Max-Planck-Institut
für Meteorologie

MAX-PLANCK-GESELLSCHAFT

International Max Planck Research School
on Earth System Modelling

Internal variability in a changing climate

A large ensemble perspective on tropical Atlantic rainfall

Sebastian Milinski

Hamburg 2019

Berichte zur Erdsystemforschung
Reports on Earth System Science

219
2019

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./Phone: +49 (0)40 4 11 73 - 0
Fax: +49 (0)40 4 11 73 - 298

name.surname@mpimet.mpg.de
www.mpimet.mpg.de

Layout

Bettina Diallo and Norbert P. Noreiks
Communication

Copyright

Photos below: ©MPI-M
Photos on the back from left to right:
Christian Klepp, Jochem Marotzke,
Christian Klepp, Clotilde Dubois,
Christian Klepp, Katsumasa Tanaka

Internal variability in a changing climate

A large ensemble perspective on tropical Atlantic rainfall

Sebastian Milinski

Hamburg 2019

Sebastian Milinski
from Bern, Switzerland

Max-Planck-Institut für Meteorologie
The International Max Planck Research School on Earth System Modelling
(IMPRS-ESM)
Bundesstrasse 53
20146 Hamburg

Universität Hamburg
Geowissenschaften
Meteorologisches Institut
Bundesstr. 55
20146 Hamburg

Tag der Disputation: 25. Januar 2019

Folgende Gutachter empfehlen die Annahme der Dissertation:
Dr. Johann H. Jungclaus
Prof. Dr. Carsten Eden

Vorsitzender des Promotionsausschusses:
Prof. Dr. Dirk Gajewski

Dekan der MIN-Fakultät:
Prof. Dr. Heinrich Graener

Cover graphic by Michael Böttinger (Deutsches Klimarechenzentrum) and Jochem Marotzke (Max Planck Institute for Meteorology): mean temperature anomaly 2080–2099 under RCP2.6, RCP4.5, and RCP8.5 relative to preindustrial conditions. Realisations from the MPI Grand Ensemble and CMIP5 models. Distance of isolines is 1 kelvin, the zero line is between blue and yellow colours.

Typeset using the classicthesis template developed by André Miede, available at:
https://bitbucket.org/amiede/classicthesis/

Berichte zur Erdsystemforschung / Max-Planck-Institut für Meteorologie 219
Reports on Earth System Science / Max Planck Institute for Meteorology 2019

ISSN 1614-1199

ABSTRACT

In this dissertation I investigate the temporal development of internal variability under global warming. Understanding internal variability is essential to understand the past and possible future trajectories of our climate, yet it is often assumed to be a property of the climate system that does not change under global warming. I use a novel large ensemble, the Max Planck Institute Grand Ensemble, to introduce a new perspective on internal variability.

Internal variability can be described as the seemingly random fluctuations of the climate system over time. Due to nonlinearity in the climate system, small perturbations may grow to large anomalies over time that are associated with anomalous or even extreme events. In my first chapter, I quantify internal variability and investigate whether it changes under global warming. The change in the external forcing is the same for all of these realisations, the initial conditions are different for each realisation. Thus, each realisation follows its own, unique trajectory. For each time step, the distribution of all realisations provides an estimate of the possible states of the climate system.

I develop an analysis framework based on a large ensemble to detect, quantify and attribute changes in internal variability in a transient climate. Rather than analysing variability over time, I use the ensemble dimension of a large ensemble to quantify internal variability. This approach allows a clean separation of the forced signal from internal variability and ensures stationarity of the statistics even when the forcing is changing with time. My non-parametric approach provides an objective quantification of changes in internal variability and their robustness.

In my second chapter I apply this analysis framework to investigate rainfall in the tropical Atlantic region in the past and its possible future trajectories. I can show that simulated internal variability in the Sahel encompasses all observed values for the 20th century. The model suggests an externally forced increase in rainfall towards the end of the 20th century. However, due to large internal variability, it is not possible to detect this forced change in a single realisation. In future projections, I find an increase in the mean rainfall over the Sahel, accompanied by an increase in the variability. This implies that the average rainfall will increase, but individual years may show deviations from this mean value that are larger than under present-day conditions.

In the tropical Atlantic region, most state-of-the-art coupled climate models show large biases in the simulated sea surface temperature and rainfall when compared to observations. These model errors

challenge the reliability of future projections. In my third chapter, I investigate the tropical Atlantic sea surface temperature bias in the Max Planck Institute Earth System Model. By using different high-resolution configurations of a climate model and dedicated sensitivity experiments, I can show that the coastal warm bias in the southeastern tropical Atlantic is caused by too low winds along the coast that result in too little upwelling of cold water masses. This error in the wind speed is mainly caused by the misrepresentation of the coastal orography in the low-resolution atmospheric model. By increasing the horizontal resolution in the atmosphere, large parts of the coastal warm bias in sea surface temperature can be eliminated. Contrary to previous hypotheses, eliminating the coastal warm bias does not affect the large scale biases in the tropical Atlantic.

ZUSAMMENFASSUNG

In dieser Dissertation untersuche ich Veränderungen der internen Variabilität unter globaler Erwärmung. Ein umfassendes Verständnis der internen Variabilität ist unerlässlich um die vergangene Entwicklung und mögliche zukünftige Trajektorien des Klimas zu verstehen. Dennoch wird die interne Variabilität häufig als eine Eigenschaft des Klimasystems gesehen, die sich unter globaler Erwärmung nicht verändert. In dieser Studie nutze ich ein großes Ensemble von Klimamodellsimulationen um eine neue Perspektive auf das Verständnis interner Variabilität zu eröffnen.

Interne Variabilität kann als die scheinbar zufälligen Schwankungen des Klimasystems im Laufe der Zeit beschrieben werden. Aufgrund der Nichtlinearitäten im Klimasystem können kleine Störungen mit der Zeit anwachsen und ungewöhnliche oder gar Extremereignisse verursachen. In meinem ersten Kapitel quantifiziere ich die interne Variabilität und untersuche, ob diese sich unter globaler Erwärmung ändert. Zu diesem Zweck verwende ich ein großes Ensemble von Klimamodellsimulationen. Alle Simulationen, auch als Realisierungen bezeichnet, haben den gleichen zeitlichen Verlauf der Randbedinungen, also der extern vorgeschriebenen Parameter wie solarer Einstrahlung oder Konzentrationen von Treibhausgasen in der Atmosphäre. Die unterschiedlichen Realisierungen unterscheiden sich jedoch in ihren Anfangsbedingungen. So folgt jede Realisierung ihrer eigenen, einzigartigen Trajektorie. Für jeden Zeitschritt liefert die Verteilung aller Realisierungen eine Abschätzung der möglichen Zustände und assoziierten Wahrscheinlichkeiten des Klimasystems.

Ich entwickle eine Analysemethode auf Basis eines großen Ensembles, um Veränderungen der internen Variabilität zu detektieren, zu quantifizieren und mit einer Änderung in den Randbedingungen zu assoziieren. Anstatt die Variabilität anhand einer Stichprobe basierend auf mehreren Jahren zu analysieren, verwende ich die Ensemble-Dimension eines großen Ensembles, um die interne Variabilität zu quantifizieren. Dieser Ansatz ermöglicht eine saubere Trennung des extern angetriebenen Signals von der internen Variabilität und gewährleistet Stationarität. Mein nichtparametrischer Ansatz bietet eine objektive Quantifizierung von Veränderungen der internen Variabilität und ihrer Robustheit.

In meinem zweiten Kapitel wende ich diese Analysemethode an, um die Niederschläge im tropischen Atlantik in der Vergangenheit und ihre möglichen zukünftigen Trajektorien zu untersuchen. Ich kann zeigen, dass die simulierte interne Variabilität in der Sahelzone alle beobachteten Werte für das 20. Jahrhundert umfasst. Die Simu-

lationen zeigen zudem einen Anstieg der Niederschläge gegen Ende des 20. Jahrhunderts, der durch die Änderung in den Randbedingungen verursacht wird. Aufgrund der großen internen Variabilität ist es jedoch nicht möglich, diese extern angetriebene Veränderung in einer einzelnen Realisierung zu identifizieren. Die Zukunftsszenarien zeigen einen Anstieg des mittleren Niederschlages über der Sahelzone, verbunden mit einer Zunahme der Variabilität. Dies bedeutet, dass die durchschnittliche Niederschlagsmenge zunehmen wird, aber einzelne Jahre Abweichungen von diesem Mittelwert aufweisen können, die größer sind als unter heutigen Bedingungen.

Im tropischen Atlantik zeigen nahezu alle aktuellen gekoppelten Klimamodelle große Fehler in der simulierten Meeresoberflächentemperatur und im simulierten Niederschlag im Vergleich zu Beobachtungen. Diese Modellfehler stellen die Zuverlässigkeit und den Wahrheitsgehalt von Zukunftsprojektionen in Frage. In meinem dritten Kapitel untersuche ich die Ursachen der Modellfehler in der Oberflächentemperatur des tropischen Atlantiks in einem gekoppelten Klimamodell. Durch die Verwendung verschiedener hochauflösender Konfigurationen eines Klimamodells kann ich zeigen, dass der küstennahe Fehler im südöstlichen tropischen Atlantik durch zu geringe Windgeschwindigkeiten entlang der Küste verursacht wird, die zu einem zu geringen Auftrieb von kalten Wassermassen führen. Dieser Fehler in der Windgeschwindigkeit wird hauptsächlich durch die Fehldarstellung der küstennhanen Orographie bei niedriger Auflösung im Atmosphärenmodell verursacht. Durch die Erhöhung der horizontalen Auflösung im Atmosphärenmodell können große Teile des küstennahen Modellfehlers in der Meeresoberflächentemperatur eliminiert werden. Im Gegensatz zu früheren Hypothesen hat der küstennahe Modellfehler jedoch keinen Einfluss auf die Modellfehler im restlichen tropischen Atlantik.

TEILVERÖFFENTLICHUNGEN DIESER DISSERTATION

Pre-published work related to this dissertation

Milinski, S., J. Bader, H. Haak, A. C. Siongco, and J. H Jungclaus (2016), "High atmospheric horizontal resolution eliminates the wind-driven coastal warm bias in the southeastern tropical Atlantic". *Geophys. Res. Lett.*, **43**, 10,455–10,462, doi:10.1002/2016GL070530

ACKNOWLEDGEMENTS

Writing this thesis was only possible through the invaluable support that I received from everyone who accompanied me on my journey.

Most of all, I want to thank my family, Manfred, Petra, and Linus, for their love, support, inspiration and encouragement.

I want to thank my supervisors, Johann Jungclaus, Jürgen Bader and Jochem Marotzke for guiding me through this scientific adventure, for giving me a lot of freedom to explore, and shaping me as a scientist. I am also grateful to Carsten Eden for hist support as my panel chair, and for always asking the right questions that encouraged me to look at my project from a slightly different perspective.

The inspiring and creative atmosphere at the Max Planck Institute for Meteorology has made the past four years a wonderful experience and I am grateful for the support of the institute. I want to thank Dian Putrasahan and Yohei Takano for their valuable feedback on an earlier version of this thesis, for quick questions that turned into long conversations about large ensembles, observations, everything in between, and—not least of all—culinary adventures. I want to thank Nicola Maher, Dirk Olonscheck, and the whole Grand Ensemble working group for inspiring discussions.

I want to thank Cheska Siongco and Helmuth Haak for the collaboration that lead to my first publication. I am grateful to Helmuth Haak for the amazing support with any technical question about modelling.

Thank you to Antje Weitz, Cornelia Kampmann, Michaela Born, and Wiebke Böhm for taking a lot of weight off my shoulders and making everything go just a little bit smoother.

Finally, I want to thank Till Baumann and Sandro Dahlke for sharing part of my scientific journey, for keeping me grounded and, despite the geographical distance, never being more than a skype call away.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACRONYMS

CMIP5	Coupled Model Intercomparison Project Phase 5
GMST	global mean surface temperature
IPCC	Intergovernmental Panel on Climate Change
ITCZ	Intertropical Convergence Zone
MPI-GE	Max Planck Institute Grand Ensemble
SST	sea surface temperature

INTRODUCTION

Understanding the past evolution and the possible future trajectories of the climate system is at the core of climate science. We use observations of the recent past and proxy-based reconstructions that go back in time for several millennia and combine these with climate model simulations to understand the mechanisms that drive the climate system in the past, present, and future. In this context, there is a particular interest in understanding the influence of external forcing, most prominently the anthropogenic contribution, on the evolution of the climate system. However, any realisation of the transient climate system, such as the observational record or a climate model simulation, is a combination of the forced signal and chaotic internal variability that arises from the nonlinear nature of the climate system. Therefore, we need to quantify the forced signal and internal variability separately, which is not possible with a single realisation (Frankcombe et al. 2018). In this study, I use a novel large ensemble of climate model simulations to overcome this problem.

Internal variability is often treated as an irreducible uncertainty that conceals the forced signal. However, internal variability itself needs to be characterised to fully understand the observed realisation, which is the climate that we experience. Internal variability determines how large and how frequent deviations from the forced signal are on different time scales. On long time scales, internal variability can modulate decadal trends of the global mean surface temperature (Marotzke and Forster 2015) and can lead to misunderstandings about our knowledge of the forced signal, which has happened for example in the global warming hiatus discussion (Hedemann et al. 2017). On interannual time scales, modes of variability, such as the El Niño-Southern Oscillation, cause anomalous rainfall patterns and surface temperature anomalies over large regions that have important implications for the population. On short time scales, extreme events are the most discernible manifestation of internal variability and can have large impacts despite being of relatively short duration and regionally confined. All of these manifestations of internal variability are not taken into account when investigating the forced signal. Therefore quantifying internal variability and differentiating it from the forced signal is crucial both for understanding the present and the future.

Characterising internal variability is not straightforward and can be further impeded by a number of challenges that depend on the context that is investigated. A widely applied assumption is that internal variability does not change in response to the forcing (e.g. Hawkins and Sutton 2010, 2012; Lehner et al. 2017; Thompson et

al. 2015). In this context, a long control simulation with constant external forcing can be used to quantify internal variability, which is then used to study internal variability in a simulation with changing external forcing. This approach does not allow for a clean separation of internal variability and the forced signal. To estimate the forced signal, we have to make assumptions, such as defining the forced signal as the low-frequency part of the time series (Hawkins and Sutton 2012). This makes it difficult to determine whether a decadal trend in the simulation with changing forcing is caused by the forcing change or internal variability. In addition, the assumption that internal variability does not change in response to the forcing might not hold. Even relatively small changes in the internal variability can have striking consequences for the occurrence of extreme events and should therefore not be excluded a priori. This confronts us with a challenge that is more difficult to overcome than the separation of internal variability and the forced signal.

A common estimator for internal variability is the variance of a time series. The time series needs to be sufficiently long to allow for an adequate sampling of the internal variability. Sampling over time is only possible when the statistics are stationary. Common practice is to detrend a time series, which requires knowledge of the forced signal, and sample from the residual. This approach does not take a change in variability into account.

In this study, I use a large number of realisations to overcome these challenges. A comprehensive climate model has been ran 100 times. Each of these realisations starts with different initial conditions sampled from different equilibrated ocean-atmosphere states. The evolution of the external forcing and the model configuration is identical, yet the trajectories for all realisations are different because variations due to internal variability unfold differently. At the same time, the changes in the external forcing cause a similar response in all realisations. By averaging over the ensemble dimension, the random variations due to internal variability cancel out and the mean response to the external forcing change is left (Frankcombe et al. 2018). In the context of a single model ensemble, this is the forced signal. With this approach, we can effectively separate internal variability and the forced signal. For quantifying internal variability in a transient climate, we can then proceed to use the ensemble dimension to overcome the challenge of potential changes of the internal variability over time. When using the ensemble dimension to quantify variability, variance across different realisation at the same time with identical boundary conditions is used, rather than variance over time. This approach ensures stationarity of the statistics, rather than having to assume that the samples distribution is stationary. I show that variance across the ensemble dimension provides a well defined estimate of the internal variability for every forcing state in a transient climate.

In the first chapter of the thesis, I develop a new analysis framework to detect and attribute changes of the internal variability of the climate system. While several large ensembles of comprehensive climate models are available, they are mostly used to investigate the forced signal rather than internal variability itself. At the same time, numerous methods for statistical analysis exist that can be used to investigate and compare the variability in samples from different populations. I combine different approaches from previous work with a large ensemble of a comprehensive climate model to develop an analysis framework for investigating internal variability and forced changes to it in response to a change in the external forcing.

I use an idealised warming scenario to exemplify the challenges we face when using a single realisation and demonstrate how they can be solved by using a large ensemble. I analyse vertically integrated global mean specific humidity, which is also described as precipitable water. Precipitable water is a quantity where I expect increased variability under warming. From the Clausius-Clapeyron relationship the water holding capacity of the air is expected to increase by 7% per °C under global warming (Held and Soden 2006). When assuming that relative humidity does not change in response to global warming, the variability of precipitable water is mainly controlled by temperature variability. Even if the temperature variability does not change in response to warming, the variability in precipitable water will increase. This is because similar variations in temperature will cause larger variations in precipitable water in a warmer atmosphere. Therefore I use precipitable water in an idealised warming scenario as an example of a quantity with increasing variability to demonstrate how we can: (i) cleanly separate the forced signal and internal variability, (ii) adequately sample the phase space of internal variability, (iii) and detect forced changes in the internal variability in a transient climate.

In the second chapter, the analysis framework is then applied to analyse changes in rainfall variability in the tropical Atlantic region under global warming. The framework allows me to perform a robust attribution of projected changes, both in the mean state and variability, to changes in the external forcing. I find that changes in rainfall variability are not spatially homogeneous over the tropical Atlantic. While there is an increase in variability over the Sahel and close to the centre of the oceanic Intertropical Convergence Zone (ITZC), there is a decrease of variability on the outer southern flank of the oceanic ITCZ.

For this study, I have to rely on a climate model because internal variability and the forced signal in a transient climate can only be studied with a large number of realisations. However, the simulated climate in the tropical Atlantic region suffers from large biases that might impair the models reliability for future projections. Therefore I investigate the origin of the mean state biases that are common to most

state-of-the-art coupled climate models. I show that the mean state coastal SST bias in the southeastern tropical Atlantic can be reduced by increasing the atmospheric horizontal resolution (Milinski et al. 2016). I then examine the representation of variability in the model configuration with a reduced mean state bias to quantify the potential benefits of increasing the atmospheric horizontal resolution.

The aim of this study is to broaden our understanding of internal variability within the limits of the currently available data and models. Using a large ensemble of a coupled climate model adds to our understanding of the single realisation available from observations. Increasing the horizontal resolution of a climate model can reduce some of the model biases, but at the same time, the increase in the computational cost does only allow for a small number of realisations. At the moment, a large ensemble is not feasible with a high-resolution climate model. I conclude by discussing the benefits of combining a coarser-resolution large ensemble with a single realisation of a high-resolution model for future modelling strategies.

THE MPI GRAND ENSEMBLE

The Max Planck Institute Grand Ensemble (MPI-GE) is a novel set of large ensemble simulations that was produced as a combined effort of all departments of the Max Planck Institute for Meteorology. It consists of five forcing scenarios with 100 realisations each. The 100 realisations only differ by their initial conditions, while the model configuration and the prescribed external forcing is the same. The idea of these simulations is to enable the investigation of internal variability in transient climate simulations to improve our understanding of future projections and reinterpret the observed record. The advantage of a single model large ensemble in comparison to a multi-model ensemble such as the Coupled Model Intercomparison Project 5 (CMIP5) is that the realisations only differ because of internal variability. This allows to cleanly separate the forced signal and internal variability for this model.

The first two chapters of this thesis are based on the MPI-GE. The purpose of this section is to provide a technical description of the model and experiments and outline some of the applications of the MPI-GE.

MODEL AND EXPERIMENTS

EXPERIMENT	TIME PERIOD	REALISATIONS	ABBREVIATION
Preindustrial Control	2000 years	1	PiControl
Historical	1850-2005	100	
RCP2.6	2006-2099	100	
RCP4.5	2006-2099	100	
RCP8.5	2006-2099	100	
1% yr^{-1} CO_2 increase	150 years	100	1% CO_2

Table 0.1: The MPI-GE experiments.

The MPI-GE is based on the Max Planck Institute Earth system model (MPI-ESM) as described by Giorgetta et al. (2013). The model version used for the MPI-GE is MPI-ESM1.1-LR (Bittner et al. 2016; Stevens 2015). To allow a large number of realisations, the low-resolution (LR) configuration of the model is used. It consists of the atmospheric component ECHAM (Stevens et al. 2013), the ocean component

MPIOM (Jungclaus et al. 2013; Marsland et al. 2003), the land component JSBACH (Reick et al. 2013) with the standard fire module, and the biogeochemistry component HAMOCC (Ilyina et al. 2013). The ocean component is run on the GR15L40 grid with a horizontal resolution of 1.5° and 40 layers, the spectral atmospheric component uses a T63L47 configuration, which translates to ~1.8° horizontal resolution and has 47 vertical levels. The output frequency is limited to monthly mean values in order to limit disk space usage to a manageable amount.

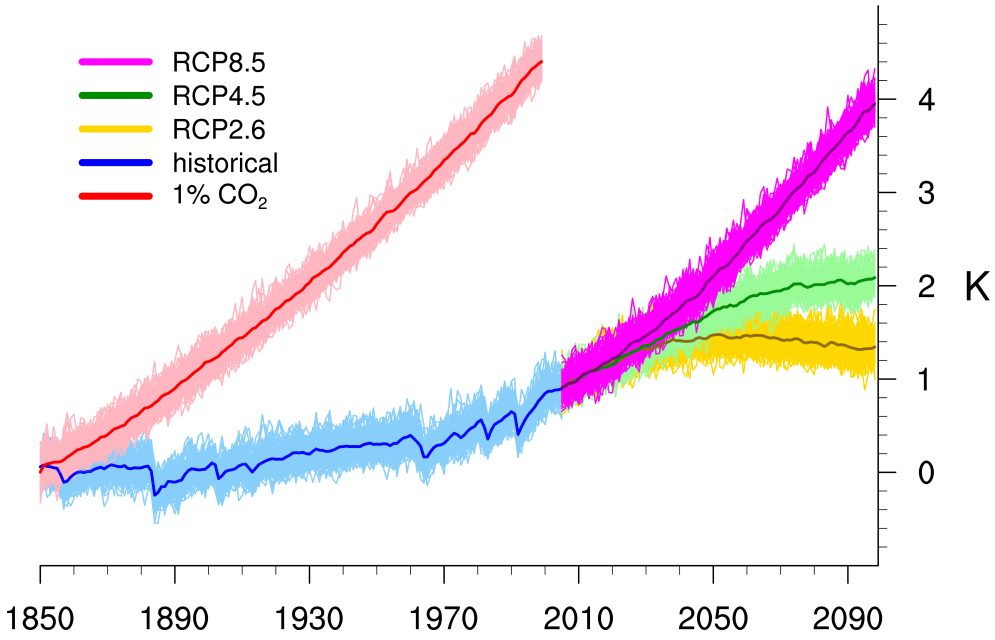

Figure 0.1: **GMST change in the MPI-GE experiments.** Annual mean global mean surface temperature change for all experiments, relative to the 1850–1900 mean GMST in historical experiments. Thin lines show individual realisations, thick lines show the ensemble mean. For illustration purposes, the 1% CO_2 experiment is plotted against years 1850–1999. The boundary conditions for 1% CO_2 are idealised and not associated with years in the historical record.

The experiments in MPI-GE follow the CMIP5 protocol (Taylor et al. 2012). A 2000-year preindustrial control simulations and five experiments with a changing forcing have been run as listed in table 0.1. The historical and 1% CO_2 experiments are initialised by sampling ocean and atmosphere initial conditions in different years of the 2000-year preindustrial control simulation. This approach ensures that the ensemble covers a wide range of possible states of the system and allows analysis of the experiments from the first year onward. The alternative approach of ensemble generation is to add small random perturbations to otherwise identical initial conditions. With this initialisation technique, all realisation are very similar in the beginning and slowly disperse, therefore the ensemble does not cover a wide range of possible states in the beginning and the first few

years cannot be used for analysis. The rate at which the realisations disperse can range from a few weeks in the atmosphere to several decades in the deep ocean. For the MPI-GE, the historical and 1% CO_2 experiments cover a wide range of possible states already in the first year. The three scenarios starting in year 2006 are initialised using the end of the 100 historical realisations. Each of the three scenarios is using the same set of 100 different initial conditions.

Figure 0.1 shows the evolution of global mean surface temperature (GMST) for all realisations. The ensemble mean shows the forced response of the model to the change in external forcing by efficiently removing the influence of internal variability (Frankcombe et al. 2018). The large spread of individual realisations indicates that internal variability is a relevant contribution to the uncertainty in future projections.

APPLICATIONS OF THE MPI-GE

Several studies have used the MPI-GE to disentangle internal variability and the forced response for various quantities. Hedemann et al. (2017) showed that the observed global warming hiatus can be explained by internal variability. Suárez-Gutiérrez et al. (2017) used the MPI-GE to show that internal variability can reconcile the apparent discrepancies between simulated and observed warming in the upper tropical troposphere. In a recent study, we investigated the response of the El-Niño Southern Oscillation to global warming (Maher et al. 2018). We found that internal variability can explain observed changes and most of the differences between CMIP5 models for the historical period. Only under strong warming, a robust change in ENSO is simulated. These studies highlight the potential of this novel ensemble to rethink and advance our understanding of the climate system, both for the past and possible future trajectories. To make use of the full potential of a large ensemble and generate new knowledge, it is crucial to modify and rethink the commonly used analysis methods. In the first chapter I present a new analysis framework for detection and attribution of changes in internal variability in a transient climate simulation, which can be used to study forced changes of internal variability for a variety of variables.

1

A FRAMEWORK FOR ANALYSING FORCED CHANGES OF INTERNAL VARIABILITY IN A TRANSIENT CLIMATE SYSTEM

1.1 LARGE ENSEMBLES AS A NEW TOOL FOR CLIMATE RESEARCH

To capture the chaotic nature of the climate system and sample a large variety of possible states of the climate system, multiple model simulations with identical boundary conditions but different initial conditions are performed, referred to as ensemble simulations. In weather forecasting, using ensemble simulations has become common practice since several decades (Lewis 2005). In the 1990s, Griffies and Bryan (1997) used a coupled global climate model with a total of 40 realisations and 4 different experiments to sample different states of the North Atlantic and show that under certain conditions, predictability on time scales longer than a decade can be achieved. In 2005, Zelle et al. 2005 used a 62-member ensemble of a global climate model to investigate changes in the El Niño-Southern Oscillation (ENSO) under global warming. The idea of using a large ensemble of a comprehensive climate model was later taken up by several modelling groups who performed large ensemble simulations with state-of-the-art climate models (Deser et al. 2012b; Fyfe et al. 2017; Kay et al. 2015; Rodgers et al. 2015; Stevens 2015). These ensembles used ensemble sizes ranging from 30 to 100 members to study the forced response and internal variability of the climate system for the past century and possible future scenarios. The objective of most studies using these large ensembles is to quantify the forced response of the climate system and investigate when this forced signal emerges from the noise of internal variability (for example Lehner et al. (2017) and Rodgers et al. (2015)). In this context, internal variability is treated as time-independent noise that conceals the forced signal. In this chapter, I introduce a new analysis framework to quantify internal variability in a changing climate and detect forced changes of internal variability.

1.2 THE CHALLENGE OF QUANTIFYING INTERNAL VARIABILITY

We perceive internal variability of the climate system as seemingly random fluctuations on a wide range of time scales. It is well established that increasing CO_2 concentration in the atmosphere lead to a warming of the Earth. Nevertheless, single years or even a number of consecutive years can deviate from this expected forced warming signal because of internal variability. The surface warming hiatus, a

period during which the observed warming was much slower than the warming projected by climate models, is only one example where internal variability has been shown to have a noticeable effect on the climate system (Hedemann et al. 2017). On shorter time scales, internal variability determines how large and frequent anomalies on seasonal or even daily time scales can be. But even on decadal time scales, internal variability has been shown to have a substantial contribution (Deser et al. 2012a; Marotzke and Forster 2015). To understand anomalous events on short time scales or anomalous trends on longer time scales, we need to isolate internal variability and the forced signal and quantify both.

A time-independent quantification of internal variability is not sufficient in a transient climate. To understand possible future trajectories of the Earth system, it is necessary to investigate potential changes of the internal variability in response to a change in the forcing because an increase of the amplitude of internal variability might have large effects on the amplitude of anomalous or even extreme events.

In a transient climate, where the external forcing is changing with time, the traditional approach to quantify internal variability from a single time series can lead to wrong conclusions. In this chapter, I introduce a new analysis framework that allows for a well-defined quantification of internal variability under different background conditions. This approach does not only provide a quantification of internal variability and enables the detection of a change in variability, but also provides a non-parametric estimate of the uncertainty to objectively describe the robustness of a detected change.

To demonstrate the limitations of the traditional approach and develop the new analysis framework, I will focus on a single quantity for which a change in variability is expected under global warming. Here, I focus on global mean precipitable water (vertically integrated specific humidity). The global mean precipitable water is only controlled by thermodynamics. Assuming that relative humidity stays constant, the amount of precipitable water must increase in a warmer atmosphere, based on the Clausius-Clapeyron relation for saturation vapour pressure (Held and Soden 2006). When adding the assumption that temperature variability does not decrease in a warmer atmosphere, the variability of global mean precipitable water must also increase under global warming, because the same temperature anomaly would result in a larger anomaly in precipitable water in a warmer atmosphere.

To test these assumptions with different analysis methods, I use climate model simulations with an idealised forcing scenario: a 1%-CO_2 increase per year. In this scenario, the atmospheric CO_2 concentration starts at preindustrial conditions and increases by 1% per year, relative to the previous year. The simulation length is 150 years (table 0.1). We performed 100 realisations with the same external forcing but

different initial conditions. Equilibrated initial states for the ocean and atmosphere are sampled from the 2000-year PiControl run.

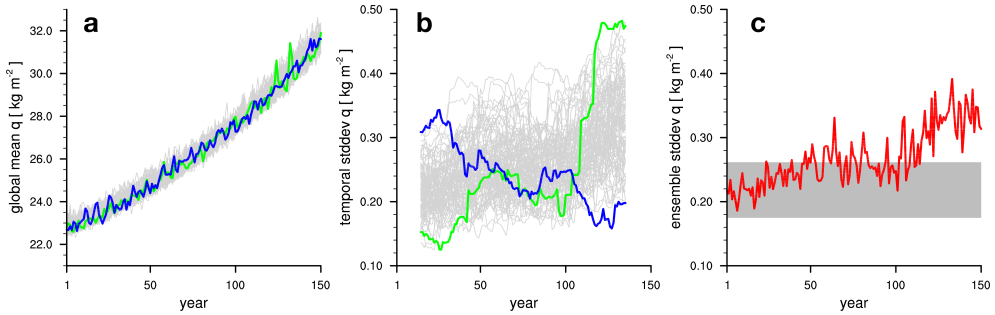

Figure 1.1: **In a transient climate, internal variability is only well defined in the ensemble dimension, not in the time dimension. (a)** global mean annual mean precipitable water in 100 realisations in a strong warming scenario, two realisations are highlighted in green and blue. **(b)** 30-year running standard deviation for each of the 100 realisations after detrending with the ensemble mean. **(c)** Ensemble standard deviation calculated for each year. The grey bar indicates the 5-95% range of standard deviations from 20 samples consisting of 100 randomly selected years from PiControl.

A common measure to quantify internal variability is to calculate the temporal standard deviation of a time series after removing the forced signal. In figure 1.1, the time series of global mean, annual mean precipitable water are shown for 100 realisation under the 1% CO_2 forcing. The forced signal in this case can be characterised as a monotonous increase in precipitable water and can be estimated using the ensemble mean. To investigate a change in internal variability, all members are detrended by subtracting the ensemble mean. Then, the temporal standard deviation in a 30-year window is calculated as a measure of internal variability. This window is then moved along the time axis in steps of one year to generate the time series of 30-year temporal standard deviation in figure 1.1b. The two highlighted realisations show very different results for the temporal evolution of internal variability: while the green realisation shows a clear increase in variability as expected from the initially formulated hypothesis, the blue realisation shows a decrease in variability.

This apparent disagreement occurs because 30 years are not long enough to sample internal variability sufficiently. A solution to this would be to increase the temporal sampling period. However, the temporal standard deviation for a long sampling window of 50 or even 100 years is even less representative for specific background conditions and might even sample from different states of internal variability if the internal variability is indeed changing over time.

The alternative approach I suggest is based on the variability across the ensemble dimension in a large initial condition ensemble of a

climate model. The ensemble domain provides a well-defined estimate of internal variability under clearly defined background conditions. The different samples are taken from different realisations with the same atmospheric CO_2 concentration and therefore sample from the same population. In the time series of the ensemble standard deviation in figure 1.1c, a clear increase in the internal variability in response to the forcing can be seen.

The examples so far have been based on standard deviation as a measure of internal variability. The standard deviation is only a good description of quantities that follow a normal distribution. For any non-normally distributed quantity, the standard deviation does not necessarily describe all changes in the distribution, in particular changes in the tails of the distribution might not be detected. Therefore, the focus of this chapter is to develop a non-parametric approach to quantify internal variability based on the ensemble domain of a large ensemble. This approach includes changes in all statistical moments of a distribution.

Statistical attribution of a change in the variability to the change in external forcing needs to be based on robust quantification of the uncertainty. The year-to-year fluctuations of the ensemble standard deviation in figure 1.1c indicate that the sample size of the ensemble is not sufficient to detect a small forced change in the variability. A robust change must be larger than the sampling uncertainty, which depends on the sample size and the variance of the underlying distribution. This sampling uncertainty can be estimated by generating alternative samples from a long control simulation, or from a much larger ensemble. In a long control run, the reference period for investigating changes in variability is fixed to preindustrial conditions. A much larger ensemble is not a feasible approach due to the large computational cost. I have therefore developed an approach to quantify the sampling uncertainty based on the available ensemble realisations.

1.3 QUANTIFYING INTERNAL VARIABILITY

Internal variability in an equilibrated climate

Internal variability in the climate system is commonly defined as deviations from the mean state that arise from the nonlinear nature of the climate system. The trajectory of internal variability is chaotic and cannot be predicted for all time scales because small perturbations of negligible magnitude may grow over time due to nonlinear processes in one component or in the interaction of different components in the climate system. There is limited predictability if the initial conditions are well known, but only as long as chaotic internal variability does not dominate the signal. While the exact trajectory of the system can only be predicted theoretically when there is perfect knowledge of the system, the initial conditions, and the boundary conditions, it is still possible to arrive at a statistical description of internal variability. For constant boundary conditions, i.e. when the external forcing does not change from year to year, the range of possible states that the climate system can assume over time can be described as a probability distribution. Most of the states will be close to the mean, while larger deviations might occur less frequently.

An idealised set-up for quantifying internal variability in this way is a climate model with preindustrial external forcing with the same seasonal cycle in insolation (incoming solar radiation) repeating every year, commonly referred to as a preindustrial control run. In this set-up, all greenhouse gas concentrations, such as CO_2, are held at a constant level. Thus, the forcing for every year is exactly the same as that of the previous year. However, the state of the Earth system is not the same in every year, which is due to internal variability. In this PiControl simulation (figure 1.2), there is substantial internal variability on interannual to multi-decadal time scales, despite the absence of any interannual variability in the forcing.

The basic measures for quantifying the statistics of a distribution are the mean and the standard deviation. For any variable in the climate system, the mean provides the expected value and the higher order statistical moments describe how large deviations from the expected value might be—i.e. internal variability. Because of the constant forcing, I do not expect a change of the population over time, thus statistics can be computed by using the entire time series as a sample. The histogram of the 2000-year time series in figure 1.3a shows the distribution of annual mean precipitable water. The discrete value of each sample is indicated above the histogram.

The histogram also visualises the shape of the distribution. In 1.3a, the right tail of the distribution seems to be heavier, which is to be expected for precipitable water because of the nonlinear dependence of vapour pressure on temperature. However, a histogram includes

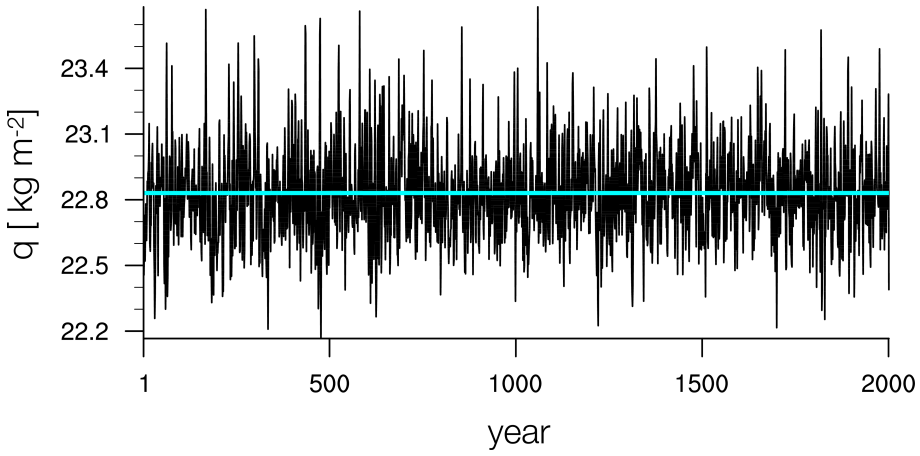

Figure 1.2: **Large internal variability on interannual to multi-decadal time scales for global mean precipitable water.** Annual mean time series of global mean precipitable water from a 2000-year simulation at constant 1850 forcing conditions. The cyan line indicates the mean.

a subjective choice of the number of bins and the bin size and is therefore not a purely objective description of the distribution. Different discrete values are aggregated in one bin. An alternative is a cumulative distribution, where the values are sorted and the fraction of samples smaller or equal to a value is plotted against the sample values, as done in figure 1.3b.

Comparing distributions

For most applications, we want to compare properties of the distribution to some reference. In the most simple case, this can be a comparison of the mean of two different distributions. However, using only the mean of a distribution excludes any information about the variability, and possible differences in the variability. In addition to the comparison of the mean, the standard deviation or higher-order statistical moments can be compared to capture the characteristics of the distributions and the variability. The drawbacks of using higher-order moments is that they are difficult to interpret, for example how a change in higher order moments translates into a change in extreme events. In addition, a robust estimate of higher order moments requires a larger sample size than for lower order moments, which might not always be available.

To highlight the limitations of using a histogram and introduce an alternative approach, I compare the distribution from the control run to a normal distribution that was fitted to the sample mean and sample standard deviation of the control run. The histogram in figure 1.4a can be used to compare these distributions. From the histogram,

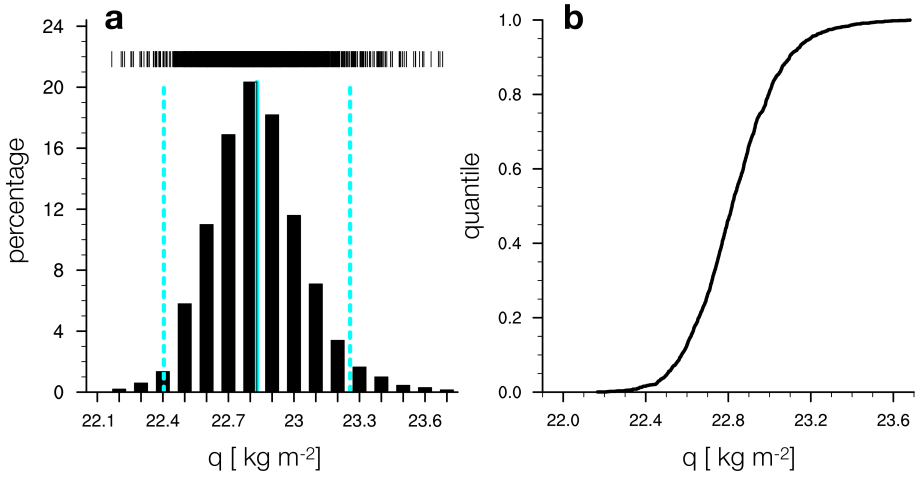

Figure 1.3: **Internal variability of precipitable water in a 2000-year preindustrial control simulation. (a)** The histogram shows the distribution of annual mean global mean precipitable water from a 2000-year simulation at constant 1850 forcing conditions. The blue solid line shows the mean, the region between the two dashed lines is indicating $\pm 2\sigma$ from the mean. There are 17 bins with a width of 0.1. The black lines above the histogram show the 2000 samples that were used to plot the histogram. **(b)** The same distribution plotted as a cumulative distribution function.

one can see that the distribution in the control run has a slightly heavier upper tail than a normal distribution. This agrees well with the expected response of precipitable water to a temperature anomaly: for warmer temperatures, the response should be larger than for colder temperatures. As mentioned before, the histogram depends on the choice of the bins and some features of the distribution might be represented in a different way when the configuration of bins is changed. Therefore a histogram is not well suited for an objective comparison of distributions, in particular when the sample size is small. Therefore, I use the cumulative distribution function to quantitatively compare the two distributions.

Some common methods for comparing distributions are based on cumulative distribution functions. The Kolmogorov–Smirnov test calculates the maximum distance between cumulative distributions (Storch and Zwiers 2001). While it seems to be useful to reduce the problem to a single number, this does not include any information about where the difference occurs and whether the rest of the distributions shows good agreement or not. The Cramér-von-Mises test (Stephens 1970) extends this approach and uses the sum of the squared distances of two cumulative distributions. While this approach includes information about how different the distributions are on average, it still discards information about where differences occur.

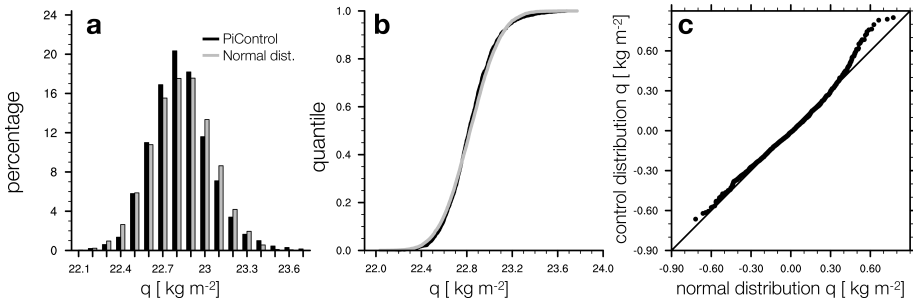

Figure 1.4: **Distribution of precipitable water in PiControl compared to a fitted normal distribution. (a)** The black histogram shows the control run as in figure 1.3, the grey histogram is a normal distribution with 10000 samples, fitted to the sample mean and standard deviation of the control run, **(b)** cumulative distributions for the control run and fitted normal distribution, **(c)** quantile-quantile plot of the control run and the normal distribution after removing the mean from both samples.

A graphical representation that illustrates all of the information that is neglected by the tests mentioned above is a quantile-quantile plot or qq plot(figure 1.4c). This plots the quantiles of a sample distribution, here the control run, against a reference distribution, which is in this case the fitted normal distribution. Each quantile level from the sample distribution is plotted against the same quantile level from the reference distribution. If the two distributions have a different number of samples, the expected value of the quantile that should be compared to the other distribution needs to be computed. Here, I group the samples from the larger distribution into chunks, so that the number of chunks equals the number of samples in the smaller distribution. For each chunk, the median is computed and used for comparing the quantile levels of the two distributions. An alternative approach for computing corresponding quantile levels is to interpolate between quantiles. For the applications in this thesis, the sample size of the reference distribution is in general much larger than that of the distribution that is compared to the reference. In this case, the median of each chunk provides a good estimate of the expected value for the corresponding quantile level.

If the two distributions are very similar, all points will lie close to the one-to-one line. In this example, most points are on the one-to-one line, except for the upper right quadrant. Here, the values from the sample distribution are consistently higher than what would be expected if the sample was normally distributed. Values in the upper right quadrant are large positive deviations from the mean and can be seen as positive extreme events. From the qq plot I can conclude that the sample is mostly normally distributed, except for a heavier upper tail.

The advantage of the qq plot in comparison to the histogram is that it does not depend on a subjective choice of bin size and it works with a smaller sample size because it is based on an integrated quantity. Estimating the statistical moments of a distribution is a feasible approach for lower-order moments, such as the sample mean or sample variance, but cannot intuitively be interpreted for distributions that deviate from normality and for higher-order moments. Furthermore, if the focus is on events in a certain part of the distribution, such as extreme events, all statistical moments may contribute to their magnitude and probability. Here, the qq plot provides an objective description for all parts of a distribution and how they deviate from a reference distribution.

1.4 INTERNAL VARIABILITY IN A TRANSIENT CLIMATE

In the idealised control simulation discussed so far, samples for all the techniques can be generated by sampling over the time dimension because all samples are coming from the same population. However, the Earth has undergone changes in the external forcing in the past, which will continue in the future. The insolation changes on long time scales, volcanic eruptions introduce short-term disturbances, and anthropogenic emissions change the composition of the atmosphere and change the radiation budget. All of these external forcing factors can affect the statistics of the Earth system, introducing a potential time-dependence of the statistics. This implies that samples from different time periods in a transient climate do not come from the same population, because the mean and other statistical moments might have changed. Thus the sampling needs to be limited to very short time periods where the statistics can be assumed to be stationary.

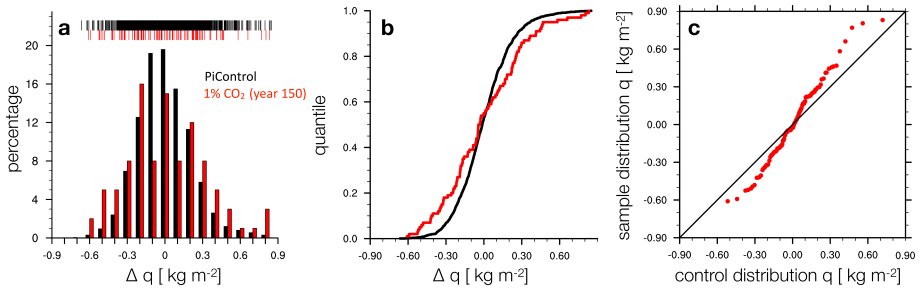

Figure 1.5: **The internal variability of precipitable water increases under strong global warming. (a)** The black histogram shows the control run as in figure 1.3, the red histogram samples from 100 ensemble members in year 150 of the 1% CO_2 experiment. The mean has been removed from both samples. **(b)** cumulative distributions for the control run and year 150 from the 1% CO_2 experiment, **(c)** quantile-quantile plot of the control run and year 150 of the 1% CO_2 experiment.

A commonly used assumption is that only the mean is changing over time while higher-order moments are not changing. By detrending a time series, the change in the mean is removed and sampling over time is used to generate a large sample. In figure 1.1b, the temporal standard deviation over 30-year periods is used to estimate sample internal variability and investigate a change of internal variability. When using this approach, the different realisations lead to different conclusions, ranging from a decrease in variability to an increase in variability. This indicates that the sample size, here 30 years, is too small to robustly estimate internal variability. In a transient climate, increasing the sample size by sampling over longer time periods does not solve the problem: if the statistics are changing over 150 years, a smaller change in the statistics might already occur over a few decades and any sample taken over time is not necessarily based on a single underlying population. The problem of analysing internal variability in a transient climate can be overcome by using a large ensemble of climate model simulations.

To understand how a large ensemble can be used to characterise internal variability in a transient climate, it is useful to revisit the definition of internal variability. In general, internal variability describes the distribution of possible states the climate system can assume. For the observed temperature record, one could for example ask: which global mean surface temperature values would have been possible for a certain year in the record, given that the Earth had followed a slightly different trajectory, but with the same anthropogenic emissions? This distribution of alternative Earths cannot be generated from the observed record, but it is possible to estimate it with an ensemble of climate model simulations that use the same evolution of the external forcing but different initial conditions. The nonlinearity of the climate system can cause even small differences in the initial conditions to grow over time. In such an ensemble, internal variability for any point in time can be estimated by the ensemble spread. This definition of internal variability in the ensemble domain provides a well-defined estimate of internal variability at different background conditions. With the same approach, we can also sample the distributions for specific future boundary conditions, representing states of the Earth that are possible under these boundary conditions when taking internal variability into account.

Combining the sampling over the ensemble dimension with the non-parametric analysis of internal variability presented in figure 1.4 is a powerful tool to robustly quantify a change in internal variability in a transient climate. Here, I use this approach to compare the internal variability at the end of the 1%-CO_2 scenario to preindustrial conditions (figure 1.5). While a comparison of the histograms already indicates that internal variability of precipitable water is increasing, the sample size is still too small to generate a smooth histogram for

the 1% CO_2 experiment. Furthermore, the subjective choice of the bin size and number of bins changes the appearance of the histogram and complicates the interpretation. In this example, there seems to be a large change in the rightmost bin. However, the small number of samples in this bin, indicated by the red lines near the top, raises the question whether this change in the percentage is robust. The cumulative distribution shows a similar picture—an increase of the variability under strong warming—without the need to structure the data into bins prior to the analysis. The quantile-quantile plot can be used to analyse how the variability is changing. All the points in the upper right quadrant are above the one-to-one line, indicating that positive anomalies grow stronger while the points in the lower left quadrant are below the diagonal, indicating that negative anomalies become more negative. From this I conclude that the distribution is widening, and by how much anomalies of a certain magnitude will be changing. However, from this analysis alone it is not possible to identify which of these changes are robust, i.e. what is caused by the change in forcing, and what is only occurring by chance and is therefore an artefact of insufficient sampling.

1.5 ROBUSTNESS OF VARIABILITY CHANGES

To quantify the robustness of the identified changes, we need independent samples of the same size from the same distribution. The range of results from these independent samples can then be used to derive confidence intervals. The clean approach in this context would be to run additional ensembles with 100 members each for the same scenario and compare the results for year 150. However, this approach is not feasible under most circumstances due to the high computational cost. A larger number of samples could also be generated by splitting the sample from the 100 members into several smaller subsamples. While this would allow us to estimate confidence intervals, it comes at the cost of reduced precision. Given that the ensemble standard deviation based on 100 samples contains a large amount of sampling uncertainty (figure 1.1c), reducing the sample size further would increase the sampling uncertainty, making the detection of a small signal more difficult. Therefore reducing the ensemble size is not the preferred approach.

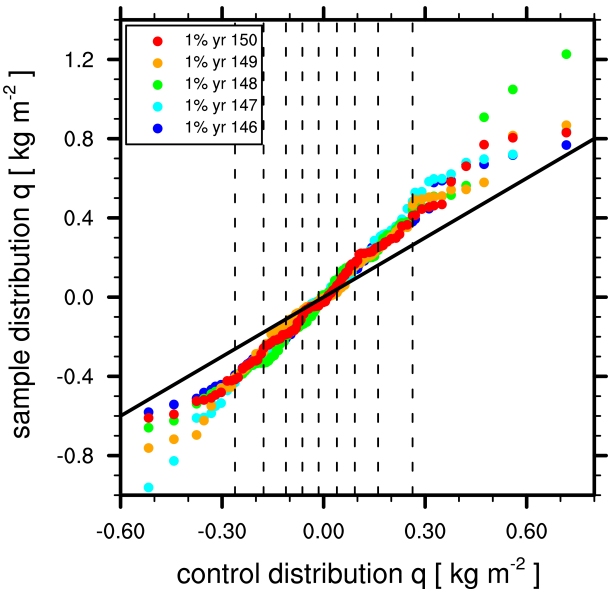

Figure 1.6: **The sign of the forced change is robust for all anomalies while the magnitude of the forced change is only robust for anomalies near the median of the distribution but not for the most extreme anomalies.** Quantile-quantile plot as in figure 1.5, but for each of the last 5 years of the 1%-CO_2 scenario separately. The vertical dashed lines indicate the deciles of the control distribution, i.e. there are 10% of the samples between every two dashed lines.

The alternative approach to generate more samples that I use here is to sample over time. In a transient simulation, this approach may potentially violate the assumption that all samples are identically

distributed. For a small number of consecutive years, I argue that the difference between the years is not dominated by a true change in the distribution but by sampling uncertainty, which I want to quantify. An abrupt change in the external forcing, for example from a volcanic eruption, may violate my assumption and the approach should only be used if the change in the external forcing is small. Here, I use the last five years of the simulation to investigate if the changes in year 150 relative to preindustrial conditions can be attributed to the change in the external forcing. For this purpose, qq plots for each of the last five years are computed and combined in figure 1.6. While the qq plots for all five years follow a similar shape, they diverge near the upper and lower ends.

The different qq plots agree very well for moderate anomalies near the centre of the distribution. For moderate anomalies I conclude that both the sign of the change and the magnitude of the change are robust and can be clearly attributed to the change in the external forcing. For the extreme events in the upper and lower tail of the distribution, the different years do not show the same magnitude of change. However, the sign of the change is robust across the samples from different years. This means that a widening of the distribution in the tails can be attributed to the change in the forcing, but the magnitude of this widening cannot be robustly estimated.

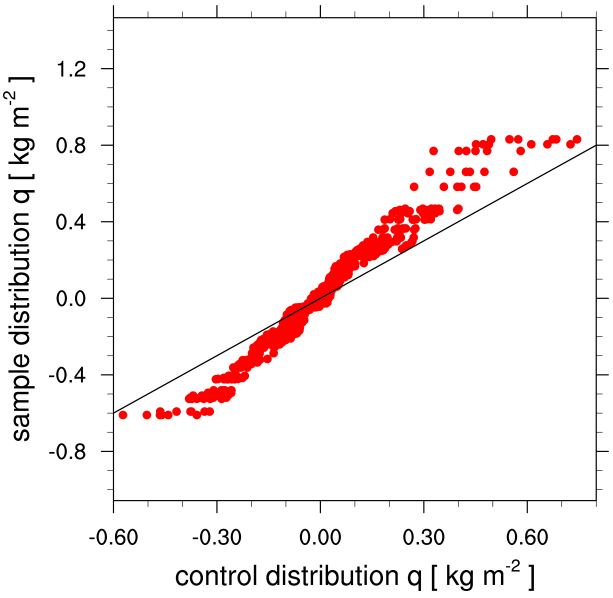

Figure 1.7: **When the sample size of the reference distribution is small, the associated sampling uncertainty can become large.** Quantile-quantile plot based on 6 different samples of size 100 for the reference distribution (sampled from the control run). The sample distribution on the y-axis is based on the last year of the 1%-CO_2 scenario for all 6 reference distributions.

Sampling uncertainty in the reference distribution was not included in the analysis up to this point, but may add to the uncertainty and therefore affect the robustness of a detected change. The reference distribution used in the examples was either a fitted normal distribution as in figure 1.4, or the distribution sampled from a 2000-year preindustrial control simulation (figure 1.5). In both cases, the sample size of the reference distribution is very large compared to the distribution to be investigated, so that sampling uncertainty for the reference distribution was not taken into account. This changes if the reference distribution has a sample size similar to the distribution sampled from a different climate. This may occur when a different time period is used as a reference. For many applications, present-day conditions might be used as a reference state. In this case, the sample for the reference distribution is much smaller because the sampling over time is limited to a sampling period for which stationarity can be assumed. Even with a large ensemble, the sample size for the reference distribution would have a similar size as the sample for a changed climate.

In the qq plot in figure 1.6, sampling uncertainty is only considered along the y-axis, that is, for the distribution in a warmer climate. But there is also a sampling uncertainty associated with the reference distribution, which is uncertainty along the x-axis. To demonstrate this sampling uncertainty for the reference distribution, the control run is randomly subsampled with a sample size of 100. In figure 1.7, six samples from the control run are plotted against the last year of the 1% CO_2 experiment. Sampling uncertainty for the distribution on the y-axis is not shown in this figure. The sampling uncertainty for the reference distribution is comparable to the sampling uncertainty shown in figure 1.6. While the reference distribution in figure 1.6 is based on a sample size of 2000 and sampling uncertainty is negligible, the analysis in figure 1.7 indicates that this may not be true for a smaller sample size. In all applications in my thesis, the sample size of the reference distribution is much larger than the sample size of the distribution that is investigated. Therefore, I will neglect the sampling uncertainty of the reference distribution and only evaluate the sampling uncertainty of the distribution that is investigated.

1.6 THE TRADE-OFF BETWEEN PRECISION AND ROBUSTNESS

An estimate of the sampling uncertainty is essential to decide if a detected signal can be attributed to the change in the forcing. To estimate the sampling uncertainty, the available samples need to be split into subsamples of equal size. For example, the last five years of the 1%-CO_2 ensemble contain 500 values. In figure 1.6, the 500 values have been split into five subsamples of size 100. An alternative approach could have been to split them into ten subsamples of size 50.

On one hand, this would constrain the sampling uncertainty better, which would then be based on ten independent samples rather than five. On the other hand, the sampling uncertainty for the sample of size 50 would be larger than the sampling uncertainty for a sample of size 100, which makes it more difficult to detect a small signal. At the same time, the precision of the statistics estimated from a smaller sample is lower than for a larger sample.

In practice, the precision should be maximised, while keeping enough independent samples to constrain the sampling uncertainty. However, for applications where internal variability is small or the signal is large, it might be possible to partition the available samples into more but smaller subsamples.

Ultimately, the choice for partitioning the available samples will determine the magnitude of the signal that can be detected, and will therefore depend on the question at hand. When applying this framework, one needs to decide what magnitude constitutes a relevant signal. From this information, the necessary sample size can be estimated. In practice, one of the major limitations will be a predetermined number of samples that is available, for example a certain number of ensemble members of a climate model have been run. In this case, the available sample size will determine the detectable magnitude of the signal and therefore determine the questions that can be answered with the available samples.

1.7 THE TEMPORAL EVOLUTION OF INTERNAL VARIABILITY

In the previous examples, I focused on the difference between the very last year of an idealised, strong warming experiment. This maximises the potential difference in the distributions and was chosen to illustrate the analysis method. For practical applications, the change of the distribution over time is of interest to identify when a detectable change first occurs. In figure 1.9, the analysis is applied to different 5-year periods in the 1% CO_2 experiment. This illustrates the temporal evolution. In the first five years, sampling uncertainty leads to a large spread in the upper and lower decile, but no robust signal can be identified. Over time, the widening of the distribution evolves.

An alternative approach is to investigate changes in the statistical moments of the distribution, as illustrated in figure 1.8. While the change in the mean and standard deviation is clearly visible, changes in higher order moments are dominated by sampling uncertainty. Compared to this approach, the non-parametric analysis in figure 1.9 has several advantages, because it aggregates changes in all statistical moments into one graphical representation. Interpreting a change in the magnitude of extreme events from a change in one statistical moment is challenging, while the qq plot allows to easily identify the magnitude of the change for an anomaly in a specific decile of the distribution. At the same time, the qq plot indicates if the change is robust or an artefact of sampling uncertainty.

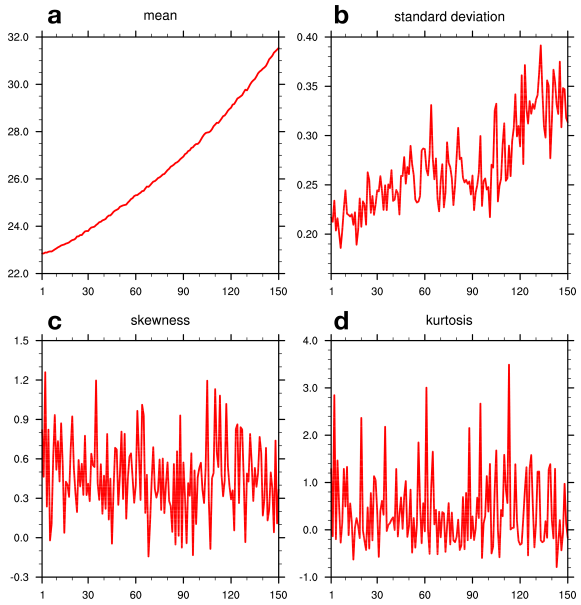

Figure 1.8: **Sampling uncertainty for higher-order statistical moments is large.** Time series of statistical moments computed over the ensemble dimension for global mean annual mean precipitable water in the 1% CO_2 experiment. **(a)** mean, **(b)** standard deviation, **(c)** skewness, and **(d)** kurtosis.

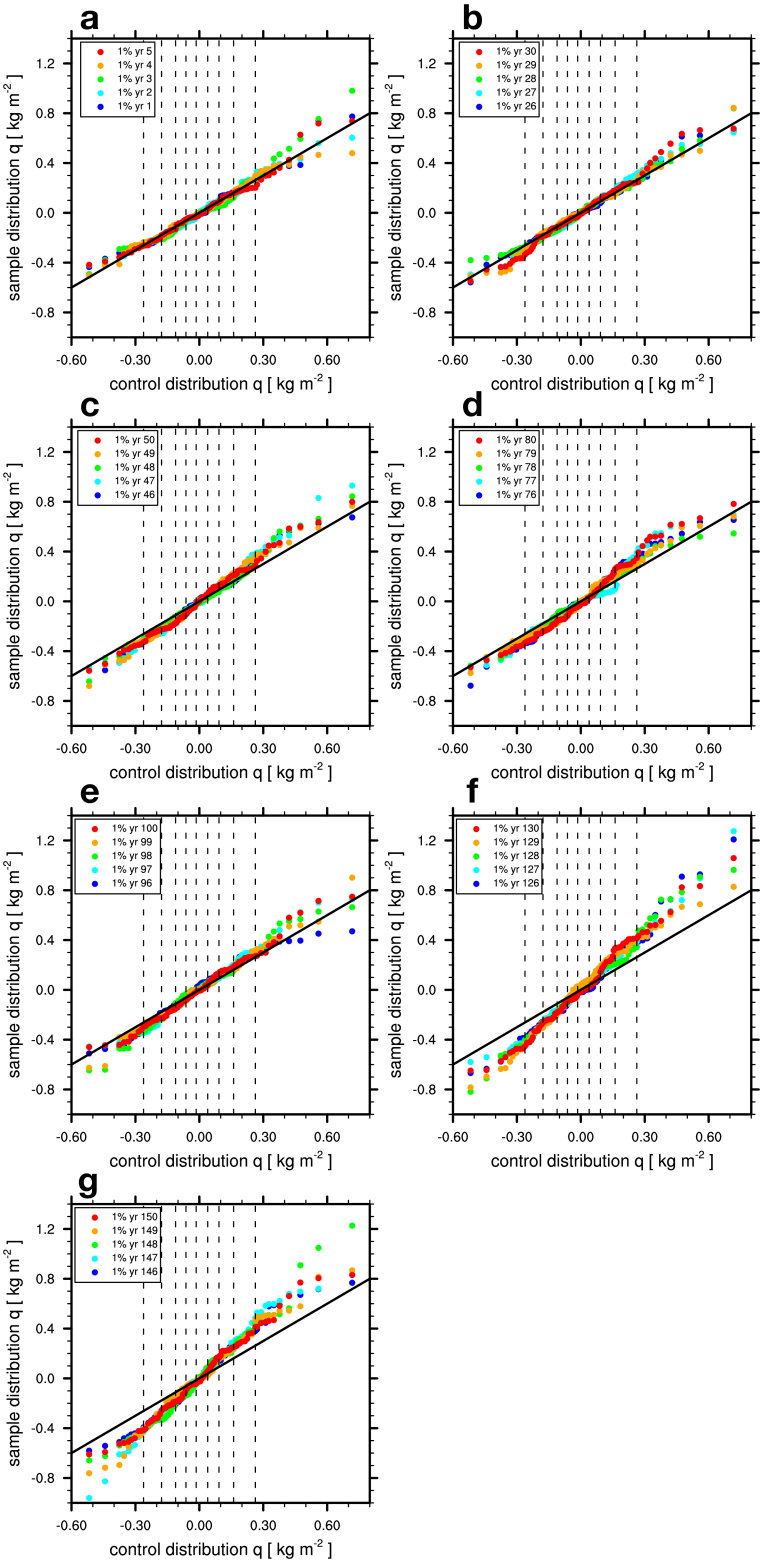

Figure 1.9: **Robust changes in the extreme events appear after 30 years, a broadening of the whole distribution occurs more gradually.** QQ plots similar to figure 1.6 for different 5-year windows over the whole length of the simulation . The vertical dashed lines indicate deciles of the control distribution. a: years 1-5, b: 26-30, c: 46-50, d: 76-80, e: 96-100, f:126-130, g: 146-150.

1.8 REGIONAL INHOMOGENEITY

In the previous sections, I used a globally averaged quantity to develop an analysis framework to detect and attribute changes in internal variability. The advantage of using globally averaged precipitable water is that the effect of the circulation is removed and only thermodynamics contribute to changes in the mean state and variability. While globally averaged quantities are often used to characterise the state or evolution of the climate system, they are not necessarily representative for changes on a regional scale. In the following, I evaluate the spatial patterns of changes in the mean and standard deviation of precipitable water under strong global warming.

Figure 1.10a shows that most of the precipitable water is in the tropics. In response to global warming, precipitable water is increasing mostly in the tropics and less at higher latitude (figure 1.10b). This follows the theory by Held and Soden (2006) and can be characterised as an intensification of existing patterns. Internal variability in annual mean water vapour is dominated by the tropical Pacific, where the El-Niño Southern Oscillation causes large variability in the surface temperature (figure 1.11a).

While the change in the mean state in response to global warming is increasing everywhere, with larger values in the tropics, the change in variability shows more spatial inhomogeneity (figure 1.11b). The tropical Pacific shows increasing internal variability under global warming, other regions show decreasing variability despite an increase in the mean state. In particular over the tropical Atlantic, regions north an south of the equator show a decrease in variability.

The spatial inhomogeneity in the distribution of precipitable water, its variability and the changes in response to global warming suggest that the regional characteristics can be very different and are not well described by a globally averaged analysis. Therefore, in the second chapter, I investigate regional changes in rainfall and its variability.

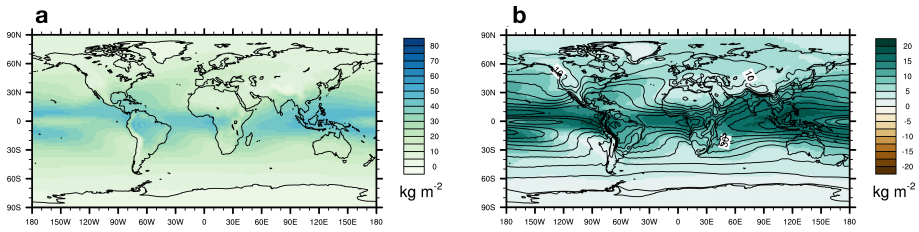

Figure 1.10: **Distribution of precipitable water is not homogeneous. In response to global warming, patterns are intensified. (a)** Mean state of annual mean precipitable water for PiControl conditions, **(b)** change from PiControl to the end of the 1% CO_2 experiment. Contour lines show the PiControl distribution as in (a), the distance between contour lines is 5 kg m^{-2}.

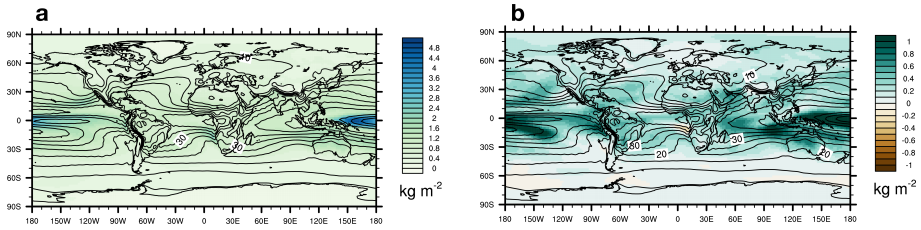

Figure 1.11: **Internal variability of precipitable water is largest in the tropics and dominated by the tropical Pacific. Changes in variability are not homogeneous. (a)** Temporal standard deviation (shading) and mean state (contour lines, spacing 5 kg m^{-2}) for PiControl, **(b)** change of variability from PiControl to the end of the 1% experiment. Contour lines as in (a)

1.9 CONCLUSIONS

The aim of this chapter was to develop a method to robustly detect and attribute changes in internal variability in response to a change in the external forcing. I show that internal variability of globally averaged precipitable water increases under global warming. This change in variability can only be robustly detected in a large ensemble of model simulations, while a single realisation may either indicate increasing or decreasing variability, even if the change in the external forcing causes an increase in variability. Furthermore, the non-parametric approach I use allows me to quantify changes in different parts of the distribution whereas parametric approaches, which mostly rely on the standard deviation to describe internal variability, are not able to clearly discriminate between changes in extreme events or anomalies closer to the centre of the distribution.

The advantage of using the ensemble domain is that this approach ensures stationarity of statistics, whereas analyses in the temporal domain need to assume that stationarity is true for the time period under investigation. The limitation of the ensemble domain is that all information about the time scales of variability is lost. If two realisations show different states of the climate system for a specific year in an ensemble simulation, this only means that these two states are possible under the given background conditions and forcing history. It cannot be deduced if these two states could occur in consecutive years in a single realisation, or if they are part of a multi-decadal oscillation and can only occur several decades apart.

Finally, I show that that the spatial distribution of changes in the internal variability of precipitable water is not homogeneous. For the purpose of this chapter, a globally averaged quantity with an expected increase in variability under global warming was chosen to develop and test the analysis framework. While this simplification is an ideal test bed for the purpose of this chapter, it should be used with care in

practical applications, because different, possibly opposing changes, are aggregated.

The analysis of the spatial distribution of precipitable water indicates that changes in regional mean precipitation and its variability might occur under global warming, in particular in the tropics. Interestingly, the variability is decreasing in some regions, despite an increase in the mean state. In the second chapter, I will use the framework from this chapter and apply it to regional rainfall changes under global warming.

2

THE INTERNAL VARIABILITY OF RAINFALL IN THE TROPICAL ATLANTIC REGION AND ITS FUTURE CHANGES

2.1 SUMMARY

Rainfall variability on regional scales determines the magnitude and frequency of anomalous events. Therefore, quantifying rainfall variability and its response to warming is crucial to understand forced changes in the past and in the future. Globally, the hydrological cycle is expected to increase by intensifying existing patterns of precipitation minus evaporation (Held and Soden 2006) and rainfall variability has been projected to increase (Pendergrass et al. 2017). On smaller regional scales like the Sahel, future projections are uncertain (Biasutti et al. 2008; Monerie et al. 2016; Rowell et al. 2015), while the observed drying in the second half of the 20th century has been hypothesised to be caused by changes in the external forcing (Biasutti and Giannini 2006). There is some agreement on the causes of past changes in the Sahel region, but possible future changes remain elusive. I show that in a large ensemble of model simulations for the 20th century, a forced increase in Sahel rainfall both for the mean and variability exists after the 1980s, but individual realisation may not show this change. Internal variability of Sahel rainfall is projected to increase under warming along with an increase in the mean rainfall and the larger internal variability in a warming world adds to the uncertainty of future projections. However, rainfall variability is not increasing everywhere in the tropical Atlantic region under global warming. On the southern flank of the oceanic ITCZ, the internal variability is decreasing, driven by changes in the circulation. By analysing different regions, I show that a change in variability does not necessarily scale with a change in the mean state.

2.2 INTRODUCTION

Understanding past and future changes in rainfall is a key question in climate science because regional changes in rainfall have a large effect on the local population. More rainfall can lead to flooding, less rainfall to water scarcity and droughts. The African Sahel, the region south of the Sahara desert and north of tropical rainforest, receives most of the annual rainfall from July to September while dry conditions dominate during the rest of the year (Nicholson 2013; Thorncroft et al. 2011). The Sahel experienced a drying in the 1960s to 1980s (Nicholson 1980; Nicholson et al. 2000) (figure 2.1), with a subsequent recovery of mean rainfall to 1960s values (Martin and Thorncroft 2013). The question if this drying can be attributed to anthropogenic influence or internal variability is not only important to understand the past, but also the future changes that might occur in response to global warming. Giannini et al. (2003) found that warm SST over the Atlantic weakens the continental convergence associated with the monsoon and thus contributed to the observed drying. Biasutti and Giannini (2006) argue that the late 20th century drying has a forced component of at least 30%. While most coupled models simulate a long-term drying over the Sahel in the 20th century, they fail to capture the timing of the 1960s to 1980s drying. This leads to the question if the drying was indeed forced and the models fail to reproduce this forced signal, or if the observed drying can be explained by internal variability.

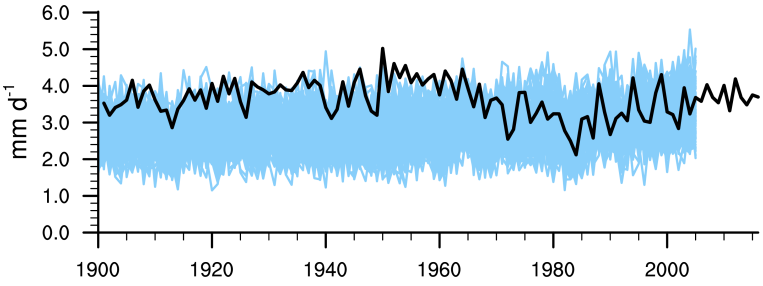

Figure 2.1: **Historical JAS Sahel rainfall shows substantial variability on multi-decadal time scales with a drying from the 1950s to 1980s and a subsequent wetting.** Seasonal mean July to September rainfall averaged for 20°W-35°E and 10°N-20°N for observations (black line, CRU TS (Harris et al. 2013)) and all MPI-GE historical realisations (blue lines).

Future model projections of Sahel rainfall diverge between simulations from different models (Biasutti 2013; Biasutti et al. 2008; Rowell et al. 2015), raising the question how the hydrological cycle responds to warming and what contributes to the uncertainty in the model projections. Held and Soden (2006) have found that the response of the hydrological cycle to global warming can be described as an intensification of the existing spatial pattern of evaporation minus

precipitation and associate this with a decrease in convective mass flux and an increase in the horizontal moisture transport. Based on observations, Greve et al. (2014) have shown that aridity changes cannot be explained by a simple intensification of the existing patterns. This is supported by the findings of Byrne and O'Gorman (2015) that a simple thermodynamic scaling of moisture convergence fails over land and tends to overestimate the precipitation response to warming. However, over the ocean, changes in precipitation can be described as an intensification of existing patterns (Byrne and O'Gorman 2015; Held and Soden 2006). The response of the oceanic ITCZ to global warming can be described as an intensification and narrowing of the convergence zone (Byrne and Schneider 2016).

In this chapter, I evaluate forced changes in Sahel rainfall for the 20th century and for future projections in the MPI-GE. Motivated by the finding from the first chapter that variability in water vapour increases under global warming, I explore forced changes in rainfall variability in response to global warming and how they may affect the past and future trajectory of Sahel and tropical Atlantic rainfall.

While the mean state changes under global warming over the land and ocean have been investigated extensively, changes of internal variability of rainfall are often not considered or internal variability is assumed to not change under global warming (Hawkins and Sutton 2010; Thompson et al. 2015). For example, Monerie et al. (2017) investigated a forced signal in Sahel rainfall under global warming and treated internal variability as time-invariant noise that conceals the forced signal and impedes the detection of forced changes. However, the characteristics of internal variability determine how large and how frequent anomalous or even extreme events are. If the internal variability changes, this can have a large effect on the occurrence of these events and the effect should not be neglected. The results from my first chapter suggest that the variability of precipitable water in the atmosphere changes in response to global warming, which indicates that the internal variability of precipitation may change as well.

Pendergrass et al. (2017) analysed internal variability of rainfall and its response to global warming. They found that internal variability of rainfall increases by 3-4% per Kelvin of global warming on a global average, with higher values over land. They attribute these changes to an increase in atmospheric moisture content and find that a weakening circulation can partially mitigate the moisture-driven increase in variability. However, they aggregate variability over large regions. The results from my first chapter and Byrne and O'Gorman (2015) suggest that the response of precipitation to global warming varies regionally. Therefore, a globally aggregated description of the changes in rainfall variability may not adequately characterise changes on a regional scale that can be dominated by different mechanisms.

I will first focus on the Sahel region and apply the analysis framework from the first chapter to disentangle the forced trend, internal variability and forced changes in the variability to address the following research questions:

- **RQ2.1:** How does the mean state and internal variability of Sahel rainfall change in response to global warming?
- **RQ2.2:** Are observed Sahel rainfall changes forced or can they be explained by internal variability?

A common assumption is that the variability scales with the change in the mean state: when the mean state increases, variability increases as well. In section 2.5, I investigate the response of the oceanic ITCZ over the tropical Atlantic to global warming, motivated by the research question:

- **RQ2.3:** Does the change in variability scale with a change in the mean state everywhere in the tropical Atlantic region?

2.3 RAINFALL IN THE MODEL AND OBSERVATIONS

Current generation climate models struggle to represent the climatological mean rainfall for the historical period. Especially in the tropics, the models have large biases (Stevens et al. 2013). In the Pacific, there are bands of rainfall simulated north and south of the equator, rather than on the equator, known as the double-ITCZ bias. In the Atlantic, rainfall over the ocean in CMIP5 models is displaced either to the west or east, but no model matches the observed rainfall distribution (Siongco et al. 2015). While Biasutti et al. (2006) attribute the precipitation biases in models mostly to simulated rainfall being too sensitive to warm SSTs south of the equator, Siongco et al. (2015) find that the models can be classified by either having too much precipitation near the coast in the west or east of the tropical Atlantic, which cannot be explained by the hypothesis of Biasutti et al. (2006) alone. For the MPI-ESM, Eichhorn and Bader (2016) find that the precipitation bias is strongly affected by the warm-SST bias in the southeastern tropical Atlantic, which affects both the mean state and seasonal cycle of precipitation. In general, Stephens et al. (2010) find that while climate models tend to provide a reasonable representation of time integrated and globally composited rainfall, on regional scales models rain twice as often as observations, but compensate this error by simulating rainfall that is too light.

Previous analyses of model biases were mostly limited to comparing a time mean from observations to a time mean from one or several model realisations and define the difference as the model bias. This approach does not take into account that observations and a model simulation represent different realisations of the Earth system and

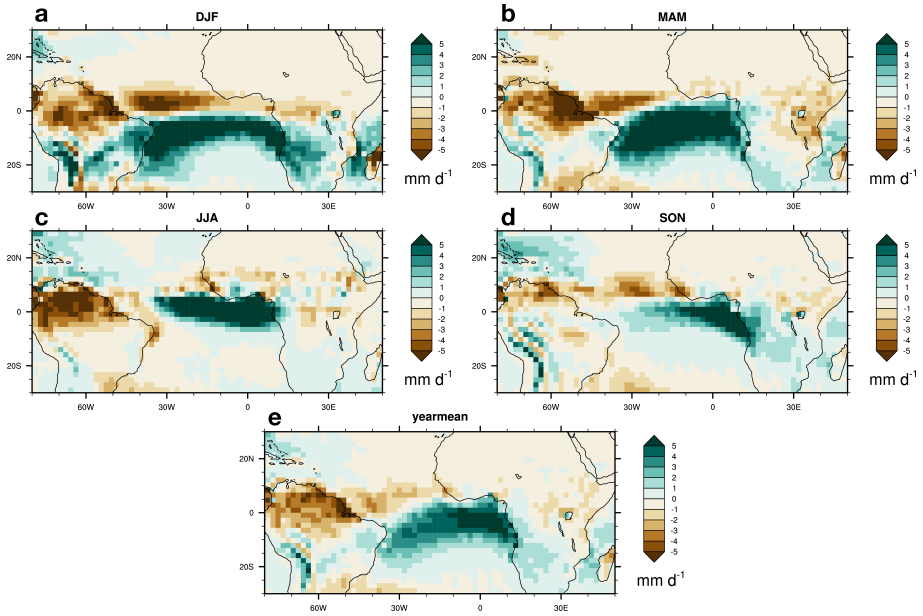

Figure 2.2: **The simulated oceanic ITCZ is shifted south relative to observations.** Seasonal and annual mean difference of GPCP and MPI-GE rainfall averaged for 1979-2005.

might even diverge if the model is an accurate representation of the climate system.

Biases in variability are more challenging to quantify. In a first step, variability must be isolated from the forced signal. This requires a precise quantification of the forced signal (Frankcombe et al. 2018). In a single time series, the forced signal is not known. Approaches such as fitting a linear trend might overestimate internal variability, while a high-pass filter might underestimate variability. Therefore, isolating variability in the single realisation of observations or a single model simulation is difficult, or even impossible without additional knowledge of the underlying physical processes. In addition, internal variability can have a decadal or even multi-decadal component (e.g. figure 1.2). Therefore, a sample based on a short time series might sample low or high internal variability for a decade. If the observed period and a simulated period differ in their internal variability, this does not necessarily imply a systematic bias in internal variability. Only when the time period that is compared is long enough to capture all states of internal variability that can occur, that is when no significant contribution from variability on time scales longer than the analysis period exists, the difference can be interpreted as a systematic bias.

For the observed record, this means we need to extrapolate from what we have observed in the past and be convinced that no event that has not been observed could occur. This approach has been used to scale simulated internal variability to make it resemble the observed

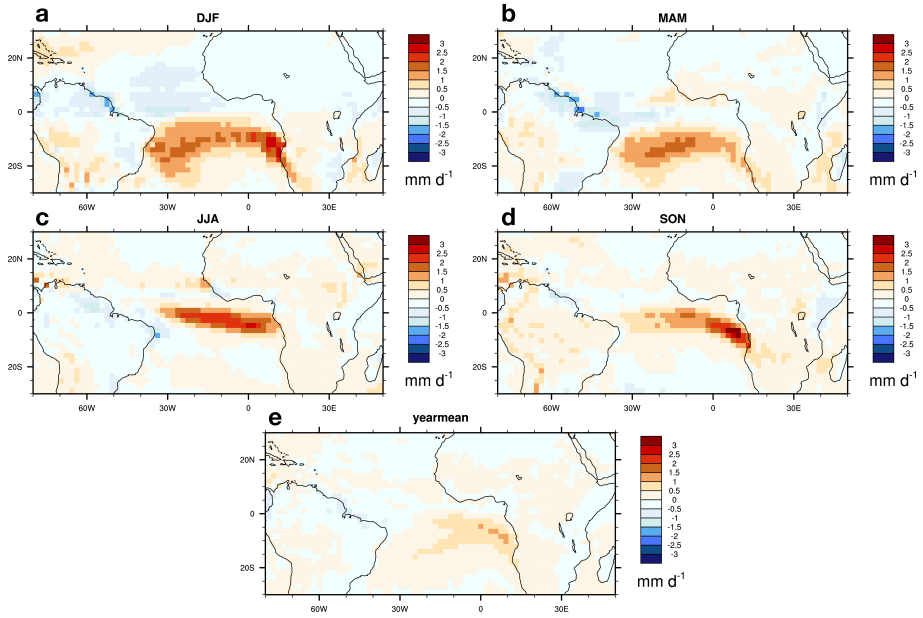

Figure 2.3: **Simulated rainfall variability on the southern flank of the oceanic ITCZ is too strong and shifted too far south.** Seasonal and annual mean difference of GPCP and MPI-GE rainfall variability for 1979-2005. For observations, temporal standard deviation is used; for the model, temporal standard deviation is computed for each member and then averaged over the ensemble. No detrending is applied because this might artificially reduce internal variability. The record shows no statistically significant trend for the analysis period.

variability (McKinnon and Deser 2018; McKinnon et al. 2017). However, there are several examples where apparent inconsistencies between observations and model simulations could be explained by internal variability (Hedemann et al. 2017; Maher et al. 2018; Suárez-Gutiérrez et al. 2017), emphasizing that internal variability should be considered when discussing differences between models and observations. It has to be noted that in some regions, the difference between simulated and observed quantities can only be explained by systematic differences between models and observations and the influence of internal variability can be disregarded until model biases are reduced to values that have the same order of magnitude as internal variability.

Here, I compare simulated rainfall in the MPI-GE to monthly GPCP precipitation (Adler et al. 2016), which is based on station data and satellite observations. I am using the period from 1979-2005, where both GPCP precipitation and model simulations with historical forcing are available. Observations are interpolated to the model grid prior to computing differences. For comparing the mean state, the temporal mean for observations and the temporal and ensemble mean for the model are computed and shown for all seasons and the annual mean in

figure 2.2. Over the ocean, the ITCZ is displaced to the south and east. This has been partially attributed to the warm SST biases in the south east Atlantic (Eichhorn and Bader 2016; Siongco et al. 2017). Rainfall over land is too low over Brazil, but reasonably well represented over Africa north of the equator.

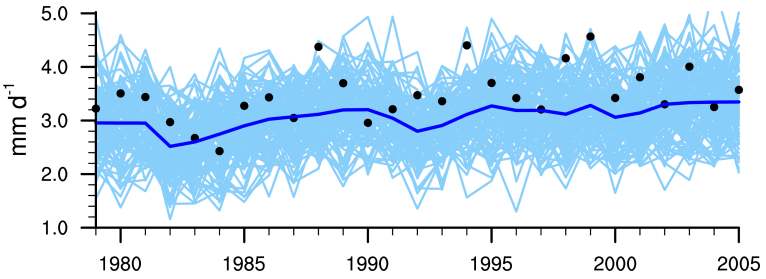

Figure 2.4: **Observed JAS rainfall over the Sahel lies within the ensemble spread of simulated rainfall.** Rainfall averaged for 20°W-35°E and 10°N-20°N. Black dots show GPCP rainfall, blue lines show all 100 model realisations of MPI-GE for the historical period. The thick blue line indicates the ensemble mean.

The bias in internal variability follows the mean state bias in general as seen in figure 2.3. Over the ocean, variability is displaced to the south, following the displacement of the ITCZ, and larger than in observations. Over land, the variability is similar in the model and observations. In the Sahel region, the main rainy season is during boreal summer when the monsoon is contributing a large fraction of the annual rainfall. Therefore, I focus on July to September (JAS) for the evaluation of Sahel rainfall (as used by Giannini et al. (2003) and others). Simulated rainfall agrees well with observations over the Sahel region (figure 2.4). However, figure 2.1 indicates that observed Sahel rainfall in the mid-20th century is near the upper end of simulated rainfall. The fact that the observations are within the ensemble range show that the model is able to capture observed rainfall values, but the consistent high rank of observations among model simulations indicates that the model may overestimate variability or not represent a forced component accurately. Figure 2.4 shows the simulated and observed rainfall for the Sahel region for 1979–2005. In July to September, Sahel rainfall is reasonably well represented. However, observed values are near the upper end of the range simulated by the ensemble, indicating that the model is biased towards lower values for rainfall in this period and might overestimate internal variability.

The spatial correlation of JAS Sahel rainfall with global rainfall as shown in figure 2.5a is qualitatively similar to the findings of Biasutti et al. (2008) (their figure 3). The correlation of Sahel rainfall with surface temperature (figure 2.5b) indicates that there might be an influence of SST on Sahel rainfall, mostly from the tropical oceans but also with contributions from the Mediterranean and North Atlantic.

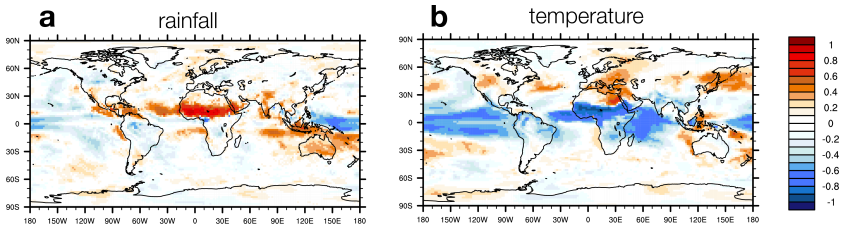

Figure 2.5: **Correlation of JAS Sahel rainfall with global rainfall and temperature indicates teleconnections with surface temperature in the tropical oceans as well as in the North Atlantic and Mediterranean.** Correlation coefficients for Sahel rainfall and **(a)** precipitation and **(b)** temperature globally. Correlation coefficients are computed over the ensemble dimension for the beginning of the historical experiments.

Previous studies have suggested that the North Atlantic (Knight et al. 2006) and Mediterranean (Park et al. 2016) affect Sahel rainfall. Therefore, the basic characteristics of Sahel precipitation seem to be represented reasonably well in the MPI-ESM, despite large biases in the oceanic ITCZ.

In my analysis, I extend the analysis season for Sahel rainfall from JAS to JJASO to account for possible shifts in the seasonality of rainfall under global warming that have been suggested by Dunning et al. (2018).

2.4 RAINFALL CHANGES IN THE SAHEL

Sahel rainfall under strong global warming

Sahel rainfall changes are first investigated under strong global warming in the 1% CO_2 experiment. By using this idealised experiment, I investigate if the mean rainfall and the internal variability of rainfall show a robust response to global warming. As shown in figure 2.6, the mean rainfall in the Sahel is increasing by 50% in response to strong global warming. The internal variability, defined as the ensemble standard deviation, is increasing by slightly less than 50% by the end of the simulation.

Because the standard deviation only captures symmetric changes of the distribution, I use the non-parametric approach from the first chapter to quantify changes in the rainfall distribution compared to the PiControl experiment (figure 2.7). Robust changes in the upper 10% of the distribution occur after about 50 simulation years (figure 2.7c), while changes in more moderate anomalies start to emerge in the second half of the simulation. By the end of the simulation, the whole distribution has widened compared to the PiControl experiment (figure 2.7g). Changes to the distribution are symmetric and changes in positive and negative anomalies contribute equally. Thus changes to the distribution are explained by changes in even-order statistical moments like the standard deviation and kurtosis. Here, these changes are dominated by the change in the standard deviation.

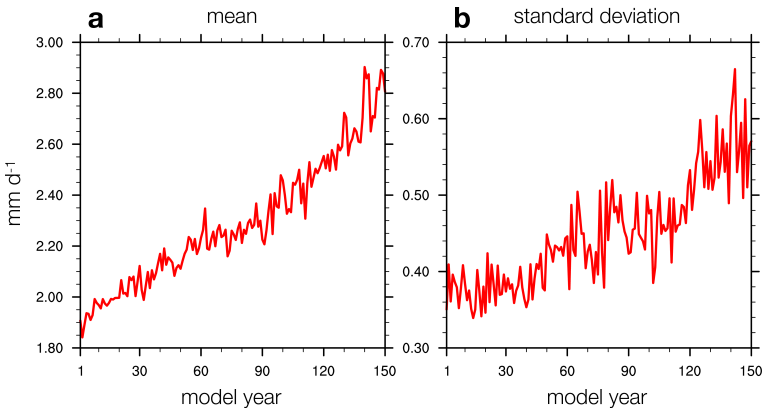

Figure 2.6: **Increase in mean rainfall and rainfall variability over the Sahel in response to global warming.** Time series of **(a)** ensemble mean and **(b)** ensemble standard deviation for JJASO Sahel rainfall in the 1% CO_2 experiment.

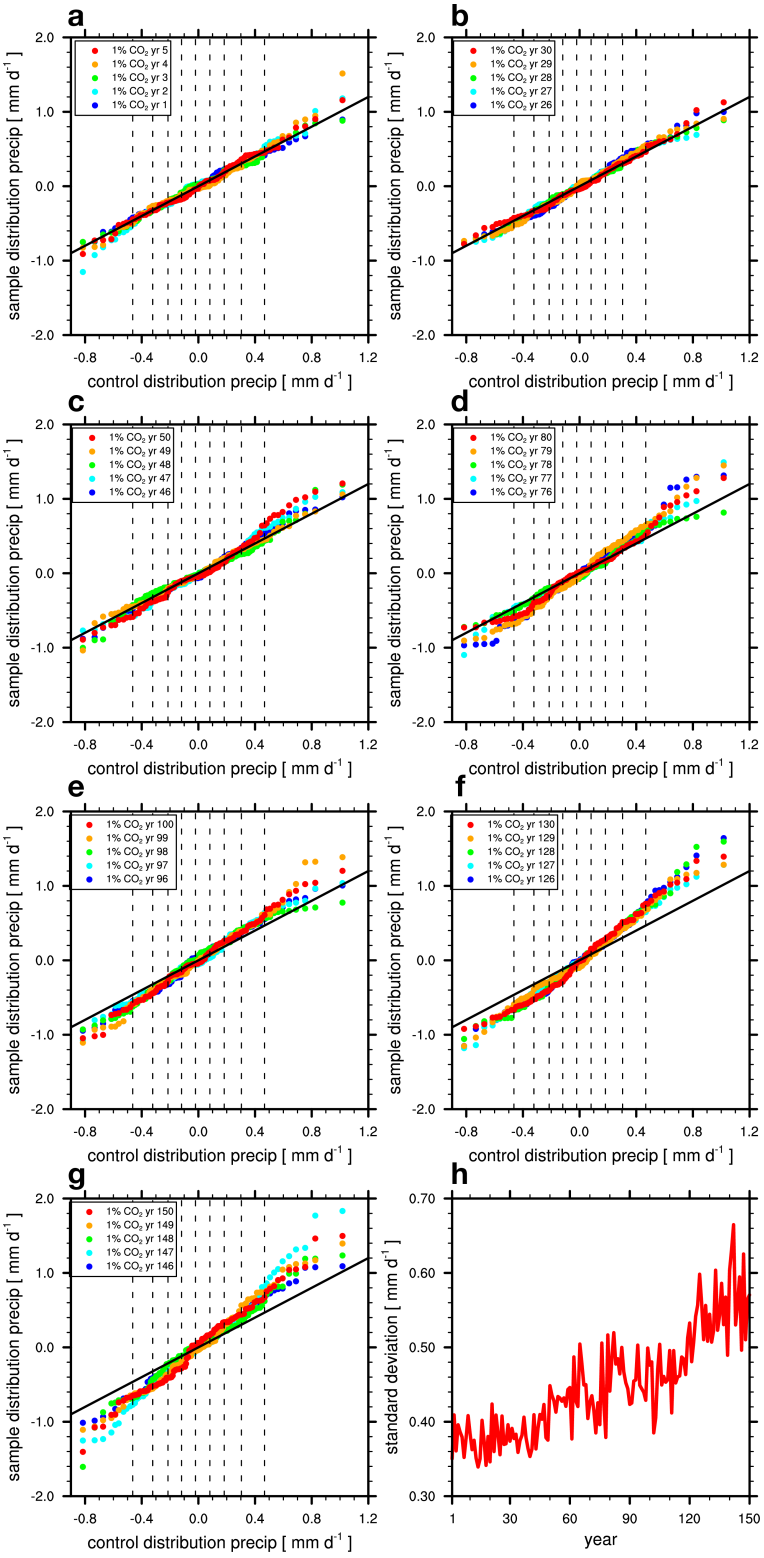

Figure 2.7: **The increasing JJASO Sahel rainfall variability in response to global warming is robust.** QQ plots for different 5-year periods from the beginning to the end of the 1% CO_2 experiment for the Sahel region in JJASO. As in figure 1.6, each of the five years is plotted individually against the preindustrial control distribution in panels (a)-(g). Dashed lines indicate deciles of the control distribution. Panel (h) shows the time series of ensemble standard deviation as in figure 2.6b.

In response to the strong warming in the 1% CO_2 experiment, the model simulates a robust forced increase in mean rainfall and a robust increase in rainfall variability. The trajectory of a single realisation may still be dominated by internal variability, making it difficult to identify the contribution of the forcing. In figure 2.8, time series of JJASO Sahel rainfall are shown for all members and the ensemble mean. The increase in the mean is clearly visible and the increase in variability can be seen as a widening of the envelope of all realisations seen as the light red lines in the background. In figure 2.8 a, two individual realisations are highlighted: the realisation with the steepest slope of a fitted linear trend (light grey), and the realisation with the weakest slope of a fitted linear trend (dark grey). Both of these realisations show a positive slope, indicating an increase of the time-mean rainfall. However, the slope can be much larger or weaker than the forced signal, which is indicated by the ensemble mean (red line). Figure 2.8b highlights two different realisations: the one with the largest trend in internal variability (light grey), and the realisation with the lowest trend in internal variability (dark grey). The trend in internal variability is calculated as the difference of the 30-year standard deviation from the end of the simulation to the beginning of the simulation. While one realisation is characterised by small interannual fluctuations in the beginning and large fluctuations in the end, the other realisation shows nearly the opposite characteristic. As shown in figure 2.7, internal variability shows a robust increase in response to global warming, but a single realisation as shown in figure 2.8b may not show this clear change.

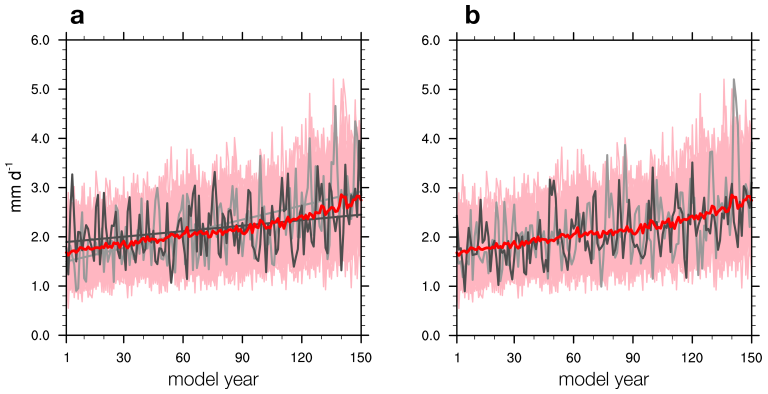

Figure 2.8: **Trajectories of individual realisations may deviate from the forced change in the mean and variability.** Light red lines show all realisations for Sahel JJASO rainfall in the 1% CO_2 experiment, the thick red line shows the ensemble mean. **(a)** realisations with maximum (light grey) and minimum (dark grey) linear trend. The fitted linear trend is shown in the same colour. **(b)** realisations with maximum (light grey) and minimum (dark grey) trend in 30-year standard deviation from the beginning to the end of the simulation.

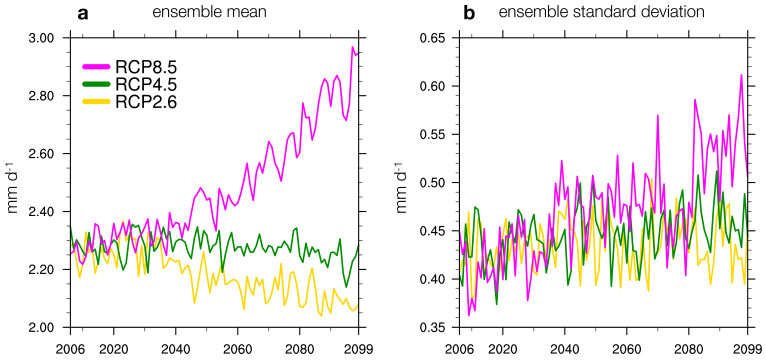

Figure 2.9: **Response of the mean and variability vary for different RCP scenarios. (a)** time series of ensemble mean JJASO Sahel rainfall in RCP2.6, RCP4.5, and RCP8.5 scenario. **(b)** Time series of the ensemble standard deviation for the scenarios.

Sahel rainfall in different future scenarios

In the previous section, I established that the model shows increasing mean rainfall and rainfall variability in the Sahel in response to an idealised, strong warming experiment. Here, I include the analysis of three different future scenarios: RCP2.6, RCP4.5, and RCP8.5. While the mean temperature increase in RCP8.5 is comparable to the 1% CO_2 experiment (figure 0.1), RCP4.5 shows a weaker GMST warming with 2K by the end of the 21st century compared to preindustrial, and the RCP2.6 scenario shows a warming of about 1.5K by the end of the 21st century. While temperature is increasing in all of these scenarios, only the RCP8.5 scenario shows an increase in the Sahel rainfall (figure 2.9a). RCP2.6 shows decreasing mean rainfall for the Sahel, while there is hardly any change in the RCP4.5 scenario.

Changes in the internal variability are symmetric as in the 1% CO_2 experiment (not shown), therefore I describe changes in the variability only by changes in the ensemble standard deviation (figure 2.9b). RCP8.5 shows an increase in variability similar to the response in the 1% CO_2 experiment, while the internal variability does not change in the RCP2.6 or RCP4.5 scenario.

Historical Sahel rainfall

The MPI-GE simulations show a robust response of Sahel rainfall to strong global warming, but a weaker response in more moderate warming scenarios such as RCP2.6 and RCP4.5. This leads to the question if the change in the external forcing over the historical period influenced Sahel rainfall. To answer this question, I analyse the 100 realisations from the historical experiment. Here, I use July to September (JAS) for the analysis to allow a direct comparison to previous studies.

Furthermore, changes in the seasonality have only been suggested for future climate change (Dunning et al. 2018), whereas JAS captures the monsoon rainfall well over the historical period.

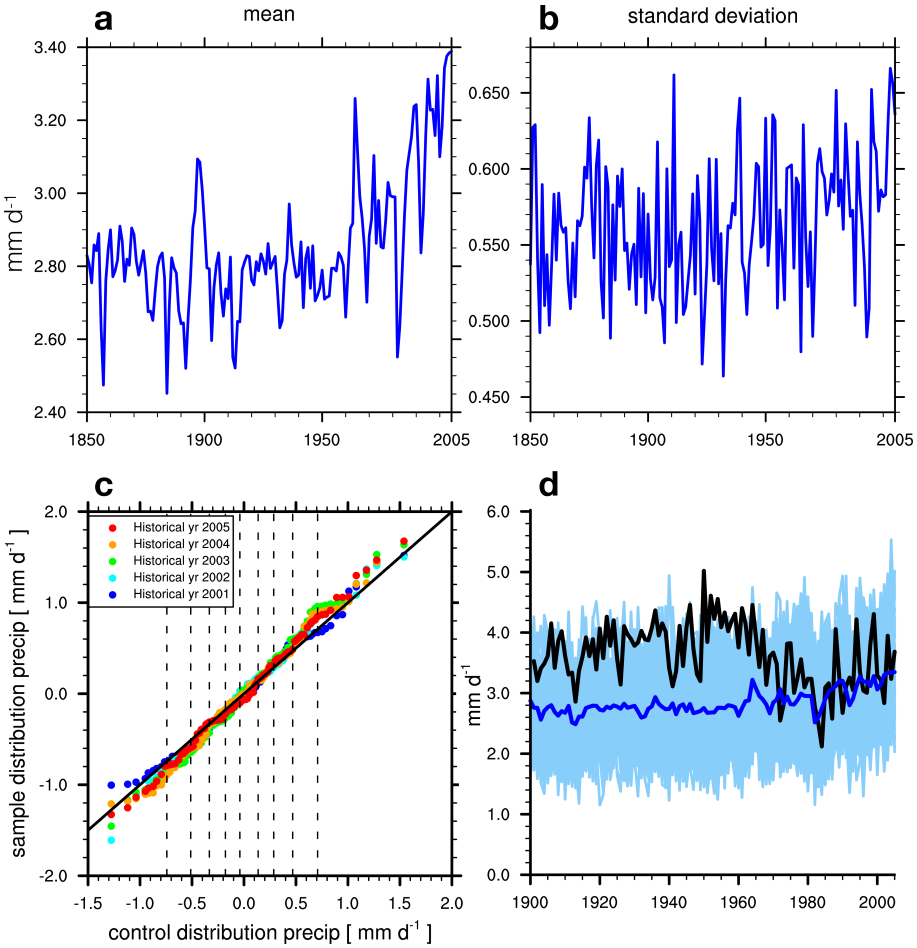

Figure 2.10: **Despite robust changes in Sahel rainfall over the 20th century, a single realisation might not show any of these changes.** JAS Sahel rainfall in the historical experiment: **(a)** ensemble mean time series, **(b)** ensemble standard deviation, **(c)** qq plot comparing years 2001–2005 to preindustrial conditions, and **(d)** all individual realisations, the ensemble mean (thick blue) and observations (CRU TS).

Figure 2.10a,b show the temporal evolution of the ensemble mean and ensemble standard deviation for the historical period. The ensemble mean is increasing after the 1950s, suggesting a forced increase in Sahel rainfall. At the same time, the ensemble mean shows large spikes, indicating either forced events or sampling uncertainty. Internal variability is dominated by sampling uncertainty and shows slightly higher values in the last years. An inspection of the qq plot for the last five years in figure 2.10c shows that hardly any robust change in the distribution can be identified and all changes are close

to the one-to-one line. Furthermore the magnitude of the changes is negligibly small.

This prompts the question if the forced change simulated by the model can be seen in the observations, and if the observed changes can be explained by the model. Figure 2.10d shows time series for all ensemble members, CRU TS observations (as in figure 2.1) and the ensemble mean. While both the model and observations show an increase in the precipitation from the 1980s onward, the model seems to underestimate the rainfall in the first half of the century. All realisations of the model show large variability on interannual time scales, but hardly any variability on multi-decadal time scales comparable to the 1960s to 1980s drying in the observations. This indicates that the model might not capture a forced drying in this period, or it might underestimate internal variability on multi-decadal time scales(e.g. Yan et al. 2018).

2.5 RESPONSE OF THE OCEANIC ATLANTIC ITCZ TO GLOBAL WARMING

The results for the Sahel region indicate that the change in variability can be understood by a scaling with the mean state: increasing mean values are accompanied by an increase in the variability. In this section, I address the third research question and investigate if this holds for all regions. Here, I focus on the oceanic ITCZ in boreal summer (JJA) in the 1% CO_2 experiment.

In the mean state, the oceanic ITCZ is located slightly north of the equator, and the largest variability can be found on the southern flank (figure 2.11a). In response to the strong warming, the rainfall is increasing in the centre of the ITCZ and reduced on the northern and southern flank as seen in figure 2.11b, which agrees with previous work (Byrne and Schneider 2016; Held and Soden 2006; Huang et al. 2013). The change in the variability shows three regions with a strong change under global warming: decreasing variability south of the ITCZ, increasing variability on the southern flank of the ITCZ and increasing variability over the Sahel region (figure 2.11c). In the following, I will focus on the two regions on the southern flank of the oceanic ITCZ (marked in figure 2.11d) that indicate a northward shift of the variability. The change over the Sahel region was discussed in the previous section.

For further analysis, rainfall is averaged over the regions indicated in figure 2.11d and analysed in the time and ensemble dimension. The change in the ensemble mean and ensemble standard deviation are shown in figure 2.12. This analysis shows that the variability is increasing over time in the northern region and decreasing in the southern region, while the mean state is decreasing in both regions.

This indicates that the change in variability cannot be assumed to scale with the mean state change.

A closer analysis of the change in variability in the northern region shows that the changes in the variability emerge after the first half of the simulation (figure 2.13d). The analysis also reveals that the distribution is skewed, with the largest negative anomalies being nearly twice as large as the most positive anomalies. Despite this skewness, changes in positive and negative anomalies show an equal magnitude and contribute to the increase in variability.

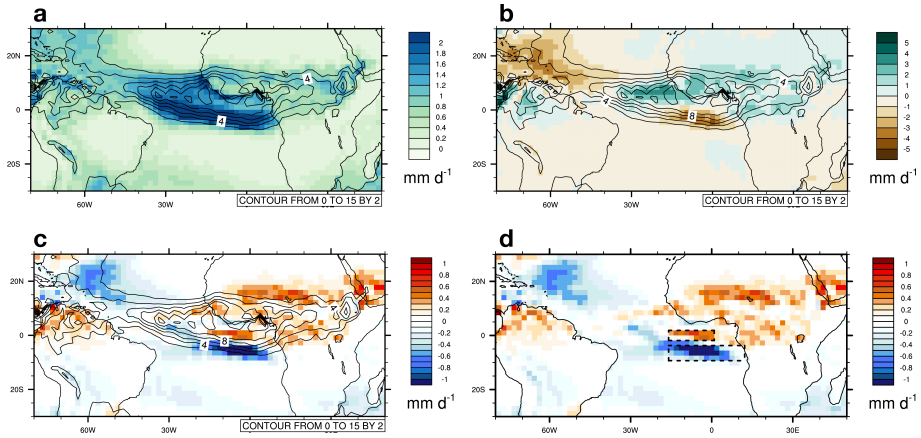

Figure 2.11: **The oceanic ITCZ is narrowing and intensifying under global warming. The southern flank of the ITCZ and therefore its variability is simulated further north.** (a) Mean state (contour lines) and variability (ensemble standard deviation, shading) averaged for years 1-5 of the 1% CO_2 experiment for JJA; (b) change in the ensemble mean rainfall from the first five years to the last five years of the 1% CO_2 experiment (shading), mean state as in a (contour lines); (c) change in variability from the first five years to the last five years (shading), mean state as in a; (d) same as c, boxes indicate regions with change in variability on the southern ITCZ flank.

In the region further to the south, which is located south of the more narrow ITCZ in a warmer climate, changes in the variability are not symmetric. Changes in the negative anomalies start to emerge after about 50 simulation years (figure 2.14c,d), while changes in positive anomalies only emerge much later (figure 2.14f) and contribute less to the overall narrowing of the distribution. The analysis shows that while the mean rainfall in this region is steadily decreasing under global warming, variability is first reduced by fewer large negative anomalies while positive anomalies remain unchanged. Reduced magnitude of positive anomalies only occurs much later. This asymmetric change can be explained by the lower threshold of the rainfall distribution: rainfall cannot be less than zero. Therefore, the largest possible negative anomaly is given by the magnitude of the mean state. In the

first years, the mean rainfall is close to 2 mm/day and the largest negative anomaly is also close to 2 mm/day. When the mean state is reduced to 1.5 mm/day, the largest possible negative anomaly is 1.5 mm/day. Here the decrease in the mean rainfall sets a lower bound for the magnitude of negative anomalies and explains the reduction of negative anomalies in figure 2.14.

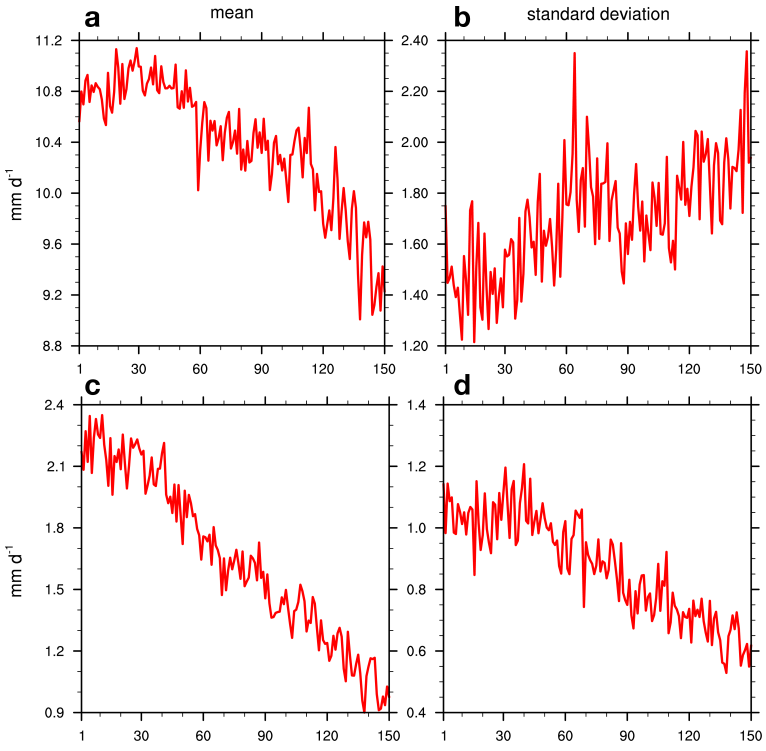

Figure 2.12: **The change in variability does not follow the change in the mean state in all regions.** Temporal evolution of the ensemble mean and ensemble standard deviation of JJA rainfall south of the centre of the ITCZ (a,b) and on the southern flank of the ITCZ (c,d) in the 1% CO_2 experiment.

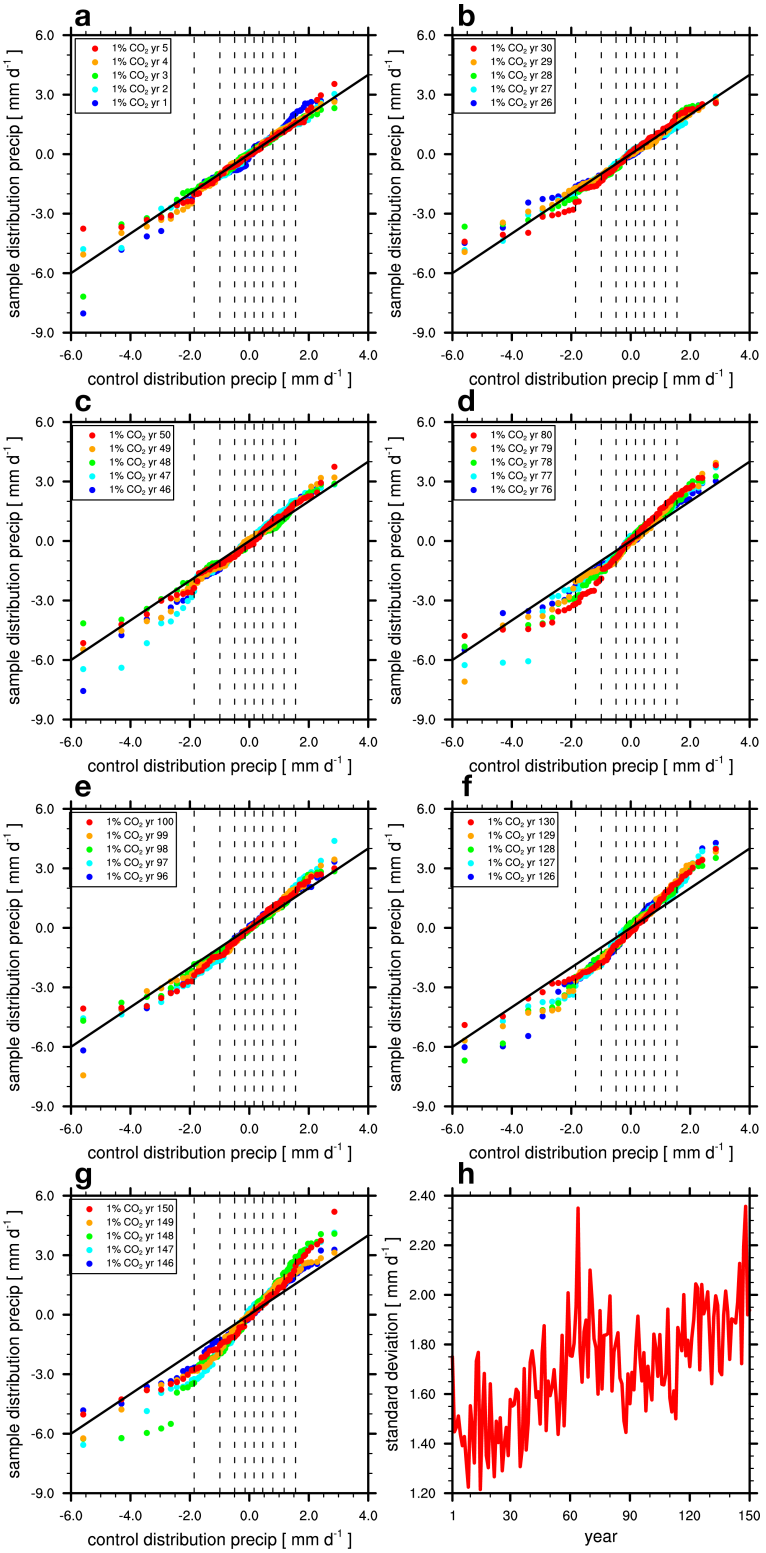

Figure 2.13: **Increasing JJA variability in response to global warming is robust in the region on the southern flank of the ITCZ.** QQ plots for different 5-year periods from the beginning to the end of the 1% CO_2 experiment for the region on the southern flank of the oceanic ITCZ in JJA. As in figure 1.6, each of the five years is plotted individually against the preindustrial control distribution in panels (a)-(g). Dashed lines indicate deciles of the control distribution. Panel (h) shows the temporal evolution of the ensemble standard deviation.

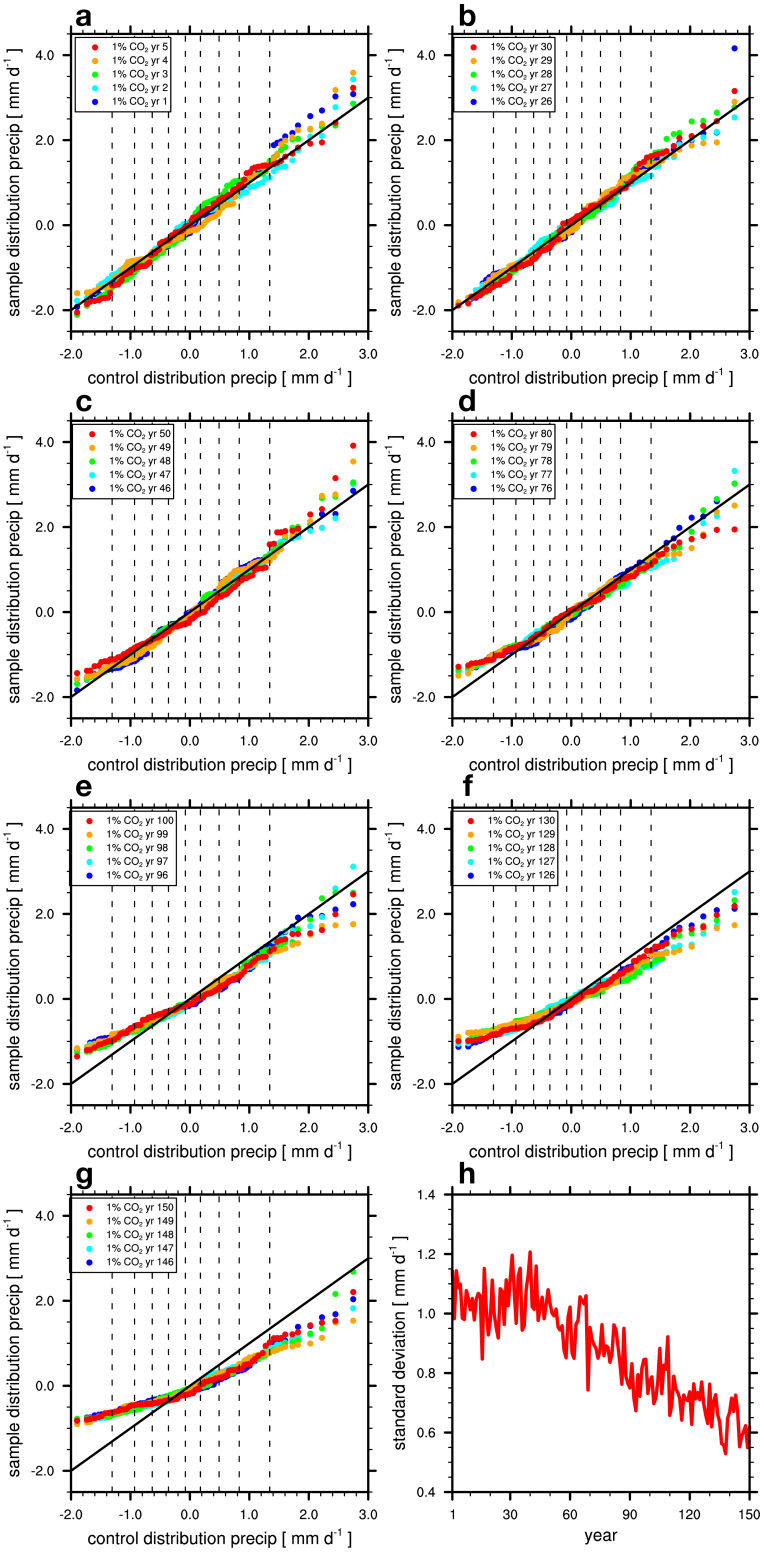

Figure 2.14: **Decreasing JJA variability in response to global warming is robust in the region south of the oceanic ITCZ, with changes in negative anomalies occurring earlier than for positive anomalies.** Same as figure 2.13, but for the region south of the ITCZ.

2.6 DISCUSSION AND CONCLUSIONS

In this chapter, I applied the analysis framework from the first chapter to answer three research questions on Sahel rainfall and the oceanic ITCZ over the tropical Atlantic ocean. I conclude by summarising the findings and discussing their implications for each of the three research questions.

RQ2.1: How does the mean state and internal variability of Sahel rainfall change in response to global warming?

Sahel rainfall is increasing in response to strong global warming in the 1% CO_2 experiment and RCP8.5 scenario according to the model used in this study. The increase in the mean state is accompanied by an increase in the internal variability. Park et al. (2016) have found that the influence of the Mediterranean on the West-African monsoon will increase under global warming. They show that increased moisture transport into the Sahel region during the onset of the monsoon leads to a stronger convection, which in turn causes an increased advection from the Atlantic into the Sahel. This leads to enhanced rainfall in the monsoon season. Sheen et al. (2017) find that the warming in the North Atlantic and the Mediterranean enhance Sahel rainfall through increasing meridional convergence of externally sourced moisture. These remote controls of mean Sahel rainfall are likely contributing to the mean state increase under strong warming seen in the MPI-GE. However, if the increase in the variability is caused by the same mechanism, then the correlation of Sahel rainfall with Mediterranean or North Atlantic SST as shown in figure 2.5b should increase, which is not the case (not shown). An alternative explanation for the increase in variability is that the moisture convergence from the Atlantic during the monsoon season is more efficient in a warmer atmosphere. Thus, if a small anomaly during the onset increases the monsoon in its early stage, the later intensification will be stronger than in a colder atmosphere. Thus small anomalies during the onset of the monsoon in some years, for example caused by moisture inflow from the Mediterranean, may develop into a much stronger monsoon season. Sheen et al. (2017) find that Sahel rainfall variability on interannual time scales is affected by the tropical Pacific and western Indian ocean SSTs. By modulating the zonal circulation, these regions affect the ascent and local recycling of Sahel moisture. This mechanism may contribute to the increase in Sahel rainfall variability under global warming in the MPI-GE. Further investigation of the mechanisms driving the increase in the variability would require additional analysis of the moisture budget to disentangle thermodynamic and dynamic contributions to the anomalies.

While the response of the model to warming was robustly identified and quantified, the implications for a single realisation are less clear as shown in figure 2.8. Assuming that the simulated realisations are adequate representations of possible future realisations, this implies that identifying a forced change in the observations will be difficult. While Monerie et al. (2017) focus on identifying a forced signal in the mean state, I can show that internal variability itself is changing in response to the change in external forcing. While the increase in variability makes identifying a forced change in the mean more difficult, the increasing internal variability itself will influence single trajectories by making the range of possible annual rainfall values in the Sahel larger than in present-day conditions. But even when there is a forced increase in the variability, such as in this model, a single realisation might show a decrease in variability over time as shown in figure 2.8b. If this would happen in future observations, it would inevitably raise doubts about the reliability of the model. This emphasises the need to evaluate forced changes of the mean and variability in a large number of realisations and quantify the likelihood that any single realisation shows this forced change or seemingly contradicts the hypothesised forced response by not providing evidence for a forced change.

RQ2.2: Are observed Sahel rainfall changes forced or can they be explained by internal variability?

While the observed past rainfall in the Sahel lies within the spread of the historical simulations, there are clear indications that the model does not capture all features of the observed trajectory. The model shows larger variability on interannual time scales, but does not reproduce multi-decadal variability similar to the observed multi-decadal trends.

Models do in general underestimate variability on multi-decadal time scales (Martin et al. 2014). Because the model may underestimate multi-decadal variability, I cannot deduce from the simulations if the observed drying trend from the 1960s to the 1980s was forced or could be explained by internal variability.

The apparent lack of simulated multi-decadal variability can also be caused by an underestimation of a forced response. In particular, Knight et al. (2006) have suggested that Sahel rainfall is influenced by the Atlantic Multidecadal Variability (AMV). While Sahel rainfall in the MPI-GE is correlated with North Atlantic surface temperature (figure 2.5b), the AMV in the MPI-GE might have a smaller forced contribution than in observations (Hand et al. 2018). The amplitude of the AMV in individual historical realisations is similar to observations, but the phase is not the same in all realisations. If the Sahel drying in the second half of the 20th century was indeed controlled by a forced change in North Atlantic surface temperature, this relationship may

not be captured by the historical simulations in the MPI-GE because the North Atlantic SST in the model does not appear to have a large forced contribution itself.

RQ2.3: Does the change in variability scale with a change in the mean state?

Changes of rainfall variability on the southern flank of the oceanic ITCZ show increasing variability, while the mean state is decreasing. This emphasises that locally, the scaling of the variance with the change in the mean does not necessarily hold.

Changes in the rainfall variability could either be caused by changes in the temperature variability or changes in the circulation-driven moisture convergence. The temperature variability over the tropical Atlantic does not change in any of the warming scenarios (not shown). Previous studies did not find any evidence for a change in temperature variability under global warming either (Huntingford et al. 2017; Lehner et al. 2018). Therefore, a change in the circulation is the most likely explanation for the change in variability. In response to global warming, the simulated ITCZ is narrowing (figure 2.11b) and the southern flank is moving further north. This can explain both the decrease in variability in the south and the increase in variability further north. Under preindustrial-like conditions, the investigated region in the south is located on the southern flank of the ITCZ. Thus, in years when the ITCZ extends further south, this region receives more rainfall and less in the years when the southern flank is further north. The narrowing ITCZ under global warming moves the southern flank further north. The southern region is now outside of the ITCZ and receives less rain.

The region further north is initially close to the centre of the ITCZ, where it receives large amounts of rainfall in most years. When the ITCZ is narrowing, this region is located on the southern flank, leading to a reduction of mean rainfall. At the same time this region is now under the influence of the variable position of the southern flank, which explains the increased variability.

The results for the oceanic ITCZ regions emphasise that investigating the changes in rainfall variability in an aggregated form as in Pendergrass et al. (2017) may conceal opposing effects in different regions. Shifts in patterns as identified over the tropical Atlantic are accompanied by changes of opposite sign and would therefore not show up when averaging over the whole region. Only an analysis of the spatial distribution of the variability can help to identify regions with different characteristics. Furthermore, I show that changes in the distribution are not necessarily symmetric. In particular for rainfall, where a reduction in the mean state can move the distribution closer to its lower bound and thereby reduce the magnitude of negative anomalies without affecting the upper part of the distribution. Using

the standard deviation to quantify internal variability does not capture these details.

While the simulated rainfall over the Sahel region is reasonably close to observations, simulated rainfall over the ocean has large systematic biases as shown in section 2.3. This strongly impedes the credibility of future projections. To improve the reliability of future projections of tropical Atlantic rainfall, these model biases must be addressed.

3

TOWARDS REDUCING MODEL BIASES IN THE TROPICAL ATLANTIC

The results from this chapter have been published in: Milinski, S., J. Bader, H. Haak, A. C. Siongco, and J. H Jungclaus (2016), High atmospheric horizontal resolution eliminates the wind-driven coastal warm bias in the southeastern tropical Atlantic, Geophys. Res. Lett., 43, 10,455–10,462, doi: 10.1002/2016GL070530.

3.1 SUMMARY

I investigate the strong warm bias in sea surface temperatures (SST) of the southeastern tropical Atlantic that occurs in most of the current global climate models and has been linked to precipitation biases. I analyse this bias in the Max Planck Institute Earth System Model at different horizontal resolutions ranging from 0.1° to 0.4° in the ocean and 0.5° to 1.8° in the atmosphere. High atmospheric horizontal resolution eliminates the SST bias close to the African coast, due to an improved representation of surface wind-stress near the coast. This improvement affects coastal upwelling and horizontal ocean circulation, as confirmed with dedicated sensitivity experiments. The wind-stress improvements are partly caused by the better represented orography at higher horizontal resolution in the spectral atmospheric model. The reductions of the coastal SST bias obtained through higher horizontal resolution do not, however, translate to a reduction of the large-scale bias extending westward from the African coast into the southeastern tropical Atlantic.

3.2 INTRODUCTION

The SST biases in the tropical Atlantic are a long-standing problem common to most climate models (Richter et al. 2014). Ding et al. (2015) find that the mean state biases affect the representation of interannual variability in the tropical Atlantic, which might not be true for other models (Richter et al. 2014). The warm bias is largest along the eastern boundary of the southeast tropical Atlantic (SETA) and, while covering large parts of the tropical south Atlantic, decreases towards the west (figure 3.1c).

In this study, I focus on the coastal SST bias that I define as the localised, strong rise of simulated SST close to the African coast in the SETA region. I show that increased horizontal resolution in the atmosphere eliminates the coastal SST bias due to a better representa-

tion of the surface wind-stress which can be partly explained by better resolved orography.

Multiple causes for the development of the warm bias in the SETA region have been suggested (Richter 2015). A local underrepresentation of low-level clouds was found to create excessive heating of the ocean by shortwave radiation (Wahl et al. 2011) but also a remote contribution from the surface wind-stress on the equator via Kelvin waves travelling southward along the coast has been suggested (Richter et al. 2011). Locally, strong winds close to the coast drive coastal upwelling, bringing cold water masses to the surface (Nicholson 2010). These surface winds are too weak in many models, leading to an underestimation of the coastal upwelling (Gent et al. 2010; Large and Danabasoglu 2006; Richter et al. 2011; Vanniere et al. 2014) and misrepresentation of horizontal ocean circulation (Small et al. 2015). At higher atmospheric horizontal resolution, these winds were found to increase, coincident with a reduction of the coastal SST bias (Doi et al. 2012; Small et al. 2014). These studies indicated that increasing horizontal resolution in the atmosphere can alleviate the persistent SST biases in the models. However, the attribution of the too weak winds close to the coast to a certain atmospheric model component remains elusive (Griffies et al. 2011; Small et al. 2015).

A possible cause for the wind bias might be the misrepresentation of the coastal orography (Harlaß et al. 2015; Large and Danabasoglu 2006). Low-resolution spectral atmospheric models fail to represent the gradients and the height of the orography in the vicinity of strong orographic gradients, such as on the African coastline in the SETA region. This misrepresentation is due to the Gibbs phenomenon that arises from the truncation of higher order terms during the transformation of the observed orography to the spectral domain. Close to strong gradients in the observed orography, the truncation of the higher order terms leads to a more gradual slope as well as overshooting and undershooting of the observed height (Washington and Parkinson 2005). These deficiencies can be reduced by including more higher order terms in the spectral domain, that is, increasing the horizontal resolution of the spectral atmospheric model. In the Pacific, the coastal low-level jet off the coast of California was found to depend on an adequate representation of coastal orography, land-sea contrast and the shape of the coastline (Ranjha et al. 2016). Although it seems plausible that the orography contributes to the wind bias and thus to the SST bias, it has not been shown that the misrepresentation of the orography in spectral models is the cause of the wind bias and how much it contributes to the SST bias.

Here I systematically investigate the influence of the orographic resolution on the surface winds by replacing the orography in a high-resolution simulation with a low-resolution orography. This isolates the effect of the low-resolution orography on the surface winds and

subsequently on the SST bias while maintaining the high-resolution for all other model components. Furthermore, I examine the effect of the improved surface winds on upwelling and advection in the ocean model by using dedicated sensitivity experiments.

3.3 MODEL, DATA AND METHODS

The Max Planck Institute Earth System Model (MPI-ESM, Giorgetta et al. 2013) is used for this study. It consists of the MPI Ocean Model (MPIOM, Jungclaus et al. 2013) version 1.5 and the spectral European Center-Hamburg (ECHAM6, Stevens et al. 2013) atmospheric model version 6.1. Both the ocean and atmosphere model are used at high and low horizontal resolution in different combinations. The high-resolution ocean model is running on an eddy-resolving 0.1° tripolar grid with 40 vertical levels (Storch et al. 2012), whereas the low-resolution model version is using a tripolar grid with 0.4° horizontal resolution but the same vertical resolution. The high-resolution atmospheric model is running at T255, denoting a triangular truncation of the spherical harmonics to 255 wave numbers, providing a horizontal resolution of approximately 40 km. The low-resolution model has a resolution of T63 (~200 km); both have 95 vertical layers. I use a set of four experiments that cover all possible combinations of high and low horizontal resolution in the atmosphere and ocean (*HR*: T255 atm. / 0.1° oc.; *LR*: T63 atm. / 0.4° oc.; *HRatm*: T255 atm. / 0.4° oc.; *HRoc*: T63 atm. / 0.1° oc.; note that *LR* here is the same as *MR* in Giorgetta et al. 2013). All simulations are initialised from a spun-up state of an *LR* control run for the ocean and use the same preindustrial forcing. The ocean and the atmosphere are coupled every hour. The 26-year-mean from each experiment is used to analyse the SST differences. The experiments have different integration lengths, ranging from 38 to 90 years. Analysis of periods from the beginning and end of each simulation revealed no evidence for SST drift in the tropical Atlantic.

As a reference SST the period 1980-2005 (26 years) from the HadISST1 dataset (Rayner et al. 2003) on a 1° grid is used. The time mean SST bias is calculated relative to HadISST by subtracting the spatially averaged SST for the tropics (30°S-30°N, all longitudes) from all experiments and the observations to account for the different mean states. All datasets are interpolated to a regular horizontal 0.25° grid.

A flux adjusted experiment with a modified surface wind-stress is carried out with the *LR* model version. The flux correction terms for the wind-stress are derived from the climatological difference between *LR* and *HR* wind-stress. This correction term is added to the momentum flux computed by the atmospheric model before it is applied to the ocean. The experiment is run for 20 years, the analysis is

using the last 10 years. The model adjusts to the different wind-stress within the first two years and does not show any drift thereafter.

A sensitivity experiment with modified mean orography, *HRatm-MOD*, is constructed based on the *HRatm* setup. I implement the mean orography from the T63 model version into the *HRatm* setup to quantify the effect of the resolution of the orography on the surface wind-stress and SST bias. To construct this orographic field, the T255 surface geopotential is truncated to T63 by setting all higher wavenumbers to zero. The subgrid-scale fields, which affect parameterisations, have not been changed. This experiment is integrated for 10 years, preceded by a 2-year spin-up.

3.4 RESULTS

In this study I differentiate between the coastal and the large-scale bias and focus mostly on the coastal bias. The large-scale bias is defined as the warm bias covering most of the south-equatorial Atlantic with an approximately linear increase towards the coastline in the east. The coastal bias is defined as the localised, strong warm anomaly close to the coast that is superimposed on the linear eastward increase of the large-scale bias.

Increased atmospheric horizontal resolution eliminates the coastal SST bias in the SETA region while it does not significantly affect the large-scale SST bias in the MPI-ESM. In a suite of experiments with different combinations of high and low horizontal resolution in the atmosphere and ocean, the coastal SST bias is eliminated in those experiments with high horizontal resolution in the atmosphere (HR, HRatm), whereas the experiments with low atmospheric horizontal resolution (LR, HRoc) exhibit a strong coastal SST bias (figure 3.1a). The observed SST is monotonically decreasing towards the eastern coast of the south Atlantic. A zonal slope similar to the observations can be seen in those experiments with a high horizontal resolution in the atmosphere albeit with a positive offset in the global average SST. The experiments with low resolution in the atmosphere (LR, HRoc) exhibit a sharp rise in SST close to the African coast. This coastal bias is even stronger in HRoc where only the ocean horizontal resolution is increased, whereas the ocean resolution has no substantial effect on the coastal bias at high atmospheric resolution. The large-scale SST bias, which is the difference in the zonal slope of the SST between observations and the model further off the coast in figure 3.1a, is not significantly affected by changes in atmospheric or oceanic horizontal resolution. This is also evident from the 2-dimensional distribution of the bias in figure 3.1b and c: the coastal bias is reduced at high atmospheric horizontal resolution while the large-scale bias is not affected.

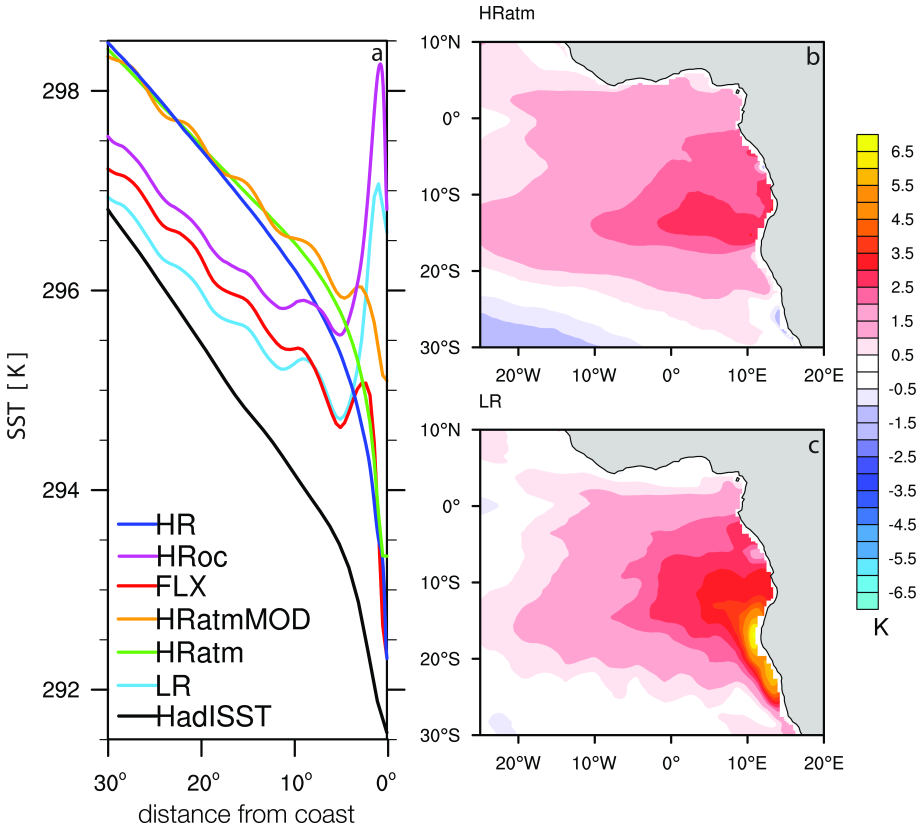

Figure 3.1: **High atmospheric horizontal resolution eliminates coastal SST bias in the SETA region. (a)** coast-following meridional mean of SST on model grid, averaged 15°S to 25°S; **(b)** time mean SST bias for HRatm (0.5° horizontal resolution); **(c)** time mean SST bias for LR (1.8° horizontal resolution)

I have established that the origin of the coastal SST bias lies within the atmospheric model component. Thus the surface fluxes that are provided by the atmosphere cause the coastal SST bias. Because the local dynamical forcing of the ocean is mainly determined by the surface wind-stress, I test the influence of the surface wind stress on the SST bias in a sensitivity experiment.

In figure 3.2a, the difference in meridional wind-stress between the reference experiment with low and the experiment with high atmospheric horizontal resolution is shown. In a region extending 1-2° off the coast, the southerly meridional wind-stress is stronger at higher atmospheric resolution. The difference between the HR and LR wind-stress is applied to a low-resolution experiment as a flux adjustment. The SST close to the coast in this flux-adjusted experiment (figure 3.1a, *FLX*, red curve) closely follows the SST in the high-resolution simulation (*HR*, dark blue curve), showing that the coastal SST bias is indeed caused by the surface wind-stress.

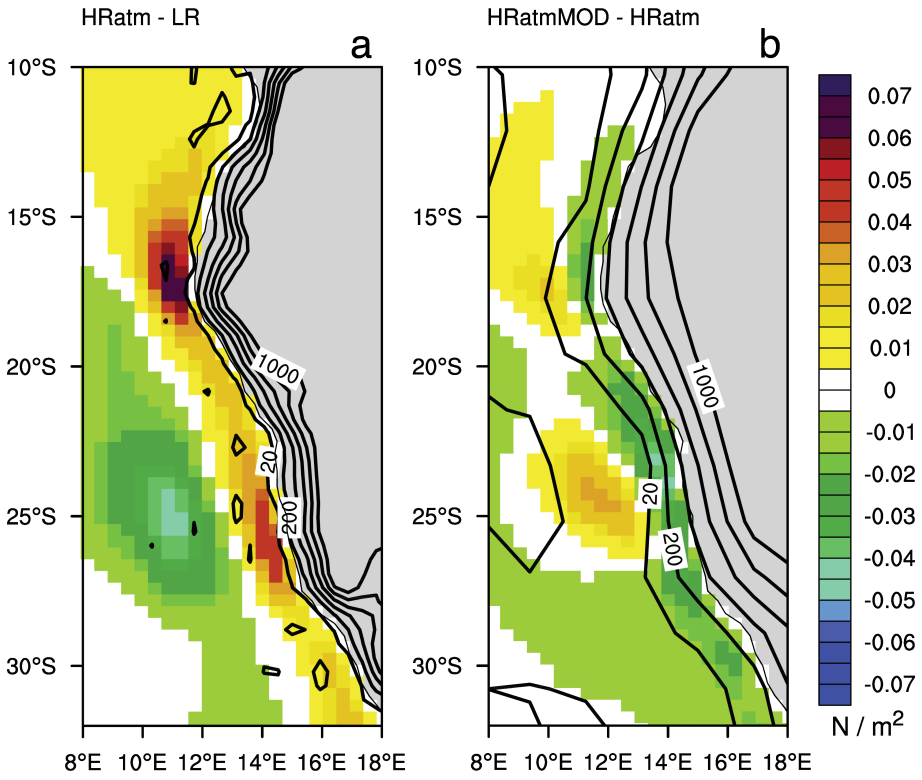

Figure 3.2: **Meridional surface wind-stress difference: (a)** meridional wind-stress difference between high and low horizontal resolution in the atmosphere; there are stronger southerly winds close to the coast at high resolution; **(b)** effect of the low-resolution orography on the surface wind-stress. The plot shows the difference between HRatm and HRatmMOD (high-resolution atmosphere, but low-resolution orography). The contour lines show the height of the orography at high resolution (a) and low resolution (b)

The coastal orography has in the past been suggested to contribute to the coastal SST bias and the accompanying too weak southerly winds. The coincidence of the too weak winds in LR with the positive elevation of the orography over the ocean close to the coast (figure 3.2b, orography contours) suggests that the orography at low resolution might cause the surface wind-stress bias and thus the coastal SST bias. I test this hypothesis by replacing the orography in the HRatm setup with the low-resolution orography. Thus, the effect of the orography on the surface wind-stress and SST can be isolated in a model setup that has hardly any coastal SST bias in its non-modified form. The meridional surface wind-stress in the modified orography experiment is reduced close to the coast (figure 3.2b). Consequently, the SST bias increases in HRatmMOD compared to HRatm (figure 3.1). However, the coastal SST bias in HRatmMOD is not as large as in the experiments with low atmospheric horizontal resolution (LR, HRoc). This means that the low-resolution orography accounts for half of the SST

bias difference between HRatm and LR. The remaining difference in the wind-stress and SST is therefore caused by other better-resolved features at high atmospheric horizontal resolution.

The surface wind-stress, which is responsible for the coastal SST bias, can affect the SST in two different ways: first, it affects the coastal upwelling and thus the cooling of the surface from below and second, it affects the horizontal oceanic circulation and thus advection in the current system of the southward Angola and northward Benguela Current.

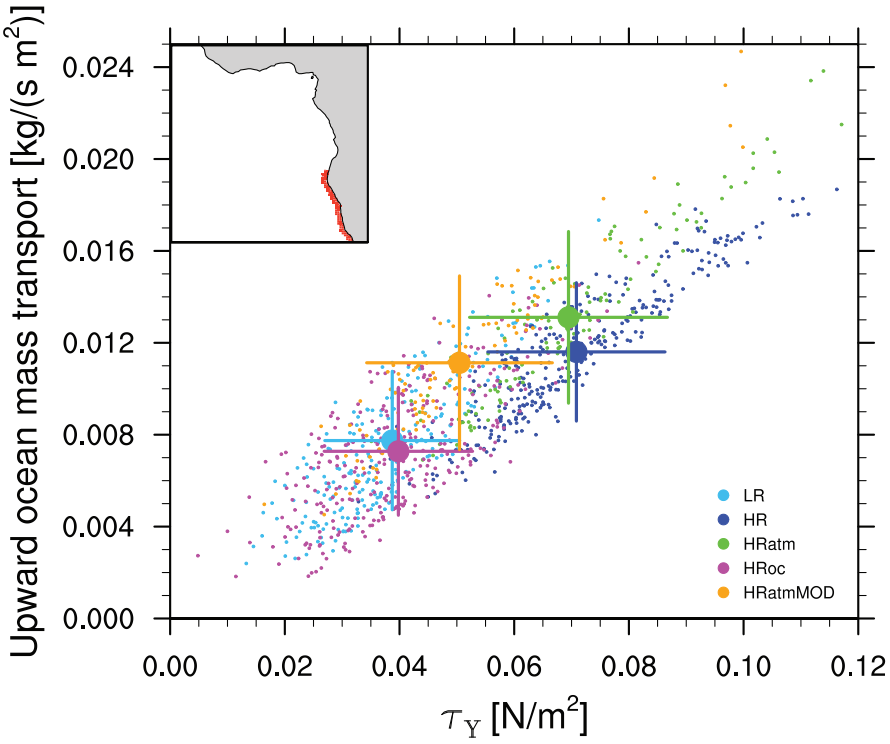

Figure 3.3: **Larger wind-stress at high atmospheric horizontal resolution causes increased upwelling:** Monthly mean meridional wind-stress and upward ocean mass transport into the uppermost layer. Spatially averaged from the coastline to 1° off the coast from 15°S to 30°S (region marked in map). The large dots indicate the mean values for each experiments, the lines mark one standard deviation in each direction.

The coastal upwelling is mainly driven by alongshore winds causing offshore Ekman transport leading to upwelling close to the coast. The meridional surface wind-stress in the model, which contributes most to the coast-parallel component of the surface wind-stress, is

well correlated with the upward ocean mass transport into the surface layer in all experiments (figure 3.3). Those experiments with low atmospheric horizontal resolution (LR, HRoc) have a weaker surface wind-stress than those experiments with high resolution in the atmosphere (HR, HRatm). The oceanic upwelling adjusts to the applied surface wind-stress, irrespective of the ocean resolution. Thus, the 0.4° ocean resolution is sufficient to represent the upwelling, provided that the corrected surface wind-stress is applied. The modified orography reduces the along-coast surface wind-stress and thus the upwelling, but not as much as in the low-resolution atmosphere experiments.

The horizontal ocean circulation is also affected by the surface wind-stress and it contributes to the wind-driven coastal SST bias. In the region where the coastal bias is most pronounced, the southward, warm Angola Current and the northward, cold Benguela Current meet to form a zonally oriented front (Shannon et al. 1987). Any shift in this current system leads to a bias in the SST due to the large horizontal temperature gradients at the oceanic front. In the low-resolution model (figure 3.4a), an unrealistically strong Angola Current is following the coast southward to 30°S. In the subsurface at 20°S, the core of the Angola Current is located close to the coast and associated with a core of warm water in the upper 30 meters (figure 3.4 d). There is no evidence for a Benguela Current at 20°S at low resolution. In contrast, at high atmospheric resolution (figure 3.4b), the Angola Current is substantially weaker and deflected to the ocean interior before reaching 20°S. Close to the coast in the south, the northward Benguela Current is co-located with colder water masses. In the subsurface near 20°S, the upward slope of the isotherms towards the coast indicates coastal upwelling (figure 3.4e). The Angola Current reaching too far south in LR is replaced by the Benguela Current in HRatm and FLX, associated with colder temperatures. In the low-resolution experiment with the adjusted surface wind-stress (FLX), both the horizontal structure of the currents (figure 3.4c) as well as the vertical structure (figure 3.4f) are very similar to the experiment with high atmospheric resolution (figure 3.4 b,e). The Benguela Current is slightly stronger in FLX than in HRatm, but still very similar to HRatm. The flux-adjusted experiment shows that improved surface wind-stress is sufficient to explain the improvements in the coastal SST bias via changes in coastal upwelling and horizontal ocean circulation.

In figure 3.5, I investigate the seasonal and annual mean rainfall bias in the HRatm model. Previous studies suggested, that the warm SST bias in the southern tropical Atlantic contributes to the misrepresentation of precipitation (Biasutti et al. 2006; Eichhorn and Bader 2016). Here, I compare the last 30 years of the HRatm simulation to observed GPCP precipitation. The HRatm simulation is only available with preindustrial forcing, therefore it has to be noted that differences between the model and observation may also arise from different

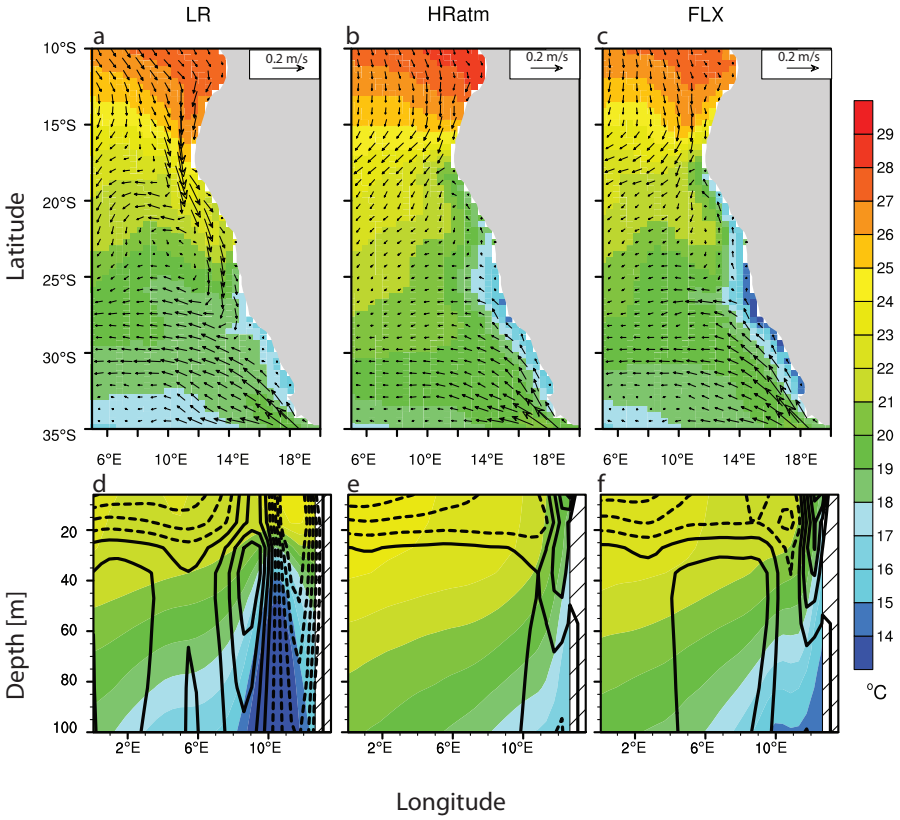

Figure 3.4: **The representation of horizontal ocean circulation mostly depends on the surface wind-stress:** (upper) maps of average velocity and temperature at 17m; (lower) sections (depth/lon) of meridional velocity and temperature at 20°S covering the realistic location of the northward Benguela Current. Velocity contours are from -0.1 m/s to 0.1 m/s with 0.02 m/s interval. Southward velocities are dashed, northward velocities are solid lines.

background conditions. Compared to the lower-resolution simulations in MPI-GE (figure 2.2), the simulated precipitation in the southeastern tropical Atlantic is improved. In the west, the bias is not improved in DJF and MAM and even turns into a wet bias in JJA and SON, which is in line with the findings of Siongco et al. (2017) that the precipitation moves towards the west when increasing the horizontal resolution in the atmospheric model ECHAM6. While the precipitation bias is improved over the southeastern tropical Atlantic, it is exacerbated over the Sahel region. The lower-resolution model simulates reasonable values for the mean precipitation over the Sahel, the high-resolution model has a strong dry bias in most seasons.

This leads to the question if the improvements in the mean state bias translate into an improved representation of rainfall variability. Figure

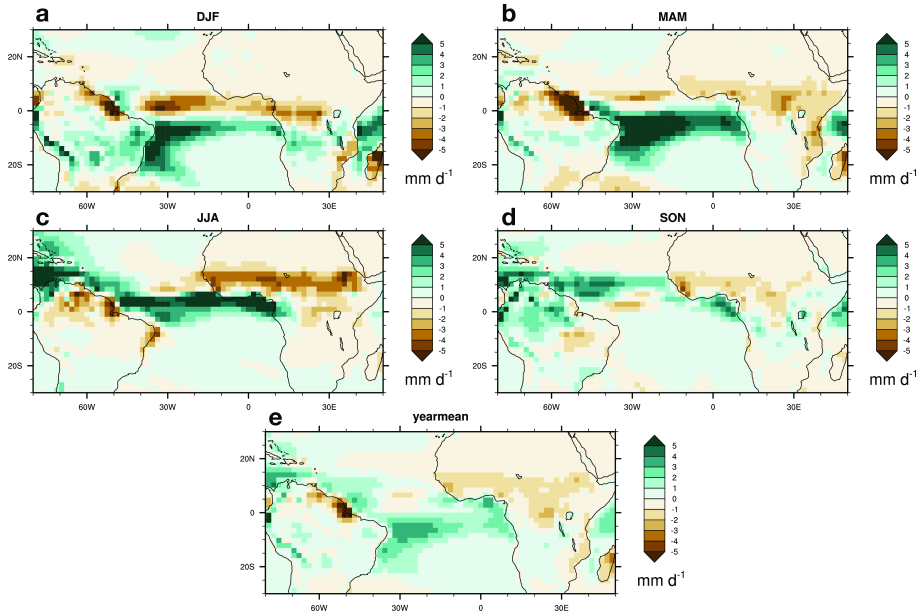

Figure 3.5: **Rainfall biases are only partly reduced for increased horizontal resolution, while some regions show even larger biases than at low resolution.** Seasonal mean rainfall biases as in figure 2.2, but for high horizontal resolution in the atmosphere.

3.6 is comparing the temporal standard deviation the the HRatm model to observations. While some improvements compared to the lower-resolution MPI-GE (figure 2.3) are visible in the southeastern tropical Atlantic, the variability bias is exacerbated in other seasons. In MAM, the positive variability bias north and south of the ITCZ is even larger than at lower resolution. Only in SON, a general improvement can be seen. At lower resolution, the variability bias is concentrated in the southeastern tropical Atlantic. When the representation of the SST is improved at high resolution, this bias in the precipitation variability is largely eliminated.

3.5 DISCUSSION AND CONCLUSIONS

The coastal SST bias in the southeastern Atlantic in the MPI-ESM can be explained by the model's deficiency in simulating surface wind-stress. At high atmospheric horizontal resolution (0.5°), the wind bias is reduced, which in turn eliminates the coastal SST bias. By adjusting the surface wind-stress along the coast in a low-resolution coupled ocean-atmosphere model, I show that the coastal SST bias is purely wind driven. The origin of the wind bias can be partly attributed to the representation of orography in the low-resolution atmospheric model. The misrepresentation of the orography at low horizontal resolution causes half of the coastal SST bias, as shown in a coupled experiment with modified orography.

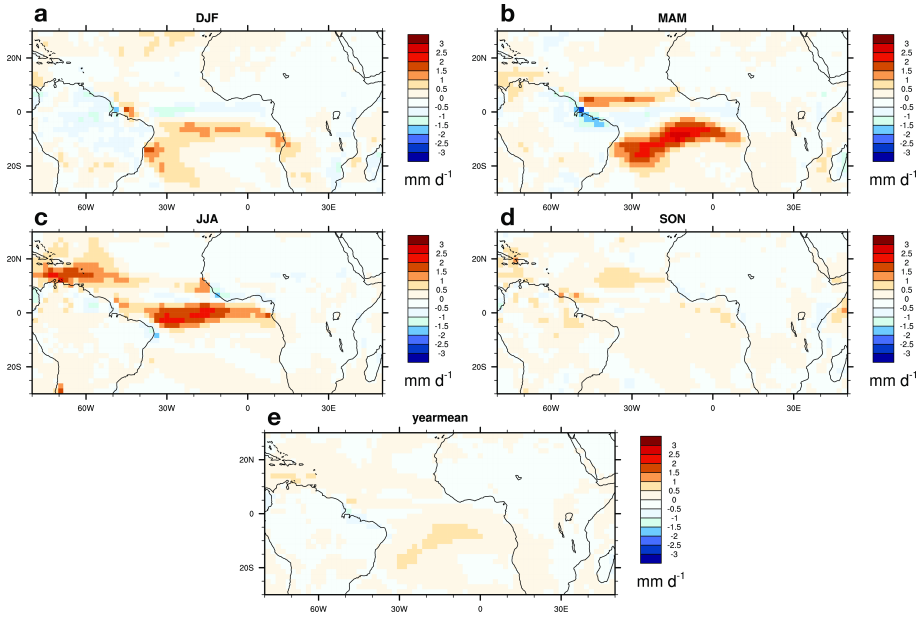

Figure 3.6: **Biases in rainfall variability are improved in SON and DJF, but worsened in MAM and JJA.** Seasonal rainfall variability biases as in figure 2.3, but for high horizontal resolution in the atmosphere.

However, the contributions of the particular deficiencies of the orography in the model cannot be quantified, which are the too weak slope at the coast, the positive elevation over the ocean, and the overshooting in the vertical in both directions. Spectral filtering can reduce the overshooting at the cost of creating even weaker slopes that would cause positive elevation over the ocean further off the coast (Navarra et al. 1994). Because the largest wind bias can be found at the coast where the slope of the orography is much weaker than observed and the elevation is still positive over the ocean, I assume these deficiencies to be the major problem, rather than the overshooting. Therefore I conclude that spectral filtering is unlikely to reduce the bias. A steeper slope can only be included by increasing the atmospheric horizontal resolution, which might not be feasible for all studies due to the increased computational cost. Furthermore, the orography accounts for approximately half of the coastal SST bias, thus I infer that the remaining coastal bias is caused by the misrepresentation of features other than the orography at low horizontal resolution.

The strength of the coastal wind jet depends not only on the steep orography but also on the shape of the coastline and the land-sea contrast close to the coast (Ranjha et al. 2016). The representation of these features is closely linked to the horizontal resolution of the atmosphere. This again suggests that the only pertinent way to reduce the coastal SST bias is to increase the atmospheric horizontal resolution, which has to be done globally for spectral atmospheric models.

Grid-point models, on the other hand, do not suffer from the same constraints on the resolution of the orography. When both types of models are run at a similar coarse horizontal resolution, grid-point models can represent a steep slope of the orography between adjacent grid cells while spectral models suffer from deficiencies such as the weaker slope and positive elevation extending over the ocean. Furthermore, grid-point models allow for a regionally increased resolution without the additional computational cost of increasing horizontal resolution globally, thus allowing a better resolution in the coastal upwelling regions in the vicinity of prominent orographic features. Despite their potential to address the typical problems of the spectral models, grid-point models suffer from similar biases in the SETA region (Grodsky et al. 2012; Patricola et al. 2012). One possible reason might be that the surface properties of ocean grid points close to the shore are affected by adjacent land grid points during the interpolation from the atmospheric to the oceanic grid. This impairs the representation of strong horizontal gradients in surface properties as pointed out for CCSM4 by Small et al. 2015. There might be further differences between spectral models like ECHAM6 (used in this study) and grid-point atmosphere models like CAM4. At 0.5° horizontal resolution, CAM4 still places the coastal wind-jet too far off the coast and some of the coastal SST bias remains while ECHAM6 places the jet closer to the coast at a similar horizontal resolution.

The large-scale SST bias is almost identical in all our simulations, neglecting the uniform offset between the untuned experiments due to slightly different global mean surface temperatures. Thus I conclude that the large-scale bias is independent of oceanic and atmospheric horizontal resolution in the examined range. Moreover, the large-scale bias does not change when the coastal bias is reduced or eliminated (HR, HRatm, FLX and HRatmMOD). From this I conclude that the coastal warm bias cannot be the root cause of the large-scale bias. If the large-scale bias was caused by the coastal bias either via ocean advection or by enhanced convection over anomalous warm SST that suppresses the formation of low level stratocumulus clouds over the SETA region, I would expect the large-scale bias to change when the coastal bias is reduced. However, it is still possible that parameterisations affecting cloud formation are not sensitive enough to the underlying SST and thus do not adjust to the colder SST at the coast. Further work concentrating on the representation of low-level clouds and their sensitivity to SST might contribute to the understanding of the large-scale SST bias. In this context, it might be necessary to increase the vertical resolution of the atmosphere as suggested by Harlaß et al. 2015.

Here, I presented evidence for local mechanisms causing the coastal SST bias. The coastal SST bias can be eliminated by adjusting the surface wind-stress applied to the ocean model. However, changes in

either the coastal upwelling or the horizontal circulation cannot be unambiguously attributed to a certain feature in the surface wind-stress based on the available experiments. This is evident from the upwelling in HRatmMOD in figure 3.3. While the meridional wind-stress is smaller than in HR and HRatm, the upwelling is stronger than one might expect from the relationship between meridional wind-stress and upwelling in the other experiments. This deviation is most likely caused by a wind-stress curl driven contribution to the upwelling. Considering the warmer SST close to the coast in HRatmMOD compared to HRatm/HR (figure 3.1a) despite the similar upwelling, I can conclude that the difference in SST between HRatmMOD and HRatm/HR is mainly caused by differences in the horizontal circulation. This highlights two limitations of my study: (1) I cannot attribute changes in upwelling and horizontal circulation to a certain feature in the wind-stress distribution and (2) the relative contributions of upwelling and horizontal circulation to the heat budget of the coastal region cannot be derived from the experiments. The bias in the coastal region is sensitive to small changes in the frontal position of the Angola-Benguela front (ABF) that is characterised by a strong meridional SST gradient. The location of the frontal position has been found to be related to the wind-stress curl (Xu et al. 2014). Furthermore, Toniazzo and Woolnough 2014 noted that small-scale features near the ABF can have a significant contribution to the heat advection, further complicating an exact quantification of the changes in heat advection due to changes in wind-stress. In addition, the heat advection is not only determined by the strength and position of the currents in the SETA region, but can also change due to differences in the properties of the advected water masses.

Previous studies suggested that remote effects from the equatorial Atlantic also contribute to the SST biases in the SETA region. Zonal wind anomalies on the equator excite downwelling Kelvin waves that suppress the coastal upwelling in the SETA region (Richter 2015; Richter et al. 2011; Toniazzo and Woolnough 2014). This mechanism might contribute to the remaining large-scale SST bias in the model used in this study because the MPI-ESM suffers from similar zonal wind and SST biases on the equator as other coupled models.

Although model biases in the precipitation already exist in uncoupled atmospheric model simulations (Biasutti et al. 2006), Eichhorn and Bader (2016) showed that part of the precipitation biases in coupled simulations can be attributed to the SST biases. I show that simulations with a reduced coastal SST bias in the southeastern tropical Atlantic show some local improvements in the representation of precipitation. However, the large scale precipitation biases over the tropical Atlantic cannot be reduced by increasing the atmospheric horizontal resolution. This implies that increasing the horizontal resolution is not sufficient

to reduce the misrepresentation of simulated precipitation, neither to improve mean state biases, nor to improve the simulated variability.

Here I show that increased horizontal atmospheric resolution can eliminate the coastal SST bias, which is purely wind-driven. The large-scale warm bias remains although the coastal bias is reduced, which implies that the patterns of the coastal and large-scale bias are superimposed and caused by different mechanisms. Because the large-scale bias is insensitive to the horizontal resolution, its cause is most likely to be found in an erroneous parameterisation. Half of the coastal SST bias can be attributed to the representation of the coastal orography at low horizontal resolution. The possibilities to modify the spectral orography are limited and not likely to reduce the problem. Half of the coastal SST bias is not caused by the orography but by other better resolved features at high horizontal resolution. Therefore, the only pertinent way to eliminate the coastal SST bias in a coupled model with a spectral atmospheric model component at this time is to increase the horizontal resolution.

4

IMPLICATIONS

In my dissertation, I argue that a large ensemble is needed to robustly detect and attribute forced changes in the mean state and internal variability in a transient climate. In recent years, several single model large ensembles of comprehensive climate models have been generated at different modelling centres. To make full use of the advantages of a single-model large ensemble, we need to rethink the analysis methods and move beyond the established procedures for analysing multi-model ensembles. In this section, I discuss how large ensembles have improved our understanding of the observed past and projected future. Large ensembles for studying climate change in comprehensive models have only become available in the last few years and have not been fully exploited at this point. Here, I want to outline open questions that can be addressed by using these large ensembles.

While large ensembles are promising to improve our understanding of specific aspects of the climate system, they are of limited use for other types of questions. I will discuss for which aspects a multi-model large ensemble should be preferred over a single model large ensemble.

In my third chapter, I demonstrate that a high-resolution model shows significant improvements compared to a low-resolution model for simulating tropical SST. While both high resolution and a large number of realisations have clear advantages, limited computing resources require to compromise between high resolution and a large ensemble size. The future modelling strategy must take both high resolution and large ensembles into account and optimise the distribution of resources to make use of the unique advantages of both.

While there is a clear need for large ensembles, there is no consensus on the minimum number of realisations or how to determine this number. The ensemble size of current large ensembles ranges from 30–100 realisations. I argue that the requirements for ensembles size can be separated into three types of application.

I will conclude this chapter by discussing how single model large ensembles, high-resolution models, and multi-model ensembles can be incorporated in a future modelling strategy that can improve our understanding of the observed climate system and its possible futures.

THE MEANING OF A SINGLE REALISATION

A single realisation of the trajectory of the climate system is the combination of the forced response to a change in the external forcing and

random internal variability. This is true for both climate models and observations. To understand why Earth's climate or a model simulation evolved the way it did, we need to disentangle the response to a change in the external forcing from the seemingly random fluctuations of internal variability. As I demonstrated, internal variability itself may change under global warming, which further complicates the separation of the forced signal from internal variability. Without any additional knowledge, this separation is not possible based on a single realisation.

The observed past trajectory is a single realisation and it is not possible to generate an additional realisation of Earth's past. Furthermore, future projections of climate models diverge and there is a need to understand if this is because the models' forced responses are different or because each model simulation is one realisation with a unique trajectory of internal variability. A commonly used approach to isolate internal variability is to apply a high-pass filter to a time series (Hawkins and Sutton 2009; Kirtman and Power 2014). However, several studies have shown that internal variability contributes substantially to variations on multi-decadal time scales (Deser et al. 2012a; Marotzke and Forster 2015). When applying a high-pass filter with a decadal cut-off, internal variability on decadal time-scales will be interpreted as part of the forced signal. A different approach to isolate internal variability is to remove a trend from the time series and interpret the residual anomaly time series as internal variability, as applied by Pendergrass et al. (2017) and others. While high-pass filtering underestimates internal variability, this approach may overestimate internal variability on multi-decadal time scales by including parts of the forced signal that deviate from the fitted trend into the estimate of internal variability.

A clean separation of internal variability from the forced signal is crucial for attribution studies that aim to explain an observed signal and attribute it to internal variability or a change in the external forcing. For Sahel rainfall, Biasutti (2013) argues that a forced response to global warming exists because idealised experiments show a response to the direct CO_2 effect on rainfall and a response of the rainfall to warming SST. To understand a signal in the observed record and attribute it to a change in forcing, it must be ensured that the signal cannot be explained by internal variability alone. If the reasoning is based on changing SST causing the signal, then the change in SST must be attributed to the change in external forcing to formulate a closed argument. While it is generally assumed that warming trends in SST can be understood as the forced response to the increasing CO_2 concentration in the atmosphere, recent studies suggest that temperature variations on decadal time scales have a substantial contribution from internal variability (Hand et al. 2018; Marotzke and Forster 2015). Several recent studies have also shown that other observed changes

that have previously been associated with a response to the external forcing can be explained by internal variability, for example for ENSO (Maher et al. 2018) or the GMST hiatus (Hedemann et al. 2017).

The findings in my thesis and the aforementioned studies emphasize that a robust quantification of internal variability is necessary to interpret a signal in the observed record or a single climate model realisation.

HOW THE USE OF LARGE ENSEMBLES CHANGES OUR UNDERSTANDING OF SINGLE REALISATIONS

Large ensembles of single climate models complement our understanding of the climate system by providing alternative realisations for the same forcing history. Every realisation can be seen as an alternative reality—a trajectory of the climate system that could have occurred instead of the observed trajectory. The assumption that a climate model realisation can be seen as a possible realisation of the Earth cannot be true unconditionally because climate models have systematic biases and fail to represent some features of the observed climate, including climate variability. Nevertheless, large ensemble are an important source of information because they allow to estimate and quantify the forced signal and internal variability.

Because all realisations in a single model large ensemble use the same model formulation and identical boundary conditions, they only differ due to internal variability. Therefore, averaging over a large number of realisations provides a robust estimate of the forced response while the residual after subtracting this forced signal only contains internal variability. Frankcombe et al. (2018) argue that the ensemble mean of a single model large ensemble is superior to alternative approaches of estimating the forced signal. It is important to note that this approach only provides a robust estimate of the forced signal in the context of a single model, not the true forced signal.

As I showed in the first and second chapter of this thesis, a large ensemble allows a robust detection of forced changes in the mean state and even forced changes in variability. While it may be possible to clearly identify a forced change in a large ensemble, the implications of such a finding for a single realisation are still challenging to formulate. Here, I summarise the implications for interpreting either a single realisation or a large ensemble.

1. **If a forced change exists in a large ensemble, this change is not guaranteed to show up in a single realisation.** This implies that projected changes may not clearly show up in the real world , even if the simulated forced change is a correct representation of the forced change in the real climate system. For example, a projected warming for a specific emission pathway may be stronger or weaker than the projected forced change because

of internal variability as illustrated in figure 2.8. This poses a communication challenge on climate science that must be faced now, rather than in a few decades when the observed record shows a different trajectory than what a current projection indicates.

2. **A forced change that was hypothesised based on observations should be verified with a large ensemble.** This can be separated into two types of questions: does any forced change exist vs. can the forced change be quantified. For proving the existence of a forced change, a significant difference from a reference must be shown. Quantifying the signal of the forced change is more difficult because a single realisation contains insufficient information to quantify the forced signal and internal variability. Therefore, a large ensemble is required to provide an additional estimate of the forced signal and verify the hypothesis based on a single realisation. However, the failure of a model to precisely reproduce the observed trajectory does not imply that the model is wrong as long as the observations lie within the range of realisations produced by the model.

INTERNAL VARIABILITY IN THE PAST AND FUTURE

To better understand the observed trajectory, it would be desirable to produce additional realisations of the the real climate system or a very close approximation. McKinnon et al. (2017) have randomly resampled the observational record and combined it with the forced signal of a large ensemble to generate alternative realisations of the past winter temperatures over North America. They argue that the variability in the model is too large and therefore use observed variability. The assumptions in their study are that (1) the model accurately represents the forced signal and (2) the observed record represents the full phase space of internal variability. In other words, they assume that all possible states of the system have already been observed and no extreme state outside of the distribution of sampled events can occur. For a sufficiently long record or sufficiently small variability, this assumption may hold. In their study, they use North American winter temperatures and argue that the variability for this quantity is low and therefore well represented by sampling the observed record.

In a second study, McKinnon and Deser (2018) extended this approach to temperature, precipitation and sea level pressure globally. In particular for precipitation, the observed record in some regions may not be long enough to represent the full distribution of possible events. In this case, a larger range of variability in the model does not necessarily indicate an overestimation of variability in the model, but could point to an undersampling of internal variability in the observations. If internal variability changes over time, sampling from the observations

alone is insufficient to estimate the full range of alternative realisation of the Earth system that could have occurred. For instance, my results from the second chapter indicate a change in the variability of Sahel rainfall in the MPI-GE historical simulations that potentially continues under global warming.

The central issue, both for the past and future, is to estimate the characteristics of unobserved events. This mainly applies to extreme events that are so rare that they did not occur in the observed record, but could have occurred in an alternative realisation. The approach by McKinnon et al. (2017) does not take this into account. However, it is precisely when a previously unobserved event occurs that an explanation for the origin of this event is asked for. When assuming that the past represents all possible states of internal variability, then a previously unobserved event can only be attributed to the change in external forcing. Examples like the global warming hiatus discussion emphasise that it is necessary to take unobserved events into account when trying to understand new observations that do not fall within the range of previous expectations. While large ensembles can be used to provide information about unobserved events, validating these simulated events is difficult. It may well be that a model generates events that could never occur in reality, or that a model is not able to resolve extreme events that could occur.

For future projections, communicating internal variability as an irreducible uncertainty is crucial (Lehmann and Rillig 2014). Even with perfect knowledge of the forced signal and the characteristics of internal variability, a range of future trajectories is possible. Therefore, the change of a quantity like GMST for a specific year in the future compared to present day can never be given without uncertainty, even if the the external forcing and all processes in the climate system are known and included in a model.

HOW LARGE DOES AN ENSEMBLE NEED TO BE?

My thesis and several other studies emphasise the need for large ensembles to understand the characteristics of a model and the difference to observations or other models. While many studies agree that there is a need for large ensembles, there is no general agreement for the required size of an ensemble. Current large ensembles have widely different ensemble sizes: the GFDL ESM2M has 30 realisations (Rodgers et al. 2015), the CESM-LE produced 40 realisations (Kay et al. 2015), the CanESM2 large ensemble 50 realisations (Fyfe et al. 2017), and the MPI-GE 100 realisations (Bittner et al. 2016). Even earlier, Zelle et al. (2005) used a 62-member ensemble to investigate ENSO under global warming. Although their model uses a lower resolution than other large ensembles, it can be considered a state-of-the-art climate model for 2005. For most of these large ensembles, the availability of

computational and storage resources may have determined the ensemble size. However, the wide range of ensemble sizes clearly shows that there is no consensus on the required number of realisations for a large ensemble to study the recent past and future scenarios. For future modelling efforts, it is necessary to determine how many realisations are required in order to use available computing resources efficiently.

Previous studies have estimated how many ensemble members are necessary to detect a certain signal. Deser et al. (2012b) and Li and Ilyina (2018) are using a signal-to-noise approach to determine how many realisations are necessary to robustly detect a specific signal. This approach first requires a signal to be present, and additionally more ensemble members than actually needed to be able to artificially reduce the ensemble size until the signal is no longer detectable. While their results provide a retrospective minimum number of realisations for a specific application, the minimum ensemble size is determined after the simulations have been carried out and it only works if a signal exists. In Maher et al. (2018), we tried to estimate the number of ensemble members required to adequately sample ENSO variability. We used the variability based on all 100 realisations to determine how many realisations are necessary to estimate this value within a certain error range. While this approach does not require a signal in the sense of a forced change, it still requires the availability of a large number of realisations to confirm that a smaller ensemble size would have been sufficient.

Before being able to determine an adequate ensemble size, the requirements need to be clearly specified. I propose three requirements, each tailored to a specific type of problem, and each calling for a different minimum ensemble size. The requirements are sorted from the lowest requirements on ensemble size to the highest.

1. identify the forced signal
2. adequate sampling of the phase space of internal variability
3. robust detection of a forced change in internal variability

In a large ensemble, the ensemble mean represent the forced signal. This is true when internal variability cancels out when averaging the ensemble members. This first requirement asks for the smallest ensemble size, because the only condition is that different realisations are randomly distributed. Even for an ensemble size of less that ten members, this requirement seems to be fulfilled for several quantities (Olonscheck and Notz 2017).

The second requirement calls for a larger number of realisations. To sample internal variability, the ensemble members need to represent the entire phase space of possible states of internal variability. If internal variability is changing over time, internal variability must be adequately represented when sampling the ensemble dimension for

any given state of the system, for example a single year or pentade. If internal variability is not changing over time and stationarity can be assumed, sampling over the ensemble dimension can be supplemented by sampling over time. This allows to generate a large sample even with a small number of realisations. While Bengtsson and Hodges (2018) show that stationarity can be assumed for many variables over the historical period, the results from my thesis indicate that this may not be true for some regional quantities in the historical period, and that stationarity may break for a large change in the external forcing.

The third requirement asks for the largest ensemble size. To detect a change in internal variability, condition (2) must be fulfilled and the ensemble must sample internal variability adequately for two different states of the system that should be compared. Additionally, the required ensemble size for (3) depends on the magnitude of the change to be detected. For a large change in variability, a reasonably good estimate of variability for the two states is needed. However, for a small change, the accuracy for the estimate of internal variability must be sufficient to detect the small change.

The first two requirements can be estimated by evaluating the saturation, i.e. when adding more realisations does not change the estimate of the forced signal (1) or internal variability (2) any more, the ensemble size is sufficient. This means that it is technically possible to generate a large ensemble for a specific question and stop adding realisations when saturation has been reached.

The third requirement must fulfil requirements (1) and (2), but in addition, the magnitude of the signal to be detected must be defined a priori. This decision should be based on the relevance of the magnitude of the signal. A small change in variability may be detectable in a large ensemble, but the implications of this change for the occurrence of anomalous or extreme events might be negligible. Limiting the ensemble size to detect changes only above a certain threshold might not reveal smaller forced changes, but will ensure an efficient use of resources.

ENSEMBLE SIZE, RESOLUTION, OR MODEL VARIETY?

For the applications presented in the second chapter of this thesis, a large ensemble size is necessary to cleanly separate internal variability from the forced signal and investigate internal variability itself. Running up to 100 realisations with a single model is only feasible at low resolution. On the other hand, the results from my third chapter indicate that only high atmospheric resolution can reduce the SST biases in the eastern boundary upwelling regions and provide a more realistic simulation of SST in these regions. With the currently available computing resources, only a small number of realisations can be performed at this high resolution. This leads to two competing

interests: on one hand, we want to increase the resolution to resolve more processes and improve the realism of simulations, on the other hand we want to have a large number of realisations to differentiate between forcing and internal variability. While there are arguments for either using high resolution or a large ensemble, the solution cannot be a choice between two options. Each one is providing context for the other, which is not available when only one option is chosen.

While higher resolutions may improve the representation of important processes, a large ensemble can provide the context to decide which improvements seen in a single realisation of a high-resolution model can be attributed to difference in resolution rather than to internal variability.

There is a strong public interest in understanding localised extreme events and their change under global warming. While the features of these small-scale, short lived events may only be well-represented in a high-resolution model, rare events can only be captured in a large sample. This is related to the problem of simulating unobserved events. The challenge is to find a way to link extreme events, that are represented in a high-resolution model, to large-scale features that are captured by a lower-resolution large ensemble.

The third aspect to this, which I have not discussed up to this point, is model variety. In the context of a single model, a large ensemble provides a robust estimate of the forced signal and internal variability. But all realisations in a single model large ensemble share the same systematic biases and might deviate from the truth, while a different model may be much closer to the truth.

Hawkins and Sutton (2009) mentioned three types of uncertainty: internal variability, model uncertainty and scenario uncertainty. In multi-model large ensembles, such as CMIP5, model uncertainty and internal variability cannot be separated because the difference between any two realisations from different models may either arise from model differences, or internal variability. A single-model large ensemble does not provide information about the model uncertainty by itself. By comparing single-model large ensembles from different models, internal variability and model differences can be cleanly separated. A comparison of the existing large ensembles will help to estimate how large internal variability and model differences are.

5 CONCLUSIONS

The aim of my thesis was to robustly quantify internal variability and detect forced changes to it. For this purpose, I developed a new analysis framework that makes use of a large ensemble of a comprehensive climate model to robustly detect changes in internal variability and attribute them to a change in the external forcing. I apply this framework to forced changes of rainfall variability in the Sahel and the Atlantic oceanic ITCZ. Finally, I address the model biases in the tropical Atlantic region and show how increasing the atmospheric horizontal resolution can reduce the coastal SST bias. I want to conclude by summarising the main findings of each chapter and discuss my results in the context of the challenges discussed in chapter 4.

Chapter 1

- In a transient climate, internal variability is only well-defined in the ensemble dimension.
- Changes in internal variability can only be robustly detected and attributed to a change in the external forcing in a large ensemble.
- With the non-parametric approach I applied, changes in variability can be quantified and tested for robustness.

Internal variability is often considered as noise concealing a forced signal, but being time-independent itself. I move beyond this perspective and show that internal variability itself can change in response to a change in the external forcing. The analysis framework I present exploits the potential of large ensemble simulations to formulate a new interpretation of the forced response both in the variability and mean state of the climate system to a change in the external forcing.

Chapter 2

- Internal variability of Sahel rainfall in the monsoon season increases in response to strong global warming.
- The observed drying trend in Sahel rainfall in the past century cannot be attributed to a change in the external forcing in the MPI-GE.
- Regional changes in the variability do not scale with the mean state change and may even have the opposite sign of change.

- Changes in variability are not necessarily symmetric about the centre of the distribution. If the distribution or its changes are not symmetric, the variance or standard deviation is not a good descriptor of internal variability.

My results show that quantifying internal variability is essential to understand if an observed trend can be attributed to a change in the external forcing. While the observed drying in the Sahel has been partly attributed to changes in the Atlantic Multidecadal Variability, this does not constitute an externally forced trend in Sahel rainfall unless the AMV can be attributed to a change in the external forcing. This emphasises the need for a careful separation of remote forcing and external forcing when attributing observed changes.

The common assumption that variability scales with a change in the mean state holds for projected Sahel rainfall. However, I show that this relationship does not hold in general: rainfall variability on the southern flank of the ITCZ increases while the mean state is decreasing, which can be explained by a shift in the circulation.

Chapter 3

- The coastal warm bias in the south-eastern tropical Atlantic can be reduced by 50% by increasing the horizontal resolution in the atmosphere. Due to a better resolved orography the simulation of the along-coast wind stress is improved. This in turn improves the representation of upwelling and the horizontal current structure, reducing the coastal SST bias.

- In contrast to previous belief, the coastal SST bias in the south-eastern tropical Atlantic is not directly causing the large scale warm bias further off the coast. Reducing the coastal bias does not resolve the large scale bias.

- Reducing the coastal SST bias only slightly improves the representation of rainfall over the tropical Atlantic.

Large scale warm biases in the tropical Atlantic are present in most state-of-the-art coupled climate models. I can show that increasing the atmospheric horizontal resolution reduces the coastal warm bias, while the resolution of the ocean model does not improve the simulated SST. In contrast to previous hypotheses, my work shows that the coastal bias is not directly causing the large-scale warm bias. Even though increasing the model resolution reduces some of the model biases, it is not sufficient to resolve all issues.

BIBLIOGRAPHY

Adler, R. et al. (2016). "Global Precipitation Climatology Project (GPCP) Climate Data Record (CDR), Version 2.3 (Monthly). National Centers for Environmental Information." DOI: 10.7289/V56971M6.

Bengtsson, L and K. I. Hodges (2018). "Can an ensemble climate simulation be used to separate climate change signals from internal unforced variability?" *Climate Dynamics* 49.6249, pp. 1–21. DOI: 10.1007/s00382-018-4343-8.

Biasutti, M (2013). "Forced Sahel rainfall trends in the CMIP5 archive." *Journal of Geophysical Research-Atmospheres* 118.4, pp. 1613–1623. DOI: 10.1002/jgrd.50206.

Biasutti, M, I. M. Held, A. H. Sobel, and A Giannini (2008). "SST Forcings and Sahel Rainfall Variability in Simulations of the Twentieth and Twenty-First Centuries." *Journal of Climate* 21.14, pp. 3471–3486. DOI: 10.1175/2007JCLI1896.1.

Biasutti, M. and A. Giannini (2006). "Robust Sahel drying in response to late 20th century forcings." *Geophysical Research Letters* 33.11, pp. 2616–4. DOI: 10.1029/2006GL026067.

Biasutti, M., A. H. Sobel, and Y Kushnir (2006). "AGCM precipitation biases in the tropical Atlantic." English. *Journal of Climate* 19.6, pp. 935–958. DOI: 10.1175/JCLI3673.1.

Bittner, M., H. Schmidt, C. Timmreck, and F. Sienz (2016). "Using a large ensemble of simulations to assess the Northern Hemisphere stratospheric dynamical response to tropical volcanic eruptions and its uncertainty." *Geophysical Research Letters* 43.17, pp. 9324–9332. DOI: 10.1002/2016GL070587.

Byrne, M. P. and P. A. O'Gorman (2015). "The response of precipitation minus evapotranspiration to climate warming: Why the "Wet-get-wetter, dry-get-drier" scaling does not hold over land." English. *Journal of Climate* 28.20, pp. 8078–8092. DOI: 10.1175/JCLI-D-15-0369.1.

Byrne, M. P. and T. Schneider (2016). "Narrowing of the ITCZ in a warming climate: Physical mechanisms." *Geophysical Research Letters* 43.21, pp. 11350–11357. DOI: 10.1002/2016GL070396.

Deser, C., R. Knutti, S. Solomon, and A. S. Phillips (2012a). "Communication of the role of natural variability in future North American climate." *Nature Climate Change* 2.11, pp. 775–779. DOI: 10.1038/nclimate1562.

Deser, C., A. Phillips, V. Bourdette, and H. Teng (2012b). "Uncertainty in climate change projections: the role of internal variability." *Climate Dynamics* 38.3-4, pp. 527–546. DOI: 10.1007/s00382-010-0977-x.

Ding, H, R. J. Greatbatch, and M Latif (2015). "The impact of sea surface temperature bias on equatorial Atlantic interannual variability in partially coupled model experiments." *Geophysical Research Letters* 42.13, pp. 5540–5546. DOI: 10.1002/2015GL064799.

Doi, T., G. A. Vecchi, A. J. Rosati, and T. L. Delworth (2012). "Biases in the Atlantic ITCZ in Seasonal–Interannual Variations for a Coarse- and a High-Resolution Coupled Climate Model." *Journal of Climate* 25.16, pp. 5494–5511. DOI: 10.1175/JCLI-D-11-00360.1.

Dunning, C. M., E. Black, and R. P. Allan (2018). "Later wet seasons with more intense rainfall over Africa under future climate change." *Journal of Climate*, JCLI–D–18–0102.1–49. DOI: 10.1175/JCLI-D-18-0102.1.

Eichhorn, A. and J. Bader (2016). "Impact of tropical Atlantic sea-surface temperature biases on the simulated atmospheric circulation and precipitation over the Atlantic region: An ECHAM6 model study." *Climate Dynamics* 4.6, pp. 1–15. DOI: 10.1007/s00382-016-3415-x.

Frankcombe, L. M., M. H. England, J. B. Kajtar, M. E. Mann, and B. A. Steinman (2018). "On the Choice of Ensemble Mean for Estimating the Forced Signal in the Presence of Internal Variability." *Journal of Climate* 31.14, pp. 5681–5693. DOI: 10.1175/JCLI-D-17-0662.1.

Fyfe, J. C. et al. (2017). "Large near-Term projected snowpack loss over the western United States." *Nature Communications* 8, p. 14996. DOI: 10.1038/ncomms14996.

Gent, P. R., S. G. Yeager, R. B. Neale, S. Levis, and D. A. Bailey (2010). "Improvements in a half degree atmosphere/land version of the CCSM." *Climate Dynamics* 34.6, pp. 819–833. DOI: 10.1007/s00382-009-0614-8.

Giannini, A, R Saravanan, and P Chang (2003). "Oceanic Forcing of Sahel Rainfall on Interannual to Interdecadal Time Scales." *Science* 302.5647, pp. 1027–1030. DOI: 10.1126/science.1089357.

Giorgetta, M. A. et al. (2013). "Climate and carbon cycle changes from 1850 to 2100 in MPI-ESM simulations for the Coupled Model Intercomparison Project phase 5." *Journal of Advances in Modeling Earth Systems* 5.3, pp. 572–597. DOI: 10.1002/jame.20038.

Greve, P., B. Orlowsky, B. Mueller, J. Sheffield, M. Reichstein, and S. I. Seneviratne (2014). "Global assessment of trends in wetting and drying over land." *Nature Geoscience* 7.10, pp. 716–721. DOI: 10.1038/ngeo2247.

Griffies, S. M. and K Bryan (1997). "Predictability of North Atlantic multidecadal climate variability." *Science* 275.5297, pp. 181–184. DOI: 10.1126/science.275.5297.181.

Griffies, S. M. et al. (2011). "The GFDL CM3 Coupled Climate Model: Characteristics of the Ocean and Sea Ice Simulations." *Journal of Climate* 24.13, pp. 3520–3544. DOI: 10.1175/2011JCLI3964.1.

Grodsky, S. A., J. A. Carton, S. Nigam, and Y. M. Okumura (2012). "Tropical Atlantic Biases in CCSM4." *Journal of Climate* 25.11, pp. 3684–3701. DOI: 10.1175/JCLI-D-11-00315.1.

Hand, R., J. Bader, D. Matei, R. Ghosh, and J. Jungclaus (2018). "Substantial contributions of the ocean circulation to the Atlantic Multidecadal Variability in the MPI Grand Ensemble under changing external forcing." *submitted to Journal of Climate*.

Harlaß, J., M. Latif, and W. Park (2015). "Improving climate model simulation of tropical Atlantic sea surface temperature: The importance of enhanced vertical atmosphere model resolution." *Geophysical Research Letters* 42.7, pp. 2401–2408. DOI: 10.1002/2015GL063310.

Harris, I, P. D. Jones, T. J. Osborn, and D. H. Lister (2013). "Updated high-resolution grids of monthly climatic observations - the CRU TS3.10 Dataset." *International Journal of Climatology* 34.3, pp. 623–642. DOI: 10.1002/joc.3711.

Hawkins, E. and R. Sutton (2009). "The Potential to Narrow Uncertainty in Regional Climate Predictions." *Bulletin of the American Meteorological Society* 90.8, pp. 1095–1107. DOI: 10.1175/2009BAMS2607.1.

– (2010). "The potential to narrow uncertainty in projections of regional precipitation change." *Climate Dynamics* 37.1-2, pp. 407–418. DOI: 10.1007/s00382-010-0810-6.

– (2012). "Time of emergence of climate signals." *Geophysical Research Letters* 39.1. DOI: 10.1029/2011GL050087.

Hedemann, C., T. Mauritsen, J. Jungclaus, and J. Marotzke (2017). "The subtle origins of surface-warming hiatuses." *Nature Climate Change* 7.5, pp. 336–+. DOI: 10.1038/NCLIMATE3274.

Held, I. M. and B. J. Soden (2006). "Robust responses of the hydrological cycle to global warming." *Journal of Climate* 19.21, pp. 5686–5699. DOI: 10.1175/JCLI3990.1.

Huang, P., S.-P. Xie, K. Hu, G. Huang, and R. Huang (2013). "Patterns of the seasonal response of tropical rainfall to global warming." *Nature Geoscience* 6.5, pp. 357–361. DOI: 10.1038/ngeo1792.

Huntingford, C., P. D. Jones, V. N. Livina, T. M. Lenton, and P. M. Cox (2017). "No increase in global temperature variability despite changing regional patterns." *Scientific Reports* 500.7462, pp. 327–330. DOI: 10.1038/nature12310.

Ilyina, T., K. D. Six, J. Segschneider, E. Maier-Reimer, H. Li, and I. Núñez-Riboni (2013). "Global ocean biogeochemistry model HAMOCC: Model architecture and performance as component of the MPI-Earth system model in different CMIP5 experimental realizations." *Journal of Advances in Modeling Earth Systems* 5.2, pp. 287–315. DOI: 10.1029/2012MS000178.

Jungclaus, J. H., N Fischer, H Haak, K Lohmann, J Marotzke, D Matei, U Mikolajewicz, D Notz, and J. S. von Storch (2013). "Characteristics of the ocean simulations in the Max Planck Institute Ocean Model (MPIOM) the ocean component of the MPI-Earth system model."

Journal of Advances in Modeling Earth Systems 5.2, pp. 422–446. DOI: 10.1002/jame.20023.

Kay, J. E. et al. (2015). "The Community Earth System Model (CESM) Large Ensemble Project: A Community Resource for Studying Climate Change in the Presence of Internal Climate Variability." *Bulletin of the American Meteorological Society* 96.8, pp. 1333–1349. DOI: 10.1175/BAMS-D-13-00255.1.

Kirtman, B. P. and S. Power (2014). *Climate Change 2013 – The Physical Science Basis: Working Group I Contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge: Cambridge University Press.

Knight, J. R., C. K. Folland, and A. A. Scaife (2006). "Climate impacts of the Atlantic Multidecadal Oscillation." English. *Geophysical Research Letters* 33.17, pp. 686–4. DOI: 10.1029/2006GL026242.

Large, W. and G. Danabasoglu (2006). "Attribution and impacts of upper-ocean biases in CCSM3." *Journal of Climate* 19.11, pp. 2325–2346. DOI: 10.1175/JCLI3740.1.

Lehmann, J. and M. Rillig (2014). "Distinguishing variability from uncertainty." *Nature Communications* 4.3, pp. 153–153. DOI: 10.1038/nclimate2133.

Lehner, F., C. Deser, and L. Terray (2017). "Toward a New Estimate of "Time of Emergence" of Anthropogenic Warming: Insights from Dynamical Adjustment and a Large Initial-Condition Model Ensemble." *Journal of Climate* 30.19, pp. 7739–7756. DOI: 10.1175/JCLI-D-16-0792.1.

Lehner, F., C. Deser, and B. M. Sanderson (2018). "Future risk of record-breaking summer temperatures and its mitigation." *Climatic Change* 146.3-4, pp. 1–13. DOI: 10.1007/s10584-016-1616-2.

Lewis, J. M. (2005). "Roots of ensemble forecasting." English. *Monthly weather review* 133.7, pp. 1865–1885. DOI: 10.1175/MWR2949.1.

Li, H. and T. Ilyina (2018). "Current and Future Decadal Trends in the Oceanic Carbon Uptake Are Dominated by Internal Variability." *Geophysical Research Letters* 45.2, pp. 916–925. DOI: 10.1002/2017GL075370.

Maher, N, D Matei, S Milinski, and J Marotzke (2018). "ENSO change in climate projections: forced response or internal variability?" *Geophysical Research Letters*, pp. 1–27. DOI: 10.1029/2018GL079764.

Marotzke, J. and P. M. Forster (2015). "Forcing, feedback and internal variability in global temperature trends." *Nature* 517.7536, pp. 565–570. DOI: 10.1038/nature14117.

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 Modelling* 5.2, pp. 91–127. DOI: 10.1016/S1463-5003(02)00015-X.

Martin, E. R. and C. D. Thorncroft (2013). "The impact of the AMO on the West African monsoon annual cycle." *Quarterly Journal of*

the Royal Meteorological Society 140.678, pp. 31–46. DOI: 10.1002/qj.2107.

Martin, E. R., C. Thorncroft, and B. B. B. Booth (2014). "The Multidecadal Atlantic SST—Sahel Rainfall Teleconnection in CMIP5 Simulations." *Journal of Climate* 27.2, pp. 784–806. DOI: 10.1175/JCLI-D-13-00242.1.

McKinnon, K. A. and C. Deser (2018). "Internal variability and regional climate trends in an Observational Large Ensemble." *Journal of Climate*, JCLI–D–17–0901.1–52. DOI: 10.1175/JCLI-D-17-0901.1.

McKinnon, K. A., A. Poppick, E. Dunn-Sigouin, and C. Deser (2017). "An "Observational Large Ensemble" to Compare Observed and Modeled Temperature Trend Uncertainty due to Internal Variability." *Journal of Climate* 30.19, pp. 7585–7598. DOI: 10.1175/JCLI-D-16-0905.1.

Milinski, S, J Bader, H Haak, A. C. Siongco, and J. H. Jungclaus (2016). "High atmospheric horizontal resolution eliminates the wind-driven coastal warm bias in the southeastern tropical Atlantic." *Geophysical Research Letters* 43.19, pp. 10455–10462. DOI: 10.1002/(ISSN)1944-8007.

Monerie, P.-A., E. Sanchez-Gomez, and J. Boé (2016). "On the range of future Sahel precipitation projections and the selection of a subsample of CMIP5 models for impact studies." English. *Climate Dynamics* 48.7, pp. 2751–2770. DOI: 10.1007/s00382-016-3236-y.

Monerie, P.-A., E. Sanchez-Gomez, B. Pohl, J. Robson, and B. Dong (2017). "Impact of internal variability on projections of Sahel precipitation change." *Environmental Research Letters* 12.11, pp. 114003–13. DOI: 10.1088/1748-9326/aa8cda.

Navarra, A, W. F. Stern, and K Miyakoda (1994). "Reduction of the Gibbs oscillation in spectral model simulations." *Journal of Climate* 7.8, pp. 1169–1183. DOI: 10.1175/1520-0442(1994)007<1169:ROTGOI>2.0.CO;2.

Nicholson, S. E. (1980). "The Nature of Rainfall Fluctuations in Subtropical West Africa." *Monthly weather review* 108.4, pp. 473–487. DOI: 10.1175/1520-0493(1980)108<0473:TNORFI>2.0.CO;2.

Nicholson, S. E., B Some, and B Kone (2000). "An analysis of recent rainfall conditions in West Africa, including the rainy seasons of the 1997 El Nino and the 1998 La Nina years." *Journal of Climate* 13.14, pp. 2628–2640. DOI: 10.1175/1520-0442(2000)013<2628:AAORRC>2.0.CO;2.

Nicholson, S. E. (2010). "A low-level jet along the Benguela coast, an integral part of the Benguela current ecosystem." *Climatic Change* 99.3-4, pp. 613–624. DOI: 10.1007/s10584-009-9678-z.

– (2013). "The West African Sahel: A Review of Recent Studies on the Rainfall Regime and Its Interannual Variability." *ISRN Meteorology* 2013.4, pp. 1–32. DOI: 10.1155/2013/453521.

Olonscheck, D. and D. Notz (2017). "Consistently Estimating Internal Climate Variability from Climate Model Simulations." English. *Journal of Climate* 30.23, pp. 9555–9573. DOI: 10.1175/JCLI-D-16-0428.1.

Park, J.-y., J. Bader, and D. Matei (2016). "Anthropogenic Mediterranean warming essential driver for present and future Sahel rainfall." *Nature Climate Change* 6.10, pp. 941–945. DOI: 10.1038/nclimate3065.

Patricola, C. M., M. Li, Z. Xu, P. Chang, R Saravanan, and J.-S. Hsieh (2012). "An investigation of tropical Atlantic bias in a high-resolution coupled regional climate model." *Climate Dynamics* 39.9-10, pp. 2443–2463. DOI: 10.1007/s00382-012-1320-5.

Pendergrass, A. G., R. Knutti, F. Lehner, C. Deser, and B. M. Sanderson (2017). "Precipitation variability increases in a warmer climate." *Scientific Reports* 7.1, pp. 1–9. DOI: 10.1038/s41598-017-17966-y.

Ranjha, R., M. Tjernström, G. Svensson, and A. Semedo (2016). "Modelling coastal low-level wind-jets: does horizontal resolution matter?" *Meteorology and Atmospheric Physics* 128.2, pp. 263–278. DOI: 10.1007/s00703-015-0413-1.

Rayner, N. A., D. E. Parker, E. B. Horton, C. K. Folland, L. V. Alexander, D. P. Rowell, E. C. Kent, and A Kaplan (2003). "Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century." *Journal of Geophysical Research: Atmospheres (1984–2012)* 108.D14, p. 4407. DOI: 10.1029/2002JD002670.

Reick, C. H., T Raddatz, V Brovkin, and V Gayler (2013). "Representation of natural and anthropogenic land cover change in MPI-ESM." *Journal of Advances in Modeling Earth Systems* 5.3, pp. 459–482. DOI: 10.1002/jame.20022.

Richter, I, S. P. Xie, S. K. Behera, T Doi, and Y Masumoto (2014). "Equatorial Atlantic variability and its relation to mean state biases in CMIP5." *Climate Dynamics* 42, pp. 171–188. DOI: 10.1007/s00382-012-1624-5.

Richter, I. (2015). "Climate model biases in the eastern tropical oceans: causes, impacts and ways forward." *Wiley Interdisciplinary Reviews: Climate Change* 6.3, pp. 345–358. DOI: 10.1002/wcc.338.

Richter, I., S.-P. Xie, A. T. Wittenberg, and Y. Masumoto (2011). "Tropical Atlantic biases and their relation to surface wind stress and terrestrial precipitation." *Climate Dynamics* 38.5-6, pp. 985–1001. DOI: 10.1007/s00382-011-1038-9.

Rodgers, K. B., J Lin, and T. L. Frölicher (2015). "Emergence of multiple ocean ecosystem drivers in a large ensemble suite with an Earth system model." *Biogeosciences* 12.11, pp. 3301–3320. DOI: 10.5194/bg-12-3301-2015.

Rowell, D. P., C. A. Senior, M. Vellinga, and R. J. Graham (2015). "Can climate projection uncertainty be constrained over Africa using metrics of contemporary performance?" *Climatic Change* 134.4, pp. 621–633. DOI: 10.1007/s10584-015-1554-4.

Shannon, L. V., J. J. Agenbag, and M. Buys (1987). "Large and mesoscale features of the Angola-Benguela front." *South African Journal of Marine Sciences* 5.1, pp. 11–34. DOI: 10.2989/025776187784522261.

Sheen, K. L., D. M. Smith, N. J. Dunstone, R Eade, D. P. Rowell, and M Vellinga (2017). "Skilful prediction of Sahel summer rainfall on inter-annual and multi-year timescales." *Nature Communications* 8. DOI: 10.1038/ncomms14966.

Siongco, A. C., C. Hohenegger, and B. Stevens (2015). "The Atlantic ITCZ bias in CMIP5 models." *Climate Dynamics* 45.5-6, pp. 1–12. DOI: 10.1007/s00382-014-2366-3.

– (2017). "Sensitivity of the summertime tropical Atlantic precipitation distribution to convective parameterization and model resolution in ECHAM6." *Journal of Geophysical Research-Atmospheres* 122.5, pp. 2579–2594. DOI: 10.1002/2016JD026093.

Small, R. J. et al. (2014). "A new synoptic scale resolving global climate simulation using the Community Earth System Model." *Journal of Advances in Modeling Earth Systems* 6.4, pp. 1065–1094. DOI: 10.1002/2014MS000363.

Small, R. J., E. Curchitser, K. Hedstrom, B. Kauffman, and W. G. Large (2015). "The Benguela Upwelling System: Quantifying the Sensitivity to Resolution and Coastal Wind Representation in a Global Climate Model." *Journal of Climate* 28.23, pp. 9409–9432. DOI: 10.1175/JCLI-D-15-0192.1.

Stephens, G. L., T. L'Ecuyer, R. Forbes, A. Gettelmen, J.-C. Golaz, A. Bodas-Salcedo, K. Suzuki, P. Gabriel, and J. Haynes (2010). "Dreary state of precipitation in global models." *J. Geophys. Res* 115.D24, pp. 1147–14. DOI: 10.1029/2010JD014532.

Stephens, M. A. (1970). "Use of the Kolmogorov-Smirnov, Cramér-Von Mises and related statistics without extensive tables." *Journal of the Royal Statistical Society Series B* 32.1, pp. 115–122.

Stevens, B. (2015). "Rethinking the Lower Bound on Aerosol Radiative Forcing." *Journal of Climate* 28.12, pp. 4794–4819. DOI: 10.1175/JCLI-D-14-00656.1.

Stevens, B. et al. (2013). "Atmospheric component of the MPI-M Earth System Model: ECHAM6." *Journal of Advances in Modeling Earth Systems* 5.2, pp. 146–172. DOI: 10.1002/jame.20015.

Storch, H. von and F. W. Zwiers (2001). *Statistical Analysis in Climate Research*. Cambridge University Press.

Storch, J.-S. von, C. Eden, I. Fast, H. Haak, D. Hernández-Deckers, E. Maier-Reimer, J. Marotzke, and D. Stammer (2012). "An Estimate of the Lorenz Energy Cycle for the World Ocean Based on the 1/10 degrees STORM/NCEP Simulation." *Journal of physical oceanography* 42.12, pp. 2185–2205. DOI: 10.1175/JPO-D-12-079.1.

Suárez-Gutiérrez, L., C. Li, P. W. Thorne, and J. Marotzke (2017). "Internal variability in simulated and observed tropical tropospheric

temperature trends." *Geophysical Research Letters* 44.11, pp. 5709–5719. DOI: 10.1002/2017GL073798.

Taylor, K. E., R. J. Stouffer, and G. A. Meehl (2012). "An Overview of CMIP5 and the Experiment Design." *Bulletin of the American Meteorological Society* 93.4, pp. 485–498. DOI: 10.1175/BAMS-D-11-00094.1.

Thompson, D. W. J., E. A. Barnes, C. Deser, W. E. Foust, and A. S. Phillips (2015). "Quantifying the Role of Internal Climate Variability in Future Climate Trends." *Journal of Climate* 28.16, pp. 6443–6456. DOI: 10.1175/JCLI-D-14-00830.1.

Thorncroft, C. D., H. Nguyen, C. Zhang, and P. Peyrillé (2011). "Annual cycle of the West African monsoon: regional circulations and associated water vapour transport." *Quarterly Journal of the Royal Meteorological Society* 137.654, pp. 129–147. DOI: 10.1002/qj.728.

Toniazzo, T. and S. Woolnough (2014). "Development of warm SST errors in the southern tropical Atlantic in CMIP5 decadal hindcasts." *Climate Dynamics* 43.11, pp. 2889–2913. DOI: 10.1007/s00382-013-1691-2.

Vanniere, B., E. Guilyardi, T. Toniazzo, G. Madec, and S. Woolnough (2014). "A systematic approach to identify the sources of tropical SST errors in coupled models using the adjustment of initialised experiments." *Climate Dynamics* 43.7-8, pp. 2261–2282. DOI: 10.1007/s00382-014-2051-6.

Wahl, S., M. Latif, W. Park, and N. Keenlyside (2011). "On the Tropical Atlantic SST warm bias in the Kiel Climate Model." *Climate Dynamics* 36.5-6, pp. 891–906. DOI: 10.1007/s00382-009-0690-9.

Washington, W. M. and C. L. Parkinson (2005). *An Introduction to Three-dimensional Climate Modeling*. University Science Books.

Xu, Z., P. Chang, I. Richter, W. Kim, and G. Tang (2014). "Diagnosing southeast tropical Atlantic SST and ocean circulation biases in the CMIP5 ensemble." *Climate Dynamics* 43.11, pp. 3123–3145. DOI: 10.1007/s00382-014-2247-9.

Yan, X., R. Zhang, and T. R. Knutson (2018). "Underestimated AMOC Variability and Implications for AMV and Predictability in CMIP Models." *Geophysical Research Letters* 45.9, pp. 4319–4328. DOI: 10.1029/2018GL077378.

Zelle, H., G. Jan van Oldenborgh, G. Burgers, and H. Dijkstra (2005). "El Niño and Greenhouse Warming: Results from Ensemble Simulations with the NCAR CCSM." *Journal of Climate* 18.22, pp. 4669–4683. DOI: 10.1175/JCLI3574.1.

VERSICHERUNG AN EIDES STATT
Declaration of oath

Hiermit versichere ich an Eides statt, dass ich die vorliegende Dissertation mit dem Titel: Internal Variability In a Changing Climate – a Large Ensemble Perspective on Tropical Atlantic Rainfall selbstständig verfasst und keine anderen als die angegebenen Hilfsmittel – insbesondere keine im Quellenverzeichnis nicht benannten Internet-Quellen – benutzt habe. Alle Stellen, die wörtlich oder sinngemäß aus Veröffentlichungen entnommen wurden, sind als solche kenntlich gemacht. Ich versichere weiterhin, dass ich die Dissertation oder Teile davon vorher weder im In- noch im Ausland in einem anderen Prüfungsverfahren eingereicht habe und die eingereichte schriftliche Fassung der auf dem elektronischen Speichermedium entspricht.

Hamburg, den 7.11.2018

Sebastian Milinski

Hinweis / Reference

Die gesamten Veröffentlichungen in der Publikationsreihe des MPI-M
„Berichte zur Erdsystemforschung / Reports on Earth System Science“,
ISSN 1614-1199

sind über die Internetseiten des Max-Planck-Instituts für Meteorologie erhältlich:
http://www.mpimet.mpg.de/wissenschaft/publikationen.html

All the publications in the series of the MPI -M
„Berichte zur Erdsystemforschung / Reports on Earth System Science“,
ISSN 1614-1199

are available on the website of the Max Planck Institute for Meteorology:
http://www.mpimet.mpg.de/wissenschaft/publikationen.html