{-2}\text{K}^{-1}$ , respectively.

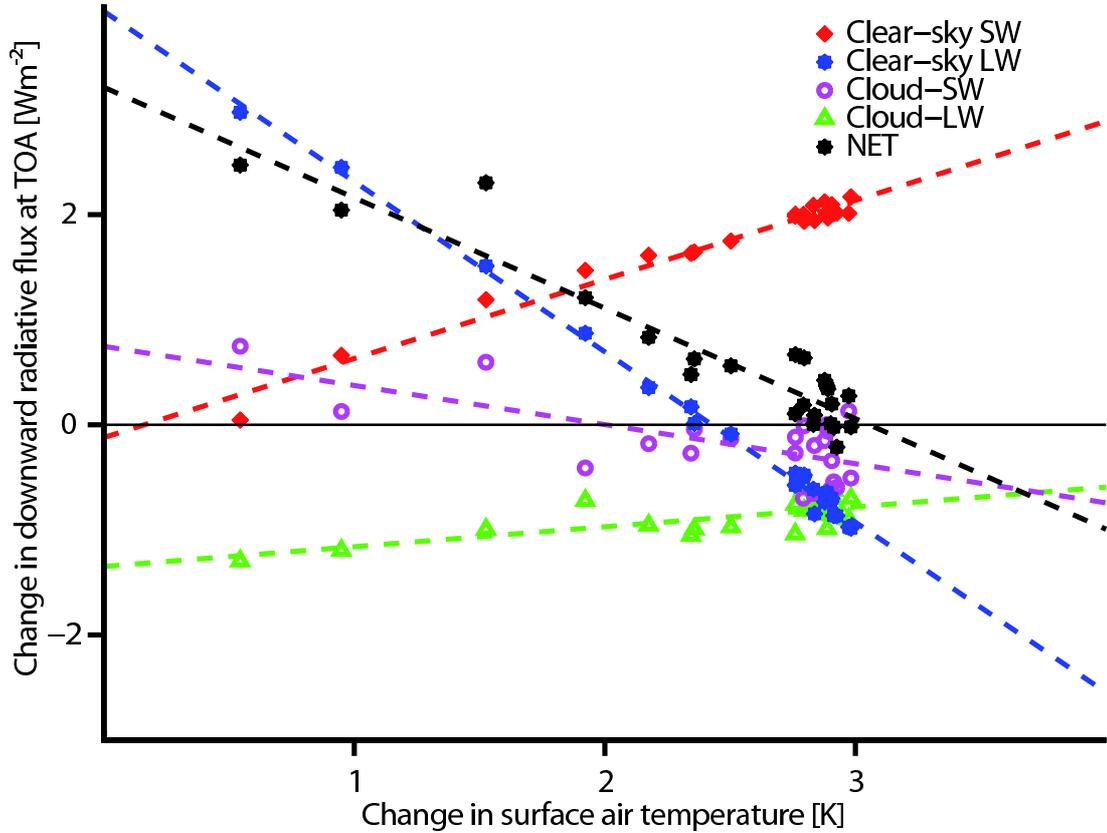
Combining the water vapor and lapse rate feedbacks,  $\lambda_{WV+LR}$ , partly compensates the discrepancies between the PRP and kernel methods. The LW components do not overlap within one standard deviation of the inter-annual variability for either  $\lambda_{LR}$  or  $\lambda_{WV}$  between the two methods, but they do in the combined feedback. This is analogous to the feedback strength difference for  $\lambda_{LR}$  and  $\lambda_{WV}$



**Figure 3.1:** Surface albedo ( $\lambda_A$ ), cloud ( $\lambda_C$ ), water vapor ( $\lambda_{WV}$ ), tropospheric temperature lapse rate ( $\lambda_{LR}$ ) and the combined  $\lambda_{WV}$  and  $\lambda_{LR}$  feedback factor calculated with different methods. Each box is the mean feedback strength  $\pm$  one standard deviation and the whiskers indicate the maximum and minimum yearly averaged feedback strength of the six analyzed years. For the Gregory method the whiskers are calculated taking the maximum and minimum deviation from the regression, while the box gives the mean regression error.

among climate models, which decreases when the two are combined (Colman 2003).

The surface albedo feedback is only affecting the SW radiation. In our simulations it is the smallest feedback, with  $0.22/0.16 Wm^{-2}K^{-1}$  using the PRP method



**Figure 3.2:** Change in net downward radiative flux regressed against the change in global mean surface temperature for yearly averages. The slope of the cloud components is proportional to the cloud feedback strength.

and  $0.17 \text{ Wm}^{-2}\text{K}^{-1}$ , if calculated with the surface albedo kernel. These measures agree within the measure of uncertainty used here, when the PRP-FW and PRP-BW are combined, yielding a  $\lambda_A$  of  $0.18 \text{ Wm}^{-2}\text{K}^{-1}$ .

The cloud feedback affects the LW and SW radiation strongly, but the globally temporally averaged feedback factors are small. While the PRP method, kernel method and Gregory method give a positive cloud feedback with  $0.34/0.16$ ,  $0.33$  and  $0.18 \text{ Wm}^{-2}\text{K}^{-1}$  respectively, the  $\Delta\text{CRE}$  is of opposite sign with  $-0.35 \text{ Wm}^{-2}\text{K}^{-1}$ . This is not surprising, because  $\Delta\text{CRE}$  does not correspond directly to  $\lambda_C$  as defined in equation 3.2 (see Section 3.2). Examining the SW and LW component separately gives no consistent picture of the cloud feedback strength across all methods used here. The PRP methods gives a positive cloud feedback factor for both components, while the  $\Delta\text{CRE}$  is of negative sign for both components. The kernel corrected  $\Delta\text{CRE}$  gives a positive LW and a negative SW cloud feedback, while the cloud feedback derived from the Gregory-method is negative in the LW

### 3.4 FEEDBACK FACTORS

**Table 3.1:** Feedback factors and their standard deviation between different years, calculated for six years with different methods for all components. Values are given in  $\text{Wm}^{-2}\text{K}^{-1}$ . The values calculated with the backward (BW) partial radiative perturbation (PRP) are multiplied with -1 to be comparable to the forward (FW) calculated PRP.

$\lambda_x$	Method	LW	LW Std	SW	SW Std	Net	Net Std
$\lambda_{\text{PL}}$	PRP (FW/BW)	-3.23/-3.17	0.1/0.1	0.0	0.0	-3.23/-3.17	0.1/0.1
	Kernel	-3.08	0.1	0.0	0.0	-3.08	0.1
$\lambda_{\text{LR}}$	PRP (FW/BW)	-0.61/-0.23	0.05/0.06	-0.01/0.0	0.0/0.0	-0.61/-0.23	0.05/0.06
	Kernel	-0.68	0.1	0.0	0.0	-0.68	0.1
$\lambda_{\text{WV}}$	PRP (FW/BW)	1.32/1.42	0.07/0.07	0.43/0.38	0.01/0.01	1.76/1.79	0.08/0.08
	Kernel	1.71	0.08	0.37	0.02	2.08	0.09
$\lambda_{\text{A}}$	PRP (FW/BW)	0.0/0.0	0.0/0.0	0.22/0.16	0.01/0.0	0.22/0.16	0.01/0.0
	Kernel	0.0	0.0	0.17	0.01	0.17	0.01
$\lambda_{\text{C}}$	PRP (FW/BW)	0.18/0.08	0.01/0.01	0.16/0.08	0.24/0.24	0.34/0.16	0.23/0.23
	Kernel	0.49	0.03	-0.17	0.17	0.33	0.18
	$\Delta\text{CRE}$	-0.24	0.04	-0.11	0.16	-0.35	0.16
	Gregory	-0.19	0.06	0.37	0.18	0.18	0.19
$\lambda_{\text{WV+LR}}$	PRP (FW/BW)	0.71/1.18	0.06/0.06	0.43/0.38	0.01/0.01	1.13/1.56	0.07/0.07
	Kernel	1.03	0.09	0.37	0.02	1.40	0.09

spectra and positive in the SW. These differences arise from the cloud masking effect in the  $\Delta\text{CRE}$  calculation, as well as in the Gregory method (see also section 3.2). Figure 3.2 shows  $\Delta\text{CRE}$  in relation to the change in surface temperature. The slopes of the regression lines for the cloud LW and cloud SW components indicate the cloud feedbacks. It is notable that both regression lines have a non-zero intercept at the Y-axis, which is in this framework part of the forcing (Gregory and Webb 2008). This explains the smaller magnitude of the net cloud feedback when estimated by the Gregory-method. The regression error is used as the sampling error, comparable to the standard deviation of the inter-annual variability and the maximum and minimum distance from the regression line are used for the whiskers in Figure 3.1. The uncertainties inferred from this are large, especially for the SW component.

## 3.5 Feedback variability

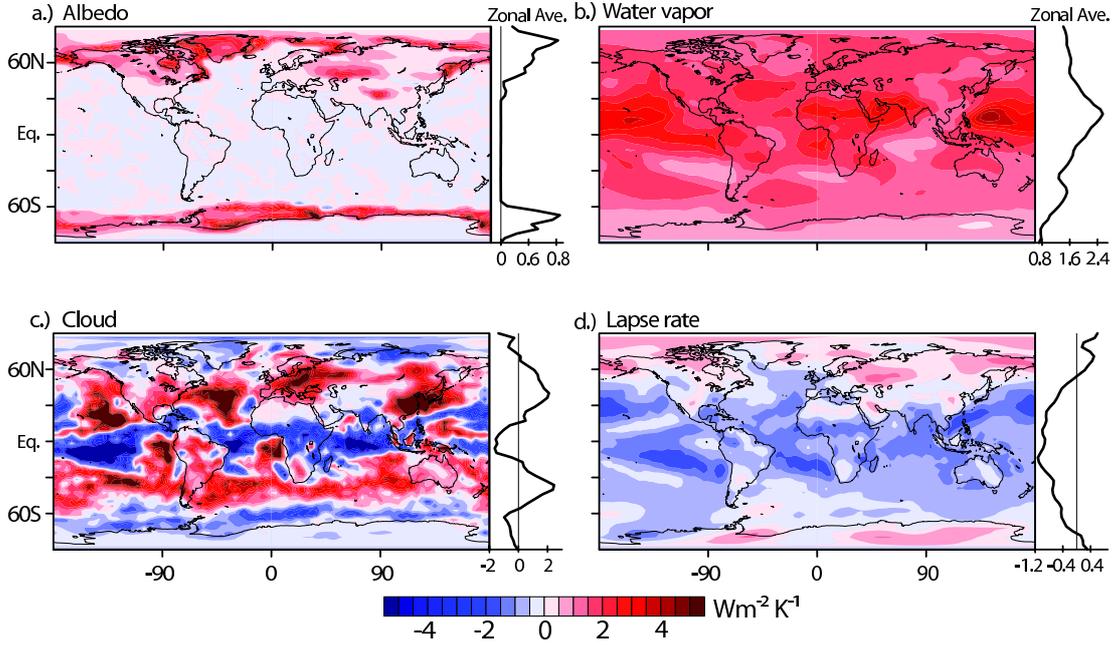
Feedback analyses are applied to differences between two simulations in stationary state. Hence feedback factors are constant by definition, if averaged over a sufficiently long period, and we discussed them as such in the previous section. For practical reasons those analyses are applied to relatively short periods which can be too short, resulting in sampling errors when estimating feedbacks. The length of the averaging period depends on the feedback of interest and the desired accuracy. In a climate model many sources of variability can be eliminated to minimize the variability and to make understanding of the feedback processes easier. For the experiments used here, we use a low resolution climate model (see section 3.3), coupled to a mixed layer ocean. Due to this simplification the absolute numbers for the feedback quantification can be seen only as an approximation to reality, but the implications and differences of the used feedback metrics are still worth analyzing. In the following we analyze spatial and temporal variability separately, which both are likely higher in a less simplified set up, when a full dynamical ocean, vegetation, sources of natural variability like volcanoes, or a varying solar constant are included. The same is true for nature, where a smaller, transient forcing and uncertainties in observations make the accurate determination of feedback factors a challenging, still unresolved, task. Analyzing the temporal and spatial variability of feedbacks helps to answer how long we need to average, in order to get accurate estimates of a feedback factors from any given method and geographical distributions help to understand the underlying processes controlling feedbacks, ultimately leading to a better understanding of the climate system and its response to perturbations.

### 3.5.1 Spatial variability

Figure 3.3 and 3.4 show maps of the geographical distributions of the surface albedo, water vapor, cloud and lapse rate feedback factors, calculated using the PRP method and radiative kernels, respectively, and Figure 3.6 shows the  $\Delta\text{CRE}$  as a proxy for the cloud feedback. All feedback maps are calculated from the same six hourly output over six years.

The Planck response (not shown) is the first order feedback, if just temperature would change uniformly when a forcing was imposed to the system. It is strongly negative everywhere with  $-3.08$  to  $-3.23 \text{ Wm}^{-2}\text{K}^{-1}$  as a global average, depending on the method (Table 3.1). As temperature rises with higher carbon dioxide concentrations, the forcing gets balanced at the ToA by radiating more energy out to space. The strongly non-linear relation ( $\sigma T^4$ ) makes the Planck response strongest in the tropics, where temperatures are already high, and weakest in

### 3.5 FEEDBACK VARIABILITY



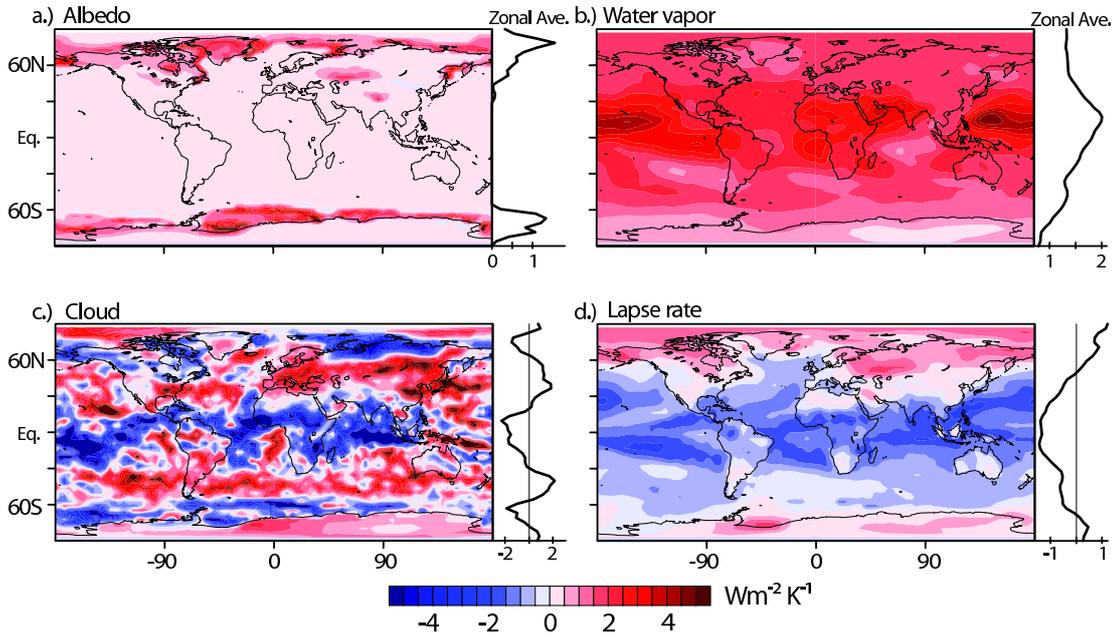
**Figure 3.3:** Geographical distribution of physical feedback factors calculated using the partial radiative perturbation method and their zonal means, as averaged over six years.

high latitudes, assuming a uniform temperature change.

The surface albedo feedback is positive nearly everywhere and only effective in mid to high latitudes (Figure 3.3a and 3.4a). Higher temperatures under doubled  $\text{CO}_2$  conditions cause less ice and snow to form in winter and lead to an earlier snow and ice melting in spring. The second is more relevant, because there is more short-wave radiation in spring to make this feedback more effective. Locally this feedback factor can exceed  $3 \text{ Wm}^{-2}\text{K}^{-1}$  ( $1 \text{ Wm}^{-2}\text{K}^{-1}$  on zonal average) and reaches its maximum in our simulations around  $70\text{--}80^\circ$  in both hemispheres.

The water vapor feedback is strongly positive everywhere (Figure 3.3b and 3.4b) and geographical structures are similar for both methods. The cold tropical tropopause and the dry subtropical subsiding branches of the upper atmosphere are most susceptible to changes in specific humidity, which leads to a maximum of  $\lambda_{\text{WV}}$  at about  $15^\circ \text{ N}$ .

The lapse rate feedback ( $\lambda_{\text{LR}}$ , Figure 3.3d and 3.4d) is positive over large regions in the mid and high latitudes, mainly continental areas. At low latitudes, the atmosphere warms more at higher altitudes than at the surface, where the vertical temperature profile remains close to the moist adiabat due to the influence of deep convection. In mid- to high-latitude continental areas the surface temperature responds strongest, leading to a positive lapse rate feedback. Here



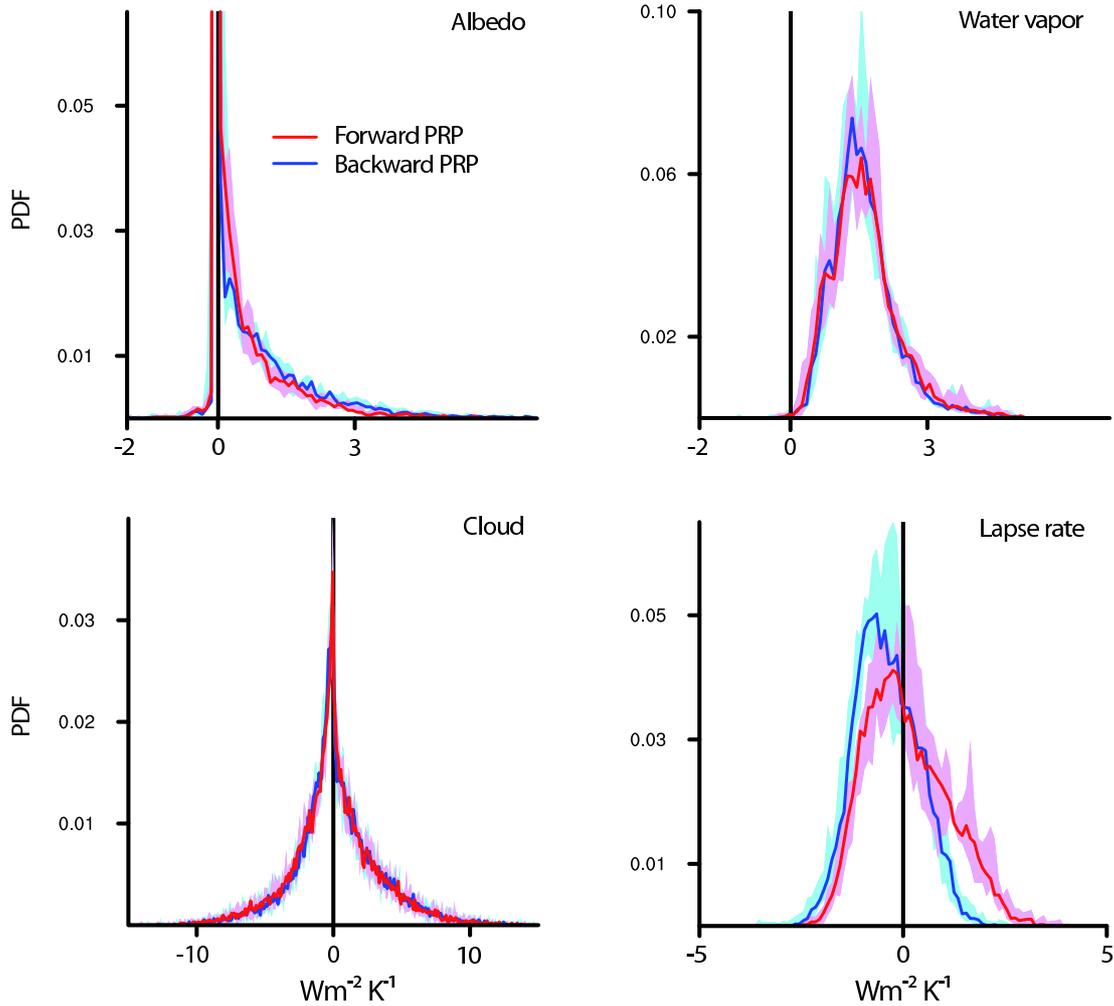
**Figure 3.4:** Geographical distribution of physical feedback factors calculated using radiative kernels and their zonal means, as averaged over six years. The cloud feedback factor in c.), is calculated from clear-sky corrected  $\Delta\text{CRE}$ .

the tropospheric temperature lapse rate is mainly controlled by baroclinic adjustment (Stone and Carlson 1979). For the temperature lapse rate feedback the difference is largest between the forward and backward calculated feedback factor (Figure 3.5d). The probability density functions for the lapse rate feedback differ throughout the whole distribution. Those differences are strongest at high latitudes where the lapse rate feedback is positive. In these regions artificial perturbations are most relevant through de-correlation of the fields in the PRP method. Also changes in cloud masking strongly influence the results obtained when the PRP method is applied only one way (i.e, only FW).

### Spatial variability of the cloud feedback

The cloud feedback (Figure 3.3c, 3.4c and 3.6) can locally be strongly negative or positive. On a global average this nearly cancels out, so that the global averaged feedback factor is close to zero (see Figure 3.1 and Table 3.1) and depending on the method, this average feedback can be positive or negative. This makes the accurate estimation of the cloud feedback particularly difficult. Clouds can change their height, depth, size, frequency, reflectivity, phase, or any combination of these. The cloud feedback thus affects the long- and short-wave spectra and is

### 3.5 FEEDBACK VARIABILITY

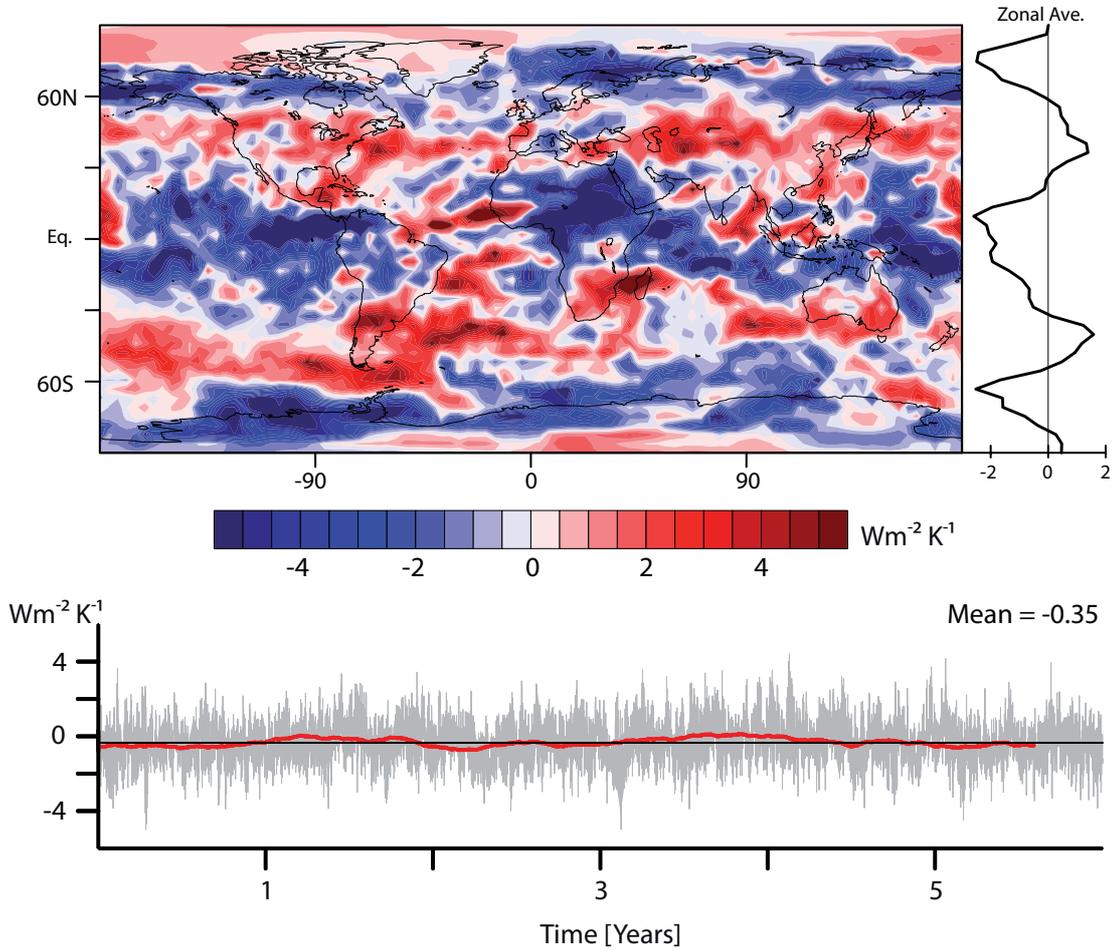


**Figure 3.5:** Probability density functions of physical feedback factors calculated for annual mean geographical distributions, using the forward PRP (red) and the backward PRP (blue, multiplied with -1 for better comparison) method. The solid lines indicate the average over the six years and the shaded area indicates the year to year variability.

highly variable in space and time.

The geographical distributions of the net cloud feedback obtained from the PRP, kernels and  $\Delta\text{CRE}$  look similar, but on a global average  $\Delta\text{CRE}$  is lower, in our case in fact of a different (negative) sign. This is due to the temperature and water vapor feedbacks in the clear-sky component. Some parts of those feedbacks are not separated from the cloud feedback when subtracting the clear-sky component, if the cloud masking changing (see Section 3.2). The zonally averaged structure is similar between the three methods, with a negative cloud feedback in the inner tropics, positive cloud feedback in the mid latitudes and again a negative feedback in the high latitudes. The cloud feedback maps differ in

their detail but the corrected  $\Delta\text{CRE}$  through the kernel method agrees better in its geographical distribution, as well in its global average, with the cloud feedback calculated using the PRP method.

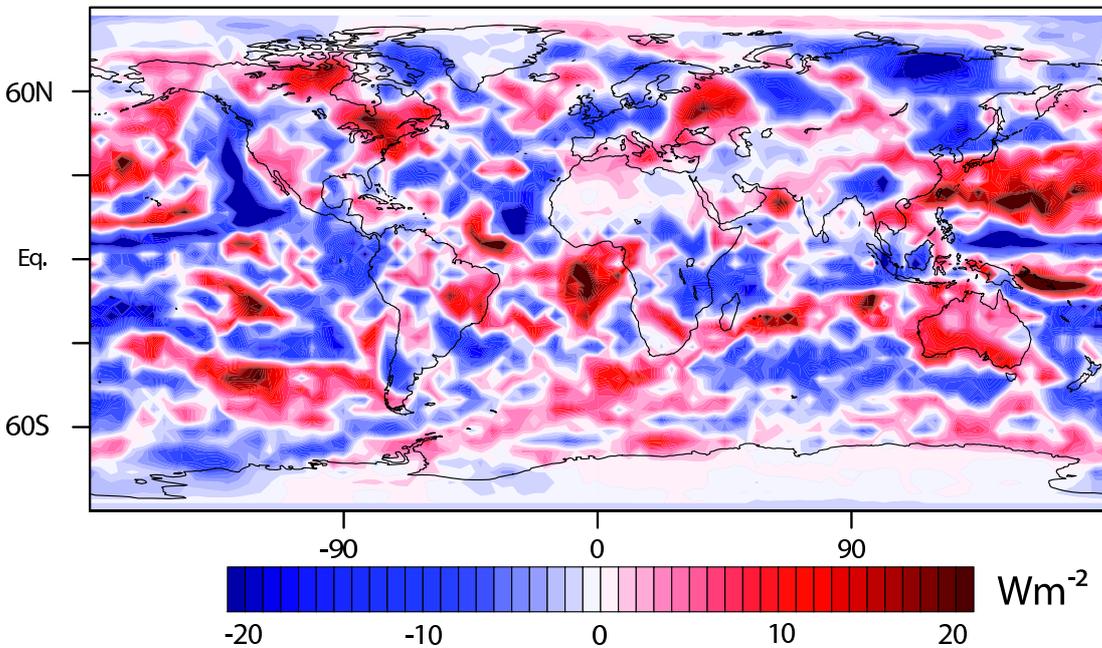


**Figure 3.6:** Change in net cloud radiative effect as time average (above) and global averages for every six hours over six years in grey and five month running mean in red (bottom).

The bulk of the geographical variability comes from the short-wave component of the cloud feedback, which shows structures similar to the net cloud feedback (Figure 3.8). Changes in the SW cloud forcing are mainly due to changes in cloud cover (rather than cloud top height or cloud water content).

The LW component of the cloud feedback is globally positive except over subtropical oceans and polar regions, where it is slightly negative (Figure 3.8). In our simulations the tropopause rises in general due to a deepening of the convective overturning, leading to an increase in upper tropospheric humidity and consequently an upward shift of the profile of tropospheric infrared cooling. This

decoupling of LW emission at the top of high anvil clouds from the surface emissions was described by Hartmann and Larson (2002) and Zelinka and Hartmann (2010), who hypothesized that through these mechanisms all models simulate a positive LW cloud feedback.



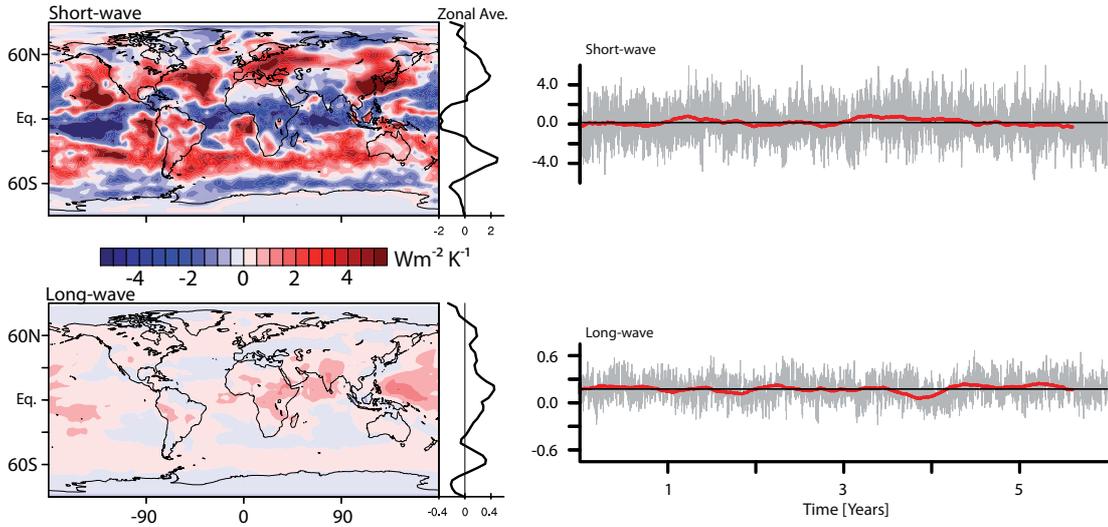
**Figure 3.7:** Cloud radiative forcing, if the PRP method is applied to cloud related fields switched from one year of a control simulation to another year.

The year-to-year variability in clouds is strong. If the PRP method is applied to two years of the same climate state, e.g. by taking cloud fields from one year and calculating their radiative perturbation as they are set in the atmospheric state of a different year, the radiative forcing can be of comparable magnitude to the cloud feedback (Figure 3.7). In this - arbitrarily chosen - case the global mean ToA radiative forcing is  $-0.55 \text{ Wm}^{-2}$ , but over several years this averages out to zero. The main feature that gives confidence in the feedback in Figure 3.3c, 3.4c and 3.6 is its structure, which is independent of the applied method. Especially the zonal structure appears robust. Little structure can be identified in the year-to-year variation of the cloud forcing, which indicates that the cloud-climate feedback can be separated from the natural variability in cloudiness.

### 3.5.2 Temporal variability

The climate feedback concepts are based on global mean, long time averages, but the geographical and temporal variability still have valuable information. Especially, if the goal is to verify feedback strength estimates from models with

## CHAPTER 3 FEEDBACK METRICS



**Figure 3.8:** Geographical distribution of cloud short-wave and cloud long-wave radiative feedback factor (left) and six hourly globally averages (right) of the components of cloud feedback factor for six years (grey lines), calculated using the FW PRP method. The red line in the time series is the five month running mean and the horizontal black lines is the mean over the entire time series. Note that the Y-axis range of the short-wave is the same as the range for the net cloud feedback factor and the range for the long-wave component is identical to all other feedback factors in Figure 3.9.

observations, or to infer feedback strength directly from observations, the variability characteristics become important as one has to derive the averaging period that is needed for accurately estimating feedback factors. Figure 3.9 shows the temporal variability of the different feedbacks as global averages, every six hours for six consecutive years, as calculated with the PRP method. The standard deviations in Table 3.1 are calculated from different yearly averages along each time series.

The time series in Figure 3.9 for the globally averaged surface albedo, lapse rate and water vapor feedback show seasonality and vary within  $0.5\text{-}1.0 \text{ Wm}^{-2}\text{K}^{-1}$  over the six years analyzed here. The variation in the lapse rate feedback and the water vapor are weakly anti-correlated on short (6 hours) time-scales (correlation coefficient  $r=-0.21$ ), but strongly correlated ( $r=0.71$ ) when averaged over three months. A weak lapse rate feedback is caused by a smaller temperature change aloft, leading to a small water vapor feedback at the same time. Huybers (2010) reports further correlation between feedbacks across climate models, which might not be entirely physical. We find that the surface albedo feedback and the lapse rate feedbacks have the strongest correlation ( $r=0.31$ ) on short time-scales, even stronger than the correlation of the tropospheric temperature lapse rate with the water vapor feedback. While on longer time-scales however, the correlation is

only  $r=0.50$  related to the seasonal cycle. The global mean values of the net cloud feedback are correlated with the lapse rate feedback ( $r=0.17$  on short time scales and up to  $r=0.31$  on the time-scale of days) and a correlation between surface albedo and water vapor feedback also exists (up to 0.5 depending on the averaging time scale), all other combinations show little to no correlation.

The global mean values of the the surface albedo, lapse rate and water vapor feedback reflect seasonality. This is especially true for the surface albedo feedback, which is strongest in northern hemisphere spring when solar radiation at high northern latitudes starts to increase.

### Temporal variability of the cloud feedback

The cloud feedback is much more variable than the other feedbacks but shows no seasonal variation in its global mean. As the geographical distribution of the cloud feedback strength in Figure 3.3, where regionally the feedback can be strongly positive, or negative, it varies in the global mean on the 6-hourly time scale by  $\pm 5 \text{ Wm}^{-2}\text{K}^{-1}$ , while the mean is close to zero.

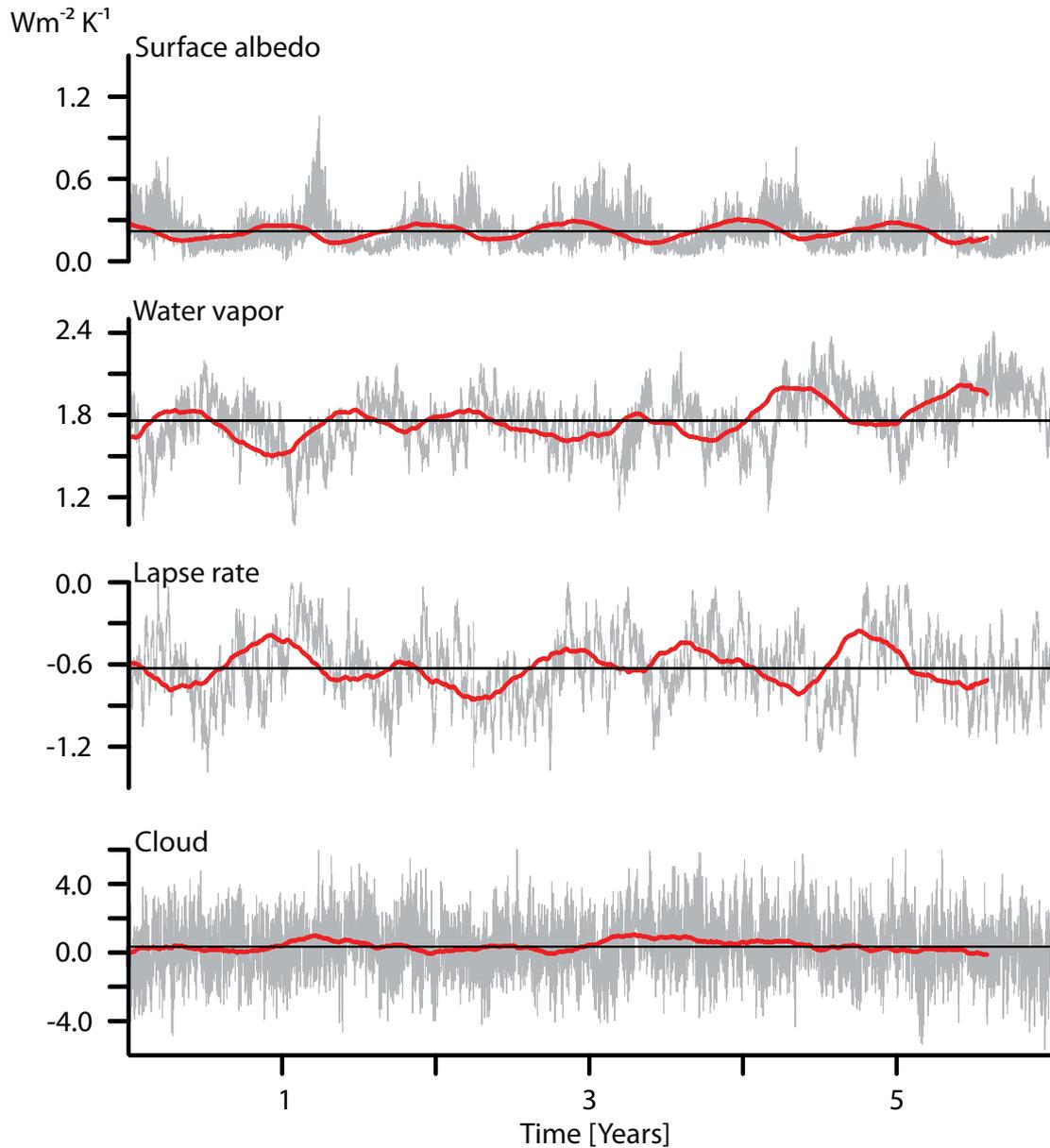
The largest part of the temporal variability comes from the SW component of the cloud feedback (Figure 3.8) that is dominated by low clouds, that have a high albedo. Due to their low thermal contrast with the surface, the impact on the LW is much smaller, and its variability is comparable to the other feedbacks. Note that the temporal variability from  $\Delta\text{CRE}$  in Figure 3.6 is much smaller than the temporal variability obtained from the PRP method.

## 3.6 Implications and conclusions

By using different methods to separate physical feedbacks in idealized climate simulations, through a consistent use of model output and radiation code, we show that feedback parameters are method-dependent. Methods differ in their definition of forcing and feedback which is affecting the results. Assumptions of how processes are disaggregated and how this is done exactly introduces further biases between methods.

Overall the geographical distributions are comparable between the methods, with robust regional features (although details differ). For the cloud feedback the geographical structure is consistent between years, giving confidence in the simulated feedback.

We also show that feedback processes vary on different timescales and with a different magnitude, even when many modes of variability are excluded in our experiments. This makes long-term averages necessary for stable estimates of feedback factors. The cloud feedback varies the most in our simulations, especially



**Figure 3.9:** Time series of physical feedback factors for six hourly (grey line) globally averaged values for six years, calculated using the FW PRP method. The five months running mean (red line) is plotted to accentuated long term variability and the mean value (horizontal black line) is plotted as reference. Note the eight times larger range on the Y-axis of the cloud feedback factor.

in the SW spectrum, thereby introducing large sampling errors (on the order of magnitude of the actual feedback) if only short global temporal averages are used. In our simulations a single year is sufficient to estimate the surface albedo feedback. The tropospheric temperature lapse rate feedback requires about three

### 3.6 IMPLICATIONS AND CONCLUSIONS

years averaging time, although absolute differences are large depending on the method. For the water vapor feedback this is five years and for the cloud feedback six years. For the latter, the sampling error can be as large as the inter-model difference in the CMIP3 ensemble and might be even larger in more complex models, let alone in reality.

Depending on one's interest, even simple estimates of the  $\Delta\text{CRE}$  can give valuable and reliable information, for example about the zonal structure of the cloud feedback. For the PRP and the kernel method, where all contributions to  $\lambda$  are known, the equation 3.1 can be evaluated for a climate in equilibrium. This would verify whether all assumptions are justified when disaggregating the feedbacks.

The forcing introduced by the doubling of the atmospheric  $\text{CO}_2$  concentration, as done in the perturbed simulation, can be calculated online as the stratospheric adjusted radiative forcing (see Stuber et al. 2001). This yields  $3.89 \text{ Wm}^{-2}$  for the model configuration used here. The forcing could also be estimated by using a  $\text{CO}_2$  kernel, analog to the other kernels for the feedback calculations, or taken from the Gregory method ( $3.91 \text{ Wm}^{-2}$ ). The equilibrium climate sensitivity for the model configuration used here is  $2.98 \text{ K}$ .

For the PRP method equation 3.1, with the feedback parameter expanded as in Eq. 3.4, and with the radiation imbalance on the right-hand-side set to zero assuming an equilibrium is attained, yields:

$$0 \approx -0.36(\pm 1.13) = 3.89 + [-3.23(\pm 0.0) - 0.42(\pm 0.06) + 1.78(\pm 0.08) + 0.19(\pm 0.01) + 0.25(\pm 0.23)] * 2.98$$

For the radiative kernel method, it yields:

$$0 \approx -0.66(\pm 0.75) = 3.89 + [-3.08(\pm 0.0) - 0.61(\pm 0.1) + 2.08(\pm 0.09) + 0.17(\pm 0.01) + 0.33(\pm 0.23)] * 2.98$$

This equation gives consistent estimates of the total feedback factor within the range of uncertainty, here defined as the sampling error due to variability, for the PRP method as well as for the radiative kernel method. The results suggests that sampling errors can easily exceed the errors introduced through the assumptions made in the different methods.



## Chapter 4

# On constraining estimates of climate sensitivity with present-day observations through model weighting<sup>1</sup>

The distribution of model-based estimates of equilibrium climate sensitivity has not changed substantially in more than 30 years. Efforts to narrow this distribution by weighting projections according to measures of model fidelity have so far failed, largely because climate sensitivity is independent of current measures of skill in current ensembles of models. Here we provide a cautionary example showing that measures of model fidelity that are effective at narrowing the distribution of future projections (because they are systematically related to climate sensitivity in an ensemble of models) may be poor measures of the likelihood that a model will provide an accurate estimate of climate sensitivity (and so degrade distributions of projections if they are used as weights). Furthermore, it appears unlikely that statistical tests alone can identify robust measures of likelihood. We consider two ensembles: one obtained by perturbing parameters in a single climate model, and a second containing the majority of the world's climate models. The simple ensemble reproduces many aspects of the multi-model ensemble, including the distributions of skill in reproducing the present-day climatology of clouds and radiation, the distribution of climate sensitivity, and the dependence of climate sensitivity on certain cloud regimes. By restricting error measures to those regimes

---

<sup>1</sup>This chapter is in review by the *Journal of Climate*: Klocke, D., R. Pincus, J. Quaas, On constraining estimates of climate sensitivity with present-day observations through model weighting, *Journal of Climate*, 2011.

we can identify tighter relationships between climate sensitivity and model error and narrower distributions of climate sensitivity in the simple ensemble. These relationships, however, do not carry into the multi-model ensemble. This suggests that model weighting based on statistical relationships alone is unfounded, and perhaps that climate model errors are still large enough that model weighting is not sensible.

## 4.1 Model error and climate sensitivity

Equilibrium climate sensitivity, defined as the response in global-mean near-surface temperature to a doubling of atmospheric CO<sub>2</sub> concentrations from pre-industrial levels, is a useful proxy for climate change because many other projections scale with it. Climate models produce a range of estimates of climate sensitivity which can themselves be sensitive to fairly small changes in model formulation (Soden et al. 2004). The distribution of these projections has remained roughly the same for more than 30 years (compare, for example, Charney 1979; Solomon et al. 2007).

One might expect that with improvements of climate models over time, projections would converge to a narrower distribution, but this has not yet proved true: successive generations of climate models have produced improved simulations of the present-day climate (Reichler and Kim 2008) but commensurate distributions of climate sensitivity (Knutti et al. 2008).

The distribution might also be narrowed by invoking Bayes's theorem and weighting each prediction of climate sensitivity by the likelihood of the corresponding model (Murphy et al. 2004; Stainforth et al. 2005; Knutti et al. 2010). This likelihood is usually modeled as a decreasing function of model error, defined as some measure of the difference between long-term averages of observations and model simulations of the present-day climate. Weighting ensembles is fraught with theoretical issues including the impact of the sampling strategy used to construct the initial ensemble (Frame et al. 2005) and questions of how to treat an ensemble in which members have varying degrees of interdependence (e.g. Knutti et al. 2010; Tebaldi and Knutti 2007). But weighting projections has so far failed to substantially narrow distributions of climate sensitivity for a more practical reason: in current ensembles of climate models, global measures of error are not systematically related to climate sensitivity or the underlying feedbacks (Knutti et al. 2006; Murphy et al. 2004; Piani et al. 2005; Sanderson et al. 2008; Collins et al. 2011).

Any observable measure of present-day error that is correlated with climate sensitivity in a given ensemble of climate projections, if used as a weight, would

narrow the distribution of climate sensitivity estimates. This makes it tempting to seek such measures. But if the systematic relationships between the present day and the future in an ensemble of models have causes which are not shared by the physical climate system, weighting by such a measure can introduce substantial projection errors (Weigel et al. 2010).

Here we provide a practical demonstration of how hard it can be to determine whether relationships between the present day and the future in a given ensemble have a more general basis. We consider two ensembles of climate models: one containing a wide range of models and another employing a single model with varied values of closure parameters. We use the simpler, single-model ensemble as a proxy for understanding the behavior of the more complicated multi-model ensemble, much as one might use the more complicated ensemble to understand the real world. Section 4.2 describes the construction of the simple ensemble; we then show that this simple ensemble reproduces several relevant aspects of the multi-model ensemble. Section 4.4 describes the construction of a metric of present-day performance that is correlated with climate sensitivity in the simple model but does not generalize to the multi-model ensemble. We conclude by exploring the implications for model weighting.

## 4.2 A simple ensemble spanning a range of errors and climate sensitivities

We construct a perturbed-parameter ensemble by varying the values of selected closure parameters (Table 4.1) in physical parameterizations of the general circulation model ECHAM5 (Roeckner et al. 2003). The parameters are uncertain in observations and are those used to adjust the model so that its energy budget is balanced at the top of atmosphere (to within observational uncertainties and accounting for ocean heat storage). Each parameter is restricted to fairly small ranges near the default and all parameters are sampled simultaneously using Latin hypercube sampling (McKay et al. 1979). Five hundred realizations of ECHAM5 are created and each model is run for a single year using present-day climatological distributions of sea ice and sea surface temperature.

For each ensemble member, we compute an aggregate measure of the error in simulating the present-day distribution of clouds, radiation, and precipitation. Because it is not known which observable aspects, if any, of the present-day climate are connected to climate sensitivity, any aggregate metric is arbitrary; we justify the narrow focus of our choice by noting that a) differences in cloud feedbacks drive much of the diversity in climate sensitivity estimates from climate models (Soden and Held 2006), particularly by affecting the radiation budget,

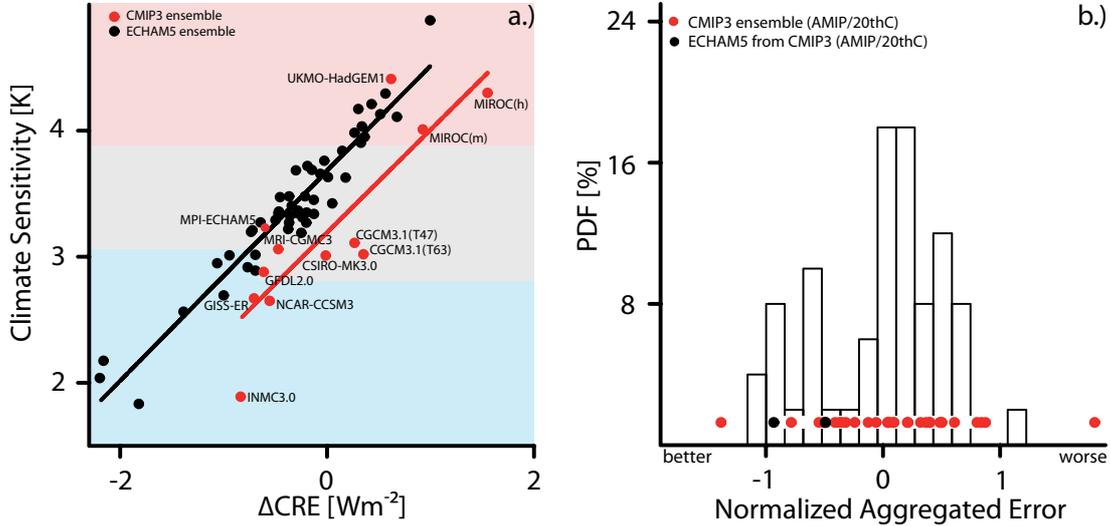
**Table 4.1:** List of perturbed parameters in the ECHAM5 ensemble, their description, default value, the range they are varied in and the percentage contribution to the variation in skill and climate sensitivity. <sup>1</sup>Default value in the atmosphere-only model. <sup>2</sup> Default value in the coupled model. \*Indicates coupled parameters, to keep top of the atmosphere radiative fluxes close to balance.

Description of parameter	Default value	Range	$R^2$ [%] Skill	$R^2$ [%] Sensitivity
Entrainment rate for shallow convection* (Tiedtke 1989)	0.0003	0.0003 - 0.001	3	44
Cloud mass flux above level of non-buoyanc* (Tiedtke 1989)	0.1 <sup>1</sup> /0.3 <sup>2</sup>	0.1 - 0.3333	3	44
Entrainment rate for penetrative convection (Tiedtke 1989)	0.0001	0.00001-0.0005	64	0
Conversion rate from cloud water to rain (Tiedtke 1989)	0.0004	0.0001-0.005	0	1
In-homogeneity of liquid clouds (Cahalan et al. 1994)	0.7	0.65 - 1	4	0
In-homogeneity of ice clouds (Cahalan et al. 1994)	0.7 <sup>1</sup> /0.8 <sup>2</sup>	0.65 - 1	20	1
Asymmetry of ice particles in clouds (Stephens et al. 1990)	0.91 <sup>1</sup> /0.85 <sup>2</sup>	0.75 - 1	0	1
Coefficient for horizontal diffusion	12	6 - 24	6	5
Gravity wave drag activation threshold (mean) (Lott 1999)	500	400 - 1000	2	0
Gravity wave drag activation threshold (stddev) (Lott 1999)	200	100 - 700	2	0
Albedo minimum of snow/ice	0.6/0.5	0.45 - 0.65	8	0
Albedo maximum of snow/ice	0.8/0.75	0.75 - 0.9	9	3

and b) a majority of the varied parameters are cloud-related. We compute the root-mean-square error relative to observations for cloud fraction, longwave and shortwave cloud radiative effects at the top of the atmosphere (e.g. Hartmann and Short 1980), and surface precipitation over each month of the annual cycle (Pincus et al. 2008). These errors are much larger in our short integrations than for long runs with well-tuned models because sampling errors are large. Still, the difference in errors based on individual years from longer runs (described below) is very small relative to the difference in error spanned by the ensemble, indicating that the diversity in error is robust. Errors in individual fields are standardized so that the distribution of each error across the ensemble has zero mean and a standard deviation of one, then added together to provide an aggregate error

## 4.2 CONSTRUCTING A SIMPLE ENSEMBLE

measure for each model, where low errors reflect greater skill relative to other members of the ensemble.

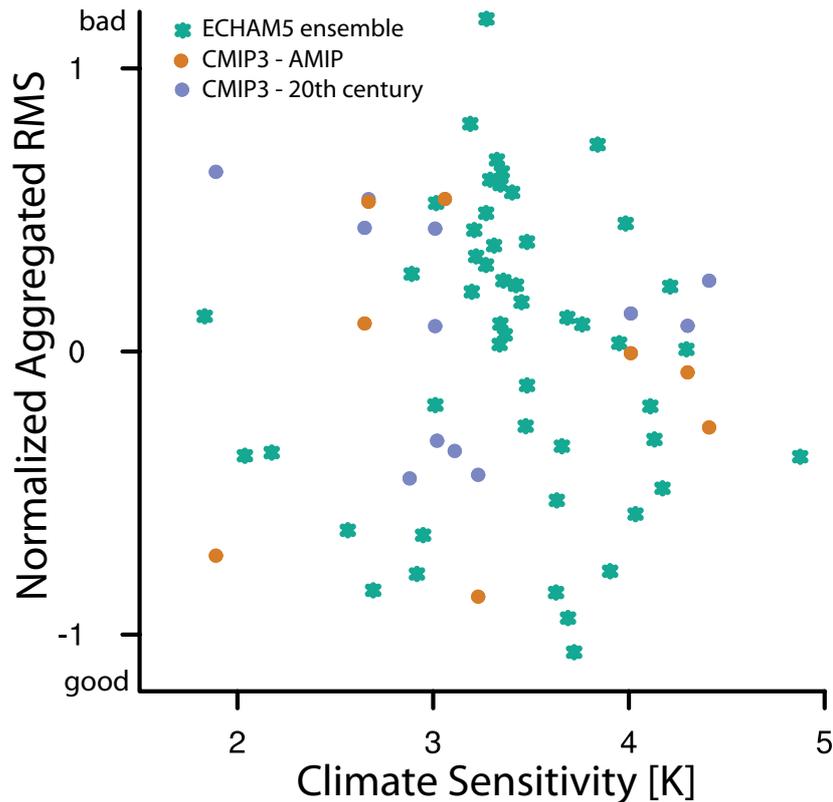


**Figure 4.1:** Climate sensitivity and skill in two ensembles of climate models. a) Equilibrium climate sensitivity as a function of the change in global annual mean net cloud radiative effect ( $\Delta\text{CRE}$ ) under doubled  $\text{CO}_2$  conditions. The CMIP3 ensemble is shown with red dots; the models are also labelled. The distribution of climate sensitivities is similar in the two ensembles, as is the mechanism driving the variability (the change in cloud radiative effect). Background colors indicate the highest (red) and lowest (blue) sensitivity models used later. b) Distributions of aggregate skill in present-day simulations of clouds, radiation, and precipitation for our perturbed-parameter ensemble (histogram) and from the CMIP3 ensemble (dots). The skill measure integrates over the annual cycle, the geographic distribution, and four variables. Black dots indicate the performance of the base ECHAM model (atmosphere-only and coupled to an ocean model) within the CMIP3 ensemble.

We sort the models according to this measure of aggregate error and compute the equilibrium climate sensitivity of every tenth model across the range of aggregate skill (so that the distribution of skill in the initial ensemble is roughly preserved). Ten-year runs are performed using a slab ocean model and present-day greenhouse gas concentrations, from which we determine the flux corrections necessary to maintain present-day sea surface temperatures. A fifty-year simulation is then performed using the same ocean heat flux corrections but with doubled carbon dioxide concentrations. Equilibrium climate sensitivity is computed as the difference in global mean surface temperature between the last ten years of the doubled  $\text{CO}_2$  and the present-day simulations.

### 4.3 The simple ensemble as proxy for the multi-model ensemble

Results from this ensemble, in which all diversity arises from parametric uncertainty, are comparable in many ways to the multi-model ensemble from the World Climate Research Programme’s Coupled Model Intercomparison Project phase 3 (CMIP3; see Meehl et al. 2007), which represents the majority of the world’s climate models and contains both parametric and structural variability. In particular, the distributions of climate sensitivity (Figure 4.1a) and our aggregated measure of model error (Figure 4.1b) are similar in both ensembles. These quantities are not systematically related to each other in either ensemble (Figure 4.2). The similarity in the distributions of error and sensitivity, as well as the lack of a connection between the two, mirror previous experiences across a wide range of perturbed-parameter ensembles (Murphy et al. 2004; Stainforth et al. 2005; Collins et al. 2011).



**Figure 4.2:** Global measure of skill, aggregated over cloud radiative effects, precipitation and cloud cover are unrelated to climate sensitivity in a simple ensemble and the multi-model CMIP3 ensemble.

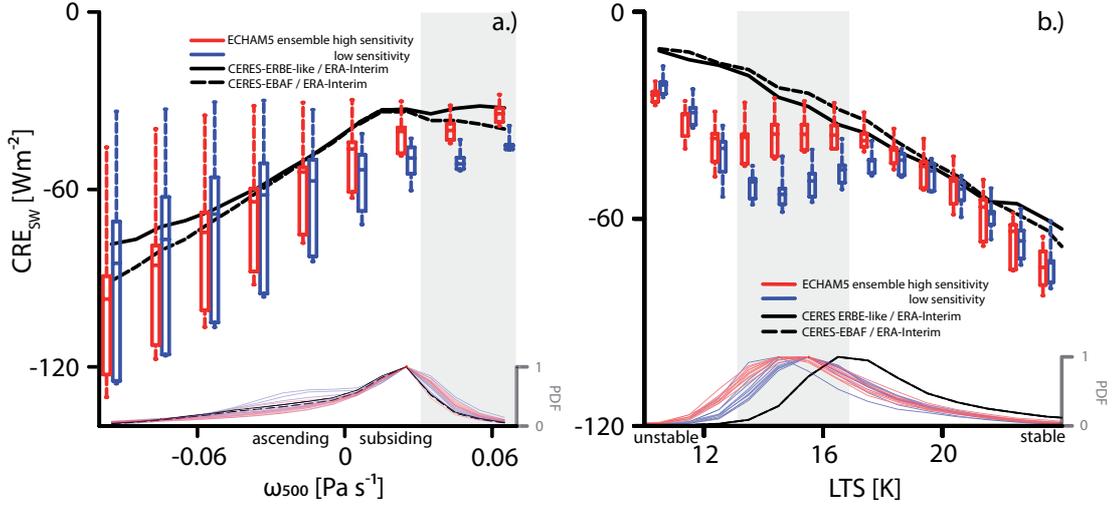
The two ensembles also share an important structural feature: the same mechanism underlies the variability in climate sensitivity. In both ensembles, models with a large change in the net cloud radiative effect under doubled CO<sub>2</sub> concentrations are those with higher climate sensitivity (Figure 4.1a). The longwave cloud radiative effect in our ensemble does not change much between present-day and doubled CO<sub>2</sub> conditions, while diversity in shortwave cloud radiative effect (CRE<sub>SW</sub>) changes, in turn, is largely driven by diversity in the response of low-latitude oceanic boundary layer clouds (Bony and Dufresne 2005).

By these measures, the perturbed-parameter ensemble is a successful proxy for the multi-model ensemble. This allows us to test the generality of model weighting techniques in two structurally distinct but statistically similar ensembles.

## 4.4 Developing measures of model error linked to climate sensitivity

We now design a measure of error in reproducing the present-day climate that is explicitly related to climate sensitivity in our simple ensemble. We identify such a measure by focussing on the low-latitude oceanic boundary layer clouds whose response is tightly linked to climate sensitivity (Bony and Dufresne 2005). Boundary layer clouds dominate CRE<sub>SW</sub> in subsidence regions, i.e. where the mid-tropospheric pressure velocity is downward ( $\omega_{500} > 0$ ), so we sort present-day CRE<sub>SW</sub> by this quantity (Bony et al. 2004). In our ensemble the present-day distribution of CRE<sub>SW</sub> in subsidence regions differs markedly between the ten highest- and ten lowest-sensitivity model variants (Figure 4.3a). Higher sensitivity models have weaker values of CRE<sub>SW</sub>, indicating that clouds are some combination of less frequent, less extensive, or less reflective than in low-sensitivity simulations. The higher sensitivity models are also more consistent with observations (here, cloud radiative effect derived from satellite observations (Wielicki et al. 1996; Loeb et al. 2009) and sorted by  $\omega_{500}$  inferred from ERA-Interim reanalysis data (Simmons et al. 2007)). Although the highest- and lowest-sensitivity models in our ensemble are distinct from each other, at the most frequent values of subsidence essentially all members over-estimate CRE<sub>SW</sub> relative to observations. In regions of large-scale ascent ( $\omega_{500} < 0$ ) the distributions of CRE<sub>SW</sub> in the highest- and lowest-sensitivity models are much broader and overlap significantly.

In nature, boundary layer clouds in subsiding regions over the oceans are further correlated (Medeiros and Stevens 2010) with lower tropospheric thermodynamic stability (LTS; see Bretherton and Wyant 1997; Klein and Hartmann 1993), here defined as the difference in the potential temperature at 1000 hPa and 700 hPa. Our simple ensemble reproduces this dependency as well (Figure 4.3b). Through

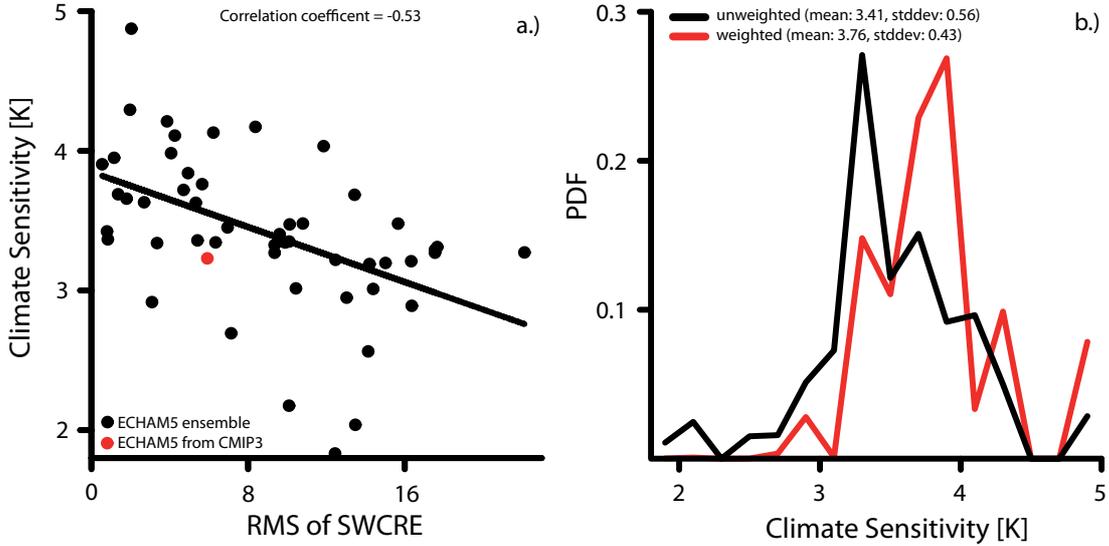


**Figure 4.3:** Relationships between present-day cloud properties and atmospheric state in a perturbed-parameter ensemble. Both figures are restricted to the tropical ( $30^{\circ}S - 30^{\circ}N$ ) oceans. The ten highest- and lowest-sensitivity models (red and blue, respectively) in the perturbed-parameter ensemble are shown; box and whisker plots summarize the medians (central lines), quartiles (box ends), and range (whiskers) of the distributions. Observations are shown in black, and the frequency distribution of models and observations in the lower part of each panel. a) Monthly-mean values of shortwave cloud radiative effect  $CRE_{SW}$  (all-sky fluxes minus clear-sky fluxes) sorted by mid-tropospheric pressure velocity  $\omega_{500}$ . Boundary-layer clouds dominate in subsiding ( $\omega_{500} > 0$ ) regions where high- and low-sensitivity models in our ensemble are distinct. Global measures of skill, though, are dominated by the errors unrelated to climate sensitivity occurring through the entire domain. The grey area indicates regions used in Figure 4.3b. b) Cloud radiative effect in subsidence regions ( $\omega_{500} > 0.03 Pa s^{-1}$ ) sorted by lower tropospheric stability. The grey background color indicates regions used for weighting in Figure 4.4b. High- and low-sensitivity models are distinct through a 4 K range of stability, though the ensemble is systematically roughly 2 K less stable than is observed.

much of the range of LTS the highest- and lowest-sensitivity models are indistinguishable, but in the range  $13 K < LTS < 17 K$   $CRE_{SW}$  in the high-sensitivity models is consistently weaker, and in better agreement with observations, than for low-sensitivity models. These are the most frequent values of LTS in subsiding regions in our ensemble.

Figure 4.3 demonstrates why global measures of skill are unrelated to model climate sensitivity: because the clouds whose systematic changes explain the diversity in sensitivity occur in a small region of the globe. Most measures of skill compare models to observations in global domains (e.g. Gleckler et al. 2008;

#### 4.4 MEASURES OF MODEL ERROR LINKED TO CLIMATE SENSITIVITY



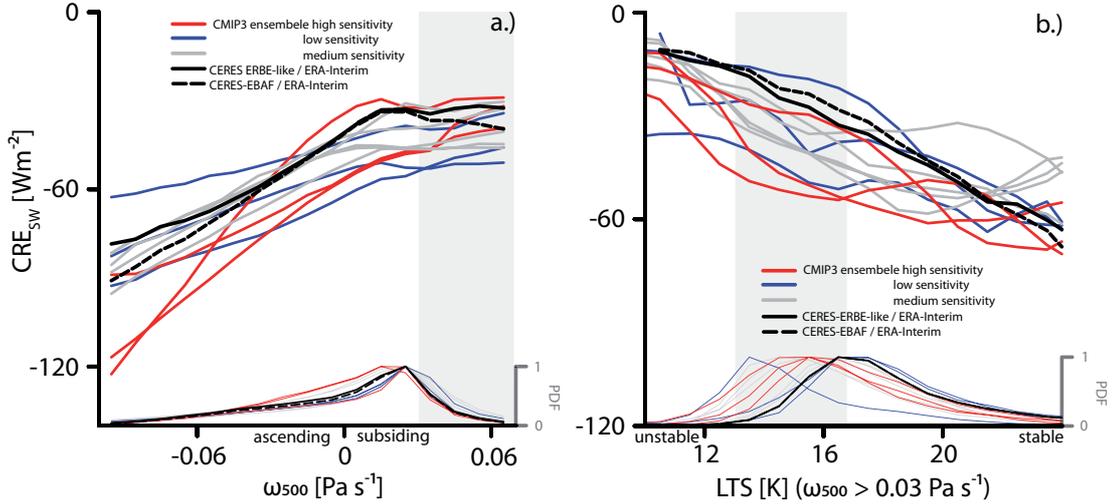
**Figure 4.4:** A tightly-focused measure of skill narrows the distribution of climate sensitivity in a simple ensemble. a) Equilibrium climate sensitivity as a function of conditionally-sampled root-mean-square error in shortwave cloud radiative effect of simulations compared to satellite observations. The error is computed only in regions of descending air ( $\omega_{500} > 0.03 \text{ Pa s}^{-1}$ ) and moderate lower tropospheric thermodynamic stability ( $13 \text{ K} < LTS < 17 \text{ K}$ ) over tropical oceans. b) Distributions of climate sensitivity estimates before (black) and after weighting by a function of the error in panel a. Weighting by this metric decreases the standard deviation of the distribution by about 23% and increases the mean by 0.35 K.

Pincus et al. 2008; Reichler and Kim 2008). Restricting the geographical domain over which errors are computed would not change this result much: even considering only the low-latitude oceans, the root-mean-square difference with observations are influenced not only by the regions controlling the sensitivity but also by ascending regions, where errors are large, and low-sensitivity models perform somewhat better, on average.

We define instead a conditioned error measure  $E_c$  as the root-mean-square difference between model simulations and observations of  $\text{CRE}_{\text{SW}}$  integrated over regions with large-scale subsidence ( $\omega_{500} > 0.03 \text{ Pa s}^{-1}$ ) and moderate lower tropospheric stability ( $13 \text{ K} < LTS < 17 \text{ K}$ ).

Regions satisfying both conditions comprise just 5% of the area of the tropics (2.5% of the globe) in the observations and somewhat more in the models. Nonetheless,  $E_c$  is a reasonably good predictor of climate sensitivity in the simple ensemble (Figure 4.4), which means it can be used to narrow the distribution of climate sensitivity estimates. Figure 4.4b shows the distribution of climate sensitivity obtained from the perturbed-parameter ensemble before and after weighting

by the likelihood  $L(E_c) = \exp(-E_c^2/2)$  (Murphy et al. 2004). The standard deviation of the posterior distribution is 3/4 of that of the prior distribution, mostly because a few models with low sensitivity have large errors and hence low weight. The mean climate sensitivity also increases by 0.35 K.



**Figure 4.5:** Relationships between present-day cloud properties and atmospheric state in a multi-model ensemble. These figures are constructed in the same way as Figure 4.3, but the distribution of cloud radiative effect as sorted by  $\omega_{500}$  (a) or lower tropospheric stability in subsiding regions (b) does not distinguish between high- and low-sensitivity models in the CMIP3 ensemble.

But despite the many similarities between the perturbed-parameter and multi-model ensembles, the systematic relationship between climate sensitivity and  $E_c$  does not carry into the multi-model ensemble (Figure 4.5), nor does the distribution of sensitivity estimates from the multi-model ensemble change when weighted by  $L(E_c)$ .

## 4.5 Implications for weighting projections from multi-model ensembles

One could conclude that we have obtained a null result and that the single-model perturbed-parameter ensemble is, after all, a poor proxy for the multi-model ensemble. Instead, we propose that these calculations are a concrete illustration of some of the issues involved in the weighting and more general interpretation of multi-model ensembles.

First, our results confirm that it is possible to obtain distributions of climate sensitivity and global measures of error as diverse as those produced by the multi-

## 4.5 IMPLICATIONS FOR WEIGHTING MULTI-MODEL PROJECTIONS

model ensemble with even modest variations about a single model. This suggests that variability in error and sensitivity at these levels is easy to come by (though why this is so remains an intriguing open question). In fact, in our ensemble diversity in skill and climate sensitivity arises from surprisingly simple parametric sensitivity: Climate sensitivity is primarily related to the entrainment rate for shallow convection, which varies along with a cloud mass flux parameter (explaining 44% of the variance in climate sensitivity; Table 4.1), while aggregate error is related to another parameter, the entrainment rate for deep convection (explaining 64% of the variance in aggregated error; Table 4.1). If broad diversity in behavior can arise from underlying simplicity then the diversity itself is uninformative. This is an illustrative reminder that the distribution of climate sensitivity from any model ensemble can not be interpreted as an estimate of the total uncertainty in climate sensitivity.

Second, while the motivation to narrow the distribution of climate sensitivity estimates is strong, our results dramatize the danger of focusing exclusively on this goal. Relationships between sensitivity and model fidelity in any ensemble emerge from an unknown mix of underlying similarity in model representation and error, statistical sampling error, and physical relationships also present in the natural world. This means that arbitrarily-chosen error measures may arise from underlying similarity not present in the physical climate system. We argue that because metrics developed from the full multi-model ensemble alone can not be falsified by comparison to more general ensembles, they can not be justified as a model likelihood purely on the basis of the strength of the statistical connection between that metric and climate sensitivity. Indeed, where observations have been used successfully to constrain model response (Hall and Qu 2006; Clement et al. 2009) statistical metrics have been bolstered by physical arguments. Much depends on the way weights are chosen, since incorrect weighting (that is, weighting not related to true model likelihood) can substantially reduce the benefits of using an ensemble of projections (Weigel et al. 2010).

Finally, it is possible that present-day models are not yet sufficiently accurate to benefit from model weighting. Weighting model projections by skill is an assertion that models are likely to produce accurate estimates of future climate in proportion to their ability to reproduce some aspects of the present-day climate; the implicit assumption is that models with higher skill are more likely to be accurate representations of the physical climate system. But by most measures, no current climate model produces distributions of the present-day climate statistically consistent with observations (Gleckler et al. 2008; Pincus et al. 2008, see also Figure 4.3 and 4.5), implying that all models are formally unlikely. Weighting an ensemble under these circumstances is essentially asserting that incorrect models are more reliable than even-more-incorrect models. But the result of Bayes's the-

## CHAPTER 4 CONSTRAINING CLIMATE SENSITIVITY BY WEIGHTING

orem is ambiguous when the system being modeled is far from the system being observed, and it may be that model weighting will be more profitable when the collection of models we have is closer to the world we observe.

## Chapter 5

# Towards the use of data assimilation for process-oriented climate model evaluation and climate sensitivity constraints

This chapter outlines how process-oriented climate model evaluation might improve our understanding of which processes are driving climate sensitivity. A possibility to challenge climate models directly with observations is to apply initial value problems, as in numerical weather prediction. Here we follow the avenue proposed by Rodwell and Palmer (2007), who use data assimilation to obtain an optimal initial state for short climate model integrations. This initial state is the most consistent with observations, and how much the model solution drifts from this state is used to analyze model errors on time scales of a few hours. Using this approach in ECHAM requires technical modifications, in particular the integration of the climate model in a data assimilation/forecast cycle, whereby the model starts from the best estimate of the atmospheric state. Here we use the same perturbed physics ensemble as in Chapter 4, where models only differ in their representation of clouds. The physical processes related to clouds act on “fast” timescales (i.e., minutes to days) and therefore the errors associated with cloud physics evolve quickly. Consequently, cloud parameterizations in climate models are well suited to be evaluated with this method.

The technical modifications to ECHAM and an interface for an ensemble data assimilation framework are presented followed by an outline of the potential use of the data assimilation technique in the spirit of results from the previous chapters. This can help in the future to

more rigorously evaluate climate models on a more process oriented level and might lead to better simulations of cloud processes which eventually improves our understanding in cloud responses to climate perturbations.

## 5.1 Introduction

Projections of climate models can be verified only on long time scales. Observational data for the climate response to large forcings is sparse (for an example see Chapter 2), hence confidence in climate models has to be built by testing them against present and past climates, for which observations of quantities considered important for climate simulations are available. This is generally done by comparing long-term statistics from climate model output to long-term statistics from observations (Reichler and Kim 2008; Gleckler et al. 2008). Although climate models continuously increased their ability to simulate the present-day and past climate over the last decades (Reichler and Kim 2008), climate projections, quantified by estimates of climate sensitivity, did not converge in the same time (compare for example Charney 1979 and Solomon et al. 2007).

The simulated atmospheric state at a given time does not necessarily correspond to the real atmospheric state, because climate models solve boundary condition problems, where usually arbitrary initial conditions are used. One reason for this is that the predictability of weather is limited to a few weeks, given the current observation systems and numerical models, while climate timescales are much longer. Therefore processes in climate models cannot be evaluated by a direct comparison with observations (an exception may be the use of “nudging” techniques to evaluate individual model parameterizations). Errors in long-term statistics in turn do not necessarily relate to errors in just one process; they could be caused by other processes interacting with the process of interest. Especially fast physical processes, like those related to clouds, react on time scales of minutes to hours to changes in environmental conditions. Comparisons of long-term statistics can identify errors, but may not untangle the processes that cause them.

Instead we can evaluate processes more directly, by challenging climate models with observations. Climate models can be initialized with the best estimate of the atmospheric state at a given time and make short-term forecasts, comparable to Numerical Weather Predictions (NWP). Phillips et al. (2004) initialized a climate model with reanalysis data to evaluate it within a five day forecast. They attributed the mean forecast error to deficiencies in physical parameterizations. Another option is to initialize a climate model with the use of data assimilation as in NWP. Based on this idea, Rodwell and Palmer (2007) adapted NWP meth-

ods for an evaluation of “fast” physical processes. They assimilated observational data in the climate version of their model and analyzed assimilation increments after every six-hour forecast, as a measure for errors in the representation of fast physical processes. Tendency diagnostics of physical parameterizations are used as error diagnostics to identify for contributions of different physical processes to the assimilation increment. This overcomes the dependence on a control model used to generate the reanalysis, as in the Phillips et al. (2004) approach, which may be inconsistent with the physics of the forecast model. The so called “initial tendency” approach requires the integration of the model in a data assimilation/forecast cycle.

The initial tendency approach can only assess fast physics perturbations, i.e. perturbations that have an impact on weather forecasts as well as the simulated climate. Recent research suggests that most of the present climate model parameter uncertainty is associated with clouds (e.g. Cess et al. 1990; Colman 2003; Klocke et al. 2011a,b), and cloud-related processes act on short time-scales, thus they are ideal to be evaluated with this approach.

Especially the response of low clouds in the tropical marine boundary layer has been identified as the main contributor to the inter-model spread in climate sensitivity within the Coupled Model Inter-comparison Project phase-3 ensemble (CMIP3) (Bony and Dufresne 2005). Klocke et al. (2011a) related present-day features of this cloud regime to climate sensitivity in a simplified ensemble. This suggests that differences in climate sensitivity are process-related and geographically restricted, especially as global measures of skill fail to relate an observable of the present-day climate to climate sensitivity.

Processes related to clouds are not directly simulated in climate models, but are represented via parameterizations. Such statistical formulations include empirical parameters, which are not well constraint by observations. Especially cloud parameters are used to tune the radiation balance of climate models, in order to obtain the best fit of the model to the present-day climate. The good results models achieve by tuning may be due to compensating errors. Large uncertainties in estimated climate sensitivities can be related to those parameters. These uncertainties can be explored systematically with large “perturbed physics” experiments (Murphy et al. 2004; Stainforth et al. 2005; Klocke et al. 2011a). Such experiments are computationally expensive, as the models need to be run in many configurations and there is a large number of parameters in a climate model. Table 4.1 in Chapter 4, for example, shows the parameter uncertainties that Klocke et al. (2011a) explored. Most of the parameters are cloud related, similar to the approaches in other perturbed physics experiments (e.g. Murphy et al. 2004).

In this study we outline a strategy to make use of data assimilation techniques, to initialize the climate model ECHAM from the best estimate of the atmospheric

state at a given time. We propose to use 3-D assimilation increments of temperature, wind components, surface pressure and specific humidity in the same set of models used by Klocke et al. (2011a) (see Chapter 4) to more rigorously identify processes that determine the skill of the present-day simulation. Can we show that errors in the climatology already manifest themselves in the first few time-steps? Do local errors in the representation of clouds have a remote influence? What we wish to achieve, is identifying a process which is related to climate sensitivity, rather than some statistical relation.

Another possible approach to use the data assimilation framework in a climate model is to systematically assess model uncertainty as investigated by Annan et al. (2005). They use an Ensemble Kalman Filter data assimilation scheme in which some model parameter values are included as part of the model state vector, and where the cost function involves a test of the model's simulation of present-day observations. For a simplified atmospheric model they are able to reproduce three out of five known parameter values in identical twin experiments where the model was assumed to be perfect (except for the unknown parameter values). Further work is required to determine the efficiency of this approach when applied to more complex, more non-linear, and less perfect models, with more tunable parameters like the climate model ECHAM used here. Such a study is currently performed as a diploma thesis project at MPI-M (Schirber, in preparation).

In section 5.2 modifications to ECHAM and the interface with the data assimilation system DART is described. In section 5.4 we show some first results obtained with this method and in section 5.5 we give a brief outlook to future work.

## 5.2 Technical notes on the integration of ECHAM in a data assimilation/forecast cycle

We use an ensemble filter based data assimilation approach, because in contrast to variational data assimilation techniques no tangent-linear, or adjoint approximation is required. The two approaches are also comparable in their performance (Kalnay et al. 2007; Buehner et al. 2010). Different ensemble filters are included in the Data Assimilation Research Toolbox (DART, Anderson et al. 2009) which is relatively easy to adapt and freely available to the research community.

DART is developed at the National Center of Atmospheric Research (NCAR). It is a flexible and comprehensive ensemble filter system, designed for research and education. Its modular structure allows to incorporate new models, filters, or observation types with relatively small effort, via interface routines.

The modifications to ECHAM and DART are necessary to link them as an ensemble data assimilation/forecasting system. On the climate model side, the first time-step after a restart needs to be modified. ECHAM restarts from two time-levels, but observations are assimilated only at one time-step. On the DART side, transformation programs are necessary to transform the ECHAM restart files into the required state vector and back, after each data assimilation/forecast cycle. DART also needs information about the ECHAM variable names, grid specifications and time stepping in order to assign suited assimilation windows and to ensure the right interpolation of observations in the vertical and horizontal (section 5.2.2). Finally, an overarching program is designed to drive the ensemble forecast system to advance the model ensemble, organize observations and transform data between state vectors for the assimilation framework and spectral space for the climate model restart from the analysis (section 5.2.3).

### 5.2.1 Modifications to the model restart

ECHAM is developed for climate simulations and has not been used as part of a data assimilation/forecast cycle, although it originally emerged from the forecast model of the European Centre for Medium-Range Weather Forecasts (ECMWF). Climate simulations are a boundary condition problem, therefore little attention was paid to the initialization of the model during the model development.

For the experiments proposed here the initialization and starting of the model plays a crucial role. In order to properly start the model from the analysis created by data assimilation, modifications to the model start are necessary. ECHAM uses a semi-implicit two time-level leapfrog time discretization scheme (see Figure 5.1). Prognostic variables are vorticity, divergence, temperature ( $T$ ), the logarithm of the surface pressure ( $P_S$ ), specific humidity ( $Q$ ) and cloud water ( $X_1$ ). From vorticity and divergence the meridional wind ( $U$ ) and zonal wind ( $V$ ) components can be calculated. The DART system assimilates observations at one time level  $t$ , but for a standard model restart, two time levels are needed. There are two options for a model start from the analysis, when incorporated in the data assimilation/forecast cycle. These are to

1. include both time levels in the state vector (i.e.,  $T(t)$ ,  $U(t)$ ,  $V(t)$ ,  $Q(t)$ ,  $P_S(t)$  and  $T(t-1)$ ,  $U(t-1)$ ,  $V(t-1)$ ,  $Q(t-1)$ ,  $P_S(t)$ ) and assimilate the observations only on time level  $t$ , but adjust the time level  $t-1$  according to the covariances across the ensemble between time level  $t$  and  $t-1$ .
2. use only the time level  $t$  in the state vector, but adjust the model start to a one time level restart, by changing the very first time step to start from the analysis.

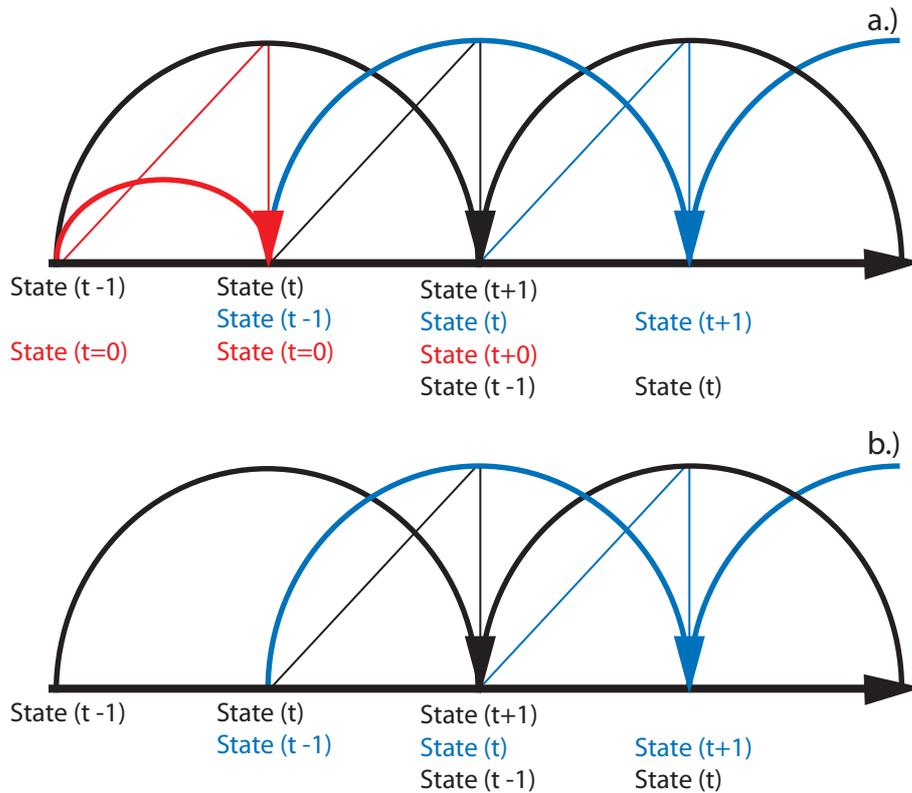
Computational costs increase with the number of state variables in the state vector and the number of observations. For this reason we use option 2. For this, the strategy used for initial model starts, which is different from a model restart, is applied.

Climate simulations with ECHAM are initialized with a subset of variables. The prognostic variables (vorticity, divergence, temperature, surface pressure and specific humidity) and some boundary conditions (e.g. sea-surface temperature, sea-ice cover, orography, vegetation) are loaded. All other variables are pre-defined by some default value (usually zero) and they develop their physically defined state within the first couple of time steps (days) of the simulation. This means that the model is not starting from a “realistic“ atmospheric state, but the realistic state is developed during the first time steps. At the model initialization variables are only available on one time level ( $t$ ) and ( $t-1$ )-values are obtained by copying the actual values (at time  $t$ ) into the first time-step. This first time-step is only a forward half Euler time-step and not a leap-frog-step, because of the equality of  $t$  and ( $t-1$ )-values (Figure 5.1a). Actual values (at time  $t$ ) are the spectral coefficients read. They are transformed to Fourier coefficients during the second time step, followed by a transformation to grid point values via inverse Fourier transformation.

This is done in the subroutine `ioinitial`. A description of the time stepping algorithm for an initial start is illustrated in Figure 5.1a. The model starts from a single time level and performs an Eulerian step forward with half the time step length of the full leapfrog time integration scheme. The calculated new model state is used to calculate the tendencies of the first real leapfrog time step.

The restart of ECHAM is handled by the subroutine `iorestart` and its descendants. The restart file contains a full description of the model state at  $t-1$  in grid point space. The prognostic variables of time level  $t$  in an intermediate symmetric/anti-symmetric split set of Fourier coefficients are used to calculate the tendencies for advancing the model to  $t+1$  in a way that is consistent with the model time discretization (see Figure 5.1b). This means that the restart information stored ensures a simulation identical to an uninterrupted one. In contrast to an initial start, a model restart starts from time level  $t-1$  and uses the available prognostic variables of time level  $t$  to advance to time level  $t+1$ .

We change the model restart for the prognostic variables (for other variables the restart remains the same) such that it is analogous to the initial model start. The spectral prognostic variables (vorticity, divergence, temperature and log surface pressure) at time-level  $t$  from the restart file are used, but need to be patched for the use in DART. The symmetric and asymmetric components of the Fourier coefficients need to be recombined and transformed to the grid point space. Zonal and meridional wind are calculated from divergence and vorticity and are written

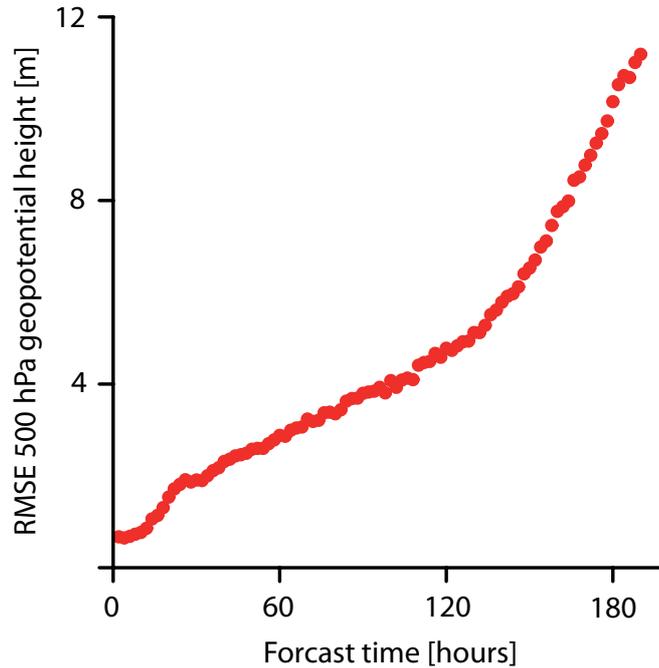


**Figure 5.1:** Schematic illustration of time stepping on an initial models start (a) and at model restart (b). Black and blue illustrate the two time levels, while the red color shows the half Euler time-step at model initialization.

together with specific humidity and temperature into a separate file. This file contains all state variables in grid space at time  $t$ . These variables are then ready to be transformed to a state vector for DART and updated by the data assimilation filter.

After DART updated the state vector, all variables are transformed back to spectral space and divergence and vorticity are calculated from the two wind components. A new IO reading routine is created (`ionwp`), derived from subroutine `ioinitial`, which reads the analysis. All other variables are read from the restart file as usually done in a restart. This is similar to the way current NWP models are starting their forecasts. The first time step is then an Euler time-step with half the time-step length of the leap frog scheme. Next, the model advances as usual, until the next assimilation window is reached.

Figure 5.2 shows the error introduced through the modified restart in comparison to the default model restart. The root mean square error (RMSE) of the 500 hPa geopotential height from the modified restart is calculated against the



**Figure 5.2:** Deviation in geopotential height at 500 hPa, of the restart for one time-level restarts, modified to initialize the model with analysis from data assimilation, compared to the standard two-time-level restart, measured as the two hourly global mean root-mean square-error (RMSE).

default two time-level restart of the model. The small error introduced by the modified model restart is negligible in comparison to the rather brutal changes to the atmospheric state introduced by the observations to the model.

### 5.2.2 The interface to DART

Ensemble filter data assimilation systems are attractive, because they derive anisotropic and flow-dependent covariance estimates from an ensemble. ECHAM is interfaced with the Data Assimilation Research Testbed (DART; Anderson et al. 2009) system developed at the National Center for Atmospheric Research (NCAR), which is a comprehensive ensemble filter system designed for research and education. We use the default ensemble adjustment Kalman filter (EAKF; Anderson 2001), which is a square root filter and is implemented in DART with serial observation processing (Anderson 2003). The EAKF includes cross covariances between different observation- and state-space variables. Because of the extensive documentation in the literature, we do not include a description of the filter here but refer the interested reader to the DART related references above. The ensemble filter is used to assimilate surface pressure and 3-D observations of

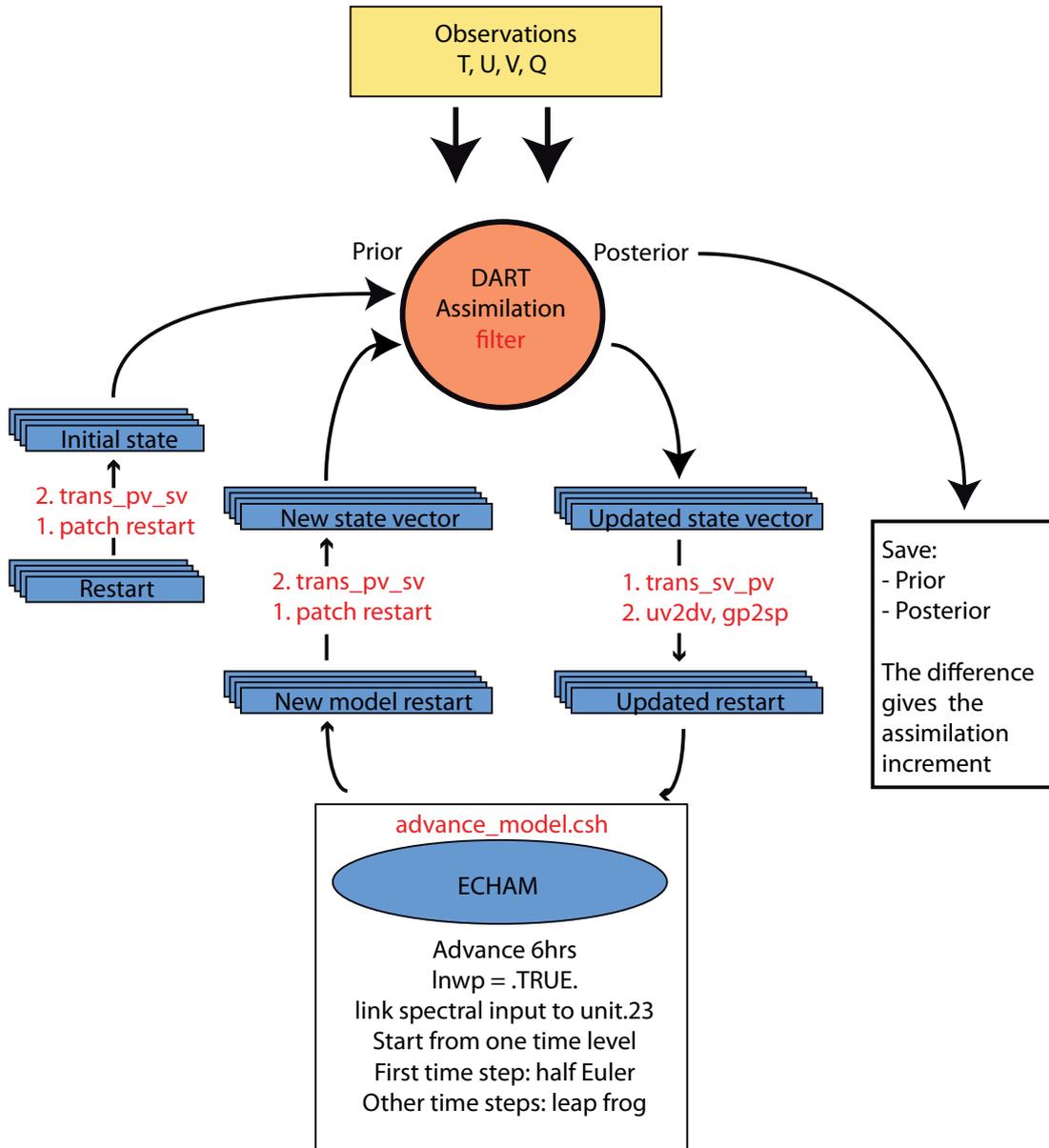
U-wind, V-wind, specific humidity and temperature to estimate the atmospheric states. The predictions analyzed are the 6-hrs ensemble forecasts initialized from the ensemble filter analyses.

In the ECHAM-DART system presented here, the initial conditions are generated by integrating the climate model for  $N$  years to get restart files for  $N$  ensemble members, each at the same time of the year. This ensures that the ensemble members are independent. Arbitrarily large ensembles can be generated, but we chose  $N=80$  members, always using the first of January and July as restart files for our experiments. Performing the experiments in boreal summer and winter ensures that the results are not season-dependent.

DART needs all variables in a state vector. Two separate programs transfer the relevant variables ( $T, U, V, Q, P_S$  plus optional tracers) from the model space to the state vector (`trans_pv_sv`) and, after the state vector is updated by DART, back (`trans_sv_pv`) to the model grid. The ECHAM-DART interface reads the state vector, determines the model time, model grid size and assimilation window, depending on the time-step length. For the low resolution T31 ECHAM5 version used here, the time step length is 40 minutes, which gives an assimilation window of  $t \pm 20$  minutes. In DART a location, a desired generic 'kind' (like `KIND_SURFACE_PRESSURE`, `KIND_TEMPERATURE`, `KIND_SPECIFIC_HUMIDITY`, ...) is assigned to each entry of the state vector and observations are interpolated to the ECHAM locations. The state variables (or observations) that are close to a given base observation are located. We assimilate  $T, U, V, Q, P_S$  from radiosondes, aircraft data and satellite data from the Global Positioning System radio occultation technique (GPS-RO; Anthes et al. 2008) and the Atmospheric Infrared Sounder (AIRS; Chahine et al. 2006).

### 5.2.3 Data assimilation/forecast cycle work flow

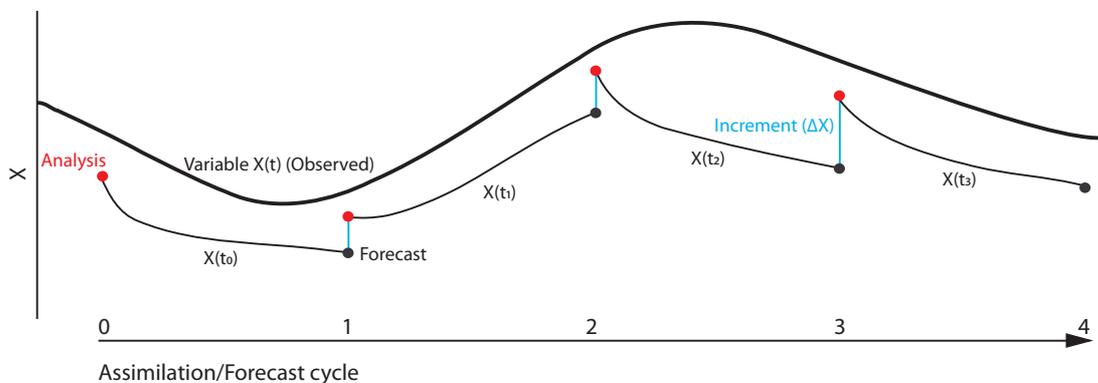
Figure 5.3 summarizes the data assimilation/forecast cycle of ECHAM-DART, including the intermediate transformation steps. A prior ensemble of  $N$  initial conditions (start files) needs to be created. To obtain these, the model is integrated for 80 years and each restart file of January and July first at midnight is saved for the initial ensemble. From those start files the initial ensemble of state vectors for DART is created, after the spectral variables are transformed to grid space and the U- and V-wind components are calculated. DART updates the initial ensemble of state vectors with observations and writes an ensemble of analyses. These updated state vectors are transformed to grid space and a second program transforms the wind components back to vorticity and divergence and writes these plus all other variables, except  $Q$ , back to spectral space. ECHAM starts from this one time level and advances the ensemble to the next assimila-



**Figure 5.3:** Work flow diagram of the ECHAM-DART data assimilation/forecast cycle. Red font indicates the separate routines used by the filter. Trans\_pv\_sv and Trans\_sv\_pv transform the prognostic variables between the model state and the state vector, respectively. Advance\_model advances the  $N$  ensemble members to the next assimilation window after each assimilation time. Patch restart transforms the model spectral variables to grid point space and calculates from divergence and vorticity the two wind components, while uv2dv and gp2sp do the reverse operation.

tion window. The new ensemble of restart files, six hours later from the initial ensemble, is the basis for the next cycle.

## 5.3 Assimilation increments as a skill measure



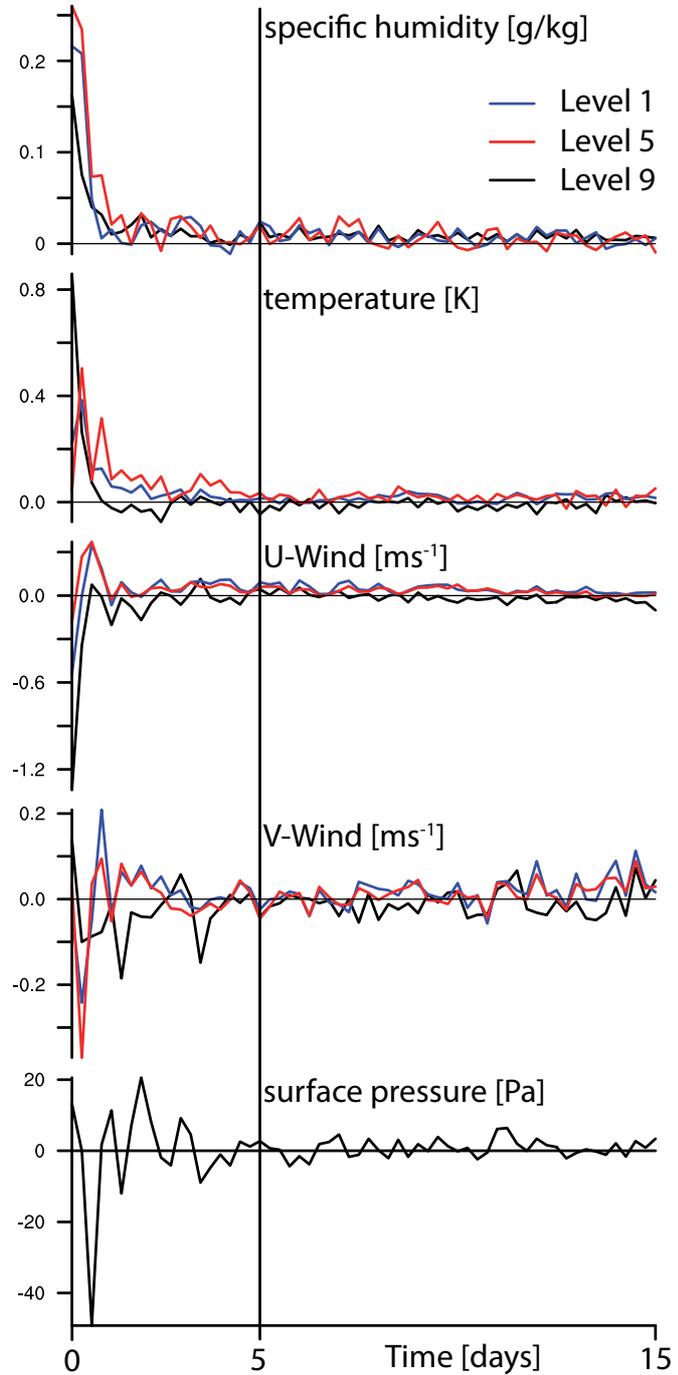
**Figure 5.4:** Schematic illustration of the data assimilation and forecast integration aspects of ECHAM-DART.  $X$  represents time series of an observable (e.g. of temperature, wind components, or specific humidity). For each  $i$ ,  $X_i(t_i)$  represents the model forecast initiated from an analysis. For the purposes of explaining the methodology, the role of systematic forecast error (negative bias) has been emphasized (adapted from Rodwell and Palmer (2007)).

In the perturbed physics ensemble used in Chapter 4, all models are identical, except for some parameters in the cloud parameterizations. The same models are initialized here every six hours with the best estimate of the atmospheric state, with the use of observations. How far a model drifts away from the observed trajectory of the atmospheric state relative to other models, depends entirely on the cloud parameterization. Maps of the assimilation increments of  $T$ ,  $U$ ,  $V$  and  $Q$  highlight regions with large increments. This may allow for relating the increments to individual processes.

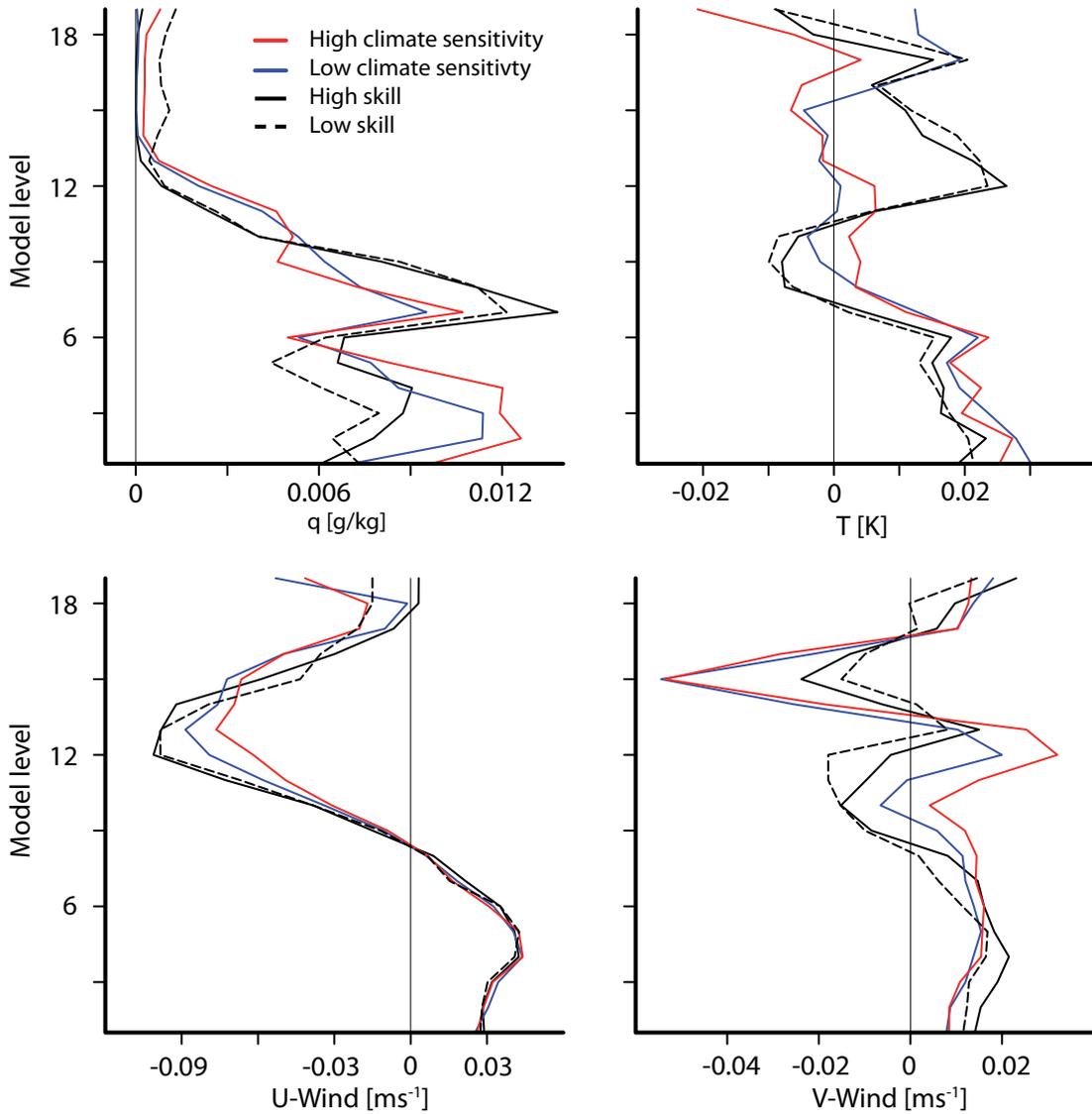
The mean assimilation increment in a six-hour forecast and  $m$  assimilation/forecast cycles for any variable  $X$  can be written as:

$$\text{INC} = \frac{1}{m} \sum_{i=1}^m \Delta X_i = \frac{1}{m} \sum_{i=1}^m (X(t_i) - X(t_{i-1} + \Delta t)) \quad (5.1)$$

This is also illustrated in Figure 5.4. The mean assimilation increment is the average departure of the model forecast from the analysis. A systematic model error would lead to a non-zero increment. A positive (negative) increment in a variable indicates that the model tends to drift to too small (large) values during the forecast, hence the data assimilation filter would pull the model back to larger (smaller) values.



**Figure 5.5:** Global mean assimilation increments from 1 January 2008 every six hours (00, 06, 12, 18 h) for specific humidity, temperature, U and V-wind components and surface pressure. Three different model levels are shown, level 1 being the lowest model level. The vertical line after five days indicates when the ensemble and observations are considered in equilibrium. All days following day five are used for model evaluation.



**Figure 5.6:** Vertical profiles of global mean assimilation increments averaged over ten days in January 2008 (00, 06, 12, 18 h) for specific humidity ( $q$ ), temperature ( $T$ ), zonal wind (U-Wind) and meridional wind (V-Wind). The models with highest (red) and lowest (blue) equilibrium climate sensitivity and the highest (black plain) and lowest (black dashed) skill in simulating the climatology of cloud related variables from the perturbed physics ensemble of Klocke et al. (2011a) are shown (see Table 5.1).

## 5.4 First results

Figure 5.5 shows the global mean assimilation increments for the surface pressure for 15 days in January 2008 and in three exemplary model levels for temperature, specific humidity and the two wind components (U, V) from ECHAM6. After

five days the ensemble has equilibrated with the observations (the assimilation increments reach stationarity) and in all remaining figures only the ten days after day five are used to obtain the mean assimilation increments.

In Figure 5.6 the mean assimilation increments for four different models from the ensemble of Klocke et al. (2011a) are shown as vertical profiles. The highest- and lowest-sensitivity model, plus the most and least skilled models (according to the cloud related skill measure defined in Chapter 4) are selected from the ensemble. For illustration purposes only four parameters are perturbed (see Table 5.1). Those four parameters are controlling most of the variability of skill and equilibrium climate sensitivity across the full ensemble used in Chapter 4.

**Table 5.1:** Parameter settings in the four members of the Klocke et al. (2011) ECHAM6 model ensemble chosen for the analysis. These are the models with the highest and lowest climate sensitivity and the best and worst aggregate skill, respectively. Parameters marked with (\*) are varied jointly in the perturbed-physics ensemble.

Description of parameter	High skill	Low skill	High sensitivity	Low sensitivity
Entrainment rate for shallow convection*	6.98E-04	6.62E-04	6.28E-04	3.94E-04
Cloud mass flux above level of non-buoyancy*	2.33E-01	2.21E-01	2.09E-01	1.31E-01
Entrainment rate for penetrative convection	1.15E-04	1.43E-04	3.20E-05	3.02E-05
Conversion rate from cloud water to rain	3.13E-03	4.91E-03	2.16E-03	4.90E-03

The vertical profiles of the specific humidity ( $q$ ) assimilation increments (Figure 5.6) are positive throughout the column for all four model versions. This indicates a too dry atmosphere, which is corrected in every assimilation cycle to higher specific humidities based on observations. Errors at higher model levels are relatively large for the low sensitivity model simply because the upper atmosphere has low specific humidity.

The specific humidity assimilation increments in the lower model levels are consistent with the profiles of temperature ( $T$ ) assimilation increments, which are positive, indicating a tendency of all model versions to a lower troposphere that is too cold. At higher model levels the two models with opposing climate sensitivity are closer to the observations while the models of opposing skill have positive assimilation increments (negative bias).

Biases in the winds are in the lower model levels, independent of the parameter values, which is especially true for the U-wind component. This suggests that this bias is unrelated to the perturbed cloud parameterizations. At higher levels the different models diverge.

## 5.5 OUTLOOK AND RESEARCH QUESTIONS

In all variables the largest biases are robust, independent of parameter settings. Interestingly the sensitivity models show comparable vertical profiles, and the two models with opposing skill have similar assimilation increments. A possible reason for this is that the parameter for the entrainment rate of penetrative convection is similar in the two model versions with opposing skill, and also similar in the two model versions with opposing climate sensitivity (see Table 5.1), but between these two sets, values are substantially different. Because deep convection in the tropics largely drives the general circulation, this parameter is expected to have a strong global mean impact. For this first pilot study the parameter values may however not be optimally chosen, which hinders some interpretations.

Because the parameter for the entrainment rate of penetrative convection is relevant for the skill in reproducing the climatology of cloud related variables, one would expect small assimilation increments for the model with high aggregate skill, and large ones for the model with low skill. Recalling that surface precipitation, cloud radiative effects and cloud cover were used, one would expect this to be true in particular in the lower troposphere and in humidity, that are most relevant for cloud formation. As can be seen in Figure 5.6, however, the opposite is found. It remains to be proven that the process-oriented evaluation proposed here is superior to the aggregate skill metric used earlier. It is expected that the value of this method improves when assimilation increments are evaluated on regional, rather than just global scales.

An explanation for the comparable assimilation increments for the two models with opposing climate sensitivity may be due to the finding that global measures of skill and climate sensitivity are broadly un-related (Chapter 4). This technique does provide 3-D assimilation increments, hence the evaluation can be refined to target certain processes in specific regions. In particular we expect a difference in assimilation increments of humidity in the subtropics, since cloudiness in these regions specifically was found to correlate with climate sensitivity. These questions will be investigated in more detail with the help of more perturbed model versions in future studies.

## 5.5 Outlook and research questions

The experiments here shown use just a few parameters from the subset of the perturbed parameter ensemble in Chapter 4. Further simulations with a larger set of models from the perturbed physics ensemble are necessary to obtain more meaningful results. Nevertheless, these preliminary results already demonstrate the potential of this method for the evaluation of fast physical processes. Especially when the mean assimilation increments are regionally confined several interesting

## CHAPTER 5 TOWARDS PROCESS-ORIENTED MODEL EVALUATION

research questions can be addressed with this approach. A few examples are:

- How do errors in the climatology manifest themselves in the first few time-steps?
- Can a certain fast process be related to assimilation increments and subsequently to errors in the climatology?
- Are there teleconnections of fast evolving errors?
- Can a specific process be linked to climate sensitivity?

# Chapter 6

## Conclusion and Outlook

**How sensitive is the Earth's climate system to perturbations?** This question intrigues many because future planning for mitigation and adaptation to a changing climate is easier the more certain we are about the answer to this questions. The range of climate sensitivity estimates did not narrow much in the past decades and it remains a great challenge to do so. Several paths to get closer to answering this question more accurately are followed and discussed in detail in this thesis.

In Chapter 2 the possibility of better constraining climate sensitivity from volcanic forcings is explored through the use of ensemble simulations of the last millennium with a comprehensive earth system model. These include all known forcings based on reconstructions. Here we simply assume that the response to the short pulse forcing is directly related to climate sensitivity. The composite of many different volcanic forcing realizations in many different states of the climate system is used to estimate how many volcanic eruptions would have to be observed to yield an estimate of climate sensitivity narrower than the one currently obtained from the CMIP3 multi-model ensemble. Few volcanoes, but with stronger radiative forcing are better suited for deriving a narrower composite normalized response, compared to many weak volcanic events. To achieve a comparable range of climate sensitivities as given by the multi-model ensemble of CMIP3, more than 45 volcanoes, all larger than Mt. Pinatubo (the eruption with the largest radiative impact in the last century) are needed. For a frequency of large volcanic eruptions comparable to the one in the last millennium, more than 10000 years are necessary to derive a tighter range of climate sensitivity from response following volcanic eruptions. This suggests that large volcanic events are in fact not useful to constrain equilibrium climate sensitivity.

Physical radiative feedbacks from idealized climate simulations are quantified in Chapter 3 using four different methods, with different levels of complexity. The results differ between the methods and differences are largest for the cloud feedback. The spatial and temporal variability of each feedback are used to estimate the averaging time scale necessary to satisfy the feedback concept of one constant

global mean value. We find that the year-to-year variability of each feedback process in this single model is comparable to the model-to-model spread in feedback strength of the CMIP3 ensemble. The strongest spatial and temporal variability is in the short-wave component of the cloud feedback. Even in our very idealized simulations, where many sources of natural variability are neglected, multi-year averages are necessary to get a reliable estimate of the simulated cloud feedback. Considering the large natural variability and relatively small forcing present in the real world, as compared to the forcing imposed by doubling CO<sub>2</sub> concentrations in the simulations, this implies that using observations to constrain the cloud feedback is a challenging task and requires reliable long-term measurements.

In Chapter 4 we provide a cautionary example showing that measures of model fidelity that are effective at narrowing the distribution of future projections (because they are systematically related to climate sensitivity in an ensemble of models) may be poor measures of the likelihood that a model will provide an accurate estimate of climate sensitivity (and so degrade distributions of projections if they are used as weights). This conclusion is achieved considering two ensembles: one obtained by perturbing parameters in a single climate model, and a second containing the majority of the world's climate models. The simple ensemble reproduces many aspects of the multi-model ensemble, such as the distributions of skill in simulating the present-day climatology of clouds and radiation, the distribution of climate sensitivity, and the correlation of climate sensitivity with the cloud feedback. By constructing an error metric for the subtropical marine low-level cloud regimes, we can identify a relationship between climate sensitivity and model error useful to obtain a narrower distribution of climate sensitivity in the simple ensemble. This relationship, however, does not carry into the multi-model ensemble. This suggests that model weighting based on statistical relationships alone is unfounded, and perhaps that climate model errors are still large enough that model weighting is not sensible.

Climate model errors are dominated by errors in fast physical processes on the sub-grid scale such as the ones related to clouds. These processes are dependent on the large scale atmospheric state. If evaluated as a climatology, errors are hard to isolate. For example could even a perfect cloud parameterization only deliver only a poor representation of clouds, if the thermodynamical conditions are erroneous. Thus an evaluation of the simulated cloud fields in this hypothetical model would come to the wrong conclusion that something is wrong with the cloud parameterization. Consequently, closure parameters in the cloud parameterizations would be adjusted to make the model fit better to the observed climate. By using data assimilation increments as skill measure this can be partly avoided and fast processes can be evaluated more directly. This approach is outlined in detail in Chapter 5 which allows to approach several interesting questions. Which errors

in fast processes are related to errors in the climatology? Is there a process which can be related to skill in the climatology and maybe even to the response of clouds to perturbations? These questions will be investigated in future studies using the introduced data assimilation/forecast framework in Chapter 5 on the perturbed parameter ensemble from Chapter 4.

The first three parts of the thesis explored several possibilities to quantify and constrain climate sensitivity but consistently concluded that these avenues fail to provide more reliable estimates of climate sensitivity. However, the data assimilation approach can potentially lead to better climate models, in conjunction with a better understanding of the climate's response to perturbations.



# Bibliography

- Anderson, J., T. Hoar, K. Raeder, H. Liu, N. Collins, R. Torn, and A. Avellano, 2009: The Data Assimilation Research Testbed: A community facility. *Bull. Am. Meteorol. Soc.*, **90**, 1283–1296, doi:10.1175/2009BAMS2618.1.
- Anderson, J. L., 2001: An ensemble adjustment Kalman filter for data assimilation. *Mon. Wea. Rev.*, **129**, 2884–2903, doi:10.1175/1520-0493(2001)129<2884:AEAKFF>2.0.CO;2.
- 2003: A local least squares framework for ensemble filtering. *Mon. Wea. Rev.*, **131**, 634–642, doi:10.1175/1520-0493(2003)131<0634:ALLSFF>2.0.CO;2.
- Annan, J. D., D. J. Lunt, J. C. Hargreaves, and P. J. Valdes, 2005: Parameter estimation in an atmospheric GCM using the Ensemble Kalman Filter. *Nonlinear Proc. Geoph.*, **12**, 363–371, doi:10.5194/npg-12-363-2005.
- Anthes, R. A., D. Ector, D. C. Hunt, Y.-H. Kuo, C. Rocken, W. S. Schreiner, S. V. Sokolovskiy, S. Syndergaard, T.-K. Wee, Z. Zeng, P. A. Bernhardt, K. F. Dymond, Y. Chen, H. Liu, K. Manning, W. J. Randel, K. E. Trenberth, L. Cucurull, S. B. Healy, S.-P. Ho, C. McCormick, T. K. Meehan, D. C. Thompson, and N. L. Yen, 2008: The COSMIC/FORMOSAT-3 mission: Early results. *Bull. Am. Meteorol. Soc.*, **89**, 313–333, doi:10.1175/BAMS-89-3-313.
- Arrhenius, S., 1896: On the influence of carbonic acid in the air upon the temperature on the ground. *Philos. Mag. and J. of Science*, **41**, 237–276.
- Bender, F., A. Ekman, and H. Rodhe, 2010: Response to the eruption of Mount Pinatubo in relation to climate sensitivity in the CMIP3 models. *Clim. Dyn.*, **35**, 875–886, doi:10.1007/s00382-010-0777-3.
- Boer, G., M. Stowasser, and K. Hamilton, 2007: Inferring climate sensitivity from volcanic events. *Clim. Dyn.*, **28**, 481–502, doi:10.1007/s00382-006-0193-x.
- Boer, G. J. and B. Yu, 2003: Climate sensitivity and climate state. *Clim. Dyn.*, **21**, 167–176, doi:10.1007/s00382-003-0323-7.

## BIBLIOGRAPHY

- Bony, S., R. Colman, V. M. Kattsov, R. P. Allan, C. S. Bretherton, J.-L. Dufresne, A. Hall, S. Hallegatte, M. M. Holland, W. Ingram, D. A. Randall, B. J. Soden, G. Tselioudis, and M. J. Webb, 2006: How well do we understand and evaluate climate change feedback processes? *J. Clim.*, **19**, 3445–3482, doi:10.1175/JCLI3819.1.
- Bony, S. and J.-L. Dufresne, 2005: Marine boundary layer clouds at the heart of tropical cloud feedback uncertainties in climate models. *Geophys. Res. Lett.*, **32**, L20806, doi:10.1029/2005gl023851.
- Bony, S., J.-L. Dufresne, H. Le Treut, J.-J. Morcrette, and C. Senior, 2004: On dynamic and thermodynamic components of cloud changes. *Clim. Dyn.*, **22**, 71–86, doi:10.1007/s00382-003-0369-6.
- Bretherton, C. S. and M. C. Wyant, 1997: Moisture transport, lower-tropospheric stability, and decoupling of cloud-topped boundary layers. *J. Atmos. Sci.*, **54**, 148–167, doi:10.1175/1520-0469(1997)054<0148:MTL TSA>2.0.CO;2.
- Buehner, M., P. L. Houtekamer, C. Charette, H. L. Mitchell, and B. He, 2010: Intercomparison of variational data assimilation and the ensemble Kalman filter for global deterministic NWP. Part I: Description and single-observation experiments. *Mon. Wea. Rev.*, **138**, 1550–1566, doi:10.1175/2009MWR3157.1.
- Cagnazzo, C., E. Manzini, M. A. Giorgetta, P. M. F. Forster, and J. J. Morcrette, 2007: Impact of an improved shortwave radiation scheme in the MAECHAM5 general circulation model. *Atmos. Chem. Phys.*, **7**, 2503–2515, doi:10.5194/acp-7-2503-2007.
- Cahalan, R. F., W. Ridgway, W. J. Wiscombe, T. L. Bell, and J. B. Snider, 1994: The albedo of fractal stratocumulus clouds. *J. Atmos. Sci.*, **51**, 2434–2455, doi:10.1175/1520-0469(1994)051<2434:TAOFSC>2.0.CO;2.
- Callendar, G. S., 1938: The artificial production of carbon dioxide and its influence on temperature. *Q. J. Roy. Meteor. Soc.*, **64**, 223–240, doi:10.1002/qj.49706427503.
- Cess, R. D., 1975: Global climate change: an investigation of atmospheric feedback mechanisms. *Tellus*, **27**, 193–198, doi:10.1111/j.2153-3490.1975.tb01672.x.
- Cess, R. D. and G. L. Potter, 1987: Exploratory studies of cloud radiative forcing with a general circulation model. *Tellus A*, **39A**, 460–473, doi:10.1111/j.1600-0870.1987.tb00321.x.

- Cess, R. D., G. L. Potter, J. P. Blanchet, G. J. Boer, A. D. Del Genio, M. Déqué, V. Dymnikov, V. Galin, W. L. Gates, S. J. Ghan, J. T. Kiehl, A. A. Lacis, H. L. Treut, Z.-X. Li, X.-Z. Liang, B. J. McAvaney, V. P. Meleshko, J. F. B. Mitchell, J.-J. Morcrette, D. A. Randall, L. Rikus, E. Roeckner, J. F. Royer, U. Schlese, D. A. Sheinin, A. Slingo, A. P. Sokolov, K. E. Taylor, W. M. Washington, R. T. Wetherald, I. Yagai, and M.-H. Zhang, 1990: Intercomparison and interpretation of climate feedback processes in 19 atmospheric general circulation models. *J. Geophys. Res.*, **95**, 16.601–16.615, doi:10.1029/JD095iD10p16601.
- Chahine, M. T., T. S. Pagano, H. H. Aumann, R. Atlas, C. Barnet, J. Blaisdell, L. Chen, M. Divakarla, E. J. Fetzer, M. Goldberg, C. Gautier, S. Granger, S. Hannon, F. W. Irion, R. Kakar, E. Kalnay, B. H. Lambrigtsen, S.-Y. Lee, J. Le Marshall, W. W. McMillan, L. McMillin, E. T. Olsen, H. Revercomb, P. Rosenkranz, W. L. Smith, D. Staelin, L. L. Strow, J. Susskind, D. Tobin, W. Wolf, and L. Zhou, 2006: AIRS: Improving weather forecasting and providing new data on greenhouse gases. *Bull. Am. Meteorol. Soc.*, **87**, 911–926, doi:10.1175/BAMS-87-7-911.
- Charney, J., 1979: Carbon dioxide and climate: A scientific assessment. *National Academy of Sciences*.
- Clement, A. C., R. Burgman, and J. R. Norris, 2009: Observational and model evidence for positive low-level cloud feedback. *Science*, **325**, 460–464, doi:10.1126/science.1171255.
- Collins, M., B. Booth, B. Bhaskaran, G. Harris, J. Murphy, D. Sexton, and M. Webb, 2011: Climate model errors, feedbacks and forcings: a comparison of perturbed physics and multi-model ensembles. *Clim. Dyn.*, to appear, doi:10.1007/s00382-010-0808-0.
- Colman, R., 2003: A comparison of climate feedbacks in general circulation models. *Clim. Dyn.*, **20**, 865–873, doi:10.1007/s00382-003-0310-z.
- Colman, R. A. and B. J. McAvaney, 1997: A study of general circulation model climate feedbacks determined from perturbed sea surface temperature experiments. *J. Geophys. Res.*, **102**, 19383–19402, doi:10.1029/97JD00206.
- Crowley, T. J., G. Zielinski, B. Vinther, R. Udisti, K. Kreutz, J. Cole-Dai, and J. Castellano, 2008: Volcanism and the little ice age. *PAGES Newsletter*, **16**, 22–23.
- Forster, P. M. F. and J. M. Gregory, 2006: The climate sensitivity and its components diagnosed from earth radiation budget data. *J. Clim.*, **19**, 39–52, doi:10.1175/JCLI3611.1.

## BIBLIOGRAPHY

- Frame, D. J., B. B. Booth, J. A. Kettleborough, D. A. Stainforth, J. M. Gregory, M. Collins, and M. R. Allen, 2005: Constraining climate forecasts: The role of prior assumptions. *Geophys. Res. Lett.*, **32**, L09702, doi:10.1029/2004GL022241.
- Friedlingstein, P., P. Cox, R. Betts, L. Bopp, W. von Bloh, V. Brovkin, P. Cadule, S. Doney, M. Eby, I. Fung, G. Bala, J. John, C. Jones, F. Joos, T. Kato, M. Kawamiya, W. Knorr, K. Lindsay, H. D. Matthews, T. Raddatz, P. Rayner, C. Reick, E. Roeckner, K.-G. Schnitzler, R. Schnur, K. Strassmann, A. J. Weaver, C. Yoshikawa, and N. Zeng, 2006: Climate-carbon cycle feedback analysis: Results from the C4MIP model intercomparison. *J. Clim.*, **19**, 3337–3353, doi:10.1175/JCLI3800.1.
- Gleckler, P. J., K. E. Taylor, and C. Doutriaux, 2008: Performance metrics for climate models. *J. Geophys. Res.*, **113**, D06104, doi:10.1029/2007jd008972.
- Gregory, J. and M. Webb, 2008: Tropospheric adjustment induces a cloud component in CO<sub>2</sub> forcing. *J. Clim.*, **21**, 58–71, doi:10.1175/2007JCLI1834.1.
- Gregory, J. M., W. J. Ingram, M. A. Palmer, G. S. Jones, P. A. Stott, R. B. Thorpe, J. A. Lowe, T. C. Johns, and K. D. Williams, 2004: A new method for diagnosing radiative forcing and climate sensitivity. *Geophys. Res. Lett.*, **31**, doi:10.1029/2003GL018747.
- Gregory, J. M., R. J. Stouffer, S. C. B. Raper, P. A. Stott, and N. A. Rayner, 2002: An observationally based estimate of the climate sensitivity. *J. Clim.*, **15**, 3117–3121, doi:10.1175/1520-0442(2002)015<3117:AOBEOT>2.0.CO;2.
- Hall, A. and X. Qu, 2006: Using the current seasonal cycle to constrain snow albedo feedback in future climate change. *Geophys. Res. Lett.*, **33**, doi:10.1029/2005GL025127.
- Hansen, J., A. Lacis, D. Rind, G. Russell, P. Stone, I. Fung, R. Ruedy, and J. Lerner, 1984: Climate sensitivity: analysis of feedback mechanisms. *Geophys. Mono.*, 130–163.
- Hansen, J., M. Sato, A. Lacis, and R. Ruedy, 1997: The missing climate forcing. *Phil. Trans. R. Soc. Lond. B.*, **352**, 231–240, doi:10.1098/rstb.1997.0018.
- Hartmann, D. L. and K. Larson, 2002: An important constraint on tropical cloud-climate feedback. *Geophys. Res. Lett.*, **29**, doi:10.1029/2002GL015835.

- Hartmann, D. L. and D. A. Short, 1980: On the use of earth radiation budget statistics for studies of clouds and climate. *J. Atmos. Sci.*, **37**, 1233–1250, doi:10.1175/1520-0469(1980)037<1233:OTUOER>2.0.CO;2.
- Held, I. and B. Soden, 2000: Water vapor feedback and global warming. *Ann. Rev. Energy Environ.*, **25**, 441–475, doi:10.1146/annurev.energy.25.1.441.
- Huybers, P., 2010: Compensation between model feedbacks and curtailment of climate sensitivity. *J. Clim.*, **23**, 3009–3018, doi:10.1175/2010JCLI3380.1.
- Jansen, E., J. Overpeck, K. Briffa, J.-C. Duplessy, F. Joos, V. Masson-Delmotte, D. Olago, B. Otto-Bliesner, W. Peltier, S. Rahmstorf, R. Ramesh, D. Raynaud, D. Rind, O. Solomina, R. Villalba, and D. Zhang, 2007: *Paleoclimate, in: Climate change 2007: the physical science basis. Contribution of working group 1 to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge, 433–497 pp.
- Jungclauss, J. H., S. J. Lorenz, C. Timmreck, C. H. Reick, V. Brovkin, K. Six, J. Segschneider, M. A. Giorgetta, T. J. Crowley, J. Pongratz, N. A. Krivova, L. E. Vieira, S. K. Solanki, D. Klocke, M. Botzet, M. Esch, V. Gayler, H. Haak, T. J. Raddatz, E. Roeckner, R. Schnur, H. Widmann, M. Claussen, B. Stevens, and J. Marotzke, 2010: Climate and carbon-cycle variability over the last millennium. *Clim. Past*, **6**, 723–737, doi:10.5194/cp-6-723-2010.
- Kalnay, E., H. Li, T. Miyoshi, S.-C. Yang, and J. Ballabrera-Poy, 2007: 4D-VAR or ensemble Kalman filter? *Tellus A*, **59**, 758–773, doi:10.1111/j.1600-0870.2007.00261.x.
- Klein, S. A. and D. L. Hartmann, 1993: The seasonal cycle of low stratiform clouds. *J. Clim.*, **6**, 1587–1606, doi:10.1175/1520-0442(1993)006<1587:TSCOLS>2.0.CO;2.
- Klocke, D., R. Pincus, and J. Quaas, 2011a: On constraining estimates of climate sensitivity with present-day observations through model weighting. *J. Clim.*, revised.
- Klocke, D., J. Quaas, M. Giorgetta, and B. Stevens, 2011b: Assessment of different feedbacks metrics. *Clim. Dyn.*, to be submitted.
- Knutti, R., M. R. Allen, P. Friedlingstein, J. M. Gregory, G. C. Hegerl, G. A. Meehl, M. Meinshausen, J. M. Murphy, G. K. Plattner, S. C. B. Raper, T. F. Stocker, P. A. Stott, H. Teng, and T. M. L. Wigley, 2008: A review of uncertainties in global temperature projections over the twenty-first century. *J. Clim.*, **21**, 2651–2663, doi:10.1175/2009JCLI3361.1.

## BIBLIOGRAPHY

- Knutti, R., R. Furrer, C. Tebaldi, J. Cermak, and G. A. Meehl, 2010: Challenges in combining projections from multiple climate models. *J. Clim.*, **23**, 2739–2758, doi:10.1175/2009JCLI3361.1.
- Knutti, R., G. A. Meehl, M. R. Allen, and D. A. Stainforth, 2006: Constraining climate sensitivity from the seasonal cycle in surface temperature. *J. Clim.*, **19**, 4224–4233, doi:10.1175/2007jcli2119.1.
- Loeb, N. G., B. A. Wielicki, D. R. Doelling, G. L. Smith, D. F. Keyes, S. Kato, N. Manalo-Smith, and T. Wong, 2009: Toward optimal closure of the earth’s top-of-atmosphere radiation budget. *J. Clim.*, **22**, 748–766, doi:10.1175/2008JCLI2637.1.
- Lott, F., 1999: Alleviation of stationary biases in a GCM through a mountain drag parameterization scheme and a simple representation of mountain lift forces. *Mon. Wea. Rev.*, **127**, 788–801, doi:10.1175/1520-0493(1999)127<0788:AOSBIA>2.0.CO;2.
- Marsland, S. J., H. Haak, J. H. Jungclaus, M. Latif, and F. Röske, 2003: The Max Planck Institute global ocean/sea ice model with orthogonal curvilinear coordinates. *Ocean Model.*, **5**, 91–127, doi:DOI: 10.1016/S1463-5003(02)00015-X.
- Mass, C. F. and D. A. Portman, 1989: Major volcanic eruptions and climate: A critical evaluation. *J. Clim.*, **2**, 566–593, doi:10.1175/1520-0442(1989)002<0566:MVEACA>2.0.CO;2.
- McKay, M. D., R. J. Beckman, and W. J. Conover, 1979: A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, **21**, 239–245.
- Medeiros, B. and B. Stevens, 2010: Revealing differences in GCM representations of low clouds. *Clim. Dyn.*, to appear, doi:10.1007/s00382-009-0694-5.
- Meehl, G. A., C. Covey, K. E. Taylor, T. Delworth, R. J. Stouffer, M. Latif, B. McAvaney, and J. F. B. Mitchell, 2007: The WCRP CMIP3 multimodel dataset: A new era in climate change research. *Bull. Am. Meteorol. Soc.*, **88**, 1383–1394, doi:10.1175/BAMS-88-9-1383.
- Murphy, J. M., D. M. H. Sexton, D. N. Barnett, G. S. Jones, M. J. Webb, M. Collins, and D. A. Stainforth, 2004: Quantification of modelling uncertainties in a large ensemble of climate change simulations. *Nature*, **430**, 768–772, doi:10.1038/nature02771.

## BIBLIOGRAPHY

- Phillips, T. J., G. L. Potter, D. L. Williamson, R. T. Cederwall, J. S. Boyle, M. Fiorino, J. J. Hnilo, J. G. Olson, S. Xie, and J. J. Yio, 2004: Evaluating parameterizations in general circulation models: Climate simulation meets weather prediction. *Bull. Amer. Meteor. Soc.*, **85**, 1903–1915.
- Piani, C., D. J. Frame, D. A. Stainforth, and M. R. Allen, 2005: Constraints on climate change from a multi-thousand member ensemble of simulations. *Geophys. Res. Lett.*, **32**, L23825, doi:10.1029/2005gl024452.
- Pincus, R., C. P. Batstone, R. J. P. Hofmann, K. E. Taylor, and P. J. Glecker, 2008: Evaluating the present-day simulation of clouds, precipitation, and radiation in climate models. *J. Geophys. Res.*, **113**, D14209, doi:10.1029/2007jd009334.
- Pinto, J. P., R. P. Turco, and O. B. Toon, 1989: Self-limiting physical and chemical effects in volcanic eruption clouds. *J. Geophys. Res.*, **94**, 11165–11174, doi:10.1029/JD094iD08p11165.
- Raddatz, T., C. Reick, W. Knorr, J. Kattge, E. Roeckner, R. Schnur, K.-G. Schnitzler, P. Wetzell, and J. Jungclaus, 2007: Will the tropical land biosphere dominate the climate-carbon cycle feedback during the twenty-first century? *Clim. Dyn.*, **29**, 565–574, doi:10.1007/s00382-007-0247-8.
- Randall, D. A., R. A. Wood, S. Bony, R. Colman, T. Fichefet, J. Fyfe, V. Kattsov, A. Pitman, H. Shukla, J. Srinivasan, R. J. Stouffer, A. Sumi, and K. E. Taylor, 2007: Climate models and their evaluation. *Climate Change 2007: The physical science basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*.
- Reichler, T. and J. Kim, 2008: How well do coupled models simulate today's climate? *Bull. Am. Meteorol. Soc.*, **89**, 303–311, doi:10.1175/BAMS-89-3-303.
- Rodwell, M. J. and T. N. Palmer, 2007: Using numerical weather prediction to assess climate models. *Q. J. Roy. Meteor. Soc.*, **133**, 129–146, doi:10.1002/qj.23.
- Roeckner, E., L. Bonaventura, R. Brokopf, M. Esch, M. Giorgetta, S. Hagemann, L. Kornbluh, U. Schlese, U. Schulzweida, I. Kirchner, E. Manzini, A. Rhodin, and A. Tompkins, 2003: The atmospheric general circulation model ECHAM5. Part I: Model description. Technical Report 349, Max-Planck-Institut für Meteorologie, Hamburg, Germany.
- Sanderson, B., C. Piani, W. Ingram, D. Stone, and M. Allen, 2008: Towards constraining climate sensitivity by linear analysis of feedback patterns in

## BIBLIOGRAPHY

- thousands of perturbed-physics GCM simulations. *Clim. Dyn.*, **30**, 175–190, doi:10.1007/s00382-007-0280-7.
- Sato, M., J. E. Hansen, M. P. McCormick, and J. B. Pollack, 1993: Stratospheric aerosol optical depths, 1850-1990. *J. Geophys. Res.*, **98**, 22987–22994, doi:10.1029/93JD02553.
- Schneider, E. K., B. P. Kirtman, and R. S. Lindzen, 1999: Tropospheric water vapor and climate sensitivity. *J. Atmos. Sci.*, **56**, 1649–1658, doi:10.1175/1520-0469(1999)056<1649:TWVACS>2.0.CO;2.
- Simmons, A., S. Uppala, D. Dee, and S. Kobayashi, 2007: ERA-Interim: New ECMWF reanalysis products from 1989 onwards. *ECMWF Newsletter*, **110**, 29–35.
- Soden, B. J., A. J. Broccoli, and R. S. Hemler, 2004: On the use of cloud forcing to estimate cloud feedback. *J. Clim.*, **17**, 3661–3665, doi:10.1175/1520-0442(2004)017<3661:OTUOCF>2.0.CO;2.
- Soden, B. J. and I. M. Held, 2006: An assessment of climate feedbacks in coupled ocean-atmosphere models. *J. Clim.*, **19**, 3354–3360, doi:10.1175/JCLI3799.1.
- Soden, B. J., I. M. Held, R. Colman, K. M. Shell, J. T. Kiehl, and C. A. Shields, 2008: Quantifying climate feedbacks using radiative kernels. *J. Clim.*, **21**, 3504–3520, doi:10.1175/2007JCLI2110.1.
- Soden, B. J., R. T. Wetherald, G. L. Stenchikov, and A. Robock, 2002: Global cooling after the eruption of mount Pinatubo: A test of climate feedback by water vapor. *Science*, **296**, 727–730, doi:10.1126/science.296.5568.727.
- Sohn, B.-J., J. Schmetz, R. Stuhlmann, and J.-Y. Lee, 2006: Dry bias in satellite-derived clear-sky water vapor and its contribution to longwave cloud radiative forcing. *J. Clim.*, **19**, 5570–5580, doi:10.1175/JCLI3948.1.
- Solomon, S., D. Qin, M. Manning, Z. Chen, M. Marquis, K. Averyt, M. Tignor, and H. Miller, 2007: Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. *Cambridge Univ. Press*.
- Stainforth, D. A., T. Aina, C. Christensen, M. Collins, N. Faull, D. J. Frame, J. A. Kettleborough, S. Knight, A. Martin, J. M. Murphy, C. Piani, D. Sexton, L. A. Smith, R. A. Spicer, A. J. Thorpe, and M. R. Allen, 2005: Uncertainty in predictions of the climate response to rising levels of greenhouse gases. *Nature*, **433**, 403–406, doi:10.1038/nature03301.

## BIBLIOGRAPHY

- Stephens, G. L., S.-C. Tsay, P. W. Stackhouse, and P. J. Flatau, 1990: The relevance of the microphysical and radiative properties of cirrus clouds to climate and climatic feedback. *J. Atmos. Sci.*, **47**, 1742–1754, doi:10.1175/1520-0469(1990)047<1742:TROTMA>2.0.CO;2.
- Stone, P. H. and J. H. Carlson, 1979: Atmospheric lapse rate regimes and their parameterization. *J. Atmos. Sci.*, **36**, 415–423, doi:10.1175/1520-0469(1979)036<0415:ALRRAT>2.0.CO;2.
- Stuber, N., R. Sausen, and M. Ponater, 2001: Stratosphere adjusted radiative forcing calculations in a comprehensive climate model. *Theor. Appl. Climatol.*, **68**, 125–135, doi:10.1007/s007040170041.
- Tebaldi, C. and R. Knutti, 2007: The use of the multi-model ensemble in probabilistic climate projections. *Phil. Trans. R. Soc. A*, **365**, 2053–2075, doi:10.1098/rsta.2007.2076.
- Tiedtke, M., 1989: A comprehensive mass flux scheme for cumulus parameterization in large-scale models. *Mon. Wea. Rev.*, **117**, 1779–1800, doi:10.1175/1520-0493(1989)117<1779:ACMFSF>2.0.CO;2.
- Timmreck, C., S. J. Lorenz, T. J. Crowley, S. Kinne, T. J. Raddatz, M. A. Thomas, and J. H. Jungclaus, 2009: Limited temperature response to the very large AD 1258 volcanic eruption. *Geophys. Res. Lett.*, **36**, doi:10.1029/2009GL040083.
- Weigel, A. P., R. Knutti, M. A. Liniger, and C. Appenzeller, 2010: Risks of model weighting in multimodel climate projections. *J. Clim.*, **23**, 4175–4191, doi:10.1175/2010JCLI3594.1.
- Wetherald, R. T. and S. Manabe, 1988: Cloud feedback processes in a general circulation model. *J. Atmos. Sci.*, **45**, 1397–1416, doi:10.1175/1520-0469(1988)045<1397:CFPIAG>2.0.CO;2.
- Wetzel, P., E. Maier-Reimer, M. Botzet, J. Jungclaus, N. Keenlyside, and M. Latif, 2006: Effects of ocean biology on the penetrative radiation in a coupled climate model. *J. Clim.*, **19**, 3973–3987, doi:10.1175/JCLI3828.1.
- Wielicki, B. A., B. R. Barkstrom, E. F. Harrison, R. B. Lee, G. Louis Smith, and J. E. Cooper, 1996: Clouds and the Earth’s Radiant Energy System (CERES): An Earth Observing System Experiment. *Bull. Am. Meteorol. Soc.*, **77**, 853–868, doi:10.1175/1520-0477(1996)077<0853:CATERE>2.0.CO;2.

## BIBLIOGRAPHY

- Wigley, T. M. L., C. M. Ammann, B. D. Santer, and S. C. B. Raper, 2005: Effect of climate sensitivity on the response to volcanic forcing. *J. Geophys. Res.*, **110**, 8pp, doi:10.1029/2004JD005557.
- Wigley, T. M. L., P. D. Jones, and S. C. B. Raper, 1997: The observed global warming record: What does it tell us? *Proc. Natl. Acad. Sci.*, **94**, 8314–8320.
- Yokohata, T., S. Emori, T. Nozawa, Y. Tsushima, T. Ogura, and M. Kimoto, 2005: Climate response to volcanic forcing: Validation of climate sensitivity of a coupled atmosphere-ocean general circulation model. *Geophys. Res. Lett.*, **32**, doi:10.1029/2005GL023542.
- Zelinka, M. D. and D. L. Hartmann, 2010: Why is longwave cloud feedback positive? *J. Geophys. Res.*, **115**, 16pp, doi:10.1029/2010JD013817.
- Zhang, M. H., J. J. Hack, J. T. Kiehl, and R. D. Cess, 1994: Diagnostic study of climate feedback processes in atmospheric general circulation models. *J. Geophys. Res.*, **99**, 5525–5537, doi:10.1029/93JD03523.

# Acknowledgements

I would like to thank Johannes Quaas for being a great advisor throughout my PhD time. The mixture of leaving me all the freedom in the world, but being involved and supportive at all times made his guidance so good. Also Bjorn Stevens and Marco Giorgetta as part of my advisory panel had large contributions to the success of my thesis. My panel meetings were very helpful and opposing to what one might expect of a panel meeting they were also fun. My supervisor team had a good balance of pushing me forward and pulling me back. More important than the scheduled meetings were the many spontaneous discussions while having a coffee, lunch, dinner, run, or a beer. I hope some of these occasions will give room for more interactions in the future.

Being part of the International Max Planck Research School for Earth System modeling was a great pleasure. The financial support made it possible for me to meet interesting colleagues all over the globe which was of great value for my work and will be of even greater value for years to come. More importantly it was great to be part of this group of people in the school, which makes being a PhD student so much more fun. Especially the dedication of Anje Weitz and Cornelia Kampmann make it such an awesome school.

The financial support of the IMPRS also allowed me to spend three months in New York to work with Robert Pincus. This was an important contribution to my PhD. Working with Robert was very instructive, great fun and opened new perspectives on being a scientist. I also want to thank the people at GISS and especially Anthony Del Genio for hosting me.

My office mates Benjamin Möbis, Torsten Weber and Gabriela Santos being closest but all other members of the atmosphere department in general and the cloud-climate feedbacks group especially made coming to work in the mornings more enjoyable.

Thank you:

Jan Härter, Lorenzo Tomassini and Jin-Song von Storch for being uncertain.

Monika Esch, Sebastian Rast, Ulli Schlese, Michael Botzet and Uwe Schulzweida for helping me to control the beast.

Erich Roeckner and Thorsten Mauritsen for sharing the experience of playing with the beast.

Jeffrey Anderson, Patrick Hoffmann, Nancy Collins and Luis Kornbluh for helping me to make two beasts dance.

Sebastian Schirber for keeping the beasts dance.

Stephan Bakan, Christian Klepp and Axel Anderson for their infectious enthusiasm.

My friends for the so important distraction.

Especially I want to thank Louise for inspiring and motivating me. Particularly during the final stages of my PhD her support essential.

Finally I want to thank my family for being understanding, patient, loving and supportive throughout my life.

## ACKNOWLEDGEMENTS

